J. Vis. Commun. Image R. 60 (2019) 44–50

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

# Research on automatic target detection and recognition based on deep learning ☆

Jia Wang [a], Chen Liu [b,*], Tian Fu [c], Lili Zheng [c]

[a] School of Automation Science and Electrical Engineering, Beihang University, China
[b] Skyinfo General Aviation (Beijing) Technology Co., Ltd, China
[c] Institute of Unmanned System, Beihang University, China

ARTICLE INFO

*Article history:*
Received 1 December 2018
Revised 10 January 2019
Accepted 10 January 2019
Available online 11 January 2019

*Keywords:*
Image processing
Target detection
Target recognition
In-depth learning

ABSTRACT

With the development of computer technology, the related achievements of image processing have been applied. Among them, the results of automatic target detection and recognition are widely used in the fields of reconnaissance, early warning and traffic control with the application of UAV. But now, the research of automatic target detection and tracking is becoming smaller and smaller. The original automatic target detection and recognition algorithm seems to be inadequate. The bottleneck of low-level feature design and optimization makes the accuracy and efficiency of automatic target detection inefficient. Therefore, based on in-depth learning, this paper establishes a method to automatically learn effective image features from images to achieve automatic target detection. Through the simulation of target detection in VEDAI database. The results show that the recognition rate of the proposed model is more than 95%. The results show that the proposed method can realize the automatic detection and recognition of targets very well.



## 1. Introduction

With the development of computer application technology and image recognition technology, people can realize the visual effect only existing in organisms by computer. Nowadays, computer vision research has become an interdisciplinary research hotspot. The main purpose is to enable the computer to make correct and meaningful judgments about the objects and scenes according to the images captured by the image collector, and the automatic detection and recognition of the objects is one of the research directions. Nowadays, the research results of automatic target detection and recognition are widely used in vehicle detection and recognition in intelligent transportation system. Intelligent Transportation System (ITS) refers to the effective application of advanced computer processing technology, information technology, data communication transmission technology and electronic automatic control technology to the traffic management system, which organically links people, vehicles and vehicles into a continuous system. So that the vehicle can travel safely and freely on the road depending on its own intelligence.

For automatic target tracking, the main related technologies include the following three points: real-time detection of target foreground, accurate target recognition and reliable target tracking. However, there are many difficulties in the actual tracking of moving objects: (1) Detection of moving objects is the premise of tracking. In addition to overcoming the disturbance of light mutation and tree disturbance, detection of moving objects in static background also needs to ensure the real-time performance of detection. (2) For the tracking of a specific target, the motion characteristics of the tracking object are mostly non-linear, so it is difficult to achieve the control only by manual operation of the staff. It can be considered to train a target recognizer to complete the tracking. (3) The background of the moving target tracking process is constantly changing. It is unavoidable that the occlusion of the target and the change of the attitude of the target will occur, which will make the tracking unstable and even lead to the loss of the tracking target.

At present, the main technical methods of automatic target recognition are statistical pattern recognition method [1,2], fuzzy pattern recognition method [3,4], artificial neural network method [5,6]. Statistical pattern recognition method is fundamentally to calculate the probability distribution function of all kinds of targets directly by using the distribution characteristics of all kinds of targets to achieve classification and recognition. Its basic technologies

☆ This article is part of the Special Issue on TIUSM.

* Corresponding author.
*E-mail address:* liu513@vip.126.com (C. Liu).

https://doi.org/10.1016/j.jvcir.2019.01.017
1047-3203/

include clustering analysis, discriminant domain interface method, statistical judgment, etc. Fuzzy pattern recognition is based on the theory and method of fuzzy mathematics to solve the problem of target recognition. The effectiveness of this kind of method mainly depends on whether the membership function is good or not. This method can be roughly divided into the maximum membership principle direct recognition method and the proximity principle indirect classification method. Artificial neural network (ANN) is a nonlinear dynamic system connected by a large number of neurons. It has strong self-learning ability and fault-tolerance. There are many kinds of models of ANN. BP neural network model [5–7] is one of the most widely used networks in pattern recognition. It uses given samples. During the training process, the weights and thresholds between the internal network layers are constantly revised, which means that the actual output and the expected output fall within a certain range of errors.

Along with the development of neural network research, various recognition technologies based on neural networks have emerged, among which in-depth learning is one of them. In 2006, Hinton and Salahutdinov of the University of Toronto published a paper in the Journal Science [8], which made a great breakthrough in deep learning and feature learning. The main idea of this paper is to train the neural network layer by layer unsupervised. Such a mechanism can learn different levels of feature expression. Each layer of feature expression is obtained by the previous expression transformation, so that all the layers are superimposed to form a deep neural network. The pre-trained parameters of each layer are used as the initialization parameters of the whole neural network, which is called Deep Boltzmann Machine (DBM) [9,10]. In-depth learning is to learn multi-level feature transformation at the same time, through which we can get better high-level features.

Deep learning develops from unsupervised multi-level pretraining initialization to direct multi-level training of deep convolution network. In the task of image classification, deep convolution network has achieved surprising results, which has also brought about the explosion of deep learning in the field of image [11–15]. As the effect of deep convolution network becomes more and more prominent, it is widely used in image tasks. Deep convolution was used by most teams in the 2013 Image Net Large-scale Visual Recognition Challenge Competition. Zeiler [16] had an error rate of 11.20% in the first five candidates for classification tasks, but because most teams used deep convolution networks, the gap between the first and second places was not large.

After the British successfully developed the world's first UAV in 1917, with the development of technology, UAV has become an indispensable part of modern intelligent transportation and modern military. Nowadays, more mature systems aiming at target detection, recognition and tracking, such as the quasi-real-time tracking system of Cornell University, USA, can be used to detect and track multiple moving targets. A detection algorithm based on model feature matching is adopted, which combines image registration and moving target location estimation. When the target is deformed, it can still keep the detection of the target, while retaining the information of target motion and size change. By comparing with historical information, whether the target is moving or not, or is temporarily occluded by other objects, it can achieve target detection and tracking. With the maturity of UAV system, more and more research results on automatic target recognition of UAV appear [22,23].

Automatic target recognition and detection based on UAV is a content of image processing, but the interference of environment in the process of UAV shooting is greater than that of ordinary images, and the target detection of UAV involves the detection of small targets. Therefore, in order to accurately monitor and track the target automatically, a new model is needed. In this paper, a target detection and recognition method based on depth learning is established by comparing and analyzing the results of light, noise and background of the image. The simulation results of the public database show that the recognition rate of the vehicle reaches 95.8%, and the recognition effect is very good.

## 2. Methods

### 2.1. Background modeling method

Target detection and processing is the key link of target tracking, target recognition and other operations [23–29]. Background subtraction is widely used in the detection of moving targets in static environment. Background subtraction, as its name implies, is to abandon the pixels that are different from the background model and retain the pixels that are similar to the background model. Therefore, the main steps of background subtraction include: background modeling, pixel comparison, deletion of pixels, and the most important of which is background modeling, that is, to establish a mathematical template for the image background. By calculating, the mathematical template can be closer to the background, and then the target foreground of the current frame can be extracted. In fact, even in the static background, there are still some interference factors such as light, shadow and branch disturbance [30–34]. In order to improve the accuracy of target detection and overcome the small interference, the background can be modeled by using Gauss distribution. The basis of the Gauss model is the Gauss distribution, which is used to model every pixel in the image. At the present moment, we first judge whether a pixel is a background, then replace the old Gauss model with the gray value of the pixel, and finally use the new background model to continue the subsequent judgment. According to the number of Gaussian distributions used, the Gaussian model can be divided into two categories: single Gaussian model and mixed Gaussian model [17–19]. Single Gauss model is used in background modeling and only uses one Gauss distribution. Its updating is mainly reflected in the modification of relevant parameters. The advantage of this model is that it has less computation and fast detection speed. It can be used to detect targets in the environment with simple background changes. However, the single Gaussian model cannot deal with target detection in background changing environment, so this paper chooses the mixture Gaussian modeling method. Before input video frames, K Gaussian distributions are used to represent the characteristics of each pixel in the current frame. Then the new frame image is matched with the mixture Gaussian model of each pixel in the current frame. If successful, the point is determined as the background point, and the mixture Gaussian model is updated, otherwise it is the foreground point.

If a pixel on an image on T-Time is represented by $X_t$, its probability density function P can be calculated by weighted summation method.

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t}\eta(X_t, \mu_{i,t}, \sigma_{i,t}) \tag{1}$$

The above $\omega_{i,t}$ is the weights of the i Gauss distribution at t-time, $\eta(X_t, \mu_{i,t}, \sigma_{i,t})$, $\mu_{i,t}$ and $\sigma_{i,t}$ they are probability density function, mean value and variance respectively. The probability density function can be calculated by the following methods:

$$\eta(X_t, \mu_{i,t}, \sigma_{i,t}) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{(X_i - \mu_{ij})^T}{2\sigma_{ij}(X_i - \mu_{ij})}\right) \tag{2}$$

Match the new input pixel gray value with the established K Gaussian distributions according to the following formula

$$|X_t - \mu_{i,t-1}| \leqslant D\sigma_{i,t-1} \tag{3}$$

The D in the above is a confidence parameter, and the range of values is D ∈ [2 3], after matching, two results will be produced: (1) If no Gaussian model matching the new input pixel values is found, the lowest weight Gaussian distribution in the Gaussian model is deleted, and then a new Gaussian distribution is established according to the new pixel values. The new pixel values are taken as the mean, and a larger variance and a smaller weight are given. The number of updated Gauss distributions is still K. (2) If the matching results of formula (3) are consistent, the parameters of formula (1) are updated according to formula (4).

$$\begin{cases} \omega_{i,j} = (1-a)\omega_{i,t-1} + \alpha M_{i,t} \\ \mu_{i,t} = (1-\beta)\mu_{i,t-1} + \beta X_{i,t} \\ \sigma^2_{i,t} = (1-\beta)\sigma^2_{i,t-1} + \beta(X_{i,t}-\mu_{i,t})^T(X_{i,t}-\mu_{i,t}) \\ \beta = \alpha\eta(X_t, \mu_{i,t}, \sigma_{i,t}) \end{cases} \tag{4}$$

M in formula (3) is a binary value. If M = 1, it represents the target, otherwise it represents the background. $\alpha$ is the learning rate of the Mixture Gauss Model, representing the update speed of the background and $\beta$ is the update factor.

### 2.2. Feature extraction method

The extraction of target feature information can lay a foundation for subsequent recognition and tracking operations, so feature selection is very important [35,36]. The following characteristics are considered in feature selection: intuition, simplicity, specificity and invariance. The feature extraction in this paper mainly includes RGB color weighted histogram and Sobel [20,21] edge weighted histogram.

The color histogram needs to calculate the color in each region, so it is necessary to quantify the three components of the color space, and divide the whole color space into some cells, which constitute the color feature interval. This paper chooses the target RGB.

Color histogram is one of the features of target extraction, where R channel is quantized to L level and G and B channel is quantized to L level.

L is a quarter of the grade, so the number of feature intervals is $L \times (L/4)^2$.

In the initial frame, the minimum region and centroid coordinates of the target are obtained by target detection. The number of pixels in the target region is n, the set of pixels in the target region is $\{x_i\}, i = 1, 2, ..., n$, the coordinates of the target central region is $x_0$, $\{u = 0, 1, 2, ..., L-1\}$ is the eigenvalues of the target template on R component, and on G component and B component eigenvalue is $\{u = 0, 1, 2, ..., L/4-1\}$. In order to make the extracted color features play a more beneficial role, the influence of the pixels nearer to the tracking frame is weighted (i.e. the corresponding probability value is given), and the weighted histogram of the kernel function is established. Therefore, the discrete density function model of the eigenvalues of each component, i.e. the color weighted histogram (5):

$$\begin{cases} p_c(u) = \sum_{i=1}^{n}\left(\left|\frac{(x_i - x_0)}{\omega_t}\right|^2\right) & b(x_i) = u \\ 0 & \text{others} \end{cases} \tag{5}$$

In formula (5), K is defined as a kernel function of the size of each pixel's weight, so that the closer the pixel's weight is to the center of the target, the higher the pixel's weight is, and the effect of the surrounding environment on the target area is reduced. $\omega_t$ is the detection window bandwidth. $b(x_t)$ Represents the eigenvalues at the pixels $x_t$.

When the attitude of the target changes, the color feature still has good robustness, but when the spatial structure of the target varies greatly, the contribution value of the color feature to the target feature is smaller, so it is not enough to rely solely on the single color feature. Because the gray levels of different images are different, there will be some relatively obvious edge data at the edge, so the edge features can fill the deficiencies caused by the color features. In practical edge feature extraction, only the first and second derivatives are needed because of noise. The common first and second edge extraction operators are Prewitt operator, Sobel operator and Canny operator. Sobel is an edge detection operator based on first derivative calculation. It uses discrete difference operator to approximate the gradient of image brightness function. Because local average is used in this operator, the influence of noise can be effectively suppressed. In the Obel operator, the edge gradient values can be divided into two categories: transverse and longitudinal. The two gradient values have their own customary templates, and the transverse gradient amplitude customary template matrix use $G_x$ is expressed as follows:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{6}$$

The conventional template matrix of longitudinal gradient amplitude use $G_y$ is expressed as follows:

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{7}$$

When the first frame is detected, the image is first binarized, and then the gradient magnitude G at the edge of the target image at the edge direction $\theta_e$ are calculated by formula (8) with the template $G_x$ and $G_y$.

$$\begin{cases} G = \sqrt{G_x^2 + G_y^2} \\ \theta_e = \arctan\left(\frac{G_y}{G_x}\right) \end{cases} \tag{8}$$

The range of the edge direction is as follows: $\theta_e \in (-90^\circ, 90^\circ)$.

In order to enhance the anti-jamming ability of edge histogram, morphological operations such as filtering, dilation and thinning can be performed on it, so that the processed edge image can not only filter out the weak and false edges in the edge, but also extract the real edge. When the first frame is detected, the number of edge pixels extracted by edge detection in the target area is $n'$, the set of edge pixels is $\{x_g\}$, $g = 1, 2, ..., n'$, and the coordinates of the target central area is $x_0$. If the edge direction is divided into equal parts $M_e$, the feature value of the target template in each equal edge direction is $\{u = 0, 1, ..., 180/M_e - 1\}$. Based on the above variables, the distribution function of the edge weighted histogram of the target is as follows:

$$\begin{cases} p_e(u) = G\sum_{i=1}^{n} K\left(\left|\frac{(x_g - x_0)}{\omega_t}\right|^2\right) & b(x_g) = u \\ 0 & \text{others} \end{cases} \tag{9}$$

Formula (9) $b(x_g)$ denotes the eigenvalues at the pixels $x_g$.

### 2.3. Deep learning method

Compared with the traditional methods, it requires high priori knowledge of the target and has poor generalization ability of the model. At present, image target recognition based on depth learning has been widely used. Deep learning originates from neural networks. As one of the deep learning networks, convolutional neural network has become a research hotspot in the field of image

target recognition. The basic structure of convolution neural network is composed of feature extraction layer and feature mapping layer. As far as feature extraction is concerned, the local receptive field of the upper layer connects the input of the next layer of neurons, and the regional feature is extracted by the neuron. For the feature mapping layer, each layer of the network consists of multiple feature mappings, and the weights of the neurons in each feature mapping plane are equal, that is, weight sharing. The activation function is Re LU function, which has displacement invariance. The network parameters are reduced while sharing weights. After each convolution layer, a computational layer is connected to compute the local mean and take the second value of the feature. Convolutional neural network can be applied to two-dimensional graphics with displacement change, scale location and other forms of distortion invariance. The learning of feature detection layer is based on training data, and its learning method avoids explicit feature extraction, which is an implicit feature extraction method. The existence of network weight sharing enables the network to learn better in parallel, which is also a major advantage of convolutional neural networks compared with fully connected networks. This structure of local weight sharing makes convolutional neural networks have unique advantages in image processing, which greatly reduces the complexity of the network.

The basic network structure of convolution neural network can be divided into five parts: input layer, convolution layer, pooling layer, full connection layer and output layer. The following sections of the network are described in detail.

Input layer: The convolution input layer can directly act on the original input data. For the input image, the input data is the pixel value of the image.

Convolution Layer: The local features of a natural image can be identical or similar to those of other local regions, which indicates that the features learned in one region are also applicable to other regions. For convolution neural networks, the output of the convolution layer is convoluted by the input characteristic graph of the filter and the previous layer, where each filter generates an output characteristic graph. In this paper, the input image $X$ is represented, the $k$ feature graph of the $i$ layer is represented by $A_i^k$, and the characteristics of the $k$ filter of the $i$ layer are determined by the weight matrix $W_i^k$ and bias $b_i^k$. Then the $k$ feature map of the $i$ layer can be obtained from the following formula (10):

$$A_i^k = f(W_i^k \otimes A_{i-1}^k + b_i^k) \tag{10}$$

Pooling layer: Pooling layer is a downsampling process, which integrates the output of adjacent neurons in the same feature map. A downsampling mechanism is introduced. The principle of this aggregation is that the pixels in each adjacent region of the image have a large similarity, which can be described by calculating the maximum value of the region as the sampling value and the average value as the sampling value.

Formula (11) can be used to represent the process of the pooling layer. The neurons in this layer adopt a down ( ) downsampling function, which is used to maximize or average pooling the feature map.

$$A_i^k = f(down(A_{i-1}^k) + b_i^k) \tag{11}$$

Full Connection Layer: It can contain multiple Full Connection Layers, which is actually the hidden layer part of the Multilayer Perceptron. Generally, the ganglion points in the posterior layer are connected with each ganglion point in the preceding layer, and there is no connection between the neuron nodes in the same layer. Each layer of neuron nodes propagates forward through the weights on the connecting line, and the weights are combined to get the input of the next layer of neuron nodes.

Output layer: The number of ganglion points in the output layer is determined according to specific application tasks. If it is a classification task, the output layer of convolutional neural network is usually a classifier.

## 3. Experiment

In order to verify the algorithm, we take aerial photographs and test flights of fixed-wing UAV in various scenarios. During aerial photography, we design various possible factors such as different shooting angles, different background environments, different illumination conditions, different target size and so on. The static and moving vehicles on roads and bridges are taken as the detection targets. There are various buildings which are easy to cause interference in UAV vision. The overall contrast of images is not high, the edges of bridges are complex and the targets are far away. While the vehicle moves, the UAV also moves at the same time, and the field of view changes more complex.

In order to further validate and compare the algorithm horizontally, the algorithm is also validated in the relevant scenarios of VEDAI (VEDAI Detection on Aerial Imagery) database. There are 1210 visible light pictures with 512 * 512 resolution in VEDAI database. In the verification process, we use the 10-fold cross-validation method to allocate test sets and training sets.

In the process of analyzing the data simulation results, we use Precision-Recall method to analyze the results, in which Precision calculation method is:

$$\Pr \textit{ecision} = t_p/t_p + f_p \tag{12}$$

Recall is calculated by:

$$\text{Re } \textit{call} = t_p/t_p + f_n \tag{13}$$

Among them, $t_p$ is the number of correct detection targets, $f_p$ is the number of false detection targets, and $f_n$ is the number of missing targets.

## 4. Results and discussion

In the process of target detection and recognition using UAV, the exposure time will be unstable due to the influence of UAV's shooting angle and light. Therefore, the same target may be considered as different targets or even missed. Therefore, it is necessary to analyze the pictures of the target in different light. Fig. 1 shows the results of the original image and the light-enhanced image. The first line of Fig. 1 shows the original image of the bridge, the exposed image of the bridge, the original image of the road and the exposed image of the road respectively. The second line shows the corresponding histogram. Histogram is originally used as a statistical report graph, which shows the distribution of data by a series of vertical stripes or lines with different heights. A statistical report graph in which data distribution is represented by a series of vertical stripes or lines of varying heights. In this paper, histogram is used to describe the distribution of image features. From the distribution of histogram, there are obvious changes in image features before and after image enhancement. Therefore, in the research of target detection and recognition using UAV, it is necessary to analyze the invariance of image brightness recognition, so that the target image is not sensitive to light, otherwise it is difficult to achieve target tracking.

The image taken by UAV is a color image. In order to extract and analyze the features, the first step is to gray the color image and analyze the image on the basis of gray level. As shown in Fig. 2, the first row is the original color image taken by UAV, and the second row is the gray level image. However, the visual effect of the image becomes worse, but the characteristics of the vehicle in

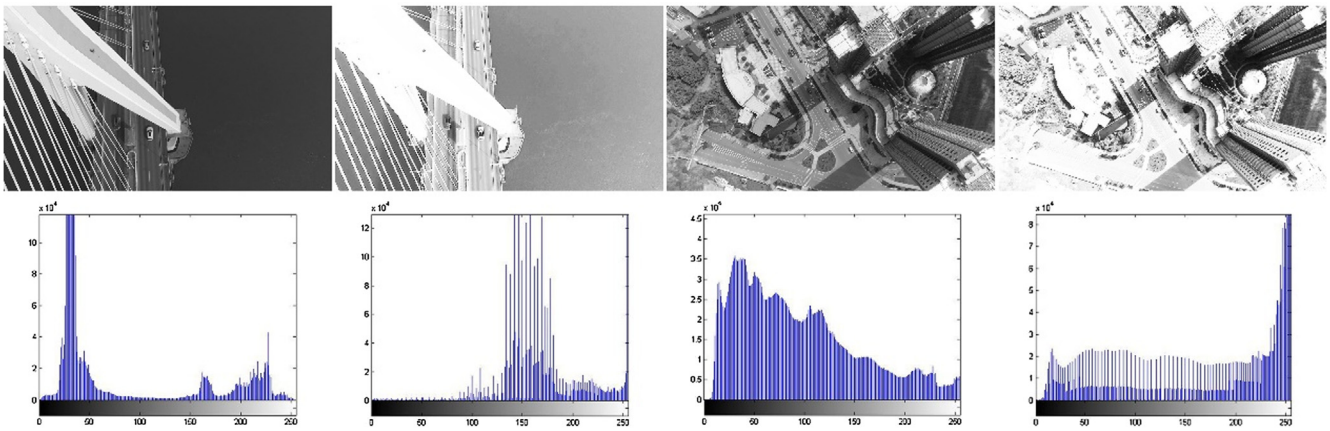

**Fig. 1.** Image histogram contrast under image enhancement.

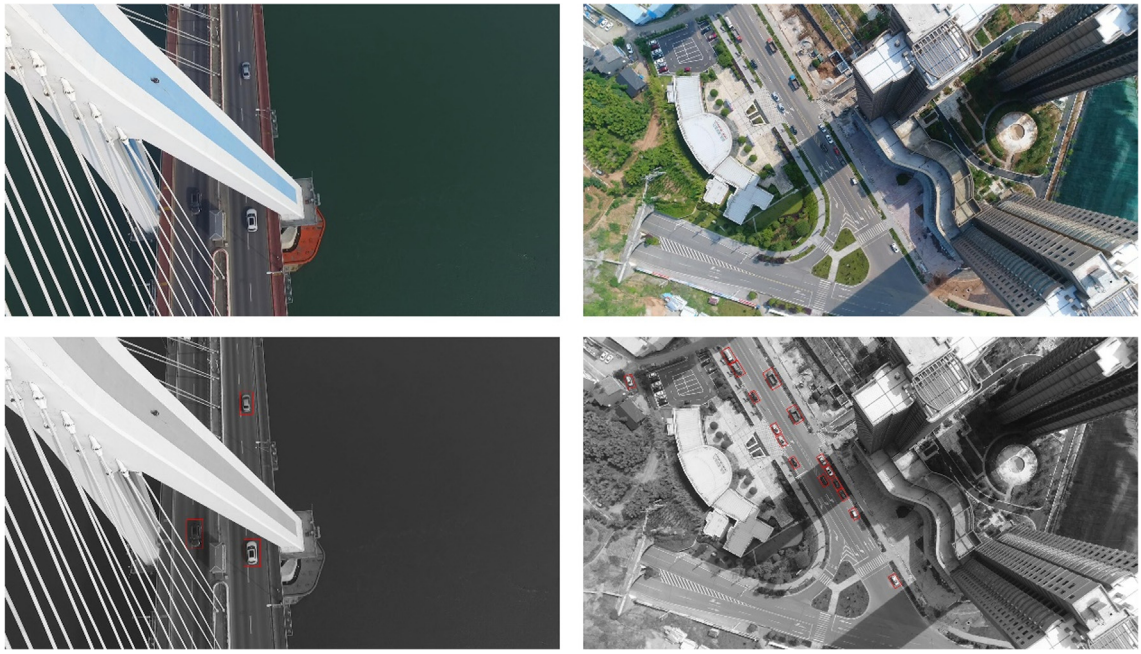

**Fig. 2.** Grayscale comparison.

the image do not change with the change of gray whiteness. Taking the cell image as an example, the color cell image is 1080 * 1920 * 3, and the gray image becomes 1080 * 1920, which is only one third of the original image. This reduces the difficulty of feature extraction.

For various reasons, the image will inevitably be disturbed by noise when taking pictures by UAV. Therefore, when automatic target detection is carried out, noise analysis of the image is needed. Fig. 3 is the contrast result after adding salt and pepper noise to the original gray scale image. From the result of Fig. 2, it can be seen that after adding salt and pepper noise, the image needs to be analyzed. The results of quantitative analysis by formula (11) and (12) are shown in Fig. 4.

From the results of Fig. 4, it can be seen that the Precision-Recall parameters are obviously compared before and after adding noise. After adding noise, the recognition rate is significantly lower than before adding noise.

From the above experimental results, it can be seen that the process of target detection and recognition is different from that of image recognition. The process of automatic target detection changes with the change of environment. Therefore, the process of processing is more complex, which requires in-depth study on the basis of image processing. This paper establishes an automatic detection and recognition model of target image based on in-depth learning. In order to prove the validity of the model used in this paper, the data of VEDAI, a public database, is used to carry out experiments. The experimental results show that the recognition rate of automatic target detection reaches 95.8%, which is much higher than the existing recognition results, which fully proves the validity of the model.

## 5. Conclusion

Target tracking is an important research field of machine vision direction. Most of the current studies are based on the theoretical environment, which is close to the target. Target occupies a large proportion in the image field of view, and the edge contour is clear and distinguishable. In practical application, especially in the initial stage of tracking, because of the long distance and the small

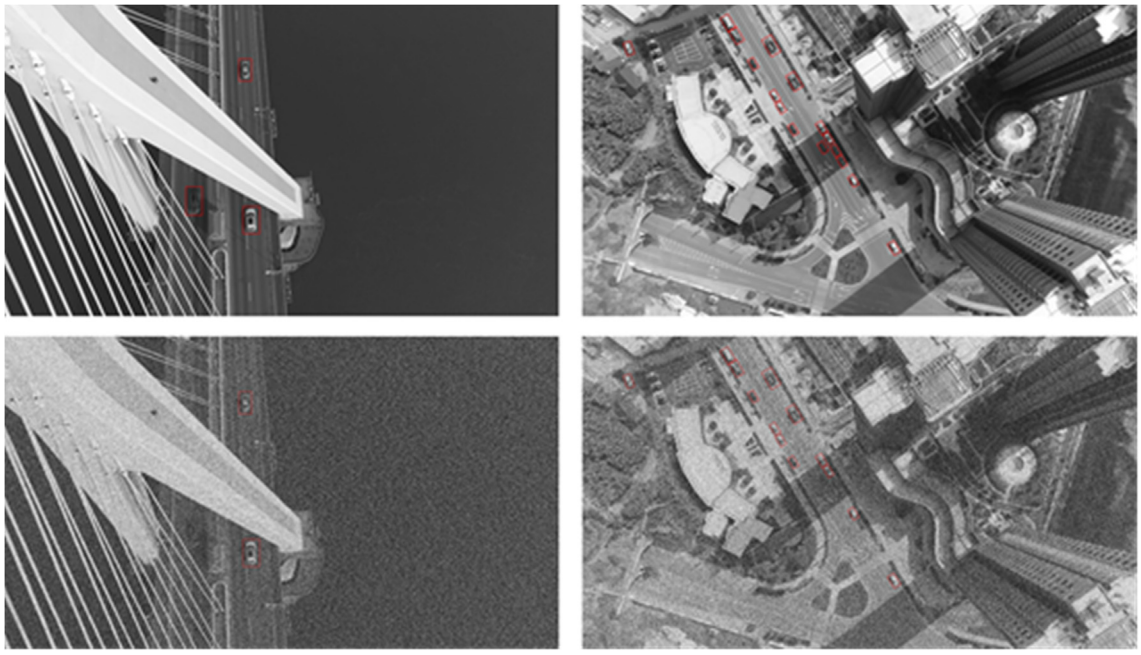

**Fig. 3.** Contrast before and after adding noise.

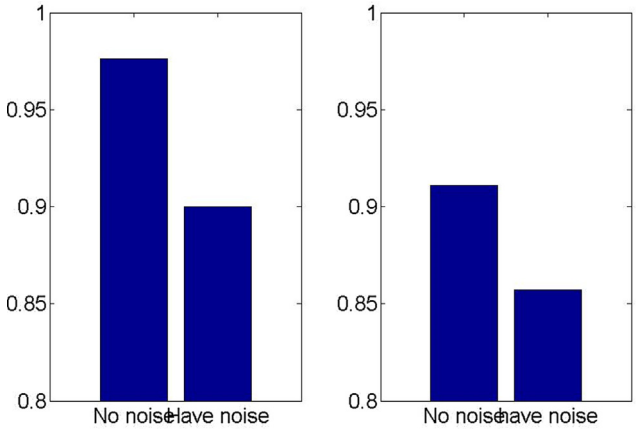


**Fig. 4.** Comparison of results before and after adding noise.

size of the target in the field of view, the edge information is less or directly unavailable, which will bring great challenges to target tracking. At the same time, there are many interferences in the field of view in complex background environment. How to eliminate the interferences will be an important prerequisite for accurate target recognition. In addition, the gray value of the image will change correspondingly due to the change of illumination intensity in the outdoor environment. General moving target detection and tracking algorithms are only suitable for static or simple moving background, but not for UAV automatic target detection and recognition. This paper establishes an automatic target detection and recognition model based on depth learning method, and tests the data of VEDAI. The experimental results show that the recognition rate of automatic target detection reaches 95.8%, which is much higher than the existing recognition results, which fully proves the validity of the model.

## Conflict of interest

There is no conflict of interest.

## References

[1] T.W. Jewitt, Automatic target detection, acquisition, and tracking via hierarchical pattern recognition, Proc. SPIE - Int. Soc. Opt. Eng. 1305 (1990) 75–86.
[2] S. Greenberg, R. Yehezkel, Y. Gurevich, et al., NLEBS: automatic target detection using a unique nonlinear-enhancement-based system in IR images, Opt. Eng. 39 (5) (2000) 1369–1376.
[3] S.M. Yamany, A.A. Farag, S.Y. Hsu, Fuzzy hyperspectral classifier for automatic target recognition (ATR) systems, Pattern Recogn. Lett. 20 (11–13) (1999) 1431–1438.
[4] I. Valova, G. Milano, K. Bowen, et al., Bridging the fuzzy, neural and evolutionary paradigms for, Appl. Intell. 35 (2) (2011) 211–225.
[5] Junwei Han, Dingwen Zhang, Gong Cheng, Lei Guo, Jinchang Ren, Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning, IEEE Trans. Geosci. Remote Sens. 53 (6) (2015) 3325–3337.
[6] Q. Futong, Y. Lihua, W. Shouhong, Performance evaluation in automatic target recognition using BP neural network, Comput. Eng. Appl. 46 (5) (2010) 148–152.
[7] L. Zhang, Y. Gao, R. Zimmermann, Q. Tian, X. Li, Fusion of multichannel local and global structural cues for photo aesthetics evaluation, IEEE Trans. Image Process. 23 (3) (2014) 1419–1429.
[8] Dingwen Zhang, Deyu Meng, Junwei Han, Co-saliency detection via a self-paced multiple-instance learning framework, IEEE Trans. Pattern Anal. Mach. Intell. 39 (5) (2017) 865–878.
[9] Kostakis J, Cooper ML, Green, TJ, et al. SPIE Proceedings [SPIE AeroSense '99 - Orlando, FL (Monday 5 April 1999)] Automatic Target Recognition IX - Multispectral sensor fusion for ground-based target orientation estimation: FLIR, LADAR, HRR 1999;3718:14–24.
[10] M. Grady, D. Campbell, K. Macleod, et al., Evaluation of a blood glucose monitoring system with automatic high- and low-pattern recognition software in insulin-using patients: pattern detection and patient-reported insights, J. Diabetes Sci. Technol. 7 (4) (2013) 970.
[11] Gong Cheng, Peicheng Zhou, Junwei Han, Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images, IEEE Trans. Geosci. Remote Sens. 54 (12) (2016) 7405–7415.
[12] Tuo Zhang, Lei Guo, Kaiming Li, Changfeng Jing, Yan Yin, Dajiang Zhu, Guangbin Cui, Lingjiang Li, Tianming Liu, Predicting functional cortical ROIs via DTI-derived fiber shape models, Cereb. Cortex 22 (4) (2012) 854–864.
[13] L. Zhang, Y. Xia, K. Mao, H. Ma, Z. Shan, An effective video summarization framework toward handheld devices, IEEE Trans. Ind. Electron. 62 (2) (2015) 1309–1316.
[14] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[15] C.Y. Zhang, C.L.P. Chen, M. Gan, et al., Predictive deep Boltzmann machine for multiperiod wind speed forecasting, IEEE Trans. Sustain. Energy 6 (4) (2017) 1416–1425.
[16] C.L.P. Chen, C.Y. Zhang, L. Chen, et al., Fuzzy restricted Boltzmann machine for the enhancement of deep learning, IEEE Trans. Fuzzy Syst. 23 (6) (2015) 2163–2173.

[17] J. Li, X. Mei, D. Prokhorov, et al., Deep neural network for structural prediction and lane detection in traffic scene, IEEE Trans. Neural Networks Learn. Syst. 28 (3) (2017) 690–703.
[18] H. Pan, B. Wang, H. Jiang, Deep learning for object saliency detection and image segmentation, IEEE Trans. Neural Networks Learn. Syst. 27 (6) (2015) 1135–1149.
[19] X. Pan, P. Rijnbeek, J. Yan, et al., Prediction of RNA-protein sequence and structure binding preferences using deep convolutional and recurrent neural networks, BMC Genomics 19 (1) (2018) 511.
[20] L. Zhang, M. Song, Q. Zhao, X. Liu, J. Bu, C. Chen, Probabilistic graphlet transfer for photo cropping, IEEE Trans. Image Process. 22 (2) (2013) 802–815.
[21] W. Tan, C. Zhao, H. Wu, Intelligent alerting for fruit-melon lesion image based on momentum deep learning, Multimedia Tools Appl. 75 (24) (2016) 1–21.
[22] J. Donahue, L.A. Hendricks, M. Rohrbach, et al., Long-term recurrent convolutional networks for visual recognition and description, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2014) 677–691.
[23] Junwei Han, Dingwen Zhang, Hu. Xintao, Lei Guo, Jinchang Ren, Feng Wu, Background prior-based salient object detection via deep reconstruction residual, IEEE Trans. Circuits Syst. Video Technol. 25 (8) (2015) 1309–1321.
[24] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European Conference on Computer Vision, 2014, pp. 818–833.
[25] J. Jiang, J. Yang, Y. Cui, et al., Mixed noise removal by weighted low rank model, Neurocomputing 151 (2015) 817–826.
[26] L. Zhang, R. Hong, Y. Gao, R. Ji, Q. Dai, X. Li, Image categorization by learning a propagated graphlet path, IEEE T-NNLS 27 (3) (2016) 674–685.
[27] Ying Wang, Solving multi-instance visual scene recognition with classifier ensemble based on unsupervised clustering, Appl. Mech. Mater. 415 (2013) 338–344.
[28] X. Liu, H. Fu, Y. Jia, Gaussian mixture modeling and learning of neighboring characters for multilingual text extraction in images, Pattern Recogn. 41 (2) (2008) 484–493.
[29] Dingwen Zhang, Junwei Han, Chao Li, Jingdong Wang, Xuelong Li, Detection of co-salient objects by looking deep and wide, Int. J. Comput. Vision 120 (2) (2016) 215–232.
[30] P. Bo, J. Yang-Sheng, P.U. Yun, A pavement cracking image recognition algorithm based on bi-layer connectivity checking, J. Highway Transport. Res. Dev. 31 (5) (2014) 21–30.
[31] P. Bixia, Study on method of pre-processing on vehicle license plate image recognition system, Eng. J. Wuhan Univ. 39 (3) (2006) 131–134.
[32] A. Sinha, Y. Barshalom, Optimal cooperative placement of UAVs for ground target tracking with Doppler radar, Proc. SPIE - Int. Soc. Opt. Eng. 5429 (2004) 95–104.
[33] Silva H, Almeida JM, Lopes F, et al. [IEEE OCEANS 2016 MTS/IEEE Monterey - Monterey, CA, USA (2016.9.19–2016.9.23)] OCEANS 2016 MTS/IEEE Monterey - UAV trials for multi-spectral imaging target detection and recognition in maritime environment; 2016:1–6.
[34] Junwei Han, King Ngi Ngan, Mingjing Li, Hong-Jiang Zhang, Unsupervised extraction of visual attention objects in color images, IEEE Trans. Circuits Syst. Video Technol. 16 (1) (2006) 141–145.
[35] L. Zhang, M. Song, Y. Yang, Q. Zhao, C. Zhao, N. Sebe, Weakly supervised photo cropping, IEEE Trans. Multimedia 16 (1) (2014) 94–107.
[36] Junwei Han, Xiang Ji, Hu. Xintao, Dajiang Zhu, Kaiming Li, Xi Jiang, Guangbin Cui, Lei Guo, Tianming Liu, Representing and retrieving video shots in human-centric brain imaging space, IEEE Trans. Image Process. 22 (7) (2013) 2723–2736.