EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Application research of image recognition technology based on CNN in image location of environmental monitoring UAV

Kunrong Zhao[1], Tingting He[2], Shuang Wu[3], Songling Wang[1], Bilan Dai[1], Qifan Yang[2] and Yutao Lei[1*]

Abstract

UAV remote sensing has been widely used in emergency rescue, disaster relief, environmental monitoring, urban planning, and so on. Image recognition and image location in environmental monitoring has become an academic hotspot in the field of computer vision. Convolution neural network model is the most commonly used image processing model. Compared with the traditional artificial neural network model, convolution neural network has more hidden layers. Its unique convolution and pooling operations have higher efficiency in image processing. It has incomparable advantages in image recognition and location and other forms of two-dimensional graphics tasks. As a new deformation of convolution neural network, residual neural network aims to make convolution layer learn a kind of residual instead of a direct learning goal. After analyzing the characteristics of CNN model for image feature representation and residual network, a residual network model is built. The UAV remote sensing system is selected as the platform to acquire image data, and the problem of image recognition based on residual neural network is studied, which is verified by experiment simulation and precision analysis. Finally, the problems and experiences in the process of learning and designing are discussed, and the future improvements in the field of image target location and recognition are prospected.

Keywords: UAV, Image recognition, CNN, Residual network

1 Introduction

"Nowadays, the development of drone technology is very rapid. This paper uses the drone to collect images and uses machine vision to identify and locate images to achieve environmental monitoring." Research on image location and recognition technology usually refers to identifying potential targets and making meaningful judgments based on the obtained image data. Deep learning technology has been gradually applied in various areas of people's lives, such as speech recognition, automatic translation, image recognition, personalized recommendation, and so on. Convolution neural network is a representative of deep learning technology, which is often used in the field of image recognition.

Artificial neural network (ANN), also known as simulated neural network, refers to an interconnected artificial neuron group that uses mathematical or computational models for information processing [1]. The learning process used to train artificial neural networks is itself a statistical technique, and White H [2] proposed some potentially useful new training methods for artificial neural networks. Zhou [3] combined artificial neural networks with remote sensing to improve the image classification performance of fragmented and heterogeneous landscapes in urban environments. Xie Y [4] based on the analysis of biological materials, using computer image analysis and artificial neural networks and selecting a set of features, describes the physical parameters that allow identification of the species. The features obtained from the image are used as learning data for the artificial neural network to train the multilayer perceptron network. Bartolome L S [5] introduced an automated vehicle test system. The system utilizes images captured from the entrance to the parking area for character recognition. The extracted images are converted into digital forms designed by researchers to

* Correspondence: leiyutao@scies.org
[1]South China Institute of Environmental Sciences, MEP, Guangzhou, Guangdong, China
Full list of author information is available at the end of the article

accommodate the requirements of artificial neural networks. Each character is then extracted from the numbered boards to produce their unique characteristics. Image denoising is a challenging task in digital image processing research and applications. Al-Sbou Y A [6] proposed a detailed performance evaluation using a neural network as a noise reduction tool. This includes using the mean and median statistical functions to calculate the output pixels of the training pattern of the neural network, using a portion of the degraded image pixels to generate the system training pattern. Image classification is one of the typical computational applications widely used in the medical field. Hemanth D J [7] solved the high convergence time and its inaccuracy caused by high-precision ANN by proposing two new neural networks, namely improved backpropagation neural network (MCPN) and improved Kohonen neural network (MKNN). Kouamo S [8] conducted experimental research on some image compression techniques based on artificial neural networks, and proposed a new hybrid method based on the use of multilayer perceptron, which combines layering and adaptive schemes. Crispim-Junior C F [9] studied the use of morphological and kinematic image features processed by artificial neural networks (ANN) to automatically detect and score behavioral events on digital video samples taken from rats placed in open fields and extract image features.

In recent years, drone technology has gained great popularity. This technology has witnessed enhanced capabilities in terms of payload, longer range operability, and hover stability [10]. Jin Y [11] discussed the difficulty of crossover drone image estimation based on theoretical analysis and image description and verified the feasibility of the rule through simulation examples. The results show that the rules provide a manual control method for astronauts. Jonghwan B [12] designed an object with four legs to be picked up during flight, using a stream image processing to verify the design of the walking drone. Srikudkao B [13] introduced the configuration and integration of various sensors based on a series of image processing techniques and their data management schemes, simulating the tasks required to estimate critical flood-related parameters. The experimental results can provide a basis for determining its potential applications in flood warning and forecasting systems and the problems that need to be addressed. Maria G [14] proposed a car inspection system prototyped in an experimental project. The video stream of UAV recorded in the urban environment is analyzed. Lee E J [15] used an UAV-based FIR camera to photograph a sunken hole at an altitude of 5002 m, and integrated classification results based on optical convolution neural network (CNN) and Boosted Random Forest (BRF) with manual features. The proposed ensemble method is successfully applied to downhole data of various sizes and depths in different environments. Skoczylas M [16] designed an autonomous UAV landing system based on CCD camera and image transmission system connected to base station. The system landed in an area of 1 m × 1 m, scanned the captured images, and detected landing marks. Xiong X [17] proposed an automatic view finding scheme that can autonomously navigate the drone to an appropriate spatial location where it can take a photo with the best composition. Oh, Jae Hong [18] used overlapping images captured by UAV to automatically reconstruct power lines in object space. Two overlapping images are selected for epipolar image resampling, and the resampled images and redundant images are extracted. The extracted lines from the polar image are matched together and reconstructed for power line primitives that are noisy due to multiple line matches.

CNN has been widely used in image analysis and speech recognition in recent years. However, the use of CNN for chart classification is still challenging. Luo Z [19] proposed the classification of convolution neural network models based on Ngram block. CNNs enable learning data-driven, highly representative, layered hierarchical image features from sufficient training data [20]. Convolution neural networks have recently shown excellent image classification performance in large-scale visual recognition challenges. Oquab M [21] showed how to use a limited amount of training data to effectively transfer the image representation learned by CNN on large annotated datasets to other visual recognition tasks. Gatys L A [22] used image representation derived from convolution neural networks optimized for object recognition, which makes advanced image information explicit. The results provide new insights into depth image representation for convolution neural networks learning, and demonstrate their potential in advanced image synthesis and operation. Milletari F [23] used convolution neural networks (CNN) to solve problems in the field of computer vision and medical image analysis and proposed a three-dimensional image segmentation method based on volume, complete convolution, and neural network. Zbontar J [24] present a method for extracting depth information from a rectified image pair; it approached the problem by learning a similarity measure on small image patches using a convolutional neural network. Ma L [25] proposed to employ the CNN for the image question answering (QA) task. Their proposed CNN provides an end-to-end framework with convolutional architectures for learning not only the image and question representations but also their inter-modal interactions to produce the answer. Pathak D [26] proposed a method to learn dense pixel labels from image level tags. Each image level label applies constraints on the output markup of a CNN classifier.

In this paper, based on the research status of UAV image processing, an image recognition technology based on CNN is proposed to recognize the images collected by UAV in the process of environmental monitoring. In other words, after the UAV is used to obtain the image, the extracted features are finally identified by CNN to solve the problem, and verified by experimental simulation and precision analysis.

2 Proposed method

2.1 UAV monitoring scheme system

UAV monitoring system includes flight control system, ground station system, and infrared vision system. The flight control system is mainly responsible for controlling the attitude and position of the aircraft; the infrared vision system is mainly responsible for monitoring environmental anomalies; the hand-held remote controller is mainly responsible for the manual control of the position of the UAV; and the ground station system functions are to display the image and video information obtained by the infrared vision system of the UAV and the flight status of the UAV.

The image recognition of the UAV monitoring based on the infrared imaging technology converts the infrared radiation acquired by the infrared detector into a gray value, thereby forming an infrared image. The main purpose is to use the difference in infrared radiation values of the target in the scene to image. The larger the infrared radiation value of the target in the scene, the higher the gray value of the target image corresponding to it, which reflects the brighter the image. Since the monitoring environment is mainly for unknown areas without fixed cameras, a spiral search strategy is generally used, as shown in Fig. 1.

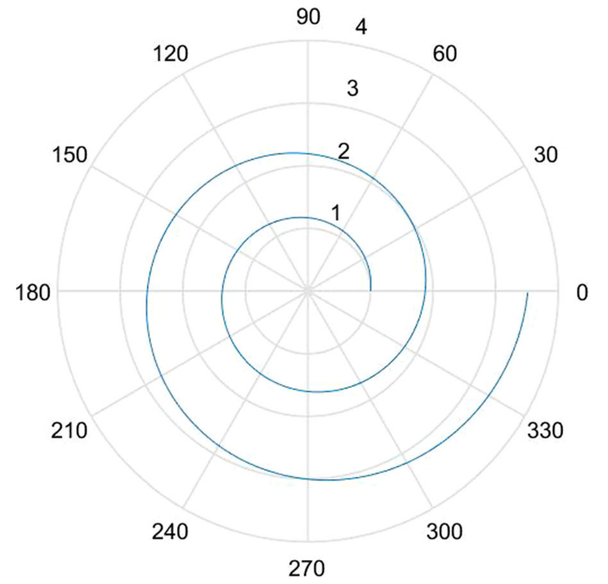

Fig. 1 Search strategy

Assume that the center of the known monitoring area is (x_0,y_0), the flying height of the drone is h, the field of view of the infrared camera is θ, and the coordinates of the generated point on the spiral is (x,y), then the spiral search algorithm is as shown in Formula 1:

$$\begin{cases} r_0 = 2 * K * h * \tan\frac{\theta}{2} \\ \phi = \frac{2\pi}{r_0} * r \\ x = x_0 + r * \cos\phi \\ y = y0 + r * \sin\phi \end{cases} \quad (1)$$

where K is a constant less than 1, r_0 represents the spacing between the tracks, ϕ is the cumulative rotation angle, and r is the current orbit radius.

The monitoring process of the drone can be divided into three phases:

1. The monitoring personnel arrive at the center of the monitoring area (x_0,y_0), and monitor the monitoring area of the drone through the ground station.
2. The monitoring drone independently performs a spiral search according to the designated area, and the density is estimated by the airborne infrared vision system. If the number is less than the threshold, the spiral search is continued; otherwise, the drone fixes the current view without performing a spiral search.
3. When the drone is fixed in position and angle of view, the abnormal monitoring is directly performed on the current viewing angle. If abnormal behavior occurs, an alarm signal is sent to the ground station, so that the monitoring personnel can take countermeasures in a timely and effective manner.

Under the condition of guaranteeing no dead-angle traversal searching of the monitoring area, the scheme avoids repeated searching of some areas, and at the same time avoids frequent turning and air braking, so that the UAV can fly at full speed and achieves the best monitoring efficiency under the condition of limited flight time.

2.2 Convolution neural network

Convolution neural network (CNN) is an artificial neural network developed on the basis of human learning ability for knowledge and computer network. The convolution neural network has more applicability for deep learning. The feature extraction and feature classification of image characters can be carried out simultaneously, and it also has the advantages of strong generalization ability and less training parameters for global optimization. It is one of the pioneering research achievements in the field of machine autonomous learning.

The structure of CNN is varied. The classical CNN structure mainly includes input layer, convolution layer, sampling layer, full connection layer, and output layer. Each layer is composed of multiple neurons, as shown in Fig. 2.

2.2.1 Convolution layer

Convolution layer is the core of the convolution neural network, and its core operation is convolution operation. The convolution layer contains many convolution kernel functions, which extract the image features from the input original image, and determine the location relationship of the image by these local features. The convolution kernel function corresponds to the image feature one-to-one. Each convolution kernel function generates a feature graph. The convolution kernel function values extracted from different feature graphs are different, but they can be shared in the same feature graph. The result of the convolution is convoluted by a learnable convolution kernel with a bias term, and the final output can be obtained by an activation function. Each output characteristic graph can be combined to convolution multiple feature maps. As shown in Formula 2.

$$\begin{aligned} x_j^l &= f\left(u_j^l\right) \\ u_j^l &= \sum\nolimits_{i\in M_j} x_i^{l-1}{}_* k_{ij}^l + b_j^l \end{aligned} \tag{2}$$

Formula 2 illustrates this convolution process, x_j^l represents the j characteristic map of the l level, f represents the excitation function, usually using functions such as sigmoid and tanh. M_j is a subset of the input feature graph, * represents convolution operation, k_{ij}^l represents the convolution kernel matrix, and b_j^l represents the offset term of the convolution feature graph. For an output characteristic graph x_j^l, the convolution kernel k_{ij}^l corresponding to each input characteristic graph x_i^{l-1} may be different.

In view of this pixel input data, there are too many parameters. When the number of parameters is too large, the learning speed of the network will be affected. The solution is a part of the adjacent regions associated with each hidden element. The method of regional connectivity is convolution on the neural network. The convolution process is shown in Fig. 3.

$$f_1 = \text{sigmoid}\left(\sum_{i=1}^{9} x_i{}_* w_i\right)$$

The input of the upper layer is the feature image (layer 0 is the original image), the output of the next layer is the feature image, and the convolution kernel is the parameter sliding window. The formula for convolution characteristics is Formula 3.

$$f_1 = \text{sigmoid}\left(\sum_{i=1}^{9} x_i{}_* w_i\right) \tag{3}$$

In general, if the input image size is i_w*i_h, the moving step size is stride_w, stride_h, and the convolution kernel size is c_w*c_h. The formula for calculating the output image size to be wide and high is 4 and 5 respectively.

$$o_\mathrm{w} = \frac{\text{i\_w}-\text{c\_w}}{\text{stride\_}w}+1 \tag{4}$$

$$o_\mathrm{h} = \frac{\text{i\_h}-\text{c\_h}}{\text{stride\_}h}+1 \tag{5}$$

2.2.2 Sampling layer

Sampling layer is mainly used to sample the feature map generated by convolution layer, so as to reduce the complexity of the image and reduce its resolution. Therefore, the real-time scaling of the feature map through the sampling layer will not affect its quality. Usually, this

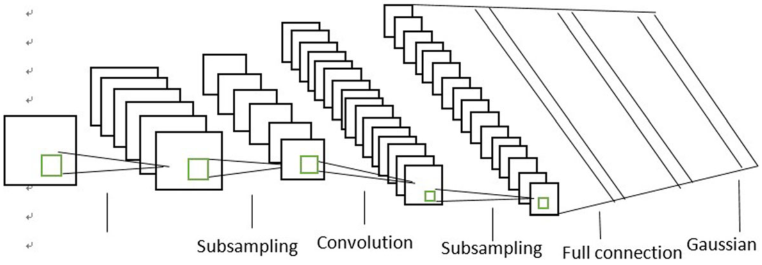

Fig. 2 CNN model

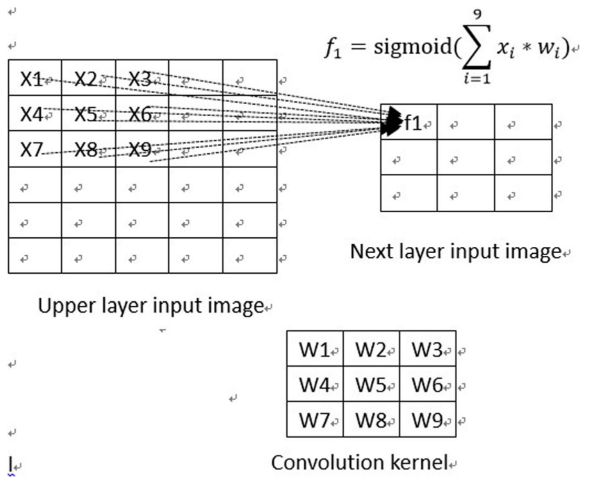

Fig. 3 Convolution operation

operation is also known as pooling operation; pooling operation is to “fuse” adjacent pixels, and the most commonly used pooling methods are maximum pooling method and average pooling method. Maximum pooling is the result of taking the maximum value in the sub-region as the mapping, while average pooling is the result of taking the average value in the sub-region as the mapping.

Figure 4 illustrates the pooling operation in the maximum pooling method. Similar to convolution, the same sliding window slides over the original image, except that each pixel taken is the maximum of the pixels in the sliding window. In general, the pool window does not overlap, that is, the mobile step is the side length of the sliding window. The mean pooling means the mean value of the current sliding window pixel value. The pooling operation is relatively simple, and a convolution feature graph produces only one pooled feature graph.

The sampling layer samples each input characteristic map through the following Formula 6, and outputs the feature map. Where u_j^l is called the net activation of the j channel of the sampling layer l. which is obtained by weighted sampling from output characteristic graph x_i^{l-1} of upper layer and biased. β_j^l is the weight coefficient, b_j^l is the bias item, and down() represents the sampling function. The input feature graph x_i^{l-1} is divided into several non-overlapping $n*n$ image blocks by sliding window method, then sum the pixels in each image block to get the maximum of the mean value, so that the output image in both dimensions is reduced by n times.

$$\begin{aligned} x_j^l &= f\left(u_j^l\right) \\ u_j^l &= \beta_j^l \mathrm{down}\left(x_i^{l-1}\right) + b_j^l \end{aligned} \tag{6}$$

2.2.3 Fully connected layer

The fully connected layer is located at the end of the convolution neural network, and after several convolution layers and pooling layers, the feature maps formed after the structure before the convolution neural network are connected to each other, so that according to the requirements of the classification, the number of output nodes is debugged, and finally the image classification task is completed. The output of the fully connected layer l can be obtained by weighting the inputs and obtaining the response of the activation function, as in Formula 7:

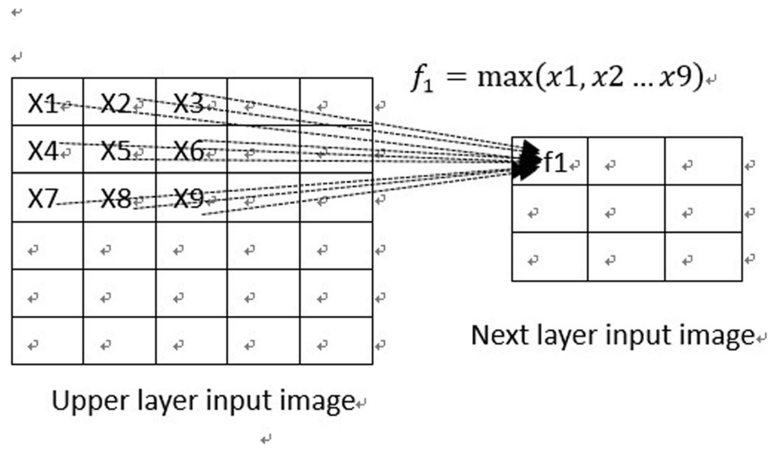

Fig. 4 Max pool operation

$$\begin{aligned} x^l &= f\left(u^l\right) \\ u^l &= \omega^l x^{l-1} + b^l \end{aligned} \tag{7}$$

Where u^l is called the net activation of the fully connected layer l, which is obtained by weighting and biasing the previous layer output characteristic x^{l-1}. ω^l is the weight coefficient of the fully connected network, and b^l is the bias item of the fully connected layer l.

2.2.4 Softmax regression

At the end of the convolution neural network is a Softmax regression classifier. Softmax is a logistic regression obtained from two-category generalization to multiple classification. The traditional logistic regression is a two-category classifier with a classification category of 0 or 1, and the result corresponds to the category probability. Softmax is a general form of logistic regression for multiple classifications. The number of categories is k. When $k = 2$ is defined, the function automatically degenerates into a traditional logistic regression. Therefore, Softmax is more suitable for multiple classification. Its function is defined as Formula 8:

$$P(y = c|x) \quad = \quad \frac{1}{1 + \sum_{k=1}^{c-1} e^{\theta_k^T x}} \tag{8}$$

where c represents the category of classification. In general, c takes integers. x represents the value of the attribute variable, and θ is the estimate parameter, and the value of θ can be obtained by the likelihood solution. At this point, the objective minimization cost function is Formula 9.

$$\begin{aligned} J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} l\left\{y^{(i)} = j\right\} \, log \frac{e_j^T x^{(i)}}{\sum_{l=1}^{k} e_l^T x^{(i)}}\right] \\ + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \end{aligned} \tag{9}$$

where m is the number of training samples, $y^{(i)}$ is category label, l{} is truth value judgment function, the second is weight attenuation term. The function of attenuation term is parameter trade-off, so that the influence of large parameter value is reduced, and λ is the influence coefficient.

The derivative of function is Formula 10.

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[x^{(i)}\left(l\left\{y^{(i)} = j\right\} - p\left(y^{(i)} = j|x^{(i)};\theta\right)\right)\right] + \lambda\theta_j \tag{10}$$

2.3 Residual neural network

The depth of the neural network is very important to its performance. But the deeper the network is, the more difficult it is to train, which means the deeper the neural network is more difficult to train. That is, the neural network has a "degeneration" phenomenon. It also means the deeper neural network will have a higher training error and test error than the shallow network. In response to this problem, there is a concept of "residual learning."

Suppose a sub-module of the neural network is mapped to $H(x)$, which may be difficult to learn. Now let the sub-module learn a residual $F(x) = H(x) - x$ instead of learning the target mapping directly, so that the original target mapping becomes $F(x) = H(x) + x$, that is, the sub-module can be composed of two parts: A linear direct mapping $x \rightarrow x$ and a nonlinear mapping $F(x)$. If the direct mapping is optimal, the learning algorithm can easily set all the weight parameters of the nonlinear mapping to zero. Residual neural network is mainly composed of multiple residual learning modules stacked together. Residual learning module is shown in Fig. 5.

In this structure, a path from x to y without weight is set up. After passing x, each module only learns the residual $F(x)$, which makes the network stable and easy to learn. With the increase of network depth, the performance will gradually improve. When the network layer is deep enough, it is easier to optimize residual function: $F(x) = H(x) - x$ than to optimize a complex nonlinear mapping $H(x)$.

In addition to the two level residual learning unit, there are three levels of residual learning units in ResNet. The two-layer residual learning unit contains two identical output channel numbers (since the residual is equal to the target output minus the input, so the input and output dimensions need to be consistent) 3*3 convolution; and the three-layer residual network uses 1*1 convolution in Network In Network and Inception

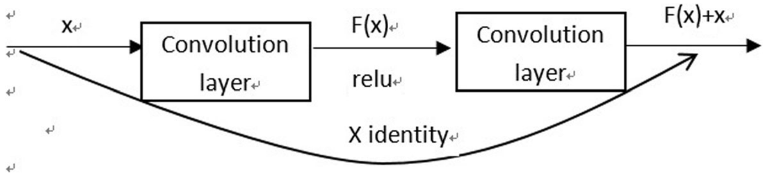

Fig. 5 Residual learning module

Net, and uses 3*3 convolution in the middle and uses 1*1 convolution before and after. The operation of reducing dimension and increasing dimension reduces computation complexity; ReLU represents a linear rectification function, and it is also called a modified linear unit which is a commonly used activation function as shown in Fig. 6.

In this paper, the residual neural network structure is used for image recognition. In the network, a 3*3 convolution kernel is used. m is a scaling parameter, which multiplies the number of convolutions in a module by m, that is, the transformation of the scaling parameter m. The width of the network becomes 3*m, and the convolution kernel of this paper becomes 3*3*16*m.

2.4 Dropout principle

Dropout is a commonly used regularization method, which can inhibit the occurrence of over-fitting phenomenon to a certain extent and improve the generalization ability of the network. The output value of some nodes is changed to zero in each training by a certain probability p, which is equivalent to "deleting" the nodes from the whole network and keeping the other neurons with probability $q = 1\text{-}p$. In this way, the corresponding parameters will not be updated when back propagation. The whole network is used for testing in the testing phase. At present, Dropout is heavily utilized in fully connected networks, usually set at 0.5 or 0.3. Dropout is equivalent to training a subnet of the whole network at each training. If there are n nodes in the network, then the number of available subnets should be 2^n. When n is large enough, the subnets used in each training will not be the same. Finally, the whole network can be regarded as the average of multiple subnet models, so as to avoid over fitting the training set in a certain subnet.

The whole process uses dropout during training, but does not use dropout during testing (that is, no discarding of network parameters). Multiply the output of dropout layer at test time by retaining probability p used in training. Suppose x is the input of dropout, y is the output of dropout, W is all the weight parameters of the upper layer, and $W \mid p$ is a subset of weight parameter obtained from retaining probability p sampling. Formula expression is $\text{train}: y = W \mid p * x$; $\text{test}: y = W * px$.

3 Simulation experimental results

The image taken by UAV in this experiment is JPEG image. The simulation experiment is carried out for the image target location and recognition system based on convolution neural network model. The convolution neural network is used for image recognition. The main process of image target localization and recognition includes three steps: image segmentation, feature extraction, and object classification. Because the network will automatically extract features that are beneficial to classification and recognition during the training process, it is not necessary to pay too much attention to the image preprocessing work, just to design the network structure and adjust the network parameters. Therefore, the simulation process is mainly to see the recognition effect of the system.

The image data of 128*128*3 is input firstly, the number of convolution kernels is 16, and 128*128*16 is output after 3*3 convolution kernel. After 6*n convolution, the size of the data is divided by 2 for each n learning modules, and the number of convolution kernels is multiplied by 2, and the down-sampling operation is realized by convolution operation of stride = 2. After convolution, the global average sampling layer is performed, followed by dropout operation, and finally the full connection layer and Sofmax. The complete network structure is shown in Fig. 7.

4 Discussion

The accuracy of the network can be greatly improved by putting the necessary preprocessing work of the image data into the network. However, in order to facilitate comparison with the existing experimental results, this paper only used the conventional preprocessing method.

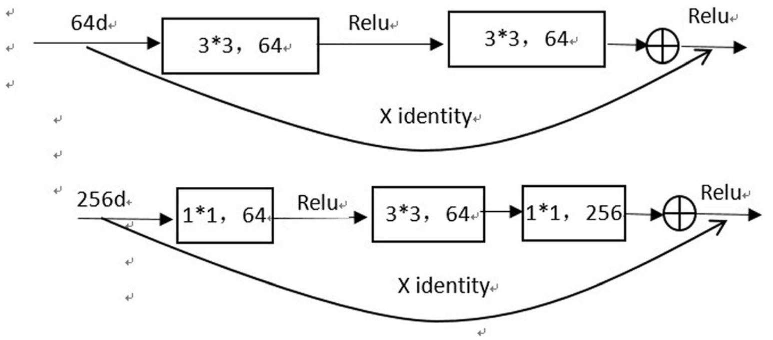

Fig. 6 Two-layer and three-layer residual learning module

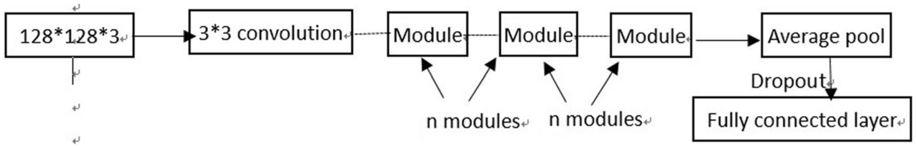

Fig. 7 Residual neural network complete network structure

The dropout is added before the full connection layer of the entire network, and the parameter p denotes that the nodes are discarded at a probability of p each time. In this experiment, the momentum random gradient descent method is used to help accelerate the vector to descend in the correct direction, thus accelerating the convergence rate. The dropout discard probability p is set to 0.3.

1. Training process

For the training process, each picture in four directions are filled with several bits 0, and then randomly cut into the original picture size, and the picture using ZCA whitening operation. In the training process, the input sequence of each epoch picture is random.

2. Test process

For the test process, the original image data is directly input in this experiment without any preprocessing operations. And there is no order disruption.

4.1 Experimental results and analysis

Following the network structure of residual learning in the previous section, each learning module has two convolutions, the whole network has $6^{*}n + 2$ layers, and each layer is composed of 3*3 convolutions. In order to explore the effects of different scaling parameters and depths on the performance of residual networks, a number of experiments were carried out. At the same time, the operation of different network models under different scaling parameters and network depths was tested, as shown in Table 1.

Table 1 Model operation under different scaling parameters and depths

Scaling parameter *m*	*N*th layer	Network depth	Parameter value	Test set error (%)	Training situation
1	*n* = 20	122	7.6M	4.89	Can be trained
1	*n* = 24	146	8.6M		Unable to train
2	*n* = 8	50	11.7M	4.21	Can be trained
2	*n* = 10	62	14.6M		Unable to train
3	*n* = 6	38	17.1M	4.23	Can be trained
3	*n* = 7	44	22.1M		Unable to train
4	*n* = 4	26	22.6M	4.1	Can be trained
4	*n* = 5	32	26.1M		Unable to train
5	*n* = 3	20	29.1M	4.06	Can be trained
5	*N* = 4	26	35M		Unable to train

As can be seen from Table 1, as the scaling parameter m increases, the number of parameters that the network can train is also increasing. When the scaling parameter is 1, when the network is added to 122 layers, the training parameter is 22.6M. If the number of layers is increased, the problem of untrained will appear. When the scaling parameter is increased to 5, the maximum training parameter is 77.1M. With the increase of parameter, the accuracy of network is higher and higher.

At the same time, the influence of different scaling parameters and depths on the error of the test set is set in the residual network, as shown in Table 2.

It can be seen from the Fig. 8 that as the scaling parameter m increases, although the network depth is decreasing, the model is still more and more complex, and more and more parameters need to be trained. At the same time, however, the classification error rate of the model on the test set has also become lower and lower. In this experimental environment, it reappeared to the 56th floor, but there was a problem of unable to train at the 152th floor.

4.2 Experimental results and discussion

Some network structures in recent years are compared with the test set in this experiment. Scale Res Net represents the network structure adopted in this experiment. The scaling

Table 2 Test set error rate for different scaling parameters and depth

Scaling parameter *m*	*N*th layer	Network depth	Parameter value	Test set error (%)
1	*n* = 12	74	4.4M	4.88
1	*n* = 20	122	7.6M	4.92
1	*n* = 16	98	8.6M	4.81
2	*n* = 8	50	11.7M	4.21
3	*n* = 5	32	16.1M	4.01
3	*n* = 6	38	19.3M	4.18
4	*n* = 4	26	22.5M	4.01
5	*n* = 3	20	17.2M	4.05

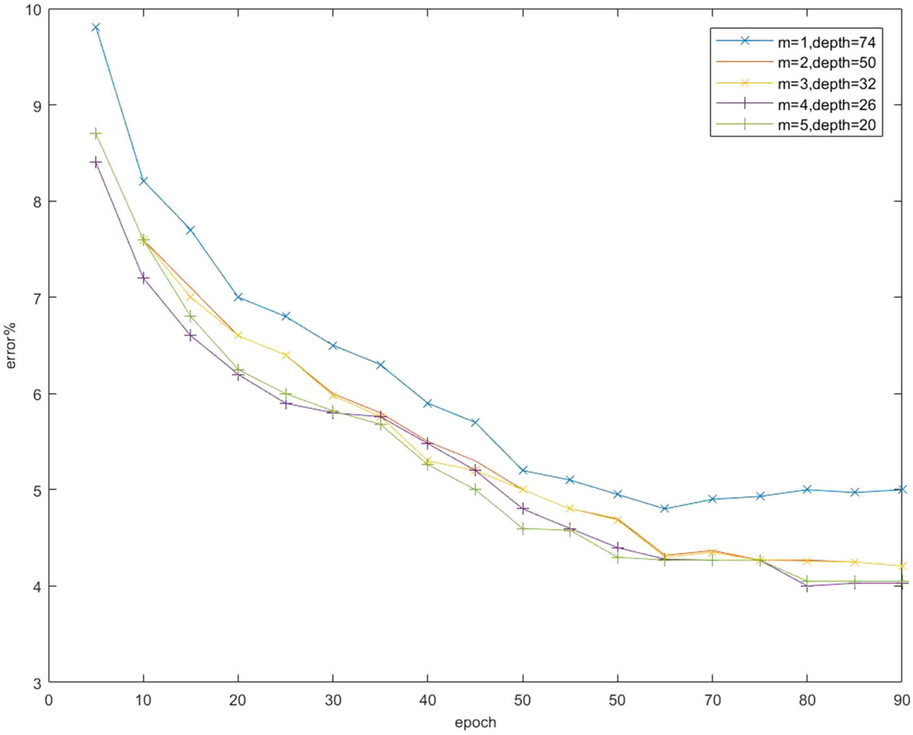

Fig. 8 Test set error rate–training times curve

parameter is 4 and the network depth is + 2 = 26. The error rate is shown below (Table 3).

As can be seen from the Fig. 9, the accuracy of Res Net-56 is 6.32%, and Res Net-152 is 4.31%. Compared with other network structures, the accuracy of residual neural network is relatively excellent. At the same time, the error rate was 6.25% when using a single model, and when using the integrated method of 100 models, the error rate drops to 4.12%, which indicates that the integration method can greatly improve the accuracy of the model.

The experimental results show that the classification error rate of the model on the test set becomes lower and lower as the scaling parameter m increases. The classification error rate of the model on the test set also becomes lower and lower, and relative to other network structures at the same time. In other words, the test accuracy of the residual network is higher. When the residual parameter network has a scaling parameter of 4 and the network depth is 26, the network structure has the highest accuracy, and the accuracy is higher than many current advanced network structures.

Table 3 Error rate of different network structures

Network structure	Test set error rate (%)
ELU	6.75
MIN	8.81
Maxout	9.21
Fractional Max-Pooling(single pass)	4.51
Res Net-110	6.32
Res Net-1082	4.31

5 Conclusions

This paper investigates the technology involved in UAV image target location and recognition. Convolution neural network model is the most commonly used model for image processing. It has more hidden layers, and its unique convolution and pooling operations are more efficient in image processing. In this paper, after analyzing the expression characteristics of CNN image features, the characteristics and related characteristics of residual nerve are analyzed. The convolution layer in the residual neural network can observe the data from many aspects and obtain more abundant input features. It not only reduces the depth of the network but also effectively suppresses the occurrence of gradient disappearance problem and reduces the difficulty of training. At the same time, by increasing the number of parameters, the learning ability of the whole network will become stronger.

In this experiment, the environmental monitoring image taken by UAV is taken as the recognition object. Through establishing the image recognition flow frame, the residual neural network based on CNN is used to recognize the image and the recognition accuracy is

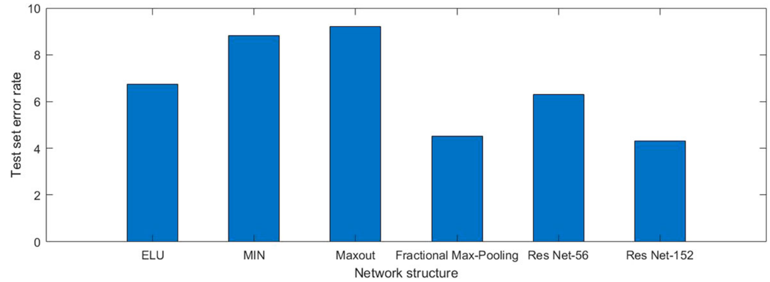

Fig. 9 Error rate of different network structures

analyzed and studied. The experiment also adjusts the network scaling parameters, network depth, dropout values, so that the accuracy of the network is further improved.

The research work of this paper still has some shortcomings: for example, it is difficult to test more complex network structure, or apply it to more complex data sets.

In recent years, image classification and recognition using convolution network has made rapid progress, and has also been applied in some industrial fields. However, convolution neural networks are complex, have long training time, and a network structure is often difficult to achieve the best results in a variety of tasks. This experiment just completed the basic assumption, and there are still some areas to optimize the engineering implementation; the training of the model is a very time-consuming process, so the contrast experiment did not do such as activation function comparison. It includes determining the size of a more accurate convolution kernel, allowing the network to adjust adaptively, and finding excitation functions with better convergence rate and accuracy.

Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

About the authors

Zhao Kunrong was born in Meizhou, Guangdong, P.R. China, in 1979. He received the Doctor's Degree from Sun Yat-sen University, P.R. China. Now, he works in South China Institute of Environmental Sciences. MEP. His research interest includes environmental engineering, computational intelligence, and information security. E-mail: zhaokunrong@scies.org
He Tingting was born in Guangzhou, Guangdong, P.R. China, in 1993. She received the bachelor's degree from Guangdong University of Finance & Economics, P.R. China. Now, she works in Guangzhou Hexin Environmental Protection Technology Co., Ltd. Her research interest include environmental assessment, big data analysis, and information security. E-mail: ivyhtt@mail.scut.edu.cn
Wu Shuang was born in Beitun, Xinjiang, P.R. China, in 1990. She received the Master's degree from Northwest Normal University, P.R. China. Now, she works in South China institute of environmental sciences, ministry of environmental protection Guangzhou Huake environmental protection engineering CO.LTD. Her research interest includes Environmental planning and management. E-mail: wushuang@scies.org
Wang Songling was born in Ledong, Hainan, P.R. China, in 1993. He received the bachelor's degree from Qingdao University of Technology, P.R. China. Now, He works in South China Institute of Environmental Sciences. MEP. His research interest include computational intelligence, information security, and big data analysis. E-mail:wangsongling@scies.org
Dai Bilan was born in Meizhou, Guangdong, China in 1995. She received a bachelor's degree from Guangdong Ocean University. At present, she is working in South China Institute of Environmental Sciences. MEP. Her research direction is the comprehensive development and utilization of environmental information resources. E-mail: daibilan@scies.org
Yang Qifan was born in Maoming, Guangdong, China, in 1995. He received the bachelor's degree from Guangdong University of Finance & Economics. Now, he works in Guangzhou Hexin Environmental Protection Technology Co., Ltd. His research interest includes cloud security, chaos encryption, and information security. E-mail:xuwj7@mail3.sysu.edu.cn
Lei Yutao was born in Huizhou, Guangdong, P.R. China, in 1978. He received the Master's Degree from Guangdong University of Technology, P.R. China. Now, he works in South China Institute of Environmental Sciences. MEP. His research interest includes environmental engineering and environmental assessment. E-mail: leiyutao@scies.org

Funding

Not applicable.

Availability of data and materials

Please contact author for data requests.

Authors' contributions

All authors take part in the discussion of the work described in this paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]South China Institute of Environmental Sciences, MEP, Guangzhou, Guangdong, China. [2]Guangzhou Hexin Environmental Protection Technology Co., Ltd, Guangzhou, Guangdong, China. [3]Guangzhou Huake Environmental Protection Engineering CO.LTD, Guangzhou, Guangdong, China.

Received: 23 September 2018 Accepted: 5 December 2018
Published online: 27 December 2018

References

1. P. Koprinkova-Hristova, V. Mladenov, N.K. Kasabov, Artificial neural networks[J]. Eur. Urol. **40**(1), 245 (2015)
2. H. White, Learning in artificial neural networks: a statistical perspective[J]. Neural Comput. **1**(4), 425–464 (2014)
3. L. Zhou, *Integrating artificial neural networks, image analysis and GIS for urban spatial growth characterization[J]* (2012)
4. K. Nowakowski, P. Boniecki, R.J. Tomczak, et al, Identification of malting barley varieties using computer image analysis and artificial neural networks[C]//Fourth International Conference on Digital Image Processing

(ICDIP 2012). International Society for Optics and Photonics. **8334**, 833425 (2012)

5. L.S. Bartolome, A.A. Bandala, C. Llorente, et al, Vehicle parking inventory system utilizing image recognition through artificial neural networks[C]// TENCON 2012 - 2012 IEEE Region 10 Conference (IEEE, Cebu, 2012), pp. 1–5
6. Y.A. Al-Sbou, Artificial neural networks evaluation as an image denoising tool. World Appl. Sci. J. **17**(2), 218–227 (2012)
7. D.J. Hemanth, C.K.S. Vijila, A.I. Selvakumar, et al., Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification[J]. Neurocomputing **130**(3), 98–107 (2014)
8. S. Kouamo, C. Tangha, in *International Joint Conference CISIS'12-ICEUTE´12-SOCO´12 Special Sessions, Vol 112.*. Image compression with artificial neural networks (Springer, Berlin Heidelberg, 2013)
9. C.F. Crispim-Junior, J. Marino-Neto, Artificial neural networks and image features for automatic detection of behavioral events in laboratory animals. 33 (2013), pp. 862–865
10. M. Nijim, N. Mantrawadi, Drone Classification and identification system by phenome analysis using data mining techniques[C]// Technologies for Homeland Security (IEEE, Waltham, 2016), pp. 1–5
11. Y. Jin, G.H. Jiang, J.G.A. Chao, Cross drone image-based manual control rendezvous and docking method. J. Astronaut. **31**(5), 1398–1404 (2010)
12. B. Jong-Hwan, J. Jin-Seong, P. Myeong-Suk, et al., Design of walking drone using image processing for logistics. Adv. Sci. Lett. **22**(9), 2288–2291 (2016)
13. B. Srikudkao, T. Khundate, C. So-In, et al. Flood warning and management schemes with drone emulator using ultrasonic and image processing[J]. Recent advances in information and communication technology. (361), 107–116 (2015)
14. G. Maria, E. Baccaglini, D. Brevi, et al, A drone-based image processing system for car detection in a smart transport infrastructure[C]// Electrotechnical Conference (IEEE, Limassol, 2016), pp. 1–5
15. E.J. Lee, S.Y. Shin, B.C. Ko, et al., Early sinkhole detection using a drone-based thermal camera and image processing. Infrared Phys. Technol. **78**, 223–232 (2016)
16. M. Skoczylas, Vision analysis system for autonomous landing of micro drone. Acta Mechanica Et Automatica **8**(4), 199–203 (2014)
17. X. Xiong, J. Feng, B. Zhou, Automatic view finding for drone photography based on image aesthetic evaluation[C]// International Conference on Computer Graphics Theory and Applications (GRAPP 2017, Porto, 2017), pp. 282–289
18. J.H. Oh, C.N. Lee, Journal of the Korean Society of Surveying, geodesy. Photogrammetry Cartography **36**(3), 127–133 (2018)
19. Z. Luo, L. Liu, J. Yin, et al., Deep learning of graphs with Ngram convolutional neural networks. IEEE Trans. Knowl. Data Eng. **29**(10), 1–1 (2017)
20. S. Hoo-Chang, H.R. Roth, M. Gao, et al., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging **35**(5), 1285 (2016)
21. M. Oquab, L. Bottou, I. Laptev, et al., Learning and transferring mid-level image representations using convolutional neural networks (2014), pp. 1717–1724
22. L.A. Gatys, A.S. Ecker, M. Bethge, Image Style Transfer Using Convolutional Neural Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society (CVPR 2016, Las Vegas, 2016), pp. 2414–2423
23. F. Milletari, N. Navab, S.A. Ahmadi, V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation(2016), pp. 565–571
24. J. Zbontar, Y. LeCun, Stereo matching by training a convolutional neural network to compare image patches[J]. J. Mach. Learn. Res. **17**(1–32), 2 (2016)
25. L. Ma, Z. Lu, H. Li, Learning to answer questions from image using convolutional neural network. AAAI **3**(7), 16 (2016)
26. D. Pathak, P. Krahenbuhl, T. Darrell, Constrained convolutional neural networks for weakly supervised Segmentation[C]// IEEE International Conference on Computer Vision. IEEE Computer Society (ICCV2015, Santiago, 2015), pp. 1796–1804