UAV target recognition algorithm based on fusion of SAE and bottom visual feature

Xie Bing, Duan Zhemin

(School of Electronics and Information, Northwestern Polytechnical University, Xi′an 710072, China)

Abstract: UAV flying in complex battlefield environment, due to the similar shape and color of the enemy UAV, and the existing algorithms can not accurately identify and classify the UAV of the enemy, resulting in false detection or even mishandling attack. To solve this problem, a feature fusion algorithm based on the combination of the bottom visual features and high-level visual features was proposed to classify the UAV target objects. The algorithm first extracted the underlying visual features and high-level visual features of the target object by using visual feature descriptors and Sparse Auto-Encoder (SAE). Then, the principal component analysis (PAC) method was used to reduce the dimensionality of the global features. Finally, the global feature response was sent to the softmax regression model to complete the recognition and classification of the target object of the UAV. Experiments show that the new algorithm has higher accuracy and robustness compared with the traditional SAE algorithm and the traditional recognition algorithm based on the underlying visual features.

Key words: UAV target object; target recognition; Sparse Auto-Encoder; underlying visual descriptor; PCA

CLC number: TP394.1 **Document code:** A **DOI:** 10.3788/IRLA201847.S126004

基于 SAE 与底层视觉特征融合的无人机目标识别算法

谢 冰,段哲民

(西北工业大学 电子信息学院,陕西 西安 710072)

摘 要: 无人机在复杂战场环境下,因敌方无人机外形、颜色等特征较为相似,现有基于底层视觉特征无法快速地对其进而准确的识别,从而造成误检测甚至误打击等事件的发生。针对这一问题,文中提出基于稀疏自动编码器融合底层视觉特征的算法,对无人机目标对象进行识别。算法首先利用底层视觉特征描述子(GIST、LBP)以及稀疏自动编码器(Sparse Auto-Encoder,SAE)提取目标对象的底层视觉特征和高层视觉特征;然后,采用主成分分析(PAC)法对全局特征进行降维融合;最后,将全局特征响应送入 softmax 回归模型完成无人机目标对象的分类。实验表明,与传统 SAE 算法及传统基于底层视觉特征描述子识别算法相比,新算法具有更高的准确性及鲁棒性。

关键词: 无人机目标对象; 目标识别; Sparse Auto-Encoder; 底层视觉描述子; PCA

收稿日期: 2018-03-16; **修订日期:** 2018-05-24

基金项目: 瞬态冲击技术重点实验室基金(61426060103162606007)

作者简介: 谢冰(1989-),男,博士,主要从事红外目标识别及红外超分辨率重建方面的研究。Email:664897555@qq.com

0 Introduction

Identification and classification of the target objects of UAV under the complex battlefield environment is the key to the successful implementation of the UAV attack task. At present, the accurate recognition and classification of target based on visual technology has become the focus of research at home and abroad. Among them, the accurate extraction of the characteristics of the target is the key to the classification of the target. Existing algorithms are usually based on statistical features such as time-domain-based LBP, color histogram and frequency-domain GIST, Fourier transform, wavelet transform, Hadamard transform, and other feature extraction methods. However, these feature extraction methods by extracting visual features such as grayscale, color, texture, shape and scene of the target image needs to build a complex feature model, which only focuses on expressing the local and superficial information of the image, and it is difficult to fully express the global details of the the target image. Therefore, the accuracy of identification and classification is poor.

In recent years, unsupervised feature learning, which has been used for massive unlabelled data is becoming a new research hotspot[1-3]. This method acquires the most essential information inside the image though simulating the human eye to scan the image, so as to facilitate the recognition and classification of the target. Among them, the SAE, an unsupervised feature learning method has been extended to the application of the limited number of labeled samples[4]. This kind of model does not need to define the feature beforehand, and only by setting the hidden layer element, we can automatically learn the hidden layer data of the image, and get the best relationship inside the image.

Therefore, the feature fusion recognition algorithm based on the combination of the bottom visual features using visual feature descriptors and high-level visual feature using SAE is proposed in this paper to classify the UAV target objects. The specific design of this algorithm is as follows: Firstly, global features of UAV target images are extracted by convolutional sparse self-encoder and traditional low-level visual descriptors such as color histogram, LBP, GIST and so on. Then, the global features obtained under different models are fused by principal component analysis (PAC)[5] to reduce the dimensionality of the feature vector. Finally, the fused global feature response is sent into the softmax regression model to complete the classification. The convolutional SAE algorithm that combined with the underlying features can complete the identification and classification of UAV target objects.

1 Algorithm design flow

The algorithm proposed in this paper mainly includes four modules: global feature extraction module based on bottom visual descriptors, global feature learning module based on SAE, the PAC feature fusion module and the target recognition and classification module.

The overall framework of the system is shown in Fig.1.

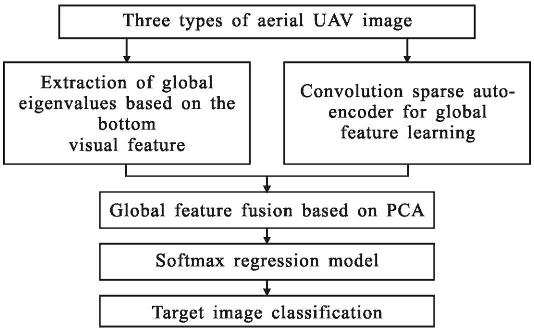
Three types of aerial UAV image
Extraction of global eigenvalues based on the bottom visual feature
Convolution sparse auto-encoder for global feature learning
Global feature fusion based on PCA
Softmax regression model
Target image classification

Fig.1 Overall block diagram of algorithm design

As shown in Fig.1, the global feature extraction module based on the botom visual feature descriptor is to use Color histogram[6], LBP descriptor[7] and GIST[8] descriptor to extract the features of the color, texture and contour of multi-frame aerial UAV images, Then, the global features of the target image obtained by each descriptor are vectorized to obtain their Global feature responses respectively. global feature learning module based on the convolutional SAE uses the SAE adding the sparsity constraint to learn the local features of each image sub-block collected in the target domain, and using whitening treatment to enhance the edges of the features; Then, using the convolutional neural network to obtain the global characteristics of each position of the whole image of each frame through convolving the learned local features learned by SAE on multi-frame aerial UAV images in the target area; Finally, the global feature response with rotation and scaling invariance is obtained by using pooling operation. Global feature fusion module based on PCA is used to reduce the dimension of global feature extracted both SAE and botom visual descriptors (color histogram, LBP descriptor and GIST descriptor), and remove the redundancy between feature vectors to obtain the best global target image expression. Global feature vectors of aerial multi-frame UAV images using PCA are sent into the softmax regression model to finish cross-validation at different iterations. The classification performance of the algorithm proposed in this paper is evaluated by recognition accuracy.

2 Algorithm design

2.1 Color histogram feature extraction

Color histogram is a common method to describe the color feature of the image, which can describe the global distribution and the proportion of different color distribution in the whole image. It is widely used in image feature extraction due to its advantages of fast computing speed, simple algorithm, and the advantages of scale, translation and rotation invariance. The color histogram can be expressed as a one-dimensional vector, as shown in equation (1).

$$H_P=[h_1,\ h_2,\ \cdots,\ h_L] \tag{1}$$

where h_k represents the the image pixel frequency in k kind of color.

$$h_k=\frac{n_k}{N}\quad k=0,\ 1,\ \cdots,\ L-1 \tag{2}$$

where k deontes the color level of an image, L denotes the number of desirable color levels, n_k is the number of pixels with color level k in the image, N denotes the total number of pixels in an image.

Since the traditional color histogram method only counts the global color information of the image and loses the spatial distribution information of the image, the retrieval efficiency will be reduced when two completely unrelated images have the same color histogram. When the feature vector in the image can not take all the values, there will be many zero values in the color histogram, which will affect the intersection operation of the histogram. As a result, the matching result can not reflect the color difference between the two images. In order to solve this problem, the cumulative color histogram is used to extract the color features of the target image. As shown in Eq.(3).

$$h_k=\frac{\sum_{i=0}^{k} n_i}{N}\quad k=0,\ 1,\ \cdots,\ L-1 \tag{3}$$

where h_k denotes the cumulative frequency of colors in pixels.

2.2 Feature extraction of LBP descriptor

Local binary patterns (LBP) proposed by OJala, which is an effective texture description operator by using statistical methods to extract the global texture features of the target image. It is widely used in license plate recognition, fingerprint recognition, face recognition because of its computational complexity, multi-scale features, rotation invariant features and other advantages.

The specific process of LBP descriptor extraction feature is as follows: first, comparing the gray value of the center pixel of the detection window with the gray value of the neighborhood pixel to obtain the binary code of the neighborhood pixel. If the gray value of the neighborhood pixel is greater than the gray value of the center pixel, The corresponding binary code is 1; otherwise 0. Then, stringing the resulting binary code in clockwise order as the new center pixel. Finally, each binary number is weighted to sum, and convert decimal number to obtain a local binary pattern of the central pixel.

The definition of the LBP descriptor is shown in formula (4):

$$\mathrm{LBP}_{P,R}=\sum_{k=0}^{P-1}S(g_k-g_c)2^k \tag{4}$$

where $s(x)=\begin{cases}1, & x>0\\ 1, & x\leqslant 0\end{cases}$, g_c denotes the gray value of the central pixel, g_k denotes the gray value of neighborhood pixels, P denotes the number of elements in neighborhood, R denotes radius of the neighborhood, 2^k denotes weighting factor for each binary number assignment.

Since the LBP operator contains two transitions from 0 to 1 or from 1 to 0. Ojala proposed LBP operator with the uniform mode, that is, when the local binary mode corresponding to the binary number from 0 to 1 or from 1 to 0 has two transitions, then the local binary number is treated as a unified mode class. The uniform pattern LBP operator is defined as:

$$\mathrm{LBP}_{P,R}^{u2}(g_c)=\begin{cases}\sum\limits_{k=0}^{P-1}\mathrm{s}(g_k-g_c), & \text{if } \mathrm{LBP}_{P,R}^{u2}(g_c)\leqslant 2\\ P+1, & \text{otherwise}\end{cases} \tag{5}$$

where $\mathrm{LBP}_{P,R}^{u2}$ (g_c) denotes the number of transitions from 0 to 1 or from 1 to 0 in local binary mode, $\mathrm{LBP}_{P,R}\leqslant 2$ denotes unified pattern. In this paper, the LBP operator based on the uniform model of rotation invariance is used to extract the texture of aerial UAV images.

2.3 Feature extraction of GIST descriptor

The GIST descriptor was first proposed by Olive, which uses Gabor filter banks with different scales and different orientations to filter the target image. Then, the filtered images were divided into regular and averaged grids in each grid to obtain the local texture features of the target image. Finally, the mean value of all the grids is cascaded into the global texture features of the target image. It is widely used to describe the global contour features of the target image due to the multi-scale Gabor filter bank can capture different spatial frequency, spatial location, orientation selectivity and other local structure information in the image, and these local structure information is not sensitive to the change of brightness. As shown in formula (6):

$$G_i(x,y)=\mathrm{cat}(f(x,y)*g_{mn}(x,y)) \tag{6}$$

where cat denotes concatenated symbol, g_{mn} (x,y) denotes filter banks, * denotes convolution operator symbol. In order to make better use of the GIST model to describe the global contour of the target image, a filter bank of 8 directions and 4 scales is used to perform convolution filtering on each target image in this paper. After the 8 ×8 grid filtering convolution image is divided, the description vector of the global outline feature for each target image is 8×8×4×8 dimension. The feature description vector not only characterizes the overall contour feature of the target image, but also describes the relationship between the local contour features within the mesh.

2.4 Global feature extraction based on convolution SAE

2.4.1 Local feature learning based on SAE

SAE is an improved form of automatic encoder, which was proposed by Bengio in 2007. The principle of SAE is as follows. First, sparse constraints are added to each hidden layer unit, so that most of the neurons in the hidden layer unit are in a "suppressed state", and only a few neurons are in an "excited state". Then, the use of back propagation training to find the minimum cost function to learn the key feature responses of the samples in the target domain, and it has a better ability to learn features of the image.

In this paper, a typical Zero-phase Component Analysis (ZCA) method is used to enhance the edge information by whitening multiple sub-blocks of image the sparse self-encoder learned form the target domain. Assuming that the size of the i image block collected from the target domain is $n\times n$, sorting them by the R, G, and B, a vector x (i) with m=n×n×3 dimensions can be obtained. The input vector after whitening is $x'(i)=W_{\text{white}}x^{(i)}$, where, W_{white} represents a matrix of whitening transform coefficients with the size of m×m. The SAE s-dimensional hidden layer response vector[9] is shown in the formula (7):

$$a^{(i)}=\sigma(Wx^{(i)}+b_1)=\sigma(W_{\text{SAE}}W_{\text{white}}x^{(i)}+b_1) \tag{7}$$

where W_{SAE} is the input weight of each image block connected with the SAE hidden layer and the whitening process, b_1 represents the input offset, and $\sigma(\cdot)$ denotes the activation function. $W=W_{\text{SAE}}W_{\text{white}}$ denotes the overall weight coefficient after whitening, which represents the mapping between the hidden layer and the original data. After the whitening process, the input value will exceed the range of [0,1], so it is not necessary to use the activation function to map the output of SAE for data reconstruction:

$$\hat{x}^{(i)}=W_{\text{SAE}}^{\mathrm{T}}a^{(i)}+b_2 \tag{8}$$

where $\hat{x}^{(i)}$ denotes i restoration sample, $W_{\text{SAE}}^{\mathrm{T}}$ donotes output weight, b_2 denotes output offset.

In order to prevent overfitting and maintain the sparseness of the hidden layer response, it is necessary to add weight attenuation term and sparse penalty term to the cost function, the overall cost function[9] is shown in the formula (9):

$$J_{\text{SAE}}(W_{\text{SAE}},b)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\|\hat{x}^{(i)}-W_{\text{white}}x^{(i)}\|^2+\lambda\|W_{\text{SAE}}\|^2+$$
$$\beta\sum_{j=1}^{s}(\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}) \tag{9}$$

where N denotes the number of unlabeled training samples, λ denotes the weight attenuation coefficient, β denotes the sparse penalty weight, s denotes the number of hidden layer units, ρ denotes the value of the target sparse, $\hat{\rho}$ denotes the average response of all training samples on the jth hidden unit. The input weight W_{SAE} obtained after training SAE is the key parameter to find the data to be self-healing, which is the weight coefficient of the corresponding image sub-block in different positions. The response of the hidden layer obtained from an image block according to the weight coefficient is a local feature response of the image block. As shown in the formula (10):

$$a_T=\sigma(W_Sx_T+b_{1S}) \tag{10}$$

where W_S denotes the local feature weight after the whitening process, b_{1S} denotes input offset, x_T denotes an image block in the target domain.

2.4.2 Global feature extraction of convolutional neural networks

Firstly, the local features of sub-blocks with the size of n×n learned by SAE from source domain database are convolutioned on each frame of image to obtain the global response value of each frame target image. Then, the responses combined into a global feature of each frame image are used for subsequent classification training. The overall sructure chart of global feature extraction is shown in Fig.2.

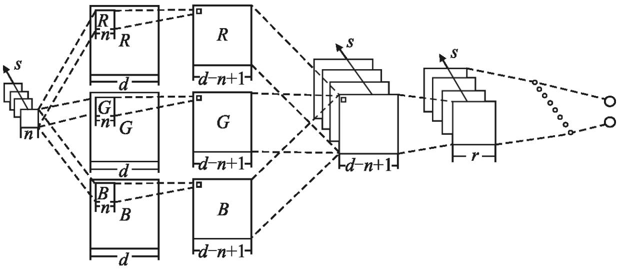

Fig.2 Global feature extraction of target image based on convolutional neural network

(1) Convolutional layer

As shown in Fig.2, assuming that SAE learned the K local features on image sub-blocks whose size is n×n (K is the number of hidden layer units in SAE), and using these local features to convolute on each-frame aerial UAV image to obtain K global feature patterns whose size is (l−n+1)×(l−n+1).

Since SAE is used to extract local features in RGB color space in this paper, in order to improve the computational efficiency, the paper firstly splits the three color channels according to the weight of each local feature learned by SAE before using convolution to extract the global features of each frame image; Then, local features split by three color channels are convolutioned on each frame aerial UAV image respectively point-by-point, and then three global feature vectors which the size is$(l-n+1)\times(l-n+1)$ can be obtained. Finally, the global feature vectors of the three components are summed to obtain the final global feature vector of each frame image. As shown in the formula (11):

$$a=f(WW_{\mathrm{ZW}}^{\mathrm{RGB}}x'+b) \tag{11}$$

where W and b represent the local feature coefficients learned by SAE, x' represents the value of each sub-region to be convoluted in the training image sample, $W_{\mathrm{ZW}}^{\mathrm{RGB}}$ denotes whitening coefficient of the ZCA in the pre-whitening stage.

(2) Pool layer

In convolutional neural networks, the purpose of the pooling layer is to converge the features of different locations obtained in the previous layer. At present, the two methods of pooling take the mean or the maximum value of the pooled area as the aggregated statistical results respectively. The literature shows that convolutional automatic encoder is more suitable for average pooling. Therefore, this paper follows this principle, selecting the average pool as the pooling mode, which can not only reduce the dimension and prevent over-fitting, but also can make the polymerization feature spatially scale and rotation invariance. As shown in Fig.2, the global feature pattern with the size of $(l-n+1)\times(l-n+1)$ will become the size of $p\times p$ after pooling operation.

2.5 Global feature fusion based on PCA

The bottom visual features based on statistics, such as a single texture, color, outline are difficult to fully express the global features of the target image. However, Convolutional SAE, which simulates the human eye to scan the image to obtain the internal information of the image, can fully express the local and global features of the target image. The features extracted from the target image using the two different mechanisms are complementary to each other, and combining the features of the two model can obtain better feature information of the target image. However, if connecting features under two different mechanisms end to end to form the new feature vectors, these feature vectors not only increase a large amount of redundant information, but also cause a "dimensionality disaster". The PCA method can take full account of the correlation between the combined feature vectors, which can not only convert the combined features into lower-dimensional eigenvectors, but also remove the redundancy between the features. In this paper, PCA is used for the merged underlying visual features and high-level visual features to reduce the dimensionality of feature vectors.

PCA known as KL transform, is an optimal data compression and characterization method based on the minimum mean square error criterion of second-order information, which can represent the original data with less interrelated feature space and achieve high-dimensional data reduction and feature extraction. This paper uses PCA to reduce dimension fusion, and algorithm flow is as follows:

(1) The feature vector matrix X is obtained by merging the color feature vector $[a_1,a_2,\cdots,a_n]$, the texture feature vector $[b_1,b_2,\cdots,b_n]$, the contour feature vector $[c_1,c_2,\cdots,c_n]$, and the high-level visual feature vector $[d_1,d_2,\cdots,d_n]$, which is shown in the formula (12).

$$X=\begin{bmatrix} a_1, & b_1, & c_1, & d_1 \\ a_2, & b_2, & c_2, & d_2 \\ \cdots, & \cdots, & \cdots, & \cdots \\ a_n, & b_n, & c_n, & d_n \end{bmatrix} \tag{12}$$

(2) The normalized matrix can be obtained by normalizing the matrix X according to the formula (12), as shown in the formula (13).

$$x^*=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{13}$$

where, x denotes the original matrix element, $x_{\min}$ denotes the minimum value of each column, $x_{\max}$ denotes the maximum value of each column, and x^*

denotes the element of normalization matrix.

(3) The autocorrelation matrix is established according to formula (13), as shown in the formula (14).

$$R=X^{*T}X^{*}/(N-1) \tag{14}$$

(4) The eigenvalues and eigenvectors $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_m$ are obtained from the correlation matrix. The eigenvalues of the autocorrelation matrix and the eigenvector $u_1,u_2,\cdots,u_m$ corresponding to each feature are obtained.

(5) Determining the number of principal components: variance contribution rate and cumulative variance contribution rate, respectively, as shown in formula (15).

$$\eta_i=\lambda_i/\sum_{i=1}^{m}\lambda_i \quad \eta_\Sigma(p)=\sum_{i}^{p}\eta_i \tag{15}$$

where p main components corresponding to the eigenvector is $U_p=[u_1,u_2,\cdots,u_p]$, then the matrix of N samples of the p main components is shown in the fomula (16):

$$Z_{\mathrm{Np}}=X^{*}U_{\mathrm{mp}} \tag{16}$$

(6) The matrix composed of N principal components of p samples is vectorized to obtain the fused global feature response, which is sent to the classification model for classification. The number of samples used in this paper is 300, and the number of principal components p is one third of the dimension of the combined features.

3 Experimental verification

In the part of experimental verification, a new algorithm proposed in this paper, a convolutional SAE classification algorithm, and a classification algorithm based on the underlying visual descriptors (color histogram, LBP descriptor, GIST descriptor) are used for recognition and classification of three types of aerial UAV images, respectively. The experimental sample set is 300 frames of three types of UAV images with different pose and flight speed, wherein the size of each frame of image is 64×64, the part of samples are shown in Fig.3.

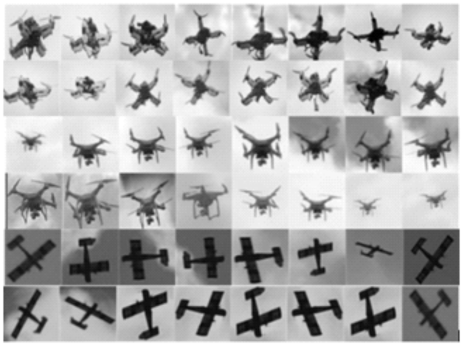

Fig.3 Three types of UAV target image

When using convolutional SAE to extract the high-level visual features of the target image, the regularization constant of the albino preprocessing stage is set as 0.1, the number of hidden layer units of SAE is 100 (corresponding to 100 self-learning features), the weight attenuation coefficient λ is 3e−3, the sparse penalty weight is 5, the target sparse value is 0.035, and the size of the pooling area of the convolutional neural network is set as 19×19. When using PCA to reduce the dimension of the underlying visual features and high-level visual features, The number of principal components is one third of the dimension of the combined features. In order to ensure the generality of the experimental results, the algorithm takes 80% of the target samples for training (240) and the remaining 20% do the tests (60) for 5 times of cross-validation at different iteration times, and the performance of the algorithm is verified from the accuracy aspect.

The bottom visual features such as single color, single texture and single contour of the target sample are extracted by using color histogram, LBP descriptor, GIST descriptor, and the high-level visual features of the target sample extracted by convolutional SAE , which are sent into the softmax classification model for classification, respectively. The classification results under different iteration times are shown in Tab.1.

As shown in Tab.1, the classification of three types of UAV target objects based on convolutional SAE can get much more better classification accuracy than the LBP descriptor classification algorithm, classification algorithm based on color histogram and GIST descriptor classification algorithm under different iterations.

Tab.1 Results of classification accuracy using convolution SAE and single bottom visual feature descriptor

The number of iterations	10	30	50	70	90
Convolutional SAE	86.72%	87.02%	87.43%	87.54%	87.68%
Color histogram	74.32%	75.66%	76.43%	76.89%	75.79%
GIST	80.12%	79.87%	81.32%	79.68%	80.78%
LBP	76.64%	77.65%	78.34%	78.85%	79.66%

Experiments show that the method of unsupervised feature learning based on convolutional SAE which simulate the human eye to scan the image to obtain the deep visual features of the target image can acquire more effective for representation of the target image than the bottom visual features captured by human experience and prior knowledge.

The combination of a single bottom visual features and high-level visual features are sent into the softmax regression model for classification, classification results under different iterations is shown in Tab.2.

As shown in Tab.2, under different iteration times, the high-level visual features of the target image combined with a single underlying visualization are applied for the recognition and classification of three types of UAV target images. The classification accuracy is much higher than classification algorithm based on color histogram, contour GIST descriptive classification algorithm, LBP descriptive classification algorithm and convolutional SAE classification algorithm. The experimental results show that there is a complementary advantage between the underlying visual features and the high-level visual features of the target image under the two different mechanisms. The combined features include more classification information, and the accuracy of the classification is improved.

Tab.2 Results of classification accuracy combined visual features with single underlying visual features

The number of iterations	10	30	50	70	90
Convolutional SAE	86.72%	87.02%	87.43%	87.54%	87.68%
Color histogram	74.32%	75.66%	76.43%	76.89%	75.79%
GIST	80.12%	79.87%	81.32%	79.68%	80.78%
LBP	76.64%	77.65%	78.34%	78.85%	79.66%
Convolutional SAE+ color histogram	87.67%	86.31%	87.91%	88.21%	88.11%
Convolutional SAE+GIST	88.58%	87.96%	88.07%	88.65%	89.03%
Convolutional SAE+LPB	87.36%	87.88%	88.24%	88.16%	88.02%

PCA is used for the merged features which include the single bottom visual features and high-level visual features to reduce dimensionality. The classification results are shown in Tab.3.

Tab.3 Results of classification accuracy using PCA to reduce dimensionality of the merged features

The number of iterations	10	30	50	70	90
Convolution SAE+color histogram+PCA	88.06%	87.28%	88.65%	88.78%	88.87%
Convolution SAE+GIST+PCA	86.95%	87.26%	87.89%	87.94%	87.58%
Convolution SAE+LBP+PCA	86.39%	86.74%	87.23%	86.79%	86.77%
Convolutional SAE+color histogram	87.67%	86.31%	87.91%	88.21%	88.11%
Convolutional SAE+GIST	88.58%	87.96%	88.07%	88.65%	89.03%
Convolutional SAE+LPB	87.36%	87.88%	88.24%	88.16%	88.02%

In the Tab.3, it can be seen that fused features obtained by using PCA to reduce the dimensionality of the merged features is sent into softmax classification model, the classification accuracy does not decrease, but is higher than the classification accuracy of the combined features. This shows that the fused feature using PCA not only retains the effective information of the original feature, but also reduces the dimensionality of the feature, and removes the redundancy between features.

4 Conclusion

This paper presents a feature extraction algorithm that combines the bottom visual features and SAE high-level visual features to classify three types of the UAV target images. The algorithm first extracts the bottom visual features and high-level visual features of the target images by using both traditional feature descriptors and SAE; and then using PAC to reduce the dimension of the combined features; Finally, the global feature response after dimension reduction are sent into softmax model for classification of three types of UAV target images. Experimental results show that the algorithm proposed in this paper is effective and accurate.

References:

[1] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. ***Nature***, 2015, 521(7553): 436-444.

[2] Chen T, Borth D, Darrell T, et al. Deepsentibank: visual sentiment concept classification with deep convolutional neural networks[OL]. http://arxiv.org/abs/1410.8586v1, 2014.

[3] You Q, Luo J, Jin H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks [C]//29th AAAI Conference on Artificial Intelligence (AAAI), Austin, USA, 2015: 381-388.

[4] Coates A, Lee H, Ng A Y. An analysis of single-layer networks in unsupervised feature learning [C]//14th International Conference on Artificial Intelligence and Statistics, 2011: 215-223.

[5] Huang H, Yang A. Face recognition based on PCA algorithm [J]. ***Department of computer Science Guangdong University of Science & Technologyn***, 2015, 8: 98-101.

[6] Nejhum S, Ho J, Yang Minghsuan. Online visual tracking with histograms and articulating blocks [J]. ***Computer Vision and Image Understanding***, 2010, 114(8): 901-914.

[7] Maenpaa T, Pietikinen M. Texture Analysis with Local Binary Patterns. Chen CH & Wang PSP (eds) Handbook of Pattern Recognition and Computer Vertion [M]. 3rd ed. Singapore: World Scientific, 2005: 197-216.

[8] Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition[J]. ***Progress in Brain Research: Visual Perception***, 2006, 155: 23-36.

[9] Eng J, Zhang Z, Eyben F, et al. Autoencoder-based unsupervised domain adaptation for speech emotion recognition [J]. ***IEEE Signal Processing Letters***, 2014, 21 (9): 1068-1072.