International Journal of Nanotechnology

ISSN online: 1741-8151 - ISSN print: 1475-7435
https://www.inderscience.com/ijnt

Image feature extraction algorithm based on parameter adaptive initialisation of CNN and LSTM

Dong Li, Zai Luo, Xingmin Ma

DOI: 10.1504/IJNT.2022.10051326

Article History:

Received:	17 March 2021
Last revised:	03 August 2021
Accepted:	14 September 2021
Published online:	31 May 2023

Image feature extraction algorithm based on parameter adaptive initialisation of CNN and LSTM

Dong Li

Gongniu Group Co., Ltd.,
Ningbo, Zhejiang, 315000, China
Email: li1032321384@163.com

Zai Luo

College of Metrology and Measurement Engineering,
China Jiliang University,
Hangzhou, Zhejiang, 310000, China
Email: luozai@cjlu.edu.cn

Xingmin Ma

North China Institute of Computing Technology,
Beijing, 100083, China
Email: 3133713064@qq.com

Abstract: Image feature selection and extraction is the basis of image processing. For different types of images and different application requirements, the selection of image features will be different. The image feature extraction algorithm based on deep learning is a good method, but its overfitting, gradient explosion and initialisation of deep learning parameters are not adaptive, which directly leads to the low accuracy of the algorithm. Therefore, this paper proposes an image feature extraction algorithm based on adaptive initialisation of convolution neural network parameters based on MMN linear activation function. The parameter initialisation algorithm based on multi-layer Maxout activation function solves the problem of poor identification effect caused by improper parameter initialisation. Based on the selective Dropout algorithm of shallow learning of long and short time memory network, the problems of overfitting and gradient explosion in deep learning are overcome. SUSAN operator and random consistency algorithm are introduced to perform fine matching and purification. The adaptive parameter initialisation CNN algorithm proposed in this paper can effectively overcome the overfitting problem in the training process of deep learning, avoid the gradient explosion caused by improper parameter initialisation, and improve the calculation speed and accuracy of image feature extraction. And through the verification and simulation, the proposed image feature extraction algorithm can obtain more image details and edge information, and can better reflect the image feature information, for the image fusion and image recognition to lay a good foundation.

Keywords: feature extraction; convolutional neural networks; long short term memory network; image matching; adaptive initialisation.

Reference to this paper should be made as follows: Li, D., Luo, Z. and Ma, X. (2023) 'Image feature extraction algorithm based on parameter adaptive initialisation of CNN and LSTM', *Int. J. Nanotechnol.*, Vol. 20, Nos. 1/2/3/4, pp.214–234.

Biographical notes: Dong Li received the Bachelor's degree in Measurement & Control Technology and Instrument from Tianjin University of Technology and Education in 2015, and the Master degree in Instrument Engineering from China Jiliang University in 2018.

Zai Luo graduated from Hefei Polytechnic University with a Doctor's degree. He is now a Professor of China University of Metrology and the Director of the Key Laboratory of testing technology and instruments for key parts of automobile braking system in mechanical industry.

Xingmin Ma graduated from the University of Science and Technology Beijing with a PhD and is currently a Senior Engineer at the North China Institute of Computing Technology. The main research directions are image processing and big data analysis.

1 Introduction

In the era of information explosion, digital image, as one of the media for people to receive and transmit information, is playing an increasingly indispensable role. The corresponding image processing and operation is also in line with the situation [1]. What kind of tool is chosen for image processing often directly determines the quality of the processing results. In recent years, the neural network, which is popular in recent years, is undoubtedly an effective means. In the process of image processing, because there are many kinds of images, and the information such as colour feature, texture feature, shape feature and spatial relation feature of a single image are complex, how to describe these image features becomes a key problem. The quality of information transmitted by the extracted image features will directly affect the results of image processing. Different from the traditional manual design of image features, neural network uses a learning based image feature extraction method, which takes image data as the input of neural network. After a series of stacked linear combination and nonlinear transformation, it is transformed into a higher-level abstract representation, and the feature information of the image is automatically extracted layer by layer, which can effectively avoid the inefficiency and tediousness of manual design features.

According to the depth network image feature extraction method mainly studied in this paper, firstly introduced the basic knowledge of neural network; secondly, introduced two depth image feature extraction tools: trestle sparse self coding and convolutional neural network, described their structure, cost function and training process, etc, the algorithm greatly reduced the amount of network weight training, thus accelerated. The training speed of the network is solved, and the binary or problem which is not solved when the perceptron is put forward is solved. Therefore, the research of neural network has entered the recovery period. Finally, a case study is carried out to illustrate the effectiveness of the proposed algorithm through the analysis of standard images and downhole measured images.

2 Related work

Krizhevsky, a student of Himon, put forward the deep convolution neural network Alex net in 2012, and won the championship in the Imagenet 2012 image recognition contest with an accuracy rate of 11% higher than the second place. In the network structure, a layer of rectifier linear units is added after the original convolution layer, and a layer of normalisation is added after the original pooling layer [3]. Fukushima proposed a dual c-metacognitive machine in 1989. This method further blurs by adding a C-element features, some unknown features can be fuzzy approximated to known features, which further improves the robustness of the improved neural cognitive machine [4]. The SUSAN detection algorithm was proposed by S.M. Smith of the University of Oxford in 1997 [5], which uses an approximately circular template to move over the image with a circular template. The corner feature of the SUSAN algorithm is accurate and the operation speed is fast, and the algorithm does not need to differentiate the image, and has strong anti-noise ability, and can detect all types of corner points. Therefore, the SUSAN detection algorithm is also applicable to many fields, such as image stitching, image registration, image segmentation, target recognition and tracking [6–9]. In 2001, Kadir and Brady [10] of the University of Oxford in the UK proposed a salient region detection algorithm. The algorithm uses information entropy to detect local feature regions in the image. The feature region has scale and affine invariance, but the algorithm has poor correctness for background complex image detection and lack of robustness. In 2004, Kadir and Brady [10] updated the algorithm [11]. In 2004, Matas et al. [12] of the Czech Machine Perception Center proposed the Maximum Stable Extremal Region (MSER) detection algorithm [12], which is based on the watershed concept to solve a kind of affine invariance of stable local regions. Algorithm, the algorithm does not have rotation without deformation, and the robustness is not strong. In 2004, Mikolajczyk and Schmid [13] of the French National Institute of Computer and Automation combined the Harris operator with the Gaussian Laplacian operator and proposed the Harris-Laplace algorithm. The algorithm selects the LoG scale response function to automatically select the feature scale of the local structure, which can detect the feature points with feature scale with a large probability. This feature can make the Harris-Laplace operator have larger resolution difference. Matching is achieved between images, but the time complexity of the algorithm is still high. In 2011, Zhang et al. [14] from Nanjing University of Science and Technology constructed a gray-scale intensity variation formula based on wavelet transform, and obtained a new Harris multi-scale corner detection algorithm based on wavelet transform. The time complexity of the algorithm is low. But the robustness of the algorithm still needs to be improved.

In 2015, Hekaimhg et al. [15] proposed a 1% deep residual network based on convolution network. Furthermore, a new residual retransmission method is proposed to ensure that the gradient propagates from the separation layer to the bottom layer, which effectively avoids the problem of the gradient disappearing from the training depth network [16]. This operator maintains image scaling, rotation and even affine transformation to keep image feature invariance. DoG only needs to subtract the adjacent images in the Gaussian scale space to obtain, and the operation is relatively simple. Therefore, the use of DoG instead of LoG is a very good choice. The robustness and correctness of the algorithm are improved, but the algorithm time complexity is too high. In 2004, Lowe [17] described the characteristics of SIFT based on the original research. Since the feature vector of the SIFT operator is 128-dimensional and the space complexity is high, it is difficult to implement real-time applications. To this end, in 2004, Yan and Sukthankar [18] of Carnegie Mellon University in the USA proposed the PCA-SIFT description operator, which uses principal components analysis (PCA) to reduce 128-dimensional vectors to 36 dimensions. Compared with the SIFT algorithm, PCA-SIFT improves the matching speed, but the uniqueness of this algorithm has decreased. In 2005, Mikolajczyk and Schmid [19] of Oxford University proposed the GLOD (Gradient Location-Orientation Histogram, GLOH) descriptor, which is also an extension of SIFT, in order to enhance the robustness and uniqueness of feature descriptors [19]. In 2011, Wang et al. [20] of the Institute of Automation of the Chinese Academy of Sciences proposed the LIOP algorithm (Local Intensity Order Pattern), which is an extension of the GLOH feature descriptor, which is a 3D histogram of edge position and direction. The edge is extracted by Canny operator, and finally a 9 × 4 = 36-dimensional feature descriptor is generated. This descriptor can effectively describe the contour information. It is widely used in shape matching and target recognition, but the algorithm is not robust enough. In 2013, Cai et al. [21] of Jimei University School of Science proposed an efficient remote sensing image registration algorithm, PSIFT image registration algorithm. Based on SIFT and ASIFT, the algorithm of perspective invariant was introduced to improve the registration. Speed, but the robustness of the algorithm is still insufficient. In 2014, Le et al. [22] of the University of Houston's computational biomedical laboratory made improvements on the basis of DAISY descriptors, proposed 3D local image descriptors, and achieved better registration results, but the time complexity of the algorithm is high. He et al. [23] of the Key Laboratory of Adaptive Optics of the Chinese Academy of Sciences in 2018 proposed a SIFT structure for matching scanned image sequences. The emergence of deep learning algorithm provides a relatively accurate and robust algorithm for image feature extraction. In 2016, Chen et al. [24] of Harbin Institute of Technology proposed a deep feature extraction algorithm based on CNN regularisation. The algorithm is robust, but there is over-fitting phenomenon of CNN algorithm, and the correct rate needs to be improved. In 2018, Ding et al. [25] of Zhejiang University proposed a non-reference stereo image quality evaluation algorithm using convolutional neural network (CNN) for feature extraction, which predicts the accuracy of stereo image quality and crosses different databases and distortion types. The robustness is superior to other state-of-the-art algorithms, but there is still a CNN algorithm overfitting phenomenon.

The advantage of CNN method is its representation learning ability, which can automatically learn the appropriate representation from the original pixels of data [26–29]. Similar to the biological neural network, the structure of weight sharing network has less weight number, avoiding the formation of complex network model. The input interface of neural network can be directly connected with the image, which reduces the complexity of feature extraction and data reconstruction in the traditional image recognition algorithm, thus simplifying the operation process of network input. The network can extract image features, including colour, texture, shape and topological structure of image, etc. it has good robustness and efficiency in recognition of displacement, scaling and other forms of distortion invariance. The feature extraction layer parameters can be obtained by training data learning, which avoids the tedious work of artificial feature extraction. The advantage of convolution network is that the neurons of the same feature graph share the weights and reduce the network parameters. The sharing of local weights is closer to the real biological neural network, which reduces the complexity of the network.

Feature point detection based on CNN algorithm is convenient and fast in practical applications, and has become a research hotspot in the field of image registration. Therefore, this paper will find the appropriate activation function to extract the features of the image, solve the problem of insufficient correctness and robustness of the CNN algorithm itself, and provide an effective feature basis for the next step of image matching. At present, the mixed use of different types of features can often achieve better recognition than a single feature, so the main research trend in recent years is multi-feature fusion. The rise of deep learning brings more options for feature extraction. Deep learning itself can be regarded as a process of feature learning, which can automatically discover and learn new features in unknown complex samples. Deep learning for image point detection and edge detection has become a research hotspot in the field of feature extraction [30,31]. Therefore, this paper will find the initial activation function of the CNN algorithm to solve the problem of high time complexity in the over-fitting, under-fitting and image matching process of the algorithm itself, and further improve the registration speed.

With the continuous development of deep neural network and artificial intelligence technology, the corresponding technology application value is also rising. Therefore, they are more and more used in the field of image recognition and semantic understanding. Using GPU to accelerate deep learning and train deep learning network can give full play to the efficient parallel computing ability of thousands of computing cores of GPU, and greatly shorten the time spent in the scenario of using massive data to train data. CNN classification network model can directly input an image into the network model, and then output the classification results directly at the output end, which is different from the traditional model in this respect. Its advantage is that it doesn't need complex preprocessing. It puts the pattern classification of feature extraction into a black box, sets the learning rate to optimise the network model for many times, so as to obtain the parameter data and classification result information. The core of the network is the network structure design and network solution. The performance of this structure is higher than that of traditional algorithms.

3 Multi-layer Maxout activation function and parameter adaptive initialisation algorithm

There are many activation functions commonly used in deep neural network algorithms. The most common ones are tangent function, S kernel function, ReLU function and Maxout function [32,33], but the first two functions are prone to supersaturation, in information propagation. The gradient information is easy to be lost in the process, which causes the problem of high time complexity of the algorithm, and ultimately affects the training efficiency of the entire image. The ReLU activation function also has obvious shortcomings. In the reverse training process, there will be a situation where the gradient is not smooth, which ultimately leads to low learning efficiency. The Maxout activation function is realised by learning the data and combining the activation function with multiple linear functions. There is no gradient in the image training process, but the training is for non-convex functions. In the case of poor fitting, that is, non-linear, there is a blind zone in the Maxout activation function. Therefore, based on the piecewise linear function, this paper proposes a multi-layer Maxout network (MMN) activation function, which is a combination of the linear combination of the original layer into a multi-layer linear function, using multiple layers. The Maxout function goes to infinitely fit to approximate the nonlinear activation function that may occur. The adaptive initialisation algorithm based on Maxout activation function is given below to solve the problems of over-fitting and gradient propagation in deep learning training, and to improve the fitting performance.

3.1 Multi-layer Maxout activation function

In the process of designing the multi-layer Maxout activation function, it is mainly for deep extraction of image features. The resulting image feature points are sufficient to represent the original image information, and avoid the above-mentioned nonlinear activation function fitting blind zone. A Maxout function that can be trained repeatedly is proposed.

Assuming that $x \in R^d$ (the original pixel point of the image or the image feature state variable of the previous layer), then the activation function of a certain neuron node in CNN is characterised as follows [32]

$$f_{i,j_1} = \max_{j_0 \in [1,k_0]} x^T W_{\cdots i_0} + b_{i_0} \tag{1}$$

$$f_{i,j_n} = \max_{j_{n-1} \in [1,k_{n-1}]} f^T_{i,j(n-1)} W_{\cdots ij_{(n-1)}} + b_{ij_{(n-1)}}, \quad n \in [2, N] \tag{2}$$

$$h_i = \max_{j_N \in [1,\, k_N]} f_{i,j_N} \tag{3}$$

In equations (1), (2), (3), k_n represents the number of nodes in the n layer, N represents the number of required MMN layers, and h_i represents the output capability of the MMN at the f node.

The initial Maxout activation function is an interval linear function, so it cannot fit the non-convex function. When the Maxout activation function is designed in multiple layers to form the MMN activation function, it has the ability to fit the non-convex function. The MMN activation function has the ability to fit non-convex functions after multiple layers of combination. For example, using the MMN activation function to fit a w-shaped non-convex function, set the two-layer MMN to complete the fitting of the non-convex function, and Maxout of each layer uses two linear functions to characterise the fit, if the image features In the process of extraction, when it is necessary to fit more complex non-convex functions, it is possible to increase the number of combinations in Maxout [33].

The MMN features are as follows:

1 Due to the nonlinear fitting ability of MMN, the nonlinear ability of CNN is strengthened. In the process of image feature extraction, deep features are easily extracted.

2 The basis of the MMN activation function is the Maxout activation function, so it is not prone to over-saturation and low learning efficiency, which speeds up the training speed and extraction speed of CNN in the image feature extraction process.

3 MMN itself has the characteristics of self-learning and self-training. In the process of CNN algorithm training, MMN can participate in the training of image feature extraction together with CNN, so that deeper image features can be extracted.

Because the training of the MMN activation function takes up extra time overhead, and the deep feature extracted at the same time occupies a certain storage overhead, the training time increases during the image feature extraction process. At the same time, a large number of image features are extracted, and the feature points are it may be a mismatch point and needs further analysis.

3.2 Parameter adaptive initialisation based on MMN activation function

CNN has a forward propagation process and a backward propagation process. The MMN activation function can effectively avoid or reduce the gradient problem in the back propagation process, and at the same time increase the feature extraction ability of the image, and participate in the CNN during the training process. The parameter optimisation process increases the number of image feature points extracted. The current Maxout parameter initialisation is not fully applicable to image extraction. Therefore, this paper proposes a parameter initialisation algorithm based on multi-layer Maxout joint optimisation. The following algorithm introduces the process of CNN training [33].

3.2.1 Beamlet transform feature extraction

In the process of Beamlet transformation feature extraction, set the input vector as x_1 and the parameter vector as x_2. The two vector parameters are subject to the same distribution function $f(x_1, x_2)$. dl is an independent distribution function symmetric about the zero point, then the response function of the Beamlet transformation can be expressed for:

$$T_f(b_z) = \int_b f(x(l))dl \quad b_z \in \beta_{n,\delta} \tag{4}$$

In formula (4), b_z is the Beamlet of a certain direction i_2 and position z under a certain scale i_1, $x(l)$ is the initial feature parameter of the image, and $\beta_{n,\delta}$ is the set of Beamlet bases under all scales. The output of each node obtained by $f(x_1, x_2)$-pair interpolation is:

$$\phi(x_1, x_2) = \sum_{i_1,i_2} f_{i_1,i_2} f(x_1, x_2) \tag{5}$$

Because the Beamlet transformation function itself takes up extra time overhead for training, and the extracted deep features take up a certain amount of storage overhead, the variance of equation (4) is:

$$\Delta T = \frac{T_f(b_z) - \Delta t}{n} \tag{6}$$

In equation (6), Δt represents the pixel-level average of the $f(x_1, x_2)$ function. Suppose there is an $n \times n$ noisy image, the image may contain line segments of unknown length, direction and position. The pixel-level data of a noisy image is expressed as:

$$y_{i_1,i_2}, 0 \le i_1, i_2 < n \tag{7}$$

In order to find the internal relationship between the mean $E\left[x_l^2\right]$ and the bias $Var[z_{l-1}]$ of each level of the vector, set x_l as:

$$x_l = \frac{z_{l-1,1} + z_{l-1,2} + \left| z_{l-1,1} - z_{l-1,2} \right|}{2} \tag{8}$$

In formula (8), $z_{l-1,1}$ and $z_{l-1,2}$ are both symmetrical distribution functions about the zero point, then the calculation of $E\left[x_l^2\right]$ has:

$$E\left[x_l^2\right] = \frac{1}{2}\left(\frac{\left[Var\left[z_{l-1,1}\right] - x_l\right]^2 + \left[Var\left[z_{l-1,2}\right] - x_l\right]^2}{n} \right) \tag{9}$$

Since $z_{l-1,1}$ and $z_{l-1,2}$ obey the same distribution function, in order to simplify the calculation, a point v_0, v_1 is taken on the boundary of the image sub-block, so that $\overline{v_0}\overline{v_1}$ is a Beamlet. Perform a test based on the largest Beamlet statistics, that is, in the sub-block, search for Beamlets that meet the following conditions:

$$z_{i_1,i_2} = \frac{A \cdot \phi\left(x_1, x_2\right)}{y_{i_1,i_2}} \tag{10}$$

In equation (10), A is a known amplitude parameter. In the Beamlet transformation process of the image along $T_y\left(b\right)$, in order that the initial state of each node conforms to the same distribution function, then set:

$$\tilde{\phi}_{i_1,i_2} = \tilde{\phi}\left(i_1, i_2; \overline{v}_0\overline{v}_1\right) \tag{11}$$

If the initial state of each node satisfies formula (11), then:

$$f_{i_1,i_2} = Ave\left\{ f \left| Pixel(i_1, i_2) \right. \right\} \tag{12}$$

According to the hypothesis analysis of Beamlet transform feature extraction, when each node obeys the same distribution function, the initialisation requirement of the feature extraction function in this paper is:

$$Y^+ = \max\left\{ T_y\left(b\right) / \sqrt{f_{i_1,i_2}} : b \in \beta_{n,\delta} \right\} \tag{13}$$

3.2.2 Image feature detection under Beamlet transform

In image feature detection, the edge of the image is the performance of the discontinuous gray value. Therefore, the key point to be considered in the Beamletb-based transformation method is the continuous transformation of each image scale space. The transformation gradient is defined as follows

$$\Delta x_l = W_l \overset{\Delta}{\Delta} h_l + Loss \tag{14}$$

In formula (14), Δx_l and Δh_l are recorded as the image edge gradient index, W_l is the weighted average of image gray levels, so that Δx_l corresponds to Δh_l, and $Loss$ is the image loss function in the Beamlet transformation process. Here, suppose Δh_l and W_l are independently distributed, and the two are symmetrical with respect to 0 during the initialisation process. Then, when 1 takes any value, there is $E\left[\Delta x_l\right] = 0$.

In the image sub-blocks, the edges of the Beamlet in various directions are searched along the transformation gradient, and the most obvious feature parameters of the edge features are compared. When there is $h_l = \max\left(z_{l,1}, z_{l,2}\right)$, there is

$$\Delta z_{l,k} = \frac{f'\left(z_{l,\mathrm{k}}\right)^{2-j}}{\theta \cdot \Delta x_{l+1}}, \quad k \in \{0, \pi\} \tag{15}$$

In formula (15), $f'\left(z_{l,\mathrm{k}}\right) = 1$, θ is the intersection angle of all the marking points. Since $f'\left(z_{l,\mathrm{k}}\right) = 1$ and Δx_{l+1} conform to the independent distribution, just take $k \in \{0, \pi\}$ and calculate the Beamlet angle in each direction in turn, then

$$E\left[\left(\Delta h_l\right)^2\right] = \frac{1}{2}\left|Bea\left[\Delta h_l\right]\right|_2^2 - \sqrt{2}\left(\Delta z_{l,k} - Bea\left[\Delta x_{l+1}\right]\right) \tag{16}$$

Establish the relationship between $Bea[\Delta h_l]$ and $Bea[\Delta x_{l+1}]$ through the above formula. If there is L beamlet in the current direction, calculate the parameters of each beamlet, then the feature parameter weight of each sub-block in the image can be obtained as

$$Bea\left[\Delta x_2\right] = \sqrt[3]{Bea\left[\Delta x_{L+1}\right]} + \frac{E\left[\left(\Delta h_l\right)^2\right]}{L \cdot T(b)} \tag{17}$$

In equation (17), $T(b)$ represents the Beamlet transformation value, combined with the Beamlet's starting point and end point coordinate distribution law, calculates the maximum feature value in the current image block

$$\hat{d}_l Bea\left[W_l\right] = \frac{\left|C_{s1} - C_{s2}\right| + \left|C_{s1} - C_{s3}\right|}{Bea\left[\Delta x_2\right]} \tag{18}$$

In equation (18), C_{s1} represents the characteristic parameter of the current Beamlet, C_{s2} and C_{s3} are the other two adjacent Beamlet characteristic parameters. In the process of edge feature detection, it is necessary to capture the sudden change of the image grey value, set the detection threshold to Tk, traverse all image sub-blocks, determine the starting point and end point pixels, use the difference calculation method to calculate all pixels, and analyse according to hypotheses, The gray value of each pixel grid corresponding pixel is

$$B_l \approx \sum_{l=\pi} \hat{d}_l \, Bea\left[W_l\right] - Tk\left(f_k + l_k\right) \tag{19}$$

Integrating the process of image feature extraction and image edge detection of Beamlet transform, all boundary points are marked at a certain interval ω in a clockwise direction, and the identification results of all characteristic pixels in the image block are obtained as:

$$Z_l = B_l - \omega\left[1, Max\left(C_{s1} + C_{s2} + C_{s3}\right)\right] \tag{20}$$

Proof of the time complexity of the algorithm.

The value of the mark interval ω will affect the accuracy and complexity of the algorithm. Increasing ω will decrease the detection accuracy; conversely, decreasing ω

will increase the detection accuracy, but also increase the computational complexity. Assuming that the digital image specification is $m \times m$, J is the maximum scale, the number of operations is set to β_i, the input image is represented as u during feature extraction, and the number of images is I, then the time cost of the algorithm T is a function related to I and u, which is recorded as $T(u,I)$. Therefore, the time cost of image feature detection based on Beamlet is

$$T(u,I) = Z_l - \sum_{i=1}^{m^2} \left(\frac{\omega}{u} \right) \beta_i(u,I) \tag{21}$$

In practical applications, the optimal value of ω is 2^{-J}, which can control the complexity of the algorithm to the greatest extent and improve the extraction accuracy. Compared with literature [34], the feature extraction algorithm proposed in this paper has lower time complexity and better results.

4 Feature extraction based on adaptive initialisation of LSTM-MMN parameters

4.1 Feature extraction of adaptive initialisation of CNN parameters

CNN has three ideas: local receptive fields (local receptive fields) shared weights pooling. In this paper, we introduce long short term memory network (LSTM) and shallow learning selective drop out (SD) algorithm [35] to avoid the over fitting problem of CNN algorithm. The data in the network database is x, and the noise reduction automatic encoder changes the input data $\tilde{x}$ into the noisy $\tilde{x}$. $\tilde{x}$ is mapped according to the mapping function of the following formula:

$$y = f\left(W\tilde{x} + b\right) \tag{22}$$

In the above formula, W and b are the parameters of the activation function, and f is the activation function. The loss function of the encoder is the cross entropy function of the following formula.

$$T = \frac{1}{m} \sum_{i=1}^{m} \left[-y_i \log a_i - \left(1 - y_i\right) \log\left(1 - a_i\right) \right] \tag{23}$$

In the above formula, m is the number of nodes in the neural network, a_i is the value of the previous activation function, and y_i is the output value of the previous node. When the encoder is used to reconstruct the data in the database, it is necessary to add a certain amount of noise data [2] artificially to avoid over fitting. Gaussian noise $\tilde{x} \middle| x - N\left(x, \Sigma\right)$ is selected as the noise reduction automatic encoder designed in this paper, and the noise is artificially increased when the encoder is running.

There is a certain correlation between the data in the network database, which makes the data appear hierarchical distribution. If S_{sgh} is an arbitrary data class and the integration relationship of data attributes is k_{drt}, the data is clustered according to the data distribution in the network database.

$$e_{ser} = \frac{f_{ser} \oplus k_{drt}}{S_{sgh}} \oplus m_{plk} \times v \tag{24}$$

In the above formula, f_{ser} is the data distribution layer of the network database, m_{plk} is the vertex set of the data distribution layer, and v is the data feature type in the database. After data clustering, the data feature types are identified according to the function shown in the following formula [4].

$$W_{kpwe} = \frac{k_{sg} * s_{fg}}{v_{lpk} \pm r_{sgj}} \oplus Q \tag{25}$$

In the above formula, k_{sg} is the data dimension, s_{fg} is the feature data, Q is the maximum divergence difference criterion function, r_{sgj} is the maximum weight value of the feature type, and v_{lpk} is the eigenvector space. The data feature cluster corresponding to the largest value of W_{kpwe} is the data feature type. According to the above process, the data features in the database can be extracted. So far, the research on feature extraction method of massive data in network database has been completed.

In order to obtain proper weight initialisation in deep LSTM algorithm. In this paper, a greedy layer by layer LSTM training algorithm is proposed, which uses the LSTM's own automatic encoder for learning, so as to ensure that the gradient of each layer is constant, and at the same time, the weight distribution of the activation function is reasonable. In the process of relative input and output, the output results of the upper layer are sent to the next LSTM automatic encoder for learning, the constant gradient between layers is adjusted by weight. Under the training and learning of supervised tasks, the problem of gradient disappearing or exploding can be avoided, which is better and faster than the random initialisation algorithm. The specific ideas are as follows:

1 The first LSTM layer is sent to the automatic encoder for training and learning, and the input sequence is also used as the input of decoding LSTM.

2 When encoding LSTM, the hidden factor of the current layer is taken as the input of the next automatic encoder. At the same time, when decoding in the current layer, the original input sequence should be recovered together.

3 The second step is executed iteratively until the required number level is initialised.

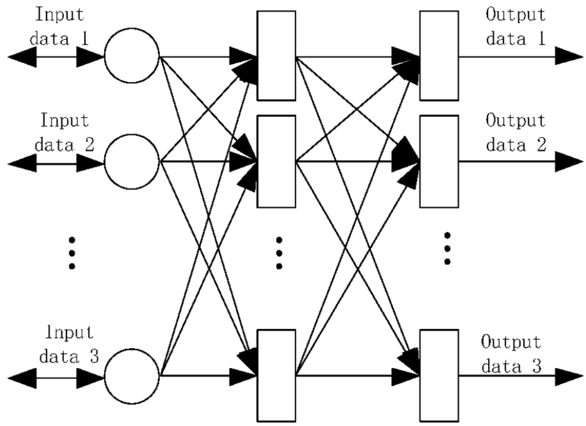
4 The hidden output of the last layer is taken as the input of the supervision layer.

5 The parameters of deep structure are fine tuned by supervised loss function.

Feature mapping is mainly used for image pre-processing and feature extraction, and then the generated feature vector is trained by CNN with parameter adaptive initialisation. The basic schematic diagram is shown in Figure 1.

The main processes of feature extraction based on adaptive LSTM-MMN algorithm include image acquisition and preprocessing, feature detection and feature extraction. Firstly, the original image is preprocessed, the sequence k of the image is extracted, the time k is selected, the image is localised and segmented by the fast convolutional neural network, and then the segmented image is convolved by the CNN of the MMN initialisation parameter. And pooling training. Then, the average k image features of each layer are averaged and sampled, and then the depth of the image features is extracted.

Figure 1 Feature mapping



4.2 Precise feature matching based on CNN-susan operator

In order to avoid the unreliability of matching caused by this problem, in order to avoid the over fitting phenomenon after the optimisation of deep learning parameters, this paper introduces a small single value segment assimilating nucleus (Susan) operator [27], which has strong anti noise ability and low time complexity, to achieve a good registration effect. At the same time, the stability of the adaptive CNN algorithm is also guaranteed [28].

From the mathematical point of view, the multi matching problem can be defined as a set, including k patterns, which can be expressed as $D=\{pat_1,\cdots,pat_k\}$ and the length is $M=\sum_{i=1}^{k}|pat_i|, 1<i\le k$. If all the models in D are linear with M, a text with length of n'' is included. Then, in a certain period of $O(n''+tocc)$, we can find patterns that match the text content, among which $tocc$ is the total number of times in D. At this time, the start and end positions of the matching pattern can be recorded.

For any X'' and Y'' strings, the distance between the two strings is described according to $D(x'',y'')$, which represents the minimum number of times to edit when string C is converted to string Y''. The corresponding editing method is to delete, replace and insert characters in X''. Therefore, the distance between X'' and Y'' is symmetric and nonnegative, and can satisfy the trigonometric inequality.

In addition to the above, we define the pattern strings $pat[1:l'']$ and $text[1:z'']$ with lengths of l'' and z'', in which $z''>l''$ has a positive integer $k(k\in[0,m''))$. At $D(x'',y'')\le k$, pat has all matched termination positions $i(i\in[1,n''])$ in $text$. Assuming that the characters are all from the finite dictionary table $\sum$, $\sum^{n''}$ describes a string of length n'', which is composed of symbols in $\sum$. To sum up, we can use the value of f'' to complete the approximate string matching, in which k differences are allowed.

$$f'':\Sigma^{m''}\Sigma^{n''}\{0,1\}n''-m''+k+1 \tag{26}$$

Among them, $f''(pat,text,k)=c_{m''-k}c_{m''-k+1}\cdots c_{n''}$, at the same time under $1<i\le j$ and $m''-k<j\le n''$ are as follows:

$$c_j = \begin{cases} 1, D\left(pat, text\left[i : j\right] \le k\right) \\ 0, \text{else} \end{cases} \tag{27}$$

Finally, according to the dynamic programming method, the edit distance matrix D of $\left(m'' + 1\right) \times \left(n'' + 1\right)$ is constructed, and the values of each element are solved to realise approximate string matching. The calculation method is as follows:

$$D_{(i,j)} = \begin{cases} 0, i = 0 \\ i, j = 0 \\ \min \begin{Bmatrix} D\left[i, j-1\right] + 1, D\left[i-1, j\right] + 1, \\ D\left[i-1, j-1\right] + c \end{Bmatrix} \\ \text{other}, c = \begin{cases} 0, pat\left[i\right] = text\left[j\right] \\ 1, pat\left[i\right] \ne text\left[j\right] \end{cases} \end{cases} \tag{28}$$

The multi pattern matching algorithm based on finite automata, namely AC algorithm, can complete the matching and traversal of intrusion data through the failure, output and goto function in the run-time preprocessing stage of the algorithm, so as to determine the location of the matching data and all the items related to it, and then realise the final matching of network intrusion patterns.

5 Results and discussion

5.1 Adaptive initialisation of CNN algorithm characteristics experiment

The feature extraction algorithm based on parameter adaptive initialisation CNN has the characteristics of extracting many feature points and high training precision. To illustrate the superiority of the algorithm, this paper uses Matlab2014 programming, computing memory is 8G. The characteristic experiment screenshot is shown in Figure 2. It can be seen from Figure 2 that the accuracy of the algorithm in the training data process is above 90%, so the training ability of the deep learning algorithm to extract features is very strong.

Figure 2 Adaptive initialisation of running parameters of CNN image feature extraction (see online version for colours)

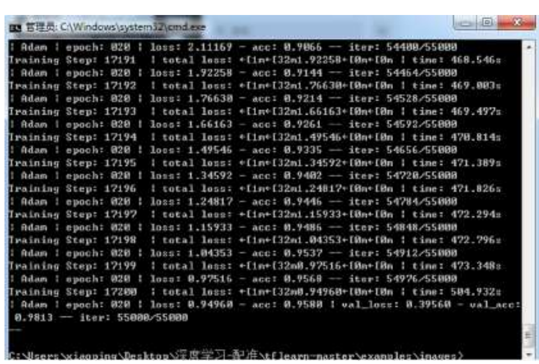
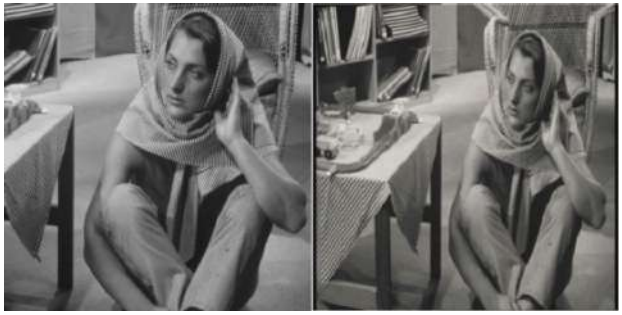


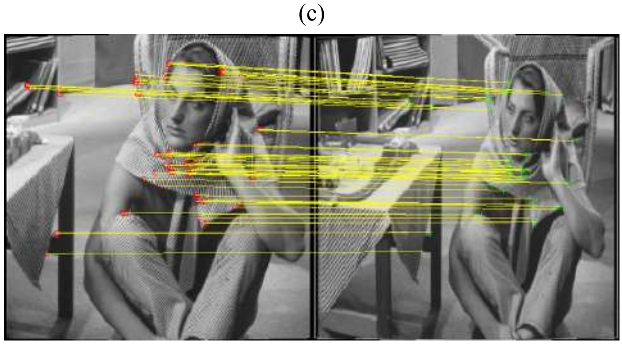
Figure 3 Matching result graph of two algorithms on standard images: (a) test image 1, (b) CNN-SUSAN algorithm, (c) match results after RANSAC purification and (d) SIFT matching results (see online version for colours)



(a)

(b)

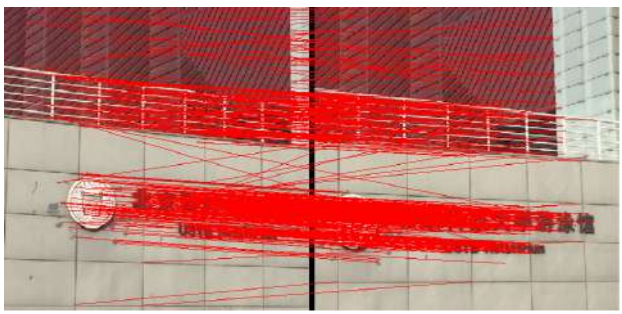
(c)



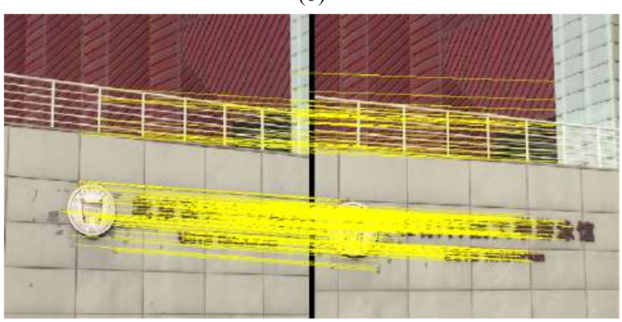
(d)

Figure 4 Matching result graph of actual algorithm image by two algorithms: (a) test image 2, (b) CNN-SUSAN algorithm, (c) match results after RANSAC purification and (d) SIFT matching results (see online version for colours)

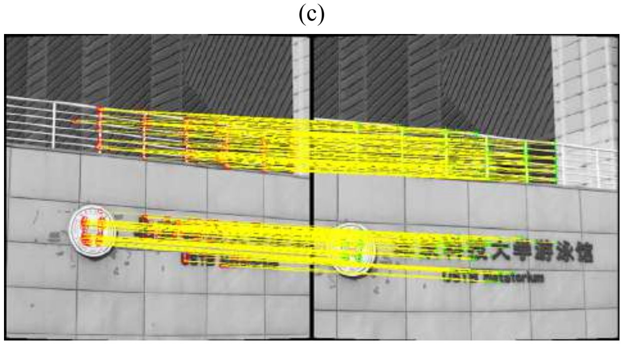
(a)



(b)



(c)



(d)

5.2 *Image test library instance analysis*

The test images in this paper are from the Brodatz image database [36] and the actual scene images. The actual scene data are taken by camera, and the natatorium of University of Science and Technology Beijing is actually captured from different angles. Then the image is standardised and the resolution is adjusted to 512 × 512. The standard library image and the actual scene image are matched by the algorithm and the SIFT algorithm. The matching results are shown in Figure 3 to Figure 4, and the relevant test data of the two algorithms for each set of test images are shown in Tables 1 and 2.

From Tables 1 and 2 and the actual matching results, the matching speed and matching accuracy of the algorithm are improved. The adaptive parameter initialisation based on the MMN linear activation function initialises the convolutional neural network to improve the matching point search speed and matching accuracy is valid. The algorithm can not only meet the accuracy and reliability requirements of image feature extraction, but also improve the image feature extraction time while improving the feature extraction accuracy.

5.3 *Analysis of actual image examples in the underground*

The test images in this paper are from the Brodatz image database [36] and the actual scene images. The actual scene data are also taken by camera. The underground roadway data of a mine of Jizhong Energy Group are collected from different angles, and then the images are standardised and the image resolution is adjusted to 512 × 512. The downhole image is matched by the algorithm and SIFT algorithm. The matching results are shown in Figures 5 and 6. The relevant test data of the two algorithms for each set of test images are shown in Table 3.

From the actual matching results of the downhole test images in Table 3 and Figure 6, the matching speed and matching accuracy of the proposed algorithm are improved compared with the SIFT algorithm. The adaptive parameter initialisation based on the MMN linear activation function is used to initialise the convolutional neural network. The use is effective for improving the matching point search speed and matching accuracy. The algorithm can not only meet the accuracy and reliability requirements of image feature extraction, but also improve the image feature extraction time while improving the feature extraction accuracy.

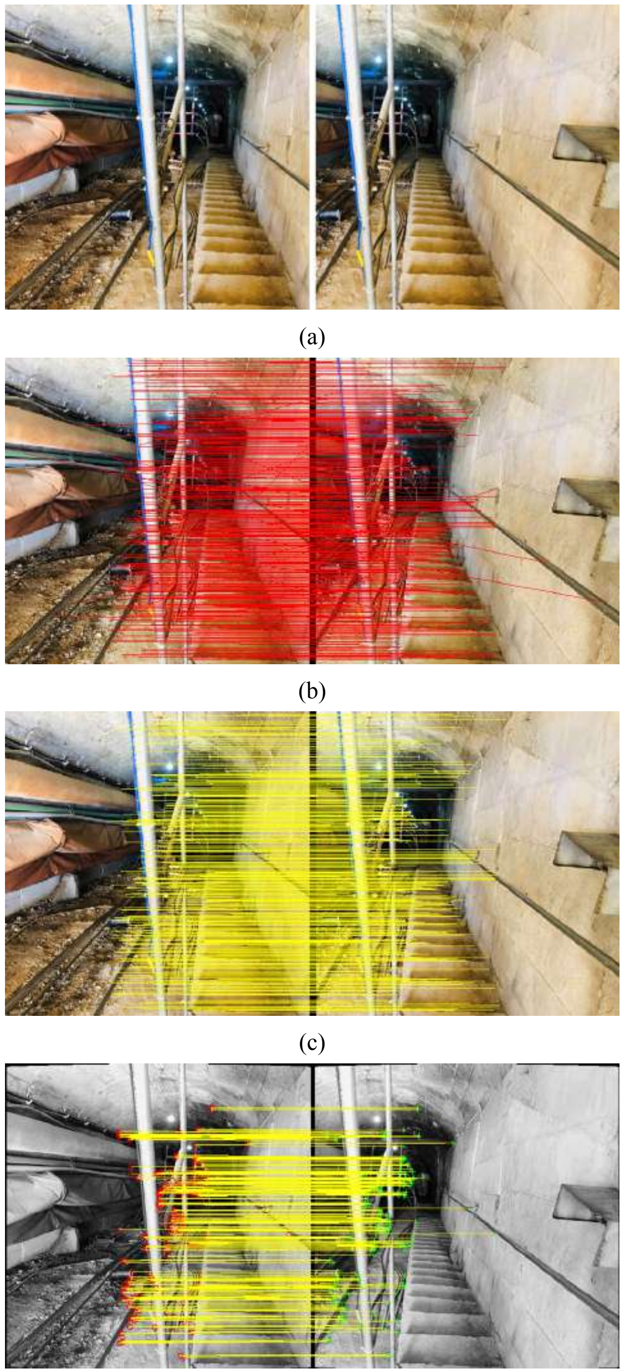
Table 1 Standard image matching data statistics

Algorithm	*Number of features*	*Time overhead* (S)	*Total number of matches*	*Correct number*	*Match rate* (%)
The paper	3278	2701.6	86	72	84
SIFT	3012	4223.1	110	60	55

Table 2 Actual scene image matching data statistics

Algorithm	*Number of features*	*Time overhead* (S)	*Total number of matches*	*Correct number*	*Match rate* (%)
The paper	4117	2709.2	91	62	68
SIFT	3922	4239.1	132	56	42

Figure 5 Matching results of two algorithms for downhole test image 1: (a) downhole test image 1, (b) downhole test image 1 matching of CNN-SUSAN algorithm, (c) registration result after RANSAC purification of downhole test image 1 and (d) SIFT matching result of downhole test image 1 (see online version for colours)



(a)

(b)

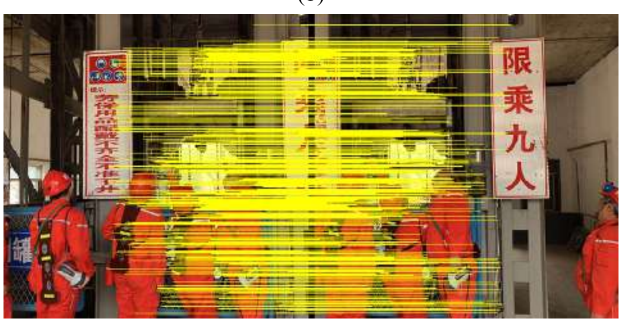
(c)

(d)

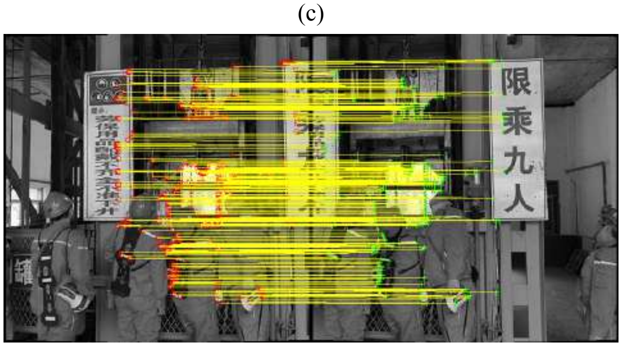
Figure 6 Matching results of two algorithms for downhole test image 2: (a) downhole test image 2, (b) downhole test image 2 matching of CNN-SUSAN algorithm, (c) registration result after RANSAC purification of downhole test image 2 and (d) SIFT matching result of downhole test image 2 (see online version for colours)

(a)

(b)



(c)



(d)

Table 3 Downhole test image matching data statistics

Image	*Algorithm*	*Number of features*	*Time overhead* (S)	*Total number of matches*	*Correct number*	*Match rate* (%)
Downhole image 1	The paper	4125	2697.2	165	142	86
	SIFT	3866	4158.7	158	98	62
Downhole image 2	The paper	4293	2756.2	187	166	89
	SIFT	4012	4365.2	162	94	58

5.4 Time complexity analysis of the algorithm

Assuming that the feature extraction algorithm in this paper is A, the input image is represented as I during feature extraction, and the number of images is A, then the time cost of algorithm A, T is a function related to I and I, denoted as $T(m,I)$. Assuming λ_i is the number of runs required for the O_i basic operation O_i, then is the function related to I and I, called $\lambda_i(m,I)$.Therefore, if the time cost of performing O_i basic operation O_i is α_i, then the time cost of feature extraction algorithm I when inputting image I is:

$$T(n,I)=\sum\nolimits_{i=1}^{m}\alpha_i\lambda_i(m,I) \tag{29}$$

According to the above formula, the time complexity of the feature extraction algorithm proposed in this paper is $O(n^2)$.

6 Conclusion

In this paper, an image feature extraction algorithm based on MMN linear activation function adaptively initialising convolutional neural network parameters is proposed. Firstly, a parameter initialisation algorithm based on multi-layer Maxout activation function is proposed, which solves the problem of poor recognition due to improper parameter initialisation. Secondly, the long-short-time memory network-shallow learning selective Dropout algorithm is introduced to solve the CNN algorithm. Over-fitting the problem, then introducing the SUSAN operator and the random consistency algorithm for fine matching and purification. Finally, the matching is verified. The adaptive parameter initialisation CNN algorithm proposed in this paper can effectively overcome the over-fitting problem in the deep learning training process, avoid the problems such as gradient explosion caused by improper parameter initialisation, improve the calculation speed and extraction precision of image feature extraction, and pass the verification and Simulation, the image feature extraction algorithm proposed in this paper can get more image detail information and edge information, and can better reflect image feature information, which lays a good foundation for image fusion and image recognition.

In the feature extraction of standard images and actual underground images, although this method has achieved good results, it has not been extended to the feature extraction of contaminated images. The next step will be the research on feature extraction and image restoration of contaminated images.

References

1 Moravec, H.P. (1977) 'Towards automatic visual obstacle avoidance', *Proceedings of the 5th International Joint Conference on Artificial Intelligence – Volume 2*, Morgan Kaufmann Publishers Inc., Ottawa, pp.584–585.

2 Poggio, T., Torre, V. and Koch, C. (1987) 'Computational vision and regularization theory', *Readings in Computer Vision*, Morgan Kaufmann, Amsterdam, pp.638–643.

3 Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) 'ImageNet classification with deep convolutional neural networks', Advances in *Neural Information Processing Systems*, Vol. 25, No. 2, pp.1097–1105.

4 Fukushima, K., Miyake, S. and lto, T. (1982) 'Neocognitron: a neural network model for a mechanism of visual pattern recognition', *Competition and Cooperation in Neural Nets*, Springer Berlin Heidelberg, pp.826–834.

5 Dickmanns, E.D. (2002) 'The development of machine vision for road vehicles in the last decade', *Intelligent Vehicle Symposium, 2002*, IEEE, Vol. 1, pp.268–281.

6 Xu, C., Lei, L., Song, J. and Li,Y. and Zhuang, Z. (2018) Corner detection and matching for infrared image based on double ring mask and adaptive SUSAN algorithm', *Opt. Quantum Electron.*, Vol. 50, No. 4, p.194.

7 Singh, V. and Misra, A.K. (2017) 'Detection of plant leaf diseases using image segmentation and soft computing techniques', *Inf. Process. Agric.*, Vol. 4, No. 1, pp.41–49.

8 Howard, M., Hock, M.C., Meehan, B.T. and Dresselhaus-Cooper, L. (2018) 'A locally adapting technique for edge detection using image segmentation', *SIAM J. Sci. Comput.*, Vol. 40, No. 4, pp.B1161–B1179.

9 Nguyen, V.D., Nguyen, H.V., Tran, D.T. and Hau, S. (2017) 'Learning framework for robust obstacle detection, recognition, and tracking', *IEEE Trans. Intell. Transp. Syst.*, Vol. 18, No. 6, pp.1633–1646.

10 Kadir, T. and Brady, M. (2001) 'Saliency, scale and image description', *Int. J. Comput. Vision*, Vol. 45, No. 2, pp.83–105.

11 Kadir, T. and Brady, M. (2003) 'Scale saliency: a novel approach to salient feature and scale selection', Vol. 18, No. 2, pp.561–568.

12 Matas, J., Chum, O., Urban, M. and Pajdla, T. (2004) 'Robust wide-baseline stereo from maximally stable extremal regions', *Image Vis. Comput.*, Vol. 22, No. 10, pp.761–767.

13 Mikolajczyk, K. and Schmid, C. (2004) 'Scale & affine invariant interest point detectors', *Int. J. Comput. Vision*, Vol. 60, No. 1, pp.63–86.

14 Zhang, J., Chen, Q., Sun, Q., Sun, H. and Xia, D. (2011) 'A highly repeatable feature detector: improved Harris–Laplace', *Multimedia Tools Appl.*, Vol. 52, No. 1, pp.175–186.

15 Zeiler, M.D. and Fergus, R. (2014) 'Visualizing and understanding convolutional networks', Vol. 8689, pp.818–833.

16 Lowe, D.G. (2004) 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vision*, Vol. 60, No. 2, pp.91–110.

17 Yan, K. and Sukthankar, R. (2004) 'PCA-SIFT: a more distinctive representation for local image descriptors', *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, Washington, pp.506–513.

18 Mikolajczyk, K. and Schmid, C. (2005) 'A performance evaluation of local descriptors', *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27, No. 10, pp.1615–1630.

19 Wang, Z., Fan, B. and Wu, F. (2011) 'Local intensity order pattern for feature description', *2011 International Conference on Computer Vision*, IEEE, Barcelona, pp.603–610.

20 Cai, G.R., Jodoin, P.M., Li, S.Z., Wu, Y.D., Su, S.Z. and Huang, Z.K. (2013) 'Perspective-SIFT: An efficient tool for low-altitude remote sensing image registration', *Signal Process.*, Vol. 93, No. 11, pp.3088–3110.

21 Le, Y.H., Kurkure, U. and Kakadiaris, I.A. (2014) '3D dense local point descriptors for mouse brain gene expression images', *Comput. Med. Imaging Graph.*, Vol. 38, No. 5, pp.326–336.

22 He, Y., Deng, G., Wang, Y. and Jay, B. (2018) 'Optimization of SIFT algorithm for fast-image feature extraction in line-scanning ophthalmoscope', *Optik*, Vol. 152, pp.21–28.

23 Chen, Y., Jiang, H., Li, C., Jia, X. and Ghamisi, P. (2016) 'Deep feature extraction and classification of hyperspectral images based on convolutional neural networks', *IEEE Trans. Geosci. Remote Sens.*, Vol. 54, No. 10, pp.6232–6251.

24 Ding, Y., Deng, R., Xie, X., Xu, X. and Krylov, A.S. (2018) 'No-reference stereoscopic image quality assessment using convolutional neural network for adaptive feature extraction', *IEEE Access*, Vol. 6, pp.37595–37603.

25 Ling, W., Du, C., Mengchao, Z., Wang, Y. and Wang, S. (2018) 'CFD simulation of low-attitude droplets deposition characteristics for UAV based on multi-feature fusion', *IFAC-PapersOnLine*, Vol. 51, No. 17, pp.648–653.

26 Chen, M., Li, Y., Zhang, Z., Hsu, C.H. and Wang, S. (2017) 'Real-time, large-scale duplicate image detection method based on multi-feature fusion', *J. Real-Time Image Process.*, Vol. 13, No. 3, pp.557–570.

27 Chen, X., Wang, S., Zhang, B., Zhang, B. and Luo, L. (2018) 'Multi-feature fusion tree trunk detection and orchard mobile robot localization using camera ultrasonic sensors', *Comput. Electron. Agric.*, Vol. 147, pp.91–108.

28 Wu, J., Zhang, B., Zhou, J., Xiong, Y., Gu, B. and Yang, X. (2019) 'Automatic recognition of ripening tomatoes by combining multi-feature fusion with a bi-layer classification strategy for harvesting robots', *Sensors*, Vol. 19, No. 3, p.612.

29 Zhang, Y. and An, M. (2017) 'Deep learning and transfer learning based super resolution reconstruction from single medical image', *J. Healthc. Eng.*, Vol. 2017, pp.1–20.

30 Wang, L., Zhang, J., Liu, P., Choo, K. and Huang, F. (2017) 'Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification', *Soft Comput.*, Vol. 21, No. 1, pp.213–221.

31 Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F.E. (2017) 'A survey of deep neural network architectures and their applications', *Neurocomputing*, Vol. 234, pp.11–26.

32 Sun, W., Su, F. and Wang, L. (2018) 'Improving deep neural networks with multi-layer maxout networks and a novel initialization method', *Neurocomputing*, Vol. 278, pp.34–40.

33 Ding, Y., Deng, R., Xie, X., Xu, X., Zhao, Y., Chen, X. and Krylov, A.S. (2018) 'No-reference stereoscopic image quality assessment using convolutional neural network for adaptive feature extraction', *IEEE Access*, Vol. 6, pp.37595–37603.

34 Gupta, N. and Kumar, M. (2017) 'Image feature detection using an improved implementation of maximally stable extremal regions for augmented reality applications', *Int. J. Image Data Fusion*, Vol. 9, No. 1, pp. 43–62.

35 Guo, X. and Yuan, S. (2013) 'A new tree pruning SD algorithm to eliminate interference', *International Journal of Digital Content Technology and its Applications*, Vol. 7, No. 7, pp.476–482.

36 Chen, Y. (2020) 'Exploring the impact of similarity model to identify the most similar image from a large image database', *J. Phys.: Conf. Ser.*, Vol. 1639, No. 1, pp.012139–012139.