International Journal of Information Technologies and Systems Approach
Volume 16 • Issue 3

The Optimization of Face Detection Technology Based on Neural Network and Deep Learning

Jian Zhao, Changzhi University, China*

ABSTRACT

Face detection is a biometric technology that automatically contains facial feature information. It integrates digital image processing, pattern recognition, and other technologies and collects images or video streams containing human faces by cameras or cameras for automatic detection and tracking. Starting from the idea of local features and deep learning, aiming at the problem that traditional convolutional neural network (CNN) only extracts features from the whole image and ignores practical local details, this article proposes a deep CNN model based on the fusion of global and local features. It explores the face detection algorithm with better performance under the interference of illumination, expression, and other internal or external factors. This method designs a suitable network structure according to the size of the training data set, and the core technology is the debugging of super parameters. The simulation results show that compared with SVM, the improved CNN has obvious advantages in the later stage of operation, and the error is reduced by 36.85%. Compared with the traditional face detection method, it can automatically extract image features and also automatically learn its model and get a higher recognition rate.

KEYWORDS

Convolutional Neural Network, Face Detection, Image Processing, Local Characteristics

1. INTRODUCTION

With the continuous growth of Internet information, there is a large amount of data on the Internet nowadays, making information security management extremely important. Therefore, the relevant departments have also strengthened their emphasis on information security. In the current situation, the most important personal information is identifying a person (Bong K, et al., 2017). Because of their reliability, universality, simplicity, stability and other advantages, biometric features have attracted more and more attention, and become a research hotspot incomputer vision and pattern recognition (Li Y, et al., 2018). Biometrics technology is to identify people by using the uniqueness and uniqueness of their biological features, including palmprint, iris, fingerprint, facial information, etc (Bong K, et al., 2017). Compared with fingerprint and iris information which are difficult to extract, face features are easy to obtain, easy to capture, easy to handle and non-contact. They are widely used as personal identity authentication information. Face detection is a biometric technology that automatically

DOI: 10.4018/IJITSA.326051 *Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

carries out facial feature information. It integrates digital image processing, pattern recognition and other technologies, and collects images or video streams containing faces by cameras or cameras for automatic detection and tracking (Long B, Yu K, Qin J, 2018).

Face detection refers to the use of face detection technology to identify people's identity information in images and mark it out. Face verification refers to using face detection technology to determine whether the current person to be identified is the same as the preset person (Yin X & Liu X, 2018). When traditional machine learning methods deal with raw data, they need to use a feature selection algorithm to transform the raw data into a more discriminating feature representation. In contrast, deep learning is an end-to-end learning method, which does not need to design features manually, and its learning process is simple, greatly reducing the time spent on feature selection (Deng J, 2017). Compared with the features extracted by traditional image description operators, the features extracted by deep learning are deep, abstract and complex. This in-depth feature can describe all aspects of human face well, and has strong robustness to the interference factors of environment and human face itself (Xiao Y & Xie X, 2019). Based on the idea of local features and deep learning, this article uses CNN to optimize face detection technology based on existing face local feature selection algorithms. It explores the face detection algorithm with better performance under the interference of light, expression and other self or external factors.

Face detection technology is to solve the problem of face image feature extraction. Like other image recognition problems, studying face detection is not only to solve the face detection problem but also to solve the significance of further target recognition. The face detection system has no high requirements for the performance of image acquisition devices. Common image acquisition devices such as mobile phones and cameras can be used for face detection, and other devices are not needed for assistance (Zhang B, 2019). Face detection is mainly based on the fact that human face features are quite different among different individuals, and it is a relatively stable measure for the same person. Because of the complexity of face changes, there are many difficulties in feature expression and feature extraction. Face detection has been widely used in human-computer interaction, entertainment, information security, video surveillance, medicine and finance. In the research of face detection optimization, the main contributions of this article are as follows:

(1) In this article, an optimization model of face detection based on deep neural network is proposed, which shifts the sample blocks with high priority, avoids the sample blocks with a large number of unknown pixels from being processed, and reduces the continuous accumulation of errors caused by matching errors.
(2) When the priority of the sample block is determined, the model only calculates the data items of the sample block to ensure the continuity of the structural features in the repair results. Moreover, the repair process is repaired layer by layer along the damaged area by marking to ensure the credibility of the repair process.

2. RELATED WORK

In the stage of face feature extraction, to prevent the loss of useful potential information, the potential information in the original input data is automatically extracted into a model with good features through certain learning strategies so that the model can abstract the hidden information of the original data layer by layer. The obtained data information becomes abstract with the deepening of layers, and the input data's deep features are transformed many times to obtain data information. Zhang(2019) was the first to propose using geometric features to solve the face detection problem. In this method, the proportion and distance between the main parts of the face are the most essential features, and then KNN algorithm is used to calculate the distance between features to classify the features. This feature extraction method is not automatic, and manual measurement is needed to participate in the stage of feature establishment. Li et al.(2019) used the technique of template matching to construct a template by measuring the distance between the main parts of the face, including eyes, nose, mouth, etc., and then used the template to classify the face images. Ono et al.(2017) further proposed a uniform local binary pattern with constant rotation by limiting the number of changes of 0/1 in the binary coding of characteristic pattern. Cherny et al.(2020) coded the difference information between the gray values of neighboring pixels and the central pixel based on LBP, and proposed a complete LBP. Striova et al.(2021) compared average pixel values of sub-blocks to describe texture features at different scales, and offered a multi-scale block LBP algorithm. Costantini et al.(2021) proposed a local directional pattern algorithm that uses edge response to encode the central pixel. Dong et al.(2017) found that singular values have good stability, and the singular values of face images have geometric invariance. Memik et al.(2016) put forward a face detection method of Fisherface based on LDA, that is, Fisher face method is also called linear discriminant analysi. To make the distance between classes of samples as large as possible and between classes of the first-class samples as small as possible, Fisherface found an optimal mapping direction.

The dimensions of face information features are usually very high, and there will be a lot of redundant information in these features. Therefore, reducing the image features can effectively reduce the computational complexity and improve the algorithm's performance. This paper proposes an optimization model of face detection based on deep neural network. When the priority of sample blocks is determined, only the data items of sample blocks are calculated to ensure the continuity of structural features in the repair results. Moreover, the repair process is repaired layer by layer along the damaged area by marking to ensure the credibility of the repair process. The comparison results are used to identify this person's identity, or the most matching face is found, and the retrieval results are output according to the matching degree.

3. METHODOLOGY

3.1. Local Feature Selection and Deep Learning

Face detection can be divided into three parts: face preprocessing, feature selection and model application. In the training and testing stage, every image is subjected to face pre-processing. The main function of face pre-processing is to extract the face frame and face key points from the face image, normalize the face image, and finally intercept the face part of the image. In order to better realize the function of feature selection, the face feature selection module needs to design the appropriate neural network structure and loss function for the face detection problem, and finally get the face detection network model through training. The key point of face detection is extracting and comparing face features from images. Early face feature selection methods include geometric feature selection and algebraic feature selection. A human face includes eyes, nose, mouth and chin. The geometric description of these organs and the structural relationship among them can be regarded as an essential feature for identification and classification. These features are called geometric features. Face detection can be classified into biometrics, and it can identify people by extracting face feature information. It integrates face detection, face alignment, face verification and other related face analysis technologies. Face detection has many unique advantages compared to other biometric identification methods, so it is widely adopted and used in biometric identification.

This advantage enables the image to be directly used as the network's input, and it is more obvious when the network's input is a multi-dimensional image, thus avoiding the complicated stage of feature selection and data reconstruction in traditional recognition algorithms. Getting the feature face is the first step of face detection, and it is imperative to get the feature face effectively. First, we need to get the set S containing M faces. Each image can be converted into a vector of N dimension. Put these M vectors into a set S :

$$S = \left\{ \Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_n \right\} \tag{1}$$

After obtaining the face vector set, calculate the average image:

$$\varphi_n = \Gamma_n - \psi \tag{2}$$

M orthogonal unit vectors u_k are found, and this kind of unit vectors are used to describe φ_n distribution. Find the eigenvalue λ_k. When u_k is determined, λ_k value is minimum:

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}\left(u_k^T \varphi_n\right)^2 \tag{3}$$

The method of extracting geometric features is to extract feature data that can classify faces according to face description information and feature distance. Its feature indexes usually include Euclidean distance, curvature or angle between face feature points, etc. (Kargioti E, Vouvoudi E, Nannou C, 2021). Linear projection representation and nonlinear projection representation are two common representation methods. The CNN model of face detection is shown in Figure 1.

Let X_i^k represent the sum of inputs of k layer neurons i, Y_i^k is the output, and the weights of $k-1$ layer neurons j to k layer neurons i are W_{ij}, then there is the following functional relationship:

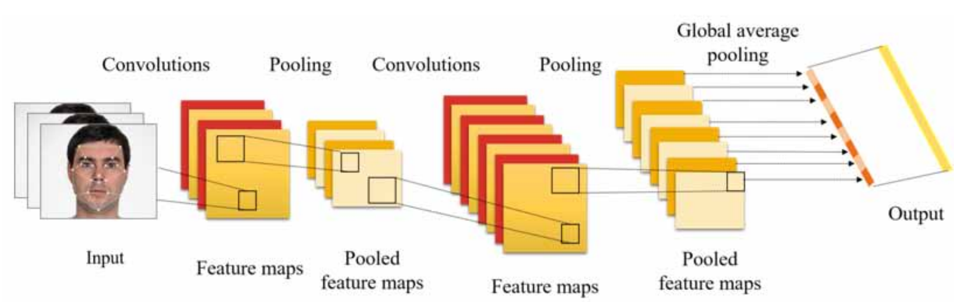
$$Y_i^k = f\left(X_i^k\right) \tag{4}$$

$$X_i^k = \sum_{j=1}^{n+1} W_{ij} Y_j^{k-1} \tag{5}$$

Generally, f A is an asymmetric Sigmoid function:

$$f\left(x_i^k\right) = \frac{1}{1 + \exp\left(-X_i^k\right)} \tag{6}$$

Figure 1. CNN model of face detection



If the output layer is the m layer, the actual output of the i neuron in the output layer is Y_i^m. Let the corresponding human body signal be Y_i, and define the error function e as:

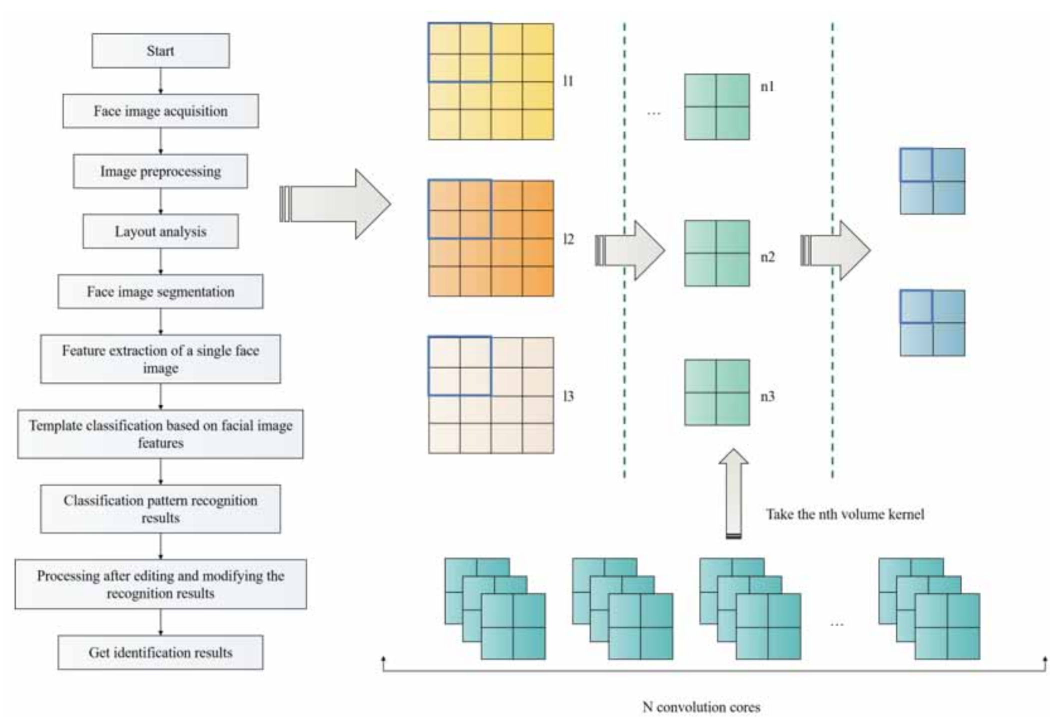
$$e = \frac{1}{2}\sum_i \left(Y_i^m - Y_i\right)^2 \tag{7}$$

Generally speaking, cameras that can be seen everywhere in our life can be used to collect facial images, which does not require the use of particularly complex proprietary equipment, thus greatly reducing the cost of recognition. Collecting face images can usually be completed within a few seconds. The method of deep internal information processing is similar to the mechanism of internal information processing in the human brain, which can learn and fit very complex functions. Importantly, this learned content has strong migration ability.

3.2. Improved Face Detection Algorithm Based on CNN

In an image, the high-level features are combined with the bottom features, and the edges can be combined locally to form a basic pattern. The core layer in CNN is the convolution layer. The local features of the previous layer are extracted by convolution operation of convolution kernel, and different convolution kernels extract different bottom features. The high-level convolution layer extracts specific overall features by combining the rich bottom-edge features, finally completing the recognition task. In CNN, the convolution layer usually appears after the input and pool layers, and is used with its activation function. The convolution operation stage of face detection is shown in Figure 2.

Figure 2. Convolution stage of face detection



In face detection, when a certain method has been selected to extract image features, it is need to determine what is in the deep network. The size of convolution kernel matrix determines the parameters and computation of the network. The convolution kernel is too small, although the network speed is fast, the visual range is small, so it is challenging to extract image features effectively. If the convolution kernel is too large, the detailed information will be ignored because the visible range is too extensive and contains too much global information. Under the condition that the two images are similar. Euclidean distance between pattern vectors and is defined as:

$$D\left(x,y\right)=\left|X-Y\right|=\sqrt{\sum_{i=1}^{n}\left(x_i-y_i\right)^2} \tag{8}$$

Where n is the dimension of the feature space. Obviously, if the samples X and Y are located in the same type area, the Euclidean distance is relatively small.

Let the width of the image be $width$, the height be $height$, the width of the minimum circumscribed rectangular frame be $rect\_width$, the height be $rect\_height$, the center be $center$, and the error redundancy be expressed by $edge$. The horizontal and vertical discrimination formulas are:

$$edge+\frac{rect\_width}{2}<center.x<width-edge-\frac{width}{2} \tag{9}$$

$$edge+\frac{rect\_height}{2}<center.y<height-edge-\frac{width}{2} \tag{10}$$

Let the gray value range of the original face image $f\left(x,y\right)$ be $\left(g_{\min},g_{\max}\right)$, select an appropriate threshold T, and:

$$g_{\min}\le T\le g_{\max} \tag{11}$$

The following formula can represent image segmentation with single threshold:

$$g\left(x,y\right)=\begin{cases}1, & f\left(x,y\right)\ge T\\ 0, & f\left(x,y\right)<T\end{cases} \tag{12}$$

$g\left(x,y\right)$ is a binary image. The object can be easily exposed from the background through binarization, and the key to binarizing the face image is the reasonable selection of the threshold T.

Let the gray function of the two-dimensional face image be $f\left(x,y\right)$. $r\left(r>0\right)$ fields that define $\left(i,j\right)$ position pixels are the following sets:

$$N_r\left(i,j\right)=\left\{\left(k,l\right)\middle|\max\left\{\left|i-k\right|,\left|j-l\right|\right\}\le r\right\} \tag{13}$$

The value defined as follows is called the interest degree of $\left(i,j\right)$ -bit pixels:

$$I\left(i,j\right)=\frac{1}{\left(2r+1\right)^{2}-1}\sum_{\left(k,l\right)\in N_{r}\left(i,j\right)}\left(f\left(i,j\right)+w\left(k,l,\sigma\right)f\left(k,l\right)\right) \tag{14}$$

Among them:

$$w\left(k,l,\sigma\right)=\psi\left(i-k,j-l,\sigma\right) \tag{15}$$

$\psi\left(x,y,\sigma\right)$ is a DOG function.

Moreover, the area that meets the above conditions is considered as the face detection target area that ultimately enters the visual field. Otherwise, the face detection target does not completely enter the field of vision. If one of the conditions is met, the face detection target has completely entered the horizon horizontally or horizontally. To increase the number of times the pixels at the corner edge of the image are used and avoid the loss of the position information of the image edge, padding is often used in the actual convolution layer (Madariaga J M, et al., 2016). In the neural network structure, the expression ability of linear model with linear activation function is not enough. To solve this problem, the nonlinear ability of the network can be enhanced by introducing nonlinear factors into the activation function. Through multi-layer nonlinear transformation of CNN, the change between classes is maximized.

4. RESULT ANALYSIS AND DISCUSSION

When face detection is applied to practical projects, in some cases, it is difficult to label the categories of collected face pictures in advance due to the lack of prior knowledge, so it is impossible to compare whether the predicted classification is the same as the real classification, so it is not suitable to adopt the supervised learning classification method. CNN provides hierarchical display of visual data, and the weight of each layer of network actually includes some components of the image [24]. The higher the number of layers, the more specific information the network contains. CNN's original image input signal is processed layer by layer, and some features of the image are slowly recognized to the whole image. In the pre-processing stage of images, various images are selected, so it is need to change images of different sizes to uniform sizes to build different network structures. In the face detection stage, image preprocessing is considered the most important step in the previous work, because the image quality will directly affect the performance of feature extraction and classification. The input image will have noise interference such as size and illumination. To remove this unnecessary interference information, the image needs to be pre-processed. The common methods are graying, histogram equalization, image smoothing and normalization. Data outlier removal processing is shown in Figure 3.

It is necessary to construct a deep neural network using different stacked modules. Different methods can be used for neural networks with different depths to speed up training, batch processing and constantly update data parameters. The sizes of the original face images in the training and testing sets are different, and the face sizes are not consistent, which greatly impacts the subsequent training and testing process. Therefore, the image preprocessing module is used to normalize the size of the images and align the faces. SVM algorithm is selected as the contrast object, and the experimental results are shown in Table 1 and Table 2.

Figure 3. Data outlier removal processing

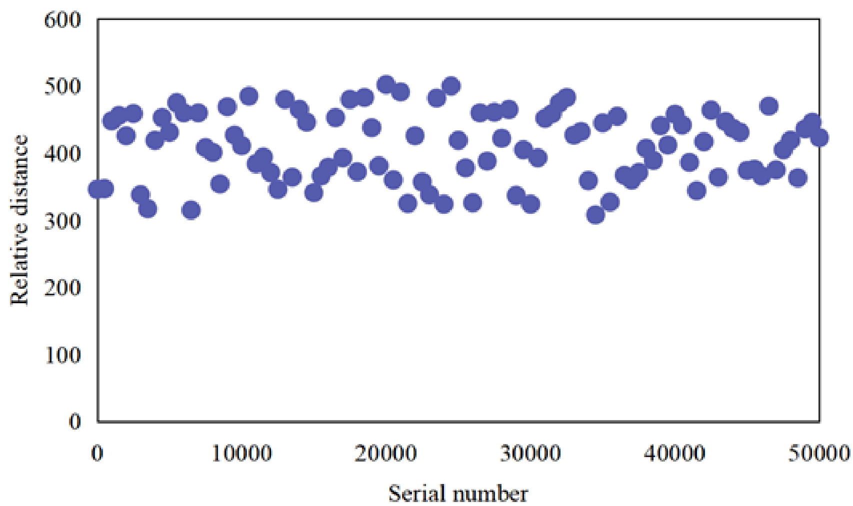


Table 1. Accuracy of human motion face detection by the method in this article

Sample Size	Face Detection Accuracy (%)
15	98.78
30	98.24
45	97.78
60	97.42
75	96.83
90	96.22
105	96.01

Table 2. Accuracy of face detection of human motion based on SVM

Sample Size	Face Detection Accuracy (%)
15	96.61
30	96.01
45	95.65
60	94.27
75	93.52
90	92.99
105	92.45

The experimental data shows that when the number of test samples starts to increase, the accuracy of face detection by the two methods shows a downward trend. However, compared with SVM, the face detection accuracy of this method is higher.

Scatter chart of predicted value and actual value using SVM algorithm is shown in Figure 4. The scatter diagram of predicted value and actual value of face detection model using improved CNN is shown in Figure 5.

It can be analyzed that the face detection model based on improved CNN is better than SVM in both accuracy and efficiency. Interference factors such as light, face posture and face occlusion

Figure 4. Scatter diagram of SVM actual value and predicted value

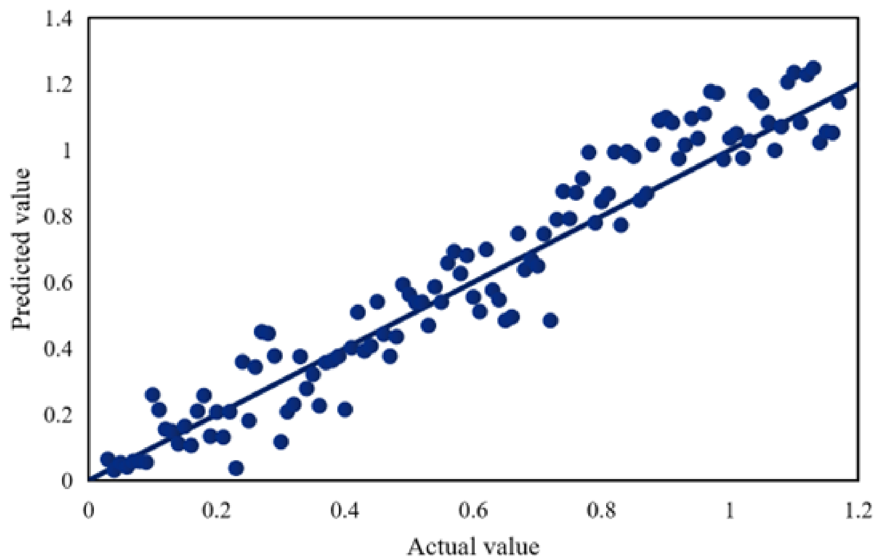
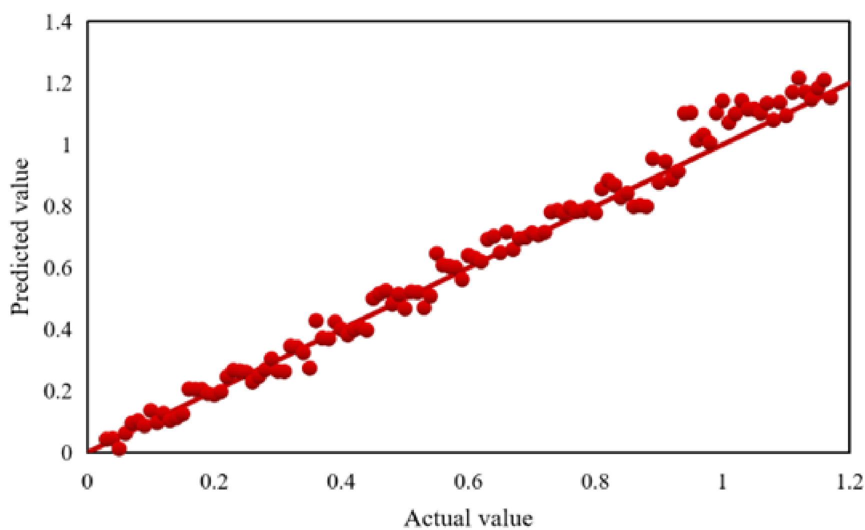


Figure 5. Scatter chart of improved CNN actual value and predicted value



will greatly reduce face detection accuracy. Moreover, the performance of the online real-time face detection system is also related to the hardware computing resources at runtime, and the images with poor computing power will be stuck. Under the condition that the training accuracy of the improved convolution kernel network is lower than that of SVM algorithm, the improved convolution kernel network achieves the same recognition rate as the improved convolution kernel network, and the highest recognition rate is obtained with the minimum loss. The depth feature of automatic learning can be well used for recognition and classification, and with the supplement of local features, the overall model has a better recognition rate, which shows that the local learning method is conducive to improving the classification effect.

The classifier module contains gradient module. First, set the neural network gradient, activate it by activating function, combine it with variables, send it into the classifier module, and save the model. After the feature selection, the output of two-channel face images is weighted and fused. Finally, the fused results are classified by softmax. The comparison of the average absolute errors of the algorithms is shown in Figure 6.

It can be seen that compared with SVM, the improved CNN has obvious advantages in the later stage of operation, and the error is reduced by 36.85%. Compared with other loss functions, the model accuracy of cross-entropy loss function is higher and the model loss is smaller. This may be because the logarithmic characteristic of the cross entropy loss function can measure the distance between image data more accurately in the task of image classification. Compare the accuracy and recall of the algorithm for face detection, as shown in Figure 7 and Figure 8.

The detection results show that the accuracy of this algorithm for face detection is higher, which is 26.55% higher than the comparison algorithm, and the recall is increased by 18.68%. It can accurately locate the edge contour of the face. The system can basically complete the task of real-time detection and analysis, further verifying the model's effectiveness in this article, and providing a foundation for further research on the face detection application.

5. CONCLUSION

With the continuous growth of Internet information, there is a large amount of data on the Internet nowadays, which makes information security management extremely important. Based on the idea

Figure 6. Comparison of average absolute error of algorithms

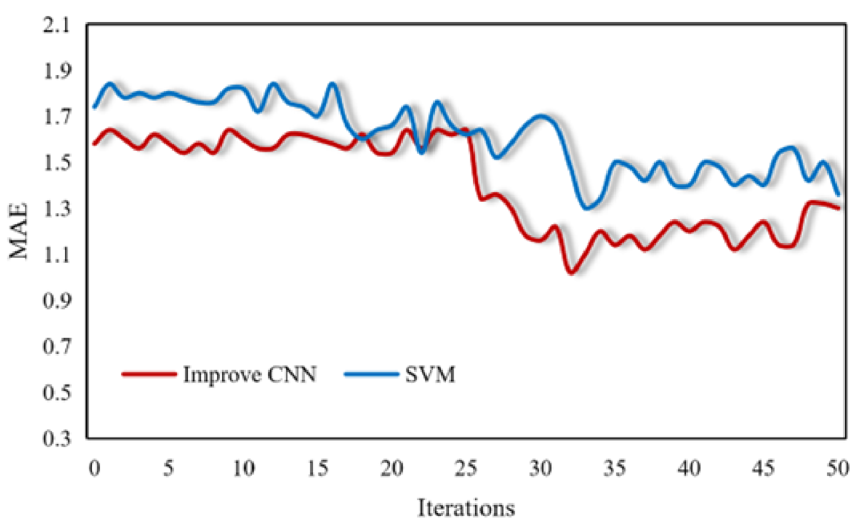


Figure 7. Comparison of face detection accuracy

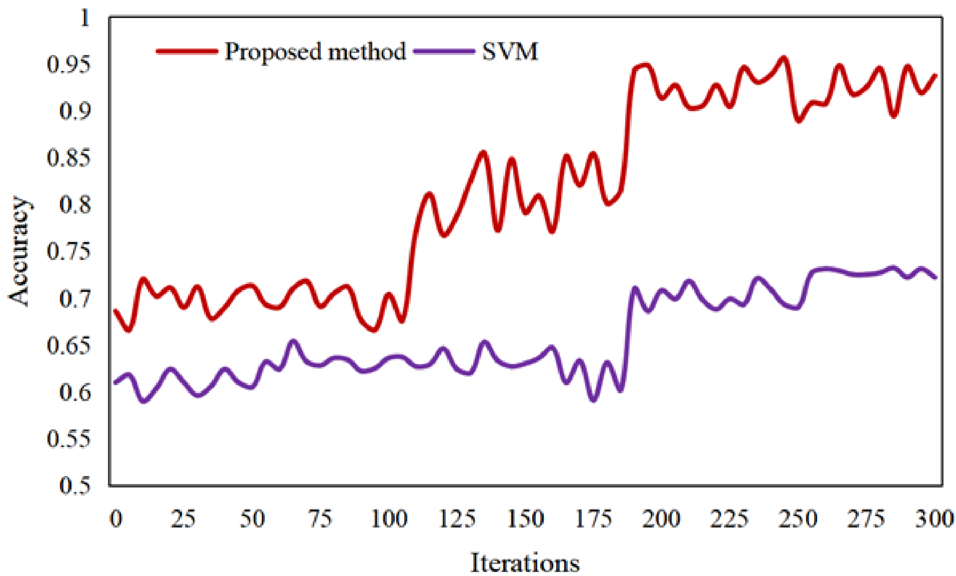
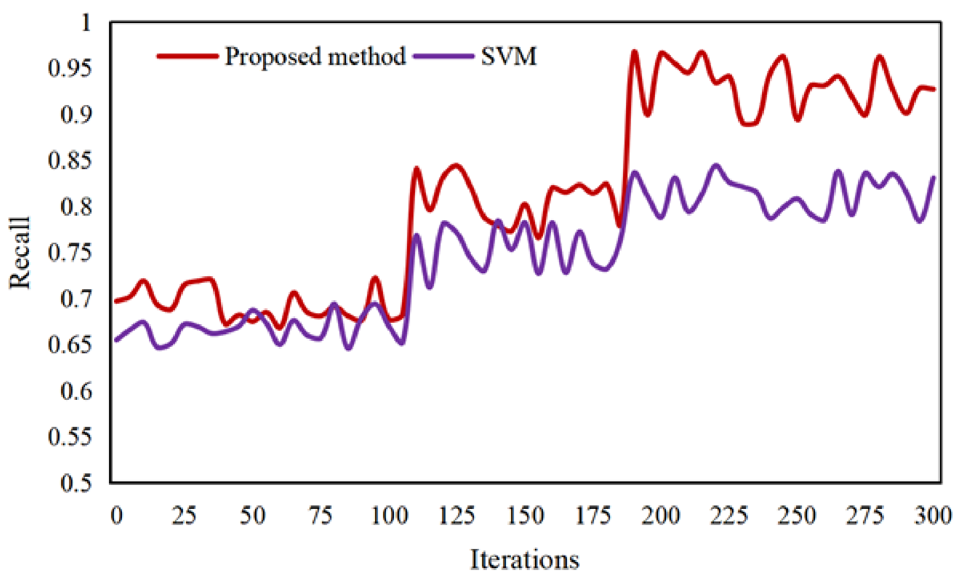


Figure 8. Comparison of recall of face detection



of local features and deep learning, this article uses CNN to optimize face detection technology based on existing face local feature selection algorithms. It explores the face detection algorithm with better performance under the interference of light, expression and other self or external factors. The improved CNN-based face detection model is better than SVM in time measurement accuracy and efficiency. Compared with SVM, the improved CNN has obvious advantages in the later stage of operation, and the error is reduced by 36.85%. The accuracy of this algorithm for face detection is higher, which is 26.55% higher than that of the comparison algorithm, and the recall is increased by 18.68%. It can accurately locate the edge contour of the face. The principle of face detection system is similar to that of natural person. It can increase the system's security by judging the accuracy and other performance of the face detection system by natural person and maintaining the system.

REFERENCES

Bong, K., Choi, S., Kim, C., Han, D., & Yoo, H.-J. (2017). A Low-Power Convolutional Neural Network Face Detection Processor and a CIS Integrated With Always-on Face Detector. *IEEE Journal of Solid-State Circuits*, *53*(1), 115–123. doi:10.1109/JSSC.2017.2767705

Bong, K., Choi, S., Kim, C., & Yoo, H.-J. (2017). Low-Power Convolutional Neural Network Processor for a Face-Recognition System. *IEEE Micro*, *37*(6), 30–38. doi:10.1109/MM.2017.4241350

Cherny, R. W. (2020). San Francisco's New Deal Murals in Long-Term Perspective: Controversy, Neglect, and Restoration. *California History (San Francisco)*, *97*(1), 3–32. doi:10.1525/ch.2020.97.1.3

Costantini, I., Aramendia, J., Tomasini, E., Castro, K., Manuel Madariaga, J., & Arana, G. (2021). Detection of unexpected copper sulfate decay compounds on late Gothic mural paintings: Assessing the threat of environmental impact. *Microchemical Journal*, *169*(1), 106542. doi:10.1016/j.microc.2021.106542

Deng, J. (2017). The application and research of computer image processing and neural network in human face detection. Boletin Tecnico/Technical Bulletin, 55(10), 43-50.

Dong, B., Shen, Z., & Xie, P. (2017). Image Restoration: A General Wavelet Frame Based Model and Its Asymptotic Analysis. *SIAM Journal on Mathematical Analysis*, *49*(1), 421–445. doi:10.1137/16M1064969

Kargioti, E., Vouvoudi, E., Nannou, C., Bikiaris, D., & Lambropoulou, D. (2021). Unraveling the origin of aged varnishes for the proper restoration of old paintings using spectroscopic and spectrometric techniques. *Microchemical Journal*, *168*(1-2), 106467. doi:10.1016/j.microc.2021.106467

Kim, Y. H., Kim, H., Kim, S. W., Kim, H.-Y., & Ko, S.-J. (2017). Illumination normalisation using convolutional neural network with application to face detection. *Electronics Letters*, *53*(6), 399–401. doi:10.1049/el.2017.0023

Li, Y., Zheng, W., Cui, Z., & Zhang, T. (2018). Face Detection based on Recurrent Regression Neural Network. *Neurocomputing*, *297*(7), 50–58. doi:10.1016/j.neucom.2018.02.037

Li, Y. K., Wu, X. J., & Kittler, J. (2019). L1-2D2PCANet: A deep learning network for face detection. *Journal of Electronic Imaging*, *28*(2), 1. doi:10.1117/1.JEI.28.2.023016

Long, B., Yu, K., & Qin, J. (2017). Data Augmentation for Unbalanced Face Detection Training Sets. *Neurocomputing*, *235*(4), 10–14. doi:10.1016/j.neucom.2016.12.013

Madariaga, J. M., Maguregui, M., Castro, K., Knuutinen, U., & Martínez-Arkarazo, I. (2016). Portable Raman, DRIFTS, and XRF Analysis to Diagnose the Conservation State of Two Wall Painting Panels from Pompeii Deposited in the Naples National Archaeological Museum, no. Italy). *Applied Spectroscopy*, *70*(1), 137–146. doi:10.1177/0003702815616589 PMID:26767639

Memik, S. O., Member, S., & Katsaggelos, A. K. (2016). Analysis and FPGA implementation of image restoration under resource constraints. *IEEE Transactions on Computers*, *52*(3), 390–399. doi:10.1109/TC.2003.1183952

Nguyen, D. T., Pham, T. D., & Min, B. L. (2019). Visible-Light Camera Sensor-Based Presentation Attack Detection for Face Detection by Combining Spatial and Temporal Information. *Sensors (Basel)*, *19*(2), 410. doi:10.3390/s19020410 PMID:30669531

Ono, S. (2017). Primal-Dual Plug-and-Play Image Restoration. *IEEE Signal Processing Letters*, *24*(99), 1108–1112. doi:10.1109/LSP.2017.2710233

Shuai, P. (2019). Bo. Image restoration and color fusion of digital microscopes. *Applied Optics*, *58*(9), 2183–2189. doi:10.1364/AO.58.002183 PMID:31044916

Striova, J., Fontana, R., Barbetti, I., Pezzati, L., Fedele, A., & Riminesi, C. (2021). Multisensorial Assessment of Laser Effects on Shellac Applied on Wall Paintings. *Sensors (Basel)*, *21*(10), 3354. doi:10.3390/s21103354 PMID:34065913

Xiao, Y., & Xie, X. (2020). Application of Novel Gabor-DCNN into RGB-D Face Detection. *International Journal of Network Security*, *22*(3), 534–541.

Yin, X., & Liu, X. (2018). Multi-Task Convolutional Neural Network for Pose-Invariant Face Detection. *IEEE Transactions on Image Processing*, *27*(2), 964–975. doi:10.1109/TIP.2017.2765830 PMID:29757739

Zhang, B. (2019). Distributed SVM face detection based on Hadoop. *Cluster Computing*, *22*(1), 1–8.

Zhang, Y., Shang, K., Wang, J., Li, N., & Zhang, M. M. Y. (2018). Patch strategy for deep face detection. *IET Image Processing*, *12*(5), 819–825. doi:10.1049/iet-ipr.2017.1085

Zhang, Z., Chen, X., Wang, B., Hu, G., Zuo, W., & Hancock, E. R. (2019). Face Frontalization Using an Appearance-Flow-Based Convolutional Neural Network. *IEEE Transactions on Image Processing*, *28*(5), 2187–2199. doi:10.1109/TIP.2018.2883554 PMID:30507505