廣東技術師範大學
Guangdong Polytechnic Normal University

国家级一流本科专业建设点信息采集表

所在高校基本情况

支撑材料

其 它

目 录

国家知识产权局

STATE INTELLECTUAL PROPERTY OFFICE OF THE PEOPLE'S REPUBLIC OF CHINA

中国专利优秀奖

名　称　实现数字版权加密的视频多播传输的方法

专利号　ZL201310074499.6

发明人　赵慧民　孟　鹏　林智勇　陈荣军　李　军　朱　立　陈小玲

申长雨

中华人民共和国国家知识产权局局长

中华人民共和国国家知识产权局

2017 年 12 月

广东省科学技术奖励

证　书

为表彰广东省科学技术奖获得者，

特颁发此证书。

项目名称：海量视频内容快速检索与深度分析的关键技术及其应用

奖励等级：一等奖

获 奖 者：林智勇

粤府证：〔2016〕0812号

项目编号：B07-0-1-01-R05

广东省人民政府

二〇一六年一月

广东省科学技术奖励

证　书

为表彰广东省科学技术奖获得者，

特颁发此证书。

项目名称：身份认证压缩感知技术及其在智能安防工程领域的产业化

奖励等级：三等奖

获 奖 者：赵慧民

证书编号：〔2018〕0925号

项目编号：D07-3-08-301

月

Chinese Conference on Pattern Recognition and Computer Vision (PRCV)
PRCV 2018: Pattern Recognition and Computer Vision pp 221-232 | Cite as

Discriminative Visual Tracking Using Multi-feature and Adaptive Dictionary Learning

Authors Authors and affiliations

Penggen Zheng, Jin Zhan, Huimin Zhao ✉, Jujian Lv

Conference paper
First Online: 02 November 2018

Part of the Lecture Notes in Computer Science book series (LNCS, volume 11259)

Abstract

We propose a novel tracking method using color features and texture features to obtain accurate target appearance model. Each feature dictionary is independent and learned by labeled consistent K-SVD. In the subsequent frames, we exploit the maximum similarity of sparse features by the minimal reconstruction error criterion to locate the best tracking result. When a significant change occurs, we propose an adaptive dictionary learning which update the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.2966858, IEEE Transactions on Multimedia

Dynamic Objectives Learning for Facial Expression Recognition

Guihua Wen, Tianyuan Chang, Huihui Li*, Lijun Jiang
School of Computer Science, Guangdong Polytechnic Normal University
Tianhe, Guangzhou, Guangdong, China

***Abstract*—Facial expression recognition has been widely used to solve the problems such as lie detection and human-machine interaction. However, due to the difficulties to control the application environments, current methods have the lower recognition accuracy in practice. This paper proposes a new method for facial expression recognition by considering several aspects. First, human beings are easy to recognize some expressions, while difficult to recognize others. Inspired by this intuition, a new loss function is proposed to enlarge the distances between samples from easily confused categories. Second, human learning is divided into many stages, and the learning objective of each stage is different. Thus, dynamic objectives learning is proposed, where each objective at different stage is defined by the corresponding loss function. In order to better realize the above ideas, a new deep neural network for facial expression recognition is proposed, which integrates the covariance pooling layer and residual network units into the deep convolution neural network so as to better perform dynamic objectives learning. The experimental results on the standard**

appearance features on each local region are extracted[12]. In local regions, the most salient region can be searched that has the higher discriminating ability[13]. On the other hand, geometrical features are also considered[14][15]. All these methods depend on the manually designed features. However, designing appropriate features is an extremely difficult task in machine learning.

B. Deep neural network methods

Deep neural network methods have achieved very good results in facial expression recognition. International Machine Learning Conference in 2013 launched the facial expression

Applied Intelligence (2019) 49:4319–4334
https://doi.org/10.1007/s10489-019-01491-8

Facial expression recognition sensing the complexity of testing samples

Tianyuan Chang[1] · Huihui Li[1] · Guihua Wen[2] · Yang Hu[1] · Jiajiong Ma[1]

Published online: 14 June 2019

Abstract
Facial expression recognition has always been a challenging issue due to the inconsistencies in the complexity of samples and variability of between expression categories. Many facial expression recognition methods train a classification model and then use this model to identify all test samples, without considering the complexity of each test sample. They are inconsistent with human cognition laws such as the principle of simplicity, so that they are easily under-learned and then are difficult to identify test samples correctly. Hence, this paper proposed a new facial expression recognition method sensing the complexity of test samples, which can nicely solve the problem of the inconsistent distribution of samples complexity. It firstly divided the training data into the hard subset and the easy subset for classification according to the complexity of samples for expression recognition. Subsequently, these two subsets are applied to train two classifiers. Instead of using the same classifier to predict all test samples, our method assigned each test sample to the corresponding classifier based on the complexity of the test sample. The experimental results demonstrated the effectiveness of the proposed method and obtained a significant improvements of the recognition performance on benchmark datasets.

Keywords Facial expression recognition · Sample complexity · Convolutional neural network · Gestalt principle

1 Introduction

FACIAL expression recognition (FER) has a wide range of research prospects in human computer interaction and affective computing, including polygraph detection, intelligent security, entertainment, Internet education, and intelligent medical treatment [1]. As we all known, facial expression is a major way of expressing human emotions. Hence, the main task in determining emotion is how to automatically, reliably, and efficiently recognize the information conveyed by facial expressions. In FER research, Ekman and Friensen first proposed the Facial Action Coding System (FACS) [2]. The six basic categories of expressions (surprise, sadness, disgust, anger, happiness, and fear) are defined in FACS, and are com-

Published: 14 June 2019

Facial expression recognition sensing the complexity of testing samples

Tianyuan Chang, Huihui Li ✉, Guihua Wen, Yang Hu & Jiajiong Ma

Applied Intelligence **49**, 4319–4334 (2019) | Cite this article

A Correction to this article was published on 27 July 2020

This article has been updated

Abstract

Facial expression recognition has always been a challenging issue due to the inconsistencies in the complexity of samples and variability of between expression categories. Many facial expression recognition methods train a classification model and then use this model to identify all test samples, without considering the complexity of each test sample. They are inconsistent with human cognition laws such as the principle of simplicity, so that they are

Research Article

INTERNATIONAL JOURNAL OF ADVANCED ROBOTIC SYSTEMS

International Journal of Advanced Robotic Systems
November-December 2019: 1–13

DOI: 10.1177/1729881419890155
journals.sagepub.com/home/arx
SAGE

Incremental learning-based visual tracking with weighted discriminative dictionaries

Penggen Zheng[1], Huimin Zhao[1], Jin Zhan[1], Yijun Yan[2], Jinchang Ren[2], Jujian Lv[1] and Zhihui Huang[1]

Abstract

Existing sparse representation-based visual tracking methods detect the target positions by minimizing the reconstruction error. However, due to complex background, illumination change, and occlusion problems, these methods are difficult to locate the target properly. In this article, we propose a novel visual tracking method based on weighted discriminative dictionaries and a pyramidal feature selection strategy. First, we utilize color features and texture features of the training samples to obtain multiple discriminative dictionaries. Then, we use the position information of those samples to assign weights to the base vectors in dictionaries. For robust visual tracking, we propose a pyramidal sparse feature selection strategy where the weights of base vectors and reconstruction errors in different feature are integrated together to get the best target regions. At the same time, we measure feature reliability to dynamically adjust the weights of different features. In addition, we introduce a scenario-aware mechanism and an incremental dictionary update method based on noise energy analysis. Comparison experiments show that the proposed algorithm outperforms several state-of-the-art methods, and useful quantitative and qualitative analyses are also carried out.

Keywords

Visual tracking, similarity weights, sparse representation, incremental update, weighted dictionary

Date received: 1 February 2019; accepted: 25 October 2019

Topic: Vision Systems
Topic Editor: Antonio Fernandez-Caballero
Associate Editor: Kaifu Yang

Introduction

As a subtask of computer vision, visual target tracking has always drawn many attentions for decades, and many advanced methods have been explored. However, complex situations such as occlusions, target deformation, rotation, scale changes, and cluttered background, and so on, make visual target tracking still a challenging task and the existing methods cannot always track the targets precisely. The current trackers can be typically divided into two types, that is, generative methods[1–5] and discriminative methods.[6–17] They usually sample a set around the target object to describe the appearance characteristics, and search for

[1] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China
[2] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK

Corresponding author:
Jin Zhan, School of Computer Science, Guangdong Polytechnic Normal University, Tianhe District Zhongshan Road West, No. 293 Guangzhou, Guangdong 510665, China.
Email: gszhanjin@gpnu.edu.cn

candidate targets by maximizing the similarity or find the decision boundaries of the target and the background.

To get the satisfied tracking performance, two key issues need to be addressed. First, since the appearance of the target changes frame by frame throughout the video sequence, the most discriminating samples in the current frame may not last for a long time and tend to result in a model overfitting. So, the improvement of long-term tracking performance is an important issue. Second, Unpredictable target deformation and background clutter in the sampling region cause a negative impact on the selection of candidate samples. Thus, the elimination of these obstacles to advance tracking performance in the case of small target samples is also an important issue.

To address both issues, sparse representation-based tracking solutions have been proposed, such as L1 tracker.[5] Because of insensitivity to the target noise, this kind of methods has a strong tracking robustness when target deformation occurs. However, single-feature and the initial discriminative dictionary do not satisfy complex tracking scenarios. Moreover, the object localization under the frequent online updating often brings drift-away problems as some negative samples are mis-tracked. These problems remain difficult in the literature of sparse representation-based trackers. Hence, a natural question is how we can augment positive samples in the feature space to capture target appearance variations in the temporal domain.

In this work, we take advantage of the recent progress in discriminative dictionary learning method label consistent K-SVD (LC-KSVD)[18,19] to facilitate the dictionary learning and to propose a novel tracking method with weighted dictionaries incremental learning and pyramidal feature selection strategy. In summary, this work has the following main steps. Firstly, we model the discriminative dictionaries from positive and negative samples based on two feature descriptors, where different features correspond to different dictionaries. Secondly, according to the center distance from the training samples to the target, we assign Gaussian weights for each basis vector in different feature dictionaries, which are used to measure the similarity of spatial structure to improve the accuracy of sparse feature selection. Finally, we select the best sample region by similarity measurement and fusion of the multiple features reconstruction error of candidate samples.

The article is organized as follows. We introduce the research background in the "Introduction" section and review the related work in the "Related work" section. Afterwards, the "Proposed method" section describes the proposed method in detail, including dictionary representation and construction, incremental dictionary updating, and adaptive feature fusion strategy. The experiments are given in "Experimental results and comparison." We conclude the article and discuss future work in the "Conclusion and future work" section.

Related work

In this section, we briefly review the relevant literature of object tracking algorithms in recent year, including deep learning-based tracking methods[11–17,20,21–23,24] and sparse representation-based tracking method.[1,3,5,6,25–28,30,31]

The main advantage of deep learning-based tracking methods lies in their powerful characterization of depth features. It brings a new research direction for solving various challenges in visual tracking. Wang and Yeung[20] proposed deep learning tracker and performed unsupervised off-line depth pretraining on large-scale natural image data sets. The idea of transfer learning reduces the requirement of training samples and improves the performance of the tracking algorithm. Then, they propose structured output-deep learning tracker[11] and use convolutional neural network (CNN) model to solve the sensitivity of model updating. Qi et al.[12] proposed a novel CNN-based tracking method, which considers the features from all CNN layers and hedge these features into a single stronger one. Furthermore, they propose a hedging deep feature-based tracking framework[13] which use correlation filters to feature maps of each CNN layer to construct a weak tracker and design a Siamese network to define the loss of each weak tracker. The tracker achieves favorable performance on challenging image sequences.

To solve the imbalance between positive and negative samples in video tracking, Zhang et al.[14] proposed an attribute-based CNN with multiple branches, where each branch is responsible for classifying the target under a specific attribute. The tracker reduces the appearance diversity of the target under each attribute and thus requires fewer data to train the model. Qi et al.[15] proposed to integrate the point-to-set/image-to-imageSet distance metric learning (DML) into visual tracking. The point-to-set DML is conducted on CNN features of the training data, and the tracking result is located by the minimal distance to the target template. Because the methods based on matching tracking cannot deal with the problem of target rotation in the plane very well, Zhong et al.[16] proposed a hierarchical tracker that learns to move and track by a coarse-to-fine verification. The coarse level utilizes a recurrent CNN-based deep Q-network to learn data-driven searching policies. The idea of learning target position from coarse to fine is helpful to deal with target scale change and improve the accuracy of tracking target border. The authors also apply this idea to multi-person tracking and propose a deep alignment network-based multi-person tracking method[17] with occlusion and motion reasoning which achieves good performance. Wang et al.[24] proposed a deep learning-based hybrid

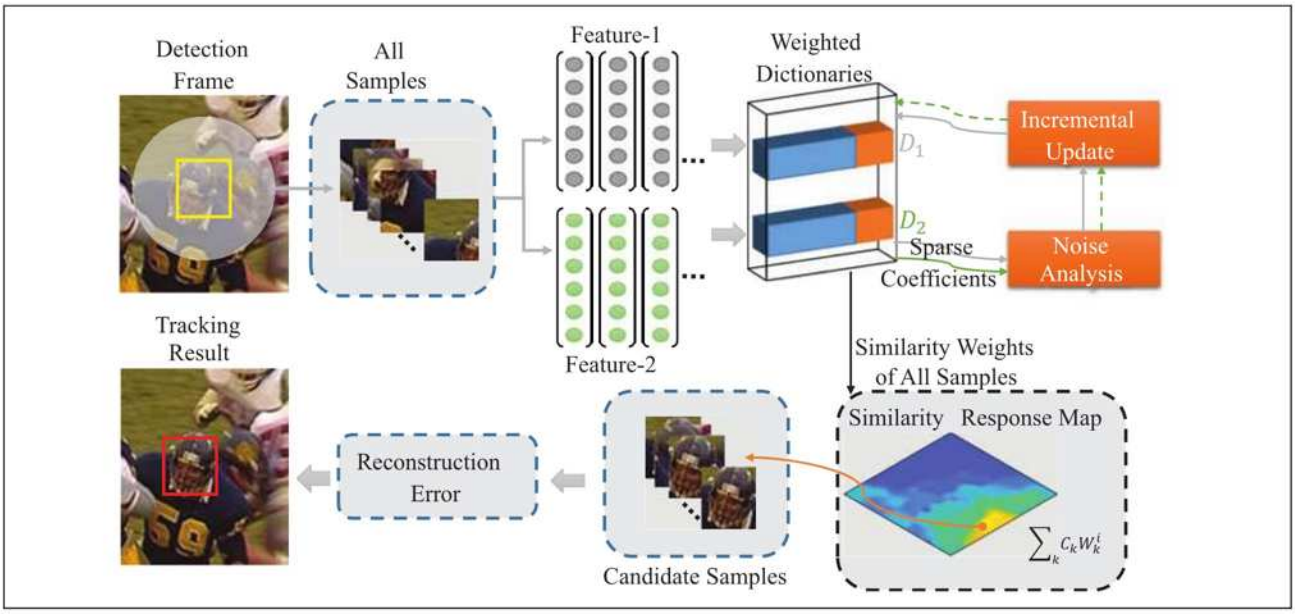


Figure 1. The main tracking process of the proposed approach.

spatiotemporal saliency feature extraction framework for saliency detection from video footages.

Sparse representation-based tracking methods show strong robustness in some tracking scenarios. Therefore, many visual tracking methods[5,25,27,31,32] based on sparse representations have been proposed. Local sparse representations are widely used in visual tracking.[28,29,33] Zhang et al.[30] summarized and evaluated some classical tracking methods based on sparse representation. The process of sparse representation-based trackers can be roughly divided into two stages. The first stage acquires a sparse sample set around the target, and the second stage uses a classifier to classify each sample as a target or background. However, the positive samples obtained from the first frame of video are far from meeting the requirement of label data volume in classifier training, and the positive and negative samples are imbalanced greatly, which makes it impossible to capture the rich appearance changes of the target. These limitations are also reflected in some deep learning-based trackers[21–23] that use this two-stage framework.

In the target tracking process, a good model update strategy can improve the tracking effect and tracking ability. Lu et al.[26] used incremental subspace learning methods to reconstruct a new template and then utilized it to replace the old one. However, the updated base vector will gradually degrade in the scene where noise or occlusion exists. In addition, Mei and Ling[5] replaced the least important template with the current template based on the frequency of use of the dictionary template. Han et al.[27] updated the dictionary template in a random replacement manner.

The combination of multiple features enhances the characterization capabilities of the model and is applied to many different classification tasks. From the perspective of visual attention saliency, Yan et al.[34,35] used Gestalt rule to guide the saliency detection by characterizing human visual system (HVS) features and forming targets and proposed a method to cognitively detect and track salient objects from videos by combining red-green-blue (RGB) image and thermal image. The proposed fusion-based approach can successfully detect and track multiple human objects in most scenes regardless of any light change or occlusion problem. Lan et al.[25] used an unreliable feature detection method to detect unreliable features. However, the representation of reliable features is still suppressed by the joint sparse framework, and different features are limited to similar sparse patterns. Mai and Ling[5] fused multiple features for appearance modeling and detect the outlier particles. The same sparse pattern is used for all features of the non-outlier particles.

In this article, we propose a novel multifeatures dictionary-based sparse tracking method, where a specific feature dictionary is built upon hybrid features with the ability of independently maintaining. Then an incremental dictionary update strategy is proposed to reduce the redundancy of sparse dictionaries while increasing the diversity of positive samples. The output of these dictionaries responses in a different sparse pattern for the final comprehensive decision during the tracking process.

Proposed method

In this section, the proposed method including three modules is introduced. The main framework of our method is shown in Figure 1. We maintain two sets of samples (positives and negatives) to construct weighted feature

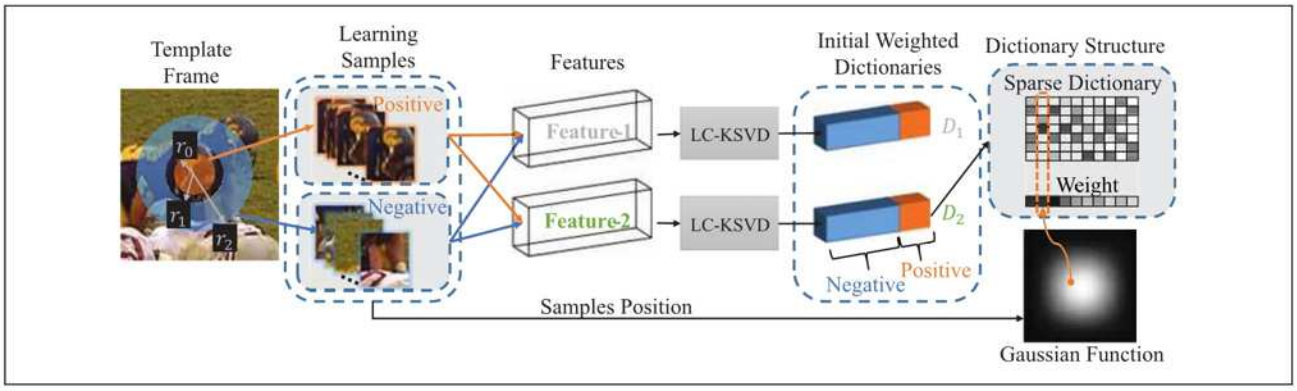


Figure 2. Initial dictionary learning and the structure of multiple dictionaries.

dictionaries. In the tracking process, the samples are sparsely decomposed by the weighted dictionaries, and the weights of the samples can be obtained and used to select candidate samples. By comparing the reconstruction errors of these candidate samples, we can select the most similar sample as the tracking result.

Dictionary representation and construction

In sparse representation theory, dictionary is composed of super-complete base vectors to obtain a more concise representation of the appearance of the target. For this purpose, three types of sets, that is, the positives T, the backgrounds B, and the noise L are integrated together. The initial dictionary D of the samples at the first frame can be represented as $D = [D^T, D^B, D^L]$, where D^T, D^B, and D^L are the sets of T, B, and L, respectively. In the tracking process, a candidate sample y can be represented by the sparse representation (equation (1))

$$y \approx D\gamma = \left[D^T, D^B, D^L\right] \begin{bmatrix} z \\ v \\ e \end{bmatrix} \tag{1}$$

where D is the discriminative dictionary, z is the target coefficient, v is the background coefficient, e is the noise coefficient, and γ is the sparse coding. In this article, the LC-KSVD[18] method is used to unify dictionary learning and classification labeling.

Figure 2 shows the construction process of the initial dictionary. The center of the initial target is set as the center of the circle, pixels in the range of radius r_0 are sampled as positive samples, and pixels in the range of radius between r_1 and r_2 are dense sampled to obtain negative samples which contain the background context around the target.

For the positive and negative samples sampled in the first frame, we extract two kinds of features to form two initial dictionaries respectively. After that, we utilize the correspondence between the sample template and the dictionary base vector and assign the Gaussian weight to each base vector by calculating the center distance $d(i)$ between sample templates and the target center. The weight of the ith base vector is defined as follows

$$W(i) = \exp(-d^2(i)/2\sigma^2) \tag{2}$$

where α is the standard deviation of normal distribution. This weight reflects the similarity between the target and the samples. Finally, we get the weighted discriminative dictionaries, and each discriminative dictionary corresponds to a weight table.

Incremental dictionary updating

In many existing tracking methods, the appearance model of target is often updated to reduce the negative impact of target and background changes in the frames. In the sparse decomposition, the coefficient γ of sample contains the most representative information, where the noise factor indicates the situation of target occlusion and tracking drift to some extent. To this end, an incremental dictionary updating strategy is proposed to measure the change of target or scene by analyzing the noise energy u (the sum of the noise coefficients e). The larger the noise energy is, the more significant the deformation of the target or the greater the change of the scene causes.

In the frame t, the average noise energy expression for all samples can be represented as $\overline{u_{k,t}} = \frac{\sum_i u_k^i}{n}$, where u_k^i is the noise energy of the ith sample, n is the number of all samples, and k is the feature tag ($k = 1$ denotes color feature and $k = 2$ denotes noncolor feature). We define a dynamic threshold to analyze changes of the target and scenes, which are defined as follows

$$P\{U_k > x_k^\alpha\} = \alpha \tag{3}$$

where x_k^α is the upper quantile of set U_k (all $\overline{u_{k,t}}$ from the first frame to current frame) and reflects the overall level of noise energy during the tracking process. If the tracked average noise energy $\overline{u_k^T}$ exceeds the threshold, it indicates that the background changes too much in the

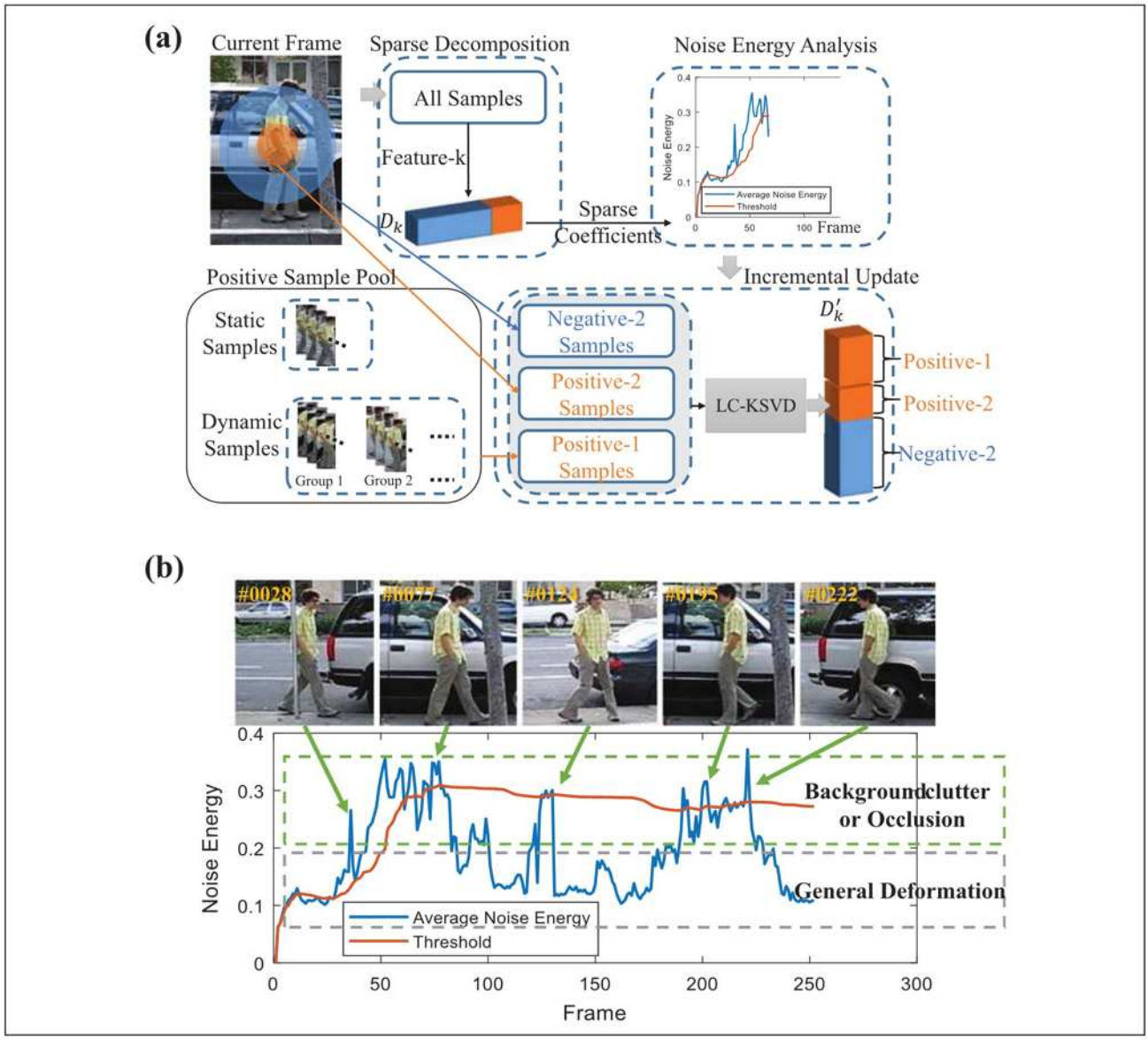


Figure 3. Discriminative dictionary incremental update (a) and noise analysis (b).

current frame. Meanwhile, we also set the minimum update interval m to make the tracking process more efficient and the interval between two updates must be more than m frames.

In scene detection, we use the dynamic threshold of noise energy to judge the intensity of the scene change. Based on the target noise energy and the average noise energy, we can determine whether to perform a dictionary update. If the update condition is met, we use the samples of the first frame and the samples of the detected frame to obtain a new weighted dictionary D_k'. The new dictionary will be used for the next frame tracking task.

The incremental dictionary update trigger mechanism is shown in Figure 3(a). We divide the positive samples into two categories: static samples and dynamic samples. The samples obtained in the first frame are static samples, and the samples obtained in the trigger update mechanism are dynamic samples. When the number of positive samples is larger than that of current negative samples, we use a new positive sample set to randomly replace one group of the dynamic samples to reduce the impact of sample imbalance and maintain the efficiency of dictionary learning.

Figure 3(b) shows the changes of threshold curve and noise energy curve in sequence *David 3*. Five frames with large changes of target pose and background are selected as examples for illustration. It can be seen that the selected examples occur when the noise energy value is higher than the threshold value. Hence, our updating strategy can detect and reduce the impact of the background change in real time through the analysis of noise energy for better tracking performance.

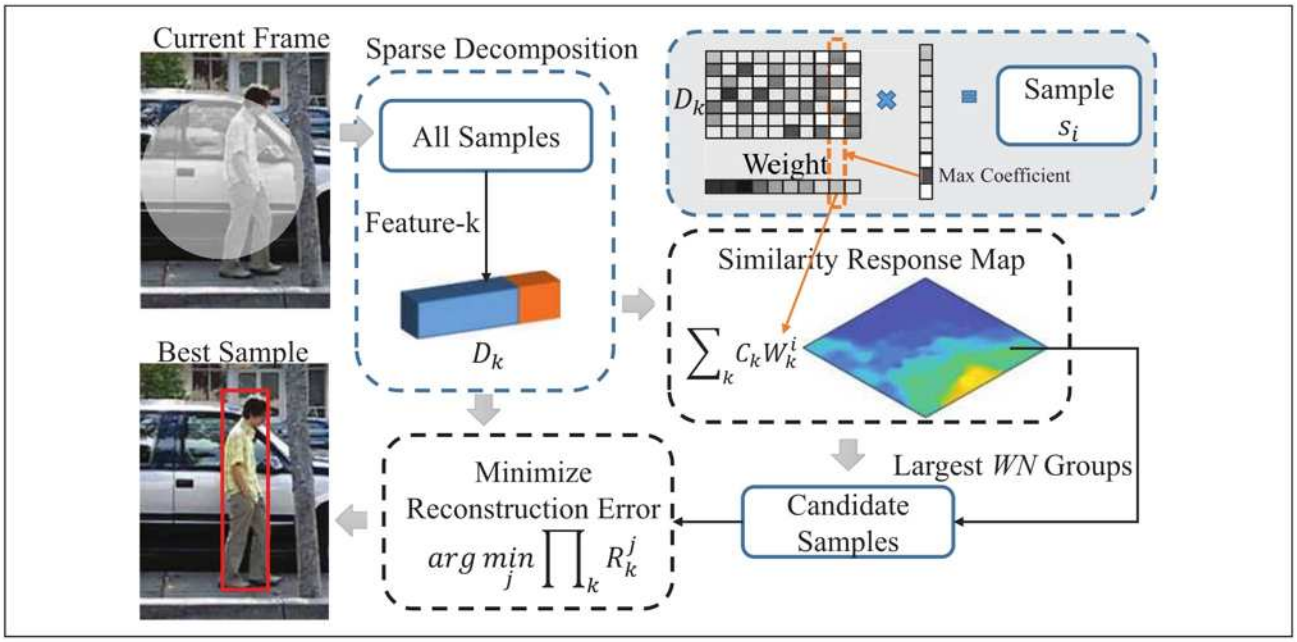


Figure 4. Feature selection process.

Adaptive feature fusion strategy

In this section, we introduce the pyramid feature selection strategy to locate the target tracking position, as shown in Figure 4. We use a pyramidal selection strategy in the feature selection. First, we select WN groups of samples with the largest similarity weights as candidate samples CS_j (j is the tag of the candidate samples). The sample similarity weight can be obtained in the sparse decomposition process. Then, we compare the comprehensive reconstruction error of the candidate samples to select the best sample as the tracking result.

In the current frame, all samples S_i ($i = 1, 2, \ldots, n$) are sparsely resolved by different feature dictionaries D_k ($k = 1, 2$) to obtain sparse coefficients γ_k^i, where k is a feature tag. The similarity weights W_k^i and reconstruction errors R_k^j are normalized into [0,1] to eliminate the inconsistency of different feature weights. Each sample S_i has k feature sparse coefficients. The weight values corresponding to the maximum values of the k feature sparse coefficients are used as the similarity weights W_k^i ($k = 1, 2$). Therefore, we fuse these two weights into a composite weight w^i, which is defined as follows

$$tW^i = \sum_k C_k W_k^i \tag{4}$$

$$C_k = 1 - \frac{\overline{u_k^T}/x_k^\alpha}{\sum_{k=1}^{2} \overline{u_k^T}/x_k^\alpha} \tag{5}$$

In equation (4), we set the dynamic feature weight parameters C_k based on the feature reliability. Then we select a few candidate samples CS_j which have the largest synthetic weights among all samples and the maximum value of synthetic weights is denoted as WN. When the noise energy is relatively large, the feature weight C_k is relatively small. The definition of C_k is shown in equation (5). $\overline{u_k^T}$ is the k-feature average noise energy of the current frame, and x_k^α is the k-feature noise energy threshold defined in equation (3).

Then we use the synthetic reconstruction error to select the best sample from candidate samples. The expression of the synthetic reconstruction error is as follows

$$R^j = \prod_k R_k^j \tag{6}$$

where R_k^j represents the reconstruction error of the sample s_j in *k-feature*, j is the label of the candidate samples. Finally, we select the one with the smallest synthetic reconstruction error in the candidate samples as our tracking result.

Experimental results and comparison

In this section, the public sequences of VOT2017[36] and OTB100[37] are used for the parameter setting and tracking performance evaluation of our method, respectively. Firstly, we experiment with eight RGB sequences of VOT2017,[36] analyze the parameter settings in the feature selection, and discuss the optimal combination of features. Then all 74 RGB sequences on the OTB100[37] are used for tracking performance evaluation. The experiment tracking results of other benchmarking methods are primarily derived from publicly available results data on the author's homepage and OTB100[37] homepage. The computer environment used by our method is Intel (R) Core (TM) i3-3.7 GHz, RAM-12 GB, and MATLAB R2017a.

Table 1. The average CLEs for different dual feature combinations.[a]

Type	Feature (s)	*ball1*	*blanket*	*butterfly*	*crossing*	*godfather*	*pedestrian1*	*sheep*	*wiper*	Average
Color feature	HSV	5.69	10.42	21.22	47.57	15.06	41.18	43.66	24.46	26.16
	RGB	43.8	17.85	27.21	40.75	16.89	21.71	35.46	216.36	52.5
	LAB	4.91	12.76	19.74	45.48	10.35	13.41	43.11	163.71	39.18
Noncolor feature	Haar-like	12.4	49.04	76.42	37.05	9.27	96.55	**6.48**	25.68	39.11
	HOG	62.21	40.11	41.9	19.67	19.56	66.26	93.36	70.48	51.69
Fusion of color feature and noncolor feature	HOG + HSV	7.15	**9.54**	22.15	**17.91**	10.13	12.08	76.79	**15.61**	21.42
	HOG + RGB	29.87	9.85	30.13	19.53	7.95	13.54	22.24	43.33	22.06
	HOG + LAB	5.28	9.75	19.77	30.41	8.65	**10.39**	41.28	81.83	25.92
	Haar-like + HSV	3.69	11.05	27.97	42.84	7.14	23.26	11.58	19.05	**18.32**
	Haar-like + RGB	4.07	16.31	44.41	28.99	**6.89**	22.14	12.38	36.81	21.5
	Haar-like + LAB	3.23	14.92	27.4	35.49	7.52	15.57	39.81	29.08	21.63

CLE: center location error.
[a]Bold data represent the best results of single video tasks.

Table 2. The average CLEs with different WN values.[a]

Sequences	WN = 1	WN = 2	WN = 3	WN = 4	WN = 5	WN = 6	WN = 7
ball1	5.066321	5.051257	4.621554	3.477649	4.148773	4.210412	3.893678
blanket	18.31654	10.40168	10.52438	11.24985	11.77508	15.84256	15.01424
butterfly	27.08312	28.87616	28.02551	30.88001	30.57644	30.84231	30.94732
crossing	45.38556	46.61894	47.04421	37.7282	42.01624	37.27262	35.52469
godfather	9.649543	7.13208	8.276668	7.635623	12.19341	7.387032	7.490803
pedestrian1	19.45411	20.88936	21.5531	19.13769	34.73961	15.42067	29.3694
sheep	35.82426	30.52743	22.04572	25.95596	27.32844	45.41071	22.90445
wiper	20.25632	24.19629	22.58765	20.48694	25.05795	26.48748	19.63991
Average	22.62947	21.71165	20.58485	19.56899	23.47949	22.85922	20.59806

CLE: center location error.
[a]*Source*: The parameter setting of the variable WN is from 1 to 7.

Implementation details and analysis

The method of this article adopts uniform parameter settings. The number of all samples obtained by Gaussian sampling during the tracking process is 500 and the sampling radius is 25. The sampling parameter of the training sample is set to: $r_0 = 4$, $r_1 = 7$, $r_2 = 15$. The Haar-like[38] feature dimension is set to 150, and the histogram bin of a single color channel is set to 36. Correspondingly, the color feature dimension of an RGB frame is set to 108. The update time interval must be greater than $m = 6$ frames, and the noise energy threshold parameter is $\alpha = 0.2$.

Feature selection. Two types of features (noncolor features and color features) are used in our proposed model. In this section, the performance of different feature fusion strategy on eight RGB sequences (*ball1*, *blanket*, *butterfly*, *crossing*, *godfather*, *pedestrian1*, *sheep*, and *wiper*) in VOT2017[32] is investigated and useful analysis is also carried out.

Table 1 shows the performance of different feature fusion strategies in terms of average center location errors (CLEs). The CLE is the Euclidean distance between the tracking result and the standard target position. In general, dual feature fusion always outperforms single feature. Feature CIE L*a*b* (LAB) performs poorly in combination with other non-color features. It is worth noting that histogram of orientation gradient (HOG) + hue-saturation-value (HSV) has the best performance in the sequences of *blanket*, *crossing*, and *wiper*, but the average performance is the second best which is 3.1 lower than the best one, that is, Haar-like + HSV. Therefore, Haar-like + HSV is selected as feature fusion strategy for our following experiments.

Candidate samples selection. In this section, we need to select a small number of candidate samples to narrow the scope of the target searching. These candidate samples are obtained by the composite similarity weights, where the optimal similarity weight values need to be determined. In this section, we discuss the influence of the maximum value of synthetic weights WN on the tracking effect. The experimental results are shown in Table 2.

In order to ensure the rationality of the experiment, we do not adopt the dictionary update strategy here. Based on the above experimental data, we can obtain the curve of CLE versus WN (Figure 5).

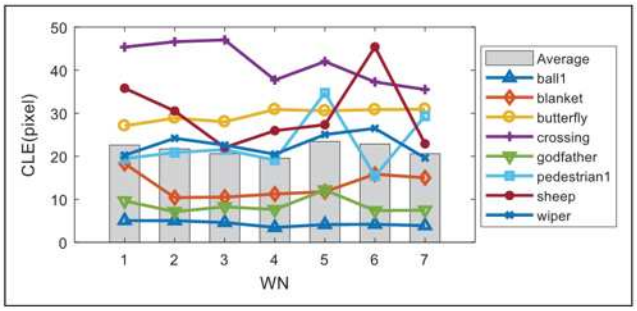


Figure 5. The average CLE variation for eight sequences. CLE: center location error.

Table 3. The distribution of 11 challenging attributes in the evaluating sequence set: IV, SV, OCC, DEF, MB, IPR, OPR, OV, BC, LR, and FM.

	IV	OPR	SV	OCC	DEF	MB	FM	IPR	OV	BC	LR
Frequency	32	47	49	42	37	26	32	34	11	24	8

IV: illumination variation; SV: scale variation; OCC: partial or full occlusion; DEF: non-rigid object shape deformation; MB: motion blur; IPR: in-plane rotation; OPR: out-of-plane rotation; OV: out of view; BC: background clutters; LR: low resolution; FM: fast motion.

In Figure 5, the broken line indicates the change in the effect of a single video tracking. The histogram shows the average tracking effect of all videos. As shown in Figure 5, the average value is significantly increased when WN is greater than 4 and the tracking results of some sequences are also significantly changed, such as *pedestrian1*, *sheep*, and so on. In the method evaluation experiment, we set WN to 3 in the experiment.

Experimental evaluation

In the performance evaluation section, we mainly compare the proposed method against eight state-of-the-art methods including adaptive local sparse appearance model-based tracker (ASLA[1]), incremental learning-based tracker (IVT[2]), L1 sparse tracker using APG (L1APG[3]), compress tracker (CT[6]), context tracker (CXT[7]), online robust image alignment tracker (ORIA[9]), online boosting tracker (OAB[8]), and tracking learning-detection tracker (TLD[10]). The qualitative and quantitative experimental results are carried out with a useful analysis. All 74 RGB sequences on OTB100[37] are used as evaluating sequence set, and the distribution of all challenging attributes in the evaluating sequence set is shown in Table 3.

Quantitative analysis. In this section, the tracking results based on precision plots and success plots are used to comprehensively evaluate the performance of different methods on OTB100.[37] The legend of precision plots shows the values at the error threshold of 20 pixels, and the legend of success plots show the area under curve (AUC) values. The overlap score is a measure of the overlap range of the tracking result and the ground truth tracking box, defined as $OS = intersection\ area/union\ area$, where intersection area and union area are the intersection and union of two regions, respectively.

Figure 6 shows the overall tracking precision plots and success plots of all nine methods on 74 RGB sequences of OTB100.[37] The precision score and success score of our approach are ranked first, higher than the second methods by 10.6% and 7.4%, respectively. As can be seen from the precision plots of one-pass evaluation (OPE), as the location error threshold increases, the precision of other trackers grows slowly, and our algorithm improves a lot. In the precision plots of OPE, the success score of our method is significantly higher than the other methods.

Table 4 shows the performance of our method and eight benchmarking methods in terms of success plots and AUC scores on different attributes. The average AUC value of our method, TLD and CXT trackers are top 3 on 11 attributes. TLD and CXT trackers perform well on attributes of fast motion (FM), motion blur (MB), out of view (OV), and low resolution (LR) due to dense sampling. ASLA tracker performs better on occlusion (OCC), scale variation (SV), and non-rigid object shape deformation (DEF) attributes by its local representation.

Figure 7 shows the ranking of success plots of all benchmarking methods on the 11 challenging attributes. On challenging attributes of SV, OCC, out-of-plane rotation (OPR), DEF, FM, MB), OV, in-plane rotation (IPR), and LR, the success plot of our method ranks the first. Despite the lack of a scale-changing mechanism, our method still has the best performance with 0.390 score on SV attribute. In similar methods, ASLA using local information also has a good score on SV attribute, but its score is lower than our method by 3%.

The AUC scores of our method are higher than the second method ASLA by 5.8% and 7.4% on the attributes OPR and DEF, respectively, which shows the effectiveness of our feature selection mechanism in the target appearance change. ASLA and TLD trackers use local information and have good scores on OCC attribute, which are 5.1% lower than our method. On attributes FM and MB, our method is 8.3% and 7.5% higher than the second method CXT, respectively. The ASLA tracker used local information and had the best results on background clutter (BC) and illumination variation (IV) attributes, and the success rate score of ASLA is 0.397 which is better than other similar methods.

Qualitative analysis. Figure 8 shows the tracking process of eight similar trackers and our method in the several RGB sequences. In Figure 8, our method has good tracking performance on the attribute of MB and FM. In sequences *Deer* and *BlurOwl*, although tracking drift sometimes

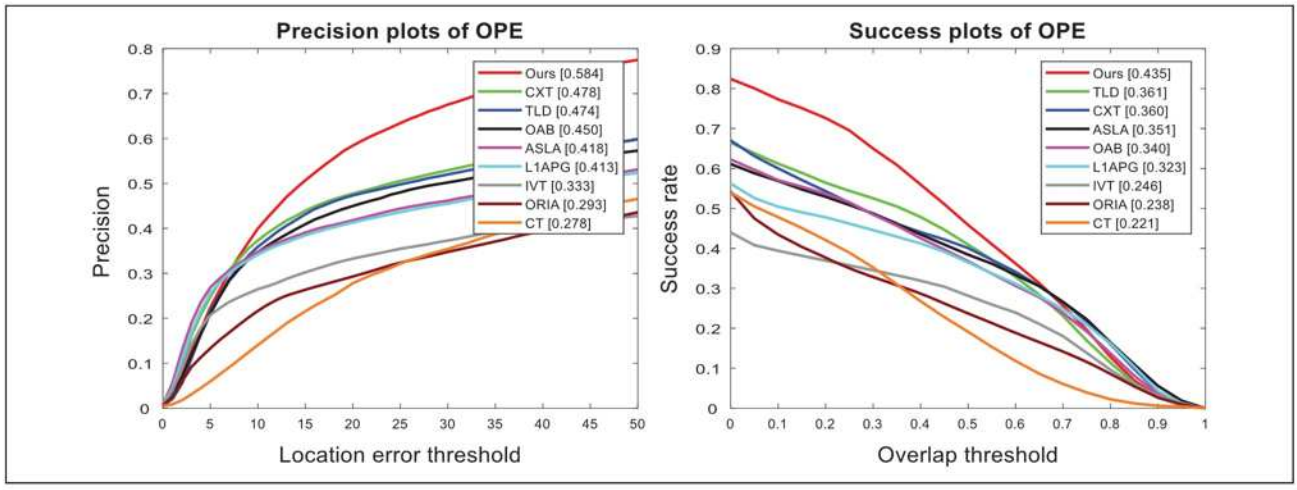


Figure 6. The comprehensive precision plots (left) and success plots (right) of comparison methods on 74 RGB sequences of OTB100.[37]

Table 4. The AUC value of all trackers in different attributes.[a]

	Ours	ASLA	IVT	OAB	L1APG	TLD	CT	ORIA	CXT
IV	**0.357**	**0.387**	0.263	0.278	0.324	0.343	0.215	0.268	**0.344**
OPR	**0.406**	**0.348**	0.232	0.290	0.264	**0.342**	0.236	0.252	0.328
SV	**0.390**	**0.360**	0.241	0.308	0.287	**0.348**	0.225	0.247	0.345
OCC	**0.408**	**0.357**	0.265	0.299	0.295	**0.357**	0.208	0.263	0.311
DEF	**0.387**	**0.313**	0.171	0.252	0.253	**0.292**	0.204	0.170	0.240
MB	**0.472**	0.213	0.188	0.363	0.322	**0.366**	0.180	0.171	**0.397**
FM	**0.452**	0.218	0.165	**0.366**	0.298	0.358	0.183	0.162	**0.369**
IPR	**0.409**	**0.356**	0.222	0.311	0.290	0.343	0.241	0.263	**0.373**
OV	**0.338**	0.264	0.190	0.217	0.201	**0.325**	0.185	0.153	**0.328**
BC	**0.380**	**0.397**	0.225	0.261	0.291	0.271	0.252	0.185	**0.293**
LR	**0.350**	0.325	0.274	0.301	0.334	**0.342**	0.208	0.229	**0.345**
Average	**0.395**	0.322	0.221	0.295	0.287	**0.335**	0.212	0.215	**0.334**

AUC: area under curve; ASLA: adaptive local sparse appearance model-based tracker; IVT: incremental learning-based tracker; OAB: online boosting tracker; L1APG: L1 sparse tracker using APG; TLD: tracking learning-detection tracker; CT: compress tracker; ORIA: online robust image alignment tracker; CXT: context tracker; IV: illumination variation; OPR: out-of-plane rotation; SV: scale variation; OCC: partial or full occlusion; DEF: non-rigid object shape deformation; MB: motion blur; FM: fast motion; IPR: in-plane rotation; OV: out of view; BC: background clutters; LR: low resolution.
[a]Bold data indicate the AUC scores are top three.

occurs, our approach can readjust the tracking position through the positive and negative templates when drifting is not severe. In general, OAB has a good tracking effect in these two videos, but it is prone to have tracking drift problems when the target moves fast, as shown by #0025 of *Deer* and #0390 of *BlurOwl*. CXT tracker can well recognize the target information in these sequences, but when the target blur and FM occur, the scale of the tracking will be abnormal.

Sequences *bolt*, *bolt2*, and *basketball* are typical of the target DEF. CT tracker is a tracking method based on compressed sensing and has good performance in target DEF, as shown in sequence *bolt2*. However, it appears that many tracking failures occur in sequences *deer* and *BlurOwl*, which indicates that CT tracker suffers from target FM easily. Our method performs well in the challenges of target DEF, but not in IV. As shown in #0700 of sequence *basketball*, our method shows significant tracking drift when there is a noticeable illumination change.

In sequence *David3*, most trackers suffer from OCC and BC, but our method can effectively deal with short-term occlusion of a large area because the adaptive dictionary update strategy minimizes occlusion interference. From the sequences *David3* and *couple* in Figure 6, we can see that OAB, CT, and the proposed approach have good tracking performance in the background changes. TLD has the problems of tracking drift and target lost. Both ORIA and CXT trackers are affected by small-range occlusion, which causes to tracking failure. In sequence *David3* #0146, a wide range of occlusions also leads to tracking failures of CT and OAB. Our method effectively identifies the target location in these cases and does a good job for the rest of tracking tasks.

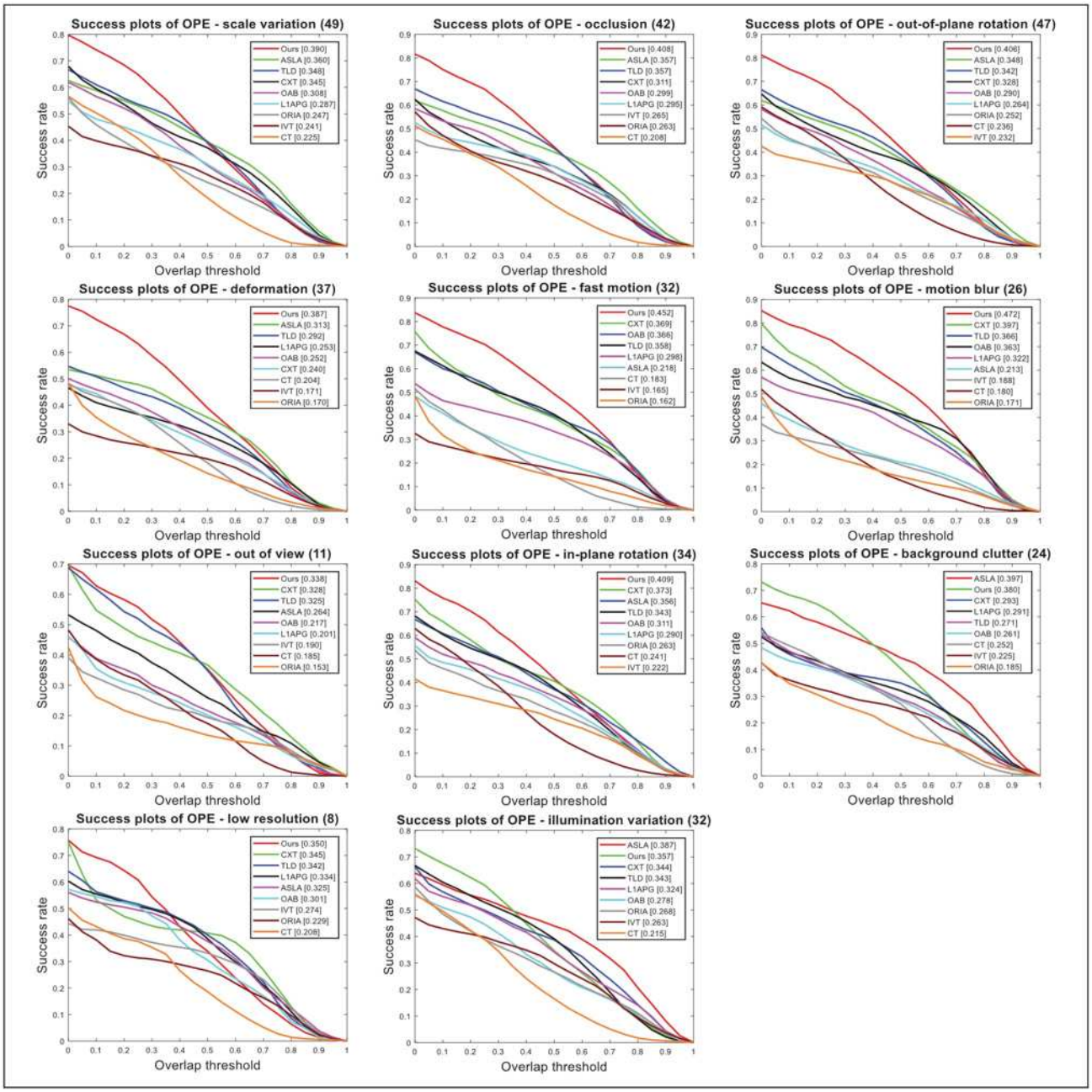


Figure 7. Success plots of 11 challenging attributes on all 74 RGB sequences of OTB100.[37]

In the last two sequences, *Lemming* and *DragonBaby* contain multiple challenge attributes such as SV, OCC, rotation (IPR or OPR), and OV. It can be seen in Figure 6 that the tracking drift is easily occurred when the target fast rotation, SV, and BC occur simultaneously. In #1010 and #1078 of sequence *Lemming*, TLD, CXT, and CT trackers have obvious tracking drift due to fast rotation, while OAB and our method do not suffer from that and perform better results. In #0084 and #0096 of sequence *DragonBaby*, our method performed well for target fast rotation and background interference. CXT tracker has tracking scale anomalies, and other methods have repeatedly experienced tracking drift and tracking failure. In sequence *human8* #0054 and #0070, most trackers have tracking failures when both illumination and scale changes occur. At the #0101 and #0126 frames of sequence *human8*, the true scale of the target is significantly smaller, and the result area selected by our method contains a large amount of background information. This situation makes the performance of our tracker unstable and prone to tracking failure.

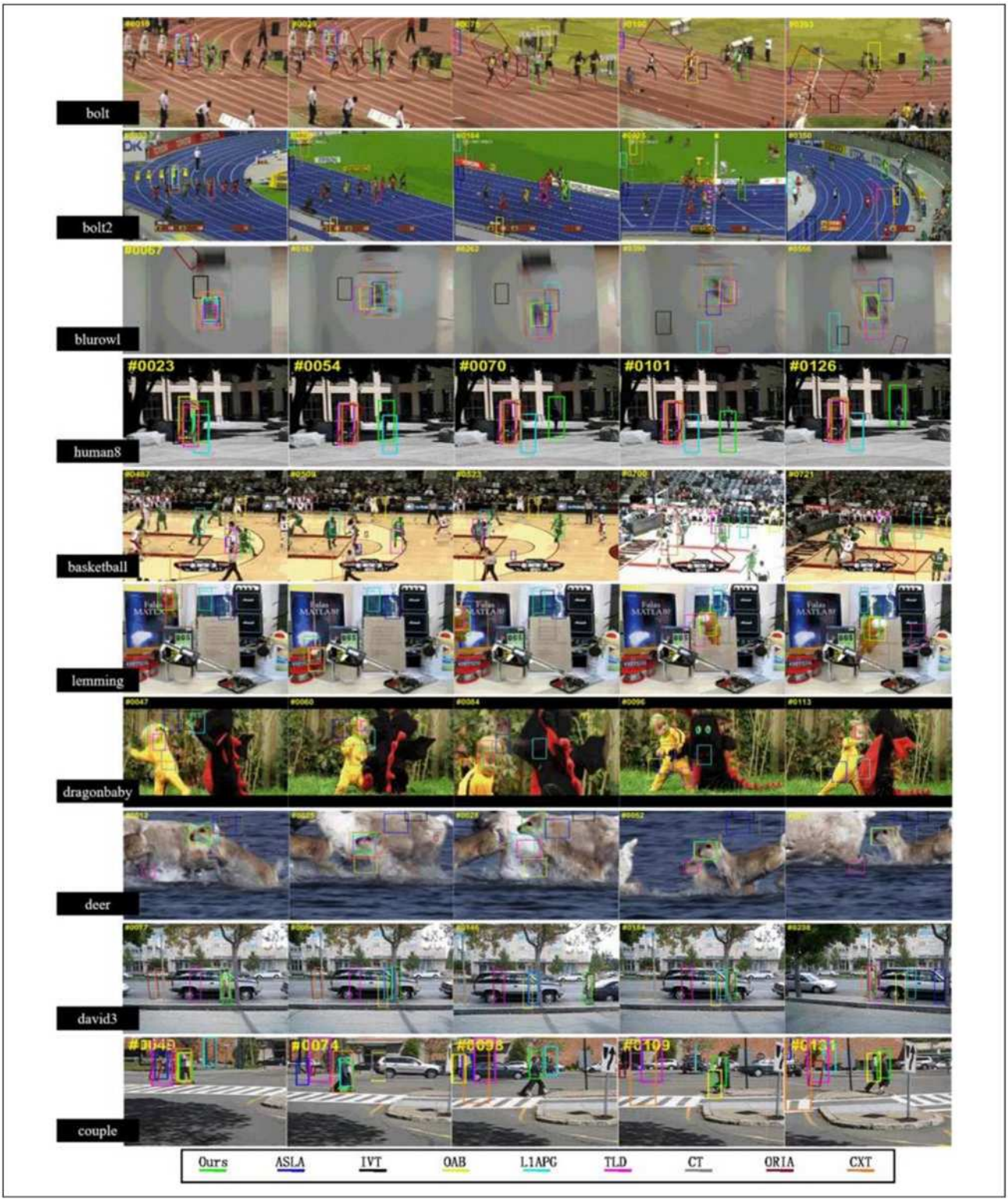


Figure 8. Comparison of the proposed approach with eight benchmarking methods ASLA, IVT, OAB, L1APG, TLD, CT, ORIA, and CXT. ASLA: adaptive local sparse appearance model-based tracker; IVT: incremental learning-based tracker; OAB: online boosting tracker; L1APG: L1 sparse tracker using APG; TLD: tracking learning-detection tracker; CT: compress tracker; ORIA: online robust image alignment tracker; CXT: context tracker.

Conclusion and future work

This article proposes a novel visual tracking method based on the weighted discriminative dictionaries and a pyramidal feature selection strategy. We utilize color features and noncolor features of the training samples to build multiple discriminate dictionaries. Then, we use the position information of samples to assign weights to the base vectors in dictionaries. These weights are used to optimize the process of target searching for selection of candidate samples, so that the frequency of abnormal samples can be effectively reduced. In the tracking process, for reducing the introduction of interference information in the dictionary and improving the tracking efficiency, we gradually update the dictionary based on noise analysis of the sparse coefficients. During the incremental update process, we sample the pool to maintain the appearance change of the target and obtain the current foreground and background information. The positive sample pool also uses a random replacement maintenance strategy to maintain the class balance of the samples. Experimental results on the all RGB sequences on OTB100[37] show that the proposed method is effective to deformation, occlusion, and other challenges in object tracking.

We will further investigate this work. First, in the video scene with cluttered background, the target is easy to be misjudged. We plan to increase the fusion of three or more features to enhance the accuracy of the target representation. Secondly, when the target scale changes, it is easy to drift away even though there are different scales of sampling. So, the mechanism of dealing with the change of target scale should be further studied.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Natural Science Foundation of China (grant numbers 61772144 and 61672008), Guangdong Provincial Application-oriented Technical Research and Development Special Fund Project (grant number 2016B010127006), Foreign Science and Technology Cooperation Plan Project of Guangzhou Science Technology and Innovation Commission (grant number 201807010059), the Scientific and Technological Projects of Guangdong Province (grant number 2017A050501039), Innovation Team Project (Natural Science) of the Education Department of Guangdong Province (grant number 2017KCXTD021), Innovation Research Project (Natural Science) of Education Department of Guangdong Province (grant number 2016KTSCX077), and Foundation for Youth Innovation Talents in Higher Education of Guangdong Province (grant number 2018KQNCX139).

ORCID iD

Penggen Zheng https://orcid.org/0000-0001-6225-7002

References

1. Lu H, Jia X, and Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: *Proceeding computer vision and pattern recognition (CVPR)*, Providence, RI, USA, 16–21 June 2012. pp. 1822–1829. IEEE.
2. Ross DA, Lim J, Lin RS, et al. Incremental learning for robust visual tracking. *Int J Comput Vision* 2008; 77(1–3): 125–141.
3. Bao C, Wu Y, Ling H, et al. Real time robust L1 tracker using accelerated proximal gradient approach. In: *Proceeding computer vision and pattern recognition (CVPR)*, Providence, RI, USA, 16–21 June 2012, pp. 1830–1837.
4. Henriques JF, Rui C, Martins P, et al. High-speed tracking with kernelized correlation filters. *IEEE Trans Pattern Anal Mach Intell* 2014; 37(3): 583–596.
5. Mei X and Ling H. Robust visual tracking using ℓ 1 minimization. In: *IEEE international conference on computer vision*, Kyoto, Japan, 29 July 2010. IEEE.
6. Zhang K, Zhang L, Yang MH, et al. Real-time compressive tracking. In: *Computer vision (ECCV)*, Florence, Italy, 7–13 October 2012, pp. 864–877.
7. Dinh TB, Vo N, and Medioni G. Context tracker: exploring supporters and distracters in unconstrained environments. In: *Proceeding computer vision and pattern recognition (CVPR)*, Colorado Springs, CO, USA, 20–25 June, 2011, pp. 1177–1184. IEEE.
8. Grabner H and Bischof H. On-line boosting and vision. In: *Proceeding computer vision and pattern recognition (CVPR)*, New York, NY, USA, 17–22 June 2006, pp. 260–267.
9. Ling H. Online robust image alignment via iterative convex optimization. In: *Proceeding computer vision and pattern recognition (CVPR)*. Providence, RI, USA, 16–21 June 2012, pp. 1808–1814. IEEE.
10. Kalal Z, Mikolajczyk K, and Matas J. Tracking-learning-detection. *IEEE Trans Pattern Anal Mach Intell* 2012; 34(7): 1409–1422.
11. Wang N, Li S, Gupta A, et al. Transferring rich feature hierarchies for robust visual tracking. In: *Computer science*, Ithaca, NY, USA, 19 January 2015, vol. 1.
12. Qi Y, Zhang S, Qin L, et al. Hedging deep features for visual tracking. *IEEE Trans Pattern Anal Mach Intell* 2019; 41(5): 1116–1130.
13. Qi Y, Zhang S, Zhang W, et al. Learning attribute-specific representations for visual tracking. In: *Thirty-third AAAI conference on artificial intelligence (AAAI)*, Honolulu, Hawaii, USA, 27 January–1 February 2019.
14. Zhang S, Qi Y, Jiang F, et al. Point-to-set distance metric learning on deep representations for visual tracking. *IEEE Trans Intell Transport Syst* 2018; 19(1): 187–198.
15. Qi Y, Zhang S, Qin L, et al. Hedged deep tracking. In: *IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 27–30 June 2016, pp. 4303–4311.

16. Zhong B, Bai B, Li J, et al. Hierarchical tracking by reinforcement learning based searching and coarse-to-fine verifying. *IEEE Trans Image Process* 2019; 28(5): 2331–2341.
17. Zhou Q, Zhong B, Zhang Y, et al. Deep alignment network based multi-person tracking with occlusion and motion reasoning. *IEEE Trans Multimedia* 2019; 21: 1183–1194.
18. Jiang Z, Lin Z, and Davis LS. Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans Pattern Anal Mach Intell* 2013; 35(11): 2651–2664.
19. Jin Z, Su Z, Wu H, et al. Robust tracking via discriminative sparse feature selection. *Visual Comput* 2015; 31(5): 575–588.
20. Wang N and Yeung DY.Learning a deep compact image representation for visual tracking. In: *Advances in neural information processing systems*, Lake Tahoe, Nevada, USA, 5–10 December 2013, pp. 809–817.
21. Song Y, Ma C, Wu X, et al. Vital: visual tracking via adversarial learning. In: *Conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, 18–22 June 2018, pp. 8990–8999.
22. Guo Q, Feng W, Zhou C, et al. Learning dynamic Siamese network for visual object tracking. In: *IEEE international conference on computer vision (ICCV)*, Venice, Italy, 22–29 October 2017, pp. 1781–1789. IEEE Computer Society.
23. Fan H and Ling H. SANet: structure-aware network for visual tracking. In: *Computer vision and pattern recognition workshops*, Honolulu, HI, USA, 21–26 July 2017, pp. 2217–2224. IEEE.
24. Wang Z, Ren J, Zhang D, et al. A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. *Neurocomputing* 2018; 287: 68–83.
25. Lan X, Ma AJ, Yuen PC, et al. Joint sparse representation and robust feature-level fusion for multi-cue visual tracking. *IEEE Trans Image Process* 2015; 24(12): 5826–5841.
26. Lu H, Jia X, and Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Providence, RI, USA, 16–21 June 2012. IEEE Computer Society.
27. Han Z, Jiao J, Zhang B, et al. Visual object tracking via sample-based adaptive sparse representation (AdaSR). *Pattern Recogn* 2011; 44(9): 2170–2183.
28. Qi Y, Qin L, Zhang J, et al. Structure-aware local sparse coding for visual tracking. *IEEE Trans Image Process* 2018; 27(8): 3857–3869.
29. Li Z, Zhang J, Zhang K, et al. Visual tracking with weighted adaptive local sparse appearance model via spatio-temporal context learning. *IEEE Trans Image Process* 2018; 27(9): 4478–4489.
30. Zhang S, Yao H, Sun X, et al. Sparse coding based visual tracking: review and experimental comparison. *Pattern Recogn* 2013; 46(7): 1772–1788.
31. Hong Z, Mei X, Prokhorov D, et al. Tracking via robust multi-task multi-view joint sparse representation. In: *Proceedings of the IEEE international conference on computer vision*, Sydney, NSW, Australia, 1–8 December 2013, pp. 649–656.
32. Zhang S, Zhou H, Jiang F, et al. Robust visual tracking using structurally random projection and weighted least squares. *IEEE Trans Circ Syst Vid Tech* 2015; 25(11): 1749–1760.
33. Zhang S, Lan X, Yao H, et al. A biologically inspired appearance model for robust visual tracking. *IEEE Trans Neural Network Learn Syst* 2017; 28(10): 2357–2370.
34. Yan Y, Ren J, Sun G, et al. Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. *Pattern Recogn* 2018; 79: 65–78.
35. Yan Y, Ren J, Zhao H, et al. Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. *Cogn Comput* 2017; 10(1): 94–104.
36. Kristan M, Eldesokey A, Xing Y, et al. The visual object tracking VOT2017 challenge results. In: *IEEE international conference on computer vision workshop*, Venice, Italy, 22–29 October 2017, pp. 1949–1972. IEEE Computer Society.
37. Wu Y, Lim J, and Yang MH. Object tracking benchmark. *IEEE Trans Pattern Anal Mach Intell* 2015; 37(9): 1834–1848.
38. Lienhart R and Maydt J. An extended set of haar-like features for rapid object detection. In: *International conference on image processing*, Rochester, NY, USA, 22–25 September 2002, *vol. 1*, pp. I-900–I-903. IEEE.

Research Article

INTERNATIONAL JOURNAL OF ADVANCED ROBOTIC SYSTEMS

International Journal of Advanced Robotic Systems
November-December 2019: 1–13

DOI: 10.1177/1729881419890155
journals.sagepub.com/home/arx
SAGE

Incremental learning-based visual tracking with weighted discriminative dictionaries

Penggen Zheng[1], Huimin Zhao[1], Jin Zhan[1], Yijun Yan[2], Jinchang Ren[2], Jujian Lv[1] and Zhihui Huang[1]

Abstract

Existing sparse representation-based visual tracking methods detect the target positions by minimizing the reconstruction error. However, due to complex background, illumination change, and occlusion problems, these methods are difficult to locate the target properly. In this article, we propose a novel visual tracking method based on weighted discriminative dictionaries and a pyramidal feature selection strategy. First, we utilize color features and texture features of the training samples to obtain multiple discriminative dictionaries. Then, we use the position information of those samples to assign weights to the base vectors in dictionaries. For robust visual tracking, we propose a pyramidal sparse feature selection strategy where the weights of base vectors and reconstruction errors in different feature are integrated together to get the best target regions. At the same time, we measure feature reliability to dynamically adjust the weights of different features. In addition, we introduce a scenario-aware mechanism and an incremental dictionary update method based on noise energy analysis. Comparison experiments show that the proposed algorithm outperforms several state-of-the-art methods, and useful quantitative and qualitative analyses are also carried out.

Keywords

Visual tracking, similarity weights, sparse representation, incremental update, weighted dictionary

Date received: 1 February 2019; accepted: 25 October 2019

Topic: Vision Systems

Incremental learning-based visual tracking with weighted discriminative dictionaries

Penggen Zheng, Huimin Zhao, Jin Zhan, more...

Show all authors

First Published December 4, 2019 | Research Article | Check for updates

https://doi.org/10.1177/1729881419890155

Article information

Altmetric 0

Abstract

Existing sparse representation-based visual tracking methods detect the target positions by minimizing the reconstruction error. However, due to complex background, illumination change, and occlusion problems, these methods are difficult to locate the target properly. In this article, we propose a novel visual tracking method based on weighted discriminative dictionaries and a pyramidal feature selection strategy. First, we utilize color features and texture features of the training samples to obtain multiple discriminative dictionaries. Then, we use the position information of those samples to assign weights to the base vectors in dictionaries. For robust

Methodologies and Application | Published: 15 April 2019

Modeling reverse thinking for machine learning

Huihui Li & Guihua Wen

Soft Computing **24**, 1483–1496 (2020) | Cite this article

267 Accesses | **2** Citations | Metrics

Abstract

Human inertial thinking schemes can be formed through learning, which are then applied to quickly solve similar problems later. However, when problems are significantly different, inertial thinking generally presents the solutions that are definitely imperfect. In such cases, people will apply creative thinking, such as reverse thinking, to solve problems. Similarly, machine learning methods also form inertial thinking schemes through learning the knowledge from a large amount of data. However, when the testing samples are vastly different, the formed inertial thinking schemes will inevitably generate errors. This kind of inertial thinking is called illusion inertial thinking. Because all machine learning methods do not consider the illusion inertial thinking, in this paper we propose a new method that uses the reverse thinking to correct the illusion inertial thinking, which increases the generalization

ds. Experimental results on benchmark data sets validated

Download PDF

Sections | Figures | References

Abstract

Introduction

Related work

Modeling reverse thinking

Experimental results

Conclusions

References

Funding

Author information

Ethics declarations

Additional information

Rights and permissions

About this article

Multimed Tools Appl (2019) 78:6847–6868
https://doi.org/10.1007/s11042-018-6279-8

Natural tongue physique identification using hybrid deep learning methods

Huihui Li[1,2] · Guihua Wen[1,2] · Haibin Zeng[1,3]

Received: 22 December 2017 / Revised: 30 May 2018 / Accepted: 15 June 2018 /
Published online: 31 July 2018

Abstract Traditional Chinese Medicine (TCM) illustrates that the physique determines the susceptibility of human to certain diseases and treatment programs for illness. Tongue diagnosis is an important way to identify the physique, but now it is performed by the doctor's professional experience and the design of a questionnaire. Consequently, accurate physique identification cannot be obtained easily. In this paper, we propose a new method to identify the physique through wild tongue images using hybrid deep learning methods. It begins with constructing a large number of tongue images that are taken in natural conditions, instead of in a controlled environment. Based on the resulting database, a new method of tongue coating detection is put forward that applies a rapid deep learning method to complete the initial tongue coating detection, and then utilizes another deep learning method, a calibration neural network, to further improve the accuracy of tongue detection. Finally, an effective deep learning method is applied to identify the tongue physique. Experiments validate the proposed method, illustrating that physique identification can be performed well using hybrid deep learning methods.

Keywords Deep learning · Tongue coating · Physique identification · Traditional Chinese Medicine (TCM)

✉ Guihua Wen
crghwen@scut.edu.cn

Huihui Li
2977756@qq.com

Haibin Zeng
873532938@qq.com

[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

[2] Guangdong Artificial Intelligence Engineering Research Center for Traditional Chinese Medicine, Guangzhou, China

[3] Guangdong Artificial Intelligence Engineering Research Center for Traditional Chinese Medicine, Shenzhen, China

1 Introduction

Traditional Chinese Medicine (TCM) is a complete medical system that plays an irreplaceable role in medical care in China [33]. It follows the holistic and systematic ideas while Western Medicine does the reductionism thinking mode. TCM has been validated by many patients and scientists to be very safe and effective in practices. Particularly it can provide alternative ways when western medicine are unable to supply effective treatment for some routine ailments such as influenza, allergies, and chronic diseases [15]. Currently TCM is recognized and practiced not only in China but also in the Southeast Asia, Japan, United States, Europe, Australia and some other countries. TCM is gradually formed and developed with a long-term clinical practice, having accumulated a large amount of clinical data and medical literature on diagnosis and treatment over thousands of years [17]. Treating the same disease with different methods and homotherapy or heteropathy are the basic principles of knowing and treating diseases in TCM [1, 18]. Differentiation is the premise and basis of treatment, which heavily depends on the individual physique. TCM categorizes the physique into nine types in the book "Classification and Decision of TCM Physique" published by the China Association of Chinese Medicine [7]. These types are "pinghe," "yangxu," "yinxu," "qixu," "shire," "tanshi," "qiyu," "xueyu," and "tebin." At the same time, a questionnaire has been designed with criteria to recognize the physique automatically. Subjects usually answer the questions in the questionnaire according to a certain experience and feeling of the frequency or severity of experience. However, this method can be easily influenced by the individual's subjective intention [34], while a large amount of time is required to complete all testing. In such cases, a more intelligent method than the questionnaire method is required to perform physique identification automatically and efficiently.

Tongue diagnosis is one of the most widely used diagnostic methods in TCM [36]. This is because the color and texture features of tongue coating reflect the health status of the patient. For example, color information of the tongue coating reflects valuable information about the state of disease and its correlation with the internal organs [8]. Clinical studies have suggested relationships between visceral cancers, heart diseases, and abnormalities of the tongue and its coating [26]. Inspection of the tongue can instantly clarify one's pathological problems, so that people seeking health care can have their tongues routinely examined [4]. Therefore, the appearance changes of the tongue coating are very helpful to TCM practitioners in monitoring the improvement or deterioration of the patient's health status. However, tongue diagnosis offers a simple, immediate, inexpensive, and non-invasive solution for various medical applications [30]. It is also side-effect free, painless, and well suited for remote diagnosis. Therefore, it has been selected to perform physique identification [34]. Actually, some tongue image analysis systems have been advanced to prove its effectiveness not only in Asian medicine, but in Western medicine as well [12, 13].

As the tongue coating reflects the environment inside the body, there is great reference value in using the tongue to determine physical characteristics. Various physiques corresponding to the tongue are mentioned in the book "Classification and Decision of TCM Physique" [7]. For example, "pale red tongue, the tongue body fat soft, side have teeth marks" is provided for the "qixu" physique. These physical characteristics can be applied to design the questionnaire for identifying physiques. However, it is based on interrogation, human subjects have great difficulties in making objective decisions. If the tongue coating picture of a subject can be automatically taken and then stored in the form of a digital image, an intelligent diagnosis can be achieved to obtain the more objective and accurate results [4,

37]. Actually, various intelligent methods have been applied to study TCM theory [5, 17], such as clustering analysis, association rules and regression analysis, and pattern recognition. Objective quantitative methods to evaluate the color, texture, and surface of the tongue and to define their relationships with patients' health conditions have also been devised [26]. Some related tongue diagnosis systems have been also proposed [9, 10, 30]. However, to the best of our knowledge, these systems generally assume that the tongue images are taken in a well-controlled environment, including lighting conditions and fixed tongue position [19]. To effectively deal with tongue images taken under different conditions, where the color of tongue images may be different, many methods have been proposed to perform color correction on the captured tongue images in TCM systems [8, 29, 37, 38]. However, they are all based on traditional machine learning methods, instead of deep learning methods.

Various factors impact the quality of captured tongue images, such as illumination conditions and random noise. These will definitely reduce the accuracy and reliability of the subsequent automatic tongue diagnosis analysis [38]. Another challenge results from variations in the tongue features used in diagnosis, such as color, texture, coating, and shape, which results in the difficulties in precisely extracting the tongue region from the entire image. Finally, few machine learning methods are applied to perform tongue physique identification. To meet these challenges, in this paper we propose a new method to identify the physique through natural tongue images using hybrid deep learning methods on the large number of training databases. The main contributions of the paper are as follows. A large number of tongue images in natural conditions are constructed instead of taking the photos in a controlled environment. Second, a method of tongue coating detection is put forward that applies a rapid deep learning algorithm to complete the initial detection of the tongue coating and then applies another deep learning method, a calibration neural network, to further improve the accuracy of tongue coating detection. Finally, an effective deep learning method is selected to identify the tongue physique, illustrating that physique identification can be performed automatically.

2 Related work

The proposed method involves in the quality of tongue images, tongue detection, and tongue physique identification. They are theoretically related to object detection and image classification,which are performed primarily using deep learning methods. For example, Li et al. used a cascade structure that trained three dierent classication models and three bounding-box regression models to transform the face detection into three "categorical-regression" cascade phases [16]. We also ever presented a new cascade structure face detection method[20], which uses a full convolution network to get face candidates quickly and the bounding-box regression model to perform face regression until the termination condition was satised. Finally, a binary classication CNN was used to obtain the nal result. However, as there are a huge number of references related to image classification, we limit the related work to the tongue image field.

2.1 Capturing conditions of tongue images

Tongue images, captured in the natural environment instead of a well-controlled one, incorporate different amounts of color distortion when compared to actual tongue images. This might result in incorrect diagnosis and prescription [37]. For example, an android-based

application based on the Canny algorithm was developed for automatic tongue diagnosis [36], but it does not discuss calibration of image colors due to various lighting conditions. Many methods can be applied to perform color correction on captured tongue images in TCM tongue digital analysis systems [14, 29, 37, 38]. For example, based on hue, saturation, and value color space, new color brightness threshold parameters have been devised to improve the efficiency of tongue coating separation procedures and to eliminate shadows [8]. The TongueDx system presents a tongue color calibration method by using teeth color as a standard color [6]. These methods aim to obtain high-quality tongue images. Different from these, the proposed method applies deep learning methods to learn features automatically from all kinds of tongue images captured in different environments, so that it can better deal with tongue images captured under any conditions when working.

2.2 Detection

An accurate segmentation of the tongue coating from a tongue image is a crucial step for automatic tongue diagnosis in TCM. Some methods have been proposed to segment the structure of interest for medical images, such as the bielliptical deformable contour approach based on active contour models [23]. Because this method is very sensitive to the initial curve, an automatic tongue segmentation method is proposed that combines a region merging strategy with initial watershed segmentation [21]. Another novel method was proposed that combines the geometrical snake model with the parameterized GVF snake model [25]. This method utilizes the prior knowledge of tongue shape and the tongue's location in tongue images. In addition, a double geo-vector flow was applied to detect the tongue edge and segment the tongue region in the image [26]. Generally, these approaches can be divided into two classes: region-based approaches and edge-based ones [32]. However, two key difficulties are encountered in their use. First, the initial contours must be close to the true boundaries; second, active contours progress into boundary concavities with difficulty. Taken as a whole, these methods fail to satisfy demands for both accuracy and robustness simultaneously, which are the basic requirements for a successful segmentation of tongue images. Therefore, a new method combining these two categories was proposed, which often performs better [32]. All these methods are based on image processing technology. Recently, some machine learning methods have been applied to tongue coating separation, such as the Gaussian conditional density model [11] and fuzzy clustering [2]. However, these are traditional machine learning methods, and they do not use the deep learning methods to automatically segment the tongue coating. In these methods, the accurate judgment of separation is still ambiguous because of under-detection or over-detection of the tongue coating.

2.3 Physique identification

Currently, physique identification is often performed by answering a personal questionnaire that is designed to recognize the individual's physique. The answer to each question in the questionnaire will be selected from five cases, including no (don't), little (a bit), sometimes (some), often (quite), and always (very). Answers to all questions are then applied to determine the physique type using a predefined computing method. However, this method can be easily influenced by the individual's subjective intention [34]. It is very difficult for human subjects to make objective decisions in answering certain questions. In addition, the questionnaire-based testing and analysis is time consuming. However, features of the tongue, such as color, texture, and geometry, are often applied to examine health conditions

in TCM [31, 35]. For example, the quantitative chromatic and textural features are first extracted from tongue images and then a Bayesian network is employed to model the relationship between these features and diseases [22]. Geometric features such as length, area, and angle of the tongue are also applied to automatically recognize and analyze tongue shapes [5]. On the whole, these tongue diagnoses are always concerned with the identification of syndromes rather than with the connection between tongue appearances and physique identification [22]. Only in one work were tongue features such as color and texture features used to recognize the physique [34]. All features used in these techniques are hand-crafted, instead of being automatically learned from a large number of training samples. Moreover, machine learning methods are not applied in them to perform physique identification automatically.

3 Framework of the proposed approach

In order to perfectly identify the physique in the natural environment, we propose a new method to identify the physique through natural tongue images using hybrid deep learning methods on large training databases. The method is illustrated in Fig. 1, and contains six steps. The first step is tongue image acquisition, where the tongue image is taken with a camera in the natural environment. The second step is initial tongue coating detection, where a deep learning method designated Faster R-CNN is applied [24]. Namely,tongue region extraction out of the whole face image is performed in this step. In the third step, another deep learning method, VGG, is applied to further calibrate the detected tongue region to obtain a more accurate tongue coating image [16, 27]. In the fourth step, the tongue coating image is segmented from the entire tongue image. The fifth step performs physique identification on the segmented tongue image using the deep learning method named GoogLeNet [28]. Finally, the recognized physique types are outputted.

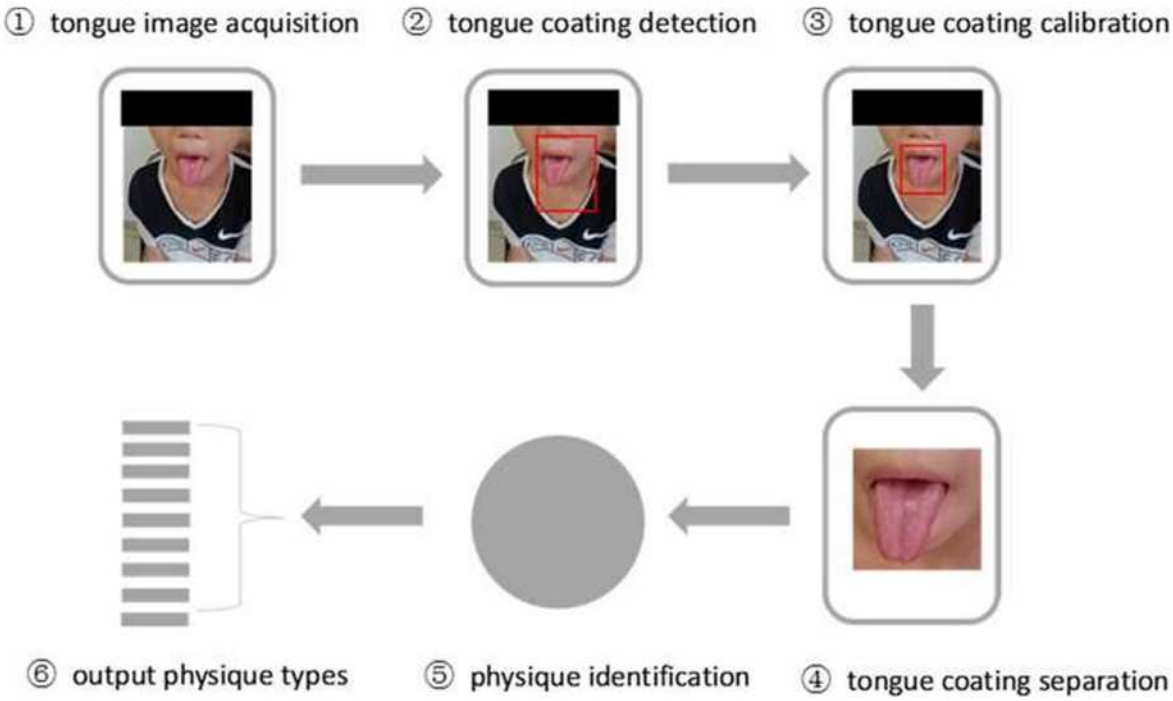


Fig. 1 Architecture of proposed method

3.1 Databases

As deep learning methods are applied to perform the physique identification, it is necessary to have a large number of tongue images as training databases with predefined labels. However, such databases are not publicly available. Therefore, we must construct them by cooperating with hospitals, in which the tongue images of patients are directly captured and then labeled physique types by Chinese medicine experts. In this way, a large number of tongue images with all kinds of qualities are collected to construct the tongue image databases for tongue coating detection, tongue coating calibration, and physique identification. Generally, a person may have several physique types, which undoubtedly increases the difficulty of physique identification. In order to reduce the difficulty of identification, the task is simplified to be a single label classification problem in this paper. That is to say, each tongue image only corresponds to one of the physique types with the largest belief degree.

Tongue coating detection database The tongue coating detection database is composed of the images containing part of the face and tongue fur. Tongue coating detection detects the tongue image from the entire image and then extracts it as a separate image. We utilize a deep neural network to perform this task. In order to train the deep neural network model, a large number of training samples are required for tongue coating detection, in which each sample contains the captured image showing the tongue and the real position information of the tongue image in the entire captured image. In this paper, a tongue coating detection database was constructed that contains 5683 tongue coating images with the corresponding real positions of the tongue coatings.

Tongue coating calibration database The tongue region image extracted by the tongue coating detection method may deviate from its correct position. These deviations can be defined as belonging to many categories. Each deviation can be calibrated to its correct position based on predefined rules. To automatically classify the deviation type of the detected tongue region image using a deep learning method, a large number of training samples composed of the detected tongue image region and its corresponding deviation type should be prepared in advance to form the tongue coating calibration database.

The tongue coating calibration method is inspired by face region calibration [16], where the deviation category comprises the proportion of the scale, deviation in the X direction, and deviation in the Y direction. The specific deviation categories are defined by Eq. (1), where (x, y, w, h) represents x and y coordinates of the upper left, width, and height of the true tongue coating region. By using the inverse operation of (1), the tongue coating region can be adjusted to the correct position by the deviation type. In this way, each tongue coating image can produce 45 new regional tongue images according to (1). Each new regional tongue image is intercepted and annotated by deviation category. All of them are applied to build the tongue coating calibration database, which in the example used in this paper, has a total of 255735 tongue images. The relevant formulas are expressed as follows:

$$\left(x - \frac{x_n w}{s_n}, y - \frac{y_n h}{s_n}, \frac{w}{s_n}, \frac{h}{s_n}\right), \tag{1}$$

$$s_n \in \{0.83, 0.91, 1.0, 1.10, 1.21\}, \tag{2}$$

$$x_n \in \{-0.17, 0, 0.17\}, \tag{3}$$

$$y_n \in \{-0.17, 0, 0.17\}. \tag{4}$$

Table 1 Sample distribution of tongue physique identification database

Data	qixu	yingxu	yangxu	tanshi	shire	qiyu	xueyu	tebin	pinghe
Number	7840	3670	1189	2514	2557	2059	933	91	1629

Tongue physique identification database In order to perform tongue physique identification, a larger tongue physique identification database is constructed, which contains 22482 images with their physique types assigned by TCM experts. Each image is the tongue coating image extracted from the captured tongue images by tongue coating detection and tongue coating calibration, thus reducing the impact of other parts of the tongue image on tongue physique identification. The sample distribution of the tongue physique identification database is illustrated in Table 1,where the number value indicates that the number of tongue images belong to each physique. The samples corresponding to each physique are presented in Fig. 2.

3.2 Tongue coating detection

Tongue coating detection can be regarded as a kind of object detection. As Faster R-CNN is a newly proposed and excellent object detection method [24], it is applied to perform tongue coating detection. Faster R-CNN, whose structure is presented in Fig. 3, is composed of a convolutional neural network (CNN), a region proposal network (RPN), and a CNET.

CNN is mainly used to extract features of tongue images based on the idea of shared weights and good nonlinear learning ability. It is implemented here by using the VGG-16 model with the first 13 layers, while the fully connected layer is removed. In this way, convolution, ReLU, and pooling (Max) operations are mainly involved.

RPN is the key component of Faster R-CNN. It is a fully convolutional network that simultaneously predicts object bounds and objectiveness scores at each position. RPN is

Fig. 2 Architecture of proposed method

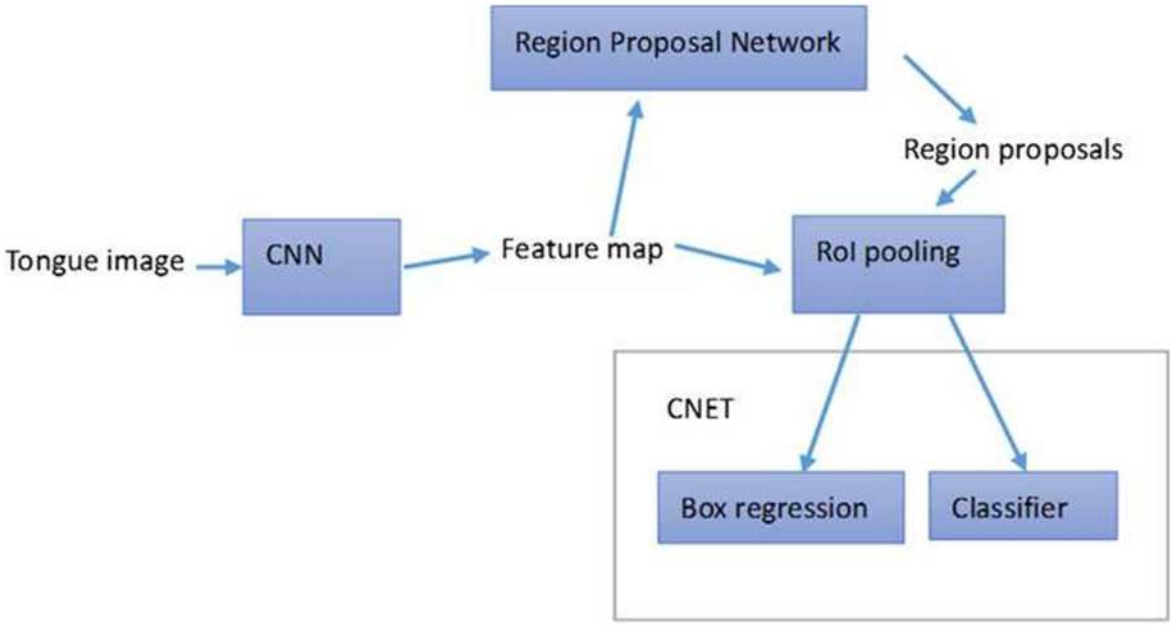


Fig. 3 Structure of Faster R-CNN in tongue detection method

trained end-to-end to generate high-quality region proposals. It uses a sliding window to generate a fully connected feature vector with 256 dimensions on a convolution feature map generated in a convolutional neural network. The feature vector is then applied to produce two branches with two fully connected layers: the *cls* layer and *reg* layer. The *cls* layer performs a binary classification task to determine whether the current region is the foreground or background, where the training database for classifying foreground or background is constructed manually in advance. The loss function of the classification is Cross Entropy. The *reg* layer is a regression task used to predict the candidate region coordinates *(x, y)* as well as the size *(w, h)* corresponding to the central anchor point of the current region. The anchor point is located at the center of the sliding window. An anchor can generate k different candidate regions with different scales and ratios between width and height, so that the same sliding window can be used to predict the candidate regions. Here, there are four kinds of scales and three kinds of ratios between width and height – see Table 2 – leading to a total of $k=12$ candidate regions. For a convolution map of size $w \times h$, $w \times h \times k$ anchor points can be obtained by using a sliding window of size 3×3 where the step size is 1.

The loss function of RPN consists of both the loss function of the classification task and the loss function of the regression task, corresponding to the *cls* layer and the *reg* layer. The loss function of the classification task is defined by cross-entropy loss, shown in (5), where x is an n-dimensional vector that contains the rating values of n categories, and $class$ is the target category:

$$L_{cls}(x, class) = -x[class] + ln\left(\sum_{j=1}^{n} e^{x[j]}\right). \tag{5}$$

Table 2 Setup parameters for anchors

Parameters	Values
Scale values	0.1, 0.2, 0.4, 0.8
Ratios between width and height	1:1, 2:1, 1:2

The loss function of the regression task uses a smooth L_1 loss, which is used to predict the coordinates, width, and height of the target region. It is defined by (6), where x is the output value of the neural network and y is the target value:

$$L_{reg}(x, y) = \frac{1}{n} \sum_i \begin{cases} 0.5 \times (x_i - y_i), & \text{if } |x_i - y_i| < 1; \\ |x_i - y_i| - 0.5, & \text{otherwise.} \end{cases} \tag{6}$$

The loss function of RPN can now be defined in (7), where $\{p_i\}$ and $\{t_i\}$ are the outputs of the *cls* layer and the *reg* layer, respectively. If the input sample is the foreground, $p_i^* = 1$, else $p_i^* = 0$. t_i^* is the correct target area. λ is set to 10. We write

$$L(\{p_i\}, \{t_i\}) = \frac{1}{n_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{n_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*). \tag{7}$$

RoI pooling is a simplified version of spatial pyramid pooling (SPP)[3], which is able to deal with inputs of different sizes, making the outputs have the same size. SPP can determine the size of the final feature vector, and then determine the size of the pool. RoI pooling is a layer of SPP using only one scale. In our case, RoI pooling is set to 6×6 output.

The CNET is mainly composed of box regression and Classifier. Box regression is composed of a fully connected layer, a batch normalization layer, and a Dropout layer, which are used to predict the location of the target area. Classifier is a neural network classifier composed of a fully connected layer, a batch normalization layer, a Dropout layer, and a Softmax layer, which is used to identify the target categories. Box regression is used as a regression task, using a smooth L_1 loss function. Classifier is a classification task using cross-entropy loss function. The loss function of the entire CNET is a weighted combination of the loss function of box regression and the loss function of Classifier.

When Faster R-CNN is applied to perform the tongue coating detection, it needs training. As it is a complex deep neural network, each component of Faster R-CNN must be trained in advance, and then the entire network is refined through training on the entire database. CNN shared by both RPN and CNET in Faster R-CNN is applied to train a binary classifier to perform the classification of tongue coating and non-tongue coating. CNN uses the trained parameters of the first 13 layers of the VGG-16 model and then fine trained on our database composed of a large number of tongue coating images and non-tongue coating images. The goal of RPN training is to classify the region of each anchor point to the foreground or background, and then to regress the target coordinates of each anchor point. Here, the minimum edge of a tongue image is fixed to 480 pixels. A large number of training samples with anchor region types are generated in accordance with the four anchor region sizes and three ratios between width and height. If the Intersection over Union (IoU) value between the generated region and the correct region is greater than 0.6, it is marked as the foreground region. If IoU is less than 0.2, the region is marked as the background region. The number of selected foreground regions and selected background regions are the same.

After training CNN and RPN, both the Classifier and box regression are added to adjust all parameters by training the entire network. The entire training process proceeds as follows.

Training process for tongue coating detection

1. Data preprocessing and generation of anchor regions are performed. Each tongue coating image is horizontally flipped over with 0.25 probability to increase training samples. The images of both the foreground and background regions are selected in terms of 1:1.
2. The tongue coating image is fed into CNN to generate feature maps.

3. RPN is applied to generate candidate regions. The sliding window in RPN is utilized to scan the feature maps generated in the second step to create features for anchor regions. Features of the foreground and background regions are respectively sent to the *cls* layer and *reg* layer. The Softmax loss for the *cls* layer (denoted L_{pcls}) and L_1 loss for the *reg* layer (denoted L_{preg}) are computed. For the foreground region, the *reg* layer is used to generate the candidate region. For the background region, the anchor region is used as the candidate region.
4. RoI pooling is applied to generate fully connected features with a fixed size based on feature maps of the selected candidate regions. This can avoid the feature size inconsistency problem due to tongue coating images with different sizes.
5. The prediction of the tongue coating region is performed. The fully connected features are passed to box regression and Classifier. Subsequently, L_1 loss of box regression (denoted L_{creg}) and the negative log-likelihood loss of Classifier (denoted L_{ccls}) are calculated.
6. Back-propagation is applied to update parameters, where CNET takes $L_{creg} + L_{ccls}$ as the loss to update its parameters, RPN takes $L_{preg} + L_{pcls}$ as the loss to update its parameters, and CNN takes the loss to update its parameters that is the sum of L_{creg}, L_{ccls}, L_{preg}, and L_{pcls}.
7. Repeat the above steps until convergence is obtained or the maximum number of iterations is reached.

Testing process for tongue coating detection The tongue coating detection process based on Faster R-CNN is slightly different from that of the training phase, and contains the following steps:

1. The data preprocessing for the tongue coating image is performed in the same way as in the training phase.
2. The preprocessed tongue coating image is fed into CNN to generate feature maps.
3. RPN is applied to create candidate regions. The sliding window of RPN is applied to scan feature maps generated in the second step to generate features of the anchor region. If it is determined to be the foreground region by the *cls* layer with confidence up to 0.95, the *reg* layer is applied to forecast the corresponding tongue coating region as its candidate region.
4. The optimal regions are selected from all candidate regions. The non-maximum suppression method is applied to eliminate redundant candidate regions so as to find the best candidate region. Non-maximum suppression always selects candidate regions with maximum reliability and then filters out those candidate regions with IoU larger than 0.25. This process is repeated to select all candidate regions with high confidence.
5. RoI pooling generates the fully connected features with fixed size based on the feature maps of all optimal regions.
6. The fully connected features are fed into Classifier and box regression to predict the tongue coating regions.
7. The non-maximum suppression method is applied to select the best one from predicted coating regions.

3.3 Tongue coating calibration

Tongue coating detection has been performed well, but not perfectly. Its performance must be further improved by adjusting the position of the detected tongue coating region.

Inspired by face region calibration [16], we use another kind of deep neural network, called a calibration network, to implement the calibration of the tongue coating position. It assumes that both the location of the tongue coating and the deviation of the tongue coating position from its real position can be predicted. Subsequently, the tongue coating position can be calibrated to its real position. This problem can be treated as a classification problem, as deviation categories from the real tongue coating position have been predefined by (2)–(4). Once the deviation category is determined, the calibration of tongue coating position can be performed according to the deviation adjustment rules defined in (1).

The calibration network is also implemented by VGG-16, as it exhibits excellent performance in tongue coating detection. The structure of the calibration network is shown in Fig. 4.

The calibration network is applied as a classifier to classify the deviation category. The training of this network is performed by the following steps: (1) Perform the data preprocessing. The mean operation is performed while each image is flipped over with probability 50% to create new samples. (2) Perform the forward propagation. After the preprocessing, each image is fed into the calibration network. After convolution, ReLU, pooling, full connectivity, and other operations, the Softmax losses of 45 categories are calculated. (3) Perform the back-propagation. All Softmax losses are back-propagated to update network parameters. (4) Repeat the above steps until convergence is obtained or the maximum number of iterations is reached.

The testing procedure of the calibration network is composed of the following steps: (1) The tongue coating region is detected by the tongue coating detection method. (2) The tongue coating region is resized to 224 × 224. (3) After the mean operation is performed, the calibration network is applied to obtain the deviation category of the tongue coating. (4) The deviation of X, the deviation of Y, and the proportion of the scale are obtained based on the obtained deviation category. (5) The inverse operation of formula (1) is applied to adjust the tongue coating region to the correct position.

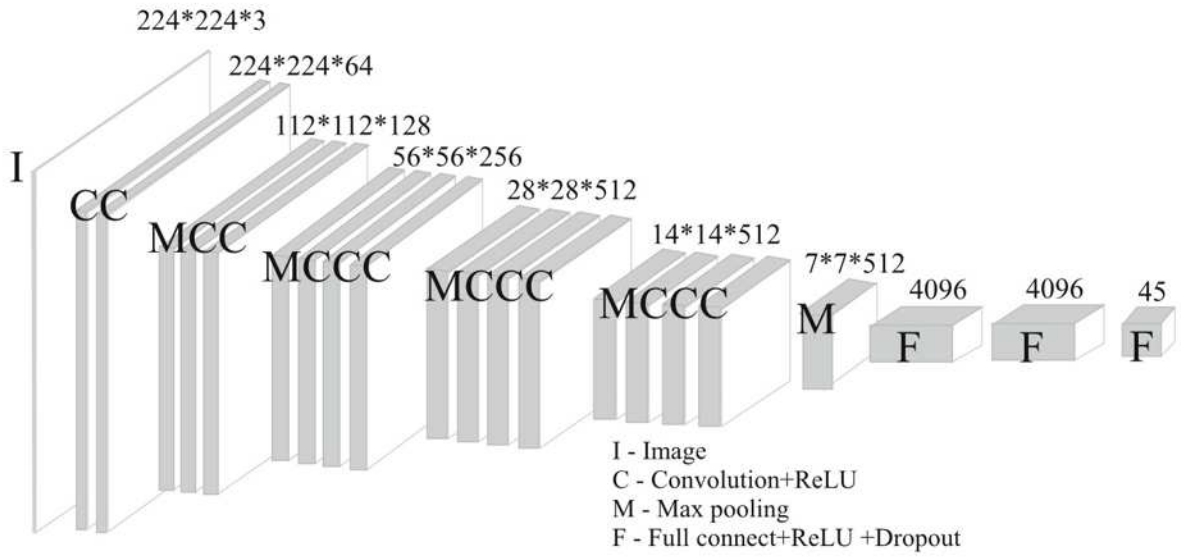


Fig. 4 Structure of calibration network

3.4 Tongue physique identification

GoogLeNet is applied to perform the tongue physique identification. It is the winner of the 2014 ILSVRC competition, and belongs to the deep convolutional neural network model [28]. Since the development of the VGG model, convolutional neural networks are being advanced toward a deeper architecture; that is, they are including increasingly more network layers and parameters, which, however, often poses challenges to computing power. In order to reduce the computational resources required while improving the neural network effectiveness, a module called Inception is proposed as a key component of GoogLeNet. The Inception module, whose structure is shown in Fig. 5, is basically a parallel combination of 1×1, 3×3, and 5×5 rolls. This ensures that the output of the Inception module is rich in characteristic information. In addition, the Inception module also uses a 1×1 convolution layer to reduce the number of features and the amount of computation in the previous layer. This layer is generally referred to as the bottleneck layer. The 1×1 convolution in the bottleneck layer can be viewed as a combination of features on the previous feature map and can be used effectively with fewer parameters. In this way, the computational overhead is greatly reduced without weakening the neural network learning. Since the input features are all related, the redundancy feature can be removed by the combination of 1×1 convolution. The GoogLeNet structure is shown in Fig. 6, depicting a large number of Inception modules.

Because tongue physique identification is a classification problem with nine physique types, the last layer of GoogLeNet is changed to be a $1 \times 1 \times 9$ fully connected layer. In addition, in order to better train GoogLeNet, it is trained on the ILSVRC training database as an initialization, and then is fine trained on our prepared tongue physique identification database.

4 Experiments and validation

To validate our approach, many experiments were conducted on tongue image databases, against which some related state-of-the-art approaches were also compared.

4.1 Experimental results on tongue coating detection

In order to validate our tongue coating detection method, we apply the Torch deep learning framework to implement it, and then perform experiments on the constructed tongue coating

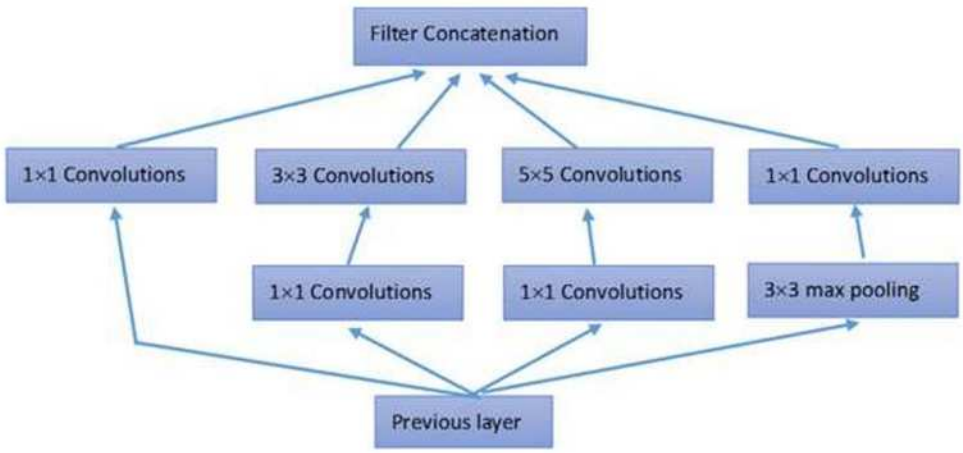


Fig. 5 Structure of Inception module

type	patch size/ stride	Output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception(3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception(3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception(4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception(4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception(4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception(4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception(4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception(5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception(5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout(40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Fig. 6 Structure of GoogLeNet for tongue physique identification

detection database, which has 5683 tongue images with each tongue image having only one tongue coating. This database is divided into a training set and testing set at a 4:1 ratio, respectively, so that the testing set is made up of 1138 tongue coating images. After 50000 iterations of training, the tongue coating detection model is formed, which will be then used to detect the tongue coating images of the testing set to evaluate detection effectiveness. The training parameters for our approach are shown in Table 3.

As the positive and negative tongue coating images are determined according to IoU value of the marked area, in order to more precisely analyze the effect of the tongue coating detection, detection results on the testing set are divided into the following categories: (1) Correct category, referring to the case that classification is correct and IoU between

Table 3 Training parameters for tongue coating detection method

Parameters	Values
Picture size	80640
Batch size	300
Candidate area	$\{48^2, 96^2, 192^2, 384^2\}$
Candidate aspect ratio	$\{1:1, 1:2, 2:1\}$
ROI pooling scale	66
Optimization method	RMSprop
Learning rate	$1e^{-5}$
ALPHA	0.9
Parameter decline factor	$1e^{-4}$
Maximum number of iterations	50000

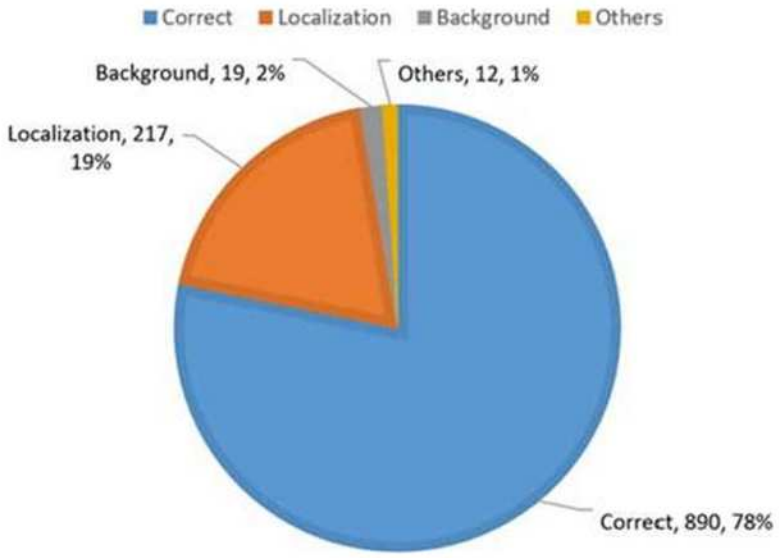


Fig. 7 Experimental results of tongue coating detection method

predicted region and true region is greater than 0.5. (2) Localization category, referring to the case that classification is correct and IoU between predicted region and true region is distributed from 0.1 to 0.5. (3) Background category, referring to the case that classification is correct and IoU between predicted region and true region is less than 0.1. (4) Others category, referring to the case that classification is in error.

The experimental results on the testing set are shown in Fig. 7. Some examples of detected tongue coating images are presented in Fig. 8. It can be seen that the method can be used to identify whether an image contains a tongue coating with high accuracy, up to 98%. The Correct category obtains an of accuracy 78%, which is the final target. The Localization category still accounts for 19%, indicating that there is room for further calibration of the tongue coating position, so as to further improve the performance of the tongue coating detection method.

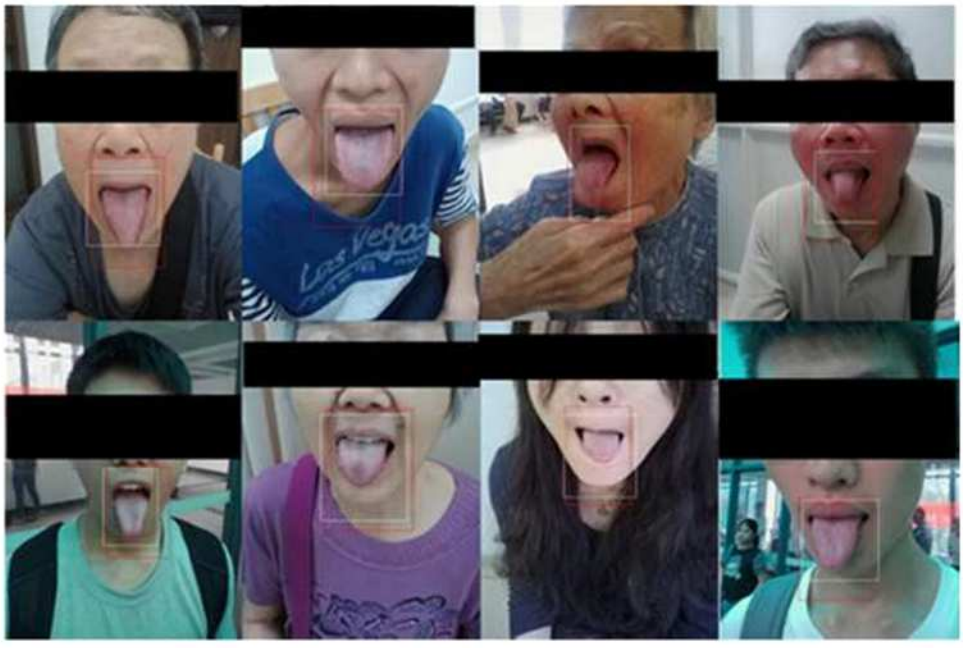


Fig. 8 Examples of detected tongue coating images

Table 4 Training parameters for tongue coating calibration method

Parameters	Values
Picture size	80640
Batch size	300
Optimization method	RMSprop
Learning rate	$1e^{-5}$
ALPHA	0.9
Parameter decline factor	$1e^{-4}$
Maximum number of iterations	50000

4.2 Experimental results on tongue coating calibration

Experiments were conducted to demonstrate the effectiveness of the calibration network. The training parameters of the network are presented in Table 4 where ALPHA is the momentum coefficient of the optimization algorithm RMSprop. In the experiments, the VGG-16 model is first trained by ImageNet images and then used to initialize the first 13 layers of the calibration network. This is why the learning rate is adjusted to a very small value.

The experimental results are shown in Fig. 9. It can be observed that the accuracy of the Correct category is up to 91% from the original 78%, while the accuracy of the Localization category reduces to 7% from the original 19%. This illustrates that the calibration network can effectively enhance the effect of tongue coating detection. Some examples of tongue coating calibration are shown in Fig. 10, where the white box is the true tongue coating region and the red box is the predicted region. It can be seen that the deviations are adjusted correctly. However, the method can be used to identify whether an image contains the tongue coating with a higher accuracy of up to 99%.

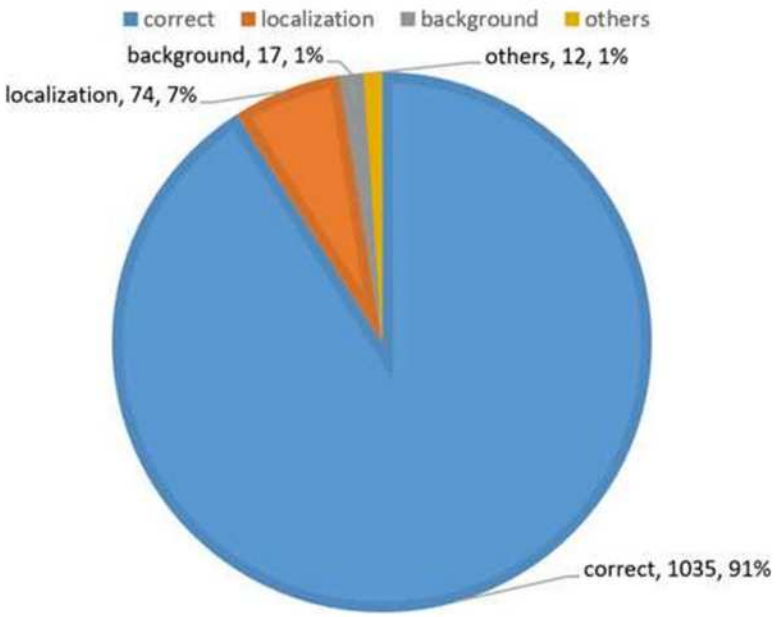


Fig. 9 Results of tongue coating calibration

Fig. 10 Examples of tongue coating calibration

4.3 Experimental results on tongue physique identification

Experiments were conducted to validate our tongue physique identification method. In order to show the rationality of the selection of GoogLeNet, VGG and ResNet are applied as two representative deep learning methods, and are compared in performing tongue physique identification, in which VGG uses the 16-layer model and ResNet the 101-layer model. Because tongue physique identification is a nine-level classification problem, the last layers of both VGG and ResNet are changed to be the fully connected layers with $1 \times 1 \times 9$. In addition, in order to obtain the better models, VGG, GoogLeNet, and ResNet are first trained on the ILSVRC as an initialization and are then fine trained on our tongue physique identification database. In order to observe the effectiveness of a training database of different sizes on our method, we divide the tongue physique identification database into five parts by randomly selecting the number of samples of 20%, 40%, 60%, and 80% respectively from each category in the training database so as to ensure that each selected training database contains all categories.

The performance evaluation criteria are Top1 error and Top3 error. Top1 error is the ratio between the number of misclassified samples and the number of total samples. Top3 error is the ratio between the number of the cases that the first three categories processed by our

Table 5 Training parameters for tongue physique identification method

Parameters	Values
Picture size	80640
Batch size	300
Optimization method	Adam
Learning rate	$1e^{-3}$
BETA1	0.9
BETA2	0.99
Maximum number of iterations	1000

Table 6 Structural parameters for VGG, GoogLeNet, and ResNet

Structural parameters	VGG	GoogLeNet	ResNet
Layers	16	22	101
Convolution layers	13	21	100
Convolution kernel sizes	3	7135	7135
Number of parameters (million)	134	6	47
Batch normalization	-	-	+

method are all wrong and the number of total cases. The reason for the evaluation of Top3 is that each person's physique may contain multiple categories rather than a single one. In our approach, we simplify the physique identification task as a single label learning task. All models are implemented using the Torch framework, the training parameters of which are presented in Table 5. In the table,BETA1 and BETA2 are first order momentum attenuation coefficient and second order momentum attenuation coefficient of the optimization algorithm Adam. The maximum number of iterations is set to 1000 because the current tongue physique identification database is not large enough. When 1000 iterations of training are performed, the error is very small.

Table 6 illustrates the basic structural parameters of VGG, GoogLeNet, and ResNet implemented in our experiments. The number of parameters of the VGG model is the highest and its model is the largest. The number of parameters of GoogLeNet is smaller, but the number of layers is deeper than that of VGG. ResNet is the deepest one, having 101 layers and a batch normalized layer. Its operation time is the longest among the three models. This is one of the reasons we selected GoogLeNet for the proposed approach to tongue physique identification.

The experimental results are shown in Table 7, where the error rates of VGG-16, GoogLeNet, and ResNet-101 are presented. It can be seen that with increasing size of the training database, the error rates of the three models decreased, indicating that the large training database is very helpful to the accuracies of the models. On each training database, compared to its former training database, the three models can reduce the error rate by approximately 4%, indicating that the current training database is not large enough and that more training samples should be prepared. Second, the GoogLeNet model performs the best one among the three models, exhibiting the lowest Top1 error rate of 39.58% and a Top3 error rate of 18.07%. VGG-16 exhibits similar results as GoogLeNet, but has more parameters and requires more computing resources. To our surprise, the effectiveness of the ResNet-101 is as not good as expected, considering it was the 2015 ILSVRC champion, and we would assume that its effectiveness should be better than that of both GoogLeNet and

Table 7 Identification results of three methods on TCM physique (error ratio%)

Training data	Testing data	VGG (Top1)	VGG (Top3)	GoogLeNet (Top1)	GoogLeNet (Top3)	ResNet (Top1)	ResNet (Top3)
3597	895	58.21	30.73	57.43	32.29	59.55	35.53
7194	1795	54.15	27.08	53.09	26.80	55.71	29.80
10794	2692	47.40	23.66	47.33	22.70	51.11	26.19
14389	3594	45.91	22.53	45.33	21.20	48.49	23.26
17990	4492	39.78	18.34	39.58	18.08	43.41	21.19

Table 8 Confusion matrix of **VGG** on largest training database (N=17990)

TCM physique	Qixu	Yingxu	Yangxu	Tanshi	Shire	Qiyu	Xueyu	Tebin	Pinghe
Qixu	0.80	0.05	0.01	0.06	0.00	0.04	0.02	0.00	0.01
Yingxu	0.30	0.52	0.01	0.06	0.03	0.04	0.02	0.00	0.02
Yangxu	0.36	0.06	0.49	0.06	0.00	0.01	0.01	0.00	0.00
Tanshi	0.34	0.08	0.02	0.45	0.02	0.04	0.02	0.00	0.01
Shire	0.28	0.11	0.01	0.07	0.43	0.05	0.03	0.00	0.03
Qiyu	0.40	0.06	0.01	0.06	0.01	0.42	0.02	0.00	0.01
Xueyu	0.35	0.09	0.01	0.04	0.01	0.03	0.45	0.00	0.02
Tebin	0.33	0.10	0.00	0.05	0.05	0.05	0.00	0.29	0.14
Pinghe	0.13	0.10	0.00	0.02	0.02	0.00	0.00	0.00	0.73

VGG-16. The reason for the poor performance of ResNet-101 is that the tongue physique identification database is not large enough, which is not conducive to learning residual errors. Secondly, it is generally hard to find the optimal parameter values for the deep learning method. Although we have cost much time to seek them, we may fail to find the optimal parameters. Finally ResNet101 model maybe do not fit our problem.

Tables 8, 9, and 10 present the confusion matrices of VGG, GoogLeNet, and ResNet, respectively, from the tongue physique identification database. In the tables, the bold numbers in the diagonal are the accuracies. The other bold numbers are greater errors. It can be seen that in the three models the best results are obtained on the "qixu" and "pinghe" categories. ResNet obtains the highest recognition rate, up to 89%, on the "qixu" category, whereas VGG obtains the highest recognition rate, up to 73%, on the "pinghe" category. Second, for the three models the recognition rate of "tebin" is the lowest, because it has much fewer samples in the training database. In practice, there are far too few people with this kind of physique, so the training samples are difficult to collect. Third, the three models easily misclassified some physique types as the "qixu" type. The reason is that the "qixu"-type training samples comprise up to one-third of the total number of samples. Finally, the effectiveness of ResNet is not as good as GoogLeNet and VGG, because it is more seriously

Table 9 Confusion matrix of **GoogLeNet** on largest training database (N=17990)

TCM physique	Qixu	Yingxu	Yangxu	Tanshi	Shire	Qiyu	Xueyu	Tebin	Pinghe
Qixu	0.81	0.05	0.01	0.04	0.01	0.05	0.00	0.00	0.02
Yingxu	0.34	0.50	0.01	0.05	0.04	0.05	0.00	0.00	0.02
Yangxu	0.35	0.10	0.49	0.02	0.01	0.03	0.00	0.00	0.01
Tanshi	0.38	0.09	0.00	0.44	0.02	0.05	0.01	0.00	0.02
Shire	0.31	0.11	0.01	0.04	0.46	0.05	0.01	0.00	0.01
Qiyu	0.39	0.05	0.00	0.04	0.04	0.46	0.00	0.00	0.01
Xueyu	0.35	0.06	0.01	0.04	0.01	0.06	0.41	0.01	0.04
Tebin	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00
Pinghe	0.12	0.09	0.00	0.02	0.03	0.02	0.00	0.00	0.72

Table 10 Confusion matrix of **ResNet** on largest training database (N=17990)

TCM physique	Qixu	Yingxu	Yangxu	Tanshi	Shire	Qiyu	Xueyu	Tebin	Pinghe
Qixu	0.89	0.05	0.01	0.01	0.02	0.01	0.00	0.00	0.02
Yingxu	0.45	0.44	0.02	0.01	0.03	0.01	0.00	0.00	0.03
Yangxu	0.45	0.06	0.44	0.01	0.02	0.00	0.01	0.00	0.01
Tanshi	0.58	0.07	0.02	0.26	0.04	0.00	0.00	0.00	0.02
Shire	0.43	0.10	0.01	0.02	0.40	0.02	0.00	0.00	0.03
Qiyu	0.59	0.09	0.01	0.02	0.03	0.25	0.00	0.00	0.01
Xueyu	0.58	0.03	0.02	0.01	0.04	0.01	0.31	0.00	0.01
Tebin	0.44	0.06	0.00	0.00	0.17	0.00	0.06	0.28	0.00
Pinghe	0.18	0.08	0.01	0.01	0.04	0.01	0.00	0.00	0.68

influenced by the "qixu" type than the other two models. This illustrates that if the interference of the "qixu" type can be mitigated or eliminated, the effectiveness of the model can be greatly improved.

5 Conclusions and future work

This paper proposes a new method to identify TCM physiques through natural tongue images using hybrid deep learning methods. It has several significant advantages. First, it can perfectly perform tongue coating detection on the tongue images captured in a natural environment instead of in a controlled one. Second, it applies deep learning methods to recognize the physique on the detected tongue coating image with good performance. Finally, it performs tongue physique identification quickly, within 1 min, while most systems, such as those based on a questionnaire, perform the same task in up to 30 min. These three advantages make it applicable to many practical fields. In the future, the training database should be enlarged as much as possible, since deep learning methods can obtain better results on larger training databases. Our planned follow-up work is multi-label learning for tongue physique identification; since each person's physique is complex, an individual may contain several physique types simultaneously.

Acknowledgements This study was supported by China National Science Foundation (Grant Nos. 60973083 and 61273363), Science and Technology Planning Project of Guangdong Province (Grant Nos. 2014A010103009 and 2015A020217002), and Guangzhou Science and Technology Planning Project (Grant No. 201504291154480,201803010088). We also thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

1. Deng T (2002) Treatment based on syndrome differentiation is the soul of clinical medicine of traditional Chinese medicine. Chin Arch Tradit Chin Med 20(4):394–395. https://doi.org/10.3969/j.issn.1673-7717.2002.04.003

2. Du JQ, Lu YS, Zhang K, Zhu MF, Ding CH (2008) A novel approach of tongue body and tongue coating separation based on FCM. In: 2008 2nd international conference on bioinformatics and biomedical engineering, Shanghai, 16–18 May, 2008, pp 2499–2503. https://doi.org/10.1109/ICBBE.2008.958
3. He K, Zhang X, Ren S, Sun J (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision – ECCV 2014. Lecture notes in computer science, vol 8691. Springer, Cham, pp 346–361. https://doi.org/10.1007/978-3-319-10578-9_23
4. Hu MC, Cheng MH, Lan KC (2016) Color correction parameter estimation on the smartphone and its application to automatic tongue diagnosis. J Med Syst Jan 40(1):18. https://doi.org/10.1007/s10916-015-0387-z
5. Huang B, Wu J, Zhang D, Li N (2010) Tongue shape classification by geometric features. Inf Sci (Ny) 180(2):312–324. https://doi.org/10.1016/j.ins.2009.09.016
6. Ini R, Siio I (2013) TongueDx: a tongue diagnosis system for personal health care on smartphone. Human Interface (Japan). http://hdl.handle.net/10083/57050
7. Institute of Chinese Medicine (2009) ZYYXH/T157-2009 Chinese medicine constitution classification and decision. Chinese Traditional Chinese Medicine Press, Beijing (In Chinese)
8. Kamarudin ND, Ooi CY, Kawanabe T, Mi X (2016) Tongue's substance and coating recognition analysis using HSV color threshold in tongue diagnosis. In: Proceedings of SPIE 10011, first international workshop on pattern recognition, 100110J (11 July 2016). https://doi.org/10.1117/12.2242404
9. Kanawong R (2012) Computer-aided tongue image diagnosis and analysis. Dissertation, University of Missouri at Columbia
10. Kanawong R, Xu W, Xu D, Li S, Ma T, Duan Y (2012) An automatic tongue detection and segmentation framework for computer-aided tongue imag+e analysis. Int J Funct Inform Personal Med 4(1):56–68. https://doi.org/10.1504/IJFIPM.2012.050420
11. Kim KH, Do JH, Ryu H, Kim JY (2008) Tongue diagnosis method for extraction of effective region and classification of tongue coating. In: 2008 first workshops on image processing theory, tools and applications, Sousse, 23–26 Nov. 2008, pp 1–7. https://doi.org/10.1109/IPTA.2008.4743772
12. Ko MM, Park TY, Lee JA, Kang BK, Lee J, Lee MS (2013) A study of tongue and pulse diagnosis in traditional Korean medicine for stroke patients based on quantification theory type II. Evid Based Complement Alternat Med 2013:508918. https://doi.org/10.1155/2013/508918
13. Lee J, Ko M, Park T, Kang B, Moon T, Lee M (2012) P02.79. Study on the basic tongue diagnosis indicators for pattern identifications in stroke using a decision tree method. BMC Complement Altern Med 12(S1):P135. https://doi.org/10.1186/1472-6882-12-S1-P135
14. Li Q, Liu Z (2009) Tongue color analysis and discrimination based on hyperspectral images. Comput Med Imaging Graph 33(3):217–221. https://doi.org/10.1016/j.compmedimag.2008.12.004
15. Li FF, Zhao C, Xia Z et al (2012) Computer-assisted lip diagnosis on traditional Chinese medicine using multi-class support vector machines. BMC Complementary Alternative Med 12:127. http://www.biomedcentral.com/1472-6882/12/127
16. Li H, Lin Z, Shen X, Brandt J, Hua G (2015) A convolutional neural network cascade for face detection. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, 7–12 June, 2015, pp 5325–5334. https://doi.org/10.1109/CVPR.2015.7299170
17. Lin F, Xiahou J, Xu Z (2016) TCM Clinic records data mining approaches based on weighted-LDA and multi-relationship LDA model. Multimed Tools Appl 75(22):14203–14232. https://doi.org/10.1007/s11042-016-3363-9
18. Liu B, Zhang Z (2004) The ancient method of syndrome differentiation of thinking. Chin J Basic Med Tradit Chin Med 10(5):325–331. https://doi.org/10.3969/j.issn.1006-3250.2004.05.002
19. Lo LC, Chen YF, Chen WJ, Cheng TL, Chiang JY (2012) The study on the agreement between automatic tongue diagnosis system and Traditional Chinese Medicine practitioners. Evid Based Complement Alternat Med 2012:505063. https://doi.org/10.1155/2012/505063
20. Luo D, Wen G, Li D, Hu Y, Huan E (2018) Iterative regression based Face Detection using deep learning. Multimed Tools Appl 2018(6):1–18
21. Ning J, Zhang L, Zhang D, Wu C (2010) Interactive image segmentation by maximal similarity based region merging. Pattern Recognit 43(2):445–456. https://doi.org/10.1016/j.patcog.2009.03.004
22. Pang B, Zhang D, Li N, Wang K (2004) Computerized tongue diagnosis based on bayesian networks. IEEE Trans Biomed Eng 51(10):1803–1810. https://doi.org/10.1109/TBME.2004.831534
23. Pang B, Zhang D, Wang K (2005) The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine. IEEE Trans Med Imaging 24(8):946–956. https://doi.org/10.1109/TMI.2005.850552
24. Ren S, He K, Girshick R, Sun J (2017) Faster r-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

25. Shi M, Li G, Li F (2013) C2g2FSnake: automatic tongue image segmentation utilizing prior knowledge. Sci China Inform Sci 56(9):1–14. https://doi.org/10.1007/s11432-011-4428-z
26. Shi MJ, Li GZ, Li FF, Xu C (2014) Computerized tongue image segmentation via the double geo-vector flow. Chin Med 9:7. https://doi.org/10.1186/1749-8546-9-7
27. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v1 [cs.CV]
28. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, 7–12 June 2015, pp 1–9. https://doi.org/10.1109/CVPR.2015.7298594
29. Wang X, Zhang D (2010) An optimized tongue image color correction scheme. IEEE Trans Inf Technol Biomed 14(6):1355–1364. https://doi.org/10.1109/TITB.2010.2076378
30. Wang X, Zhang D (2013) A high quality color imaging system for computerized tongue image analysis. Expert Syst Appl 40(15):5854–5866. https://doi.org/10.1016/j.eswa.2013.04.031
31. Wang X, Zhang B, Yang Z, Wang H, Zhang D (2013) Statistical analysis of tongue images for feature extraction and diagnostics. IEEE Trans Image Process 22(12):5336–5347. https://doi.org/10.1109/TIP.2013.2284070
32. Wu K, Zhang D (2015) Robust tongue segmentation by fusing region-based and edge-based approaches. Expert Syst Appl 42(21):8027–8038. https://doi.org/10.1016/j.eswa.2015.06.032
33. You M, Ge L, Li GZ, Xu L, Huang S (2009) A TCM platform for maters' experience sharing. In: International joint conference on bioinformatics, systems biology and intelligent computing (IJCBS09). IEEE Computer Society, Washington, DC, pp 388–391
34. Zhang J, Hu G, Zhang X (2015) Extraction of tongue feature related to TCM physique based on image processing. In: 2015 12th international computer conference on wavelet active media technology and information processing (ICCWAMTIP), Chengdu, China, 18–20 December 2015, pp 251–255. https://doi.org/10.1109/ICCWAMTIP.2015.7493986
35. Zhang B, Kumar BV, Zhang D (2014) Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features. IEEE Trans Biomed Eng 61(2):491–501. https://doi.org/10.1109/TBME.2013.2282625
36. Zhang Q, Shang HL, Zhu JJ, Jin MM, Wang WX, Kong QS (2013) A new tongue diagnosis application on Android platform. In: 2013 IEEE international conference on bioinformatics and biomedicine (BIBM), 18–21 December 2013, Shanghai, China, pp 334–327. https://doi.org/10.1109/BIBM.2013.6732705
37. Zhuo L, Zhang J, Dong P, Zhao Y, Peng B (2014) An SA–GA–BP neural network-based color correction algorithm for TCM tongue images. Neurocomputing 134:111–116. https://doi.org/10.1016/j.neucom.2012.12.080
38. Zhuo L, Zhang P, Qu P, Peng Y, Zhang J, Li X (2016) A k-PLSR-based color correction method for TCM tongue images under different illumination conditions. Neurocomputing 174:815–821. https://doi.org/10.1016/j.neucom.2015.10.008

Huihui Li received the M.S degree in South China University of Technology. She is currently working towards the Ph. D degree from the Department of Computer Science&Technology of South China University of Technology. Her research area includes facial expression recognition, artificial intelligence and Machine Learning in Traditional Chinese Medicine.

Guihua Wen is now professor, doctoral supervisor at the school of computer science &technology of South China University of Technology. His research area includes Cognitive & affective computing, Machine Learning and data mining.

Haibin Zeng received the M.S degree in South China University of Technology. His research area includes Cognitive & affective computing, Machine Learning and data mining.

Multimed Tools Appl (2019) 78:6847–6868
https://doi.org/10.1007/s11042-018-6279-8

Natural tongue physique identification using hybrid deep learning methods

Huihui Li[1,2] · Guihua Wen[1,2] · Haibin Zeng[1,3]

Received: 22 December 2017 / Revised: 30 May 2018 / Accepted: 15 June 2018 /
Published online: 31 July 2018

Abstract Traditional Chinese Medicine (TCM) illustrates that the physique determines the susceptibility of human to certain diseases and treatment programs for illness. Tongue diagnosis is an important way to identify the physique, but now it is performed by the doctor's professional experience and the design of a questionnaire. Consequently, accurate physique identification cannot be obtained easily. In this paper, we propose a new method to identify the physique through wild tongue images using hybrid deep learning methods. It begins with constructing a large number of tongue images that are taken in natural conditions, instead of in a controlled environment. Based on the resulting database, a new method of tongue coat-

Published: 31 July 2018

Natural tongue physique identification using hybrid deep learning methods

Huihui Li, Guihua Wen & Haibin Zeng

Multimedia Tools and Applications **78**, 6847–6868 (2019) | Cite this article

Abstract

Traditional Chinese Medicine (TCM) illustrates that the physique determines the susceptibility of human to certain diseases and treatment programs for illness. Tongue diagnosis is an important way to identify the physique, but now it is performed by the doctor's professional experience and the design of a questionnaire. Consequently, accurate physique identification cannot be obtained easily. In this paper, we propose a new method to identify the physique through wild tongue images using hybrid deep learning methods. It begins with constructing a large number of tongue images that are taken in natural conditions, instead of in a controlled environment. Based on the resulting database, a new method of tongue coating detection is put forward that applies a rapid deep learning method to complete the initial tongue coating

International Conference on Brain Inspired Cognitive Systems

BICS 2019: Advances in Brain Inspired Cognitive Systems pp 96-105 | Cite as

Real-Time Visual Tracking Base on SiamRPN with Generalized Intersection over Union

Authors Authors and affiliations

Zhihui Huang, Jin Zhan, Huimin Zhao ✉, Kaihan Lin, Penggen Zheng, Jujian Lv

Conference paper

First Online: 01 February 2020

Part of the Lecture Notes in Computer Science book series (LNCS, volume 11691)

Abstract

Deep learning-based tracking methods have great challenges to handle larger training data with aiming to be invariant to all sorts of appearance variations. In this paper, we incorporate a novel Generalized intersection over union (GIOU) as bounding box regression loss into Siamese framework based tracker, and propose a visual tracking method based on Siamese region proposal network (SiamRPN) with generalized intersection over union. We set out to bridge the

Robust Visual Tracking via Sparse Feature Selection and Weight Dictionary Update

Penggen Zheng[1,2], Jin Zhan[1,2], Huimin Zhao[1,2(✉)], and Hefeng Wu[3]

[1] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China
penggengg@gmail.com,
{gszhanjin, zhaohuimin}@gpnu.edu.cn
[2] Guangzhou Key Laboratory of Digital Content Processing and Security Technologies, Guangzhou, China
[3] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China
wuhefeng@gmail.com

Abstract. Sparse representation-based visual tracking methods do not adapt well to changes in the target and backgrounds, and the sparseness of samples does not guarantee optimality. In this paper, we propose a robust visual tracking algorithm using sparse multi-feature selection and adaptive dictionary update based on weight dictionaries. We exploit the color features and texture features of the learning samples to obtain different discriminative dictionaries based on the label consistent K-SVD algorithm, and use the position information of those samples to assign weights to the dictionaries' base vectors, forming the weight dictionaries. For robust visual tracking, we adopt a novel feature selection strategy that combines the weights of dictionaries' base vectors and reconstruction errors to select the best sample. In addition, we introduce adaptive noise energy thresholds and establish a dictionary updating mechanism based on noise energy analysis, which effectively reduces the error accumulation caused by dictionary updating and enhances the adaptability to target and background changes. Comparison experiments show that the proposed algorithm performs favorably against several state-of-the-art methods.

Keywords: Visual tracking · Similarity weights · Sparse representation
Adaptive update · Multi-feature selection

1 Introduction

Visual tracking is an important problem of computer vision. In the past few years, we have witnessed rapid advancements in visual tracking, but it is still a challenging task due to complex situations, such as occlusions, target deformation, rotation, scale changes and cluttered background. Most of existing methods can be roughly divided into generative methods [1–4] and discriminative methods [5–9]. The generative methods describe the appearance characteristics of the target and search for candidate targets by minimizing the reconstruction error. Based on classifiers, the discriminative methods mainly find the decision boundaries of the target and the background.

J. Ren et al. (Eds.): BICS 2018, LNAI 10989, pp. 484–494, 2018.
https://doi.org/10.1007/978-3-030-00563-4_47

In recent years, there have been many target tracking algorithms based on sparse representation [1–3], correlation filters (CF) [10–13] and deep learning [14–17], where deep learning and correlation filters are research hotspots now. The main advantage of deep learning-based tracking methods lies in their powerful characterization of depth features. SO-DLT [16] and MDnet [17] use non-task video datasets for pre-training, and then adjust the tracking model to make the model adapt to the current tracking task. However, these methods are computationally intensive and depend strongly on pre-trained samples. CF-based tracking methods show strong computational efficiency and tracking robustness. Heriques et al. [10] proposed an efficient tracking method based on HOG features using cyclic transform and kernel transform. Danelljan et al. [11, 12] later used multi-scale models to solve the problem of target scale change. The correlation filters have a great advantage in tracking efficiency. However, the tracking effect is easily affected by the boundary effect, and the target background information cannot be fully utilized.

Sparse representation-based tracking methods [1–3] select the target location by comparing the reconstruction errors of the features. Because of insensitivity to the target noise, this kind of method has a strong tracking robustness when target deformation occurs. Based on the Label Consistent K-SVD (LC-KSVD) method [18], the work [19] used the positive and negative samples together to train a discriminative sparse dictionary, making the model have stronger discriminative performance during the tracking process. However, fixed discriminative dictionary ignores the variations cues between foreground and background. Also, it used single-feature which lacks of considering color distribution of target.

The main innovations of our method are as follows:

- **Sparse Multi-feature Selection.** According to the center distance from the learning sample to the target, we assign Gaussian weights to the basis vectors of different feature dictionaries. We use the multi-feature weight of the samples to measure the similarity between the samples and the target to obtain candidate samples, and then to select the best sample by synthesizing the multiple features reconstruction error of them. The complementary effect between the various features improves the stability of the tracking.
- **Adaptive Weight Dictionary Update.** We analyze the changes of the average noise energy during tracking. In order to select the best update time for the dictionary, we use the quantile threshold of all previous average noise energies to determine the anomalous changes in the tracking scene.

2 Proposed Approach

The overall framework of our approach is shown in Fig. 1. It can be divided into two parts in general. (1) *Update-related*. In frame t, we set a threshold for the noise energy of the tracking results based on all previous tracking results, combined with the target noise and samples average noise to determine whether the update conditions are satisfied. The updated dictionary is learned from the positive and negative samples of frame t and the positive sample of frame 1, and used for the tracking detection of the

next frame. (2) *Selection-related*. In frame $t + 1$, we perform sparse coding on samples obtained from Gaussian sampling using the latest weight dictionary, and then compare sample similarity weights and reconstruction errors to select the best sample.

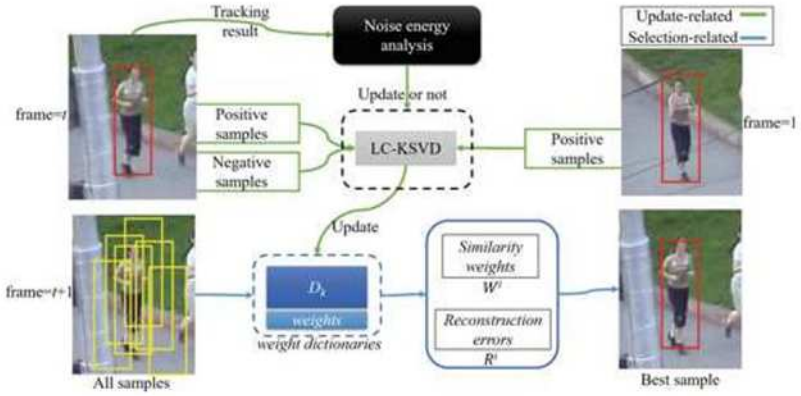


Fig. 1. The tracking flowchart of the proposed approach.

2.1 Weight Dictionary Model

Discriminative Dictionary Composition. In this paper, we use three types of templates (the target template T, the background template B and the noise template I) to collectively represent the target appearance. In the process of template sampling in first frame, all pixels in the range of radius r_0 are sampled to obtain the positive samples. And then dense sampling is performed in the range of radius between r_1 and r_2 to obtain negative samples. If the dimension of a sample feature is m, then the target templates and the background templates define as:

$$T = \{F_i | d(i) \leq r_0\}^{m \times p}, \; B = \{F_i | r_1 < d(i) \leq r_2\}^{m \times q}, \tag{1}$$

where F_i represents the sample i, $d(i)$ is the center distance from sample F_i to the target, i is the label of the samples, p and q are the number of positive samples and negative samples. The three radiuses in Eq. (1) satisfy the relationship: $0 < r_0 \leq r_1 < r_2$.
The sparse discriminative dictionary D learned from all samples consists of three parts: $D = [D^T, D^B, D^I]$, then D^T, D^B and D^I respectively corresponding to template T, B and I. In the tracking process, the sparse representation formula for the feature y is:

$$y \approx D\gamma = \left[D^T, D^B, D^I\right] \begin{bmatrix} z \\ v \\ e \end{bmatrix}, \tag{2}$$

where D is a discriminative dictionary, z is target coefficients, v is background coefficients, e is noise coefficient, and γ is sparse coding.

Weight Dictionary Learning. In this paper, the LC-KSVD [18] method is used to unify dictionary learning and classification labeling. The solution to Eq. (2) can be

transformed into solving four local dictionaries respectively. The solution process is expressed as follows:

$$arg \min_{D,A,\gamma} \|Y - D\gamma\|_2^2 + \beta\|G - A\gamma\|_2^2 + \lambda\|\gamma\|_1. \tag{3}$$

Matrix G is the discriminative sparse coding of the initial template classification, so that γ approximates the sparse coding of the initial label, and the transfer matrix A makes γ in the sparse feature space has stronger discriminant ability of samples category. β is a range control coefficient that is consistent with the regular term contribution in Formula (3). According to the samples category of the target template T, the background template B and the noise template I, matrix G is defined as follows:

$$G = \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix}, g_1 \in R^{p\times p}, g_2 \in R^{(q+m)\times(q+m)}, \tag{4}$$

where g_1 and g_2 are all-one-element matrices. In order to obtain the optimal solution of Formula (3), we can solve the following expression:

$$arg \min_{D_{new},\gamma} \|Y_{new} - D_{new}\gamma\|_2^2 + \lambda\|\gamma\|_1, \tag{5}$$

where $Y_{new} = \left(Y^T, \sqrt{\beta}G^T\right)^T$, $D_{new} = \left(D^T, \sqrt{\beta}A^T\right)^T$. Formula (5) can be solved using the K-SVD algorithm. The learning process of the dictionary D_{new} will generate a sparse coding value γ, and we can obtain the dictionary D.

When we have got the sparse dictionary, we assign weights to each base vector of the dictionary, according to the center distance from the samples to the target. We use the Gaussian function centered on the target as the weight function, and the weight of the dictionary's base vector is only related to the center distance $d(i)$. The weight of the base vector is defined as follows:

$$W(i) = exp(-d^2(i)/2\sigma^2), \tag{6}$$

where σ is the standard deviation of normal distribution. This weight reflects the similarity between the target and the samples. Finally, we can get a discriminative dictionary with a weight table, namely the weight dictionary, as shown in Fig. 2.

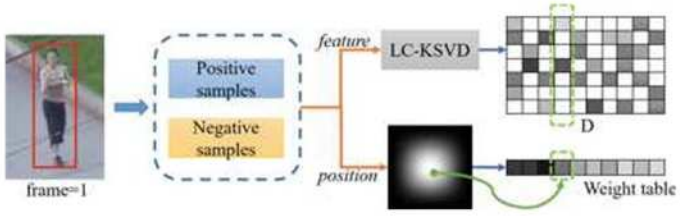


Fig. 2. Initial dictionary learning and weight assignment of the dictionary's base vectors

2.2 Multi-feature Selection Strategy

In the frame t, we perform Gaussian sampling centered on the target position of the previous frame to obtain samples S_i ($i = 1, 2, \ldots, n$). Using two feature dictionaries D_k ($k = 1, 2$) to perform sparse decomposition on all samples, we can get sparse coefficients γ_k^i, where k is a feature tag ($k = 1$ denotes color histogram, $k = 2$ denotes Haar-like features [20]) and i is the sample index.

In the sparse coefficients γ_k^i, all coefficient values reflect the correlation strength between the sample feature and the dictionary's base vector. A base vector (sparse feature) corresponding to the maximum value has the strongest correlation with the sample. In the tracking process, we combine the weights of two features to get the synthetic weights W^i. Its definition is as follows:

$$W^i = \prod_k W_k^i, \tag{7}$$

where W_k^i represents the *k-feature* weight of sample s_i. The candidate samples CS_j are the samples with the largest feature synthetic weight among all samples.

Then we use the synthetic reconstruction error to select the best sample from candidate samples. The expression of the synthetic reconstruction error is as follows:

$$R^j = \prod_k R_k^j, \tag{8}$$

where R_k^j represents the reconstruction error of the sample s_j in *k-feature*, j is the label of the candidate samples. Finally, in the candidate samples, the sample label of the best tracking result can be expressed as follows:

$$\hat{j} = arg\, min_j\, R^j. \tag{9}$$

The multi-feature selection process is shown in Fig. 3.

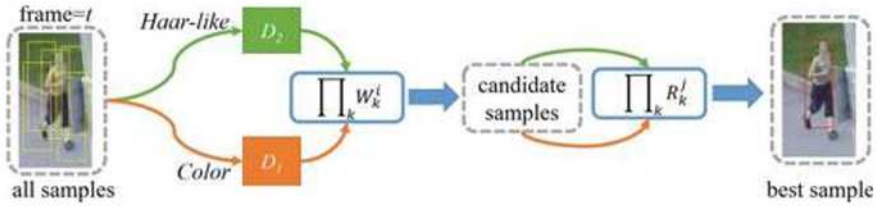


Fig. 3. Multi-feature selection process

2.3 Adaptive Dictionary Update

In the sparse coefficient γ_k^i, the maximum sparse coefficient represents most information of the sample and the noise factor e (see Eq. (2)) is much smaller than the maximum coefficient. Noise coefficient e can reflect the situation of target occlusion and tracking

drift to some extent. We should not update the dictionary for these two situations. Our method analyzes the noise energy u (the sum of the noise coefficients e) to determine the time of the dictionary update.

In frame t, the average noise energy expression for all samples in the current frame is $\overline{u_k^T} = \sum_i u_k^i$, where u_k^i is the noise energy of sample s_i. The threshold x_k^α is the upper quantile of set U_k (all $\overline{u_k^T}$ from frame 1 to frame t). The threshold is defined as:

$$P\{U_k > x_k^\alpha\} = \alpha. \tag{10}$$

We only perform the weight dictionary update when the following three conditions are met. (1) The noise energy of the current best sample is less than the average noise energy of all samples; (2) The average noise energy change curve intersects the threshold curve at the last two frames; (3) The interval time between two updates must be more than tm. The update result is shown in Fig. 4.

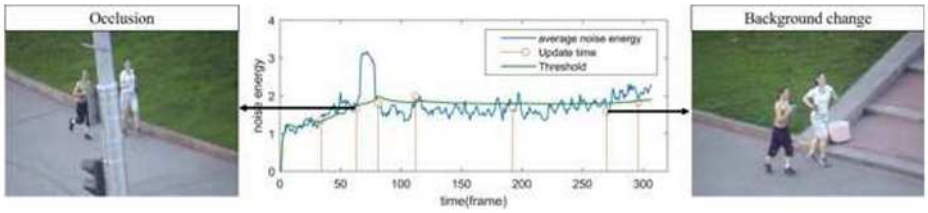


Fig. 4. Update time and corresponding scene

The learning process of the new dictionary is shown in Fig. 1. We sample the positive and negative samples in the current frame and form a new training dataset with the positive samples of the first frame. We use this new training dataset to learn a new dictionary based on LC-KSVD method [19] and replace the old one.

3 Experimental Results and Comparison

3.1 Experimental Settings

In this paper, 14 representative video sequences in the OTB100 datasets [21] are selected as the evaluation test set of the algorithm. These standard test videos are: basketball, Panda, Skater, Shaking, jumping, football, BlurFace, BlurOwl, couple, Man, jogging-1, david3, david2, carDark. These video clips cover a variety of challenging attributes, including target occlusion, target deformation, illumination changes, target rotation, background interference. In the comparative experiment, we not only selected 8 tracking models related to our method (ASLA [1], IVT [2], L1APG [3], CT [5], CXT [6], ORIA [8], OAB [7], TLD [9]), but also selected 9 excellent methods (DCF [10], KCF [10], DSST [11], SRDCF [12], LMCF [13], SiamFC [14], ACFN [15], SAMF [22], STAPLE [23]) in recent years including methods based on CF and deep-learning.

We use the open code and default parameters of these methods to experiment in the same computer environment. The computer environment used in the experiment is: Intel (R) Core (TM) i3-3.7 GHz, RAM-12 GB, matlabR2017a.

The method of this paper adopts uniform parameter settings. The number of Gaussian samples in the tracking process is 500 and the sampling radius is 25. The sampling parameter of the training sample is set to: $r_0 = 4$, $r_1 = 7$, $r_2 = 15$. The Haar-like [20] feature dimension is set to 150, and the bin of the color statistics histogram is set to 32. So the color-feature dimension of the single-channel image is 32, and the RGB image is 96. Update time interval is $tm = 6$ frames, and noise energy threshold parameter is $\alpha = 0.2$.

3.2 Experimental Results and Analysis

We use the average center error and the average overlap score [21] as evaluation criteria for comparative experiments. The center error is the Euclidean distance between the tracking result and the standard target position. The overlap rate is a measure of the overlap range of the tracking result and the standard result tracking box. The calculation formula is defined as: $score = \frac{intersection\ area}{union\ area}$. The average center error and overlap rate for our method and the related methods are shown in Tables 1 and 2.

Table 1. The average overlap rate of the related methods and our method on 14 different videos (bold indicates the best two methods). DSST and KCF were added as evaluation baseline.

Video	Ours	DSST	KCF	ASLA	IVT	OAB	TLD	L1APG	ORIA	CT	CXT
Basketball	**0.715**	0.606	**0.676**	0.277	0.123	0.026	0.310	0.202	0.065	0.212	0.019
Panda	0.512	0.142	0.159	**0.518**	0.138	0.152	**0.539**	0.285	0.109	0.489	0.182
Skater	**0.632**	**0.629**	0.611	0.548	0.551	0.532	0.542	0.567	0.597	0.597	0.426
Shaking	**0.668**	**0.706**	0.040	0.427	0.034	0.034	0.101	0.276	0.468	0.171	0.308
Jumping	**0.652**	0.070	0.275	0.086	0.122	0.059	0.646	0.105	0.082	0.031	**0.773**
Football	**0.628**	0.560	0.552	**0.590**	0.560	0.529	0.542	0.462	0.512	0.403	0.538
BlurFace	0.772	**0.841**	0.796	0.218	0.158	0.481	**0.864**	0.450	0.176	0.221	0.835
BlurOwl	**0.743**	0.190	0.195	0.194	0.055	**0.768**	0.610	0.253	0.069	0.067	0.312
Couple	**0.594**	0.092	0.201	0.100	0.074	0.420	0.188	**0.471**	0.047	0.297	0.467
Man	**0.878**	**0.883**	0.831	0.836	0.745	0.859	0.793	0.874	0.848	0.074	0.873
Jogging-1	0.701	0.187	0.186	0.186	0.177	0.612	**0.767**	0.184	0.223	0.160	**0.766**
David3	**0.657**	0.458	**0.772**	0.314	0.512	0.286	0.276	0.292	0.132	0.400	0.122
David2	0.698	0.830	0.828	**0.899**	0.702	0.754	0.782	0.839	0.468	0.003	**0.854**
carDark	0.587	0.845	0.615	**0.854**	0.663	0.767	0.446	**0.852**	0.419	0.007	0.540

In Tables 1 and 2, we can see that our method show favorably performance in comparison with the methods based on dense sampling or sparse representation.

Table 2. The average center error of the related methods and our method on 14 different videos (bold indicates the best two methods). DSST and KCF were added as evaluation baseline.

Video	Ours	DSST	KCF	ASLA	IVT	OAB	TLD	L1APG	ORIA	CT	CXT
Basketball	**6.9**	10.9	**7.9**	71.2	81.5	160.9	65.2	111.9	110.7	112.9	188.3
Panda	**6.8**	43.6	42.1	6.9	51.2	143.2	**5.9**	26.9	69.9	6.9	81.4
Skater	8.8	**8.4**	10.7	**7.5**	8.5	20.7	11.3	16.1	11.7	11.8	40.5
Shaking	**10.6**	**8.4**	112.5	27.0	87.2	155.3	68.5	37.9	24.4	65.0	122.1
Jumping	**6.2**	36.9	26.1	49.5	61.6	69.6	7.6	57.6	57.7	52.0	**3.5**
Football	**9.0**	15.8	14.6	15.0	13.9	15.8	14.3	41.3	**13.2**	19.3	14.3
BlurFace	10.9	**5.2**	8.4	117.2	148.9	53.6	**3.8**	62.8	75.6	109.9	6.5
BlurOwl	**8.9**	196.1	183.4	64.6	167.2	11.8	30.3	120.6	179.4	164.7	**6.9**
Couple	**8.9**	125.6	47.6	93.9	109.4	**26.5**	64.3	28.4	98.8	29.3	49.2
Man	1.8	1.6	2.3	**1.3**	3.4	2.3	3.1	**1.4**	1.9	45.8	2.1
Jogging-1	**6.9**	110.7	88.3	104.0	88.3	13.7	7.2	83.3	45.2	91.8	**5.6**
David3	**12.0**	88.2	**4.3**	104.6	53.0	91.1	135.7	86.0	182.6	68.5	221.8
David2	4.8	2.0	2.1	1.5	**1.2**	3.5	2.6	1.5	19.5	78.2	**1.3**
carDark	7.6	1.5	6.0	**1.4**	8.4	3.5	26.9	**1.2**	25.9	118.7	18.6

In order to objectively evaluate our approach, we also use the tracking precision and success rate [21] under different thresholds to compare our method with 17 different methods (ASLA [1], IVT [2], L1APG [3], CT [5], CXT [6], ORIA [8], OAB [7], TLD [9], DCF [10], KCF [10], DSST [11], SRDCF [12], LMCF [13], SiamFC [14], ACFN [15], SAMF [22], STAPLE [23]). The overall tracking accuracy and success rate for all test videos is shown in Fig. 5. In comparison with some advanced methods in recent years, our method also shows excellent tracking performance.

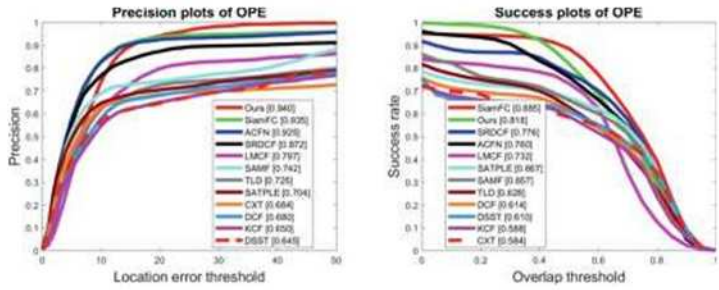


Fig. 5. Precision and success plots of overall performance comparison for 14 videos

The tracking performance evaluation of different video attributes is showed in Fig. 6. Our tracking performance is second (only below SiamFC) on attribute *deformation*, *rotation* and *scale variation*, and we achieve the best performance on attribute *background clutter* (see Fig. 6).

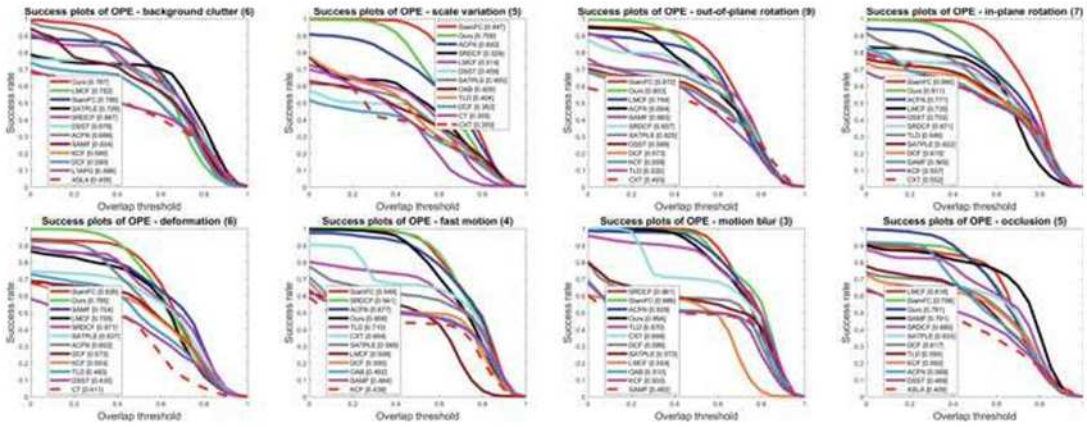


Fig. 6. Attribute based evaluation. Success plots compare our tracker with other 17 trackers on all test videos. Our tracker has favorably tracking effects in most attributes.

The color features have properties that are insensitive to the deformation of the target. The use of color feature dictionary greatly enhances the tracking effect of our tracking model when the target is blurred. In comparison with the related methods, our tracker has the best tracking effect on attribute *motion blur* and *fast motion* (see Fig. 6). The tracking results in videos *BlurOwl*, *jumping* and *couple* are shown in Fig. 7.

Fig. 7. Tracking results of 6 related methods and our method in videos BlurOwl, jumping and couple

From the comparison of these results we can see that the tracking effect of our method has reached the level of some state-of-the-art methods (e.g. LMCF, ACFN and SRDCF).

4 Conclusion

This paper proposes a novel weight dictionary model with multi-feature selection and adaptive dictionary learning to represent the appearance of the target, and noise energy analysis and sample similarity weights are introduced to improve the performance of sparse feature selection. Using different features can complement the feature representation capabilities of the target and reduce background noise interference to increase tracking accuracy. The sample similarity weight is used to narrow the search range of the optimal sample to select candidate samples, which can effectively reduce the interference of abnormal samples. A new updating algorithm has been proposed via noise energy analysis which compares the noise energy with the dynamic threshold.

Experimental results on the OTB100 dataset show the effectiveness of the proposed method. In the future, we will further solve the issue of scale change, improve our algorithm by investigating more sophisticated sparse representation-based methods [24], and extend our method to multi-camera target tracking application scenarios [25].

Acknowledgment. This research is supported by National Natural Science Foundation of China (61772144, 61672008), Innovation Research Project of Education Department of Guangdong Province (Natural Science) (2016KTSCX077), Foreign Science and Technology Cooperation Plan Project of Guangzhou Science Technology and Innovation Commission (201807010059), Guangdong Provincial Application-oriented Technical Research and Development Special Fund Project (2016B010127006), the Natural Science Foundation of Guangdong Province (2016A030311013), and the Scientific and Technological Projects of Guangdong Province (2017A050501039). The corresponding authors are Jin Zhan and Huimin Zhao.

References

1. Lu, H., Jia, X., Yang, M.H.: Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 1822–1829. IEEE (2012)
2. Ross, D.A., Lim, J., Lin, R.S., et al.: Incremental learning for robust visual tracking. Int. J. Comput. Vis. (IJCV) **77**(1–3), 125–141 (2008)
3. Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 1830–1837 (2012)
4. Liu, Q.: Decontaminate feature for tracking: adaptive tracking via evolutionary feature subset. J. Electron. Imaging **26**(6), 1 (2017)
5. Zhang, K., Zhang, L., Yang, M.H.: Real-time compressive tracking. In: Computer Vision (ECCV), pp. 864–877 (2012)
6. Dinh, T.B., Vo, N., Medioni, G.: Context tracker: Exploring supporters and distracters in unconstrained environments. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 1177–1184. IEEE (2011)
7. Grabner, H., Bischof, H.: On-line boosting and vision. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 260–267 (2006)
8. Ling, H.: Online robust image alignment via iterative convex optimization. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 1808–1814. IEEE (2012)
9. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1409–1422 (2012)
10. Henriques, J.F., Rui, C., Martins, P., et al.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2014)
11. Danelljan, M., Häger, G., Khan, F.S., et al.: Accurate scale estimation for robust visual tracking. In: British Machine Vision Conference, pp. 65.1–65.11 (2014)
12. Danelljan, M., Hager, G., Khan, F.S., et al.: Learning spatially regularized correlation filters for visual tracking. In: Proceedings of International Conference on Computer Vision (ICCV), et al, pp. 4310–4318. IEEE Computer Society (2015)
13. Wang, M., Liu, Y., Huang, Z.: Large margin object tracking with circulant feature maps. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 4800–4808. Honolulu, Hawaii (2017)

14. Bertinetto, L., Valmadre, J., Henriques, J.F., et al.: Fully-convolutional siamese networks for object tracking. In: Proceedings of European Conference on Computer Vision (ECCV), pp. 850–865. Springer, Cham (2016)
15. Choi, J., Chang, H.J., Yun, S., et al.: Attentional correlation filter network for adaptive visual tracking. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 4828–4837. IEEE (2017)
16. Wang, N., Li, S., Gupta, A., et al.: Transferring rich feature hierarchies for robust visual tracking. Comput. Sci. (2015)
17. Nam, H., Han, B.: Learning multi-domain convolutional neural networks for visual tracking. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 4293–4302, IEEE (2016)
18. Jiang, Z., Lin, Z., Davis, L.S.: Label consistent K-SVD: learning a discriminative dictionary for recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(11), 2651–2664 (2013)
19. Jin, Z., Su, Z., Wu, H., et al.: Robust tracking via discriminative sparse feature selection. Vis. Comput. **31**(5), 575–588 (2015)
20. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: International Conference on Image Processing, vol. 1, pp. I-900–I-903. IEEE (2002)
21. Wu, Y., Lim, J., Yang, M.H.: Object Tracking Benchmark. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1834–1848 (2015)
22. Li, Y., Zhu, J.: A scale adaptive kernel correlation filter tracker with feature integration. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8926, pp. 254–265. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16181-5_18
23. Bertinetto, L., Valmadre, J., Golodetz, S., et al.: Staple: complementary learners for real-time tracking. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 1401–1409. IEEE (2016)
24. Qiao, T., Yang, Z., Ren, J., et al.: Joint bilateral filtering and spectral similarity-based sparse representation: a generic framework for effective feature extraction and data classification in hyperspectral imaging. Pattern Recognit. (2017)
25. Ren, J., Orwell, J., Jones, G.A., et al.: Tracking the soccer ball using multiple fixed cameras. Comput. Vis. Image Underst. **113**(5), 633–642 (2009)

Robust Visual Tracking via Sparse Feature Selection and Weight Dictionary Update

Penggen Zheng[1,2], Jin Zhan[1,2], Huimin Zhao[1,2(✉)], and Hefeng Wu[3]

[1] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China
penggengg@gmail.com,
{gszhanjin, zhaohuimin}@gpnu.edu.cn

[2] Guangzhou Key Laboratory of Digital Content Processing and Security Technologies, Guangzhou, China

[3] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China
wuhefeng@gmail.com

Abstract. Sparse representation-based visual tracking methods do not adapt well to changes in the target and backgrounds, and the sparseness of samples does not guarantee optimality. In this paper, we propose a robust visual tracking algorithm using sparse multi-feature selection and adaptive dictionary update based on weight dictionaries. We exploit the color features and texture features of the learning samples to obtain different discriminative dictionaries based on the label consistent K-SVD algorithm, and use the position information of those samples to assign weights to the dictionaries' base vectors, forming the weight dictionaries. For robust visual tracking, we adopt a novel feature selection strategy that combines the weights of dictionaries' base vectors and reconstruction errors to select the best sample. In addition, we introduce adaptive noise energy thresholds and establish a dictionary updating mechanism based on noise energy analysis, which effectively reduces the error accumulation caused by dictionary updating and enhances the adaptability to target and background changes. Comparison experiments show that the proposed algorithm performs favorably against several state-of-the-art methods.

Keywords: Visual tracking · Similarity weights · Sparse representation
Adaptive update · Multi-feature selection

International Conference on Brain Inspired Cognitive Systems

BICS 2018: Advances in Brain Inspired Cognitive Systems pp 484-494 | Cite as

Robust Visual Tracking via Sparse Feature Selection and Weight Dictionary Update

Authors Authors and affiliations

Penggen Zheng, Jin Zhan, Huimin Zhao, Hefeng Wu

Conference paper

First Online: 06 October 2018

Part of the Lecture Notes in Computer Science book series (LNCS, volume 10989)

Abstract

Sparse representation-based visual tracking methods do not adapt well to changes in the target and backgrounds, and the sparseness of samples does not guarantee optimality. In this paper, we propose a robust visual tracking algorithm using sparse multi-feature selection and adaptive dictionary update based on weight dictionaries. We exploit the color features and texture features of the learning samples to obtain different discriminative dictionaries based on the

Cognitive Computation
https://doi.org/10.1007/s12559-019-09662-y

Salient Superpixel Visual Tracking with Graph Model and Iterative Segmentation

Jin Zhan[1] · Huimin Zhao[1] · Penggen Zheng[1] · Hefeng Wu[2,3] · Leijun Wang[1]

Received: 29 December 2018 / Accepted: 13 June 2019

Abstract

Visual object tracking is to locate an object of interest in a sequence of consecutive video frames, which is widely applied in many high-level computer vision tasks such as intelligent video surveillance and robotics. It is of great challenges for visual tracking methods to handle large target appearance variations caused by pose deformation, fast motion, occlusion, and surrounding environments in real-time videos. In this paper, inspired by human attention cognitive saliency model, we propose a visual tracking method based on salient superpixels which integrates the target appearance similarity and cognitive saliency, and helps to location inference and appearance model updating. The saliency of superpixel is detected by graph model and manifold ranking. We cluster the superpixels of the first four target boxes into a set corresponding to object foreground and model the target appearance with color descriptors. While tracking, the relevance is computed between the candidate superpixels and the target appearance set. We also propose an iterative threshold segmentation method to distinguish the foreground and background of superpixels based on saliency and relevance. To increase the accuracy of location inference, we explore particle filter in both confidence estimation and sampling procedures. We compared our method with the existing techniques in OTB100 dataset in terms of precision based on center location error and success rate based on overlap, and the experimental results show that our proposed method achieved substantially better performance. Promising results have shown that the proposed salient superpixel-based approach is effective to deformation, occlusion, and other challenges in object tracking.

Keywords Visual tracking · Superpixel · Saliency detection · Correlation matrix · Graph model

Introduction

In recent years, visual tracking methods based on different theoretical frameworks are applied in many high-level visual tasks, such as pedestrian detection [5], visual surveillance [12, 48] and human perception in intelligent environment [31]. Although great progress has been made in tracking accuracy, it is still a challenging task when the tracking object undergoes great pose variations and heavy occlusions in complex scenes. The existing tracking methods mainly deal with the target representation, matching and location by target features in the tracking process. For example, tracking-by-detection methods [7, 9–11, 18] use target texture or other high-level features, build classifiers to distinguish the target and the background. Li et al. [7] fused the color, texture, edge features to get a weighted multi-feature, combined with spatio-context correlation to update feature weights adaptively by information entropy, and greatly enhanced the adaptability to environment variations. These kinds of tracking methods typically depend on bounding boxes for target representations. To better preserve target appearance details, tracking-by-segmentation methods have been proposed [3, 6, 8, 27, 30, 35]. In 2011, Wang et al. [3] proposed superpixel-based tracking method to establish an effective target appearance model, and trained a Bayesian classifier by using Meanshift clustering. Then, Yang et al. [27] incorporated particle filtering in superpixel tracking,

✉ Huimin Zhao
zhaohuimin@gpnu.edu.cn

Jin Zhan
gszhanjin@gpnu.edu.cn

1 School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China

2 School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

3 School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China

Published online: 28 June 2019

and proposed occlusion detection mechanism in classifier updating. Both methods achieve high accuracy but the tracking speed is slow because of the large number of clustering operations. To make better use of superpixels, Yeo et al. [8] used a graph for Absorbing Markov Chain (AMC) to treat superpixels as graph nodes and edges are encoded by similarity of scores. Wang et al. [35] also proposed a constrained graph labeling algorithm for visual tracking. The edges of graph are encoded by the underlying spatial, temporal and appearance fitness constraints. They both facilitate more accurate online update when a target involves substantial non-rigid or articulated motions.

Because real-time performance is one of the indicators of visual tracking performance, hence the Correlation Filtering (CF)–based tracking methods [19, 20, 45–47] have become one of the research hotshots in recent years due to its simplicity and strong computational efficiency. Danelljan et al. [46] used spatial regularization function that penalizes filter coefficients residing outside the target region to solve the boundary effect. Lukezic et al. [47] utilized channel and space reliability to set the weights for different feature channels and update the filters with space constraints. However, conventional CF-based trackers fail to handle the scale variation that occurs when the target object is moving, so Ding et al. [19] proposed a scalable visual tracking algorithm based on kernelized correlation filters (QKCF), which can estimate the scale of the object based on the positions of its four corners by a new Gaussian training output matrix. Zhang et al. [20] proposed a latent constrained correlation filters (LCCF) by mapping the correlation filters to a given latent subspace, and then developed a new learning framework that embeds distribution-related constraints in the latent subspace. These methods have achieved good performance in resolving target scales.

Visual tracking task mainly solves the continuous attention to sequential moving objects, but many tracking methods focus on traditional statistical learning and machine learning, and ignore the fact that human visual system greatly exceeds these methods. Most tracking models are not discriminative and adaptive enough in dealing with complex and harsh environments. Therefore, how to develop tracking methods from the perspective of visual cognition has become an urgent and challenging task. The mechanism of human biological vision is not very clear, but the related theory of visual computing has carried out a lot of research since the 1990s. The visual attention cognitive computing model is a computer-based mathematical model of human visual cognitive mechanism to obtain efficient and robust visual processing results. It has been widely used in various types of tasks in the past two years, such as natural language processing, image classification [43]. Among them, saliency detection is a kind of fundamental cognitive capability by simulating human visual attention mechanism. Since the 1990s, experts in cognitive psychology, neurophysiology, and computer vision have carried out extensive and in-depth research on saliency. Considering the development process of saliency and the ultimate application direction, it can be divided into two categories: visual focus prediction [40] and salient target detection [42]. The former tends to capture the most attractive key points in the image, while the latter pursues the integrity and uniformity of the whole target.

Most salient target detection methods use the local or global contrast of low-level features such as color, gradient, edge, and texture within a local context or the entire image. Chen et al. [41] proposed a classical regional contrast saliency method, which divides the image into several regions by color sparse histogram and graph-based segmentation method. The combination of visual attention model and depth network has become more and more popular. Mnih V et al. [44] regarded attention problem as a series of goal-based decision-making processes of agent interacting with visual environment, and proposed an end-to-end optimization process of back propagation for training neural networks. Wang et al. [25] proposed a deep learning based hybrid spatiotemporal saliency feature extraction framework for saliency detection from video footage.

To reduce the computational complexity of saliency contrast at the pixel level, superpixels have been used to represent objects of interest at the district-level in fundamental applications, such as object recognition [4, 21], human posture estimation [22, 23, 28], visual tracking [32, 33] and saliency detection [1, 25]. Yang et al. [1] proposed a classified saliency detection based superpixels, which represent the image as a close-loop graph with superpixels as nodes, and rank the similarity of the image elements via graph-based manifold ranking. Yan et al. [24] proposed a Gestalt law–guided saliency detection via characterizing HVS and forming objects, and smoothed at superpixel and object levels by fusing bottom-up and top-down mechanisms. Gao et al. [14–17] proposed a new bottom-up full-resolution saliency maps for river detection in high-resolution SAR images. The superpixels at the same layers are merged by feature-based distance, then the final full-resolution saliency map is generated by fusion algorithm via tree structured graphical modeling. This shows that saliency detection can provide cognitive guidance in complex computer vision tasks and achieve relatively good detection results. The methods of using deep network to improve saliency have also attracted more and more attention, including solving unsupervised saliency labeling [13] and adaptively fusing feature maps of different depth network levels [29].

Inspired by the achievements in cognitive computation and visual saliency, in this paper, we propose a tracking

method that combines the saliency and relevance of superpixels, and obtains the rough range of the target's foreground by iterative segmentation method. In summary, this work has the following contributions. Firstly, we incorporate cognitive saliency detection based on graph model into superpixel-based tracking, which is conducive to find the target area roughly at the search scope of frames. Secondly, we use relevance to measure the relationship between each superpixel and the target, and fuse the relevance and saliency values of each superpixel to obtain confidence. Finally, we exploit particle filter sampling and select the sample with the maximum sum of confidence as the best tracking result.

The paper is organized as follows. We introduce the research background and review the related work in the "Introduction" section. Afterwards, the "Proposed Method" section describes the proposed method in detail, including target appearance model initialization, salient superpixel detection, iterative segmentation method, and target locating. The experiments are given in the "Experimental Results" section. We conclude the paper and discuss future work in the "Conclusion and Future Work" section.

Proposed Method

We will describe the details of the proposed method in this section. Our tracking method fuses the saliency and relevance detection for superpixels to obtain the rough target foreground. Then, we use an iterative segmentation method with threshold setting to segment the foreground more accurately. The final optimal target is located in particle filter framework. Figure 1 shows the process of the proposed method. The initial target appearance set is constructed by target foreground superpixel from the first four frames by using Lab color histogram features. In the search box of subsequent frames, superpixels are segmented by SLIC algorithm [2] and their saliency are detected by graph-based manifold ranking method [1]. Then, we calculate the relevance matrix between these superpixles and the initial target appearance set. The relevance and the saliency of each superpixel are fused to find the foreground of target by iterative segmentation method. We also update the target appearance set according to updating threshold.

Appearance Model Initialization

Before tracking, we construct a set of target appearance from the first four frames for initialization. For the purpose, first of all, we find out the target in the first four frames with a simple tracking result box. Considering the compactness and efficiency of superpixel segmentation, this paper uses a simple linear iterative clustering algorithm (SLIC) proposed by Achanta et al. [2] to segment image frames into several superpixel blocks that preserve color, spatial information and boundary characteristics. The SLIC method calculates the color distance and Euclidean distance between each pixel and the superpixel center (seed point) in the image, and takes their linear combination as total similarity. The

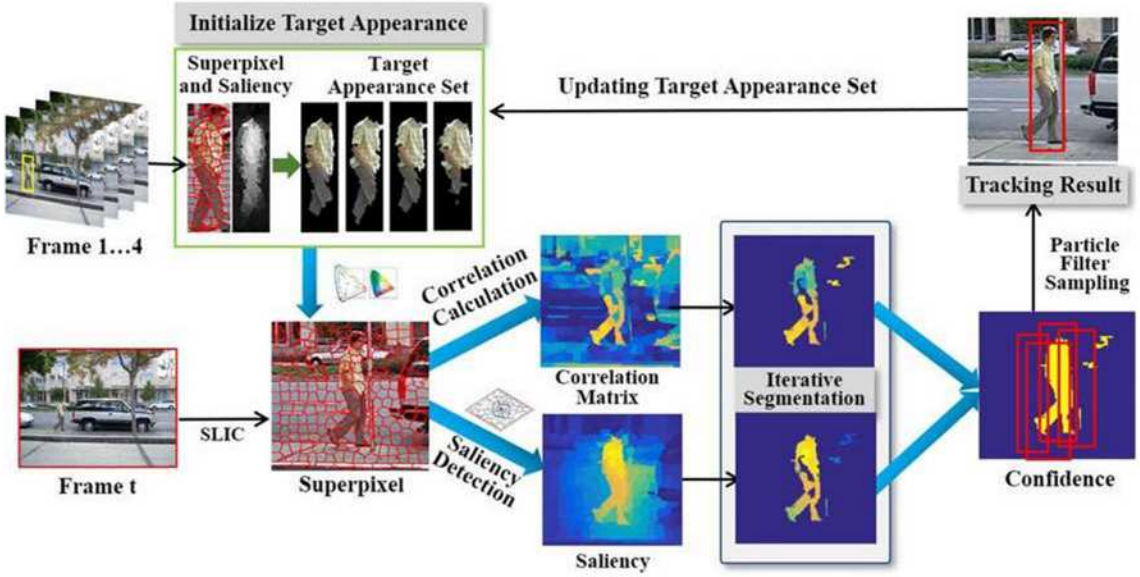


Fig. 1 Flowchart of the proposed method. The initial target appearance set is constructed by the salient superpixels in the first four frames. At frame t, the relevance of superpixels are calculated between the foreground appearance set, and the motion model predicted by saliency detection to suggest potential centers. To further estimate the object state, we use the iterative segmentation algorithm after fusing the value of relevance and saliency, and assign confidence map for each superpixels. The final optimal target is located in particle filter framework

image is clustered step by step within the range of two times the distance between the seeds. Because this method does not consider the distant pixels, it not only reduces the calculation time but also makes the size of the superpixels more regular and compact, and each superpixel can maintain the image boundary.

In frame t, we describe the set of superpixels is $P^t = \{s_i^t\}_{i=1}^N$, where N is the number of superpixels in current frame, t is the subscript of image frame. Each superpixel is described as parameter set $s_i^t = (x_i^t, y_i^t, f_i^t, \theta_i^t, \delta_i^t, l_i^t)$, where x_i^t and y_i^t is the Cartesian coordinates of the center of the superpixel, f_i^t is the feature descriptor, θ_i^t indicates whether it is at the image boundary ($\theta_i^t = 1$ means at the boundary, otherwise, $\theta_i^t = 0$), δ_i^t is the flag of foreground or background ($\delta_i^t = 1$ means foreground, otherwise, $\delta_i^t = 0$), l_i^t is the superpixel label.

The first four tracking boxes contain a small amount of background information besides the tracked target, and the superpixels of the target foreground and background are not overlapped. Then, we use saliency detection algorithm [1] to find the significant motion of superpixels, which have the greatest similarity of target foreground among four frames. By setting an adaptive threshold, the foreground and background superpixels can be separated. Although this threshold may remove the parts belonging to the target, it ensures that the foreground superpixels must be the target part.

We put the foreground superpixels of the four frames into a set to form the initial target appearance set $H = \{h_i^t | \delta_i^t = 1, t \in [1,4]\}_{i=1}^M$, which is used to calculate the relevance coefficient matrix to estimating the object state and be updated dynamically in the tracking process. Here, M is the total number of foreground superpixels in the first four frames. Considering that the color distribution of target is relatively stable and is almost unchanged except occlusions in the video sequences, hence, we use the Lab color histogram to represent the feature of superpixels.

Salient Superpixel Detection by Graph Construction

In the graph model–based saliency detection method, superpixels are nodes of graph, and the distance between adjacent superpixels (such as geometric distance or color distance) is the weight of the edge in the graph. Based on the graph-based manifold sorting algorithm [1], we calculate the saliency for each superpixel in frame.

Before calculating saliency based on graph model, the superpixels are divided into two subsets: some vertices are defined for query point; the remaining vertices are ranked according to their relevance with the query points. A graph is defined as $G = (V, E)$, where the graph vertices V correspond to superpixels. In the frame t, given the superpixel set $P^t \subset R^{d \times N}$, where d is the feature dimension and N is the number of superpixels. Let $f : P \rightarrow R^N$ denotes a ranking function which assigns a ranking value to each superpixel, and f can be viewed as a vector $f = [f_1, \ldots, f_N]^T$. The edges E are weighted by an affinity matrix $W = [w_{ij}]^{N \times N}$. Given G, the degree matrix is $D = diag\{d_1, \ldots, d_N\}$, where $d_i = \sum_j w_{ij}$, and the relationship matrix W is a sparse matrix because of the structural relationship of vertices, adjacent vertices and vertices with shared border. The optimal rankings of queries $f^\star$ is computed by solving the following optimization problem:

$$f^\star = arg\min_f \frac{1}{2}\left(\sum_{i,j=1}^N w_{ij}\left\|\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right\|^2 + \mu\sum_{i=1}^N \|f_i - y_i\|^2\right) \tag{1}$$

where the parameter μ controls the balance of the smoothness constraint and the fitting constraint, $Y = [y_1, \ldots, y_N]^T$ denotes an indication vector, in which $y_i = 1$ if s_i is a query point, and $y_i = 0$ otherwise. Set the derivative of the above function to be zero, the resulted ranking function can be written as:

$$f^\star = (D - \varepsilon w)^{-1} Y, \varepsilon = \frac{1}{1+\mu} \tag{2}$$

The weight of each edge is given by the similarity of the relevance with vertices. Considering that the superpixels with the same color and the same spatially connected area will have more consistent ranking fits with the target appearance, and more uniformly highlight the whole target. Therefore, we learn the edge weights by maximizing the scores between superpixels in target foreground with color distance and geometric distance. Let $\|c_i - c_j\|$ and $\|p_i - p_j\|$ are Lab color distance and Euclidean distance of the i-th and j-th superpixels, respectively, w_{ij} is computed by the following formula:

$$w_{ij} = e^{-\left(\frac{\|c_i - c_j\|}{\sigma^2} + \lambda\frac{\|p_i - p_j\|}{\sigma^2}\right)} \tag{3}$$

where λ is coefficient of balance and set to be 0.5 in the experiments, and σ is set to be 10.

In tracking problem, target is given in the bounding box at the first frame. The superpixels in bounding box contain target foreground and a small amount of background that do not overlap with the foreground. The center superpixel is the most important foreground location, so the spatial relationship information around with the center superpixel is an important cue for saliency detection. Based on this consideration, we exploit two-stage scheme for saliency detection in which the superpixels located at the boundary and the center are used as query points respectively to compute the ranking function.

In the first stage, four saliency maps are constructed using boundary priors and then integrated as the final map.

Taking top image boundary as an example, the superpixels on this side are the queries, and other superpixels are the unlabeled data. According to the ranking scores of Eq. (2), the saliency of the i-th superpixel in the graph relative to top image boundary is calculated by $S_t(i) = 1 - \overline{f}^{\star}(i)$, where $\overline{f}^{\star}(i)$ is the ranking score normalized to [0,1]. In the same way, we can get the saliency of the i-th superpixel relative to the bottom, left, and right boundary queries, denoted $S_b(i)$, $S_l(i)$, $S_r(i)$. Then, the saliency of the i-th superpixel in the first stage is:

$$S_1(i) = S_t(i) \times S_b(i) \times S_l(i) \times S_r(i) \tag{4}$$

The second stage is to improve the saliency map via ranking with foreground queries, for some background superpixels may not be adequately suppressed. We select the superpixel containing the target center of the previous frame as the new query point. The new ranking vector $\overline{f'}^{\star}(i)$ is calculated by Eq. (2) and normalized to [0,1], then the saliency of the i-th superpixel in the second stage is:

$$S_2(i) = \overline{f'}^{\star}(i) \tag{5}$$

The final saliency of superpixels are obtained by multiplying and normalizing the saliency obtained in the first stage and the second stage. The superpixels with high saliency value can provide sufficient cues to identify the interesting target and give the rough range of target motion, we call them salient superpixels.

Iterative Segmentation Method

In order to further distinguish whether the superpixel belongs to the foreground or background, we calculate the relevance between the superpixels in current frame and the initial target appearance set H with the correlation coefficient matrix. We first define a general formula for finding correlation coefficients as follows:

$$c_{ab} = \frac{\sum ab}{\sqrt{\sum a^2}\sqrt{\sum b^2}} \tag{6}$$

where $c_{ab} \in [0, 1]$ indicates the correlation coefficients of superpixels a and b. The closer the c_{ab} is to 1, the more relevant the two superpixels are.

In our tracking method, we define $P = \{p_i\}_{i=1}^{N} \in R^{d\times N}$ is the superpixels set in current frame, $H = \{h_j\}_{j=1}^{M} \in R^{d\times M}$ is the initial target appearance set collected in the first four frames, where d is the dimension of feature. Then, we can calculate the correlation matrix $C_{ov} = [c_{ij}]$ by Eq.(6) for superpixels in sets P and H.

$$C_{ov} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1M} \\ c_{21} & c_{22} & \dots & c_{2M} \\ & \dots & \dots & \\ c_{N1} & c_{N2} & \dots & c_{NM} \end{bmatrix} \in R^{N\times M} \tag{7}$$

where the i-th row in matrix C_{ov} indicates that the relevance between the superpixel p_i in set P and all superpixels in target set H. In other words, the superpixels of set H represent the appearance block of the target, which is the reference standard for calculating the correlation between p_i and the target.

The advantage of using relevance to measure the current superpixels is that it does not consider the change of target background, and reduces the influence caused by the change of target posture change or occlusion. However, some background superpixels may have similar colors to the foreground, they are usually calculated with a relatively high relevance values. Meanwhile, the saliency map calculated in the "Salient Superpixel Detection by Graph Construction" section also has background superpixels with high saliency. In order to remove these background superpixels in correlation matrix and saliency map, we propose an iterative segmentation method with threshold setting to segment the foreground superpixels.

Take the correlation matrix C_{ov} as an example, firstly, we find the maximum value of each row of C_{ov} to form a new vector $C_m \in R^{N\times 1}$, that is, the maximum relevance between the current frame superpixels and the target superpixels. Then, we define the separation thresholds α and β as follows.

$$\begin{cases} \alpha = \max(C_m) - \frac{\max(C_m)}{\eta} \\ \beta = \min(C_m) + \frac{\min(C_m)}{\eta} \end{cases} \tag{8}$$

where η is an empirical parameter and set to 50 in the experiments. Secondly, we maintain two sets of superpixels, S_f for object foreground and S_b for background. If $C_m(i) \geq \alpha$, the i-th superpixel in set P is highly relevance with the target, that is, p_i is the foreground and is put into set S_f. If $C_m(i) \leq \beta$, the p_i has little relevance with the target, and is put into set S_b. For the remaining values in C_m that do not satisfy the above two conditions, new α and β are computed according to Eq. (8), and the threshold separation process is iterated until all the superpixels in P are assigned as foreground or background. In the experiments, the maximum number of iterations is set to 25.

By using the above iteration segmentation method, we separate the superpixels in current frame in terms of relevance and saliency to get corresponding foreground superpixels, respectively. In Fig. 2, we select the images at the frame 25 in videos *Basketball* and *MountainBike* and segment them to superpixels by SLIC method [2]. Then, we calculate the relevancy and the saliency maps for current frame superpixels. In the relevancy map, the superpixels which are very similar to the target appearance have high similarity, but some background superpixels which are similar to the color of the target are also calculated to have high values, so the relevancy map is more obvious to the contour of the target and sensitive to the background. In the

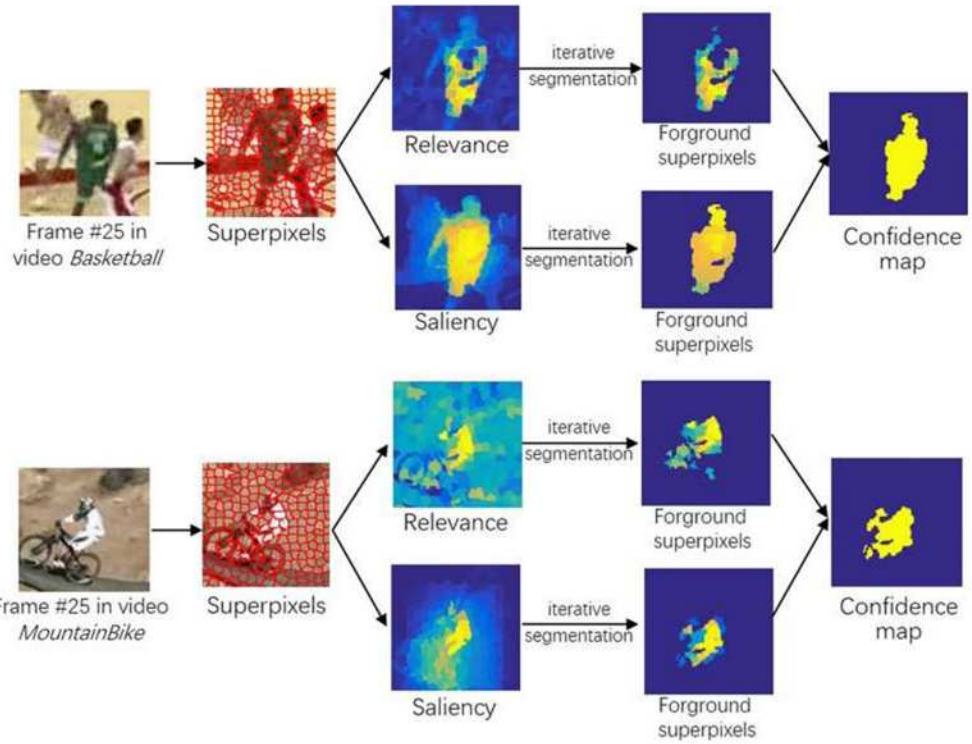


Fig. 2 Illustration of the target foreground segmented by iterative segmentation method. Images at the frame 25 in videos *Basketball* and *MountainBike* are segmented to superpixels by SLIC method [2]. Then, we calculate the relevancy and the saliency maps for current frame superpixels; To further find the foreground superpixels, we use iterative segmentation method for the relevancy and the saliency maps, respectively. The final confidence map of current frame superpixels is integrated by forgeround superpixels of the relevancy and the saliency

saliency map, it reflects the trend of the target motion, and the target with its surrounding area can be well detected, and the background has very low value. So the saliency map can better show the approximate center of the target. Therefore, to further find the foreground superpixels, we use iterative segmentation method for the relevancy and the saliency maps, respectively. The final confidence map of current frame superpixels is integrated by foreground superpixels of the relevancy and the saliency.

Target Locating by Confidence Map

When a new frame arrives, we first set a search region around the previous object location, and segment superpixels in the search region by SLIC method. Then, we calculate saliency values and correlation coefficients of superpixels, and find the foreground superpixels by iterative segmentation methods. The superpixel set after iterative segmentation is denoted as S_s and S_c, respectively.

In order to find the possible target location of the current frame, we calculate the confidence map for superpixels by combining the foreground superpixels of S_s and S_c. The confidence of superpixel i is given by:

$$\text{conf}(i) = S_s(i) \bigcup S_c(i) \tag{9}$$

Furthermore, we use particle filter to sample around the target in the previous frame. The number of particles in the experiment is 600. For each particle, we calculate the number of foreground and background superpixels, and then select the particle $\widehat{l_k}$ with the largest confidence sum of foreground superpixel as the best tracking position, as given in:

$$\widehat{l_k} = \arg\max_k \sum_{j=1}^{N_k^f} \text{conf}(j) \tag{10}$$

where k is the index of particle box and N_k^f is the foreground number of superpixels in the k-th particle.

Target Appearance Model Updating

Compared with the background, the target exhibits similar characteristics in a short period of time, that is, the appearance of the target is relatively stationary in short term. But in the entire video time domain, the video target and background are typically non-stationary signals, and their distribution parameters or distribution laws will change with time. The superpixel-based model we have described makes effective use of target foreground in the tracking frames, but the overall appearance of the object changes with time and the dynamic representation can not be ignored. We introduce an adaptive strategy into updating the feature set of target appearance.

For the candidate particles X^t in frame t, their confidences are the sum of the foreground superpixel confidences. We set a threshold to detect whether occlusion

or pose variation have taken place in the frame. The formula is as follows.

$$C(\varphi - \max(\text{conf}(X^t))) > \rho \tag{11}$$

where φ is the average confidence of the target estimation in the existing frames, $\max(\text{conf}(X^t))$ represents the maximum confidence value obtained from X particle estimates. The larger the $(\varphi - \max(\text{conf}(X^t)))$ value is, the closer the particle region is to the background. $C(.)$ is a normalization coefficient, and ρ is the empirical threshold and is set as 0.515 in the experiment.

If the value on the left side of the Eq. (11) is larger, it means that there is little difference between the two confidence values and no significant change in the appearance of the target. We add the superpixels of the predicted sample to the set of the target appearance. Conversely, occlusion or distortion is likely to occur. In this case, all the superpixels of the current frame are regarded as background and take the estimation result of the previous frame as the current result. The average value is also taken as the confidence value of the current frame. Through this adaptive update strategy, the proposed method can quickly adapt to the changes of object during tracking and enhance the robustness of object tracking.

Experimental Results

In this section, we evaluate the proposed tracking method with extensive experiments in OTB100 dataset [36], and compare our method with similar tracking methods. We conduct performance analysis both quantitatively and qualitatively in various scenarios, such as occlusions, target deformation, target scale change, and target external rotation.

Experimental Setup

We implement the proposed method in MATLAB and run it on an Intel(R) core(TM) i7@4GHz machine with 8 GB RAM. In the experiments, the number of superpixel segmentation is set to 200 in the first four frames of the video, and the remaining frames are set to 300. The parameters λ and σ in edge weight of salient superpixel detection are 0.5 and 10, respectively. The parameter η of threshold setting in iteration foreground segmentation stage is 50 in Eq. (8), and the maximum number of iterations is 25. The update threshold ρ in Eq. (11) is 0.5.

We evaluate our algorithm on popular OTB100 dataset [36] with 8 selected challenging video sequences (*FaceOcclu1*, *Basketball*, *Jogging*, *Singer2*, *Shaking*, *Boy*, *CarScale*, *David3*). These sequences cover various challenges that object tracking often confronts. We compare our method with 7 similar methods, namely SPT [3], FRAG [39], CT [37], MIL [10], TLD [9], LSK [38], and Struck [11]. For fair evaluation, we evaluate the proposed tracker against those methods using the source codes provided by the authors. Each tracker is run with adjusted parameters.

Quantitative Evaluation

We employ two commonly used evaluation criteria: the success rate based on Overlap Rate and the precision based on Center Location Error. The Overlap Rate (OLR) is the area ratio of the intersection and union of the tracking result R_m and the groundtruth R_g, calculated as $\frac{\text{area}(R_m \bigcap R_g)}{\text{area}(R_m \bigcup R_g)}$. It indicates the stability of tracker as it takes the size and pose of the target object into account. The total success rate is a percentage of successfully tracked frames to all frames, where the successfully tracked frames refer to those the overlap rate of the tracking results with the groundtruth is bigger than 50%. Table 1 shows the overall tracking success rate (higher is better) of the comparative approaches on eight sequences to quantitative estimation. Because the overlap rate can not accurately compare the tracking methods with or without scale change processing, we also use the Center Location Error (CLE) to summarize the overall performance. The CLE measures the distance between the estimated center location and the ground truth in each frame. It indicates the accuracy of tracking the target

Table 1 Tracking Success Rate (higher is better) of the comparative approaches on eight sequences. The best results are shown in italic fonts

Sequences	SPT	FRAG	CT	MIL	TLD	LSK	Struck	Ours
FaceOcclu1	0.130	*0.625*	0.235	0.375	0.398	0.343	0.507	*0.625*
Jogging	0.623	0.667	0.201	0.218	*0.896*	0.211	0.537	0.859
David3	0.561	0.584	0.317	0.568	0.110	0.137	0.522	*0.867*
Boy	0.910	0.495	0.550	0.630	0.851	0.393	0.901	*0.919*
Singer2	0.034	0.199	0.031	0.043	0.047	0.036	0.045	*0.402*
CarScale	*0.763*	0.579	0.615	0.596	0.479	0.558	0.563	0.762
Basketball	*0.816*	0.775	0.307	0.257	0.098	0.022	0.379	0.658
Shaking	0.074	0.131	0.031	*0.719*	0.412	0.031	0.042	0.230

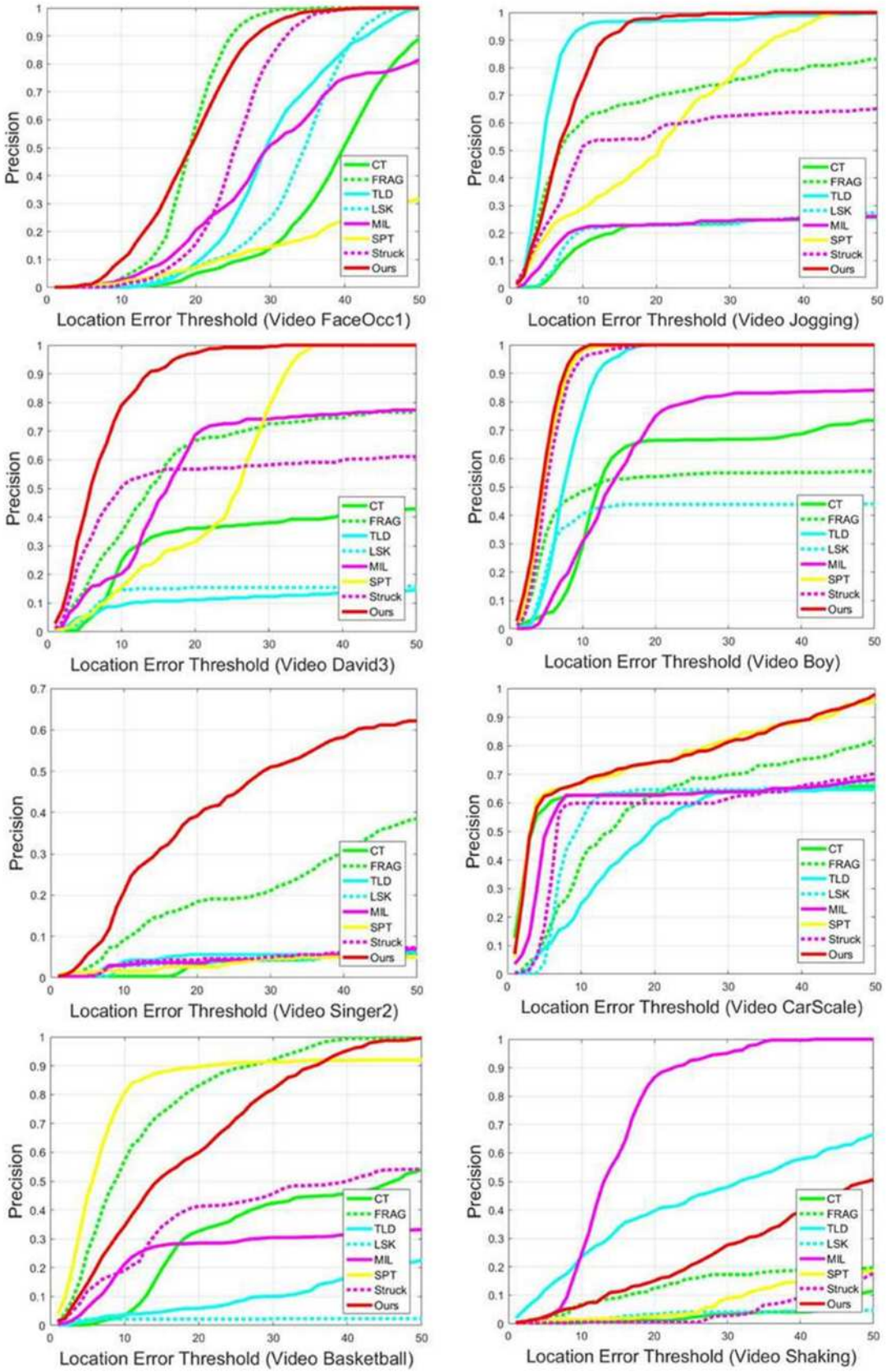


Fig. 3 The precision plots by center location errors evaluation. The location error threshold is 20 which roughly corresponds to at least 50% overlap between the tracker and the ground truth

center. Precision is defined as the ratio between the number of frames with CLE less than a given threshold and the total frames. The overall tracking performance is reported in Fig. 3. The precision is plotted as the threshold changes, when the position error threshold is 20, it is approximately equivalent to at least 50% overlap between the tracker and the ground truth.

As can be observed from the Table 1 and Fig. 3, our method performs the best in video sequences *FaceOcclu1*, *Singer2*, *Boy*, *David3*). This shows that our algorithm can effectively deal with occlusion and posture changes. This is mainly because the saliency detection in our method can estimate the approximate position of the target and help to reduce the center location error. We note that, regardless of whether the target object is occluded or not, the pose changes of the target can be approximately detected by the relevance detection of the target appearance, and thus help to improve the overlap rate of tracking results. However,

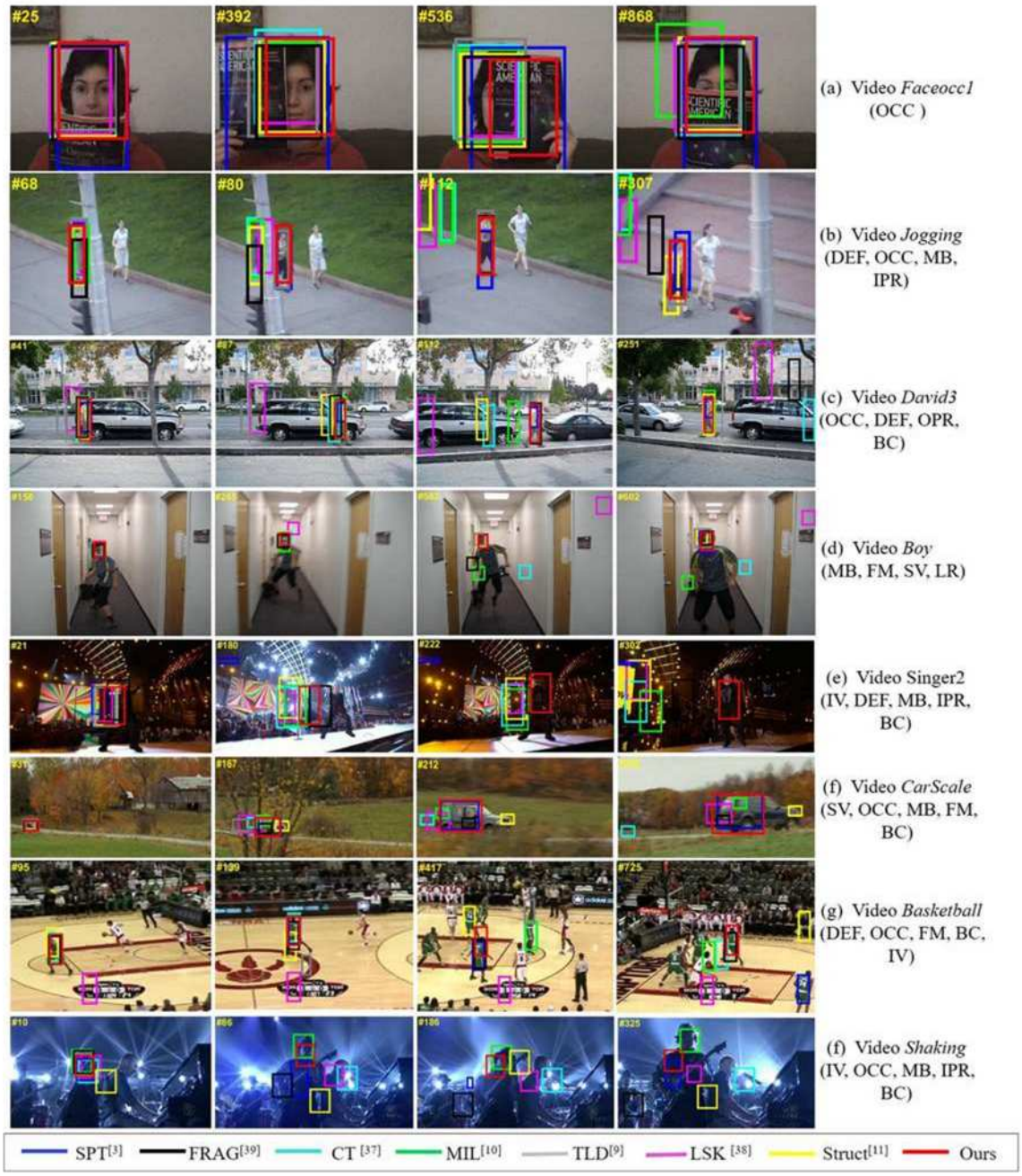


Fig. 4 Quantitative comparison with different methods on eight test video sequences. These sequences cover various challenges that object tracking often confronts, such as Illumination Variation (IV), Scale Variation (SV), Occlusion (OCC), Deformation (DEF), Motion Blur (MB), Fast Motion (FM), In-Plane Rotation (IPR), Out-of-Plane Rotation (OPR), Background Clutter (BC) and Low Resolution (LR)

our algorithm is not robust enough to deal with illumination changes and similar group movements, which makes the overlap rate of *Basketball* and *Shaking* sequences relatively low.

Qualitative Evaluation

In this subsection, we will qualitatively evaluate our method by observing its behavior in different sequences. Figure 4 shows example tracking results of the trackers on test sequences with various challenges. We analyze our tracking results on these sequences with respect to different challenging situations below.

Heavy Occlusion The *Faceocc1*, *Jogging*, and *David3* sequences contain noticeable and heavy occlusion. FRAG and our method performed well in these sequences by making use of the segmented inner structure of the object. The challenge of the *Faceocc1* sequence is that there are different degrees of occlusion. Most of the methods can track the target. SPT tracker had the worst results and regarded the occlusion as part of the target from the #25 frame. In the *Jogging* sequence, the heavy occlusion occurred from frame #68 to frame #79, TLD tracker performed best due to its long-time detection scheme and relocated the target after the target appears at frame #80. The MIL, LSK, CT, FRAG, and Struck trackers drifted away after frame #79. Then, FRAG and Struck trackers relocated the target at frame #112 and #191, respectively. In the *david3* sequence, the proposed tracker performed best and the FRAG tracker performed second best. The TLD and LSK trackers lost the target at frame #30 and #41, respectively, when target was obscured by the roadside indicator. The CT and Struck trackers lost the target at frame #87 due to tree occlusion, but the Struck tracker relocated the target as it moved back, while the CT tracker failed to track it again.

Fast Motion and Scale Change The main challenges of *Boy*, *Singer2*, and *CarScale* sequences are motion blur caused by fast motion and target scale changes. The target object also undergoes some occlusions. The proposed tracker and SPT tracker achieved best performance in the entire sequence. This is mainly because our superpixel-based tracking algorithm can get the structure and motion estimation of the object. In *Boy* sequence, all methods can track the target when the target did not move very fast. The LSK, CT, and MIL trackers failed when fast motion and motion blur occurred at the #265, #401, and #508 frames, respectively. The LSK tracker performed poorly for its local appearance model could not represent the target appropriately. The main difficulty of *CarScale* sequence is the fast moving car with tree occlusion from #159 to #176 frames and scale change. Because the color of the tree is similar to the target, CT and TLD trackers drifted away at the #167 frame. SPT tracker performed best because it had a superpixel clustering mechanism to solve the scale change problem.

Pose Variation and Background Clutter The *Basketball*, *shaking* sequences have drastic pose variation with motion blur in cluttered background. In *basketball* sequence, there are many similar individuals around the target, and human body tends to deform irregularly as non-rigid motion. SPT tracker performed best, FRAG tracker and the proposed tracker performed second best. This is mainly because superpixel-based tracking algorithms are effective to follow articulated or deformable objects and are more robust to interference from similar targets. LSK and TLD trackers drifted away when the target had noticeable pose variation at the #95 and #139 frames. In *shaking* sequence, the main challenge is the drastic pose variation with illumination and scale change. The MIL tracker performed best. The Struck tracker drifted away at the #3 frame, and FRAG, SPT, LSK, and CT trackers lost the target when full drastic illumination occurred at the #57 frame and drifted away.

Conclusion and Future Work

In this paper, we proposed a visual tracking method based on salient superpixel with iterator segmentation, where the target appearance is represented by middle-level superpixel blocks and associated color feature descriptors. The target appearance set is used to measure the correlation between candidate samples in subsequent frames. We detected the saliency of superpixles by graph model and manifold ranking to handle the importance of target foreground and estimate the approximate motion information of the target. For accurately cutting the target components, we separated the target foreground and background superpixels by an iteration segmentation algorithm to improve initial target appearance. In the tracking process, we integrated the correlation value and saliency value of the superpixel to get the confidence map of the current frame, and used the confidence as the observation value in the particle filter tracking framework to achieve robust performance. We compared our method with the existing methods on OTB100 dataset, and the experimental results show that the proposed method is effective to deformation, occlusion and other challenges in object tracking.

We will further investigate this work. First, the correlation value is calculated by Lab color features. When the target color is similar to the background, it is easy to mistake the background for the foreground. Therefore, a more

better representation of the target appearance characteristics should be studied. Second, when the background is cluttered, it is easy to mistake the background as a salient object. We plan to add the superpixel-based optical flow [26] information to improve the saliency accuracy of the moving target.

Funding Information This study was partly supported by National Natural Science Foundation of China (61772144, 61672008, 61876045), Foreign Science and Technology Cooperation Plan Project of Guangzhou Science Technology and Innovation Commission (201807010059), Guangdong Provincial Application oriented Technical Research and Development Special Fund Project (2016B010127006), the Scientific and Technological Projects of Guangdong Province (2017A050501039), Innovation Team Project (Natural Science) of the Education Department of Guangdong Province (2017KCXTD021), National Natural Science Foundation of China Youth Fund (61602116), Innovation Research Project (Natural Science) of Education Department of Guangdong Province (2016KTSCX077), and Zhujiang Science and Technology New Star Project of Guangzhou (201906010057).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

1. Yang C, Zhang L, Lu H, Ruan X, Yang M. Saliency detection via graph-based manifold ranking. In: 2013 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2013. p. 3166–3173. https://doi.org/10.1109/CVPR.2013.407.
2. Radhakrishna A, Appu S, Kevin S, Aurelien L, Pascal F, Sabine S. Slic Superpixels Compared to State-of-the-art Superpixel Methods. IEEE Trans Pattern Anal Mach Intell (PAMI). 2012;34(11):2274–2282. https://doi.org/10.1109/TPAMI.2012.120.
3. Wang S, Lu H, Yang F, Yang M. Superpixel Tracking, In: IEEE International Conference on Computer Vision (ICCV); 2011. p. 1323–1330. https://doi.org/10.1109/ICCV.2011.6126385.
4. Carsten R, Vladimir K, Grabcut AB. Interactive foreground extraction using iterated graph cuts. ACM Trans Graph (TOG). 2004;23(3):309–314. https://doi.org/10.1145/1186562.1015720.
5. Yan Y, Ren J, Zhao H, Sun G, Wang Z, Zheng J, Stephen M, John S. Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. Cogn Comput. 2017;(9):1–11. https://doi.org/10.1007/s12559-017-9529-6.
6. Vasileios B, Falk S, Nassir N, Slobodan I. Segmentation based particle filtering for real-time 2d object tracking. In: European Conference on Computer Vision (ECCV); 2012. p. 842–855. https://doi.org/10.1007/978-3-642-33765-9_60.
7. Li G, Wang ZY, Luo J, Chen X, Li H. Spatio-context-based target tracking with adaptive multi-feature fusion for real-world hazy scenes. Cogn Comput. 2018;10(4):545–557. https://doi.org/10.1007/s12559-018-9550-4.
8. Son J, Jung I, Park K, Han B. Tracking-by-segmentation with online gradient boosting decision tree. In: IEEE International Conference on Computer Vision (ICCV); 2015. p. 3056–3064. https://doi.org/10.1109/ICCV.2015.350.
9. Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. IEEE Trans Pattern Anal Mach Intell (PAMI). 2012;34(7):1409–1422. https://doi.org/10.1109/TPAMI.2011.239.
10. Xu C, Tao W, Meng Z. Robust visual tracking via online multiple instance learning with fisher information. Pattern Recogn. 2015;48(12):3917–3926. https://doi.org/10.1016/j.patcog.2015.06.004.
11. Hare S, Saffari A, Torr P. Struck: structured output tracking with kernels. In: IEEE International Conference on Computer Vision (ICCV); 2011. p. 263–270. https://doi.org/10.1109/ICCV.2011.6126251.
12. Zhang P, Zhuo T, Xie L, Zhang Y. Deformable object tracking with spatiotemporal segmentation in big vision surveillance. Neurocomputing. 2016;204(C):87–96. https://doi.org/10.1016/j.neucom.2015.07.149.
13. Zhang J, Zhang T, Dai Y, Harandi M, Hartley R. Deep unsupervised saliency Detection: A multiple noisy labeling perspective. In: 2018 I.E. Conference on Computer vision and pattern recognition (CVPR). 2018.
14. Gao F, Ma F, Wang J, Sun J, Yang E, Zhou H. Visual saliency modeling for river detection in high-resolution SAR imagery. IEEE Access. 2017;6:1000–1014. https://doi.org/10.1109/ACCESS.2017.2777444.
15. Gao F, You J, Wang J, Sun J, Yang E, Zhou H. A novel target detection method for SAR images based on shadow proposal and saliency analysis. Neurocomputing. 2017;267(C):220–231. https://doi.org/10.1016/j.neucom.2017.06.004.
16. Gao F, Zhang Y, Wang J, Sun J, Yang E. Amir h. Visual attention model based vehicle target detection in synthetic aperture radar images a novel approach. Cogn Comput. 2015;7(4):434–444. https://doi.org/10.1007/s12559-014-9312-x.
17. Gao F, Ma F, Zhang Y, Wang J, Sun J, Yang E, Amir H. Biologically inspired progressive enhancement target detection from heavy cluttered SAR images. Cogn Comput. 2016;8(5):955–966. https://doi.org/10.1007/s12559-016-9405-9.
18. Liu Q, Wang Y, Yin M, Ren J, Li R. Decontaminate feature for tracking: adaptive tracking via evolutionary feature subset. J Electron Imaging. 2017;26(6):1–10. https://doi.org/10.1117/1.JEI.26.6.063025.
19. Ding G, Chen W, Zhao S, Han J, Liu Q. Real-time scalable visual tracking via quadrangle kernelized correlation filters. IEEE Trans Intell Trans Syst. 2018;19(1):140–150. https://doi.org/10.1109/TITS.2017.2774778.
20. Zhang B, Luan S, Chen C, Han J, Wang W. Latent constrained correlation filter. IEEE Trans Image Process (TIP). 2018;27(3):1038–1048. https://doi.org/10.1109/TIP.2017.2775060.
21. Amir R, Daphna W. Extracting foreground masks towards object recognition. In: International Conference on Computer Vision (ICCV); 2011. p. 1371–1378. https://doi.org/10.1109/ICCV.2011.6126391.
22. Chai Y, Ren J, Zhao H, Li Y, Ren JC, Paul M. Hierarchical and multi-featured fusion for effective gait recognition under variable scenarios. Pattern Anal Appl. 2016;19(4):905–917. https://doi.org/10.1007/s10044-015-0471-5.
23. Ezrinda MZ, Kamarul HG, Ren J, Mohd ZS. A hybrid thermal-visible fusion for outdoor human detection. Journal of Telecommunication, Electronic and Computer Engineering (JTEC. 2018.
24. Yan Y, Ren J, Sun G, Zhao H, Han J, Li X, Stephen M, Zhan J. Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. Pattern Recogn. 2018;79:65–78. https://doi.org/10.1109/TIP.2017.2775060.

25. Wang Z, Ren J, Zhang D, Sun M, Jiang J. A Deep-Learning based feature hybrid framework for spatiotemporal saliency detection inside videos. Neurocomputing. 2018;287:68–83. https://doi.org/10.1016/j.neucom.2018.01.076.
26. Shi J, Carlo T. Good features to track. In: 1994 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 1994. p. 593–600. https://doi.org/10.1109/CVPR.1994.323794.
27. Yang F, Lu H, Yang M. Robust superpixel tracking. IEEE Trans Image Process (TIP). 2014;23(4):1639–1651. https://doi.org/10.1109/TIP.2014.2300823.
28. Perera A, Law Y, Chahl J. Human pose and path estimation from aerial video using dynamic classifier selection. Cognitive Comput. 2018;10:1019–1041. https://doi.org/10.1007/s12559-018-9577-6.
29. Zhang L, Dai J, Lu H, He Y, Gang W. A bi-directional message passing model for salient object detection. In: 2018 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2018. p. 1741–1750.
30. Zhou X, Li X, Hu W, Learning A. Superpixel-driven speed function for level set tracking. IEEE Trans Cybern. 2016;46(7):1498–1510. https://doi.org/10.1109/TCYB.2015.2451100.
31. Han J, Eric J, Paul M, Peter H. Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment. IEEE Trans Consum Electron. 2012;58(2):255–263. https://doi.org/10.1109/TCE.2012.6227420.
32. Hong Z, Wang C, Mei X, Prokhorov D, Tao D. Tracking using multilevel quantizations. In: European Conference on Computer Vision (ECCV); 2014. vol 8694. p. 155–171. https://doi.org/10.1007/978-3-319-10599-4_11.
33. Xiao J, Stolkin R, Ales L. Single target tracking using adaptive clustered decision trees and dynamic multilevel appearance models. In: 2015 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2015. p. 4978–4987.
34. Yeo D, Son J, Han B, Han JH. Superpixel-based tracking-by-segmentation using Markov chains. In: 2017 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2017. p. 511–520. https://doi.org/10.1109/CVPR.2017.62.
35. Wang L, Lu H, Yang M. Constrained superpixel tracking. IEEE Trans Cybern. 2018;48(3):1030–1041. https://doi.org/10.1109/TCYB.2017.2675910.
36. Wu Y, Lim J, Yang M. Object tracking benchmark. IEEE Trans Pattern Anal Mach Intell (PAMI). 2015;37(9):1834–1848. https://doi.org/10.1109/TPAMI.2014.2388226.
37. Zhang K, Zhang L, Yang M. Real-time compressive tracking. In: European Conference on Computer Vision (ECCV); 2012. p. 864–877. https://doi.org/10.1007/978-3-642-33712-3_62.
38. Liu B, Huang J, Yang L, Casimir K. Robust tracking using local sparse appearance model and k-selection. In: 2011 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2011. p. 1313–1320. https://doi.org/10.1109/CVPR.2011.5995730.
39. Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram. In: 2006 I.E. Conference on Computer vision and pattern recognition (CVPR). 2006.
40. Borji A, Sihite D, Itti L. Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. IEEE Trans Image Process (TIP). 2013;22(1):55–69. https://doi.org/10.1109/TIP.2012.2210727.
41. Cheng M, Zhang G, Niloy J, Huang X, Wu S. Global contrast based salient region detection. In: 2011 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2011. p. 409–416. https://doi.org/10.1109/CVPR.2011.5995344.
42. Borji A, Cheng M, Jiang H, Li J. Salient object detection: a benchmark. IEEE Trans Image Process (TIP). 2015;24(12):5706–5722. https://doi.org/10.1109/TIP.2015.2487833.
43. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X. Residual attention network for image classification. In: 2017 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2017. vol 1. p. 6450–6458. https://doi.org/10.1109/CVPR.2017.683.
44. Mnih V, Heess N, Graves A, Kavukcuoglu K. Recurrent models of visual attention. In: The 27th International Conference on Neural Information Processing Systems (NIPS); 2014. vol 2. p. 2204–2212.
45. Henriques J, Rui C, Martins P, Batista J. High-speed tracking with kernelized correlation filters. IEEE Trans Pattern Anal Mach Intell (PAMI). 2014;37(3):583–596. https://doi.org/10.1109/tpami.2014.2345390.
46. Danelljan M, Hager G, Khan F, Felsberg M. Convolutional features for correlation filter based visual tracking. In: IEEE International Conference on Computer Vision Workshop; 2015. p. 621–629. https://doi.org/10.1109/ICCVW.2015.84.
47. Lukezic A, Vojir T, Zajc L, Jiri M, Matej K. Discriminative correlation filter with channel and spatial reliability. In: 2017 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2017. vol 1. p. 4847–4856. https://doi.org/10.1109/CVPR.2017.515.
48. Benfold B, Reid I. Stable multi-target tracking in real-time surveillance video. In: 2011 I.E. Conference on Computer Vision and Pattern Recognition (CVPR); 2011. p. 3457–3464. https://doi.org/10.1109/CVPR.2011.5995667.

Cognitive Computation
https://doi.org/10.1007/s12559-019-09662-y

Salient Superpixel Visual Tracking with Graph Model and Iterative Segmentation

Jin Zhan[1] · Huimin Zhao[1] · Penggen Zheng[1] · Hefeng Wu[2,3] · Leijun Wang[1]

Received: 29 December 2018 / Accepted: 13 June 2019

Abstract

Visual object tracking is to locate an object of interest in a sequence of consecutive video frames, which is widely applied in many high-level computer vision tasks such as intelligent video surveillance and robotics. It is of great challenges for visual tracking methods to handle large target appearance variations caused by pose deformation, fast motion, occlusion, and surrounding environments in real-time videos. In this paper, inspired by human attention cognitive saliency model, we propose a visual tracking method based on salient superpixels which integrates the target appearance similarity and cognitive saliency, and helps to location inference and appearance model updating. The saliency of superpixel is detected by graph model and manifold ranking. We cluster the superpixels of the first four target boxes into a set corresponding to object foreground and model the target appearance with color descriptors. While tracking, the relevance is computed between the candidate superpixels and the target appearance set. We also propose an iterative threshold segmentation method to distinguish the foreground and background of superpixels based on saliency and relevance. To increase the accuracy of location inference, we explore particle filter in both confidence estimation and sampling procedures. We compared our method with the existing techniques in OTB100 dataset in terms of precision based on center location error and success rate based on overlap, and the experimental results show that our proposed method achieved substantially better performance. Promising results have shown that the proposed salient superpixel-based approach is effective to deformation, occlusion, and other challenges in object tracking.

Keywords Visual tracking · Superpixel · Saliency detection · Correlation matrix · Graph model

Introduction

In recent years, visual tracking methods based on different

task when the tracking object undergoes great pose variations and heavy occlusions in complex scenes. The existing tracking methods mainly deal with the target

Published: 28 June 2019

Salient Superpixel Visual Tracking with Graph Model and Iterative Segmentation

Jin Zhan, Huimin Zhao, Penggen Zheng, Hefeng Wu & Leijun Wang

Cognitive Computation **13**, 821–832 (2021) | Cite this article

Abstract

Visual object tracking is to locate an object of interest in a sequence of consecutive video frames, which is widely applied in many high-level computer vision tasks such as intelligent video surveillance and robotics. It is of great challenges for visual tracking methods to handle large target appearance variations caused by pose deformation, fast motion, occlusion, and surrounding environments in real-time videos. In this paper, inspired by human attention cognitive saliency model, we propose a visual tracking method based on salient superpixels which integrates the target appearance similarity and cognitive saliency, and helps to location inference and appearance model updating. The saliency of superpixel is detected by graph model and manifold ranking. We cluster the superpixels of the first four target boxes into a set

Applied Intelligence (2019) 49:2956–2969
https://doi.org/10.1007/s10489-019-01427-2

Sample awareness-based personalized facial expression recognition

Huihui Li[1] · Guihua Wen[1]

Published online: 13 February 2019

Abstract
The behavior of the current emotion classification model to recognize all test samples using the same method contradicts the cognition of human beings in the real world, who dynamically change the methods they use based on current test samples. To address this contradiction, this study proposes an individualized emotion recognition method based on context awareness. For a given test sample, a classifier that was deemed the most suitable for the current test sample was first selected from a set of candidate classifiers and then used to realize the individualized emotion recognition. The Bayesian learning method was applied to select the optimal classifier and then evaluate each candidate classifier from the global perspective to guarantee the optimality of each candidate classifier. The results of the study validated the effectiveness of the proposed method.

Keywords Facial expression recognition · Personalized classification · Dynamic selection · Bayesian

1 Introduction

Widely applied in mental health and human–computer interaction, emotion recognition is currently a popular research topic in the fields of computer vision and artificial intelligence [1–3] because it involves multiple disciplines, such as image processing, pattern recognition, and psychology. However, the diversity of facial expressions makes the emotion recognition difficult. For example, the collected facial images might be unidentifiable because of the lighting environment [4]. Moreover, the facial expressions of human beings are complicated and diverse, with fairly significant individual differences in skin color, age, and appearance. These differences place an added burden on machine learning.

Currently there are many emotion recognition methods, including deep learning and ensemble learning methods. They train an emotion classification model and then use this model to identify all test samples. This trained emotion classification model remains unchanged, without considering the practical conditions of each test sample. However, these methods are inconsistent with human cognition laws [5] in the real world. They model the inertial thinking and thus easily misclassify test samples [6]. Human beings change their methods dynamically based on the current test samples, instead of identifying all test samples with the same method. For example, human thinking follows the principle of simplicity (the Gestalt principle) [7]. Simple object recognition only needs simple methods, while complex object recognition needs complex methods [8]. However, most of the existing machine learning methods only consider the complexity of the whole dataset [9], or the complexity of the local neighborhood [10], without distinguishing the complexity of the object to be identified. In addition, for the same test sample, each person's emotional recognition ability is different, which is also true for classifiers. As the ensemble classifier emphasizes, the base classifiers should be diverse, indicating that many classifiers have different capabilities and complementarity [4, 10, 11]. In experiments, a classifier may work well for some test samples, but may often make mistakes for other test samples. In particular, when two classifiers are used to classify test samples, their classification ability may be totally opposite. Thus, it is rational to select the classifier dynamically in such circumstances [12, 13]. This can be implemented by first searching for the local neighborhood of each test sample, and then evaluating the classifier's capability through the samples in the local neighborhood in order to choose the most suitable classifier by which to classify the test samples [14]. The key issue of this method is that a set of candidate classifiers should be generated with high accuracy and diversity. The diversity of two classifiers is reflected in terms of the ability of each to classify the different samples. Ideally, classifiers should complement each other so that the most appropriate classifier can be selected for

✉ Guihua Wen
crghwen@scut.edu.cn

[1] School of Compute Science and Engineering, South China University of Technology, Guangzhou 510641, China

each new test sample [10, 11]. This is different from methods with static selection of classifiers, which occurs during model selection. During model selection, once the classifier is selected on the training set, it will classify all test samples without considering the differences among them. The study of dynamic classifier selection shows that it is a very effective tool for overcoming pathological classification problems, e.g., when the training data are too small or there are insufficient data by which to build the classification model [9].

The primary problem of dynamic classifier selection is measuring the ability of each classifier in classifying test samples. The most common methods for solving the problem are individual-based metrics and group-based metrics [13]. The former performs the measurement based on the classifier's individual information, such as rankings, accuracy, probability, and behavior, while the latter considers the relationship between the candidate classifiers. However, both measurement methods select the classifier according to the neighborhood of the test samples in the training set. It is difficult to obtain the globally considered performance using local estimation. Secondly, it is time-consuming to find the neighborhood of each test sample from a large training set. Cruz et al. proposed a method to dynamically select classifiers based on machine learning [14]. Using meta-features to describe the capabilities of each classifier in a local neighborhood, this method first dynamically selects classifiers for test samples through machine learning, and then uses the selected classifier to classify the test samples. The other type of methods not only consider the accuracy of the classifier but also the complexity of the problem, e.g., the complexity of the neighborhood of the test samples [9].

Based on the local neighborhood of the test samples, both aforementioned methods have two disadvantages. It is time-consuming to seek the neighborhood of a given test sample under large training data. Second, the performance of the classifier is limited to the local optimum rather than the global optimum. Hence, this paper proposes the sample awareness-based personalized (SAP) facial expression recognition method. SAP used the Bayesian learning method to select the optimal classifier from the global perspective, and then used the selected classifier to identify the emotional class of each test sample. The main contributions are that the idea of sample awareness is introduced to the field of emotion recognition, and a new emotion recognition method is proposed.

2 Related works

The SAP method proposed in this study is new in the field of emotion recognition. It selects the classifier dynamically for each test sample, which is different from the current dynamic classifier selection methods. The current dynamic classifier selection methods can be categorized into four types, which will be compared and analyzed in this paper. The recently developed methods for facial expression recognition are also presented, such as those based on 3D information of face and ensemble learning methods.

2.1 Dynamic classifier selection methods

2.1.1 Classification accuracy based on local neighborhood

These methods are based on the classification accuracy of the local neighborhood of the test sample, where the neighborhood is defined by the k nearest neighbors (KNN) algorithm [15] or the clustering algorithm [16]. For example, the overall local accuracy (OLA) selects the optimal classifier based on the accuracy of the classifier in the local neighborhood [17]. Another method is the local class accuracy (LCA), which uses posteriori information to calculate the performance of the base classifier for particular classes [18]. In addition, another method was proposed to sort the classifiers based on the number of consecutive correct classifications of samples in the local neighborhood. The larger the number, the higher the classifier is ranked to be selected [19].

There are two methods: A Priori (APRI), and A Posteriori (APOS) [20]. APRI selects the classifier based on the posterior probability of classes of the test sample in its neighborhood, which considers the distance from each neighborhood to the test sample. Unlike APRI, APOS considers each classifier's classification label for the current test sample. Based on these two methods, two new methods were proposed: KNORA-Eliminate (KE) and KNORA-Union (KU) [21]. KE only selects the classifier that correctly classifies all neighborhoods, whereas KU only selects the classifier that correctly classifies at least one neighborhood. Xiao et al. proposed a dynamic classifier ensemble model for customer classification with imbalanced class distribution. It utilizes the idea of LCA, but the prior probability of each class is used to deal with imbalanced data when calculating the classifier's performance [22]. The difference between these methods is that the local information is used in different ways, but they are both based on the local neighborhood of the test sample.

2.1.2 Decision template methods

Decision template methods are also based on the local neighborhood, but the local neighborhood is defined in the decision space [23] rather than in the feature space. The decision space consists of the classifier output of each sample, where each classifier output vector is a template. The similarities between the output vectors are then compared. For example, the K-nearest output profile (KNOP) method first defines the local neighborhood of the test sample in the decision space, and then uses a method similar to that by KNORA-E to select the classifiers that correctly classified test samples in the

neighborhood in order to form an ensemble by voting [24]. The multiple classifier behavior (MCB) method also defines the neighborhood in the decision space, but the selection is determined based on a threshold. Classifiers larger than the given threshold are used for the ensemble [25]. Although such methods are defined in the decision space, they are still based on the local neighborhood of the test samples.

2.1.3 Selection of candidate classifiers

The composition of candidate classifiers is very important for a dynamic classifier selection method since it must be accurate and diverse. In addition to methods that generate candidate classifiers using common ensemble classifier methods, there are also methods that focus on selecting training subsets for each candidate classifier [26]. For example, the particle swarm method directly selects a training set for each candidate classifier using the evolutionary algorithm [27]. The reason why a candidate classifier is generated by adopting different training subsets in the ensemble classifier is that it is easy to generate a large number of candidate classifiers that are likely to be similar rather than different. There are some methods that use heterogeneous candidate classifiers to make maintaining diversity easier.

2.1.4 Machine learning methods

The recently proposed method for dynamic selection of classifiers is based on machine learning and uses the local neighborhood features (such as meta-features of the test samples, the classification accuracy of the neighborhood samples, and the posterior probability of classes of the classified test samples) as the training samples for machine learning [14]. In the other method, the genetic algorithm was applied to divide training sets into subsets, each of which is used to train a classifier. The fitness function was defined as the accuracy of each classifier combined with the complexity of each training set [28]. Unlike these two methods, the method proposed in this study directly assigned each training sample to the classifier based on the Bayesian theorem. That is, the classifier was used as the class label of the training sample so that there was no need to calculate the neighborhood of the test sample and the machine learning could be global.

From the literatures mentioned above, it is discovered that dynamic classifier selection has not yet been applied to emotion recognition. The SAP proposed in this study is also different from currently available methods. It directly selected the candidate classifier according to the posterior classification accuracy calculated based on the Bayesian theorem. The evolutionary method was not used, and meta-features were not calculated. Instead, the proposed method directly endowed the training samples with classifier labels so that there was no need to calculate the neighborhood of the test samples. Since the learning was conducted throughout the training set, it was also global in nature.

2.2 Face images for facial expression recognition

When facial images are transformed into feature vectors, any single classifier can be used for expression recognition, such as support vector machines and neural networks. One of the differences among these methods is the application of facial image information. Expression recognition can be performed based on 2D static images, or expression recognition can be performed based on 3D or 4D images. Because of the sensitivity to illumination and head posture changes, the use of 2D static images is unstable. By contrast, facial expressions are the result of facial muscle movement, resulting in different facial deformations that can be accurately captured in geometric channels [29, 30]. In such cases, using 3D or 4D images are the trend because they enable use of more image information.

Previous 3D expression recognition methods focus on the geometric representation of a single face image [31–34]. Currently, 3D video expression recognition methods emphasize modeling dynamic deformation patterns through facial scanning sequences. For example, a heuristic deformable model for static and motion information of the video was constructed, and then the hidden Markov model (HMM) was applied to recognize expressions [35]. Another method extracted motion features between adjacent 3D facial frames, and then utilized HMM to perform facial expression recognition [36]. Temporal deformation clues of 3D face scanning can also be captured using dynamic local binary pattern (LBP) descriptors, and then an SVM can be applied to perform the expression recognition [37]. Another novel method is the conditional random forest, which aims to capture low-level expression transition patterns [38]. When testing on a video frame, pairs are created between this current frame and previous ones, and predictions for each previous frame are applied to draw trees from pairwise conditional random forests (PCRF). The pairwise outputs of PCRF are averaged over time to produce robust estimates. A more complex approach is to use a set of radial curves to represent the face, to quantify the set using Riemann-based shape analysis tools, and to then classify the facial expressions using LDA and HMM [39, 40]. There are also methods for facial expression recognition using 4D face data. For example, scattering operators are expanded on key 2D and 3D frames to generate text and geometric facial representations, and then multi-kernel learning is applied to combine different channels of facial expression recognition to obtain the final expression label [41, 42].

Deep learning has also been applied to recognize facial expressions [43]. For example, a novel deep neural network-driven feature learning method was proposed and applied to multi-view facial expression recognition [44]. The input of the network is scale invariant feature transform (SIFT) features

that correspond to a set of landmark points in each facial image. There is a simple method to recognize facial expressions that uses a combination of a convolutional neural network and specific image preprocessing steps [45]. It extracts only expression-specific features from a face image, and explores the presentation order of the samples during training. A more powerful facial feature method called deep peak–neutral difference has also been proposed [46]. This difference is defined as the difference between two deep representations of the fully expressive (peak) and neutral facial expression frames, where unsupervised clustering and semi-supervised classification methods automatically obtain the neutral and peak frames from the expression sequence. With the development of deep learning, some studies emphasize the modeling of dynamic shape information of facial expression motion, and then adopt end-to-end deep learning [41, 42, 47–49], where a 4D face image network for expression recognition uses a number of generated geometric images. A hybrid method uses a contour model to implement face detection, uses a wavelet transform-based method to extract facial expression features, and uses a robust nonlinear method for feature selection; finally, the HMM is used to perform facial expression recognition [50].

The SAP method is different from the above expression recognition methods. These methods are thus taken as candidate classifiers for SAP so as to further improve SAP's performance. This also allows SAP to easily exceed them.

2.3 Ensemble learning for facial expression recognition

Ensemble learning is also used for facial expression recognition, which can be implemented by data integration, feature integration, and decision integration. Data fusion refers to the fusion of facial, voice, and text information. For example, the fusion of video and audio is applied to recognize emotions [51]. Meanwhile, the combination of facial expression data and voice data is utilized to identify emotions [52]. Another approach combines thermal infrared images and visible light images, using both feature fusion and decision fusion [53]. This approach extracts the active shape model features of the visible light image and the statistical features of the thermal infrared image model, and then uses a Bayesian network and support vector machine to make respective decisions. Finally, these decisions are fused in the decision layer to obtain the final emotion label. There is an automatic expression recognition system that extracts the geometric features and regional LBP features, and fuses them with self-coding. Finally, a self-organizing mapping network is used to perform expression recognition [54]. When the face image is divided into several regions, and the features of each region are extracted using the LBP method, the evidence theory can be used to fuse these features [55]. Furthermore, the fusion of both Gabor features and LBP features can be applied to recognize expressions [56]. Some methods also use SIFT and deep convolution neural networks to extract features, and then use neural networks to fuse these features [57]. The decision level integrates the final decision information of multiple learning models. Each learning model participates in the processes of preprocessing, feature extraction, and decision-making. The fusion layer makes the final inference by evaluating the reliability of each member's decision-making information. For example, Wen et al. fused multiple convolutional neural network models by predicting the probability of each expression class for the test sample [4]. Zavaschi et al. extracted Gabor features and LBP features for facial images, and then generated a number of SVM classifiers. Finally, some classifiers were selected by a multi-objective genetic algorithm, and the final expression label was obtained by integrating these selected classifiers [58]. Moreover, Wen et al. proposed an integrated convolutional echo state network and a hybrid ensemble learning approach for facial expression classification [10, 11].

The SAP method is different from these ensemble learning methods for emotion recognition. SAP dynamically selects a classifier from multiple classifiers for the test sample. When a large number of candidate classifiers are available, SAP is more likely to find the most suitable classifier for the test sample. These aforementioned ensemble learning methods can be taken as candidate classifiers for SAP so that SAP's performance can be further improved and easily exceed that of the existing ensemble learning methods.

3 Proposed method

In the real world, different experts may have different abilities to identify the same sample. For example, it is justifiable to see the best doctor, but the "best doctor" is different for each disease. Similarly, each person wants to attend the best school, but different people have different definitions of the "best school." Therefore, this study proposed the SAP method for facial expression recognition.

Figure 1 shows the structure of the method. The method differs from the ensemble method that averages all classifiers and weakens the strongest classifier so that it is theoretically inferior to the best classifier. SAP also differs from the model selection method that seeks the best classifier from all training samples rather than each individual sample. SAP considers each test sample to have its own optimal classifier because each expert has his own strengths.

The SAP method calculates the ability of each candidate classifier to classify each sample on the training set to find the most suitable classifier for each training sample based on the Bayesian theorem. Using this approach, a new training set, $\Phi\{(x_i, c_i)\}$, $c_i \in C$, was constructed; that is, a label was assigned to each training sample as the optimal classifier by

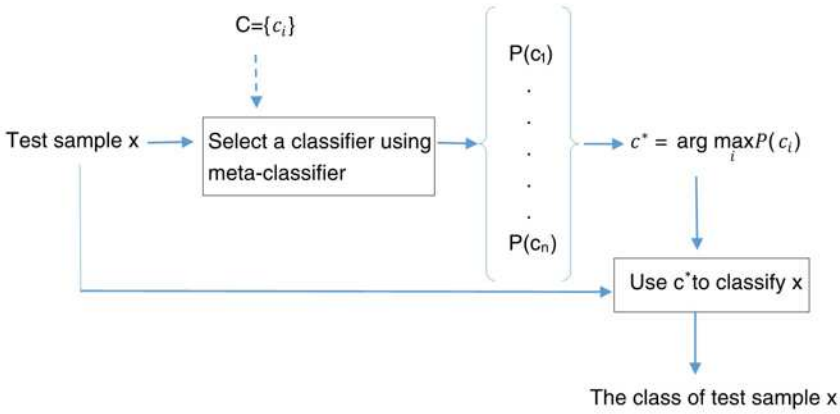


Fig. 1 Classification process of SAP

which to classify this sample. On this new training set, a new classifier was then trained to assign the most suitable classifier for each test sample.

3.1 Labeling each sample with the classifier name

$X = \{x_i | x_i \in \mathbb{R}^n\}$ is a training sample set, $Y = \{y_i | y_i \in L\}$ is the corresponding label set, and L is the set of the labels of the samples. There is a classifier set $C = \{c_i | c_i \in \mathbb{Z}\}$, where classifier $c \in C$ was used to classify sample x and calculate the probability that it would correctly classify x. The k-fold cross-validation method was applied to train the classifiers with some training samples, and then the classifiers were used to classify the test sample. If the test sample was classified correctly, $P(x|c)$ could be easily calculated. The k-fold cross-validation method was used to divide the training set into subsets as follows:

$$X = X_1 \cup \cdots \cup X_i \cup \cdots \cup X_k, \tag{1}$$

$$X_i \cap X_j = \varnothing, |X_i| = |X_j|, \tag{2}$$

$$Y = Y_1 \cup \cdots \cup Y_i \cup \cdots \cup Y_k, \tag{3}$$

$$|X_i| = |Y_i|. \tag{4}$$

Suppose that the discriminant function of classifier c in the training set X_j is defined as $g_{c,X_j \subseteq X} : X_j \rightarrow Y_j$. The prior probability of classifier c was calculated as follows. The higher the classification accuracy, the more likely it was to be selected as the optimal classifier:

$$p(c) = \frac{1}{k} \sum_{j=1}^{k} \sum_{i}^{|X|} 1 g_{c,X \setminus X_j}(x_i) = y_i. \tag{5}$$

The prior probability for classifier c to correctly classify x_i was calculated using the following equation:

$$P(x_i|c) = \frac{1}{k} \sum_{j}^{k} 1 g_{c,X \setminus X_j}(x_i) = y_i. \tag{6}$$

The goal was to calculate $P(c|x)$, which is the probability that each classifier will be selected based on the test sample. This allows us to select the most suitable classifier from the candidate classifier set to classify the test sample.

According to the Bayesian theorem, the following equation can be obtained:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}. \tag{7}$$

This is similar to the assumption of the Naive Bayesian classifier, $\text{allp}(x_i) = p(x_j)$. According to the above formula, each training sample was labeled with the classifier name to construct a new training dataset. When the probability of the classifier chosen based on x is greater than a certain threshold,

$$D_i = \{(x, c_i) | P(c_i|x) > \delta_i, x \in X, c_i \in C\}, \tag{8}$$

$$S = \cup_{i=1}^{|C|} D_i. \tag{9}$$

The candidate classifiers were constructed by D:

$$D = \left\{ (x, c_i) | x \in S, c_i = \arg\max_i P(c_i|x) \right\} \tag{10}$$

Once the training sample set D was labeled with the classifier name, another classification algorithm, φ, was selected to be trained on this set so as to obtain a new classification function as follows:

$$h_{\varphi,D} : X \rightarrow 2^C, \tag{11}$$

$$c = \arg\max_i P\left(c_i \in h_{\varphi,D}(x)\right). \tag{12}$$

Given a test sample x, we selected a suitable classifier, c, to classify the test sample.

3.2 SAP for emotion recognition

Given the inputs of the training set X, the validation set X_v, the classifier set $C = \{c_i\}$, the threshold parameter σ, the test

Fig. 2 Sample images from the experimental databases

sample x, as well as the output y (the label of the test sample), the SAP algorithm was described as follows:

1. |C| classifiers were trained on training set X.
2. Training set X was divided into k groups using the k-fold cross-validation method.
3. For $j = 1$ to k:

   (a) The jth fold of the training set was taken from the training set to train each classifier c.
   (b) The classifier c was used to classify each sample x_i in the validation set X_v.
   (c) The number of times that each sample in the validation set X_v was correctly classified in all folds was calculated, and then the probability $p(x_i|\, c)$ was computed.

   End

4. The probability $p(x_i|\, c)$ was normalized.
5. The probability $p(c|\, x_i)$ was calculated based on the Bayesian theorem so as to assign a classifier name to each training sample as the label.
6. For $i = 1$ to |C|:

$$D_i = \{(x, c_i) | P(c_i|x) > \sigma \,\&\, P(c_i|x) > P(c_j|x), x \in Xv\}.$$

   End

7. $S = \cup D_i$
8. The classification algorithm φ was used to train a meta-classifier $h_{\varphi,\, D} : D \to 2^C$.
9. The classifier $c_i = \arg\max_l P(c_i \in h_{\varphi,D}(x))$ was selected.
10. The classifier c_i was used to classify the test samples x so as to obtain the class y.

3.3 Time complexity analysis

As in Step 3 of SAP training $k \times |C|$ classifiers, which involved a complexity of $k \times \max(O(c_i))$, the other steps of SAP were linear. The greatest complexity of the algorithm laid in training or testing a classifier, and therefore the complexity of the entire algorithm was $\max(O(c_i))$. SAP spent the most time on training the classifiers using the k-fold cross-validation method. However, this calculation was only performed once during the training. The trained model was used to directly classify the test samples, and there was no need for a recalculation. Therefore, SAP was less complex than all dynamic algorithms based on the local neighborhoods.

4 Experimental results

4.1 Objective

The effectiveness of the proposed method was demonstrated by conducting experiments on two standard datasets. In principle, there are many alternative classifiers for the

Table 1 The distribution of samples in the two experimental databases

Emotions Databases	Angry	Disgust	Fear	Happy	Sad	Surprised	Neutral	Total
FER2013-TRAIN	3995	436	4097	7215	4830	3171	4965	28,709
FER2013-PUBLIC	467	56	496	895	653	415	607	3589
FER2013-TEST	491	55	528	879	594	416	626	3589
RAF2017-TRAIN	705	717	281	4772	1982	1290	2524	12,271
RAF2017-TEST	162	160	74	1185	478	329	680	3068

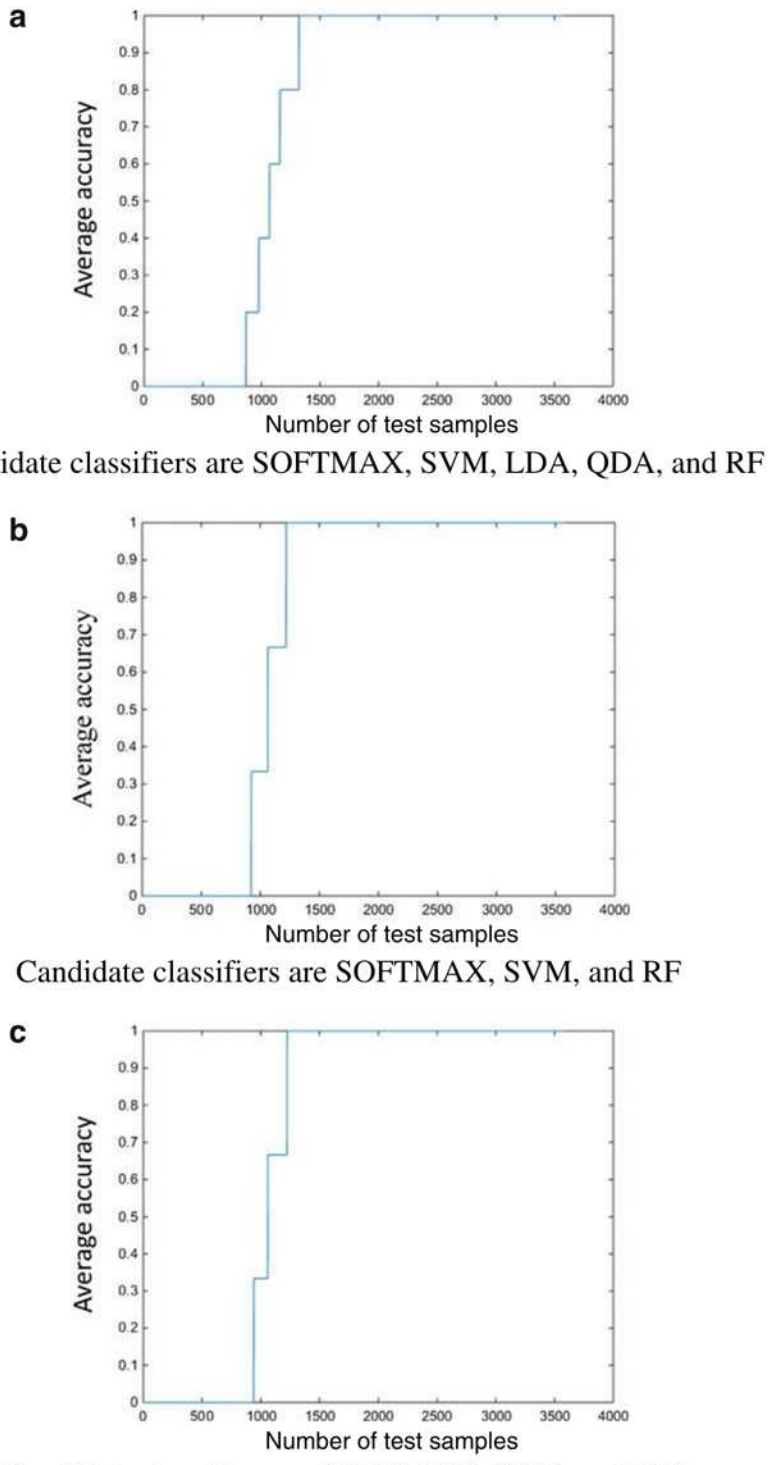


Fig. 3 Distribution of test samples against the classification satisfiability

proposed method. However, in the experiments, the most representative methods were chosen, i.e., SOFTMAX [4, 59], SVM [60], LDA [60], QDA [60], and RF [61]. Since the SOFTMAX classifier is a widely used classifier for deep learning, SAP can be applied to deep learning with the SOFTMAX classifier chosen. SVM is one of the best classifiers for small training samples. LDA and QDA are the simplest linear classifiers, whereas RF is the most representative ensemble classifier. For these candidate classifiers, default parameters were used in the experiments. The

LDA algorithm was used as the meta-classifier because it is simple and fast. In this way, two objectives will be obtained. One is to prove that the dynamic selection of classifiers is superior to the constant use of a single classifier. The other is to illustrate that the proposed method outperforms some ensemble algorithms.

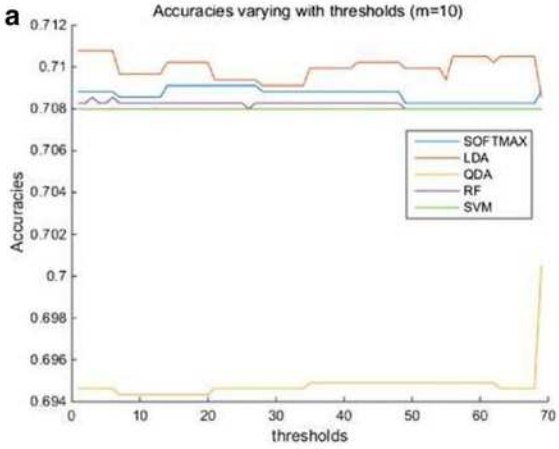
4.2 Experimental data

The deep neural network is currently the most effective approach for extracting the features of images, but it requires a large amount of training data. Therefore, FER2013 [62] and RAF [63] are selected as the experimental data. They are generally recognized as benchmark databases. Sample images from these databases are shown in Fig. 2.

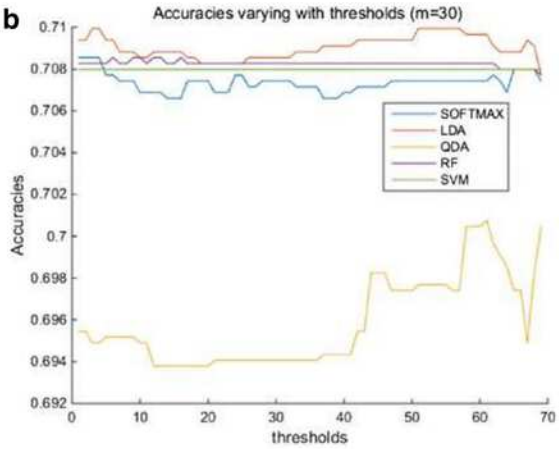
FER2013 has the larger amount of data and its images are the most difficult to distinguish. Each sample in the database has great differences in age, facial orientation, and so on. It is also closest to real world data, with the human emotion recognition rate in this database is 65 ± 5%. At the same time, the images in the database are all gray-scale images with a size of 48 × 48 pixels. The samples are divided into seven categories: anger, disgust, fear, happiness, neutral, sadness, and surprise. This database consists of three parts: FER2013-TRAIN for training a deep neural network, FER2013-PUBLIC as the validation set, and FER2013-TEST as the test set. Their sample distributions are shown in Table 1.

The Real-world Affective Faces Database (RAF 2017) was constructed by analyzing 1.2 million labels of 29,672 greatly diverse facial images downloaded from the Internet. Images in this database vary greatly in subject age, gender, ethnicity, head poses, lighting conditions, and occlusions. For example, the subjects in the database range in age from 0 to 70 years old. Fifty two percent are female, 43% are male, and 5% ambiguous; meanwhile, 77% are Caucasian, 8% are African-American, and 15% are Asian [62]. Therefore, it has large diversity across a total of 29,672 real-world images, with seven classes of basic emotions and 12 classes of compound emotions. To be able to objectively measure the performance for the following testing. In our experiments, the database with seven basic emotions is considered; these emotions are anger, disgust, fear, happiness, neutral, sadness, and surprise. This database is split into a training set RAF2017-TRAIN with 12,271 samples and a test set RAF2017-TEST with 3068 samples.

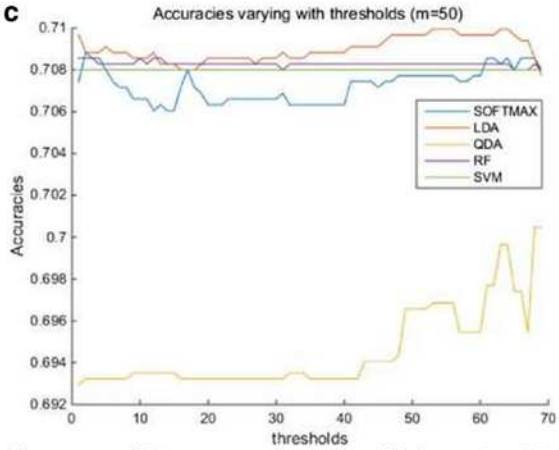
The features of all datasets were extracted using the deep neural network model [59]. Parameter analysis and time complexity analysis were performed on FER2013 since it is harder to be classified. In SAP, the *j*-th fold of training samples was taken from the training set to train the classifier, and FER2013-PUBLIC was taken as the validation set.



The average of ten test accuracies on the validation set (m= 10), which varies with the threshold



The average of 30 test accuracies on the validation set (m = 30), which varies with the threshold



The average of 50 test accuracies on the validation set (m= 50), which varies with the threshold

Fig. 4 Relationship between classification accuracy of SAP and the threshold σ

Table 2 The number of samples assigned to each classifier with the optimal threshold ($m = 10$)

	Number of samples assigned to be classified to each candidate classifier					
Meta-classifier	SOFTMAX	LDA	QDA	RF	SVM	Accuracy (%)
SOFTMAX	50	15	42	36	3446	70.91
LDA	133	75	112	167	3102	**71.08**
QDA	562	0	0	3027	0	70.05
RF	2		1	2	3584	70.86
SVM	0	0	0	0	3589	70.80

Bold data indicates the best meta classifier with the best accuracy

4.3 Evaluation on complementarity among candidate classifiers

The key to SAP is the complementarity among the candidate classifiers. To objectively evaluate the complementarity among the candidate classifiers, the concept of classification satisfiability was proposed. The probability measure for any sample to be correctly classified is referred to as classification satisfiability, which can be calculated using the following equation:

$$\mu(x) = \frac{\sum_1^n f_i(x)}{n} \tag{13}$$

where n is the number of classifiers. If classifier f_i can correctly classify x, then $f_i(x) = 1$; otherwise $f_i(x) = 0$. The greater the classification satisfiability, the more likely the sample is to be correctly classified.

Figure 3 shows the distribution of the classification satisfiability of the test samples for a given set of candidate classifiers, where FER2013 was used. The samples were ranked according to classification satisfiability from high to low. In Fig. 3a, when the candidate classifiers SOFTMAX, SVM, LDA, QDA, and RF were used, 868 samples were classified completely incorrectly, 2270 samples were correctly classified, and 451 samples were correctly classified by at least one classifier. Figure 3b shows that when the candidate classifiers SOFTMAX, SVM, and RF were used, 922 samples were classified completely incorrectly, 2371 samples were correctly classified, and 296 samples were correctly classified by at least one classifier. Figure 3c illustrates that when the candidate classifiers SOFTMAX, SVM, and LDA were used, 939 samples were classified completely incorrectly, 2366 samples were correctly classified, and 284 samples were correctly classified by at least one classifier.

In Fig. 3, there were approximately 900 samples whose classification satisfiability was 0, indicating that these samples could not be correctly classified by any classifier. It was inevitable for them to be misclassified. This indicated that the candidate classifier set is incomplete and needs to be extended so as to reduce the occurrence of such situations. As shown in Fig. 3, the number of erroneously classified samples was different for different sets of candidate classifiers. Since there was a maximum number of candidate classifiers in Fig. 3, a minimum number of misclassified samples was expected. Moreover, the greater the number of candidate classifier sets, the greater the number of samples whose classification satisfiability was greater than zero. This indicates that some candidate classifiers can correctly classify these samples. In these cases, the accuracy of the meta-classifier is extremely important. Ideally, the meta-classifier should be able to select the candidate classifier that can correctly classify these samples.

4.4 Parameter performance analysis

Since the SAP algorithm used the machine learning method (meta-classifiers) to assign classifiers to test samples, the meta-classifiers needed to be trained by the samples

Table 3 The number of samples assigned to each classifier with the optimal threshold ($m = 30$)

	Number of samples assigned to be classified to each candidate classifier					
Meta-classifier	SOFTMAX	LDA	QDA	RF	SVM	Accuracy (%)
SOFTMAX	58	31	42	156	3302	70.86
LDA	135	105	77	176	3096	**70.99**
QDA	35	357	98	0	3099	70.08
RF	1	1	1	2	3584	70.86
SVM	0	0	0	0	3589	70.80

Bold data indicates the best meta classifier with the best accuracy

Table 4 The number of samples assigned to each classifier with the optimal threshold ($m = 50$)

Meta-classifier	Number of samples assigned to be classified to each candidate classifier					Accuracy (%)
	SOFTMAX	LDA	QDA	RF	SVM	
SOFTMAX	43	34	49	131	3332	70.88
LDA	110	124	98	54	3203	**70.99**
QDA	2	418	39	3130	0	70.05
RF	1	1	0	2	3585	70.86
SVM	0	0	0	0	3589	70.80

Bold data indicates the best meta classifier with the best accuracy

whose labels were candidate classifier names. The labels for these samples were automatically completed on the training and verification sets, and their classification satisfiability was found to be the average of the test accuracy on the cross-validation set. The greater the classification satisfiability, the more reliable the classifier name that was labeled on the test sample. Therefore, a classification satisfiability below the threshold may have been wrong and therefore should be removed from the training samples of the meta-classifier.

FER2013-TRAIN was divided into 100 pieces for cross-validation, 99 of which were used as the training set each time. FER2013-PUBLIC was used as the validation set, with the validation results taken out m times. For example, $m = 10$ means that the validation results obtained for the first ten times were taken out, and then the average of the test accuracy on the validation set was calculated to obtain the classification satisfiability for each sample on the validation set. Based on the given threshold parameters, the samples in the validation set with values larger than the threshold were selected as the training samples of the meta-classifier. After the meta-classifier was trained, each test sample in FER2013-TEST would be assigned a candidate classifier.

The classification effect of SAP was related to m and the threshold σ of the classification satisfiability. The results in Fig. 4 demonstrate that different thresholds affected the classification accuracy of SAP. However, the range of the best results was relatively large and stable. This indicated that the optimal threshold σ could be easily obtained experimentally. Secondly, the optimal thresholds corresponding to different meta-classifiers were different. Although the classification accuracy of SAP varied with different values of m, its change with threshold σ was similar, which indicated that a relatively small m could be selected as the threshold parameter to reduce the time cost of the experiment.

Figure 4 also shows that the effectiveness of different meta-classifiers was different because the number of test samples assigned to each candidate classifier was different. As shown in Tables 2, 3 and 4, the more dispersed the assigned test samples, the more complementary they were and the more effective the classification. Additionally, the assignments were unbalanced. Effective candidate classifiers were in the majority. However, when all were assigned to the majority, the classification became ineffective. This behavior was associated with unbalanced data, which could be further improved with methods that are good at dealing with classification of unbalanced data.

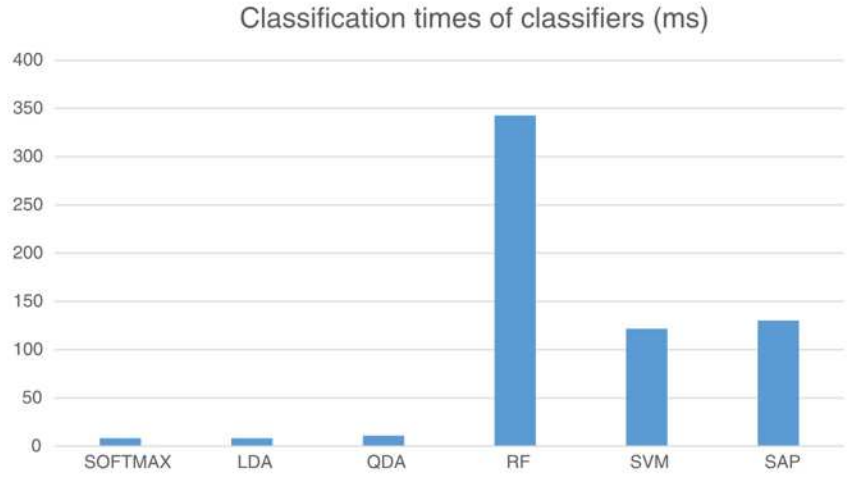


Fig. 5 Comparison of candidate classifiers and SAP in terms of classification time

Table 5 Recognition rates of SAP and the candidate classifiers on the three test sets

	Candidate classifiers					Ensemble			
Data	SOFTMAX	LDA	QDA	RF	SVM	Ens1	Ens2	Ens3	**SAP**
FER2013	0.6996	0.6999	0.6949	0.6941	*0.7080*	0.7052	0.7035	0.7069	**0.7108**
RAF2017	*0.8165*	0.8132	0.8145	0.8136	0.8132	**0.8184**	0.8158	0.8171	**0.8181**

The bold entry shows that it is the best result in the compared methods

The experimental results show that LDA as the optimal meta-classifier was not only effective but also fast. In later experiments, only LDA was used as the meta-classifier. SVM as the meta-classifier led to the worst effect since it assigned all the test samples to itself.

4.5 Time complexity analysis

When classifying the test samples, SAP first used a meta-classifier to assign a candidate classifier to each test sample, and then used the selected candidate classifier to classify the test sample, which added to the classification time. However, LDA was applied as meta-classifier in this study. Since it worked quickly, the time it added to classification was negligible. As shown in Fig. 5, it was much smaller than the maximum RF but larger than the minimum LDA and QDA. This is because SAP assigned many samples to SVM and RF, which thereby improved the emotion recognition accuracy. Among all the candidate classifiers, SVM had the highest accuracy; however, SAP was more accurate than SVM, and its classification time was only slightly bigger. Therefore, the comprehensive advantages of SAP are noteworthy.

4.6 Comparison of standard datasets

SAP only selected the optimal classifier from the candidate classifiers. We addressed the question of whether it was better than the single and ensemble versions of these candidate classifiers. For FER2013, each method adopts FER2013-TRAIN as the training set and FER2013-TEST as the test set. For RAF2017, each method adopts RAF2017-TRAIN as the training set and RAF2017-TEST as the test set.

All the results are shown in Table 5, where Ens1 denotes the combination of SOFTMAX, LDA, QDA, RF, and SVM; Ens2 indicates the combination of SOFTMAX, RF, and SVM; and Ens3 denotes the combination of SOFTMAX, LDA, and SVM. It can be observed that SAP is better than both the ensemble classifier and single candidate classifier for the FER2013 database. The ensemble classifier is not better than the best candidate classifier SVM, but it is more stable. Besides, the ensemble method and selective ensemble method were relatively effective in emotion recognition; however, as shown as in Table 6, the SAP method was shown to be superior to some ensemble methods, where the accuracy rate of ensemble methods comes directly from the original literature. Due to different techniques used in ensemble methods, such as feature extraction, the comparison of effectiveness here should only be used as a reference.

On RAF2017, SAP still outperforms any single candidate classifier. However, it seems that SAP is slightly worse than Ensemble 1, which contains all candidate classifiers, but it works faster.

Table 6 Recognition results obtained by the selective ensemble methods on FER2013

Selective integration algorithm	Accuracy (%)
Kappa [64]	68.74
QSEP [65]	68.49
DFEP [65]	68.82
Inconsistent EP [65]	69.38
DREP [66]	70.05
Complementarity method [67]	68.82
OO [68]	70.52
MRMREP [59]	70.66
ECNN [4]	69.96
SAP	**71.08**

The bold entry shows that it is the best result in the compared methods

5 Conclusion

The SAP method proposed in this study is innovative because it adopts a global approach to dynamically selecting the optimal classifier for each test sample. It used the Bayesian theorem to calculate the posterior probability of each sample, and then labeled the candidate classifier name to each sample according to its posterior probability. As a global method, SAP can be used to avoid the effects of noise and to reduce the time it takes to search for local neighborhoods when classifying the test samples. The meta-classifier, which was linear, was shown to be efficient and fast.

Although SAP requires a large number of basic classifiers, it is different from ensemble learning. The ensemble classification method needs to run multiple classifiers simultaneously to classify the test samples, which makes their work comparatively slower. It is the same for all test samples. SAP selects

the classifier most suitable to classify a given test sample from the given basic classifiers. This is more consistent with human cognition laws. In experiments, SAP's effectiveness in emotion recognition was shown to be significantly better than that of any candidate classifier, and the same was nearly true for the recognition effect of the ensemble of these candidate classifiers. Secondly, SAP is different from the traditional model selection method. Model selection involves selecting a suitable model by testing on the training data, and then this model is used to classify all test samples. In the process of classification, this model is unchanged. SAP changes dynamically according to the test sample, and therefore has a personalized classification ability.

The key technique of SAP is that it requires a method to select a suitable classifier for any given test sample. This classifier is critical for ensuring the accuracy of SAP. At present, a linear classifier is selected. In the future, we will choose a more suitable classifier to finish this task, and nonlinear classifiers may be considered. Secondly, SAP depends on a large number of candidate classifiers being available. The more candidate classifiers available, the more suitable a classifier can be selected for the given test samples, thus leading to greater classification accuracy. In the future, more candidate classifiers will be considered, and these candidate classifiers should be diverse. Finally, the advantage of SAP is that it makes full use of global information, but the disadvantage is that it fails to utilize local information. In the future, we will consider both global and local information simultaneously so as to select a more accurate classifier to classify a given test sample. Therefore, the accuracy of SAP can be further improved.

Acknowledgments This study was supported by China National Science Foundation (Grant Nos. 60973083 and 61273363), Science and Technology Planning Project of Guangdong Province (Grant Nos. 2014A010103009 and 2015A020217002), and Guangzhou Science and Technology Planning Project (Grant No. 201504291154480, 201604020179, 201803010088).

References

1. Zhang KH, Huang YZ, Du Y, Wang L (2017) Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Trans Image Process 26(9):4193–4203
2. Zeng NY, Zhang H, Song BY, Liu WB, Li YR, Dobaie AM (2018) Facial expression recognition via learning deep sparse autoencoders. Neurocomputing 273:643–649
3. Choi I, Ahn H, Yoo J (2018) Facial expression classification using deep convolutional neural network. J Electr Eng Technol 13(1): 485–492
4. Wen GH, Hou Z, Li HH, Li DY, Jiang LJ, Xun EY (2017) Ensemble of deep neural networks with probability-based fusion for facial expression recognition. Cogn Comput 9(5): 597–610
5. Wen G, Wei J, Wang J, Zhou T, Chen L (2013) Cognitive gravitation model for classification on small noisy data. Neurocomputing 118:245–252
6. Corcoran K, Hundhammer T, Mussweiler T (2009) A tool for thought! When comparative thinking reduces stereotyping effects. J Exp Soc Psychol 45:1008–1011
7. Baruchello G (2015) A classification of classic, gestalt psychology and the tropes of rthetoric. New Ideas Psychol 26:10–24
8. Smith MR, Martinez T, Giraud-Carrier C (2014) An instance level analysis of data complexity. Mach Learn 95:7225–7256
9. Brun AL, Alecus B Jr, Oliveira LS, Enembreck F, Sabourin R (2018) A framework for dynamic classifier selection oriented by the classification problem difficulty. Pattern Recogn 76:175–190
10. Wen GH, Li HH, Li DY (2015) An ensemble convolutional echo state networks for facial expression recognition. In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xian, China, pp 873–878
11. Li D, Wen G, Hou Z, Huan E, Hu Y, Li H (2018) RTCRelief-F: an effective clustering and ordering-based ensemble pruning algorithm for facial expression recognition. Knowl Inf Syst:1–32
12. Krawczyk B (2016) Dynamic classifier selection for one-class classification. Knowl-Based Syst 1307:43–53
13. Britto AS Jr, Sabourin R, Oliveira LES (2014) Dynamic selection of classifiers—a comprehensive review. Pattern Recogn 47:3665–3680
14. Cruz RMO, Sabourin R, Cavalcanti GDC, Ren TI (2015) META-DES: a dynamic ensemble selection framework using META-learning. Pattern Recogn 48:1925–1935
15. Ko AHR, Sabourin R, Britto Jr AS (2008) From dynamic classifier selection to dynamic ensemble selection. Pattern Recogn 41:1735–1748
16. Kuncheva L (2002) Switching between selection and fusion in combining classifiers: an experiment. IEEE Trans Syst Man Cybern 32(2):146–156
17. Mendialdua I, Martínez-Otzeta JM, Rodriguez-Rodriguez I, Ruiz-Vazquez T, Sierra B (2015) Dynamic selection of the best base classifier in one versus one. Knowl-Based Syst 85:298–310
18. Didaci L, Giacinto G, Roli F, Marcialis GL (2005) A study on the performances of dynamic classifier selection based on local accuracy estimation. Pattern Recogn 38(11):2188–2191
19. Sabourin M, Mitiche A, Thomas D, Nagy G (1993) Classifier combination for handprinted digit recognition. In: Second International Conference on Document Analysis and Recognition, pp 163–166
20. Giacinto G, Roli F (1999) Methods for dynamic classifier selection. In: 10th International Conference on Image Analysis and Processing, pp 659–664
21. Ko AHR, Sabourin R, Britto AS Jr (2008) From dynamic classifier selection to dynamic ensemble selection. Pattern Recogn 41:1735–1748
22. Xiao J, Xie L, He C, Jiang X (2012) Dynamic classifier ensemble model for customer classification with imbalanced class distribution. Expert Syst Appl 39:3668–3675

23. Kuncheva LI, Bezdek JC, Duin RPW (2001) Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recogn 34:299–314
24. Cavalin PR, Sabourin R, Suen CY (2012) Logid: an adaptive framework combining local and global incremental learning for dynamic selection of ensembles of HMMs. Pattern Recogn 45(9): 3544–3556
25. Giacinto G, Roli F (2001) Dynamic classifier selection based on multiple classifier behavior. Pattern Recogn 34:1879–1881
26. Szepannek G, Bischl B, Weihs C (2009) On the combination of locally optimal pairwise classifiers. Eng Appl Artif Intell 22:79–85
27. de Souza BF, de Carvalho A, Calvo R, Ishii RP (2006) Multiclass svm model selection using particle swarm optimization. In: Sixth International Conference on Hybrid Intelligent Systems, IEEE, p 31
28. Brun AL, AlceuS B Jr, Oliveira LS, Enembreck F, Sabourin R (2018) A framework for dynamic classifier selection oriented by the classification problem difficulty. Pattern Recogn 76:175–190
29. Fang T, Zhao X, Ocegueda O, Shah SK, Kakadiaris IA (2011) 3D facial expression recognition: a perspective on promises and challenges. In: IEEE International Conference on Automatic Face and Gesture Recognition, vol 28, pp 603–610
30. Zhen Q, Huang D, Wang Y, Chen L (2016) Muscular movement model-based automatic 3D/4D facial expression recognition. IEEE Trans Multimedia 18(7):1438–1450
31. Zhao X, Huang D, Dellandra E, Chen L (2010) Automatic 3D facial expression recognition based on a Bayesian belief net and a statistical facial feature model. In: IEEE/IAPR International Conference on Pattern Recognition
32. Li H, Chen L, Huang D, Wang Y, Morvan J-M (2012) 3D facial expression recognition via multiple kernel learning of multi-scale local Normal patterns. In: IEEE/IAPR International Conference on Pattern Recognition
33. Zhen Q, Huang D, Wang Y, Chen L (2015) Muscular movement model based automatic 3D facial expression recognition. In: International Conference on MultiMedia Modeling
34. Li H, Ding H, Huang D, Wang Y, Zhao X, Morvan J-M, Chen L (2015) An efficient multimodal 2D + 3D feature-based approach to automatic facial expression recognition. Comput Vis Image Underst 140:83–92
35. Yin L, Chen X, Sun Y, Worm T, Reale M (2008) A high-resolution 3D dynamic facial expression database. In: IEEE International Conference on Automatic Face and Gesture Recognition
36. Sandbach G, Zafeiriou S, Pantic M, Rueckert D (2012) Recognition of 3D facial expression dynamics. Image Vis Comput 30(10):762–773
37. Fang T, Zhao X, Shah SK, Kakadiaris IA (2011) 4D facial expression recognition. In: IEEE International Conference on Computer Vision Workshops, pp 1594–1601
38. Dapogny A, Bailly K, Dubuisson S (2017) Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests. IEEE Trans on Affect Comput 99:1–14
39. Drira H, Ben Amor B, Daoudi M, Srivastava A, Berretti S (2012) 3D dynamic expression recognition based on a novel deformation vector field and random forest. In: IEEE International Conference on Pattern Recognition, pp 1104–1107
40. Ben Amor B, Drira H, Berretti S, Daoudi M, Srivastava A (2017) 4D facial expression recognition by learning geometric deformations. IEEE Trans Cybern 44(12):2443–2457
41. Yao Y, Huang D, Yang X, Wang Y, Chen L (2018) Texture and geometry scattering representation based facial expression recognition in 2D+3D videos. In: ACM Transactions on Multimedia Computing and Applications
42. Joan B, Stephane M (2013) Invariant scattering nonvolution networks. IEEE Trans Pattern Anal Mach Intell 35(8):1872–1886
43. Ding H, Zhou SK, Chellappa R (2017) Facenet2expnet: regularizing a deep face recognition net for expression recognition. In: 12th IEEE International Conference on Automatic Face & Gesture Recognition, pp 118–126
44. Zhang T, Zheng W, Cui Z, Zong Y, Yan J (2016) A deep neural network-driven feature learning method for multi-view facial expression recognition. IEEE Trans Multimedia 18(12):2528–2536
45. Lopes AT, Aguiar ED, Souza AFD, Oliveira-Santos T (2017) Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recogn 61: 610–628
46. Chen J, Ruyi X, Liu L (2018) Deep peak-neutral difference feature for facial expression recognition. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-5909-5
47. Yang X, Huang D, Wang Y, Chen L (2015) Automatic 3D facial expression recognition using geometric scattering representation. In: IEEE International Conference on Automatic Face and Gesture Recognition
48. Liu Y, Zeng J, Shan S, Zheng Z (2018) Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition. In: 13th IEEE International Conference on Automatic Face & Gesture Recognition
49. Li W, Huang D, Li H, Wang Y (2018) Automatic 4D facial expression recognition using dynamic geometrical image network. In: 13th IEEE International Conference on Automatic Face & Gesture Recognition
50. Siddiqi MH (September 2018) Accurate and robust facial expression recognition system using real-time YouTube-based datasets. Appl Intell 48(9):2912–2929
51. Xu C, Du PF, Feng ZY, Meng ZP, Cao TY, Dong CC (2013) Multimodal emotion recognition fusing video and audio. Appl Math Inform Sci 7(2):455–462
52. Wang Y, Yang X, Zou J (2013) Research of emotion recognition based on speech and facial expression. Institute of Advanced Engineering & Science 11(1):83–90
53. Wang SF, He S, Wu Y, He MH, Ji Q (2014) Fusion of visible and thermal images for facial expression recognition. Front Comput Sci-Chi 8(2):232–242
54. Majumder A, Behera L, Subramanian VK (2018) Automatic facial expression recognition system using deep network-based data fusion. IEEE Trans Cybern 48(1):103–114
55. Wang WC, Chang FL, Liu YL, Wu XJ (2017) Expression recognition method based on evidence theory and local texture. Multimed Tools Appl 76(5):7365–7379
56. Sun YC, Yu J (2017) Facial expression recognition by fusing Gabor and local binary pattern features. In: International Conference on Multimedia Modeling, MMM, vol 10133. Springer, Cham, pp 209–220
57. Sun B, Li LD, Zhou GY, He J (2016) Facial expression recognition in the wild based on multimodal texture features. J Electron Imaging 25(6):061407
58. Zavaschi THH, Britto AS, Oliveira LES, Koerich AL (2013) Fusion of feature sets and classifiers for facial expression recognition. Expert Syst Appl 40(2):646–655
59. Li D, Wen G (2017) MRMR-based ensemble pruning for facial expression recognition. Multimed Tools Appl 10:1–22
60. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, Berlin
61. Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20(8):832–844
62. Goodfellow LJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang YC, Thaler D, Lee DH (2015)

Challenges in representation learning: a report on three machine learning contests. Neural Netw 64:59–63
63. Li S, Deng W, Junping D (2017) Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild, CVPR
64. Kuncheva LI (2013) A bound on kappa-error diagrams for analysis of classifier ensembles. IEEE Trans Knowl Data Eng 25(3):494–501
65. Kunchava LI, Whitaker CJ (2003) Measures of diversity in classifier ensemble and their relationship with the ensemble accuracy. Mach Learn 51(2):181–207
66. Li N, Yu Y, Zhou ZH (2012) Diversity regularized ensemble pruning. In: Machine Learning and Knowledge Discovery in Databases, Proceedings of the European Conference (ECML PKDD 2012). Springer Verlag, Bristol, pp 330–345
67. Dai Q, Han XM (2016) An efficient ordering-based ensemble pruning algorithm via dynamic programming. Appl Intell 44(4):816–830
68. Oleg O Giorgio V (2009) Applications of supervised and unsupervised ensemble methods [M]. Springer Berlin Heidelberg

Huihui Li received the M.S degree in South China University of Technology. She is currently working towards the Ph. D degree from the Department of Computer Science and Technology of South China University of Technology. Her research area includes facial expression recognition, artificial intelligence and Machine Learning in Traditional Chinese Medicine.

Open Access | Published: 13 February 2019

Sample awareness-based personalized facial expression recognition

Huihui Li & Guihua Wen

Applied Intelligence **49**, 2956–2969 (2019) | Cite this article

2135 Accesses | **13** Citations | **3** Altmetric | Metrics

Abstract

The behavior of the current emotion classification model to recognize all test samples using the same method contradicts the cognition of human beings in the real world, who dynamically change the methods they use based on current test samples. To address this contradiction, this study proposes an individualized emotion recognition method based on context awareness. For a given test sample, a classifier that was deemed the most suitable for the current test sample was first selected from a set of candidate classifiers and then used to realize the individualized emotion recognition. The Bayesian learning method was applied to select the optimal classifier and then evaluate each candidate classifier from the global perspective to guarantee the optimality of each candidate classifier. The results of the study

Download PDF

Sections Figures References

Abstract
Introduction
Related works
Proposed method
Experimental results
Conclusion
References
Acknowledgments
Author information
Additional information
Rights and permissions
About this article

Applied Intelligence (2019) 49:2956–2969
https://doi.org/10.1007/s10489-019-01427-2

Sample awareness-based personalized facial expression recognition

Huihui Li[1] · Guihua Wen[1]

Published online: 13 February 2019

Abstract
The behavior of the current emotion classification model to recognize all test samples using the same method contradicts the cognition of human beings in the real world, who dynamically change the methods they use based on current test samples. To address this contradiction, this study proposes an individualized emotion recognition method based on context awareness. For a given test sample, a classifier that was deemed the most suitable for the current test sample was first selected from a set of candidate classifiers and then used to realize the individualized emotion recognition. The Bayesian learning method was applied to select the optimal classifier and then evaluate each candidate classifier from the global perspective to guarantee the optimality of each candidate classifier. The results of the study validated the effectiveness of the proposed method.

Keywords Facial expression recognition · Personalized classification · Dynamic selection · Bayesian

1 Introduction

Widely applied in mental health and human–computer interaction, emotion recognition is currently a popular research topic in the fields of computer vision and artificial intelligence [1–3] because it involves multiple disciplines, such as image processing, pattern recognition, and psychology. However, the diversity of facial expressions makes the emotion recognition difficult. For example, the collected facial images might be unidentifiable because of the lighting environment [4]. Moreover, the facial expressions of human beings are complicated and diverse, with fairly significant individual differences dynamically based on the current test samples, instead of identifying all test samples with the same method. For example, human thinking follows the principle of simplicity (the Gestalt principle) [7]. Simple object recognition only needs simple methods, while complex object recognition needs complex methods [8]. However, most of the existing machine learning methods only consider the complexity of the whole dataset [9], or the complexity of the local neighborhood [10], without distinguishing the complexity of the object to be identified. In addition, for the same test sample, each person's emotional recognition ability is different, which is also true for classifiers. As the ensemble classifier emphasizes, the base classifiers

Taylor & Francis
Taylor & Francis Group

International Journal of Computer Integrated Manufacturing

ISSN: 0951-192X (Print) 1362-3052 (Online) Journal homepage: http://www.tandfonline.com/loi/tcim20

Transfer learning for aluminium extrusion electricity consumption anomaly detection via deep neural networks

Peng Liang , Hai-Dong Yang, Wen-Si Chen, Si-Yuan Xiao & Zhao-Ze Lan

To cite this article: Peng Liang , Hai-Dong Yang, Wen-Si Chen, Si-Yuan Xiao & Zhao-Ze Lan (2017): Transfer learning for aluminium extrusion electricity consumption anomaly detection via deep neural networks, International Journal of Computer Integrated Manufacturing, DOI: 10.1080/0951192X.2017.1363410

To link to this article: http://dx.doi.org/10.1080/0951192X.2017.1363410

第46卷 第6期
Vol. 46 No. 6

计 算 机 工 程
Computer Engineering

2020年6月
June 2020

·图形图像处理·

文章编号：1000－3428(2020)06－0274－07 文献标志码：A 中图分类号：TP391

基于 Mask R-CNN 的人脸检测与分割方法

林凯瀚，赵慧民，吕巨建，詹 瑾，刘晓勇，陈荣军

（广东技术师范大学 计算机科学学院，广州 510665）

摘 要：针对现有主流的人脸检测算法不具备像素级分割，从而存在人脸特征具有噪声及检测精度不理想的问题，提出了一种基于 Mask R-CNN 的人脸检测及分割方法。通过 ResNet-101 结合 RPN 网络生成候选区域，再利用 RoIAlign 算法实现像素级的特征点定位，旨在提高定位精度。根据全卷积网络生成相应的人脸二值掩码，实现图像中人脸信息与背景的分割。此外，构建了一个具有分割标注信息的人脸数据集用于训练相应模型。在通用人脸检测数据集的实验结果表明，该方法具有较好的人脸检测效果，并能在准确检测的同时实现像素级的人脸信息分割。

关键词：人脸检测；Mask R-CNN 算法；实例分割；RoIAlign 算法；全卷积网络

开放科学（资源服务）标志码（OSID）：

中文引用格式：林凯瀚，赵慧民，吕巨建，等. 基于 Mask R-CNN 的人脸检测与分割方法[J]. 计算机工程，2020，46(6)：274-280.

英文引用格式：LIN Kaihan，ZHAO Huimin，LÜ Jujian，et al. Face detection and segmentation method based on Mask R-CNN[J]. Computer Engineering，2020，46(6)：274-280.

Face Detection and Segmentation Method Based on Mask R－CNN

LIN Kaihan，ZHAO Huimin，LÜ Jujian，ZHAN Jin，LIU Xiaoyong，CHEN Rongjun

（School of Computer Science，Guangdong Polytechnic Normal University，Guangzhou 510665，China）

【Abstract】 Face detection is an important research direction in computer vision and information security，which has been widely studied over the past few decades. In the traditional face detection method，there is no pixel-level segmentation process，which leads to the problem of face features with noise and unsatisfactory detection accuracy. In order to overcome this shortcoming，a face detection and segmentation method based on Mask R-CNN is proposed in this paper. In this method，ResNet-101 and RPN is used to generate RoIs，and RoIAlign faithfully retains the exact spatial locations to generate binary mask through Fully Convolution Network. In order to train the model，this paper constructs a face dataset with

文内图片

计算机工程，2020,46(06) 北大核心 CSCD

记笔记

基于Mask R-CNN的人脸检测与分割方法

林凯瀚 赵慧民 吕巨建 詹瑾 刘晓勇 陈荣军
广东技术师范大学计算机科学学院

摘要： 针对现有主流的人脸检测算法不具备像素级分割,从而存在人脸特征具有噪声及检测精度不理想的问题,提出了一种基于Mask R-CNN的人脸检测及分割方法。通过ResNet-101结合RPN网络生成候选区域,再利用RoIAlign算法实现像素级的特征点定位,旨在提高定位精度。根据全卷积网络生成相应的人脸二值掩码,实现图像中人脸信息与背景的分割。此外,构建了一个具有分割标注信息的人脸数据集用于训练相应模型。在通用人脸检测数据集的实验结果表明,该方法具有较好的人脸检测效果,并能在准确检测的同时实现像素级的人脸信息分割。

关键词： 人脸检测; Mask R-CNN算法; 实例分割; RoIAlign算法; 全卷积网络;

基金资助： 国家自然科学基金（61772144）；广东省教育厅创新团队项目（2017KCXTD021）；广东省自然科学基金-博士启动基金（2016A030310335）；广东省普通高校青年创新人才类项目（2018KQNCX139）；广州市对外科技合作计划项目（201807010059）；

DOI： 10.19678/j.issn.1000-3428.0054566

专辑： 信息科技

专题： 计算机软件及计算机应用; 自动化技术

分类号： TP183;TP391.41

引证文献 7
被引频次

[1] 许德刚（H指数：9）;王露;李凡;
[2] 赵建敏（H指数：9）;文博;李琦;
[3] 张超 文传博 .
[4] 季晨颖 赵鸣博 李嘲 .
[5] 林凯瀚 .
[6] 孟科 .
[7] 何娃 .

·图形图像处理· 文章编号：1000－3428(2020)06－0274－07 文献标志码：A 中图分类号：TP391

基于 Mask R-CNN 的人脸检测与分割方法

林凯瀚，赵慧民，吕巨建，詹 瑾，刘晓勇，陈荣军

（广东技术师范大学 计算机科学学院，广州 510665）

摘 要：针对现有主流的人脸检测算法不具备像素级分割，从而存在人脸特征具有噪声及检测精度不理想的问题，提出了一种基于 Mask R-CNN 的人脸检测及分割方法。通过 ResNet-101 结合 RPN 网络生成候选区域，再利用 RoIAlign 算法实现像素级的特征点定位，旨在提高定位精度。根据全卷积网络生成相应的人脸二值掩码，实现图像中人脸信息与背景的分割。此外，构建了一个具有分割标注信息的人脸数据集用于训练相应模型。在通用人脸检测数据集的实验结果表明，该方法具有较好的人脸检测效果，并能在准确检测的同时实现像素级的人脸信息分割。

关键词：人脸检测；Mask R-CNN 算法；实例分割；RoIAlign 算法；全卷积网络

开放科学（资源服务）标志码（OSID）：

中文引用格式：林凯瀚，赵慧民，吕巨建，等. 基于 Mask R-CNN 的人脸检测与分割方法[J]. 计算机工程，2020，46(6)：274-280.

英文引用格式：LIN Kaihan，ZHAO Huimin，LÜ Jujian，et al. Face detection and segmentation method based on Mask R-CNN [J]. Computer Engineering，2020，46(6)：274-280.

Face Detection and Segmentation Method Based on Mask R－CNN

LIN Kaihan，ZHAO Huimin，LÜ Jujian，ZHAN Jin，LIU Xiaoyong，CHEN Rongjun

（School of Computer Science，Guangdong Polytechnic Normal University，Guangzhou 510665，China）

【Abstract】 Face detection is an important research direction in computer vision and information security，which has been widely studied over the past few decades. In the traditional face detection method，there is no pixel-level segmentation process，which leads to the problem of face features with noise and unsatisfactory detection accuracy. In order to overcome this shortcoming，a face detection and segmentation method based on Mask R-CNN is proposed in this paper. In this method，ResNet-101 and RPN is used to generate RoIs，and RoIAlign faithfully retains the exact spatial locations to generate binary mask through Fully Convolution Network. In order to train the model，this paper constructs a face dataset with segmentation annotation information. The experimental results of well-known face detection dataset show that the proposed method has better face detection effect and can achieve pixel-level face information segmentation at the same time.

【Key words】 face detection；Mask R-CNN algorithm；instance segmentation；RoIAlign algorithm；Full Convolutional Network（FCN）

DOI：10.19678/j.issn.1000－3428.0054566

0 概述

人脸检测是计算机视觉和信息安全领域的一个重要研究方向，也是目标检测技术的关键分支，具有重要的研究意义及应用价值。人脸检测一般包括人脸的识别和定位两个过程，通过利用图像处理以及机器学习等技术从图像或视频中检测定位出人脸，从而获取人脸信息。近年来，人脸检测在模式识别领域得到了广泛的研究与发展，现有的人脸检测研究主要包括基于传统机器学习的方法和基于深度学习的方法。

在传统机器学习方法中，主要是通过手工特征结合分类器实现检测过程。Viola-Jones[1] 算法是最早实时有效人脸检测的算法，标志着人脸检测开始进入实用阶段。但是，该方法存在特征维度较大、对复杂的背景情况识别率较低等问题。为解决以上问

基金项目：国家自然科学基金（61772144）；广东省教育厅创新团队项目（2017KCXTD021）；广东省自然科学基金-博士启动基金（2016A030310335）；广东省普通高校青年创新人才类项目（2018KQNCX139）；广州市对外科技合作计划项目（201807010059）。

作者简介：林凯瀚（1994—），男，硕士研究生，主研方向为计算机视觉、图像处理；赵慧民，教授、博士；吕巨建，讲师、博士；詹 瑾，副教授、博士；刘晓勇，教授、博士；陈荣军，副教授、博士。

收稿日期：2019-04-10 修回日期：2019-06-17 **E-mail**：jujianlv@gpnu.edu.cn

题，研究人员提出了更为复杂的手工设计特征，如HOG[2]、SIFT[3]、SURF[4]、LBP[5]等。其中一个重要的进展是文献[6]提出的DPM（Deformable Part Model）[6]，DPM在HOG和SVM的基础上进行性能扩展，充分利用两者的优势，在人脸识别、行人检测等任务上取得重要突破。然而传统的机器学习方法仍然存在两个主要缺陷，即基于滑动窗口的区域选择策略针对性较弱，设计较为复杂，手工设计特征对复杂情景检测稳定性较差。

自从AlexNet[7]在ImageNet中使用深度卷积神经网络，并且在图像分类的准确率上有了大幅提高以来，人们尝试将深度学习应用于人脸检测领域，取得了很好的应用效果。文献[8]利用深度卷积神经网络进行特征提取，对没有被Adaboost过滤过的人脸和非人脸的图像提取出更有价值的特征，取得了较好的实现效果。文献[9]提出一种基于深度卷积神经网络提取的归一化特征的形变部件模型。近年来，随着基于候选区域（region-based）的R-CNN系列目标检测算法的快速发展，R-CNN系列算法在人脸检测领域的应用研究逐渐兴起。文献[10]在WIDER数据集训练了Faster R-CNN模型，在FDDB和IJB-A数据集实现了人脸检测。文献[11]针对小尺度人脸检测精度不高的问题，在Faster R-CNN的基础上提出了DSFD算法。文献[12]通过特征拼接、困难样本挖掘、多尺度训练等策略改进了Faster R-CNN模型，提高了模型性能，取得较好的检测效果。基于深度学习的人脸检测方法主要采用卷积神经网络进行特征提取，在准确性和多目标检测方面具有良好的实现效果，并能够以较少的时间花费换取大幅的准确率提升，因此，基于深度学习的人脸检测算法已成为人脸检测的主流研究方向。

上述的研究工作都取得较好的人脸检测效果，但是缺乏对人脸图像分割的关注。如需在人脸图像上获取更精准的人脸信息，则现有的方法存在提取的人脸目标特征维度大、空间量化较为粗糙、具有背景噪声等问题，导致一些实用的图像处理技术（如多姿态人脸矫正[13]、人脸图像的超分辨率重构[14]、遮挡人脸姿态识别[15]等）在安防监控所抓拍的视频图像上难以应用。因此，需要一种适用于图像的人脸检测及分割方法。在检测分割方面，Mask R-CNN是文献[16]基于Faster R-CNN提出的一种改进算法，它增加了对实例分割的关注。除了分类和边界框回归之外，Mask R-CNN还为每个RoI添加并行分支以进行实例分割，并对上述3种损失进行联合训练，在目标检测数据集中取得了良好的效果。

在以上研究工作的基础上，本文提出了一种基于Mask R-CNN的人脸检测及分割方法，该方法在现有人脸检测仅实现人脸定位的基础上增加了分割分支，能够在人脸检测的同时实现像素级的人脸信息分割。同时，为了训练相应模型，本文从FDDB[17]和ChokePoint[18]数据集中随机选择了5 115张图像进行标注，构建了一个具有分割标注的新数据集。

1 相关工作

基于候选区域的CNN检测方法是目标检测领域的主流研究方向，例如R-CNN[20]、Fast R-CNN[21]、Faster R-CNN[22]以及Mask R-CNN[16]等，这些方法不仅检测精度高，且具有较快的检测速度。

R-CNN算法由Ross Girshick等人于2014年提出，该算法在VOC 2007数据集中取得了66%（mAP）的成绩，掀起了基于候选区域（region-based）CNN的研究热潮。首先由于R-CNN特征提取环节存在重复计算的问题，导致检测速度偏慢，因此Fast R-CNN基于R-CNN，在卷积层对待检测图像整图进行一次特征提取；其次引入RoI Pooling层进行特征尺度的统一化，生成固定长度的特征向量；然后用softmax（归一化指数函数）代替了SVM，并将分类和边框回归进行了合并，在提升准确率的同时，提高了检测速度。Faster R-CNN是在Fast R-CNN基础上提出的改进算法，该算法引入了区域建议网络（Region Proposal Network，RPN）模型，通过在特征图上划窗，使用不同尺寸、不同长宽比的9种锚点框映射到原图上，得到候选区域。该算法将候选框提取合并到深度网络中，大幅提升了检测的速度与精度。

为了能够在目标检测的同时，实现像素级的实例分割，满足计算机视觉任务中更为精准的检测定位需求，文献[16]提出了Mask R-CNN算法。该算法在Faster R-CNN的基础上，添加实例分割分支，通过RoIAlign获得更精准的像素信息，利用全卷积网络生成相应的二值掩码，实现了目标检测及实例分割技术的结合应用，在目标检测领域的COCO数据集和自动驾驶领域的Cityscapes数据集上的实验结果表明，该算法均取得了较好的效果。

2 人脸检测及分割方法

Mask R-CNN 算法主要由两大分支组成，即检测分支和分割分支。检测分支实现了对图像中目标的定位及分类，分割分支则通过全卷积网络（Fully Convolutional Network，FCN）[23] 生成二值掩码实现实例分割，达到像素级的区分效果。

本文模型的整体框架如图 1 所示，首先将待检测图像传入该模型，通过卷积神经网络对整张图像进行特征的提取，得到对应的特征图，利用区域建议网络（RPN）在特征图上迅速生成候选区域，再通过候选区域匹配（RoIAlign），得到固定尺寸的特征图输出，然后在分类分支做出目标框的定位及分类，在分割分支通过全卷积网络对人脸图像绘出相应的二值掩码实现实例分割，最后输出系统预测的图像。

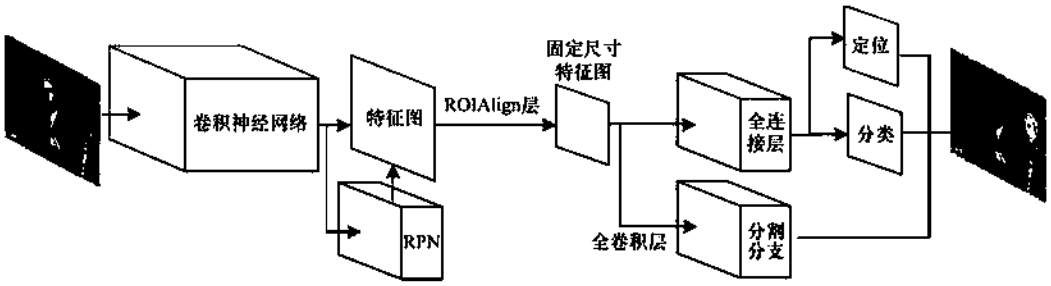


图 1 本文算法模型框架

Fig. 1 The model framework of the proposed algorithm

2.1 RPN 网络

区域建议网络（RPN）通过倍数和长宽比例不同的窗口或锚点（Anchors）在特征图上进行滑窗从而迅速生成候选区域。RPN 算法示意图如图 2 所示，图中背景图像表示经卷积神经网络提取特征后的特征图，虚线表示窗口为基准窗口，假设基准窗口大小为 16 个像素点，其包含 3 个分别表示长宽比例为 1∶1、1∶2、2∶1 的窗口，则点划线及实线分别表示 8 像素和 32 像素点大小的窗口，同理，其各有 3 个长宽比例为 1∶1、1∶2、2∶1 的窗口。RPN 利用上述 3 种倍数和 3 种比例的共 9 种尺度窗口的方法对特征图进行滑窗，当 IoU ≥ 0.5 时，认为其为正例，并对其进行回归。

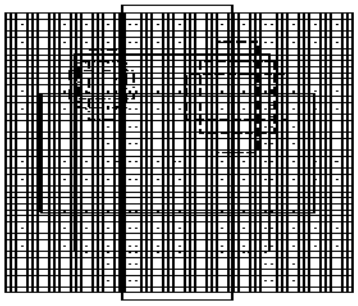


图 2 RPN 算法示意图

Fig. 2 Schematic diagram of RPN algorithm

交并比（Intersection over Union，IoU）计算公式如下：

$$\mathrm{IoU}=\frac{S_{A\cap B}}{S_{A\cup B}} \tag{1}$$

其中，A、B 分别为 RPN 网络生成的候选框及训练集中正确的目标框，$S_{A\cap B}$ 为 A、B 的相重叠处面积，$S_{A\cup B}$ 为 A、B 并集面积。

2.2 RoIAlign 层

RoIAlign 层实现对产生的候选区域（Region of Interest，RoI）进行池化，从而将不同尺度的特征图通过 RoIAlign 层池化为固定的尺度特征图的效果，RoIAlign 算法流程如图 3 所示。在 Faster R-CNN 模型中候选区域池化（Region of Interest Pooling，RoI Pooling）存在两次取整操作，从而产生了量化误差，导致了图像的像素点定位精准度较低。

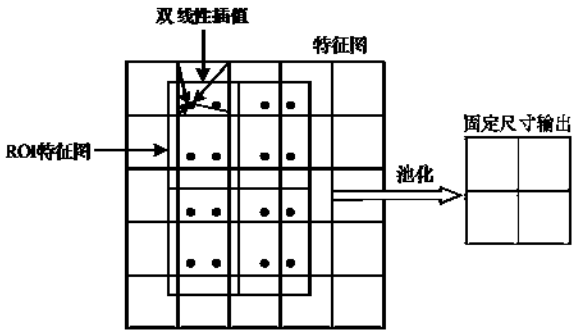


图 3 RoIAlign 算法流程

Fig. 3 RoIAlign algorithm procedure

在面对小目标检测及实例分割的任务时，RoI Pooling 达不到精准的特征点定位要求。针对上述问题，RoIAlign 在生成的 RoI 特征图上取消量化过程，采用双线性插值算法，保留浮点型坐标，避免了量化误差，可以使原图像像素与特征图像素相匹配。双线性插值算法公式如下：

1）对 x 方向进行线性插值：

$$f(R_1)\approx\frac{x_2-x}{x_2-x_1}f(Q_{11})+\frac{x-x_1}{x_2-x_1}f(Q_{21}),R_1=(x,y_1) \tag{2}$$

$$f(R_2)\approx\frac{x_2-x}{x_2-x_1}f(Q_{12})+\frac{x-x_1}{x_2-x_1}f(Q_{22}),R_2=(x,y_2) \tag{3}$$

2) 对 y 方向进行线性插值：

$$f(P)=f(x,y)\approx\frac{y_2-y}{y_2-y_1}f(R_1)+\frac{y-y_1}{y_2-y_1}f(R_2) \tag{4}$$

如图 4 所示，$f(x,y)$ 为待求解点 P 的像素值，$f(Q_{11})$、$f(Q_{12})$、$f(Q_{21})$、$f(Q_{22})$ 分别为已知四点 $Q_{11}=(x_1,y_1)$、$Q_{12}=(x_1,y_2)$、$Q_{21}=(x_2,y_1)$ 及 $Q_{22}=(x_2,y_2)$ 的像素值，$f(R_1)$、$f(R_2)$ 为 x 方向插值得到的像素值。

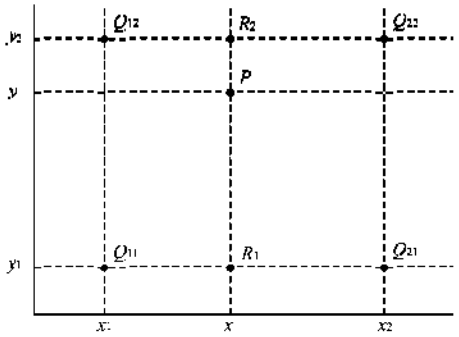


图 4　双线性插值算法流程

Fig. 4　Procedure of bilinear interpolation algorithm

2.3　全卷积网络

对于传统的 CNN 网络架构，为了得到固定维度的特征向量，其卷积层后大多连接若干全连接层，最后输出为针对输入的一个数值描述，因而普遍适用于图像的识别分类、目标的检测定位等任务。全卷积网络（FCN）与传统的 CNN 网络相似，网络同样包含了卷积层和池化层，不同之处在于全卷积网络对末端一个卷积层的特征图使用了反卷积进行上采样，使输出图像尺寸恢复到原图像尺寸，然后通过 Softmax 分类器进行逐像素预测，从而预测出每个像素点的所属类别。

2.4　损失函数定义

本文的 Mask R-CNN 模型完成了 3 个任务，即人脸框的检测定位、人脸与背景的分类、人脸与背景的分割。因此，损失函数的定义包括了定位损失、分类损失及分割损失三部分。因此，损失函数定义如下：

$$L=L_{\rm cls}^*+L_{\rm box}^*+L_{\rm mask}^* \tag{5}$$

其中，$L_{\rm cls}^*$ 为分类损失，$L_{\rm box}^*$ 为定位损失，$L_{\rm mask}^*$ 为分割损失。

分类损失公式如下：

$$L_{\rm cls}^*=L(\{p_i\})=\frac{1}{N_{\rm cls}}\sum_i L_{\rm cls}(p_i,p_i^*) \tag{6}$$

其中，i 对应第 i 个锚点，$N_{\rm cls}$ 为分类样本数，p_i 为锚点预测为目标的概率值，p_i^* 为标记的标签值，$p_i^*=1$ 为正例，$p_i^*=0$ 为负例，$L_{\rm cls}$ 为两个类别的对数损失：

$$L_{\rm cls}(p_i,p_i^*)=-\lg\left[p_i^*p_i+(1-p_i^*)(1-p_i)\right] \tag{7}$$

定位损失公式如下：

$$L_{\rm box}^*=L(\{t_i\})=\frac{1}{N_{\rm reg}}\sum_i p_i^*L_{\rm reg}(t_i,t_i^*) \tag{8}$$

其中，i 对应第 i 个锚点，$N_{\rm reg}$ 为回归的样本锚点个数，t_i 为预测框位置参数，t_i^* 为标记的真实框位置参数，$L_{\rm reg}$ 为定位的回归损失，$p_i^*L_{\rm reg}$ 则表示当锚点检测为正例时，才计算其回归损失，其公式为：

$$L_{\rm reg}(t_i,t_i^*)=\text{smooth}_{L1}(t_i-t_i^*) \tag{9}$$

分割损失即二值掩码的损失，如果候选框检测出为某一类别，则使用该类别的交叉熵作为误差值进行计算，其他类别损失值不计入，从而避免了类间的竞争，其公式为：

$$L_{\rm mask}^*=-\frac{1}{m^2}\sum_{1\le i,j\le m}\left[y_{ij}\lg\hat{y}_{ij}^k+(1-y_{ij})\lg(1-\hat{y}_{ij}^k)\right] \tag{10}$$

其中，y_{ij} 为 $m\times m$ 区域坐标点 (i,j) 的标签值，$\hat{y}_{ij}^k$ 为该点第 k 类的预测值。

3　实验与结果分析

3.1　实验环境及数据集构建

实验环境及数据集构建主要有以下 3 个方面：

1) 实验环境配置

本文实验所使用的环境配置如下：操作系统为 Ubuntu 16.04；CPU 为 Xeon E5-2690 v4 @ 2.20 GHz；GPU 为 NVIDIA GeForce GTX 1080Ti；深度学习框架为 Tensorflow 1.9.0；编程语言为 Python 3.6。

2) 数据集

由于公开的人脸数据集中大多数只有目标框的定位信息，不具备分割标注信息，因此本文从 FDDB[17] 原始图像数据集和 ChokePoint[18] 数据集中随机选择了 5 115 张图片并对其进行分割标注作为新数据集。FDDB (Face Detection Data Set and Benchmark) 数据集是马萨诸塞大学公开的人脸检测数据集，该数据集由原始图像数据集和相应的部分标签数据组成，标签数据包含了 2 845 张图像（其中包含了 5 171 张人脸）。ChokePoint 数据集是由澳大利亚信息与通信技术研究中心（NICTA）公开的人脸数据集，该数据集通过在不同场景出入口处安装不同角度监控摄像头，从而采集到的人脸数据集，该数据集包括 48 个视频序列，共计约 64 204 张人脸图像。

3) 参数设置

本文采用 ResNet-101 网络作为特征提取器，首先使用 COCO 预训练模型[24]进行参数的泛化，使模型具有一定特征提取能力，从而减少训练时间。然后利用已标注的数据集进行模型训练，参考文献[16]并通过实验调试，设置迭代次数为 50，迭代步数为 3 000，学习率为 0.001，权值衰减率为0.000 1。

3.2 结果分析

3.2.1 实验效果与性能分析

在训练出的模型上利用现有的公开数据集进行测试实验，测试图像检测结果如图 5 所示。可以看出，相较于其他人脸检测算法，本文算法在完成人脸目标框定位的同时，通过彩色的二值掩码将人脸信息与背景分割开来，实现了对人脸图像检测定位和分割的效果。

图 5 测试图像检测结果

Fig. 5 Test image detection results

为有效检测模型的性能，对比实验采用本文所构建数据集作为训练集，再进行模型性能的评测。实验在 FDDB 数据集上对比了本文算法与 Faster R-CNN[10]、Pico[25]、Viola-Jones[26]、koestinger[27]等算法，性能曲线如图 6 所示。

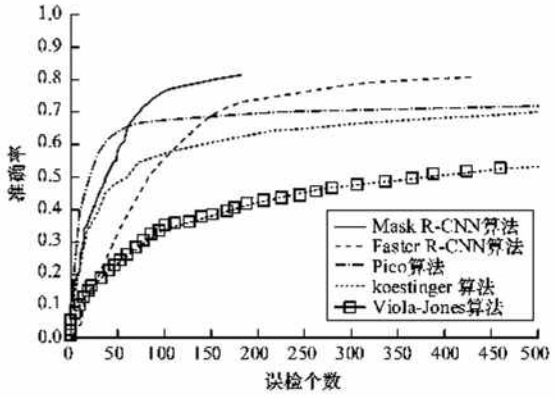


图 6 FDDB 数据集检测性能曲线

Fig. 6 FDDB datasets detection performance curve

从图 6 可以看出，本文算法误检个数明显减少，且准确率高于其他算法。

3.2.2 不同数据集的有效性分析

为了验证模型在不同数据集的有效性，本文在 AFW[28]数据集上进行了模型性能测试实验，实验对比了文献[29]中部分算法，实验结果如图 7 所示。虽然受限于训练数据集数据规模相对较小，以及标注规则存在差异，但本文模型仍取得了 94.08% AP 值（平均精度）的成绩，优于大部分对比算法，对比表现最优的 SquaresChnFtrs-5AP 值仅低了 1.16，证明了本文算法的有效性。

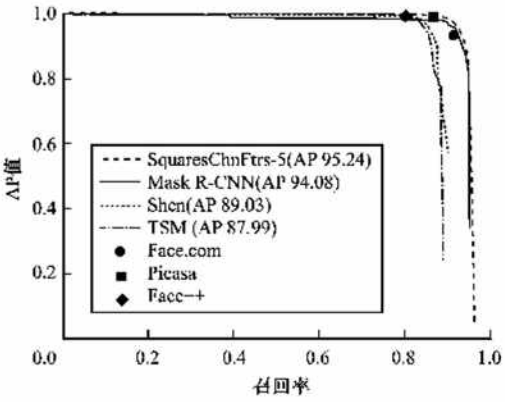


图 7 AFW 数据集检测性能曲线

Fig. 7 AFW datasets detection performance curve

此外，为进一步验证不同数据集的有效性，针对本文模型及在 FDDB 数据集性能表现较好的 Faster R-CNN 模型，本文分别在 ChokePoin 数据集、Caltech 10k Web Faces 数据集[30]、WIDER[31]数据集进行了测试，实验分别从以上 3 个数据集随机抽取 100 张图像进行测试，通过计算其 mAP（均值平均精度）值进行比较，结果如表 1 所示，其中加黑字体为最优成绩。

表 1 不同数据集性能对比

Table 1 Performance comparison of different dataset %

方法	Caltech	ChokePoint	WIDER
Faster R-CNN 方法	90.46	89.43	90.80
本文方法	**98.28**	**98.39**	**97.91**

通过不同数据集性能对比实验结果可以看出，本文方法比性能表现较好 Faster R-CNN 在 Caltech 10k Web Faces 数据集、ChokePoint 数据集、WIDER 数据集分别提高了 7.82%、8.96%、7.11%，验证了本方法在不同数据集上的有效性。

3.2.3 检测时间分析

时间测试实验同样在 ChokePoint 数据集、Caltech 10k Web Faces 数据集、WIDER 数据集进行

验证，从以上数据集分别随机抽取 100 张图像进行测试，比较其平均耗时，结果如表 2 所示，其中加黑字体为最少用时。

表 2　不同数据集检测时间对比

Table 2　Comparison of detection time of different datasets　　s

方法	Caltech	ChokePoint	WIDER
Faster R-CNN 方法	**0.320**	**0.285**	**0.308**
本文方法	0.347	0.331	0.320

与 Faster R-CNN 算法对比，由于本文算法的 RoIAlign 层计算复杂度较高以及模型添加了分割分支，因此在 Caltech 10k Web Faces 数据集、ChokePoint 数据集、WIDER 数据集实验中，检测时间分别多了 0.027 s、0.046 s 和 0.012 s。但本文算法能够以较少的时间花费换取更高的准确率，并能实现在人脸准确定位的同时将人脸信息从背景中分割出来，达到满意的检测与分割效果。

4　结束语

本文提出一种基于 Mask R-CNN 的人脸检测及分割方法，构建一个具有分割标注信息的人脸数据集，并在建立的数据集上训练模型。该模型利用 RoIAlign 算法使得人脸图像特征点定位精准度达到像素级的效果，结合 ResNet-101 网络与 RPN 网络，提高了检测精度，并通过全卷积网络生成相应的人脸二值掩码，实现了人脸信息与背景图像的分割。在公开数据集上的对比实验结果表明，本文方法能够在不显著增加计算和模型复杂度的情况下，取得较好的检测效果。下一步将对数据集进行扩充，并在现有研究的基础上对模型进行优化，进一步提高模型检测性能。

参考文献

[1] VIOLA P, JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.

[2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 886-893.

[3] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[4] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features [J]. Computer Vision and Image Understaning, 2008, 110(3): 346-359.

[5] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: application to face recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 2037-2041.

[6] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Oject detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Mchine Intelligence, 2010, 32(9): 1627-1645.

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//Proceedings of IEEE NIPS' 12. Washington D. C., USA: IEEE Press, 2012: 1097-1105.

[8] ZHAN Shu, TAO Qinqin, LI Xiaohong. Face detection using representation learning [J]. Neurocomputing, 2016, 187: 19-26.

[9] RANJAN R, PATEL V M, CHELLAPPA R. A deep pyramid deformable part model for face detection [C]// Proceedings of the 7th IEEE International Conference on Biometrics Theory, Applications and Systems. Washington D. C., USA: IEEE Press, 2015: 1-8.

[10] JIANG H, LEARNED-MILLER E. Face detection with the faster R-CNN [C]//Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE Press, 2017: 650-657.

[11] WU Wenqi, YIN Yingjie, WANG Xingang, et al. Face detection with different scales based on faster R-CNN [J]. IEEE Transactions on Cybernetics, 2018, 49(11): 4017-4028.

[12] SUN X, WU P, HOI S C H. Face detection using deep learning: an improved faster rcnn approach [J]. Neurocomputing, 2018, 299: 42-50.

[13] GUO Xiaobo, ZHOU Zhaoyong, LI Songyang. Nasal tip detection and posture correction of 3D face based on effective energy [J]. Computer Engineering, 2018, 44(9): 236-242. (in Chinese)
郭小波，周兆永，李松阳. 基于有效能量的 3D 人脸鼻尖点检测与姿态矫正 [J]. 计算机工程，2018，44(9)：236-242.

[14] ZOU Guofeng, FU Guxia, GAO Mingliang, et al. Pose varied face recognition based on self learning deep convolutional neural network [J]. Journal of Chinese Computer Systems, 2018, 39(6): 1156-1162. (in Chinese)
邹国锋，傅桂霞，高明亮，等. 基于自学习深度卷积神经网络的姿态变化人脸识别 [J]. 小型微型计算机系统，2018，39(6)：1156-1162.

[15] QIU Yiming, LIAO Haibin, CHEN Qinghu. Occluded face pose recognition based on dictionary learning with discrimination performance [J]. Geomatics and Information Science of Wuhan University, 2018, 43(2): 275-281, 288. (in Chinese)

邱益鸣,廖海斌,陈庆虎.基于鉴别字典学习的遮挡人脸姿态识别[J].武汉大学学报(信息科学版),2018,43(2):275-281,288.

[16] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 2961-2969.

[17] JAIN V, LEARNED-MILLER E. Fddb: a benchmark for face detection in unconstrained settings [EB/OL]. [2019-03-04]. https://www.researchgate.net/publication/.

[18] WONG Y, CHEN S, MAU S, et al. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2011: 74-81.

[19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.

[20] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 580-587.

[21] GIRSHICK R. Fast R-CNN [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 1440-1448.

[22] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of IEEE NIPS' 15. Washington D. C., USA: IEEE Press, 2015: 91-99.

[23] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 3431-3440.

[24] LIN T Y, MAIRE M, BELONGUE S, et al. Microsoft coco: common objects in context [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 740-755.

[25] MARKUS N, FRLJAK M, PANDZIC I S, et al. A method for object detection based on pixel intensity comparisons organized in decision trees [EB/OL]. [2019-03-10]. https://www.researchgate.net/publication/.

[26] JENSEN O H. Implementing the Viola-Jones face detection algorithm [D]. Lyngby, Denmark: Technical University of Denmark, 2008.

[27] KOESTINGER M, WOHLHART P, ROTH P M, et al. Robust face detection by simple means [C]// Proceedings of IEEE DAGM' 12. Washington D. C., USA: IEEE Press, 2012: 125-137.

[28] RAMANAN D, ZHU X. Face detection, pose estimation, and landmark localization in the wild [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 2879-2886.

[29] MATHIAS M, BENENSON R, PEDERSOLI M, et al. Face detection without bells and whistles [C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 720-735.

[30] FINK M, FERGUS R, ANGELOVA A. Caltech 10 000 Web faces [EB/OL]. [2019-03-10]. http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces.

[31] YANG S, LUO P, LOY C C, et al. Wider face: a face detection benchmark [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 5525-5533.

编辑 索书志

基于深度学习的视觉目标跟踪算法浅析

黄智慧，詹 瑾，赵慧民*，吕巨建，郑鹏根，林凯瀚
（广东技术师范大学 计算机科学学院，广东 广州 510665）

摘 要：视觉目标跟踪是计算机视觉研究领域的一个基础性问题，随着近年来人工智能的迅速兴起，目标跟踪技术的研究得到越来越多的关注．深度学习技术具有强大的特征表征能力，在图像分类、物体识别、自然语言处理等应用上比传统方法取得了更好的效果，因此逐渐成为图像视频研究的主流技术．简要概述传统目标跟踪算法及其存在的问题，然后从深度模型的角度，对近几年基于深度学习的视觉目标跟踪算法进行分类，重点介绍、分析和对比主要的深度学习跟踪框架及其优缺点，并使用OTB2015数据集测试各算法的性能，最后总结深度学习跟踪方法现有的挑战，并对进一步研究提出展望．

关键词：视觉跟踪；深度学习；特征表示；计算机视觉
中图分类号：TP391 **文献标识码**：A **文章编号**：1672-402X(2019)03-0028-09
DOI:10.13408/j.cnki.gjsxb.2019.03.006

0 引 言

视觉目标跟踪是计算机视觉领域一个重要基础应用研究，主要决解在视频序列中对目标进行自动检测、跟踪和识别，精确地估计目标的位置、尺寸和运动轨迹等问题，应用非常广泛，如自动驾驶、智能监控系统、智能交通系统以及人机交互等．但是由于视频场景存在复杂的非线性变化，目标自身状态也具有很大的不确定性，例如目标被严重遮挡、姿态尺度变化、运动模糊、光照变化、背景杂乱等情形．因此，要实现到广泛应用．2006年，加拿大多伦多大学教授Geoffrey Hinton[2]提出了深度学习，打破了传统神经网络研究瓶颈，利用深度网络对图像或文本等数据进行解释分析．深度学习体系结构可以通过学习非线性网络结构，表征输入数据，实现复杂函数逼近，在机器学习研究中已取得巨大的成功．鉴于其强大的特征表观性能，研究者们开始将深度学习引入视觉跟踪中，文献[3，4]报道了高效的实验结果．近几年，深度学习在视觉目标跟踪领域的发展速度越来越快，但国

广东技术师范学院学报，2019,40(03)

基于深度学习的视觉目标跟踪算法浅析

黄智慧 詹瑾 赵慧民 吕巨建 郑鹏根 林凯瀚

广东技术师范大学计算机科学学院

摘要： 视觉目标跟踪是计算机视觉研究领域的一个基础性问题,随着近年来人工智能的迅速兴起,目标跟踪技术的研究得到越来越多的关注.深度学习技术具有强大的特征表征能力,在图像分类、物体识别、自然语言处理等应用上比传统方法取得了更好的效果,因此逐渐成为图像视频研究的主流技术.简要概述传统目标跟踪算法及其存在的问题,然后从深度模型的角度,对近几年基于深度学习的视觉目标跟踪算法进行分类,重点介绍、分析和对比主要的深度学习跟踪框架及其优缺点,并使用OTB2015数据集测试各算法的性能,最后总结深度学习跟踪方法现有的挑战,并对进一步研究提出展望.

关键词： 视觉跟踪; 深度学习; 特征表示; 计算机视觉;

基金资助： 广州市科创委对外科技合作计划项目（项目编号:201807010059）；广东省教育厅特色创新科研项目（自然科学）（项目编号:2016KTSCX077）；广东省普通高校青年创新人才类项目（2018KQNCX139）；广东省协同创新平台环境建设国际合作领域（自然科学）（2017A050501039）；

DOI： 10.13408/j.cnki.gjsxb.2019.03.006

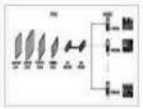
专辑： 社会科学II辑; 信息科技

专题： 计算机软件及计算机应用; 自动化技术

分类号： TP391.41;TP183

文章目录

文内图片



基于深度学习的视觉目标跟踪算法浅析

黄智慧，詹 瑾，赵慧民*，吕巨建，郑鹏根，林凯瀚
（广东技术师范大学 计算机科学学院，广东 广州 510665）

摘 要：视觉目标跟踪是计算机视觉研究领域的一个基础性问题，随着近年来人工智能的迅速兴起，目标跟踪技术的研究得到越来越多的关注。深度学习技术具有强大的特征表征能力，在图像分类、物体识别、自然语言处理等应用上比传统方法取得了更好的效果，因此逐渐成为图像视频研究的主流技术。简要概述传统目标跟踪算法及其存在的问题，然后从深度模型的角度，对近几年基于深度学习的视觉目标跟踪算法进行分类，重点介绍、分析和对比主要的深度学习跟踪框架及其优缺点，并使用OTB2015数据集测试各算法的性能，最后总结深度学习跟踪方法现有的挑战，并对进一步研究提出展望。

关键词：视觉跟踪；深度学习；特征表示；计算机视觉
中图分类号：TP391 **文献标识码：**A **文章编号：**1672-402X(2019)03-0028-09
DOI:10.13408/j.cnki.gjsxb.2019.03.006

0 引 言

视觉目标跟踪是计算机视觉领域一个重要基础应用研究，主要决解在视频序列中对目标进行自动检测、跟踪和识别，精确地估计目标的位置、尺寸和运动轨迹等问题，应用非常广泛，如自动驾驶、智能监控系统、智能交通系统以及人机交互等。但是由于视频场景存在复杂的非线性变化，目标自身状态也具有很大的不确定性，例如目标被严重遮挡、姿态尺度变化、运动模糊、光照变化、背景杂乱等情形，因此，要实现对目标准确和实时的鲁棒跟踪是一项具有挑战性的难题。传统视觉目标跟踪方法主要从目标外观表示、目标特征、目标位置搜索和定位、纠正漂移以及更新策略等方面进行了研究，往往根据特定的应用需求和限定的场景对视觉跟踪问题进行约束和简化，从而取得在鲁棒性、准确性和实时性三方面性能的一个折中研究结果。

深度学习是机器学习研究领域的一个新的分支，高性能和高度灵活性决定其在目标检测、图像分类、语音识别和自然语言处理等方面得到广泛应用。2006年，加拿大多伦多大学教授Geoffrey Hinton[2]提出了深度学习，打破了传统神经网络研究瓶颈，利用深度网络对图像或文本等数据进行解释分析。深度学习体系结构可以通过学习非线性网络结构，表征输入数据，实现复杂函数逼近，在机器学习研究中已取得巨大的成功。鉴于其强大的特征表现性能，研究者们开始将深度学习引入视觉跟踪中，文献[3,4]报道了高效的实验结果。近几年，深度学习在视觉目标跟踪领域的发展速度越来越快，但国内外该类方向的综述文献相对较少。

本文旨在介绍和分析深度学习在目标跟踪领域近年来的发展动态，简要地对传统目标跟踪算法进行分类论述，总结了传统算法存在的局限性。然后阐述了视觉目标跟踪算法引入深度学习后的发展现状，从网络结构上归类详述了基于深度学习的各代表性跟踪算法的原理，分析其在速度、精度和鲁棒性等方向上做出的改进，并指出了对应的不足和优点。接着在OTB2015数据集上对目前较流行的方法进行性能评

收稿日期：2019-04-21
基金项目：广州市科创委对外科技合作计划项目（项目编号：201807010059）；广东省教育厅特色创新科研项目（自然科学）（项目编号：2016KTSCX077）；广东省普通高校青年创新人才类项目（2018KQNCX139）；广东省协同创新平台环境建设国际合作领域（自然科学）（2017A050501039）。
作者简介：黄智慧，广东技术师范大学2018级硕士研究生。
*** 通讯作者：**赵慧民，广东技术师范大学教授，zhaohuimin@gpnu.edu.cn

估．最后，总结全文，同时展望深度学习在目标跟踪应用的进一步发展．

1　传统跟踪算法

传统的视频跟踪方法通常由目标外观模型[5,6]、运动模型[7,8]和搜索策略[9]三部分构成，外观模型对目标在特定位置的可能性状态进行估计，运动模型旨在描述跟踪对象随时间变化的各状态，而搜索策略是在当前帧搜索目标可能性最高的位置．从目标搜索策略的角度考虑，现有的传统目标跟踪算法大致可分为两类：生成式方法和判别式方法．

1.1　生成式方法

生成式跟踪算法通过最大化相似性或最小化重构误差来描述目标的外观特征，在目标外观模型确定的情况下，通过特征相似性搜索方法，在后续帧的候选对象中找到可能性最高的目标．文献[10]基于增量的主成分分析法（PCA）进行模型更新，学习低维子空间的性能表示，进而确定有效适应目标的外观变化，并使用粒子滤波法替代传统的梯度下降法估计模型的运动参数．文献[11]基于稀疏表示方法，将跟踪问题转变成模板子空间的系数相似问题，提出通过加强非负性限制条件进行稀疏表示，并对稀疏表示的系数调整权重完成模型动态更新．文献[5]提出结合稀疏表示法和PCA方法学习目标外观模型，用在线学习更新的稀疏原型表示跟踪对象．生成式跟踪算法的准确性取决于目标特征表示，忽略了图像的背景信息，并且对目标外观变化、光照变化、遮挡等情况的鲁棒性较差．

1.2　判别式方法

判别式跟踪算法将跟踪过程视为目标与背景二值分类问题．与生成式模型不同，判别式模型将背景信息考虑在内，因此它在应对目标的强遮挡及外观变化时，具有更高的鲁棒性．代表性的判别式跟踪方法有多示例学习法（MIL）[12]、支持向量机（SVM）[13,14]、在线AdaBoost（OAB）算法[15]和结构化输出法[16]．文献[12]提出了使用多示例学习法训练目标外观分类器，可处理训练数据产生歧义的问题，进而增强跟踪过程对目标被部分遮挡和外观变化的鲁棒性．文献[14]将跟踪过程当作在线半监督分类问题，在协同训练框架上同时进行新数据分类和分类器更新工作，半监督学习法根据目标各个特征的融合置信图产生新样本，并用这些新样本更新在线SVM，以完成跟踪过程．许多应用深度学习的跟踪算法最终都可归结为判别式法．

尽管上述方法在改进视觉跟踪性能方面取得较好的效果，但由于采用的是浅层特征提取方法，因此在复杂环境中（例如运动模糊、变形、旋转和光照等）对目标的表征能力表现得较弱．而具有强大的特征学习能力的深度学习可提取跟踪目标的深度特征，为解决视觉跟踪中存在的各种挑战带来了新的研究方向．

2　深度学习在目标跟踪上的应用

深度学习的出现，将机器学习推向一个更高的热潮，其最大的优势是经过大量数据的训练可以得到有效的语义特征．特征的有效性决定了跟踪系统的性能高低，优化过程体现在单特征到多特征融合．深度学习应用到跟踪领域面临两个难题：首先，仅利用视频第一帧获得的先验知识，远未能满足深度模型训练对标签数据量的要求，针对这个问题，研究者使用图像分类的大型数据集（例如ImageNet）预训练跟踪模型[18][19][20]，但是这种数据集与视频跟踪所需的数据具体存在很大的差异，会降低跟踪效果；其次，是深度网络的多层结构导致计算的复杂度增加，大量的计算会降低实际跟踪过程的实时性．

表1　基于深度学习的视觉目标跟踪算法的性能对比

算法名称	采用的深度网络	优点	缺点
DLT[21]	SAE	首次引用深度学习，与传统算法对比，提高速度和精准度	训练图像分辨率低、普适性差
文献[23]	SAE	融合多层分类器，增强性能	目标形变、全遮挡和快速运动鲁棒性差
SO-DLT[3]	CNN	最早使用CNN进行特征提取	在干扰和目标形状不规则下，易发生跟踪漂移
MDNet[4]	CNN	类间分类效果好	受相似干扰项影响大
SANet[25]	CNN+RNN	目标类似干扰项识别能力强	跟踪速度慢
VITAL[28]	CNN+GAN	增强正样本，解决分类失衡	逐帧更新，耗时大
SINT[72]	Siamese	不需要模型在线更新，速度快	遮挡和旋转情形鲁棒性差
SINT++[38]	Siamese+VAE+DQN	提高训练样本多样性和样本量	模型复杂，计算量大

即便如此，近十年，研究者不断优化深度模型与跟踪算法的结合方式，获得了比传统算法更好的准确性和实时性实验结果. 目前，跟踪算法引用的深度学习框架主要有3种，分别是堆栈自编码器(Stacked Autoencoders, SAE)，卷积神经网络(CNN)和孪生网络(Siamese Network). 表1对本文讨论的基于深度学习的视觉目标跟踪算法性能进行了对比.

2.1 基于 SAE 的跟踪算法

2013年，Wang等人提出的DLT方法[21]首次将深度模型成功应用到视觉跟踪领域中，并首次提出离线预训练与在线微调相结合的思路. 首先在离线的状态下使用栈式降噪自编码(Stacked Denoising Autoencoders,SDAE)方法学习图像通用特征表示，完成无监督预训练阶段，利用SDAE的编码部分构造判别神经网络，并使用该神经网络训练所需的特征提取器和分类器. 跟踪过程结合粒子滤波法，选出置信度最高的为最终预测目标. 在更新策略上，采用阈值判断是否更新整个网络. DLT具有高跟踪准确度和低计算复杂度的优点，但是DLT本身也存在一些不足. 首先，用于离线训练的数据图像分辨率低，很难提取到足够强的表观特征；其次，由于自编码获得的训练图像属于重构图像，且数据集的不同视频序列包含各类物体对象，会对实际跟踪过程的分类造成干扰和误判.

文献[23]提出结合深度学习技术和在线AdaBoost框架的跟踪算法，继承DLT的SDAE网络结构，并用于学习多层图像表观特征. 在包含4个隐含层的SDAE网络中，利用在线AdaBoosting框架形成一个高精准率和低错误率的强分类器. 该算法得到显著的实验成果，包括：第一，实现深度学习网络在不同层能够自动学习目标的有效通用特征；第二，Boosting框架可在多个特征层中自动识别最适用于外观建模的特征层.

2.2 基于 CNN 的跟踪算法

CNN网络结构能在比例缩放、旋转、平移等非线性变化中学习更加丰富的不变性，这种强大的特征提取能力正符合目标跟踪的要求. 因此，自2015年CNN模型被首次用于视觉跟踪并取得引人注目的效果后[3]，相继出现许多对CNN网络进行深入研究的目标跟踪算法，且在成功率和精准度的单程评估上取得不断提升的实验效果.

2.2.1 单网络应用

文献[4]提出一个基于CNN多域学习框架MDNet，模型由共享层和特定域多分支层构成. 训练过程从共享层获取目标的一般特性，比如遮挡、形变等，在线学习跟踪过程提取特定域的信息. MDNet将训练中的每一个视频序列当成一个域，每一个域对应一个二分类器，用于寻找目标和背景的决策边界. 测试时，共享层和新判别层结合构成新网络，通过初始帧在线微调识别新的视频序列，并通过long-term和short-term完成更新. 为避免出现跟踪漂移，文中借助目标检测中的硬负样本挖掘法[24]获取对分类器训练有效的负样本. MDNet较为简单，所需模型计算复杂度比图像识别低. 该算法在2016提出并产生了开创性的影响，后续有研究者对基于此跟踪算法的网络结构进行不断改进.

MDNet网络如图1所示，五个隐含层(三层卷积层和两层全连接层组成)的共享层和K个分支的特定域层组成MDNet的网络结构，其中K个分支对应K个域，每个域拥有一个二分类层，在每一个域中，黄色表示正样本，紫色表示负样本.

由于采用在线微调的结构，MDNet方法速度较慢，实时性和精度上有待提升. 文献[25,26,27,28]在MDNet基础上提出了多个改进算法，本文利用OTB100数据集对这些算法进行了测试，相应的实验结果如表2所示. 文献[26]提出简单而有效的正则化法(BranchOut)优化在线集成跟踪算法，基于CNN的多分支结构使用多层表示进行目标外观建模. 对比验证使用BranchOut表示MDNet集成的一般特性，BranchOut-MDNet比原来MDNet法具有显著优化效果. 文献[27]继承MDNet结构的嵌入自适应RoIAlign层提取特定特征，有效地区分跨多域并具有类似语义的对象实例，在保持算法MDNet精度下，速率提升了25倍，具有较好的实时性.

表 2　基于 MDNet 改进的目标跟踪算法

Trackers	改进点	测试数据集	Succ(%)	Prec(%)
MDNet		OTB100	67.7	90.9
SANet[25]	提高对类内干扰项的识别度	OTB100	69.2	92.8
BranchOut-MDNet	通过适当的有效正则化以及多层级目标表示，增强学习目标多样化外观鲁棒能力	OTB100	68.3	91.9
RT-MDNet[27]	在与MDNet几乎保持相同准确度下，提高实时性	OTB100	65.0	88.5
VITAL[28]	产生高质量正样本，提高分类网络训练强度	OTB100	68.2	91.7

SO-DLT[3]延续了非跟踪数据预训练和在线微调相结合的策略，提出使用CNN网络模型替代DLT的栈式降噪编码模型，但改变了跟踪候选区域的大小，利用了图片本身的结构化信息输出像素级概率图；还采用了短期(short-term)和长期(long-term)两个CNN组成更新模型，用集成的思路解决模型更新存在的敏感性问题；通过多尺度框寻找目标策略确定当前帧目标边界盒的位置和大小.方法在CVPR2013提出的OTB50数据集上OPE(One-pass evaluation)准确度和成功率远超当时其它方法.但是SO-DLT离线预训练依然使用的是ImageNet数据集，使用时序关联数据是一个更好的选择.图2和图3分别是SO-DLT网络结构图和在线跟踪过程示意图.

2.2.2　多网络应用

CNN网络对不同类对象(目标与背景)判别较强，但是对于类内的干扰不具备很好的处理

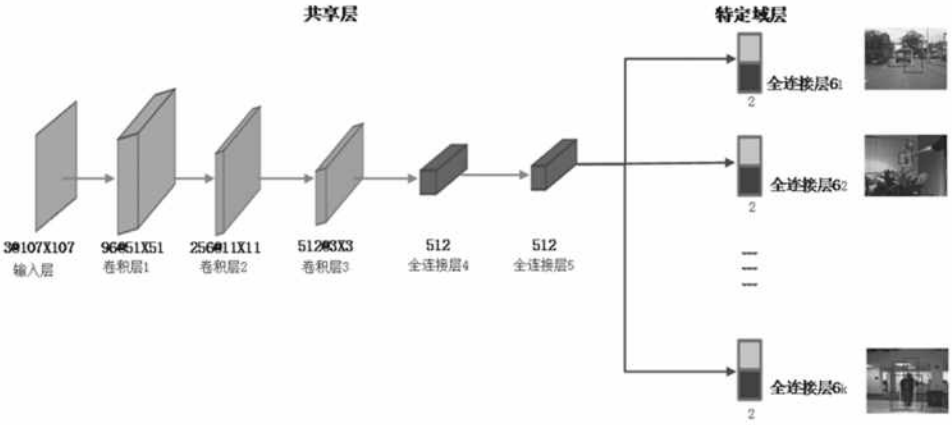


图 1　MDNet 网络结构示意图

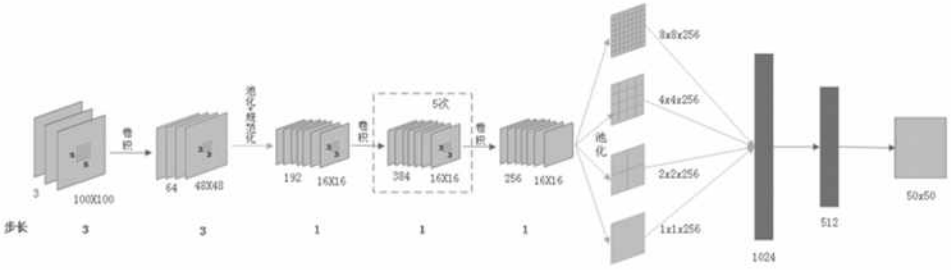


图 2　SO-DLT 网络结构示意图

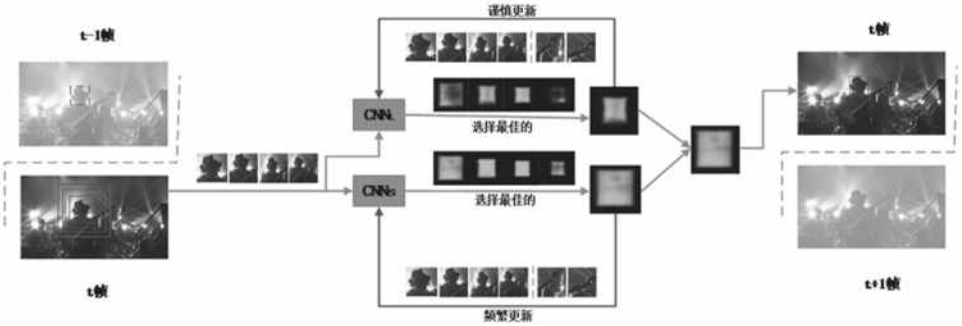


图 3　SO-DLT 在线跟踪过程示意图

能力，因此，在相似对象的干扰情况下，很容易发生漂移. 文献[25]利用目标自身结构信息区分与目标相似的干扰项，提出使用RNNs[29]网络对跟踪对象自身结构进行建模，从而提高识别相似干扰对象的鲁棒性. 为了提供丰富的训练信息，SANet[25]采用跃级式策略组成多网络结构，借鉴MDNet的方法，文中结合长短期网络完成在线更新，以及通过硬负样本挖掘法利用有效负样本训练分类器.

文献[28]针对多个正样本在空域中容易出现重叠现象，在CNN结构嵌入生成对抗网络(GAN)[30]生成高质量的正样本，另外对于正负样本分类失衡的问题，文中采用高阶成本敏感损失函数降低简易负样本对分类模型训练的影响. 在传统分类模型(VGG-M)的最后一层卷积层和第一层全连接层之间嵌入一个生成对抗网络，输入特征与Mask经过dropout操作产生时间鲁棒特征，再利用这些时间鲁棒特征训练分类器. 与MDNet的硬负样本挖掘法相比，VITAL[28]使用高阶成本损失函数获取有用负样本，不仅提高目标跟踪精度，且加快了训练到最优状态的速度. OTB100的测试结果显示，虽然在成功率评估上ECO[31]取得最优，但是在跟踪精度上VITAL超过ECO和MDNet.

2.3 基于 Siamese 的跟踪算法

Siamese网络是一种相似度量法，从数据中学习一个相似性度量，并用这个相似度量学习或匹配新的对象. Siamese网络适合数据类别多而每一种类别的样本量少的情况. 训练数据的不足一直是深度学习应用到目标跟踪中有待解决的问题，这造成深度网络的性能优化存在很大限制. 2016年，文献[32,33]首次将Siamese网络，引入目标跟踪领域取得很高的关注度，近几年涌现一系列的相关工作[34,35,36,37,38].

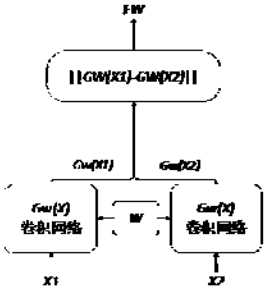


图 4 Siamese 网络结构图

图4是朴素Siamese网络基本模型，其中X_1和X_2是一对输入样本，有可能是同类或是不同类. $G_w(X_1)$和$G_w(X_2)$分别是样本X_1和X_2经过网络映射得到的低维输出结果，E_w是两个输出结果的能量表示函数，也就是相似度量，当X_1和X_2属于同一个类别时，E_w值较小，否则，其值增大.

文献[32]借助匹配函数的强大性能，通过学习外部的视频数据获得适合在线跟踪的通用匹配函数. 匹配函数一旦完成学习，就无需再调整. 不经过目标更新[39]、跟踪算法结合[40]及遮挡检测[41]等处理，即可将第一帧初始目标与新帧的候选图像进行匹配，相似度最高确定为预测目标. 并且，该算法可重新识别完全消失后的目标，减少漂移情况出现. 该算法实现相对简单，在很大程度上降低了跟踪过程时间消耗，但对目标类似干扰项的辨别能力差，容易发生定位错误现象，具有很大的改进空间.

文献[38]整体结构由三个主要模块组成，分别是正样本生成网络(PSGN)、硬正样本生成网络(HPTN)和Siamese实例搜索网络(SINT). 其中PSGN是在图像的流行结构基础上，使用变分自编码器(Variational Auto-Encoder, VAE)遍历出具更加丰富的正样本，这些解码而得的图像与原始目标图像相似但具有一定程度的形变、运动模糊和更高的多样性. HPTN网络训练输出硬正样本的过程也称为马尔可夫决策问题，并通过深度强化学习优化出更符合实际跟踪中发生的遮挡情形，以此增强跟踪的鲁棒性. 最后使用Siamese深度网络结合PSGN和HPTN两个部分组成完整的SINT++算法[38]. 虽然该算法计算复杂度较高，但在稳定性和与实际情况符合度上都获得比A-Fast-RCNN[43]更好的效果.

图5是SINT++的训练过程示意图，由正样本生成网络(PSGN)获得丰富正样本，再通过硬正样本(HPTN)的强化训练组成最优的遮挡目标图像，该硬正样本输入Siamese网络，进一步计算得损失结果，并给HPTN模块完成反馈.

从深度学习在目标跟踪的发展历程可见，跟踪精度随网络加深而不断提高，目前算法优化主要方向为鲁棒性和实时性. 深度模型对目标特征的学习具有很大依赖性，而标签数据的缺

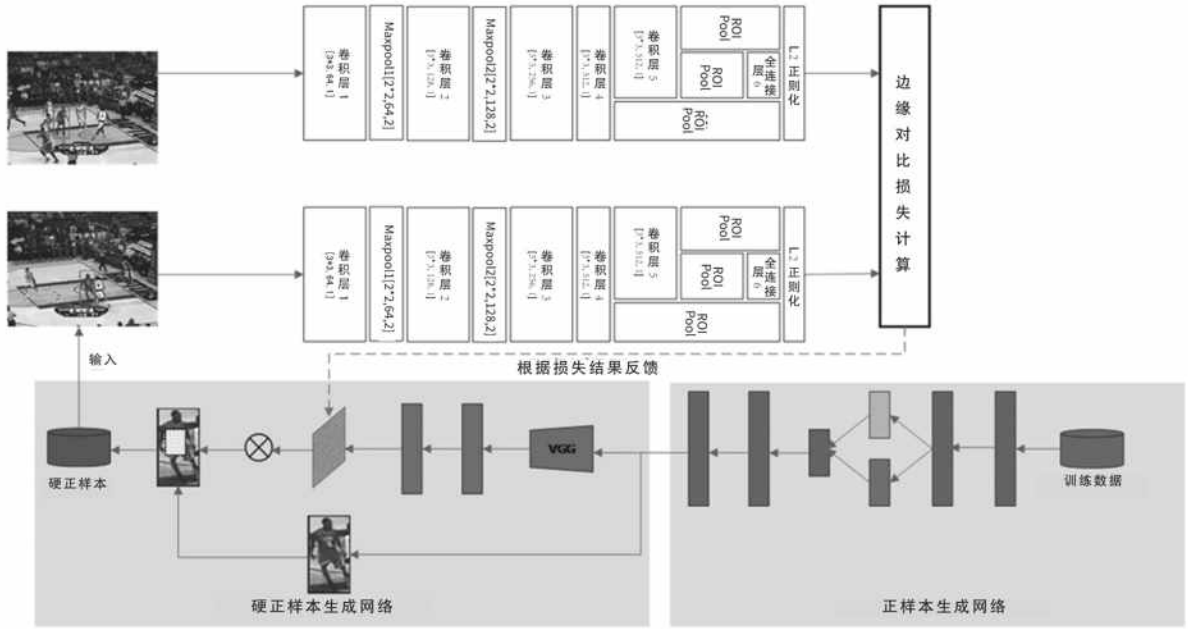


图 5　SINT++ 训练过程示意图

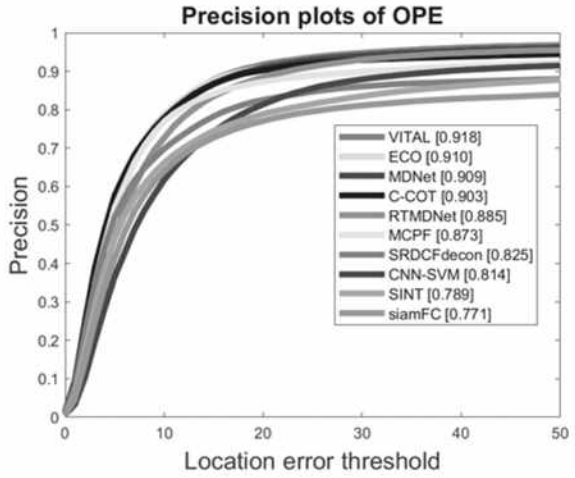
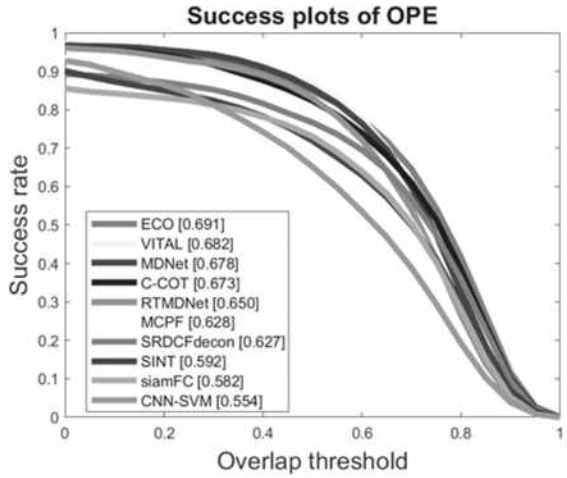


图 6　(a) 平均精度曲线



(b) 平均成功率曲线

乏和训练数据的低普适度，使得要充分发挥深度学习的优势依旧存在挑战性. 另外，大部分深度学习跟踪算法采用离线训练和在线微调相结合的策略，实时性低体现在测试阶段的更新过程，这是一项十分耗时的工作. 不同的更新策略影响跟踪结果精度的同时也会改变跟踪速率，因此模型更新是跟踪过程的重要部分.

图 6(a) 和 6(b) 给出 10 种视觉目标跟踪算法在 OTB100[44] 上的精准度和成功率. 其中，图 6(a) 表示基于中心位置误差平均精度图，阈值设为 20 像素，图 6(b) 表示边界框重叠率的平均成功率曲线图. 根据对比结果可知，在精度上，结合深度学习的视觉目标跟踪算法比相关滤波类算法取得更好的结果. 其中，采用 CNN 网络的 VITAL 和 MDNet 算法相较于基于 Siamese 类的方法表现出更高的成功率和精准度. 由此可见，目前基于深度学习的目标跟踪算法虽面临各种挑战，但仍存在很大发展潜能，值得深入探索和讨论.

3　总　结

本文主要分析了近年来深度学习在目标跟踪领域的发展动态，从深度网络结构角度对基于深度学习的视觉跟踪方法进行了归类，详细

阐述了各代表性跟踪方法的工作原理；在OTB 2015数据集上对目前较流行的方法进行速度、精度和鲁棒性等性能评估，分析其不足和优点。深度学习在视觉跟踪领域具有高效性能和巨大发展优势，取得了突破性的跟踪效果，但不同深度结构模型具有不同的优缺点。例如，CNN类方法跟踪的准确度不断提高，但速度一直是个瓶颈，而孪生网络(Siamese)突破了深度视觉跟踪算法在速度上的滞后，但其跟踪精度成为另一短板。因此，结合不同深度网络的优点，实现跟踪精度和实时性的同时提升，是未来视觉跟踪方法的主要研究内容。此外，通过不同深度网络生成高质量样本，解决视频跟踪问题的小样本学习问题，提高分类器在复杂背景下识别类内和类间干扰项的能力，这也是未来值得关注的研究内容。

参考文献：

[1]Hebb D O. The organization of behavior[M]. Neurocomputing: foundations of research. 1988.

[2]Hinto G E,OsinderoS,The Y W.A Fast Learning Algorithm For Deep Belief Nets[J].Neural Computation,2006,18(7):1527-1554.

[3]WangN,LiS,Gupta A. Transferring Rich Feature Hierarchies for Robust Visual Tracking[J].2015, arXiv:1501.04587.

[4]NamH,Han B. Learning Multi-Domain Convolutional Neural Networks for Visual Tracking [J].2016, arXiv: 1510.07945.

[5]D.Wang, H. Lu, and M. Yang.Online Object Tracking with Sparse Prototypes[J].IEEE Transactions on Image Processing, 2013, 22(1): 314-325.

[6]Babenko B, Yang M H,BelongieS.Robust Object Tracking with Online Multiple Instance Learning [J]. IEEE Transactions on Pattern Analysis &.Machine Intelligence,2010,33(8):1619-1632.

[7]ComaniciuD,RameshV,MeerP.Kernel-based Object Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2003,25(5): 564 - 575.

[8]Li Y, Ai H, Yamashita T, Lao S, and KawadeM. Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Life Spans[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008,30(10):1728 1740.

[9]GrabnerH , BischofH.On-line Boosting and Vision[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision &.Pattern Recognition. 2006: 260-267.

[10]Ross D A,Lim J et al.Incremental Learning for Robust Visual Tracking[J].International Jornal of Computer Vision,2008,77(1-3):125-141.

[11]MeiX,LingH.RobustVisual Tracking Using L1 Minimization[C]//Proceedings of the IEEE International Conference on Computer Vision,United States: IEEE,2009:1436-1443.

[12]Babenko B, Yang M, and Belongie S.Robust Object Tracking with Online Multiple Instance Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,33(8):1619 - 1632.

[13]AvidanS.SupportVector Tracking[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2004, 26(8):1064 - 1072.

[14]Tang F, Brennan S, Zhao Q, and TaoH.Co-tracking using Semi-supervised Support Vector Machines[C]// Proceedings of IEEE 11thInternational Conference on Computer Vision,Brazil:IEEE,2007:1 - 8.

[15]Grabner H, Grabner M, and Bischof H. Real-time Tracking Via On-line Boosting[C]//Proceedings ofBMVC 2006,2006: 47 - 56.

[16]Hare S, Saffari A, and Torr P H. Struck: Structured Output Tracking with Kernels[C]//Proceedings of ICCV 2011,2011:263 270.

[17]Wang L, Ouyang W, Wang X, et al. Visual Tracking with FullyConvolutional Networks [C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Chile: IEEE, 2016:3119-3127.

[18]DanelljanM,Robinson A, Khan F S, and Felsberg M.Beyond Correlation Filters: Learning Continuous Convolution[C]//Proceedings of ECCV 2016,2016: 472-488.

[19]Hong S, You T, Kwak S, and Han B. Online Tracking by Learning Discriminative Saliency Map with Convolutional Neural Network[J]. arXiv:1502.06796,2015.

[20]MaC,Huang J -B, Yang X, and Yang M -H.Hierarchical Convolutional Features for Visual Tracking[C]// Proceedings of IEEE International Conference on Computer Vision(ICCV),Chile:IEEE,2015: 3074-3082.

[21]Wang N, Yeung D -Y.Learning a Deep Compact Image Representation for Visual Tracking[C]//

Proceedings of NIPS, USA,2013: 809-817.

[22]Fan J, Xu W, Wu Y, et al. Human Tracking Using Convolutional Neural Networks [J]. IEEE Transactions on Neural Networks, 2010,21(10): 1610-1623.

[23]ZhouX,XieL,ZhangP,et al. An Ensemble of Deep Neural Networks for Object Tracking [C] //Proceedings of the IEEE International Conference on Image Processing(ICIP), France:IEEE,2014:843-847.

[24]Sung K -K,Poggio T. Example-based Learning for ViewbasedHuman Face Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence,1998,20(1):39 - 51.

[25]Fan H,Ling H.SANet: Structure-Aware Network for Visual Tracking[J].arXiv:1611.06878,2017.

[26]Ilan B,SimJ,Adam H. BranchOut: Regularization for Online Ensemble Tracking with Convolutional Neural Networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),USA:IEEE,2017: 3356-3365.

[27]Jung I,SonJ,BackM,and Han B. Real-Time MDNet[J].arXiv:1808.08834,2018:1-16.

[28]Song Y, Ma C, Wu X,et al.VITAL: VIsual Tracking via Adversarial Learning[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. USA:IEEE,2018: 8990-8999.

[29]Elman J. L. Finding Structure in Time[J]. Cognitive science, 1990,14(2):179 - 211.

[30]Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, CourvilleA,andBengio Y. Generative Adversarial Nets[J]. arXiv:1406.2661,2014:1-9.

[31]Danelljan M,Bhat G,Khan F S, and Felsberg M. Eco:Efficient Convolution Operators for Tracking[J]. 2017, arXiv: 1611.09224.

[32]Tao R, Gavves E,et al. Siamese Instance Search for Tracking [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE,2016: 850-865.

[33]Bertinetto L, Valmadre J, et al. Fully-Convolutional Siamese Networks for Object Tracking[C]// Proceedings of European Conference on Computer Vision(ECCV). Netherlands,2016.

[34]Guo Q,FengW,ZhouC,et al. Learning Dynamic Siamese Network for Visual Object Tracking[C]// Proceedings of IEEE International Conference on Computer Vision (ICCV). Italy:IEEE,2017: 1763-1771.

[35]HeA,LuoC,TianX,Zeng W. A Twofold Siamese Network for Real-Time Object Tracking[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. USA:IEEE,2018: 4834-4843.

[36]LiB,YanJ,et al. High Performance Visual Tracking with Siamese Region Proposal Network[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. USA:IEEE,2018: 8971-8980.

[37]Zhang Y,WangL,QiJ,et al. Structured Siamese Network for Real-Time Visual Tracking[C]// Proceedings of European Conference on Computer Vision(ECCV). Germany,2018: 351-366.

[38]Wang X,LiC,LuoB,Tang J. SINT++: Robust Visual Tracking via Adversarial Positive Instance Generation[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.USA:IEEE,2018: 4864-4873.

[39]Hare S, Saffari A, and Torr P H,et al. Struck: Structured Output Tracking with Kernels[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016, 38(10): 2096 - 2109.

[40]Zhang J, Ma S, and Sclaroff S. Meem: Robust Tracking via Multiple Experts using Entropy Minimization[J]. European Conference on Computer Vision, 2014: 188-203.

[41]Pernici F, Bimbo A D. Object Tracking by Oversampling Local Features. Pernici F, Del Bimbo A. Object Tracking by Oversampling Local Features[J]. IEEETransactions on Pattern Analysis and Machine Intelligence, 2014, 36(12):2538-2551.

[42]Hong Z, Chen N Z, Wang C, et al. MUlti-Store Tracker (MUSTer): A Cognitive Psychology Inspired Approach to Object Tracking[C]// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015:749-758.

[43]Girshick R. Fast R-CNN[J]. Computer Science, 2015:1440 1448.

[44]Wu Y, Lim J, and Yang M. Object tracking benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1834 - 1848.

[45]贾静平，覃亦华．基于深度学习的视觉跟踪算法研究综述[J]. 计算机科学，2017(S1):29-33.

[46]孟琭，杨旭．目标跟踪算法综述[J]. 自动化学报，2019，DOI:10.16383/j.aas.c180277.

[责任编辑：刘向红]

Analysis of Object Tracking Algorithms Based on Deep Learning Technologies

HUANG Zhi-hui, ZHAN Jin, ZHAO Hui-min, ZHENG Peng-gen, LV Jv-jian,LIN Kai-han

(College of Computer Science, Guangdong Polytechnic Normal University, Guangzhou Guangdong 510665)

Abstract: Visual tracking is one fundamental problem in computer vision research. With the rapid development of artificial intelligence in recent years, more and more attention has been paid to the research of target tracking technology. At the meantime, deep learning has developed rapidly and gradually becomes a mainstream research direction because of its powerful ability of representing features and outstanding performance over traditional algorithm in image classification, object recognition, natural language processing and so on. Firstly, a brief overview of the traditional target tracking algorithm and its existing problems are summarized. Then, from the perspective of the depth model, the visual tracking algorithms based on deep learning in recent years are classified. The framework, advantages and disadvantages of some main depth learning-based tracking algorithms are introduced, analyzed and compared, and OTB 2015 dataset is used to test the performance of each algorithm. Finally, the existing challenges of in-depth learning tracking method are discussed and prospects for further research are given.

Key words: visual tracking; deep learning; features representing; computer vision

面向应用型人才培养的数据库课程混合教学改革实践

司 瑜,唐 瑾

(广东技术师范学院计算机科学学院,广东 广州 510665)

[摘 要] 为适应信息技术产业规模发展的需求及产业发展的国际化,培养多层次、工程应用型人才已成为普通高校的主要任务。以 Oracle 为代表的数据库技术是现代信息化管理系统的重要部分,分析了现阶段高校开展 Oracle 数据库技术课程的教学现状和问题,针对应用型人才培养方案对课程教学改革具体策略进行了探究,围绕案例教学法、项目驱动法、翻转课堂教学模式展开了一系列的应用能力培养改革,并将这些教学法有机融合在一起,进行了课程教学改革的合宜探索。

[关 键 词] 混合教学改革;数据库课程;应用能力;翻转课堂模式

[中图分类号] G642 [文献标志码] A [文章编号] 2096-0603(2018)16-0165-03

一、引言

关系型数据库系统拥有较好的稳定性、安全性和可靠性等诸多优点,其庞大的存储容量和高效的管理能力是现代化信息管理的重要技术基础。随着经济的不断增长、全民信息化水平的提高,银行、电信、工业控制、航空、保险等专业领域需要大量熟悉 Oracle 数据库的技术开发人员。因此如何培养出符合社会需求的数据库管理系统专业人才是目前计算机教育从业者必须要考虑的实际问题。数据库系统课程作为计算机科学与技术专业重要的专业基础课程之一,是管理信息系统开发必须运用的核心技术,在计算机相关专业中占据着不容忽视的地位,能提高学生的信息化开发管理和应用能力。

Oracle 数据库技术是一门理论性和应用性相结合的专业课程,学生在掌握 Oracle 体系架构、存储结构等理论的基础上,要更加重视对 Oracle 软件的实际操作能力的培养,提高数据库综合应用开发能力。该课程所涉及的知识点多而杂,涉及数据库管理员和数据库开发人员两种角色操作,例如数据库的创建、安全管理,需求分析、概念结构设计、逻辑结构设计、物理结构、数据库实施、备份恢复等教学内容,每个知识点下又包括若干个子知识点,学生难以抓住知识的重点和难点,很难将知识系统形象化地整合,从而不能真正地掌握或记忆所学到的内容。当今社会更

(一)教学目标不明确

教师在教学过程中,如果不重视 Oracle 应用实践操作而局限于教材理论讲解,易出现与先导数据库课程重复教学的情况,重复教学浪费课时,在学生相应的知识已经巩固的情况下,会影响学生学习的积极性,导致学生理论与实践脱节。因此,在该课程教学过程中,学生和教师都应明确本次课程的学习目标和教学目标,分析 Oracle 技术课程的知识框架和应用操作。

(二)教学内容与行业脱节

计算机产业发展迅速,Oracle 作为全球市场占有率最高的主流数据库技术,其软件产品每年都在更新换代,而教师在教学过程中通常会延用历年所选教材和软件,这就容易脱离 Oracle 实际使用版本,不能突出课程内容的先进性、实践性、操作性。学生的技术应用水平和教师对 Oracle 技术专业引导有不可分割的关系,教师如果没有掌握最新的专业知识技术,在教学效果方面虽然没有直接的影响,但当学生进入行业实习或就业时,则很容易出现数据库应用技术基础薄弱的问题,间接反映教学效果受限。

(三)教学方法不完善

Oracle 数据库作为一门技术性知识联系紧凑的课程,前后知

现代职业教育．2018,(16)

面向应用型人才培养的数据库课程混合教学改革实践

司瑶　詹瑾

广东技术师范学院计算机科学学院

摘要： 为适应信息技术产业规模发展的需求及产业发展的国际化,培养多层次、工程应用型人才已成为普通高校的主要任务。以Oracle为代表的数据库技术是现代信息化管理系统的重要部分,分析了现阶段高校开展Oracle数据库技术课程的教学现状和问题,针对应用型人才培养方案对课程教学改革具体策略进行了探究,围绕案例教学法、项目驱动法、翻转课堂教学模式展开了一系列的应用能力培养改革,并将这些教学法有机融合在一起,进行了课程教学改革的合宜探索。

关键词： 混合教学改革; 数据库课程; 应用能力; 翻转课堂模式;

来源数据库： 高等教育;

分类号： G434;TP311.13-4

面向应用型人才培养的数据库课程混合教学改革实践

司　瑶，詹　瑾

（广东技术师范学院计算机科学学院，广东　广州 510665）

[摘　　要] 为适应信息技术产业规模发展的需求及产业发展的国际化，培养多层次、工程应用型人才已成为普通高校的主要任务。以 Oracle 为代表的数据库技术是现代信息化管理系统的重要部分，分析了现阶段高校开展 Oracle 数据库技术课程的教学现状和问题，针对应用型人才培养方案对课程教学改革具体策略进行了探究，围绕案例教学法、项目驱动法、翻转课堂教学模式展开了一系列的应用能力培养改革，并将这些教学法有机融合在一起，进行了课程教学改革的合宜探索。

[关　键　词] 混合教学改革；数据库课程；应用能力；翻转课堂模式

[中图分类号] G642　　[文献标志码] A　　[文章编号] 2096-0603(2018)16-0165-03

一、引言

关系型数据库系统拥有较好的稳定性、安全性和可靠性等诸多优点，其庞大的存储容量和高效的管理能力是现代化信息管理的重要技术基础。随着经济的不断增长、全民信息化水平的提高，银行、电信、工业控制、航空、保险等专业领域需要大量熟悉 Oracle 数据库的技术开发人员。因此如何培养出符合社会需求的数据库管理系统专业人才是目前计算机教育从业者必须要考虑的实际问题。数据库系统课程作为计算机科学与技术专业重要的专业基础课程之一，是管理信息系统开发必须运用的核心技术，在计算机相关专业中占据着不容忽视的地位，能提高学生的信息化开发管理和应用能力。

Oracle 数据库技术是一门理论性和应用性相结合的专业课程，学生在掌握 Oracle 体系架构、存储结构等理论的基础上，要更加重视对 Oracle 软件的实际操作能力的培养，提高数据库综合应用开发能力。该课程所涉及的知识点多而杂，涉及数据库管理员和数据库开发人员两种角色操作，例如数据库的创建、安全管理、需求分析、概念结构设计、逻辑结构设计、物理结构、数据库实施、备份恢复等教学内容，每个知识点下又包括若干个子知识点，学生难以抓住知识的重点和难点，很难将知识系统形象化地整合，从而不能真正地掌握或记牢所学到的内容。当今社会更需要应用型人才，应以培养应用型人才为教学目标制订教学计划，结合以上课程教学现状，以培养学生实践应用能力为目标，对课程教学模式、教学方法展开了一系列改革，进行了课程教学改革的探索。

二、课程教学现状和问题

Oracle 数据库技术课程的先导课程是数据库系统原理等专业基础课，学生需要有一定的 SQL 查询语言、关系型数据库理论等知识，Oracle 数据库技术则是对数据库原理知识的拓展与应用层面上的提高。绝大部分学生上课前没有听说过 Oracle 数据库管理系统，也不明白掌握 Oracle 技术对计算机应用开发的重要性。当前该课程在教学过程中主要存在如下问题：

（一）教学目标不明确

教师在教学过程中，如果不重视 Oracle 应用实践操作而局限于教材理论讲解，易出现与先导数据库课程重复教学的情况，重复教学浪费课时，在学生相应的知识已经巩固的情况下，会影响学生学习的积极性，导致学生理论与实践脱节。因此，在该课程教学过程中，学生和教师都应明确本次课程的学习目标和教学目标，分析 Oracle 技术课程的知识框架和应用操作。

（二）教学内容与行业脱节

计算机产业发展迅速，Oracle 作为全球市场占有率最高的主流数据库技术，其软件产品每年都在更新换代，而教师在教学过程中通常会延用历年所选教材和软件，这就容易脱离 Oracle 实际使用版本，不能突出课程内容的先进性、实践性、操作性。学生的技术应用水平和教师对 Oracle 技术专业引导有不可分割的关系，教师如果没有掌握最新的专业知识技术，在教学效果方面虽然没有直接的影响，但当学生进入行业实习或就业时，则很容易出现数据库应用技术基础薄弱的问题，间接反映教学效果受限。

（三）教学方法不完善

Oracle 数据库作为一门技术性知识联系紧凑的课程，前后知识点联系很紧密，传统的教学模式通常采取理论课和实践课分开的授课形式。由于学生个体之间存在个体能动性等差异，每位学生的接受程度不同，学生掌握实践操作时不能统一进度，个别学生在遇到技术较强的问题而又缺乏及时的辅导时，会影响学习的兴趣，打击学习的信心。这种传统的教授法会导致本课程缺乏针对性和可行性，不能有效地达到预期目的，因此应该选取更适合学生发展的教学模式，多采用案例教学法、“讲演练用”（讲解、演示，练习、应用）等多元一体化教学方法。

（四）课程考核不全面

本课程考核若只安排一次期末考试，学生容易产生懈怠心理，不能适时跟进课程阶段内容。但学生的学习时间有限，短时

作者简介：司瑶（1994—），女，硕士，主要研究方向：教育信息技术。

詹瑾（1976—），女，博士，讲师，主要研究方向：图像视频处理、机器学习。

间内涉及的项目太复杂，也不能考察出学生对 Oracle 数据库的应用系统设计、开发与管理的掌握情况，容易出现同学之间相互抄袭的情况，教师无法给予客观公平的评分。本课程只有采用多元化的考核方式，采用课堂评价、阶段性实验设计考核、上机实践考核等方式，教师才能控制教学过程中的不可知因素。

三、数据库课程教学改革实践

（一）明确教学目标和教学思路

在该课程教学过程中，学生和教师都应明确本次课程的学习目标和教学目标。教学目标为学生能够熟练操作 Oracle 数据库管理系统，并能以开发人员的角色进行数据库开发应用。在授课前，教师要了解学生的基础情况，对教学内容进行合理筛选，分析数据库课程的知识框架并分层归类。课程的实验设计要结合考核具体要求，明确知识点，如数据库系统安装与配置、表空间管理与表维护、PL/SQL 编程应用，要以具体的案例为载体，将知识分解到案例相对应的任务中。教学实践使用的开发版本最好能紧跟 Oracle 软件产品的更新，保证学生在实习或就业时能很快熟悉企业开发软件。

教师明确教法，学生明确学法，师生都要了解课程的重点和难点，并结合企业对人才的需求有侧重地进行应用实践。学生在学期结束后能够结合 JAVA 等开发语言独立完成一个小型信息管理项目，将 Oracle 数据库课程内容贯穿此项目之中，将之前所学计算机技术融会贯通，也可以在课程结束后根据自己能力选择参加 OCP 等 Oracle 认证考试。

（二）基于案例和项目驱动法的翻转课堂教学实践运用

Oracle 数据库课程的前后知识点承接联系很紧密，前一个知识点如果没有掌握牢靠，就会影响下一个知识点的学习。我们在课堂教学中以学生为主体，引入了多种教学改革方法，如案例教学法、项目驱动法、翻转课堂教学法，并将这些教学法有机融合在一起，达到较好的混合教学效果。

1.案例教学法

课堂中，数据库教学案例包括若干教学单元，每个单元对应一定的教学任务。教师根据实际情况设计不同的案例，并在教学中对部分案例进行讲解、演示、分析，指导学生对案例进行练习、应用，进一步理解和巩固知识，把理论与实践有机结合起来。案例教学有助于把理论知识以深入浅出的方式进行说明，例如，教师在讲解教务管理系统数据库表设计的案例时，可以嵌入 Oracle 数据库表字段逻辑关系、数据文件存储结构等理论知识，让学生对接下来的实际操作形成知识概念，自己动手进行上机操作完成像“学生—成绩”的数据库库表，激发学生的学习兴趣，理论与实践同时进行，有助于短时记忆变成长时记忆，培养学生动手能力和独立思考的精神。

本课程采用的案例有人力资源管理系统、教务管理系统，分别设计了不同的课程实验，如表 1 所示。案例教学法是分单元、分阶段、局部的教学法，在课堂中的案例小测验多采用验证式实验，案例之间大多是并行的，关联度不大。表 1 是实验项目设置与内容要求。

表 1 与案例相关的 Oracle 数据库课程实验列表

序号	实验项目名称	内容要求（了解、理解、掌握）	实验学时	每组人数	实验属性	实验要求	备注
1	系统登录和表的创建	掌握	3	1	综合	必做	
2	Exp imp 和数据查询	掌握	3	1	综合	必做	
3	表空间、数据文件和数据查询	掌握	3	1	综合	必做	
4	数据库链接和数据查询	掌握	3	1	综合	必做	
5	初始化文件和数据查询	掌握	3	1	综合	必做	
6	控制文件和 PL/SQL	掌握	3	1	综合	必做	
7	重做日志和函数	掌握	3	1	综合	必做	
8	归档模式和游标	掌握	3	1	综合	必做	
9	归档模式恢复和批量提取	掌握	3	1	综合	必做	
10	不完全恢复和存储过程	掌握	3	1	综合	必做	
11	Rman 和存储过程	掌握	3	1	综合	必做	

2.项目教学法

为了仿真软件企业的项目开发环境，让学生投入完整的Oracle数据库项目操作中，我们从全局上设计了一个一体化教学方法，即将项目驱动法与案例教学法相结合，使学生能参与项目的各个阶段从而更好地掌握知识，又能将数据库与开发语言技术相结合。在教学中需要重新筛选最合理的教学内容，将项目任务分解为具有综合性和实践性的工程教学内容，完善教师和学生之间的互动方式，以培养具有牢固基础、实践能力、符合企业实际需求的应用型人才。

项目教学法将制定、实施、总结三维一体，通过学生投身项目设计、实践、管理的一系列过程，通过项目引导学生学习数据库基础知识，完成各阶段案例教学任务，更新学生的知识建构，获得策划、操作和管理数据库应用系统的技能，提高了实际数据库应用技能，体会到做中学的乐趣。在这个过程中，实践项目驱动教学一定要安排合理的项目任务，不仅要把课程知识点、项目流程、工程要素和管理要素融会贯通，还要将教学内容和教学目标隐含在项目任务之中。表 2 是“人力资源管理系统”的项目任务设计示例。

表 2 项目任务设计示例

项目名称：人力资源管理系统			
模块一：员工管理			
任务	内容	能力目标	知识点
1	员工信息的插入和更新	数据表设计	1.熟悉字段常用的数据类型 2.掌握表的设计和创建 3.掌握表的增删改和主外键设置

在表2的任务中，学生要设计出员工信息表的结构、字段长度、数据类型，确立该表与其他表的逻辑关系、约束关系等，这一过程锻炼了学生的独立设计能力。在项目任务设计中，任务设计要由简单到复杂，保证学生实施后续任务的承接性，因此，设计任务时必须要条理清晰逐层设计，让学生逐步由简到繁地完成教学目标。

(三)翻转课堂教学模式设计方案

翻转课堂的教学模式要求学生先对课程进行复习，再在课堂上通过小组讨论的形式促使学生之间资源共享，促进知识内化，其重点是学生自身的发展，不仅是掌握知识和训练技能。本课程的翻转课堂教学活动结合上述案例和项目驱动教学法，具体实施翻转课堂教学模式的教学实践如图1所示。

1.知识串讲

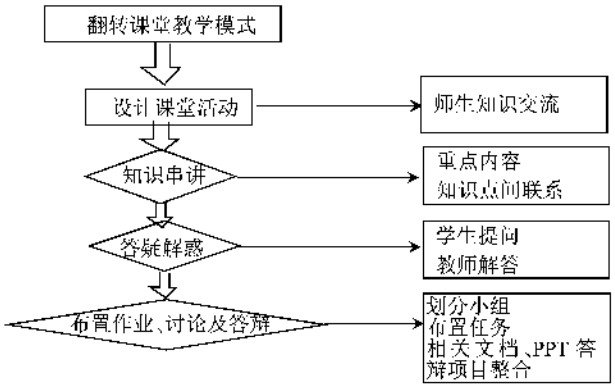
教师根据学生的学习情况和自学能力，在学习每一章之前安排课程导学，由助教(或研究生)给学生讲解该章节的核心概念及哪些部分是重点内容，降低学生自主学习的难度。将与本章节教学有关系的知识点适当提及，方便学生对知识点的理解，用原有的知识去学习新知识。

2.答疑解惑

每周定期给学生集中提出问题的时间，对于大部分学生都不懂的知识点统一讲解，有助于学生解决学习过程中自己遇到的问题，培养学生独立思考、多项思维认识事物的能力，培养学生不盲目崇拜，用批判的眼光，取其精华去其糟粕地深刻理解数据库技术内容，提高创新精神和创新能力。有些问题教师需要跨章节或沿知识的历史发展情况去解答。

3.布置作业、讨论及答辩

根据本课程项目驱动的案例设计和分组要求，学生以小组为单位完成项目内容，每组选出一名组长协调团队，分配组员作业任务，每个组员可进行课外讨论或请求教师在线指导完成自己的作业任务。在学期最后一周由组长对项目进行汇总，每个组员将自己所完成的内容制作成PPT及相关文档，进行展示答辩和提交Word文档。如下图所示：



Oracle数据库技术课程翻转课堂教学模式的教学实践

(四)完善课程考核方式

课程的考核方式一定程度上决定了学生努力的方向，影响了教学内容的安排和教学方法的选择。本课程教学过程中，教师采取多元化的考试方式，从课堂小实验、随堂测试、阶段实验报告、期末项目汇报、期末上机测试等五方面了解学生学习情况，最后按平时成绩(30%)、实验报告(30%)和期末成绩(40%)的比例作为学生的总评成绩，这三部分考核情况为：(1)平时成绩为学生在课堂上小实验、随堂测试成绩、平时考勤和提问题表现的综合情况；(2)实验报告是完成项目阶段性任务的实验报告成绩，从实验写作水平(条理性)、实验准确性、问题归纳等方面进行打分；(3)期末成绩则包括期末项目汇报、期末上机测试的成绩，项目汇报由小组成员共同完成，另外每个学生要在十分钟内上机完成指定数据库命令和操作，考核的是学生对常用数据库操作的熟练程度。成绩总分在期末进行汇总后加权平均，采用这种考核机制能调动学生的学习兴趣，培养动手能力。

四、结束语

作为计算机学院的Oracle数据库技术课程，重点要培养学生的动手能力和实践能力，培养出社会需要的应用型人才。本课程的改革实践以应用型人才培养为目标，采用了案例分析和项目驱动相结合的教学方法，能够加深学生对数据库知识的理解，使学生灵活运用所学知识完成项目作业，在教学中逐步体现出数据库的实用价值。同时，采用翻转课堂的教学模式，以学生为主体，激发学习兴趣，实现教学相长，提高学生解决实际问题、动手和分析的能力。应用型教学理念的引入更加凸显了数据库教学要注重实践能力的培养。学生具有实践能力才能更好地适应社会生活和工作，在激烈的竞争下拥有一技之长，成为受企业青睐的应用型、高层次计算机人才。

参考文献：

[1]万常选，吴京慧，廖国琼，等.数据库系列课程中创新性学习模式的探索与实践[J].计算机教育，2009(15)：13-15，22.

[2]胡文瑜.数据库开发技术课程建设和教学改革[J].计算机教育，2010(20).

[3]万常选，刘喜平，廖国琼，等.数据库系统原理课程的翻转课堂教学模式探索[J].计算机教育，2015(20)：53-57，63.

[4]陈晓菲.翻转课堂教学模式的研究[D].华中师范大学，2014：25-28.

[5]周溢辉.项目导向任务驱动的数据库应用课程改革探讨[J].软件，2012(4).

[6]林先念.试论项目式教学在高职数据库教学中的探索与实践[J].信息通信，2013(6).

[7]计成超，严红丽.项目教学法在Oracle数据库教学中的应用[J].滁州学院学报，2013(2)：117-119.

[8]杨进，郑光荣，唐德玉.数据课程教学改革探讨与实践[J].教育教学论坛，2011(32)：136-137.

[9]屈武江.高职数据库原理与应用课程教学改革的探讨[J].林区教学，2013(3).

[10]孟宪虎.面向应用型人才培养的数据库系列课程体系和教学模式改革探索[J].计算机教育，2009(24)：6-9.

[11]林川，潘盛辉，黄庆南.数据库课程教学改革研究与探讨[J].中国电力教育，2010(27)：74-75.

一种过完备字典实现指纹图像的鲁棒水印算法

赵若妍，黄智慧，肖　冰，赵慧民*，詹　瑾
（广东技术师范大学 计算机科学学院，广东 广州　510665）

摘　要：针对图像的信息取证与高精度恢复要求，结合分块压缩感知理论及其过完备字典的稀疏表示，提出一种鲁棒的自适应图像水印实现算法．基于水印图像特征结构的分布特点，将其分解为平滑、纹理和边缘的同性结构区域，组成多成分过完备字典；并结合可变的采样率，构建了一种结构自适应多成分稀疏表示和水印实现模型．同时，利用过完备字典的多成分结构和稀疏表示系数的先验知识，有效实现数字水印提取和原始指纹图像重构，实验结果表明该方法具有更高的准确性和较好的鲁棒性．

关键词：压缩感知；过完备字典；指纹图像；结构；水印
中图分类号：TP309　　**文献标识码：**A　　**文章编号：**1672-402X（2019）06-0001-07
DOI:10.13408/j.cnki.gjsxb.2019.06.001

0　引言

数字水印是实现信息隐藏技术的一种重要技术，数字水印的数据主要通过酉变换对图像进行处理后，再结合随机扰乱和加密扩频产生．但数字水印具有不可见性、稳健性和随机检测性挑战，限制其在载体信息的有效性[1-3]．从信息隐藏十几年的研究发展来看，研究人员关注的焦点是如何把数字水印作为秘密数据与载体信号进行调制/解调的实现，以及如何在拒电

础，提出一种新的指纹水印生成算法——BCSW算法(Block Compressed Sensing Watermark，BCSW)．同时，为了通过水印高精度重构指纹图像，本文通过稀疏表示设计了一种有效的指纹数据过完备字典，并提出了一种新的图像重构算法——BCSW-SPL算法(Block-based compressed sensing watermark with smooth projected Landweber reconstruction)[9]．近年来，冗余字典的研究取得了突破性的进展，

广东技术师范学院学报 . 2019,40(06)

记笔记

一种过完备字典实现指纹图像的鲁棒水印算法

赵若妍 黄智慧 肖冰 赵慧民 詹瑾

广东技术师范大学计算机科学学院

摘要： 针对图像的信息取证与高精度恢复要求,结合分块压缩感知理论及其过完备字典的稀疏表示,提出一种鲁棒的自适应图像水印实现算法.基于水印图像特征结构的分布特点,将其分解为平滑、纹理和边缘的同性结构区域,组成多成分过完备字典;并结合可变的采样率,构建了一种结构自适应多成分稀疏表示和水印实现模型.同时,利用过完备字典的多成分结构和稀疏表示系数的先验知识,有效实现数字水印提取和原始指纹图像重构,实验结果表明该方法具有更高的准确性和较好的鲁棒性.

关键词： 压缩感知; 过完备字典; 指纹图像; 结构; 水印;

DOI： 10.13408/j.cnki.gjsxb.2019.06.001

专辑： 社会科学II辑; 信息科技

专题： 计算机软件及计算机应用

分类号： TP309.7

一种过完备字典实现指纹图像的鲁棒水印算法

赵若妍，黄智慧，肖 冰，赵慧民*，詹 瑾
（广东技术师范大学 计算机科学学院，广东 广州 510665）

摘 要：针对图像的信息取证与高精度恢复要求，结合分块压缩感知理论及其过完备字典的稀疏表示，提出一种鲁棒的自适应图像水印实现算法．基于水印图像特征结构的分布特点，将其分解为平滑、纹理和边缘的同性结构区域，组成多成分过完备字典；并结合可变的采样率，构建了一种结构自适应多成分稀疏表示和水印实现模型．同时，利用过完备字典的多成分结构和稀疏表示系数的先验知识，有效实现数字水印提取和原始指纹图像重构，实验结果表明该方法具有更高的准确性和较好的鲁棒性．

关键词：压缩感知；过完备字典；指纹图像；结构；水印
中图分类号：TP309 **文献标识码：**A **文章编号：**1672-402X（2019）06-0001-07
DOI:10.13408/j.cnki.gjsxb.2019.06.001

0 引言

数字水印是实现信息隐藏技术的一种重要技术，数字水印的数据主要通过西变换对图像进行处理后，再结合随机扰乱和加密扩频产生．但数字水印具有不可见性、稳健性和随机检测性挑战，限制其在载体信息的有效性[1-3]．从信息隐藏十几年的研究发展来看，研究人员关注的焦点是如何把数字水印作为秘密数据与载体信号进行调制/解调的实现，以及如何在折中鲁棒性和安全性的前提下通过提高嵌入水印容量有效恢复隐藏的信息等问题[4-6]．

最近几年兴起的压缩感知理论(Compressed Sensing，CS)是信号处理和数学领域中的重要理论，其本质是对稀疏或可压缩信号(如图像)能够以高概率由少量测量数据恢复整个信号[7-8]．CS主要思想是采用过完备字典代替传统的正交基函数揭示信号的内在结构，并通过稀疏表示、非相干测量和重构算法(目前主要是凸优化算法和贪婪算法)恢复原始信号．

本文以生物指纹图像为研究对象，以分块压缩感知(Block Compressed Sensing，BCS)为基础，提出一种新的指纹水印生成算法——BCSW算法(Block Compressed Sensing Watermark，BCSW)．同时，为了通过水印高精度重构指纹图像，本文通过稀疏表示设计了一种有效的指纹数据过完备字典，并提出了一种新的图像重构算法——BCSW-SPL算法(Block-based compressed sensing watermark with smooth projected Landweber reconstruction)[9]．近年来，冗余字典的研究取得了突破性的进展，有效实现单一成分的稀疏表示，主要有正弦函数的级联、局部余弦字典、Grassmann框架，Anisotropic Reefinement — Gaussian(AR-G)混合字典[10]等，虽然上述字典能达到一定的表示效果，但难以用于细节特征的二维指纹图像．为此，本文构造了自适应结构化的过完备稀疏表示模型及BCSW生成算法，与早期使用整个多形态字典对高维图像进行稀疏分解不同，该算法以L_1范数作为信号稀疏性的能量度量标准分解指纹图像．针对指纹图像中平滑、边缘、纹理等多种结构成分，将同性结构分为同一区域，各区域以类型一致的单一子成分字典实现匹配

收稿日期：2019-11-02
作者简介：赵若妍，广东技术师范大学2018级硕士研究生．
*** 通讯作者：**赵慧民，广东技术师范大学教授，E-mail：zhaohuimin@gpnu.edu.cn

跟踪稀疏分解. 实验结果验证该算法提高了指纹水印图像提取质量和重构精度，同时不增加计算复杂度.

本文的研究丰富和发展了数字图像水印和压缩感知相关的研究理论，实现指纹图像有效取证并进行稀疏分解，提高指纹图像在刑侦、安全、宣传等实际应用中的精确度和可信度，研究成果在数字签名、智能身份信息验证和打击罪犯等领域能发挥重要作用.

1 BCSW 指纹图像水印算法与过完备字典的实现

1.1 BCS 观测实现原理

BCS技术是指通过图像分块采样，并行实现指纹水印数据的提取和观测值检测，高效完成图像传输过程. 把 $N\times N$ 的二维图像使用压缩感知技术分解成若干 $d\times d$ 图像块[11-13]，好的采样方法可有效收集图像重要信息，以随机投影矩阵 Φ_d 对每个图像块完成信息采样，矩阵大小为 $m_d\times d^2(m_d\in M)$，则对于第 i 个图像块 x_i 的可压缩信号 α，在稀疏基函数 Ψ_i 作用下，其观测向量可表示为：

$$y_i=\Phi_d\cdot x_i=\Phi_d\cdot\psi_i\cdot\alpha,\quad i=1,\cdots,\frac{N^2}{d^2} \tag{1}$$

基于(1)式分块观测后，当 $Y=\{y_i,i=1,...,M\}$，$X=\{x_i,i=1,...,N^2/d^2\}$，$\Psi=[\psi_1,\cdots,\psi_N]$ 时，整幅指纹图像的采样过程可表示为：

$$Y=\Phi\cdot X \tag{2}$$

其中，对角矩阵 Φ 是 Φ_d 转变而成.

1.2 指纹图像水印数据的生成

CS理论突破了信号采样过程中奈奎斯特采样定律限制，减少图像采样数据的存储量，基于压缩感知的稀疏性和可压缩性，将图像映射为更加简洁的表示形式，并结合线性投影实现图像重构. 图像压缩是指用少量的采样数据尽可能多地包含重构原始图像的重要信号，通过观测矩阵将稀疏系数映射到低维观测向量中，映射得到的观测值与原图像是唯一对应且没有损失，包含原始数字图像整体形态结构全部特征信息[14-15]. 根据这一特性，在压缩感知理论下，结合指纹水印技术，提出一种应用BCS水印实现技术——BCSW，其实现流程如图1所示.

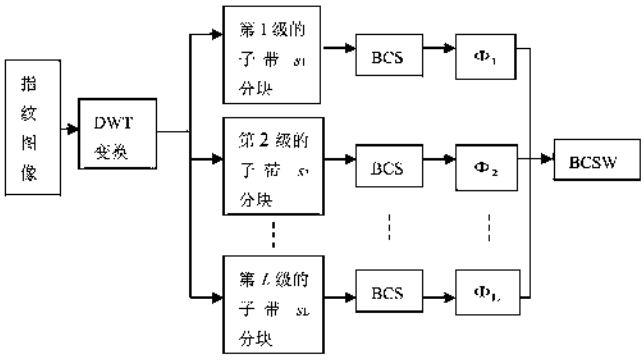


图 1 BCSW 构造流程图

假设指纹图像的大小为 $M_1\times M_2$，BCSW指纹水印的数据量为 $M\times M$，那么第 l 级指纹图像子块的测量数据为 $m=\lfloor Md_l^2/M_1\times M_2\rfloor$. 如果 l 级子带 s 上的第 i 块图像数据 $\tilde{x}_{l,s,i}$ 进行BCS的测量样本为 $y_{l,s,j}(j=1,2,...,d_l^2)$，则所有指纹图像子块的观测值组合并归一化后最终生成的BCSW数据形式为：

$$Y_M=\begin{bmatrix} y_{1,1,1} & y_{2,1,1} & \cdots & y_{L,1,1} \\ \vdots & \vdots & \cdots & \vdots \\ y_{1,2,1} & y_{2,2,1} & \cdots & y_{L,2,1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1,4,d_1^2} & y_{2,4,d_2^2} & \cdots & y_{L,4,d_L^2} \end{bmatrix}=\{W(l,s,j),W(l,s,j)\in(0,1)\} \tag{3}$$

这里，生成的BCS观测矩阵 Φ_l 的种子和DWT各级子带上的分块大小分别作为BCSW的密钥 k_1 和 k_2. 指纹图像在BCS域中根据不同形态结构分块后，由BCS在各图像子块完成自适应采样，形成数据量少但信息丰富的观测值，再进一步生成相应随机数据，结合各个随机数据生成BCSW数据，这是线性投影过程，因此，BCSW生成速率受观测矩阵 Φ_l 和各级子带上的分块大小 B_l 影响，其复杂性为 $O(mB_l^2)$.

1.3 BCSW 的过完备字典表示模型

为了高质量提取指纹水印数据和高精度恢复原始指纹图像，根据BCSW值 Y_M 的能量 $E=\|Y_M\|^2$ 大小及其确定的阈值 T，BCSW的平滑区域、纹理区域和边缘区域划分依据下公式(4)：

$$\begin{cases} T\le 0.4\cdot E_m, \text{平滑子块}\Omega_{smooth} \\ 0.4\cdot E_m<T\le 0.85\cdot E_m, \text{纹理子块}\Omega_{texture} \\ T>0.85\cdot E_m, \text{边缘子块}\Omega_{edge} \end{cases} \tag{4}$$

其中，观测值能量E的均值定义为$E_m=\frac{\sum_i E_i}{N}$，而阈值T的大小由$N\times N$的二维指纹图像$X(i,j)$的信息熵决定：

$$T=-\sum_{i=1}^{N}\sum_{j=1}^{N}X(i,j)\log_2 X(i,j) \tag{5}$$

考虑对输入图像做一次2D的DWT，在其逼近子带LL上保留了原图像的近似信息，但大小仅为原来的1/4. 对LL子带做重叠子块的分割，得到的子块数量也是对原图做重叠子块分割的1/4. 由于子块和特征向量一一对应，这样参与排序的特征向量自然就少了，从而减少了排序过程使用的时间. 在对子块的特征表示上，该方法使用了SVD(singular value decomposition).

由于平滑图像子块和纹理图像子块以及边缘图像子块的结构特征不同，按照(4)式的分类并借助于集合的概念，BCSW可对应分解为互不重叠的平滑、纹理和边缘三个类型：

$$\Omega=\Omega_{smooth}\cup\Omega_{texture}\cup\Omega_{edge} \tag{6}$$

其中，Ω是包含了图像各类特征的BCSW的总和，BCSW进一步划分为不同形态结构子集，并选用对应的字典表示，形成更具有针对性自适应稀疏的BCSW信号表示形式. 过完备BCSW多成分字典综合了原始指纹图像所有几何结构字典(如平滑、纹理和边缘)，并可在低维空间实现有效跟踪匹配，其组成如下所示：

$$D=D_{smooth}\cup D_{texture}\cup D_{edge} \tag{7}$$

这里，字典D_{smooth}，$D_{texture}$，D_{edge}分别是依据BCSW数据类型构成的平滑、纹理和边缘形态结构子字典，各字典相互补充，形成过完备字典D，能够有效匹配指纹图像中的多种几何结构，优化指纹图像的稀疏表示效果. 字典采用最小加权L_1范数度量其稀疏性，则Ω依据字典结构形成的自适应稀疏表示模型为：

$$\begin{aligned}&\min(\|\alpha_s\|_1+\|\alpha_t\|_1+\|\alpha_e\|_1)\\&s.t.\quad \Omega=\Phi_s\alpha_s\cup\Phi_t\alpha_t\cup\Phi_e\alpha_e\end{aligned} \tag{8}$$

其中，$\{\Phi_s,\Phi_t,\Phi_e\}\in\Phi$代表了指纹图像平滑、纹理与边缘结构子成分字典$D_{smooth}$，$D_{texture}$，$D_{edge}$中的原子作为列向量通过DCT和DWT变换所形成的字典矩阵，而$\{\alpha_s,\alpha_t,\alpha_e\}\in\{\alpha=\langle y,\Psi\rangle\}$为子集$\Omega_{smooth}$，$\Omega_{texture}$，$\Omega_{edge}$在对应子成分字典$D_{smooth}$，$D_{texture}$，$D_{edge}$中的稀疏分解系数. 即，$\alpha_s$代表了通过$\Phi_s$观测获得的平滑水印稀疏系数，$\alpha_t$代表了$\Phi_t$观测获得的纹理水印稀疏系数(指纹图像的脊线水印稀疏系数)，α_e代表了Φ_e观测获得的边缘水印稀疏系数(即指纹图像的谷线水印稀疏系数).

1.4　BCSW过完备字典的构造

设当前BCSW数据对应像素点的坐标为(i,j)，其邻域为$R=\{(m,n)\|m-i|\le r,|n-j|\le r\}$，$r$为邻域长度的一半[16]，则该点BCSW数据的方向一致性度量定义为：

$$C_{or}=\left[1-\frac{\left|\sum_{(m,n)\in R}\sin(2\theta_{m,n})\right|}{\sum_{(m,n)\in R}|\sin(2\theta_{m,n})|}\right]\times\left[1-\frac{\left|\sum_{(m,n)\in R}\cos(2\theta_{m,n})\right|}{\sum_{(m,n)\in R}|\cos(2\theta_{m,n})|}\right] \tag{9}$$

式中$\sin(2\theta)$，$\cos(2\theta)$分别表示方向$\theta(\theta\in(\frac{-\pi}{2},\frac{\pi}{2}))$的垂直和水平分量，$2\theta$是为了克服$\theta$从$\frac{-\pi}{2}\to\frac{\pi}{2}$时的不连续问题.

通过(9)式，我们将指纹图像库中的图像块按其方向性从$\frac{-\pi}{2}$到$\frac{\pi}{2}$共分为12类，采用最常用的字典学习算法K-SVD(K-Singular Value Decomposition)对每一类数据进行训练，得到满足条件的不同类型字典[16-18]，每类字典包括64个原子，另外再加上正交DCT字典一共形成包括832个原子的过完备BCSW字典.

这样，由子成分字典D_{smooth}，$D_{texture}$，D_{edge}中的原子(BCSW)稀疏表示和分类模型可见，第i个图像块x_i的水印数据y_i在BCSW子集Ω_{smooth}，$\Omega_{texture}$，Ω_{edge}中的形式为：

$$\begin{aligned}y^s{}_i&=\Phi_s\cdot D_{smooth}\alpha^i{}_s\\y^t{}_i&=\Phi_t\cdot D_{texture}\alpha^i{}_t\\y^e{}_i&=\Phi_e\cdot D_{edge}\alpha^i{}_e\end{aligned} \tag{10}$$

其中，第i个图像块x_i的水印信号稀疏系数$\alpha_s^i,\alpha_t^i,\alpha_e^i\in\alpha$. 通过求解(10)式获得最优稀疏分解字数$\alpha$，为过完备字典D提供特征向量有效表示，形成最佳水印数据. 本文根据最小加权L_1范数的约束条件，通过(11)式分别计算D中各个子成分字典D_{smooth}，$D_{texture}$，D_{edge}下第i个图像块x_i的水印稀疏系数：

$$\begin{aligned}\alpha^i{}_s&=\arg\min_{\alpha_s}\{\|y_i^s-\Phi_s\cdot D_{smooth}\alpha_s\|_2^2+\rho\|\alpha_s\|_1\}\\\alpha^i{}_t&=\arg\min_{\alpha_t}\{\|y_i^t-\Phi_t\cdot D_{texture}\alpha_t\|_2^2+\rho\|\alpha_t\|_1\}\\\alpha^i{}_e&=\arg\min_{\alpha_e}\{\|y_i^e-\Phi_e\cdot D_{edge}\alpha_e\|_2^2+\rho\|\alpha_e\|_1\}\end{aligned} \tag{11}$$

式中 ρ 表示加权系数，其求解过程可参考Lasso算法实现. 通过调用SparseLab库中的SolveLasso函数求解，ρ 的取值具体形式为 $\rho=\max\{|\Phi_d^T\cdot y_i|, y_i=(y_i^s,y_i^t,y_i^e), i=1,2,\cdots^{N^2}_{d^2}\}$. 并通过计算稀疏系数 $\alpha_s^i,\alpha_t^i,\alpha_e^i$ 的最小加权 L_1 范数，可分别得到第 i 个图像块 x_i 的平滑、纹理与边缘结构的BCSW数据向量 $y_i=\{(y_i^s,y_i^t,y_i^e), i=1,\ldots,\frac{N^2}{d^2}\}$，并将其按照(4)计算的能量大小对应排列在各个子集 Ω_{smooth}，$\Omega_{texture}$，Ω_{edge}，最后得到对应的各个子成分字典 D_{smooth}，$D_{texture}$，D_{edge} 下的BCSW原子. 据此，可得到整幅图像的结构化BCSW数据及其字典D.

同时，为了进一步确保安全性，本文用随机种子函数 s 对BCSW数据原子进行 cr 扩频变换，设计了BCSW数据向量基 W：

$$y_i=\{y_1,y_2,\cdots,y_n\},\ y_i\in\{0,1\}, i\in M \quad (12)$$

$$EX=\{e_1,e_2,\cdots,e_n\},\ e_i=y_i,$$

$$j\cdot cr\le i\le(j+1)\cdot cr,\ cr>1, i,j,cr\in M \quad (13)$$

$$w_i=\{w_0,w_2,\cdots,w_{M-1}\}, w_i=e_i\cdot s,\ i\in M \quad (14)$$

$$W=\{w_i, i=1,\ldots,N^2/d^2\in M\} \quad (15)$$

这里，我们把随机种子作为密钥 k_3. 由此，结合BCS观测矩阵的密钥 k_1 和图像分块大小密钥 k_2，我们能够通过密钥形成对BCSW的控制.

1.5 BCSW数据的嵌入与提取

在BCSW的实现过程中，依据扩频技术和HVS特征模型，嵌入的数据水印W具有更强的安全性和鲁棒性，嵌入形式如下所示：

$$\vec{V}=V+\beta\cdot W \quad (16)$$

其中，β 为数据嵌入的强度，V 为载体图像变换后频率系数，$\vec{V}$ 作为嵌入结果，表示生成的混合图像频率系数. 包含水印的混合图像不能直接提取数据，需要逆变换处理获得其估计值，估计条件下的数字水印值为：

$$\overline{W}=\frac{\vec{V}-\overline{V}}{\beta} \quad (17)$$

其中，$\overline{V}$ 是频率系数 V 的估计值，并由 $\overline{V}$ 获得嵌入强度 β，原始指纹图像的观测估计数据 $\overline{y_i}=\{(\overline{y_i^s},\overline{y_i^t},\overline{y_i^e}), i=1,\ldots,\frac{N^2}{d^2}\}$ 可根据公式(12)-(15)获取水印估计数据 $\overline{W}$ 的逆过程获得.

1.6 基于BCSW字典的指纹图像重构

为了从 $\overline{y_i}$ 中寻找各个子成分字典 D_{smooth}，$D_{texture}$，D_{edge} 下第 i 个图像块 x_i 的估计数据 $\overline{x_i}$，结合文献[9]的BCS-SPL算法实现指纹图像重构. 同时，采用了3×3的Wiener滤波器消除图像重构时的分块边界效应.

平滑成分恢复：对于第 i 个平滑图像块 x_i 的重构可通过优化平滑图像块的观测估计值 $\overline{y_i^s}$ 进行求解：

$$\overline{x}_i=\arg\min_{x_i}\{\|\overline{y_i^s}-\Phi_l x_i\|_2^2+\lambda_s\|D^T{}_{smooth}\cdot x_i\|_1\} \quad (18)$$

$$\overline{x_i^{k+1}}=T_s(\overline{x_i^k}+\Phi_l{}^T(\overline{y_i^s}-\Phi_l\cdot\overline{x_i^k}) \quad (19)$$

式中 λ_s 为收敛调节因子，T_s 为平滑图像块的阈值门限，一般地，$T_s=\lambda_s\sigma_s\sqrt{2\log K}$（$K$ 为平滑块非0元素个数，而 $\sigma_s=\frac{medium(|\overline{x_i}|)}{0.6745}$）.

纹理和边缘恢复：纹理和边缘部分的特征细节结构较复杂，DWT变换域过程中，在多尺度分块方向字典下精确逼近原始图像，设 x_i^t,x_i^e 分别为邻域 $R=\{(m,n)\|m-i|\le r,|n-j|\le r\}$ 内与第 i 个纹理或边缘图像块 x_i 的相似图像块，残差 $e_i^t=\|x_i-x_i^t\|_2^2$ 和 $e_i^e=\|x_i-x_i^e\|_2^2$ 分别为判断相似性的依据. 用 d 个相似块的线性投影组合形成对 x_i 的估计值为：

$$\overline{x_i^t}=\sum_{i=1}^{d}C_{or}^t\cdot x_i^t$$

$$\overline{x_i^e}=\sum_{i=1}^{d}C_{or}^e\cdot x_i^e \quad (20)$$

其中，C_{or}^t，C_{or}^e 分别为公式(9)中对应像素点BCSW数据的方向一致性度量值. 用分块残差最小化作为 L_1 范数的约束条件，指纹图像的纹理和边缘在方向字典下的稀疏表示逼近系数为：

$$\alpha_t^i=\arg\min_{\alpha_t}\{\frac{1}{2}\|y_i^t-\Phi_t\cdot D_{texture}\alpha_t^{i-1}\|_2^2+\lambda_l|\alpha_t\|_1+\frac{\gamma}{2}|D_{texture}\alpha_t^{i-1}-\overline{x_i^t}\|_2^2\}$$

$$\alpha_e^i=\arg\min_{\alpha_e}\{\|y_i^e-\Phi_e\cdot D_{edge}\alpha_e^{i-1}\|_2^2+\lambda_l\|\alpha_e\|_1+\frac{\gamma}{2}\|D_{texture}\alpha_e^{i-1}-\overline{x_i^e}\|_2^2\} \quad (21)$$

式中 λ_l 为DWT变换的第 l 级权重，γ 为稀疏表示迭代控制因子. 通过(21)求出 α_t^i 和 α_e^i 后，分别对其求稀疏逆变换得到重构的图像：$\overline{X}=D\cdot\alpha$. 本文通过交替迭代优化求解(21)，令

$x_i = D_i\alpha_i, D_i \in \forall(D_{texture}, D_{edge}), \alpha_i \in \forall(\alpha_t^i, \alpha_e^i)$，(21)

可分解为两个优化的共性问题：

$$x_i^{k+1} = \arg\min_{x_i}\{\frac{1}{2}\| y_i - \Phi_l x_i \|_2^2 + \frac{\gamma}{2}\| x_i - \bar{x}_i \|_2^2\} \quad (22)$$

式中 γ 与迭代次数 k 成正比，实验取值为 $\gamma = 0.02\cdot k$. $\bar{x}_i \in \forall(x_i^t, x_i^e)$，代表了 x_i 的估计值. 根据分块相似性误差最小的原则，采用 Lasso 算法求解方向字典中的最稀疏表示系数：

$$\alpha_i^{k+2} = \arg\min_{\alpha_i}\{\| x_i^{k+1} - D_i\alpha_i \|_2^2 + \lambda_l \| \alpha_i \|_1\} \quad (23)$$

$$x_i^{k+2} = D_i \cdot \alpha_i^{k+2} \quad (24)$$

采用梯度投影下降算法对（22）求解有：

$$x_i^{k+1} = x_i^k + \Phi_l^T(y_i - \Phi_l x_i^k) - \gamma(x_i^k - \bar{x}_i) \quad (25)$$

(25)式的求解需要先确定迭代的初始值 x_i^0. 为此，我们用 Lasso 算法首先求解字典中的初始最优稀疏表示系数：

$$\alpha_i^0 = \arg\min_{\alpha_i}\{\frac{1}{2}\| y_i - \Phi_l D_i\alpha_i \|_2^2 + \lambda_l \| \alpha_i \|_1\} \quad (26)$$

再通过稀疏逆变换即可得到指纹图像块 x_i 的初始值：$x_i^0 = D_i \cdot \alpha_i^0$.

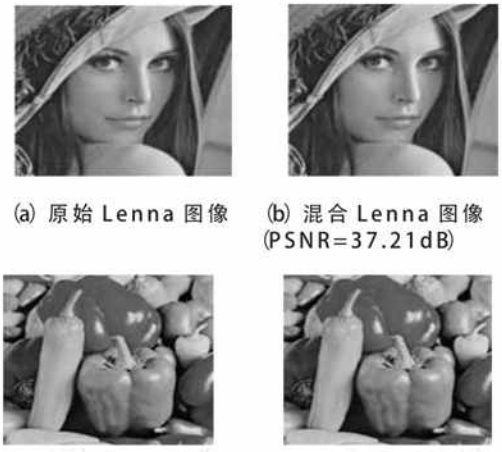
2　实验结果与讨论

为了与文献[19]的DCT水印和文献[20]DWT水印结果进行比较，本文选用同样大小的 512×512 标准灰度图像Lena和Peppers作为载体图像. 指纹水印图像实验采用FVC2002指纹库作为试验的数据集. 稀疏基 Ψ 选择3级9/7 DWT，并在DWT域每级子带上使用文献[21]的块DCT SRM作为观测矩阵. 实验中，当 l = 1, 2,3 时，对应分块大小 d=32,16,8，并分别选用S=0.1,0.2,0.3三种BCS的采样率进行观测. BCSW水印数据的嵌入使用常用的扩频调制技术. 指纹图像重构，在字典下采用本文的BCSW-SPL算法实现，并用相关性NC作为原始指纹X与提取指纹 $\overline{X}$ 的相似性度标准：

$$NC = \frac{\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} X(i,j)\cdot\overline{X}(i,j)}{\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} X^2(i,j)} \quad (27)$$

2.1　透明性实验

在Lenna和Papper图像中分别嵌入图3的指纹观测数据，形成包含水印信息的混合图像，量化步长为Q=35，图2（a）与（c）是原始图像，（b）和（d）是嵌入指纹后的混合载体图像，比较水印数据嵌入前后的载体图像，没有产生明显的变化，实验结果体现了图像具有较好的透明性.



(a) 原始 Lenna 图像　(b) 混合 Lenna 图像 (PSNR=37.21dB)

(c) 原始 Pepper 图像　(d) 混合 Pepper 图像 (PSNR=35.52dB)

图 2　图像数字水印的隐藏性

2.2　数字水印的鲁棒性实验

设定S=0.1、Q=35，分别采用压缩，中值滤波，加入Salt&Pepper噪声，旋转，剪裁和缩放等攻击行为进行水印鲁棒性检测. 表1分别说明了攻击后混合载体图像的PSNR与提取的指纹图像BCR值. 同等条件下，表1也和[19][20]的DWT水印算法进行了比较.

分析表(1)的PSNR实验结果，针对压缩和几何变换的攻击行为，BCSW算法表现出最好的鲁棒性. 对比BCR值，说明在指纹图像重构过程中，BCSW算法具备更好的一致性，并表明该算法分类训练获得的过完备字典对原始指纹图像具有更加精确的稀疏逼近. 综上分析，本文算法在安全性上具有更高的可靠性和完整性.

3　结论

在DWT变化域下，结合数字水印和压缩感知理论，本文提出了一种基于过完备字典分块压缩感知的自适应图像水印实现算法. 通过不同攻击行为，对指纹图像嵌入水印数据，并通过压缩感知构建一种结构自适应多成分稀疏表示模型，将指纹图像剖分为具有平滑、纹理和边缘结构同性的子块，每个子块由单一子成分字典

表 1 攻击后混合图像的 PSNR(dB) 与提取的指纹图像的 BCR 值

攻击行为	Linna						Pepper					
	BCSW		DWT-DCT		DWT		BCSW		DWT-DCT		DWT	
	PSNR	BCR	PSNR	BCR	PSNR	BCR	PSNR	BCR	PSNR	BCR	PSNR	BCR
Jpeg2000 (rate=0.1bpp)	37.96	0.999	36.02	0.961	35.89	0.957	36.89	0.996	35.76	0.952	35.21	0.945
Jpeg (Quality=20%)	36.30	0.992	35.28	0.985	35.01	0.979	36.16	0.990	34.93	0.978	34.42	0.972
中值滤波 (window 7×7)	33.26	0.990	32.68	0.982	18.61	0.968	32.92	0.985	31.94	0.938	31.27	0.961
Salt&Pepper 噪声 (σ = 0.05)	23.68	0.961	18.96	0.949	28.53	0.933	22.82	0.957	18.79	0.979	18.23	0.930
图像缩放 (scale=1/4)	29.96	0.993	28.76	0.982	27.36	0.965	29.28	0.989	28.65	0.823	28.42	0.978
图像旋转 (deg= 0.5°)	28.76	0.890	27.97	0.865	27.36	0.852	28.21	0.812	26.68	0.823	27.12	0.799
中心剪裁 (40%)	12.58	0.992	11.64	0.971	10.58	0.936	10.73	0.987	0.97	0.958	9.54	0.943
周围剪裁 Crop(45%)	12.21	0.972	10.78	0.953	10.23	0.932	9.92	0.986	8.96	0.973	8.73	0.969

进行低维子空间的匹配追踪稀疏分解，可以简洁而全面地表征图像的全部特征. 这种方法构建的水印不仅降低了指纹图像的维度，同时降低了搜索字典的复杂度，并且每一子块可并行实现水印数据，具有操作简单的优点，同时优化了其保密性、鲁棒性、抵抗攻击能力和精确度.

参考文献：

[1]Tao B, Dickinson B. Adaptive watermarking in the DCT domain[C]//1997 IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1997, 4: 2985-2988.

[2]Kesavan Gopall, Dr. M. Madhavi Latha. Watermarking of Digital Video Stream for Source Authentication[J]. IJCSI International Journal of Computer Science Issues, July 2010,7 (1):18-25.

[3]Chen W M, Lai C J, Wang H C, et al. H. 264 video watermarking with secret image sharing[J]. IET Image Processing, 2011, 5(4): 349-354.

[4]Kutter M, Bhattacharjee S K, Ebrahimi T. Towards second generation watermarking schemes[C]//Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348). IEEE, 1999, 1: 320-323.

[5]Swanson M D, Zhu B, Tewfik A H. Transparent robust image watermarking[C]//Proceedings of 3rd IEEE International Conference on Image Processing. IEEE, 1996, 3: 211-214.

[6]Podilchuk C I, Zeng W. Image-adaptive watermarking using visual models[J]. IEEE Journal on selected areas in communications, 1998, 16(4): 525-539.

[7]Tsaig Y, Donoho D L. Extensions of compressed sensing[J]. Signal processing, 2006, 86(3): 549-571.

[8]Candes E J, Romberg J. Quantitative robust uncertainty principles and optimally sparse decompositions[J]. Foundations of Computational Mathematics, 2006, 6(2): 227-254.

[9]Mun S, Fowler J E. Block compressed sensing of images using directional transforms[C]//2009 16th IEEE international conference on image processing (ICIP). IEEE, 2009: 3021-3024.

[10]I Ventura R M F, Vandergheynst P, Frossard P. Low-rate and flexible image coding with redundant representations[J]. IEEE Transactions on Image Processing, 2006, 15(3): 726-739.

[11]Candes E J, Wakin M B. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine,2008, 25(2):21-30.

[12]Kim Y, Nadar M S, Bilgin A. Compressed sensing using a Gaussian scale mixtures model in wavelet domain[C]//2010 IEEE International Conference on Image Processing. IEEE, 2010: 3365-3368.

[13]Fowler J E, Mun S, Tramel E W. Multiscale block compressed sensing with smoothed projected landweber reconstruction[C]//2011 19th European Signal Processing Conference. IEEE, 2011: 564-568.

[14]杨海蓉，张成，丁大为，韦穗. 压缩传感理论与重构算法[J]. 电子学报，2011,39(1):142-148.

[15]练秋生，周婷．结合字典稀疏表示和非局部相似性的自适应压缩成像算法[J]. 电子学报，2012，40(7)：1416-1422.

[16]田捷，何余良，陈宏，杨鑫．一种基于相似度聚类方法的指纹识别算法[J]. 中国科学E辑，信息科学 2005, 35(2): 186-199.

[17]Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on signal processing, 2006, 54(11): 4311-4322.

[18]Duarte M F, Wakin M B, Baraniuk R G. Fast reconstruction of piecewise smooth signals from incoherent projections[J]. Proc. SPARS, 2005.

[19]Al-Haj A. Combined DWT-DCT digital image watermarking[J]. Journal of computer science, 2007, 3(9): 740-746.

[20]Zhang D, Wu B, Sun J, et al. A new robust watermarking algorithm based on DWT[C]//2009 2nd International Congress on Image and Signal Processing. IEEE, 2009: 1-6.

[21]Do T T, Tran T D, Gan L. Fast compressive sampling with structurally random matrices[C]//2008 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2008: 3369-3372.

[责任编辑：刘　昱]

Robust Image Watermarking Algorithm Based on Block Compressed Sensing and Overcomplete Dictionary

ZHAO Ruo-yan, HUANG Zhi-hui, XIAO Bing, ZHAO Hui-min, ZHAN Jin

(College of Computer Science, Guangdong Polytechnic Normal University, Guangdong Guangzhou 510665)

Abstract: The structural characteristics of overcomplete dictionary have a vital influence on the performance of image recognition, therefore, the design of overcomplete dictionary is a fundamental task in image sparse representations. With the requirements of information forensics and high precision restoration in the image, this study proposes an adaptive robust image watermarking algorithm, which consists of the block compressed sensing and the sparse representation of overcomplete dictionary. A structure adaptive multi-component sparse representation and watermark implementation model are constructed by the size distribution features of block random projection energy of watermarking image, and are classified into smooth, texture and edge structure isotropic regions. In addition, a multi-component overcomplete dictionary is designed with consistent morphology. The extract of watermark and fingerprint image reconstruction both benefited from the prior knowledge about multi-component construction and sparse representation of dictionary. The experiment results verify the effectiveness of the proposed method.

Key words: compressed sensing; overcomplete dictionary; fingerprint image; structure; watermark

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

基于摄像头图像视频混合题材的滤波处理方法及系统 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：基于摄像头图像视频混合题材的滤波处理方法及系统/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

基于摄像头图像视频混合题材的滤波处理方法及系统 [发明] 有效

申请号：CN201711108390.4 申请日：2017.11.09 公开号：CN107809638A 公开日：2018.03.16
公告号：CN107809638B 公告日：2020.03.17 主分类：H04N19/117
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：詹瑾 郑鹏根 赵慧民 司瑶

摘要： 本发明公开了一种基于摄像头图像视频混合题材的滤波处理方法及系统，其方法包括：将基于摄像头所获取的混合题材数据按照先进先存的模式暂存在第一缓存单元中；基于先进先出模式提取第一缓存单元中的每一个数据对象，并对提取的第一缓存单元中的每一个数据对象进行模式识别；按照先进先存的模式重新将混合题材数据中的带有非指令识别的图像数据和带有非指令识别的视频数据暂存在第二缓存单元；并对所述第二缓存单元中的每一个数据对象进行模式识别；在判断数据对象是带有非指令识别的图像数据还是带有非指令识别的视频数据后，基于用于设置的环路滤波模式生成参数集；并根据生成的参数集进行环路滤波。通过本发明实施例针对不同图像和视频元素基于相应的控制实现及相应的编解码处理过程。

收藏 导出

检索结果共1条 < 1 >

主办单位:国家知识产权局中国专利信息中心 | 技术支持:北京新发智信科技有限责任公司
京ICP备11033758号-7 | 京公网安备 11010802035318号

进行处理，最后形成数字指纹数据根据树结构的不同分支相互发生共谋攻击的概率决定数字指纹嵌入的能量大小及其关系；经过CDMA/TDMA组合正交调制处理形成混合数据并形成加密数据流，最后通过信道传输；后者包括下列步骤：根据密钥K^m，$K^{(i)}$解密、压缩，恢复图像和视频信源数据，同时，进行解调，经过处理，相关性函数检测树结构各级发生共谋攻击，最后得到数字指纹数据。本发明安全性高，降低通信成本，同时增强系统追踪非法者的权利及其有效性。"/>

实现数字指纹加密的视频多播传输的方法

摘要

本发明公开了实现数字指纹加密的视频多播传输的方法，包括数字指纹加密传输与数字指纹加密接收，前者包括下列步骤：图像和视频的载体信源数据经过MPEG-2标准压缩；对基于图像或视频的数字指纹数据进行处理，最后形成数字指纹数据根据树结构的不同分支相互发生共谋攻击的概率决定数字指纹嵌入的能量大小及其关系；经过CDMA/TDMA组合正交调制处理形成混合数据并形成加密数据流，最后通过信道传输；后者包括下列步骤：根据密钥K^m，$K^{(i)}$解密、压缩，恢复图像和视频信源数据，同时，进行解调，经过处理，相关性函数检测树结构各级发生共谋攻击，最后得

著录项

公开/公告号	CN103152608A	专利类型	发明专利
公开/公告日	2013-06-12		PDF

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种多变量二次方程的乱序加密方法 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种多变量二次方程的乱序加密方法/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种多变量二次方程的乱序加密方法 [发明] 有效

申请号：CN201610053791.3 申请日：2016.01.26 公开号：CN105610571A 公开日：2016.05.25
公告号：CN105610571B 公告日：2019.02.05 主分类：H04L9/00
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：李伟键

摘要：本发明公开了一种多变量二次方程的乱序加密方法，包括：在计算每个多变量二次方程时，乱序生成n(n+1)/2个单项式下标值(i,j)；1≤i≤j≤n；按照单项式下标值(i,j)生成的顺序，依次计算所述多变量二次方程的n个单项式αijxixj；其中，αij为明文，xi为密钥，或者，αij为密钥，xi为明文；将所述n个单项式αijxixj累加后写入到寄存器中，获得密文。采用本发明实施例，能够有效提高密钥的安全性，抵御侧行道攻击。

收藏 导出

检索结果共1条 < 1 >

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

一种多变量二次方程的乱序加密装置 重新检索 二次检索

当前检索式：一种多变量二次方程的乱序加密装置/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种多变量二次方程的乱序加密装置 [发明] 有效

申请号：CN201610053768.4 申请日：2016.01.26 公开号：CN105681033A 公开日：2016.06.15
公告号：CN105681033B 公告日：2019.02.05 主分类：H04L9/08
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：李伟键

摘要： 本发明公开了一种多变量二次方程的乱序加密装置，包括：单项式下标生成器，用于在计算每个多变量二次方程时，乱序生成n(n+1)/2个单项式下标值(i,j)；1≤i≤j≤n；单项式乘法器，用于按照单项式下标值(i,j)生成的顺序，依次计算所述多变量二次方程的n个单项式αijxixj；其中，αij为明文，xi为密钥，或者，αij为密钥，xi为明文；以及，累加器，用于将所述n个单项式αijxixj依次累加后写入到寄存器中，获得密文。采用本发明实施例，能够有效提高密钥的安全性，抵御侧行道攻击。

收藏 导出

检索结果共1条 < 1 >

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种多变量二次方程的掩码加密方法 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种多变量二次方程的掩码加密方法/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种多变量二次方程的掩码加密方法 [发明] 有效

申请号：CN201610053795.1 申请日：2016.01.26 公开号：CN105721150A 公开日：2016.06.29
公告号：CN105721150B 公告日：2018.11.20 主分类：H04L9/14
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：李伟键

摘要： 本发明公开了一种多变量二次方程的掩码加密方法，包括：随机生成n个掩码；n≥1；将所述n个掩码与n个密钥一一对应进行异或加密；根据加密后的n个密钥、所述n个掩码和明文，采用掩码算法，计算获得多变量二次方程中n个单项式的加密值；将所述n个单项式的加密值依次累加到寄存器中，获得密文。采用本发明实施例，能够有效提高密钥的安全性，抵御侧行道攻击。

收藏 导出

检索结果共1条 < 1 >

主办单位:国家知识产权局中国专利信息中心 | 技术支持:北京新发智信科技有限责任公司
京ICP备11033758号-7 | 京公网安备 11010802035318号

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种多变量二次方程的掩码加密装置 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种多变量二次方程的掩码加密装置/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种多变量二次方程的掩码加密装置 [发明] 有效

申请号：CN201610053691.0 申请日：2016.01.26 公开号：CN105743644A 公开日：2016.07.06
公告号：CN105743644B 公告日：2019.02.05 主分类：H04L9/14
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：李伟键

摘要： 本发明公开了一种多变量二次方程的掩码加密装置，包括：第一掩码生成器，用于随机生成n个掩码；n≥1；第一异或计算器，用于将所述n个掩码与n个密钥一一对应进行异或加密；第一掩码型乘法器，用于根据加密后的n个密钥、所述n个掩码和明文，采用掩码算法，计算获得多变量二次方程中n个单项式的加密值；以及，第一累加器，用于将所述n个单项式的加密值依次累加到寄存器中，获得密文。采用本发明实施例，能够有效提高密钥的安全性，抵御侧行道攻击。

收藏 导出

检索结果共1条 < 1 >

主办单位:国家知识产权局中国专利信息中心 | 技术支持:北京新发智信科技有限责任公司
京ICP备11033758号-7 | 京公网安备 11010802035318号

一种基于目标联合外观模型的视觉跟踪方法

摘要

本发明公开了一种基于目标联合外观模型的视觉跟踪方法，包括如下步骤：构建目标局部外观表示模型，基于超像素中层特征，根据颜色、亮度及纹理对目标进行合理有效的局部区域划分；根据显著性计算超像素重要性图，估计每个粒子在局部外观下的置信度；对目标整体外观进行粒子采样，获取目标整体外观下的判别似然估计值；利用概率滤波，从目标的稀疏整体外观角度获取稀疏似然估计图；对目标局部外观下的置信度和目标整体外观下的稀疏似然估计图进行线性加权，得到目标状态的最优估计，确定最佳目标跟踪位置。通过本发明实施例兼顾了目标的底层特征信息和中层

著录项

公开/公告号	CN106157330B	专利类型	发明专利
公开/公告日	2020.03.17	原文格式	PDF
申请/专利权人	广东技术师范学院;	获取专利	CN201610519784.8
发明设计人	詹瑾;唐晓辛;	申请日	2020.03.17

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

请输入关键词 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种基于实例分割的监控视频多目标人脸检测方法及系统/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种基于实例分割的监控视频多目标人脸检测方法及系统 [发明] 有效

申请号：CN201811504897.6 申请日：2018.12.10 公开号：CN109670429A 公开日：2019.04.23

公告号：CN109670429B 公告日：2021.03.19 主分类：G06K9/00

申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：林凯瀚 赵慧民 吕巨建 詹瑾 陈荣军

摘要： 本发明公开了一种基于实例分割的监控视频多目标人脸检测方法及系统，其中，所述方法包括：采集待训练的人脸图像数据集，基于标注工具对人脸图像数据集进行标注，获取训练数据集；基于监控设备条件设置用于训练人脸检测系统的模型参数；基于标记后的人脸图像数据集和模型参数对所述人脸检测系统进行训练，获取训练后的人脸检测系统；将监控视频中的每一帧图像输入训练后的人脸检测系统中，对输入的每一帧图像进行多目标人脸检测，获取检测结果。在本发明实施例中，提高了对人脸目标图像的识别准确率。

收藏 导出

检索结果共1条 < 1 >

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种基于稀疏判别学习的目标跟踪方法

重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种基于稀疏判别学习的目标跟踪方法/YY

检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种基于稀疏判别学习的目标跟踪方法 [发明] 有效

申请号：CN201610519851.6 申请日：2016.07.01 公开号：CN106203495A 公开日：2016.12.07

公告号：CN106203495B 公告日：2020.03.17 主分类：G06K9/62

申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：詹瑾 肖政宏

摘要： 本发明公开了一种基于稀疏判别学习的目标跟踪方法，包括：对第一帧采样目标及其周围背景，构建目标外观模型；提取目标外观模型的二维图像特征并进行归一化处理，获得初始字典；引入有监督的判别字典学习方法，在重构误差项的基础上增加分类误差项，训练出判别字典；在稀疏度约束下求解最小重构误差项，采用迭代交换优化策略更新字典和稀疏编码；采用欧氏距离来度量样本之间的相似性，将相似性最高的样本作为跟踪目标。通过本发明实施例，在建立目标外观模型时，加入了与目标有空间相关性的周围背景作为线索模板，处理目标姿态变化比较鲁棒。

收藏 导出

检索结果共1条 < 1 >

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种数字图像特征感知的通用隐写分析方法 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种数字图像特征感知的通用隐写分析方法/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种数字图像特征感知的通用隐写分析方法 [发明] 有效

申请号：CN201610508701.5 申请日：2016.06.30 公开号：CN106157232A 公开日：2016.11.23
公告号：CN106157232B 公告日：2019.04.26 主分类：G06T1/00
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：赵慧民 戴青云 魏文国 任金昌
摘要： 本发明的目的在于提供一种数字图像特征感知的通用隐写分析方法，包括如下内容：(1)在数据采样层，对于含密图像的多维信号，通过CS实现图像局部数据和全局数据，并用层次化进程求解稀疏优化问题；(2)在特征感知层，对于含密图像数据，通过CS结构性测量矩阵实现纹理特征、边缘特征和平滑特征的数据提取，并实现特征之间、特征内部的冗余性与互补性的统计关系；(3)设计SVM分类器、隐写分析CS盲检测器的过程。本发明通过对目前的基于Markov模型的多域特征通用隐写分析的现状进行分析，解决JPEG图像在不同嵌入率下秘密信息的识别精度问题。

收藏 导出

检索结果共1条 < 1 >

主办单位:国家知识产权局中国专利信息中心 | 技术支持:北京新发智信科技有限责任公司
京ICP备11033758号-7 | 京公网安备 11010802035318号

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

一种通用隐写分析的深度学习堆栈式自动编码方法

重新检索 二次检索

当前检索式：一种通用隐写分析的深度学习堆栈式自动编码方法/YY

检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种通用隐写分析的深度学习堆栈式自动编码方法 [发明] 有效

申请号：CN201610642550.2 申请日：2016.08.03 公开号：CN106251375A 公开日：2016.12.21

公告号：CN106251375B 公告日：2020.04.07 主分类：G06T9/00

申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：赵慧民 戴青云 魏文国 任金昌 容淳铭

摘要： 本发明涉及一种通用隐写分析的深度学习堆栈式自动编码方法，其特征在于包括以下步骤：第一步，图像分块处理；第二步，USoSAE特征数据输入，第三步，USoSAE特征数据的映射，第四，USoSAE特征数据输出，第五，训练最后的优化参数值。本发明通过对隐写检测图像的概率密度函数统计，把图像数据按照方差统计进行分割，并结合SAEs的数据模型，降低输入层的计算开销。本发明通过把不同属性的数据输入到不同的隐含层进行特征提取和训练，改善多维数据摘要的表示效率。在USoSAE隐含层，本发明通过加入稀疏性限制的先验条件，进一步降低特征数据的维数，提高隐写分析数据分类的计算精度。

收藏 导出

检索结果共1条

< 1 >

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种图像内容感知的鲁棒通用隐写分析方法 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种图像内容感知的鲁棒通用隐写分析方法/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种图像内容感知的鲁棒通用隐写分析方法 [发明] 有效

申请号：CN201610567405.2 申请日：2016.07.15 公开号：CN106228505A 公开日：2016.12.14
公告号：CN106228505B 公告日：2019.06.18 主分类：G06T1/00
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：赵慧民 戴青云 魏文国 任金昌

摘要： 本发明的目的在于提供一种图像内容感知的鲁棒通用隐写分析方法，本发明的方法能够更好地利用CS技术处理数字图像的优点,进一步提高隐写分析的数据处理效率和分类判断精度。具体包括如下步骤：(1)设定数字图像X的大小N，分块大小B和测量值大小M；(2)确定9/7-HaarDWT稀疏矩阵Ψ；(3)确定BCS的哈达玛测量矩阵ΦB；(4)把数字图像分成B×B块，并根据哈达玛测量矩阵ΦB；计算第i个图像块的测量向量yi和整个图像X的测量数据Y；(5)把图像分为平滑部分和非平滑部分作为感知结果；(6)训练过程；(7)分类过程；(8)判断、分析过程；(9)重复迭代搜索，直到检测完整个图像数据。

收藏 导出

检索结果共1条 < 1 >

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

一种图像隐写的方法及装置 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：一种图像隐写的方法及装置/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

一种图像隐写的方法及装置 [发明] 有效

申请号：CN201810082685.7 申请日：2018.01.26 公开号：CN108307086A 公开日：2018.07.20

公告号：CN108307086B 公告日：2019.05.07 主分类：H04N1/32

申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：赵慧民 谢国楳 吕巨建 傅仁轩

摘要： 本发明公开了一种图像隐写的方法及装置，涉及图像处理技术领域，用以解决现有技术中存在嵌入容量比较小，且隐蔽性比较弱的问题。该方法包括：确定输入的待隐秘图像的稀疏化矩阵，采用测量矩阵对所述稀疏化矩阵进行随机测量，得到所述待隐秘图像的测量向量以及所述测量向量的二进制长度；将所述测量向量生成长度为第一秘钥的随机序列，并确定所述随机序列的列行量；根据所述随机序列的元素和设定阈值的关系，确定所述随机序列在待载入图像的嵌入位置，并将所述测量向量嵌入到确定的所述嵌入位置中，得到载入图像。

收藏 导出

检索结果共1条 < 1 >

主办单位:国家知识产权局中国专利信息中心 | 技术支持:北京新发智信科技有限责任公司
京ICP备11033758号-7 | 京公网安备 11010802035318号

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

云计算CCS细粒度数据控制方法 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：云计算CCS细粒度数据控制方法/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

云计算CCS细粒度数据控制方法 [发明] 有效

申请号：CN201610294486.3 申请日：2016.04.26 公开号：CN105827632A 公开日：2016.08.03
公告号：CN105827632B 公告日：2019.03.26 主分类：H04L29/06
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：赵慧民 戴青云 魏文国 蔡君 雷方元 罗建桢

摘要： 本发明公开一种云计算CCS细粒度数据控制方法，包括如下步骤：采用压缩感知技术，在得到初始密钥之后，利用及N级移位寄存器及反馈函数计算公式产生密钥；对产生的密钥建立索引序列，并根据该索引序列得到编码矩阵；当云计算平台在噪声和攻击条件下，根据N级移位寄存器、编码矩阵及密钥实现云计算数据访问控制过程；所述N大于等于1。本发明的一种云计算CCS细粒度数据控制方法的安全性高。

收藏 导出

检索结果共1条 < 1 >

主办单位:国家知识产权局中国专利信息中心 | 技术支持:北京新发智信科技有限责任公司
京ICP备11033758号-7 | 京公网安备 11010802035318号

UHF型RFID的通信记录分析及RFID模拟实现方法

摘要

本发明公开了一种UHF型RFID的通信记录分析及RFID模拟实现方法，其基于通信记录分析系统实施，该通信记录分析系统包括通信侦听模拟装置，通信侦听模拟装置包括天线、应用终端发射信号接收模块、RFID返回信号接收模块、RFID模拟信号发射模块、CPU控制模块、通信模块和存储模块；CPU控制模块能够受控将通信侦听模拟装置设置为记录模式，对UHF型RFID与RFID应用终端之间的交互数据进行侦听；CPU控制模块能够受控将通信侦听模拟装置设置为模拟模式，对UHF型RFID进行功能模拟；计算机能够对接收到的数据对进行分析处理。本发明能够侦听UHF型RFID在使用

著录项

公开/公告号	CN104901748B	专利类型	发明专利
公开/公告日	2018.03.16	原文格式	PDF
申请/专利权人	广东技术师范学院;		CN201510184766.4
发明设计人	杨振野;赵慧民;吴达义;梁鹏鸿;魏	申请日	2018.03.16

专利之星检索系统 测试版

智能检索 表格检索 专家检索 号单检索 分类检索 huyuqin...

筛选条件

多载波调制的信息隐藏通信方法及其装置 重新检索 二次检索

专利类型

- 全部
- 发明 1
- 实用新型 0
- 外观 0

法律状态

- 全部
- 有效 1
- 审中 0
- 失效 0

筛选

当前检索式：多载波调制的信息隐藏通信方法及其装置/YY 检索结果共1条 每页 5 条

全选 批量导出 统计分析 申请日↓

多载波调制的信息隐藏通信方法及其装置 [发明] 有效

申请号：CN201410301104.6 申请日：2014.06.23 公开号：CN104270240A 公开日：2015.01.07
公告号：CN104270240B 公告日：2019.04.02 主分类：H04L9/00
申请人：广东技术师范学院 当前权利人：广东技术师范大学 发明人：赵慧民 蔡君 魏文国

摘要： 本发明涉及一种多载波调制的信息隐藏通信方法及其装置，所述方法包括：获取信源信号、信息隐藏通信的密钥及隐藏数据，对信源信号统计处理形成随机离散的子信源信号；根据随机离散的子信源信号及密钥生成隐藏通信的边信息；根据子信源信号的方差间隔将传输信道统计处理成若干个并行的传输子信道；将各子信源信号根据统计条件分别分配在相应的传输子信道上；并计算混合数据；将混合数据进行编码处理得到信道编码信号进行传输。在接收端，结合信道编码信号、边信息及密钥对隐藏数据提取操作实现该隐藏数据的保密通信。本发明的多载波调制的信息隐藏通信方法及其装置能够保证在隐藏数据容量较大的情况下，系统传输的不可见性和鲁棒性效果都较好。

收藏 导出

检索结果共1条 < 1 >

荣誉证书

广东技术师范学院

何光煜 陈浩男 余镇霖 罗晓东 同学：

你（们）的作品《广州市半山文化传播有限责任公司》荣获2018年“挑战杯·创青春”广东大学生创业大赛

金奖

特发此证，以资鼓励。

指导老师：黄明睿

共青团广东省委员会　广东省教育厅　广东省科学技术厅　广东省人力资源和社会保障厅　广东省科学技术协会　广东省学生联合会

二〇一八年五月

"挑行杯"第六届 中国国际"互联网+"
大学生创新创业大赛广东省分赛

荣誉证书

王伟永、李雪如、黄雯愉、冯泽婵、谢庭韦、刘淑娟、陈嘉麟、麦梓洋、林泽宇、覃嘉茵 同学：

你（们）的项目 苹苹回首——用技术换优果，用电商拓销路，携农民奔小康 在第六届中国国际"互联网+"大学生创新创业大赛广东省分赛"青年红色筑梦之旅"赛道决赛中表现优异，荣获

银奖

指导老师：梁立基,陈晓桁,杨燕佳,彭炜锋

特发此证，以资鼓励。

广东省教育厅

2020年8月

以苹扶贫——“用技术换优果、用电商拓销路、携农民奔小康”

	项目名称	以苹扶贫——“用技术换优果、用电商拓销路、携农民奔小康”
	项目类型	创新训练项目
	学科类别	工学——计算机类
	项目编号	202010588028X

项目成员：

姓 名	年 级	专 业	是否主持人
林泳	18级	计算机科学与技术师范	第一主持人
冯泽婵	17级	通信工程	否
吴嘉玲	17级	通信工程	否
李雪如	18级	财务管理	否

指导老师：

姓 名	职 称	研究方向
彭炜锋	讲师	创新创业教育，创业孵化
陈晓桁	讲师	创新创业教育，企业管理
林兴初	无	深圳市赛诺尔包装印刷材料有限公司 职务财务经理

获奖证书

16th 挑战杯

广东技术师范大学

黄 娴 毕远桥 张云琛 林 泳 林思瀚 李龙明 同学：

你（们）的作品《密码医生—密码算法的侧信道泄露智能测评系统》荣获第十六届“挑战杯”广东大学生课外学术科技作品竞赛

一等奖

指导老师：李伟健 谢桂园

特颁此证，以兹鼓励。

共青团广东省委员会 广东省教育厅 广东省科学技术厅 广东省科学技术协会 广东省社会科学院 广东省学生联合会

二〇二一年七月

获奖证书

15th 挑战杯 广东大学生课外学术科技作品竞赛 挑战30载

广东技术师范大学：

何娃、邵绮玲、李豪　同学：

你（们）的项目《基于物联网及大数据分析的智能机器人家庭监护预警系统》在第十五届“挑战杯”广东大学生课外学术科技作品竞赛中荣获 一等奖 。

指导教师：郝刚、梁鹏

特颁此证，以兹鼓励。

共青团广东省委员会　广东省教育厅　广东省科学技术厅　广东省科学技术协会　广东省学生联合会　佛山市人民政府

2019年5月

获奖证书

15th 挑战杯 广东大学生课外学术科技作品竞赛 挑战30载

广东技术师范大学：

陈昌海、陈森鹏、刘欢进、林晓鹏、谢毓　同学：

你（们）的项目《一种基于深度学习的房角图像智能分级系统》在第十五届“挑战杯”广东大学生课外学术科技作品竞赛中荣获 一等奖 。

指导教师：贾西平、刘少鹏、梁立基

特颁此证，以兹鼓励。

共青团广东省委员会　广东省教育厅　广东省科学技术厅　广东省科学技术协会　广东省学生联合会　佛山市人民政府

2019年5月

获奖证书

15th 挑战杯 广东大学生课外学术科技作品竞赛 挑战30载

广东技术师范大学：

吴玉婷、陈荣杰　同学：

你（们）的项目《基于视觉技术的智能卡高速生产控制系统》在第十五届“挑战杯”广东大学生课外学术科技作品竞赛中荣获 二等奖 。

指导教师：郝刚、梁鹏

特颁此证，以兹鼓励。

共青团广东省委员会　广东省教育厅　广东省科学技术厅　广东省科学技术协会　广东省学生联合会　佛山市人民政府

2019年5月

获奖证书

广东技术师范大学

林　泳、王伟永、冯泽婵、李雪如、吴嘉玲、刘淑娟、麦梓洋 同学：

你（们）的作品《羊羊回首——静宁县康源青果品农民专业合作社》荣获第十二届“挑战杯”广东大学生创业大赛创业实践挑战赛

银奖

特发此证，以资鼓励。

指导老师：陈晓桁、梁立基、吴凤池

共青团广东省委员会　　广东省教育厅　　广东省科学技术厅　　广东省科学技术协会　　广东省学生联合会

二〇二〇年七月

共青团广东省委员会

关于2021年广东省科技创新战略专项资金（“攀登计划”专项资金）拟资助立项项目的公示

根据《广东省科技创新战略专项资金（大学生科技创新培育）管理办法》有关要求，按照2021年度“攀登计划”立项工作安排，经校省逐级审核评选，拟确定中山大学《融合知识图谱的垂直领域智能搜索、推理及可视化平台》等1050个项目拟立项项目（详见附件1）。现对拟立项项目进行为期五天的公示，公示为1月29日—2月2日。

各高校应核对公示中本校拟立项项目的申报信息，若相关信息存在错误，请根据要求填写拟立项项目信息勘误表并提供相关证明。其中，指导老师与项目作者的人员信息、人员数量和顺序不作调整。请校团委将电子版文件（word版和PDF签名扫描版）汇总后于2月2日17:30前报送至学校部邮箱（邮件命名：学校名称+“攀登计划”项目信息更正）；纸质版文件加盖校团委公章后，于3月10日前寄送至团省委学校部。

如对公示内容有异议，请在公示期内以书面形式向团省委反映，并提供相关证明材料。以个人名义反馈情况的，请提供真实姓名、联系方式；以单位名义反映情况的，请提供

单位名称（加盖公章）、联系人、联系方式。公示期内未报送勘误信息或异议的，视为对公示内容无异议。

附件：1.2021年广东省科技创新战略专项资金（“攀登计划”专项资金）拟立项项目

2.拟立项项目信息勘误表

联 系 人：罗　珂、王靖玮

联系方式：020-87185614

工作邮箱：tsw_xxb@gd.gov.cn

联系地址：广州市越秀区寺贝通津一号大院

共青团广东省委办公室

2021年1月29日

作品编号	学校	大类	小类	评审结果	资助金额（单位：万）	项目名称	主要完成人	指导老师	项目成员
pdjh2021b0280	广州体育学院	哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	乡村振兴战略背景下乡村全民健身公共服务体系构建现状调查——以广州为例	向艳梅	周结友 阳家鹏	钱春艳 李洁贤 陈 亮 李逸飞 田凤琴
pdjh2021b0281	广州体育学院	哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	后疫情时代体育网红对大学生健康健美行为影响研究	赵婷婷	武学军	李梦杰 鞠家诚 王东玲 刘 鹏 左宏伟
pdjh2021b0282	广州体育学院	哲学社会科学类社会调查报告和学术论文	教育	一般项目	1	建设体育强国背景下基于可穿戴运动设备数据的个性化训练周期报告研究	姚鸿彬	罗 智	赖培鑫 邢敏敏 黄一忻 赵凡醇 王展耀 杨 珂
pdjh2021b0283	广州体育学院	哲学社会科学类社会调查报告和学术论文	教育	一般项目	1	重复经颅磁刺激联合运动训练对提高自闭症谱系障碍儿童基本运动技能的干预研究	张舒昊	杜熙茹 黄俊豪	吴佳睿 王雪媛 邓伟基 麦友年 陈 琅 李兰妹 李梓晴 汲嫘萌 刘姚玲
pdjh2021a0284	广东技术师范大学	科技发明制作类	机械与控制	重点项目	6	风力发电机组大尺寸法兰平面度智能装备研发	刘普京	肖苏华 陈郁芬	卢琦文 罗嘉瑞 黄嘉帆 戴智彬
pdjh2021a0285	广东技术师范大学	科技发明制作类	机械与控制	重点项目	6	可分离式声控智能轮椅床垫	苏 蒲	刘大维 陈郁芬	高 敏 刘裕涛 黄泽聪 刘铭杰 朱锐梁 陈文杰 汤国林 张金祥 黄志萍
pdjh2021a0286	广东技术师范大学	科技发明制作类	信息技术	重点项目	6	IC卡芯片封装及信息安全写入关键技术研究	吴玉婷	梁 鹏 李伟键	齐建阳 吴启超 沈响响 秦嘉欣 董鹏龙 杨子煌 吴光鸿 马启键

作品编号	学校	大类	小类	评审结果	资助金额（单位：万）	项目名称	主要完成人	指导老师	项目成员
pdjh2021a0287	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	社会	重点项目	3	构建佛山“人”城市形象品牌——关于佛山形象推广及其品牌战略建设的调查报告	陈思茵	陈南先	黎欣宜 廖 晴 利 斌 周浩铖 郑芷欣
pdjh2021a0288	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	教育	重点项目	3	粤港澳大湾区高等职业教育的布局结构研究	江雪儿	陶 红	高思茵 黄文韬
pdjh2021b0289	广东技术师范大学	科技发明制作类	机械与控制	一般项目	2	面向“机电热”多参数表达的智能电池管理系统	罗郁金	许 铀 吴彦潮	罗金伦 邹贺伟 揭光景 张紫妮 刘秋香 姚锦祥 魏美清
pdjh2021b0290	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	基于物联网的智能机器人家庭看护系统	齐建阳	郝 刚 梁 鹏	彭琛琦 吴玉婷 董丽荣 张玉龙 郭一帆 黎杨威 贺 韬
pdjh2021b0291	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	面向无接触操作的电梯交互系统设计与应用	许熙熙	周 卫 马锐军 庄鑫财	丁 韩 苏 乐 陈文霞 李蔚敏 蔡梓漫
pdjh2021b0292	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	渔水新生态：集装箱养殖水质在线监测与智能调控集成方案	蓝 坚	余剑生 韩 克	陈映欣 莫明璋 梁梓蕾 黄诗琳 颜茹花 陈浩文 方辰宇 张凤珍
pdjh2021b0293	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	基于 AIoT 技术的智能椅及健康分析平台	黄 岳	陈荣军 赵慧民	张新地 黄浩民 罗小梅 于永兴 李琦桥 蓝凯霖 魏志勇 冯心瑜
pdjh2021b0294	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	家居照明与人脸识别交互融合技术的研究	张嘉玲	陈湛旭 万 巍 许鹏程	陈智辉 蔡怡冰 钱程亮 林华康 曹雨亭 温志辉 袁双立 许文缤 张颖馨
pdjh2021b0295	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	基于 STM32 的智能助盲穿戴设备	朱锐梁	郑振兴 刘大维	卢咏珊 刘裕涛 吴思进 张金祥 吴紫泰 高 敏 黄泽聪 苏 蒲 余明晖

共青团广东省委员会

关于2020年广东省科技创新战略专项资金（“攀登计划”专项资金）拟资助立项项目的公示

根据《广东省科技创新战略专项资金（大学生科技创新培育）管理办法》有关要求，按照2020年度“攀登计划”立项工作安排，经过校、省两级审核，拟确定中山大学《基于眼底图像的糖尿病视网膜病变病灶智能识别和分类系统》等1450个项目为拟立项项目（详见附件1）。现对拟立项项目进行为期七天的公示，公示期为2月18日—2月24日。

各高校需核对本校拟立项项目的作品名称、作者姓名、指导老师姓名（所有信息以网上申报平台的作品信息为准），指导老师及作者人员、数量不得调整，只作校对。若相关信息存在错误，请根据要求填写拟立项项目信息勘误表（附件2）并提供相关证明。纸质版文件需由校团委主要负责同志或团委相关工作负责老师签名并寄送至团省委学校部。电子版文件（word版和PDF签名扫描版）请校团委汇总并于2月24日17:30前报送至学校部邮箱（邮件命名：学校名称+“攀登计划”项目信息更正）。

如对公示内容有异议，请在公示期内以书面形式向团省委反映，并提供相关证明材料。以个人名义反馈情况的，请

提供真实姓名、联系方式；以单位名义反映情况的，请提供单位名称（加盖公章）、联系人、联系方式。公示期内未报送勘误信息或异议的，视为对公示内容无异议。

附件：1.2020年广东省科技创新战略专项资金（“攀登计划”专项资金）拟立项项目

2.拟立项项目信息勘误表

团省委学校部

联 系 人：罗　珂、徐婧滢

联系方式：020-87185614

工作邮箱：tswkjcx@163.com

联系地址：广州市越秀区寺贝通津一号大院

共青团广东省委办公室

2020年2月18日

项目编号	学校	大类	小类	评审结果	资助金额（单位：万）	项目名称	主要完成人	指导老师	项目成员
a0332	大学							林　军	陈岱杰
pdjh2020a0333	广东技术师范大学	自然科学类学术论文	信息技术	重点项目	4.5	面向石化机组故障诊断的相关系数与深度学习理论智能复合故障诊断方法与技术攻关	王水泽	熊建斌 岑　健	陈志翔 林景涛 李俊煌 郭碧华 刘家林
pdjh2020a0334	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	社会	重点项目	3	新媒体时代涉民族因素的网络舆情研究	张小霞	李尚旗	何　英 林月坚 胡美玲 朱林风 慎效磊
pdjh2020a0335	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	法律	重点项目	3	行政还是自治：乡村振兴背景下村级治理向何处去？---广东省村民自治制度实施现状调查	张杰锋	沈理平 毛学群	王凯文 伍泳欣 林　涵 麦贝翘 戴琳涵 李岚琦 周瑾瑭
pdjh2020b0336	广东技术师范大学	科技发明制作类	机械与控制	一般项目	2	基于可调频式的磁性阻尼动力吸振器的设计	刘倩君	曾祥坤 秦太兴	卓娟铃 彭兴武 刘仕辉 陈宇杰 蔡嘉欣
pdjh2020b0337	广东技术师范大学	科技发明制作类	机械与控制	一般项目	2	基于风光互补发电系统的隧道节能装置	李英汉	刘大维	谭秀静 罗泳棋 林吴照 刘欣莉 刘铭杰 陈浩民 陈文杰 朱雯静 高　敏
pdjh2020b0338	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	基于BIM和NB-IoT的智能配电房运维系统	谢泰荣	刘　军 刘克江	黄　娴 张俊源 邓广安
pdjh2020b0339	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	一种基于YOLO-v3与CE-Net的角膜-虹膜分割系统研究	张杰东	廖秀秀 贾西平	张杰东 陈晓静 林煜皓 刘海宏
pdjh2020b0340	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	基于“区块链+物联网”的农产品溯源系统	冯泽婵	彭炜锋 梁立基	王伟永
pdjh2020b0341	广东技术师范大学	科技发明制作类	信息技术	一般项目	2	基于药品成分冲突的安全购药服务机器人开发	林煜森	戴青云 李　亚	曾　曦 庄楚鑫 易思宇 刘宏宇 袁　涛 庄楚源 蔡璟雯 侯简清 连楷滨
pdjh2020b0342	广东技术师范大学	科技发明制作类	数理	一般项目	2	周期与随机纳米形貌对LED光提取效率影响的研究	丁润宏	万　巍 陈湛旭	张建华 曾子轩 黄雄霆 赵马楠 林婉琪 吕晓晓 蔡培鹏 吴凌峰
pdjh2020	广东技术师范	自然科学类学术	信息技术	一般项目	1.5	基于非线性多模干涉的同步被动锁模双色光纤激光器	郑婉莹	李　艳	唐　诗 兰天野 李贤惠

项目编号	学校	大类	小类	评审结果	资助金额（单位：万）	项目名称	主要完成人	指导老师	项目成员
b0343	大学	论文						马秋月	刘付武兴 缪伟业 谢芷璇
pdjh2020b0344	广东技术师范大学	自然科学类学术论文	信息技术	一般项目	1.5	基于人工电磁结构的透明超宽带微波吸波结构设计	黄丽欣	赖森锋	张怡聪 张少添 卢沛锋
pdjh2020b0345	广东技术师范大学	自然科学类学术论文	信息技术	一般项目	1.5	面向智慧农业的资源环境精准监测与智能决策平台研究	刘兆珂	熊建斌 岑 健	周振辉 林佳超 林泽山 黄子书
pdjh2020b0346	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	乡村振兴计划之农电商发展现状及机制研究	钟婉茹	廖卢琴 袁裕辉	陈丹妮 杨媛君 刘 敏 黄晓琳 赵晓宜 罗永明
pdjh2020b0347	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	互联网+时代下广东乡镇地区儿童预防接种工作进一步发展的机遇与挑战	覃嘉茵	杨燕佳 王迎吉	方嘉锐 李 涛 孔嘉欣 汤宝蓝 吴超琰 周颖琪
pdjh2020b0348	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	教育	一般项目	1	新时代强化基层党建背景下大学生政治意识培育研究	黄泽孙	黄文浩	林镇宏 杨来发
pdjh2020b0349	广东技术师范大学	哲学社会科学类社会调查报告和学术论文	教育	一般项目	1	文化自信视域下广东高校传承传统手工技艺与创造性转化研究	钱天一	张璐璐 曾丽娟	吴雨彤 李各力 龙雨馨 莫 然 张渝佩 陈康劲 谢茗锋 崔鹏翔
pdjh2020a0350	岭南师范学院	科技发明制作类	信息技术	重点项目	6	基于物联网的养老社区健康数据采集跟踪监管平台	陈心辉	曾绍庚 吴 涛	吴紫晴 赵芳俪 何婉晴 陈冠宏 林汪文 林培峰 李科德 林子杰 陈延寿
pdjh2020a0351	岭南师范学院	自然科学类学术论文	生命科学	重点项目	4.5	槟榔碱改善创伤后应激障碍（PTSD）的生理机制研究	邓梅萍	张长征 周佩玲	叶晓雯 邬嘉善 陈丽娟 江晓霞 符江柳 廖 格 钟秋琳
pdjh2020a0352	岭南师范学院	哲学社会科学类社会调查报告和学术论文	哲学	重点项目	3	出门一把锁，进门一盏灯——留守老人生命故事的访谈研究	刘吉婕	郑剑虹	郑嘉汉 钟峰强 李 宇 梁惠琳 黄广诚 刘红梅 朱晓曼 赵博也 黄 莉

共青团广东省委员会

关于对2019年广东大学生科技创新培育专项资金（“攀登计划”专项资金）拟资助立项项目公示的通知

各高校团委：

根据2019年广东大学生科技创新培育专项资金（“攀登计划”专项资金）立项工作安排，经过校、省两级审核，拟确定中山大学《肿瘤微环境敏感型IFNγ/PD-L1抗体靶向递送技术增强RFA术后残癌治疗效果》等1000个项目为拟立项项目（详见附件1）。现对拟立项项目进行为期七天的公示，公示期为1月10日-1月16日。

各高校需核对本校拟立项项目的作品名称、作者姓名、指导老师姓名（所有信息以网上申报平台的作品信息为据），指导老师及作者人员、数量不得调整，只作校对。若相关信息存在错误，请根据要求填写拟立项项目信息勘误表（附件2）并提供相关证明，电子版文件（word版和PDF盖章扫描版）请于1月16日17：00前报送至学校部邮箱（邮件名：“学校名称+‘攀登计划’项目信息更正”），并同时寄送纸质版至学校部。

如对公示内容有异议，请在公示期内以书面形式向团省

委反映，并提供相关证明材料。以个人名义反馈情况的，请提供真实姓名、联系方式；以单位名义反映情况的，请提供单位名称（加盖公章）、联系人、联系方式。公示期内未报送勘误信息或异议的，视为对公示内容无异议。

附件：1.2019年广东大学生科技创新培育专项资金（“攀登计划”专项资金）拟立项项目

2.拟立项项目信息勘误表

联 系 人：罗珂、唐显锋、梁铭升

联系方式：020-87185614

工作邮箱：tswkjcx@163.com

联系地址：广州市越秀区寺贝通津一号大院团省委学校部

共青团广东省委办公室

2019年11月9日

项目编号	学校	大类	小类	拟立项等级	拟资助金额（单位：万）	作品名称	主要完成人	指导老师	项目成员
pdjh2019b0284		科技发明制作类	信息技术	一般项目	2	一种基于深度学习的房角图像智能分级系统研究	陈昌海	贾西平 刘少鹏	陈昌海　陈桂君　陈森鹏 林晓鹏　谢　毓　谢文嵘
pdjh2019b0285		科技发明制作类	信息技术	一般项目	2	基于车流量动态配时的智能交通灯	严予甫	杜灿谊 吴　劲	严予甫　张晶晶　陈志杰 何召庭　冯文涛　孙逸钿 何婷玉　戴中铭　谭晓婷
pdjh2019b0286		科技发明制作类	信息技术	一般项目	2	基于云平台和机器视觉的自动识别跟踪巡逻无人机	梁金威	袁　飞 梁　东	梁金威　陈伦澍　张　芸 梁浩潮　何焯正　赖鸿辉
pdjh2019b0287		科技发明制作类	信息技术	一般项目	2	基于物联网与人工智能技术的智能宠物管理系统研究与实现	于永兴	陈荣军 赵慧民	于永兴　李绮桥　徐献圣 杨　敏　黄智慧　梁柏强 陈伟坚
pdjh2019b0288		自然科学类学术论文	信息技术	一般项目	1.5	智能家居 APP 花管+的研究与设计	梁妙瑜	王竹君	梁妙瑜　林玉妙　黄依桐 沈晓莹　卢俊妍
pdjh2019b0289		哲学社会科学类社会调查报告和学术论文	经济	一般项目	1	广东省新兴产业投融资现状调研与对策研究 ——以中小型企业为例	杨芷瑗	袁裕辉 李　丹	杨芷瑗　陈漫珊　李瑜滨 苏映静　刘晓婉　郑邦欣 王　馨
pdjh2019b0290		哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	水源保护区统筹型生态补偿机制研究——以广东省为例	刘润怡	陈　宁 林　晨	刘润怡　佘可漫　佘宝婵 陈奕桦　黄惠莹　曾金牛
pdjh2019b0291		哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	一带一路背景下在粤中非混血儿童社会化研究	林月坚	冯　润	林月坚　黎坤秀　李秋萍
pdjh2019b0292		哲学社会科学类社会调查报告和学术论文	社会	一般项目	1	新时代高校共青团促进大学生创业的路径优化	杨丽云	陈小花 刘雯璇	杨丽云　黄　怡　林　钰 陈绮琪　段诗晴　吴泽群
pdjh2019b0293	岭南师范学院	科技发明制作类	信息技术	一般项目	2	电饭锅智能开发测试平台的研究与实现	叶津宏	龙世瑜 梁启文	叶津宏　谭淼泉　陈智茹 韦勤彬　罗　珏

关于推荐2020年度国家级、省级大学生创新创业训练计划立项项目的公示

发布日期：2020-06-28　浏览: 1720

各单位：

根据《广东省教育厅关于报送2020年度国家级、省级大学生创新创业训练计划立项项目的通知》要求，经学院申报、专家评审，我校拟推荐《潜龙机器视觉智能系统》等100个项目为2020年度国家级、省级大学生创新创业训练计划立项项目。现予以公示（项目详情见附件）。

公示时间：自本公示发布之日起5个工作日。公示期间，如有异议，请以电话或书面形式向我校大学生创新创业训练计划工作小组（校本部广东工业实训中心402）反映。反映情况时要自报或签署真实姓名，要有具体事实；不报或不签署真实姓名的，以及不提供具体事实材料的，一律不予受理。

联系人：王老师、张老师

联系电话：020-38765279

电子邮箱：gsddachuang@163.com

联系地址：校本部广东工业实训中心402

附件:广东技术师范大学关于推荐2020年度国家级、省级大学生创新创业训练计划立项项目公示

广东技术师范大学大学生创新创业训练计划工作小组

2020年6月28日

附件【广东技术师范大学关于推荐2020年度国家级、省级大学生创新创业训练计划立项项目公示.xls】已下载775次

	A	B	C	D	E	F	G	H	I
1	广东技术师范大学关于推荐2020年度国家级、省级大学生创新创业训练计划立项项目公示								
2	序号	学院	项目名称	项目负责人	项目负责人学号	指导教师	项目类型	项目级别	
27	25	计算机科学学院	扶贫——“用技术换优果、用电商拓销路、携农民奔小	林 泳	2018035144004	梁立基、彭炜锋	创业训练	国家级	
28	26	计算机科学学院	一种用于无人机电力巡检智能缺陷检测的轻量级深度学习模型	林凯墅	2018035043045	刘晓勇	创新训练项目	国家级	
29	27	计算机科学学院	基于NB-IoT的网格化大气环境监测系统	徐开涛	2017035144024	陈荣军、赵慧民	创新训练项目	国家级	
30	28	计算机科学学院	喵~少儿编程	饶瑜芳	2018035043012	杨燕佳	创业训练	国家级	
31	29	计算机科学学院	基于生成式对抗网络的乳腺癌筛查方法	梁健峰	2018034743148	刘少鹏、梁立基	创新训练	国家级	
89	87	计算机科学学院	基于深度学习的青光眼眼底图筛查系统	肖徽森	2018035643031	梁立基、刘少鹏	创业训练	省级	
90	88	计算机科学学院	基于LoRa技术及快速跳频算法的电力应急通信系统	张颖杰	2017034243041	魏纵横、温剑丰、林智勇	创新训练项目	省级	
91	89	计算机科学学院	基于物联网技术的健康智能椅系统	廖金源	2018035743019	陈荣军、赵慧民	创新训练项目	省级	
92	90	计算机科学学院	幼智猿—编程动画教学平台	罗文茵	2018035144013	彭炜锋、梁立基、陈晓桁	创业训练	省级	
93	91	计算机科学学院	基于IPv6的NB-IoT高校实验室资产管理系统	陈 利	2018035144017	魏纵横、温剑丰、林智勇	创新训练	省级	
94	92	计算机科学学院	微信小程序的学生日常事务管理系统——辅导员小	陈嘉濠	2018034243030	张 倩、沈雯婷	创新训练	省级	
95	93	计算机科学学院	基于物联网技术的森林火灾预防系统	李国浩	2019035743041	王磊军、陈荣军	创新训练项目	省级	
96	94	计算机科学学院	基于树莓派的lora网关设计	黄振雄	2019034243048	盘茂杰、魏纵横、温剑丰	创新训练项目	省级	
103									

广东技术师范大学关于推荐2021年度国家级、省级大学生创新创业训练计划立项项目公示

序号	所属学院	项目名称	项目负责人	负责人学号	项目类别	指导教师	项目级别	项目类别	备注
13	计算机科学学院	小快通——守护独居老人安全的智能应急报警器	李雅妍	2019035643021	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁清、王磊、王文俊	国家级	一般项目	
14	计算机科学学院	水产自动喂养及水质监控智能无人船	谭天凌	2019035144036	创新训练项目	詹瑾、李伟健、陈润伟	国家级	一般项目	
15	计算机科学学院	基于LoRa跳频数字电台的电力应急多跳组网通信研究	李昊	2019114243039	创新训练项目	魏纵横、温剑丰、林智勇、彭铁清、尹以鹰、白洋	国家级	重点支持领域项目	直推
16	计算机科学学院	一种基于前房角图像的青光眼分级算法与系统	陈佳洋	2018034243016	创新训练项目	贺西平	国家级	一般项目	直推
59	计算机科学学院	无人机在森林灭火中的应用	赖炳宾	2019035743042	创新训练项目	王磊军、陈荣军、陈昕叶	省级	一般项目	
60	计算机科学学院	基于AIoT的智能垃圾桶系统	魏志勇	2019035743108	创新训练项目	陈荣军、赵慧民、卢旭	省级	一般项目	
61	计算机科学学院	基于计算机视觉的物品分拣系统	曾景睿	2019035043033	创新训练项目	盘茂杰、魏纵横	省级	一般项目	
62	计算机科学学院	基于Wi-Fi和UWB的融合室内定位系统研究	曹志坚	2018034743127	创新训练项目	魏纵横、罗特元、林智勇，彭铁清、孔倩倩、祝捷	省级	一般项目	
63	计算机科学学院	基于STM32的小型四旋翼无人机飞行控制系统算法研究与设计	陈远烜	2019034743030	创新训练项目	魏纵横、林智勇、陈昕叶、彭铁清、孔倩倩、祝捷	省级	一般项目	
64	计算机科学学院	基于树莓派的RFID和4G双模实验室资产管理系统	卢依思	2019035043018	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁清、王磊、王文俊	省级	一般项目	直推
65	计算机科学学院	高校毕业生招聘一体化管理系统	方嘉祺	2018034843115	创新训练项目	杨燕佳	省级	一般项目	直推
66	计算机科学学院	广东高校新疆少数民族大学生适应、融合与认同现状的调查研究	汤宝蓝	2018034843136	创新训练项目	迪娜古丽、杨燕佳	省级	一般项目	直推
		古南粤·新活力——电商乡村振兴战略计划，打造岭南特色							

广东技术师范大学关于推荐2021年度国家级、省级大学生创新创业训练计划立项项目公示

所属学院	项目名称	项目负责	负责人学号	项目类别	指导教师	项目级	项目类别	备注
计算机科学学院	小快通——守护独居老人安全的智能应急报警器	李雅妍	2019035643021	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁涛、王鑫、王文俊	国家级	一般项目	
计算机科学学院	水产自动喂养及水质监控智能无人船	谭天浪	2019035144036	创新训练项目	詹瑾、李伟健、陈润伟	国家级	一般项目	
计算机科学学院	基于LoRa跳频数字电台的电力应急多跳组网通信研究	李昊	2019114243039	创新训练项目	魏纵横、温剑丰、林智勇、彭铁涛、尹以雁、白洋	国家级	重点支持领域项目	直推
计算机科学学院	一种基于前房角图像的青光眼分级算法与系统	陈佳洋	2018034243016	创新训练项目	贾西平	国家级	一般项目	直推
计算机科学学院	无人机在森林灭火中的应用	赖剑宾	2019035743042	创新训练项目	王磊军、陈荣军、陈昕叶	省级	一般项目	
计算机科学学院	基于AIoT的智能垃圾桶系统	魏志勇	2019035743108	创新训练项目	陈荣军、赵慧民、卢旭	省级	一般项目	
计算机科学学院	基于计算机视觉的物品分拣系统	曾景豪	2019035043033	创新训练项目	盘茂杰、魏纵横	省级	一般项目	
计算机科学学院	基于Wi-Fi和UWB的融合室内定位系统研究	曹志坚	2018034743127	创新训练项目	魏纵横、罗梓元、林智勇，彭铁涛、孔倩倩、祝捷	省级	一般项目	
计算机科学学院	基于STM32的小型四旋翼无人机飞行控制系统算法研究与设计	陈远焙	2019034743030	创新训练项目	魏纵横、林智勇、陈昕叶、彭铁涛、孔倩倩、祝捷	省级	一般项目	
计算机科学学院	基于树莓派的RFID和4G双模实验室资产管理系统	卢依思	2019035043018	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁涛、王鑫、王文俊	省级	一般项目	直推
计算机科学学院	高校毕业生招聘一体化管理系统	方嘉锐	2018034843115	创新训练项目	杨燕佳	省级	一般项目	直推
计算机科学学院	广东高校新疆少数民族大学生适应、融合与认同现状的调查研究	汤宝蓝	2018034843136	创新训练项目	迪娜古丽、杨燕佳	省级	一般项目	直推
	古南号·新活力——电商乡村振兴战略计划，打造岭南特色							

广东技术师范大学关于推荐2021年度国家级、省级大学生创新创业训练计划立项项目公示

序号	所属学院	项目名称	项目负责	负责人学号	项目类别	指导教师	项目级	项目类别	备注
13	计算机科学学院	小快通——守护独居老人安全的智能应急报警器	李雅妍	2019035643021	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁涛、王鑫、王文俊	国家级	一般项目	
14	计算机科学学院	水产自动喂养及水质监控智能无人船	谭天浪	2019035144036	创新训练项目	詹瑾、李伟健、陈润伟	国家级	一般项目	
15	计算机科学学院	基于LoRa跳频数字电台的电力应急多跳组网通信研究	李昊	2019114243039	创新训练项目	魏纵横、温剑丰、林智勇、彭铁涛、尹以雁、白洋	国家级	重点支持领域项目	直推
16	计算机科学学院	一种基于前房角图像的青光眼分级算法与系统	陈佳洋	2018034243016	创新训练项目	贾西平	国家级	一般项目	直推
59	计算机科学学院	无人机在森林灭火中的应用	赖剑宾	2019035743042	创新训练项目	王磊军、陈荣军、陈昕叶	省级	一般项目	
60	计算机科学学院	基于AIoT的智能垃圾桶系统	魏志勇	2019035743108	创新训练项目	陈荣军、赵慧民、卢旭	省级	一般项目	
61	计算机科学学院	基于计算机视觉的物品分拣系统	曾景豪	2019035043033	创新训练项目	盘茂杰、魏纵横	省级	一般项目	
62	计算机科学学院	基于Wi-Fi和UWB的融合室内定位系统研究	曹志坚	2018034743127	创新训练项目	魏纵横、罗梓元、林智勇，彭铁涛、孔倩倩、祝捷	省级	一般项目	
63	计算机科学学院	基于STM32的小型四旋翼无人机飞行控制系统算法研究与设计	陈远焙	2019034743030	创新训练项目	魏纵横、林智勇、陈昕叶、彭铁涛、孔倩倩、祝捷	省级	一般项目	
64	计算机科学学院	基于树莓派的RFID和4G双模实验室资产管理系统	卢依思	2019035043018	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁涛、王鑫、王文俊	省级	一般项目	直推
65	计算机科学学院	高校毕业生招聘一体化管理系统	方嘉锐	2018034843115	创新训练项目	杨燕佳	省级	一般项目	直推
66	计算机科学学院	广东高校新疆少数民族大学生适应、融合与认同现状的调查研究	汤宝蓝	2018034843136	创新训练项目	迪娜古丽、杨燕佳	省级	一般项目	直推
		古南号·新活力——电商乡村振兴战略计划，打造岭南特色							

广东技术师范大学关于推荐2021年度国家级、省级大学生创新创业训练计划立项项目公示

序号	所属学院	项目名称	项目负责人	负责人学号	项目类别	指导教师	项目级别	项目类别	备注
13	计算机科学学院	小快通——守护独居老人安全的智能应急报警器	李雅妍	2019035643021	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁清、王磊、王文俊	国家级	一般项目	
14	计算机科学学院	水产自动喂养及水质监控智能无人船	谭天浪	2019035144036	创新训练项目	詹瑾、李伟健、陈润伟	国家级	一般项目	
15	计算机科学学院	基于LoRa跳频数字电台的电力应急多跳组网通信研究	李昊	2019114243039	创新训练项目	魏纵横、温剑丰、林智勇、彭铁清、尹以雁、白洋	国家级	重点支持领域项目	直推
16	计算机科学学院	一种基于前房角图像的青光眼分级算法与系统	陈佳洋	2018034243016	创新训练项目	樊西平	国家级	一般项目	直推
59	计算机科学学院	无人机在森林灭火中的应用	帕尔宾	2019035743042	创新训练项目	王磊军、陈荣军、陈昕叶	省级	一般项目	
60	计算机科学学院	基于AIoT的智能垃圾桶系统	魏志勇	2019035743108	创新训练项目	陈荣军、赵慧民、卢旭	省级	一般项目	
61	计算机科学学院	基于计算机视觉的物品分拣系统	曾景豪	2019035043033	创新训练项目	盘茂杰、魏纵横	省级	一般项目	
62	计算机科学学院	基于Wi-Fi和UWB的融合室内定位系统研究	曹志坚	2018034743127	创新训练项目	魏纵横、罗梓元、林智勇，彭铁清、孔倩倩、祝捷	省级	一般项目	
63	计算机科学学院	基于STM32的小型四旋翼无人机飞行控制系统算法研究与设计	陈远煌	2019034743030	创新训练项目	魏纵横、林智勇、陈昕叶、彭铁清、孔倩倩、祝捷	省级	一般项目	
64	计算机科学学院	基于树莓派的RFID和4G双模实验室资产管理系统	卢依思	2019035043018	创新训练项目	魏纵横、盘茂杰、林智勇、彭铁清、王磊、王文俊	省级	一般项目	直推
65	计算机科学学院	高校毕业生招聘一体化管理系统	方嘉钱	2018034843115	创新训练项目	杨燕佳	省级	一般项目	直推
66	计算机科学学院	广东高校新疆少数民族大学生适应、融合与认同现状的调查研究	汤宝蓝	2018034843136	创新训练项目	迪娜古丽、杨燕佳	省级	一般项目	直推
		古南号·新活力——电商乡村振兴战略计划，打造岭南特色							

广东技术师范大学关于推荐2021年度国家级、省级大学生创新创业训练计划立项项目公示

序号	所属学院	项目名称	项目负责	负责人学号	项目类别	指导教师	项目级	项目类别	备注
13	计算机科学学院	小快通－－守护独居老人安全的智能应急报警器	李雅妍	2019035643021	创新训练项目	魏纵横、盘茂杰、林智勇、彭秋涛、王磊、王文俊	国家级	一般项目	
14	计算机科学学院	水产自动喂养及水质监控智能无人船	谭天浪	2019035144036	创新训练项目	詹瑾、李伟健、陈润伟	国家级	一般项目	
15	计算机科学学院	基于LoRa跳频数字电台的电力应急多跳组网通信研究	李昊	2019114243039	创新训练项目	魏纵横、温剑丰、林智勇、彭秋涛、尹以晨、白洋	国家级	重点支持领域项目	直推
16	计算机科学学院	一种基于前房角图像的青光眼分级算法与系统	陈佳洋	2018034243016	创新训练项目	樊西平	国家级	一般项目	直推
59	计算机科学学院	无人机在森林灭火中的应用	赖剑宾	2019035743042	创新训练项目	王磊军、陈荣军、陈昕叶	省级	一般项目	
60	计算机科学学院	基于AIoT的智能垃圾桶系统	魏志勇	2019035743108	创新训练项目	陈荣军、赵慧民、卢旭	省级	一般项目	
61	计算机科学学院	基于计算机视觉的物品分拣系统	曾景豪	2019035043033	创新训练项目	盘茂杰、魏纵横	省级	一般项目	
62	计算机科学学院	基于Wi-Fi和UWB的融合室内定位系统研究	曹志坚	2018034743127	创新训练项目	魏纵横、罗梓元、林智勇，彭秋涛、孔倩倩、祝锾	省级	一般项目	
63	计算机科学学院	基于STM32的小型四旋翼无人机飞行控制系统算法研究与设计	陈远煊	2019034743030	创新训练项目	魏纵横、林智勇、陈昕叶，彭秋涛、孔倩倩、祝锾	省级	一般项目	
64	计算机科学学院	基于树莓派的RFID和4G双模实验室资产管理系统	卢依里	2019035043018	创新训练项目	魏纵横、盘茂杰、林智勇、彭秋涛、王磊、王文俊	省级	一般项目	直推
65	计算机科学学院	高校毕业生招聘一体化管理系统	方嘉锐	2018034843115	创新训练项目	杨惠佳	省级	一般项目	直推
66	计算机科学学院	广东高校新疆少数民族大学生适应、融合与认同现状的调查研究	汤宝莲	2018034843136	创新训练项目	迪娜古丽·杨惠佳	省级	一般项目	直推
		古南粤·新活力－－电商乡村振兴战略计划，打造岭南特色							

10	美术学院	“Fashion Buffet”工作室自助搭配服装定制系统研究开发	林紫斌	2016074543002	鹿新杰	创业训练	国家级
11	美术学院	基于环保理念的植物染工艺创新与产品开发	黄依蓉	2016074543024	董雪丹	创新训练	国家级
12	教育科学与技术学院	广州农村留守儿童创新型互联网美育教育产品的情感化设计	林玉妙	2016084643026	王竹君	创新训练	国家级
13	教育科学与技术学院	农村经纪人－－做好农村文创产业的推广者	周颖	2016084643033	王孜	创业训练	国家级
14	教育科学与技术学院	基于人工智能的教学设计应用研究	张钰团	2017085144028	吴仕云	创新训练	国家级
15	计算机科学学院	智能机器人家庭看护系统	蓝锦清	2016034743016	郝刚、梁鹏	创新训练	国家级
16	计算机科学学院	基于大数据分析预警的特殊病人监护系统	邵绮玲	2016034843012	郝刚、梁鹏	创新训练	国家级
17	计算机科学学院	基于动态图像的人脸微表情识别	邱致琳	2016034743001	林智勇	创新训练	国家级
18	计算机科学学院	基于物联网与人工智能技术的智能宠物管理系统设计	梁柏强	2016035743047	陈荣军、赵慧民	创新训练	国家级
19	机电学院	文化自信视域下广府文化培养与传承现状分析	马俊杰	2016094243037	陈郁芬	创新训练	国家级
20	机电学院	智能高速镖球机	叶俊志	2016094143009	杨勇	创新训练	国家级
21	机电学院	外骨骼机械辅助手	李盛明	2016094143035	刘大维	创新训练	国家级
22	机电学院	工科”视域下广东省职业教育师范生培养现状调查及发展策略	马俊杰	2016094243037	陈郁芬	创新训练	国家级
23	光电工程学院	实验智能助手	王政华	2016105843037	杨光洲、罗君	创新训练	国家级
24	管理学院	岭南大师文创	罗宇晨	2018024743010	廖丽平	创业训练	国家级
25	电子与信息学院	水产养殖远程监控系统	蔡铁群	2016044543132	韩克、王春安	创新训练	国家级

广东技术师范大学创新创业学院
SCHOOL OF INNOVATION AND ENTREPRENEURSHIP

网站首页 | 学院概况 | 活动竞赛 | 创新创业基地 | 教学科研 | 成果展示 | 党建之窗 | 广师大科创中心 | 对外交流

通知公告

当前位置：网站首页 >> 通知公告 >> 正文

关于我校推荐2019年省国级大学生创新创业训练计划项目的公示

发布时间:2019年05月08日 浏览量: 482

各二级学院:

根据《广东省教育厅关于报送2019年度国家级、省级大学生创新创业训练计划立项项目的通知》（粤教高函[2019]40号）要求，经学院申报、校外专家评审，拟推荐“基于机器视觉的自动识别跟踪无人机”等120个项目为2019年度省国级大学生创新创业训练计划项目，现予以公示。（项目详情见附件）

公示期2019年5月8日—5月14日。公示期间，如有异议或问题反映，可通过电话、来信、邮件等方式向创新创业学院实名反映，反映问题需实事求是。

受理地点：校本部工业中心402办公室
联系电话：38765279
电子邮件：
562206851@qq.com

附件：
广东技术师范大学2019年省国级大学生创新创业训练计划项目申报一览表.xls

61	机电学院	基于python的儿童英语早教游戏	麦世兴	2017094243036	罗永顺	创新训练	省级	
62	机电学院	工业级３Ｄ打印机关键零部件的智能控制及系统集成	李芯怡	2016096143023	建、周莉、郑振	创新训练	省级	
63	计算机科学学院	基于生成式对抗网络的眼底图杯盘分割方法	林晓鹏	2016034743141	刘少鹏	创新训练	省级	
64	计算机科学学院	一种基于深度学习的青光眼辅助分级方法研究与实现	陈昌海	2016034743142	贾西平	创新训练	省级	
65	计算机科学学院	基于深度学习的青光眼辅助诊断应用	胡　娟	2016035643007	鹏、贾西平、曾	创新训练	省级	
66	计算机科学学院	广州道恩光学电子科技有限公司	王伟永	2016034243026	杨燕佳、黄明睿	创业实践	省级	
67	计算机科学学院	大学生创新创业训练项目管理系统	方俊杰	2018035043007	杨燕佳、彭炜煌	创新训练	省级	
68	计算机科学学院	智能公交站牌系统	彭志鑫	2018035743032	王磊军、陈荣军	创新训练	省级	
69	计算机科学学院	基于移动互联网的便携式大学生体质检测系统	李晓莺	2016034243006	卢旭	创新训练	省级	
70	计算机科学学院	关于高校新疆少数民族职教师范生职业认同感的调研	叶汝婧	2018034843130	杨燕佳	创新训练	省级	
71	计算机科学学院	带机器视觉自主行驶的智能快递小车	林佳超	2016035743018	梁琼	创新训练	省级	
72	计算机科学学院	基于物联网技术的智能快递物流系统	马勇枝	2016035743039	吕巨建、陈荣军、赵慧民	创新训练	省级	
73	计算机科学学院	康源青果品农民专业合作社	陈锦涛	2016034243003	彭炜锋、杨燕佳	创业训练	省级	
74	计算机科学学院	裹仓公司	何清莹	2018035244005	徐青云	创业训练	省级	
75	计算机科学学院	智能卡表面质量视觉检测系统	陈荣杰	2016034743051	郝刚、梁鹏	创新训练	省级	
76	教育科学与技术学院	智慧家居APP的情感化体验设计与研究	梁妙瑜	2016084643020	王竹君	创新训练	省级	

广东技术师范大学创新创业学院
SCHOOL OF INNOVATION AND ENTREPRENEURSHIP

网站首页 | 学院概况 | 活动竞赛 | 创新创业基地 | 教学科研 | 成果展示 | 党建之窗 | 广师大科创中心 | 对外交流

通知公告

当前位置：网站首页 >> 通知公告 >> 正文

关于我校推荐2019年省国级大学生创新创业训练计划项目的公示

发布时间:2019年05月08日 浏览量: 482

各二级学院:

根据《广东省教育厅关于报送2019年度国家级、省级大学生创新创业训练计划立项项目的通知》（粤教高函[2019]40号）要求，经学院申报、校外专家评审，拟推荐“基于机器视觉的自动识别跟踪无人机”等120个项目为2019年度省国级大学生创新创业训练计划项目，现予以公示。（项目详情见附件）

公示期2019年5月8日—5月14日。公示期间，如有异议或问题反映，可通过电话、来信、邮件等方式向创新创业学院实名反映，反映问题需实事求是。

受理地点：校本部工业中心402办公室
联系电话：38765279
电子邮件：
562206851@qq.com

附件：
广东技术师范大学2019年省国级大学生创新创业训练计划项目申报一览表.xls

A	B	C	D	E	F	G	H	I
61	机电学院	基于python的儿童英语早教游戏	麦世兴	2017094243036	罗永顺	创新训练	省级	
62	机电学院	工业级３D打印机关键零部件的智能控制及系统集成	李芯怡	2016096143023	建、周莉、郑振	创新训练	省级	
63	计算机科学学院	基于生成式对抗网络的眼底图杯盘分割方法	林晓鹏	2016034743141	刘少鹏	创新训练	省级	
64	计算机科学学院	一种基于深度学习的青光眼辅助分级方法研究与实现	陈昌海	2016034743142	贾西平	创新训练	省级	
65	计算机科学学院	基于深度学习的青光眼辅助诊断应用	胡　娟	2016035643007	鹏、贾西平、曾	创新训练	省级	
66	计算机科学学院	广州道恩光学电子科技有限公司	王伟永	2016034243026	杨燕佳、黄明睿	创业实践	省级	
67	计算机科学学院	大学生创新创业训练项目管理系统	方俊杰	2018035043007	杨燕佳、彭炜煌	创新训练	省级	
68	计算机科学学院	智能公交站牌系统	彭志鑫	2018035743032	王磊军、陈荣军	创新训练	省级	
69	计算机科学学院	基于移动互联网的便携式大学生体质检测系统	李晓萱	2016034243006	卢旭	创新训练	省级	
70	计算机科学学院	关于高校新疆少数民族职教师范生职业认同感的调研	叶汶靖	2018034843130	杨燕佳	创新训练	省级	
71	计算机科学学院	带机器视觉自主行驶的智能快递小车	林佳超	2016035743018	梁琼	创新训练	省级	
72	计算机科学学院	基于物联网技术的智能快递物流系统	马勇枝	2016035743039	吕巨建、陈荣军、赵慧民	创新训练	省级	
73	计算机科学学院	康源青果品农民专业合作社	陈锦涛	2016034243003	彭炜锋、杨燕佳	创业训练	省级	
74	计算机科学学院	裹仓公司	何清莹	2018035244005	涂青云	创业训练	省级	
75	计算机科学学院	智能卡表面质量视觉检测系统	陈荣杰	2016034743051	郝刚、梁鹏	创新训练	省级	
76	教育科学与技术学院	智慧家居APP的情感化体验设计与研究	梁妙瑜	2016084643020	王竹君	创新训练	省级	

2021

Mathematical Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Cai Weibin
Huang Yinfeng
Chao Lumeng

With Faculty Advisor

Pan Maojie

Of

Guangdong Polytechnic Normal University

Was Designated As

Successful Participant

Steven B. Horton, Contest Director

Administered by

COMAP

With support from

AMS ASA informs MAA siam

Kelly Black, Head Judge

2020

Interdisciplinary Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Senpeng Huang
Weidian Chen
Guolingfeng Jiang

With Faculty Advisor

Jinghua Jiang

Of

Guangdong Polytechnic Normal University

Was Designated As

Honorable Mention

Amanda Beecher, Contest Director

Administered by COMAP

Kayla Blyman, Head Judge

With support from

AMS ASA informs MAA siam

2020

Mathematical Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Junyong Han
Miaoyu Wang
Jianxin Liu

With Faculty Advisor

Hui Fu

Of

Guangdong Polytechnic Normal University

Was Designated As

Successful Participant

Patrick Driscoll, Contest Director

Administered by COMAP

With support from

David H. Olwell, Head Judge

AMS ASA informs MAA siam

2020

Interdisciplinary Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Senpeng Huang
Weidian Chen
Guolingfeng Jiang

With Faculty Advisor

Jinghua Jiang

Of

Guangdong Polytechnic Normal University

Was Designated As

Honorable Mention

Amanda Beecher, Contest Director

Administered by COMAP

Kayla Blyman, Head Judge

With support from

AMS ASA informs MAA siam

2020

Mathematical Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Li Huiqiong
Zhang Lei
Zhong Zhihui

With Faculty Advisor

Sun Zhongju

Of

Guangdong Polytechnic Normal University

Was Designated As

Successful Participant

Patrick Driscoll, Contest Director

Administered by

COMAP

With support from

Robert Burks, Head Judge

AMS ASA informs MAA siam

序号	报名号	获奖选手			所在院校	指导老师
461	202119016014	苏悦	李大伟	陈杰森	深圳技术大学	陈寅
462	202119016017	李世杰	林佳锋	陈锦聪	深圳技术大学	陈寅
463	202119017001	蔡云翔	严焯专	黄辉泉	深圳大学（与深圳技术大学联合培养）	曹志强
464	202119018002	陈昌荣	张利富	谢淮润	广州航海学院	刘洋
465	202119018006	周森尧	陈永贵	杨俊峰	广州航海学院	郑珊
466	202119018007	肖虹蕾	吴晓锦	黄琳	广州航海学院	张子文
467	202119018008	郑宇捷	祝仕劲	张志亮	广州航海学院	梁英
468	202119018010	王秋小	欧兴坤	袁伟鸿	广州航海学院	张晓梅
469	202119018011	何飞龙	陈愉悦	陈子霖	广州航海学院	陈宇环
470	202119018012	陈彦全	陈奕旭	盘卓	广州航海学院	陈宇环
471	202119019009	丁婧可	祁佩瑶	冯思鹏	广东技术师范大学	陈艳美
472	202119019014	廖雪丹	唐明意	张依玲	广东技术师范大学	陈月红
473	202119019017	郑治法	罗越雄	李昊	广东技术师范大学	付辉
474	202119019[illegible]	马晓纯	[illegible]奕	林少恒	广东技术师范大学	付辉
475	202119019022	郑渊琼	梁彦琪	赖馨婷	广东技术师范大学	付辉
476	202119019023	叶燕琴	谭子琦	陈越	广东技术师范大学	黄凤英
477	202119019025	谢辉月	李远亮	赵浩男	广东技术师范大学	黄凤英
478	202119019032	宋世豪	项文慧	黎姿	广东技术师范大学	蒋经华
479	202119019040	黎文	谢宝谊	陈欢欢	广东技术师范大学	蒋经华
480	202119019054	蓝栋鑫	刘采欣	麦绮玲	广东技术师范大学	蒋经华
481	202119019065	曾繁龙	黄豪彬	赵彬宇	广东技术师范大学	李振彰
482	202119019071	吴小迪	吕炜诗	杨嘉倩	广东技术师范大学	李振彰
483	202119019074	周泽理	刘秋花	翁奕鑫	广东技术师范大学	李振彰
484	202119019104	黄楷孜	林婉彤	蔡家裕	广东技术师范大学	娄曼丽
485	202119019105	黄丹丽	詹婕妤	周楚颖	广东技术师范大学	娄曼丽
486	202119019109	陈一鸣	吴汉杰	林泽伟	广东技术师范大学	盘茂杰
487	202119019113	李惠婵	石耀安	陈浚光	广东技术师范大学	盘茂杰
488	202119019114	曾凡海	梁文新	李思欣	广东技术师范大学	彭雪
489	202119019116	黄耀贤	林钦龙	陈玮杰	广东技术师范大学	彭雪
490	202119019118	吴佳纯	陈新悦	刘春玲	广东技术师范大学	彭雪
491	202119019121	郭惠虹	刘智琳	黄宽豪	广东技术师范大学	彭雪
492	202119019122	柯宜鹏	李奕欣	陈晓琪	广东技术师范大学	彭雪
493	202119019123	赖泽洋	潘子豪	莫晓宇	广东技术师范大学	彭雪
494	202119019125	陈铭俐	钟雨婷	郑梓霈	广东技术师范大学	彭雪
495	202119019132	李诗然	周颖琪	胡晶丽	广东技术师范大学	彭雪
496	202119019138	陈绵希	吴雅思	郑怀瑾	广东技术师范大学	彭雪
497	202119019139	詹多谋	张千禧	徐怡婷	广东技术师范大学	彭雪
498	202119019141	卢薛桦	曾俊森	谢彤宇	广东技术师范大学	彭雪
499	202119019145	黄熙桐	梁祎晴	林欣语	广东技术师范大学	邱宝华
500	202119019149	吴佳佳	王静莉	郑梓欣	广东技术师范大学	孙中举
501	202119019150	胡冠鑫	徐嘉玫	许志远	广东技术师范大学	孙中举

479	202119019040	黎文	谢宝盛	陈欣欣	广东技术师范大学	高经华
480	202119019054	蓝畅鑫	刘采欣	麦靖玲	广东技术师范大学	高经华
481	202119019065	曾繁龙	黄锋彬	赵彬宇	广东技术师范大学	李振彰
482	202119019071	吴小迪	吕伟诗	杨惠倩	广东技术师范大学	李振彰
483	202119019074	周泽理	郑民花	曲奕鑫	广东技术师范大学	李振彰
484	202119019104	黄煜欣	林婉彤	蔡家熙	广东技术师范大学	姜曼丽
485	202119019105	黄丹丽	詹婕妤	周楚莉	广东技术师范大学	姜曼丽
486	202119019109	陈一鸣	吴汉杰	林泽伟	广东技术师范大学	盘茂杰
487	202119019113	李惠婷	石耀安	郑茂光	广东技术师范大学	盘茂杰
488	202119019114	曾凡海	梁文静	李思欣	广东技术师范大学	彭雪
489	202119019116	黄耀贤	林钦龙	陈炜杰	广东技术师范大学	彭雪
490	202119019118	吴佳纯	陈新悦	刘春玲	广东技术师范大学	彭雪
491	202119019121	郭惠虹	刘智琳	黄宽豪	广东技术师范大学	彭雪
492	202119019122	柯宜鹏	李奕欣	陈晓琪	广东技术师范大学	彭雪
493	202119019123	赖泽洋	潘子豪	莫晓宇	广东技术师范大学	彭雪
494	202119019125	陈铭纲	钟惠婷	郑梓霖	广东技术师范大学	彭雪
495	202119019132	李诗然	周颖琪	胡晶丽	广东技术师范大学	彭雪
496	202119019135	[illegible]	[illegible]	[illegible]	广东技术师范大学	彭雪

序号	报名号	获奖选手			所在院校	指导老师
502	202119019153	刘颖	邱淑慧	林智慧	广东技术师范大学	孙中举
503	202119019154	肖少萍	何玉山	莫玉帅	广东技术师范大学	孙中举
504	202119019155	陈馥婷	曾敏	黄润杰	广东技术师范大学	孙中举
505	202119019188	麦颖欣	陈菀灵	陈栩	广东技术师范大学	王燕
506	202119019192	吴梓健	戴柳茹	刘光林	广东技术师范大学	翁伟明
507	202119019207	邓慧婷	陈思慧	罗晓莹	广东技术师范大学	张冰
508	202119019209	史志毅	王之鸿	冯志扬	广东技术师范大学	张冰
509	202119019212	李玉敏	李智莹	钟文畅	广东技术师范大学	张广亮
510	202119019214	黄志聪	赖雪渝	朱嘉欣	广东技术师范大学	张广亮
511	202119019219	袁旻罡	凌林燕	罗翊郡	广东技术师范大学	张姣玲
512	202119020004	李美萱	徐图意	刘伯鸿	岭南师范学院	张宏礼
513	202119020006	李怡慧	赖竹凤	尹欣明	岭南师范学院	刘雄
514	202119020008	王艺霖	戴杰银	麦东琳	岭南师范学院	李雨江
515	202119020012	邓充盛	曾钰秋	郑凯钊	岭南师范学院	李雨江
516	202119020013	何赵敏	梁艺	彭武贤	岭南师范学院	麦继芳
517	202119020021	陈慧滢	杨佳丽	朱键锋	岭南师范学院	黄福员
518	202119020027	江展蓉	陈忠勇	李凌瑜	岭南师范学院	施业成
519	202119020047	彭思琦	林佳仪	董豪杰	岭南师范学院	周秀香
520	202119020057	李晓文	李烨俊	张丹萍	岭南师范学院	张宏礼
521	202119021002	王玉娴	林淇韦	张体威	韩山师范学院	肖刚
522	202119021008	苏文豪	吴艳丽	朱昌辉	韩山师范学院	麦桂珍
523	202119021012	高文娟	黄道捷	李嘉瑜	韩山师范学院	数模组
524	202119021015	曾小红	孙依婷	吕雯雯	韩山师范学院	数模组
525	202119021017	吴潮嘉	单瀚林	陈灏婷	韩山师范学院	数模组
526	202119022012	欧康婷	俞彩云	戴丽诗	广东石油化工学院	伍思敏
527	202119023004	陈小端	黄诚钢	罗秋怡	广东金融学院	柴富杰
528	202119023009	潘雨娟	吴晓韩	邹嘉琪	广东金融学院	陈员龙
529	202119023011	陈绮璐	蔡佳杰	连楚鑫	广东金融学院	葛仁良
530	202119023016	马创杰	黄钊贤	熊可彬	广东金融学院	胡蓉
531	202119023017	邝振杨	谭文凯	柯佳纯	广东金融学院	黄嘉莹
532	202119023019	黄荣华	符弘锟	梁家肇	广东金融学院	黄嘉莹
533	202119023024	黄诗琪	谢嘉浩	方俊铭	广东金融学院	李广析
534	202119023035	何琰	邓姮	张嘉琪	广东金融学院	刘炎
535	202119023036	赖鹤鸣	吴炫壮	倪梓熙	广东金融学院	刘炎
536	202119023037	王虹榆	聂敏怡	陈羽欣	广东金融学院	刘炎
537	202119023045	欧宇珩	黎禹琦	林晓花	广东金融学院	王芬
538	202119023051	李泳怡	蔡丹仪	李金鸿	广东金融学院	谢家泉
539	202119023054	郑炽辉	李沛苇	王宏博	广东金融学院	徐鹏程
540	202119024002	陈炳耿	陈明通	陈志锋	广东理工学院	李亚男
541	202119024005	赖灿炜	林壮峰	钟雅平	广东理工学院	裴阳
542	202119024006	朱铭奎	谭若冰	邱远成	广东理工学院	赵戊梅
543	202119024025	李雪怡	冯洁宜	张方霖	广东理工学院	于晓曼

2021年全国大学生数学建模竞...

序号	报名号	获奖选手			所在院校	指导老师
461	202119016014	苏悦	李大伟	陈杰森	深圳技术大学	陈富
462	202119016017	李世杰	杨佳锋	陈雅丽	深圳技术大学	陈富
463	202119017001	蔡启阳	严绿冬	黄耀康	深圳大学（与深圳技术大学联合培养）	曹志强
464	202119018002	陈昌荣	张利富	谢雁四	广州航海学院	刘玮
465	202119018006	周森鑫	陈永贵	杨俊峰	广州航海学院	郑璐
466	202119018007	肖红雪	吴晓玲	黄琳	广州航海学院	张子文
467	202119018008	郑宇捷	倪仕勃	张志高	广州航海学院	谢英
468	202119018010	王秋小	欧兴坤	袁伟鸿	广州航海学院	张晓梅
469	202119018011	何飞龙	陈愉悦	陈子霞	广州航海学院	陈宇环
470	202119018012	陈彦全	陈奕斌	袁卓	广州航海学院	陈宇环
471	202119019009	丁靖珂	郑佩瑶	冯思鹏	广东技术师范大学	陈艳贵
472	202119019014	廖雪丹	唐明康	张钒玲	广东技术师范大学	陈月红
473	202119019017	邢雨涵	尹顺越	李晨	广东技术师范大学	付辉
474	202119019019	马晓均	肖鑫	林少恒	广东技术师范大学	付辉
475	202119019022	郑靖琪	梁彦琪	赖馨婷	广东技术师范大学	付辉
476	202119019023	叶浩骏	潘子祺	陈毓	广东技术师范大学	黄凤英
477	202119019025	谢晖月	李远亮	赵尚勇	广东技术师范大学	黄凤英
478	202119019032	宋世康	项文慧	黎安	广东技术师范大学	蒋经华
479	202119019040	黎大	谢宝盛	陈欣欣	广东技术师范大学	蒋经华
480	202119019054	蓝皓鑫	刘宇航	常晓玲	广东技术师范大学	蒋经华
481	202119019065	曾歌凯	黄惠林	赵彬宇	广东技术师范大学	李振彰
482	202119019071	吴小迪	冯伟涛	杨嘉俊	广东技术师范大学	李振彰
483	202119019074	周泽健	刘祝远	翁奕鑫	广东技术师范大学	李振彰
484	202119019104	黄煜政	钟婉彤	蔡家桢	广东技术师范大学	佘慧丽
485	202119019105	詹丹凯	詹捷妤	周慧妍	广东技术师范大学	佘慧丽
486	202119019109	陈一鸣	吴汉杰	杨锦伟	广东技术师范大学	袁茂杰
487	202119019113	李惠婷	石耀安	陈茂九	广东技术师范大学	袁茂杰
488	202119019114	曾凡海	梁文彬	李思欣	广东技术师范大学	彭雪
489	202119019116	黄耀贤	林钦龙	陈炜杰	广东技术师范大学	彭雪
490	202119019118	吴佳均	陈新悦	刘春玲	广东技术师范大学	彭雪
491	202119019121	郭惠红	刘智婷	黄宽豪	广东技术师范大学	彭雪
492	202119019122	柯宜鹏	李奕欣	陈晓琪	广东技术师范大学	彭雪
493	202119019123	赖泽洋	潘子豪	吴俊宇	广东技术师范大学	彭雪
494	202119019125	陈铭树	钟雨婷	郑梓霖	广东技术师范大学	彭雪
495	202119019132	李诗然	梁颖琪	胡晶慧	广东技术师范大学	彭雪
496	202119019138	陈德寿	吴雅君	郑仔曦	广东技术师范大学	彭雪
497	202119019139	詹多谋	张千曦	张钰婷	广东技术师范大学	彭雪
498	202119019141	卢霖楷	曾俊森	谢彬宇	广东技术师范大学	彭雪
499	202119019145	黄熙桐	梁伟锋	林欣浩	广东技术师范大学	邱宝华
500	202119019149	吴佳佳	王静莉	郑梓欣	广东技术师范大学	孙中举
501	202119019150	胡冠峰	徐嘉欣	许志远	广东技术师范大学	孙中举

2021年全国大学生数学建模竞赛广东省分赛获奖名单（初稿）_Page35.jpg

获奖证书

广东技术师范大学 彭泽蔓 荣获第十一届全国大学生数学竞赛（非数学类）三等奖，特发此证。

编号：CMS（粤）F20190513

中国数学会普及工作委员会

二〇一九年十一月

2019年

全国大学生数学建模竞赛

广东省分赛

获 奖 证 书

广东技术师范大学

学生 林崇宇 苏佳诚 吴静怡

指导老师 彭 雪

荣获

三等奖

广东省教育厅 中国工业与应用数学学会

Mathematical Contest in Modeling

2019年

全国大学生数学建模竞赛

广东省分赛

获 奖 证 书

广东技术师范大学

学生 吴浩贤 韩军勇 周晓玲

指导老师 付 辉

荣获

三等奖

广东省教育厅 中国工业与应用数学学会

Mathematical Contest in Modeling

2019年

全国大学生数学建模竞赛

广东省分赛

获 奖 证 书

广东技术师范大学

学生 黎颖琦 吴龙凯 陈林婷

指导老师 孙中举

荣获

一等奖

广东省教育厅 中国工业与应用数学学会

Mathematical Contest in Modeling

2019年

全国大学生数学建模竞赛

广东省分赛

获　奖　证　书

广东技术师范大学

学生　杨宇平　黄泽涛　刘英杰

指导老师 娄曼丽

荣获　**优胜奖**

中国工业与应用数学学会

获奖证书

队　　号　B20044795

学　　校　广东技术师范大学

学　　生　王泽华　彭婉玲　郑美惠

指导教师　彭雪

荣获 2020 年第十七届五一数学建模竞赛

二等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2020 年 5 月

获奖证书

队　　号　B20043107

学　　校　广东技术师范大学

学　　生　丘敏　吴少春　蓝美琴

指导教师　孙中举

荣获 2020 年第十七届五一数学建模竞赛

二等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2020 年 6 月

获奖证书

队　　号　T210925853844448

学　　校　广东技术师范大学

学　　生　陈权升　林思瀚　陈晓业

指导教师　廖秀秀

荣获 2021 年第十八届五一数学建模竞赛

二等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2021 年 6 月

获奖证书

队　　号　B20041846

学　　校　广东技术师范大学

学　　生　陈嘉濠　黄万新　张嘉雯

指导教师　盘茂杰

荣获 2020 年第十七届五一数学建模竞赛

三等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2020 年 6 月

获奖证书

队　　号　B20041846

学　　校　广东技术师范大学

学　　生　陈嘉濠　黄万新　张嘉雯

指导教师　盘茂杰

荣获 2020 年第十七届五一数学建模竞赛

三等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2020 年 6 月

获奖证书

队　　号　B20041846

学　　校　广东技术师范大学

学　　生　陈嘉濠　黄万新　张嘉雯

指导教师　盘茂杰

荣获 2020 年第十七届五一数学建模竞赛

三等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2020 年 6 月

获奖证书

队　　号　T21488557031936

学　　校　广东技术师范大学

学　　生　陈晓婷　许嘉麟　苏明昆

指导教师　李振彰

荣获 2021 年第十八届五一数学建模竞赛

三等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2021 年 6 月

获奖证书

队　　号　B19091366

学　　校　广东技术师范大学

学　　生　吴浩贤　蓝苏琴　周晓玲

指导教师　陈艳美

荣获 2019 年第十六届五一数学建模竞赛

三等奖

江苏省工业与应用数学学会　　中国矿业大学　　徐州市工业与应用数学学会

2019 年 6 月

May Day Mathematical Contest in Modeling

2020年MathorCup高校数学建模挑战赛——大数据竞赛

荣誉证书

广东技术师范大学

参赛队员：罗冰棋 袁芝芸 付曾艳

指导教师：孙中举

荣获2020年MathorCup高校数学建模挑战赛——大数据竞赛 本科组

三等奖

2020

Mathematical Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Li Huiqiong
Zhang Lei
Zhong Zhihui

With Faculty Advisor

Sun Zhongju

Of

Guangdong Polytechnic Normal University

Was Designated As

Successful Participant

Patrick Driscoll, Contest Director

Administered by

COMAP

With support from

Robert Burks, Head Judge

AMS ASA informs MAA siam

序号	报名号	获奖选手			所在院校	指导老师
461	202119016014	苏悦	李大伟	陈杰森	深圳技术大学	陈寅
462	202119016017	李世杰	杨佳臻	陈雅聪	深圳技术大学	陈寅
463	202119017001	蔡启阳	严绿专	黄辉彦	深圳大学（与深圳技术大学联合培养）	曹志强
464	202119018002	陈昌焕	张利富	谢鹰润	广州航海学院	刘璋
465	202119018006	周森奇	陈永贵	杨俊峰	广州航海学院	郑璐
466	202119018007	肖虹雪	吴晓玲	黄琳	广州航海学院	张子文
467	202119018008	邢宇捷	倪壮勃	张志高	广州航海学院	谢晨
468	202119018010	王秋小	欧兴坤	袁伟鸿	广州航海学院	张晓梅
469	202119018011	何飞龙	陈愉悦	陈子晨	广州航海学院	陈宇环
470	202119018012	陈彦全	陈奕斌	曲卓	广州航海学院	陈宇环
471	202119019009	丁婧珂	祁佩瑶	吕思鹏	广东技术师范大学	陈艳侠
472	202119019014	廖雪丹	唐明康	张钦玲	广东技术师范大学	陈月红
473	202119019017	邢雨涵	尹靓颖	李昊	广东技术师范大学	付辉
474	202119019019	马晓均	肖燕	林少恒	广东技术师范大学	付辉
475	202119019022	邓辉欣	梁彦琪	赖馨婷	广东技术师范大学	付辉
476	202119019023	叶语琴	潘子璇	陈韬	广东技术师范大学	黄凤英
477	202119019025	谢辉月	李远亮	赵炳亮	广东技术师范大学	黄凤英
478	202119019032	宋世泰	项文慧	黎安	广东技术师范大学	蒋绍华
479	202119019040	黎文	谢宝焕	陈欣欣	广东技术师范大学	蒋绍华
480	202119019054	雷炜鑫	刘宇凯	李婉玲	广东技术师范大学	蒋绍华
481	202119019065	曾繁龙	黄奕彬	赵树宇	广东技术师范大学	李振彰
482	202119019071	吴小迪	吕炜诗	杨嘉倩	广东技术师范大学	李振彰
483	202119019074	周泽婕	刘枫云	翁奕鑫	广东技术师范大学	李振彰
484	202119019104	黄程胶	钟婉彤	蔡家桢	广东技术师范大学	姜曼丽
485	202119019105	詹丹韵	詹婕妤	周慧轩	广东技术师范大学	姜曼丽
486	202119019109	陈一鸣	吴汉杰	林泽伟	广东技术师范大学	盘茂杰
487	202119019113	李惠婷	石耀安	陈茂六	广东技术师范大学	盘茂杰
488	202119019114	曾凡海	梁文轩	李思秋	广东技术师范大学	彭雪
489	202119019116	黄耀斌	林钦龙	陈炜杰	广东技术师范大学	彭雪
490	202119019118	吴佳鸣	陈毓瑜	刘春玲	广东技术师范大学	彭雪
491	202119019121	郭惠红	刘智婷	黄楚豪	广东技术师范大学	彭雪
492	202119019122	柯宜鹏	李奕凯	陈锦琪	广东技术师范大学	彭雪
493	202119019123	赖泽洋	潘子豪	吴俊宇	广东技术师范大学	彭雪
494	202119019125	陈铭州	钟雨婷	郑梓霖	广东技术师范大学	彭雪
495	202119019132	李诗然	梁颖琪	胡晶菁	广东技术师范大学	彭雪
496	202119019138	陈绵寿	吴雅思	郑仟禧	广东技术师范大学	彭雪
497	202119019139	詹多谋	张千曦	徐贻婷	广东技术师范大学	彭雪
498	202119019141	卢森楠	曾俊森	谢彬宇	广东技术师范大学	彭雪
499	202119019145	黄熙桐	梁炜锋	林欣浩	广东技术师范大学	邱宝华
500	202119019149	吴佳佳	王静莉	郑梓欣	广东技术师范大学	孙中华
501	202119019150	胡冠羲	徐嘉欣	许志远	广东技术师范大学	孙中华

序号	报名号	获奖选手			所在院校	指导老师
502	202119019153	刘颖	邱淑慧	林智慧	广东技术师范大学	孙中举
503	202119019154	肖少萍	何玉山	莫玉帅	广东技术师范大学	孙中举
504	202119019155	陈馥婷	曾敏	黄润杰	广东技术师范大学	孙中举
505	202119019188	麦颖欣	陈菀灵	陈栩	广东技术师范大学	王燕
506	202119019192	吴梓健	戴柳茹	刘光林	广东技术师范大学	翁伟明
507	202119019207	邓慧婷	陈思慧	罗晓莹	广东技术师范大学	张冰
508	202119019209	史志毅	王之鸿	冯志扬	广东技术师范大学	张冰
509	202119019212	李玉敏	李智莹	钟文畅	广东技术师范大学	张广亮
510	202119019214	黄志聪	赖雪渝	朱嘉欣	广东技术师范大学	张广亮
511	202119019219	袁旻罡	凌林燕	罗翊郡	广东技术师范大学	张姣玲
512	202119020004	李美萱	徐图意	刘伯鸿	岭南师范学院	张宏礼
513	202119020006	李怡慧	赖竹凤	尹欣明	岭南师范学院	刘雄
514	202119020008	王艺霖	戴杰银	麦东琳	岭南师范学院	李雨江
515	202119020012	邓充盛	曾钰秋	郑凯钊	岭南师范学院	李雨江
516	202119020013	何赵敏	梁艺	彭武贤	岭南师范学院	麦继芳
517	202119020021	陈慧滢	杨佳丽	朱键锋	岭南师范学院	黄福员
518	202119020027	江展蓉	陈忠勇	李凌瑜	岭南师范学院	施业成
519	202119020047	彭思琦	林佳仪	董豪杰	岭南师范学院	周秀香
520	202119020057	李晓文	李烨俊	张丹萍	岭南师范学院	张宏礼
521	202119021002	王玉娴	林淇韦	张体威	韩山师范学院	肖刚
522	202119021008	苏文豪	吴艳丽	朱昌辉	韩山师范学院	麦桂珍
523	202119021012	高文娟	黄道捷	李嘉瑜	韩山师范学院	数模组
524	202119021015	曾小红	孙依婷	吕雯雯	韩山师范学院	数模组
525	202119021017	吴潮嘉	单瀚林	陈灏婷	韩山师范学院	数模组
526	202119022012	欧康婷	俞彩云	戴丽诗	广东石油化工学院	伍思敏
527	202119023004	陈小端	黄诚钢	罗秋怡	广东金融学院	柴富杰
528	202119023009	潘雨娟	吴晓韩	邹嘉琪	广东金融学院	陈员龙
529	202119023011	陈绮璐	蔡佳杰	连楚鑫	广东金融学院	葛仁良
530	202119023016	马创杰	黄钊贤	熊可彬	广东金融学院	胡蓉
531	202119023017	邝振杨	谭文凯	柯佳纯	广东金融学院	黄嘉莹
532	202119023019	黄荣华	符弘锟	梁家肇	广东金融学院	黄嘉莹
533	202119023024	黄诗琪	谢嘉浩	方俊铭	广东金融学院	李广析
534	202119023035	何琰	邓姮	张嘉琪	广东金融学院	刘炎
535	202119023036	赖鹤鸣	吴炫壮	倪梓熙	广东金融学院	刘炎
536	202119023037	王虹榆	聂敏怡	陈羽欣	广东金融学院	刘炎
537	202119023045	欧宇珩	黎禹琦	林晓花	广东金融学院	王芬
538	202119023051	李泳怡	蔡丹仪	李金鸿	广东金融学院	谢家泉
539	202119023054	郑炽辉	李沛苇	王宏博	广东金融学院	徐鹏程
540	202119024002	陈炳耿	陈明通	陈志锋	广东理工学院	李亚男
541	202119024005	赖灿炜	林壮峰	钟雅平	广东理工学院	裴阳
542	202119024006	朱铭奎	谭若冰	邱远成	广东理工学院	赵戊梅
543	202119024025	李雪怡	冯洁宜	张方霖	广东理工学院	于晓曼

HONORARY CREDENTIAL

荣誉证书

基于 A-IoT 技术的区域大气环境检测系统 项目：

荣获 2021 年第七届移动互联创新大赛 华南赛区 高校组 金奖 。

特颁此证，以此鼓励。

所属院校：广东技术师范大学

项目成员：温振涛　林欣颂　黄楚基　张颂林　李宏湛　罗蔼静　谢雪莹

指导教师：赵慧民　陈荣军　张 帆

全国移动互联创新大赛组委会

二〇二一年七月

中国高校计算机大赛

GPLT

证书

2019 团体程序设计天梯赛

全国总决赛

团队铜奖

参赛学校：华山论剑

参赛队伍：枪炮&玫瑰

组　　别：广东技术师范大学

指导教师：刘晓勇

参赛队员（排名不分先后）：

刘健鑫　叶勇森　廖碧莹　李伟健　李开宇

杨仁峰　梁健峰　梁祖铭　罗周鹏　黄柱轩

证书编号：CCCC2019GPLT02606

组委会主任签名：

2019 年 4 月 15 日

获奖证书

Certificate of Achievement

颁给/Awarded to

广东技术师范大学 / Guangdong Polytechnic Normal University

黄柱轩 / HUANG Zhuxuan, 梁健峰 / LIANG Jianfeng, 叶勇森 / YE Yongsen

刘晓勇(教练) / LIU Xiaoyong(Coach)

三等奖/ The Third Prize

2019年中国大学生程序设计竞赛-广东省大学生程序设计竞赛

2019 China Collegiate Programming Contest - Guangdong Provincial Collegiate Programming Contest

中山大学 2019年5月 / Sun Yat-sen University, May 2019

S2019GZT114U022B1

广东省计算机学会

Computer Academy of Guangdong

获奖证书

Certificate of Achievement

颁给/Awarded to

广东技术师范大学 / Guangdong Polytechnic Normal University

黄柱轩 / HUANG Zhuxuan, 梁健峰 / LIANG Jianfeng, 叶勇森 / YE Yongsen

刘晓勇(教练) / LIU Xiaoyong(Coach)

三等奖/ The Third Prize

2019年中国大学生程序设计竞赛-广东省大学生程序设计竞赛

2019 China Collegiate Programming Contest - Guangdong Provincial Collegiate Programming Contest

中山大学 2019年5月 / Sun Yat-sen University, May 2019

S2019GZT114U022B1

广东省计算机学会

Computer Academy of Guangdong

GDCA

2019“华资杯”广东省大学生计算机作品赛
暨泛珠三角+大学生计算机作品赛
广东选拔赛(本科组)

奖励证书

获奖项目：萌宝（疫苗）笔记——基于小程序的农村儿童预防免疫接种辅助系统

获奖等级：三等奖

获 奖 者：覃嘉茵 黄铭杰 梁祖铭 韩军勇 黄子言

指导老师：杨燕佳

获奖学校：广东技术师范大学

证书编号：2019GCFFZSGDBK33

（证书查询）

广东省计算机学会

2019年6月