

PAPER • OPEN ACCESS

Research on Face Image Encryption Based on Deep Learning

To cite this article: Yanyan Qin *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **252** 052007

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

Research on Face Image Encryption Based on Deep Learning

Yanyan Qin^a, Chennan Zhang^b, Rui Liang^c and Mingrui Chen^{*}

School of Information Science and Technology, Hainan University, Haikou, China

^{*}Corresponding author e-mail: mrchen@hainu.edu.com, ^a736393722@qq.com, ^b785829754@qq.com, ^c844508341@qq.com

Abstract. With the development of artificial intelligence and big data technology, the requirements for information security are increasing, and the role of biometrics in network security and information security authentication has also increased. Face recognition technology has been widely applied in many Internet payment platforms. This paper proposes a face recognition algorithm based on improved deep network automatic extraction feature, which can extract the discriminative features of the target more accurately and encrypt the face image to ensure the privacy and security of face recognition. In this paper, an automatic deep feature extractor is generated by preprocessing and fine-tuning, and then the hyperchaotic image is encrypted. Several common face databases are used to test in this algorithm and this results show that the algorithm has more availability than the traditional and general deep learning methods in terms of performance.

1. Introduction

Face recognition is one of the most challenging topics in the field of computer vision and machine learning in recent years, and has received extensive attention from researchers. Successful and effective face recognition has broad application prospects and can play a huge role in scenarios such as defense security, video surveillance, human-computer interaction and video indexing. On the one hand, traditional face recognition research is mostly based on handcrafted features. For example, Yang et al. [1] proved the validity of the fusion of geometric features and texture features in face recognition, which is difficult to construct. Deng et al. [2] proposed a PCA (Principal Component Analysis eigenface) with strong adaptability to change, which is used for facial feature representation, but its recognition performance is significantly reduced when the illumination and attitude change are large. The above shallow algorithm is training. It is difficult to express complex functions effectively when samples and computational units are limited [3], revealing the limitations of shallow networks. Experimental studies show that deep network structures are more efficient than shallow structures [4], and literature [5] uses CNN methods. The detection of key points of the face, because of the supervised learning algorithm, requires a large number of labeled samples. On the other hand, Yao et al [6] proposed a chaotic image encryption algorithm, which uses hyperchaos to generate binary sequences to preprocess the image. An encryption algorithm was based proposed by Gao et al. [7] on hyperchaotic system. The core idea of this algorithm is to use the Logistic chaotic map to encrypt and scramble the pixel matrix, and then encrypt the gray value by the key stream generated by hyperchaos.

Based on this, this paper builds an improved deep network based on the deep learning method, and proposes a new effective face image encryption algorithm. Firstly, the face database is preprocessed to



reduce feature correlation. Reduce network computing complexity. Then, an improved deep network feature extractor is constructed, including convolution layer, pooling layer and double-layer sparse automatic coding layer. Before the network training, this paper samples the pictures, uses the unsupervised learning method to learn the network parameters, and obtains various convolution kernels. The face feature obtained by this feature extractor can depict the face in all directions and at multiple levels, which has strong robustness. Finally, using the obtained features, the improved hyperchaotic encryption processing of the face image. The improved deep network structure model can extract the feature information of the face more accurately, and then HIE encryption can effectively avoid the choice of plaintext attack and ciphertext attack.

2. Algorithm Overview

Figure 1 is the algorithm framework diagram of this paper. The improved face network encryption algorithm based on deep network consists of three parts: preprocessing module, deep network feature extractor and encryption processing. The samples are preprocessed by histogram equalization, scale normalization and ZCA whitening, and transformed into image sets with size 64×64 and gray value range $[0, 1]$. The deep network feature extractor extracts the deep feature of the sample and then performs encryption processing.

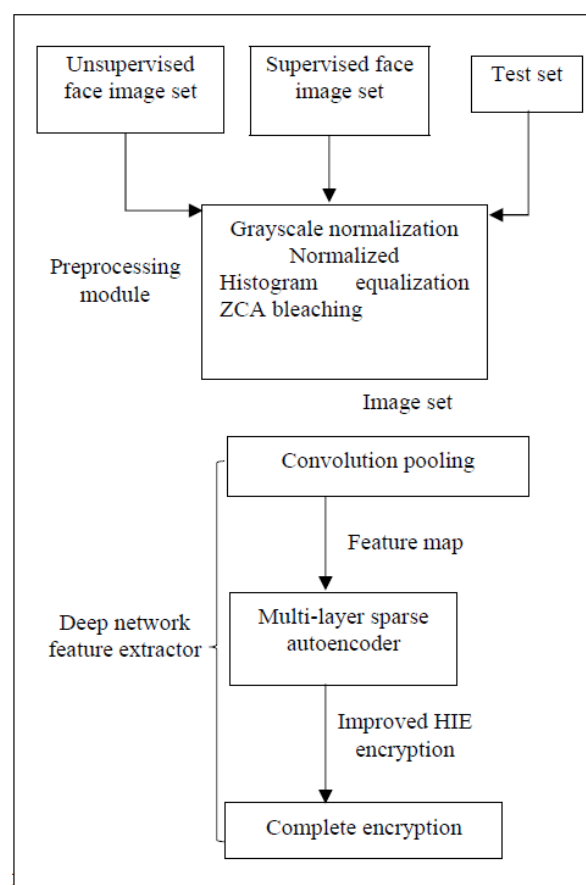


Figure 1. Algorithm Framework Diagram

2.1. Preprocessing Module

The preprocessing module is mainly composed of normalization processing and ZCA whitening. The normalization process uses scale normalization, grayscale normalization and histogram equalization to reduce the computational complexity of the overall network, while reducing the useless information on

the basis of retaining the original information. ZCA whitening is first shown in equation (1), and the correlation between the features is removed by PCA transformation; then the output feature has unit variance using equation (2); then the data is rotated back to obtain the processing result of ZCA whitening. As shown in equation (3), reduce the redundancy of the input.

$$x_{rot,i} = U^T x_i \quad (1)$$

$$x_{PCAWhite,i} = \frac{x_{rot,i}}{\sqrt{\lambda_i}} \quad (2)$$

$$x_{ZCAWhite} = U x_{PCAWhite} \quad (3)$$

The input is obtained by the preprocessing module to obtain an image set with a size of 64×64 and a gray value range of [0, 1]. The image data of the face to be processed thus obtained has a unified format, which can effectively reduce the complexity and computational difficulty of the network, thereby saving network training time; at the same time, reducing the correlation between features, and laying the foundation for the extraction of ideal features.

2.2. Deep Network Feature Extractor

Deep network implementation Deep network is a multi-layer neural network structure with excellent feature learning ability, and its learned features have more essential characterization of data, which helps to classify and visualize. As shown in Fig.2., the deep network includes a convolution layer C1, a pooling layer P2, a noise reduction sparse automatic coding layer E3, a sparse automatic coding layer E4, and an output layer O5.

2.2.1. Structure and function of convolutional layer and pooling layer. In the traditional CNN network, the convolution kernel of the convolutional layer is randomly initialized, and this paper learns through a single-layer noise reduction sparse autoencoder, and obtains the parameters of the network, namely various convolution kernels. At the same time, in the convolution and pooling process, the modified linear unit [14] (ReLU) is used to correct the traditional results, so that the features with stronger characterization ability can be obtained, and the accuracy of recognition is improved. Fig.3. shows the process of generating various convolution kernels. For the input image set through the preprocessing module, the first sampling is performed to obtain 30,000 8×8 image sub-blocks; then, these image sub-blocks are input into the noise reduction sparse automatic encoder, and the L-BFGS optimization algorithm is used for unsupervised. Learn, get the parameters of the network; finally, output these parameters, the result is a variety of convolution kernels we have learned.

In the deep feature extraction process, the convolution kernel obtained by the above learning is used to realize the function of the convolutional layer. In the convolution operation, a convolution kernel K is used to convolve an input image x , plus an offset b_x , and then a modified linear unit (ReLU) is used to obtain the feature map C_x of the convolutional layer, as in equation (4). Show:

$$C_x = \max(0, K * x + b_x) \quad (4)$$

The size of the feature map obtained after convolution is 57×57, and each pixel is connected to the 8×8 neighborhood (local receptive field) in the input image. In FIG. 2, $P2$ is a pooling layer, and the feature map obtained after convolution is non-overlapping pooled through a 3×3 pooling window, that is, a neighborhood composed of nine pixels.

Summing to get a new pixel, then weighting by scalar w_{x+1} , then increasing the offset b_{x+1} , and finally generating a 19×19 size map by the ReLU activation function.

$$S_{x+1} = \max(0, \sum C_x * w_{x+1} + b_{x+1}) \quad (5)$$

2.3. Improved HIE algorithm encryption process

The core idea of the improved hyperchaotic system image encryption algorithm is: firstly, using the hyperchaotic system to scramble the pixels obtained after the above steps, to resist the deciphering method of the general low-dimensional chaotic system; secondly, the ciphertext feedback control algorithm. The key stream in the key enables the parameters required for encryption to be related to the plaintext through ciphertext feedback, spreading the effect of one plaintext byte into more ciphertext bytes. Theoretical analysis and experimental simulations show that the improved algorithm not only can effectively avoid the choice of plaintext attack and ciphertext attack, but also has better statistical difference characteristics. The algorithm mainly includes two aspects: pixel chaos and image diffusion and chaos. The following describes the encryption process of HIE algorithm.

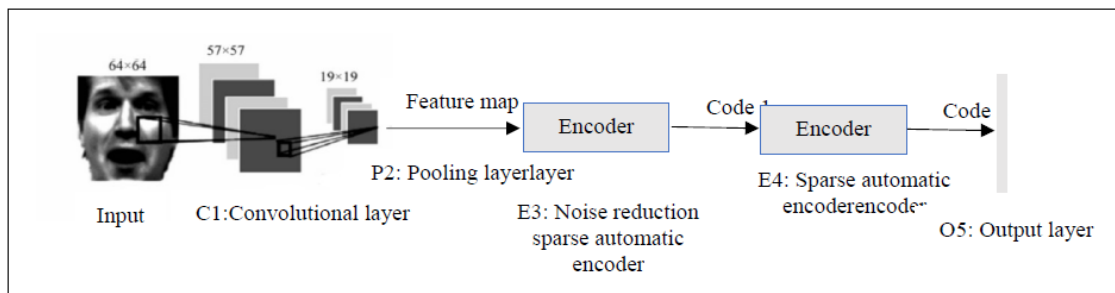


Figure 2. Deep network structure

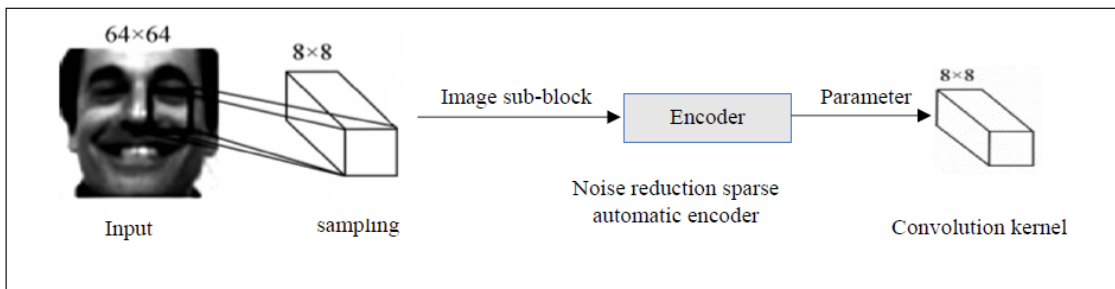


Figure 3. Convolution nuclear production

2.3.1. Pixel Scrambling. Let $P(m \times n)$ denote an image of size $M \times N$, and the pixel matrix is represented as

$$P = \begin{pmatrix} p_1 & p_2 & \cdots & p_N \\ \vdots & \vdots & \ddots & \vdots \\ p_{N(M-1)+1} & \cdots & \cdots & p_{MN} \end{pmatrix} \quad (6)$$

Step 1 Use the Runge-Kutta algorithm to iterate the hyperchaotic system N_0 times to prevent the transition effect. For a given system, the number of iterations N_0 may be related to the initial conditions and system parameters. This paper will discard the data generated by the previous $N_0=200$ iterations, and then calculate.

$$r = \text{mod}((\text{abs}(x_1) - \text{floor}(\text{abs}(x_1))) \times 10^{14}, M) \quad (7)$$

Obviously, $r \in [0, M-1]$. Iterate the hyperchaotic system until M completely different r values are generated, denoted as $\{c_j, j = 0, 1, \dots, M-1\}$. According to $\{r_i, i = 0, 1, \dots, M-1\}$ performs a row transformation on the matrix P , and the result is recorded

$$P^r = \begin{pmatrix} p_1^r & p_2^r & \cdots & p_N^r \\ \vdots & \vdots & \vdots & \vdots \\ p_{N(M-1)+1}^r & \cdots & \cdots & p_{MN}^r \end{pmatrix} \quad (8)$$

Step 2 Similarly, calculate $c = \text{mod}((\text{abs}(x_1) - \text{floor}(\text{abs}(x_1))) \times 10^{14}, N)$ where $c \in [0, N-1]$. Iterate until N is generated A completely different value of c , denoted as $\{c_j, j = 0, 1, \dots, N-1\}$. According to $\{c_j, j = 0, 1, \dots, N-1\}$ for the line of (21) the matrix P^r is used for column permutation, and the transformed

Matrix is recorded as

Matrix is recorded as

$$P^{rc} = \begin{pmatrix} p_1^{rc} & p_2^{rc} & \cdots & p_N^{rc} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N(M-1)+1}^{rc} & \cdots & \cdots & p_{MN}^{rc} \end{pmatrix} \quad (9)$$

P^{rc} is the matrix after the original matrix P is scrambled by the hyperchaotic system, and P^{rc} is encrypted below.

2.3.2. Image diffusion and confusion

Diffusion requires that the impact of a single plaintext or ciphertext be extended to more ciphertexts, making it more difficult for attackers to seek explicit redundancy. Chaos requires obscuring the relationship between ciphertext statistical properties and plaintext statistical properties. However, the random sequence generated by Logistic mapping and hyperchaotic systems in the literature is only related to the initial value and system parameters, and does not depend on plaintext, so that a plaintext byte can only affect one encrypted ciphertext byte. Based on this shortcoming, this paper proposes a new diffusion chaos method to overcome this weakness and obtain higher security.

Step 1 Calculate the random sequence generated by the hyperchaotic system

$$x_i = \text{mod}((\text{abs}(x_i) - \text{floor}(\text{abs}(x_i))) \times 10^{14}, 256) \quad i = 1, 2, 3, 4 \quad (10)$$

Obviously, $x_i \in [0, 255]$. Then calculate

$$\bar{x}_1 = \text{mod}((x_1 + x_2 + x_3 + x_4), 4) \quad (11)$$

Step 2 According to $\bar{x}_1 \in [0.3]$, select the corresponding combination from Table 1 to encrypt the row-column permutation matrix P^{rc} generated in Section 2.4.1, that is,

$$\begin{aligned} C_{3 \times (i-1)+1} &= P^{rc}_{3 \times (i-1)+1} \oplus D_{x_1}, \\ C_{3 \times (i-1)+2} &= P^{rc}_{3 \times (i-1)+1} \oplus D_{x_2}, \\ C_{3 \times (i-1)+3} &= P^{rc}_{3 \times (i-1)+1} \oplus D_{x_3}, \end{aligned} \quad (12)$$

Where D_{x_1} , D_{x_2} and D_{x_3} are as follows:

$$\begin{aligned} D_{x_1} &= \text{mod}((B_{x_1} \oplus C_{3 \times (i-1)+1}, 256), \\ D_{x_2} &= \text{mod}((B_{x_1} \oplus C_{3 \times (i-1)+2}, 256), \\ D_{x_3} &= \text{mod}((B_{x_1} \oplus C_{3 \times (i-1)+3}, 256), \end{aligned} \quad (13)$$

Obviously, $D_x \in [0, 255]$, where $i = 1, 2, \dots$ represents the i -th hyperchaotic iteration; \oplus denotes XOR, $P_i, i = 1, 2, \dots, M \times N$ represents the pixel value of the scrambled image; B_{x_1} , B_{x_2} and B_{x_3}

represent the corresponding combinations in Table 1 selected according to \bar{x}_1 , $C_i, i = 1, 2, \dots, M \times N$, representing ciphertext pixel values.

Step 3 If all the plaintexts are encrypted, the encryption process ends, otherwise go to step 1.

The decryption process is similar to the encryption process. First, use the parameters and initial values to generate the same hyperchaotic sequence, and replace (12) with

$$\begin{aligned} P_{3 \times (i-1)+1}^{rc} &= C_{3 \times (i-1)+1} \oplus D_{x_1}, \\ P_{3 \times (i-1)+2}^{rc} &= C_{3 \times (i-1)+2} \oplus D_{x_2}, \\ P_{3 \times (i-1)+3}^{rc} &= C_{3 \times (i-1)+3} \oplus D_{x_3}, \end{aligned} \quad (14)$$

Then, according to $\{r_i, i = 0, 1, \dots, M - 1\}$ and $\{c_j, j = 0, 1, \dots, N - 1\}$, the matrix is inversely transformed, and the original image can be restored.

Table 1. Different combinations of hyperchaotic sequences

\bar{x}_1	Corresponding combination
0	(x_1, x_2, x_3)
1	(x_1, x_2, x_4)
2	(x_1, x_3, x_4)
3	(x_2, x_3, x_4)

3. Experimental PART

In this paper, the common face recognition database CMU-PIE is used as the experimental object to evaluate the performance of the proposed algorithm. The hardware configuration is: Intel(R) Core(TM) i5-7200UCPU@2.50GHz, 8.00GB memory.

3.1. Experimental data and network parameter settings

This paper first collects 10,000 unmarked faces to complete the pre-training of the network, that is, obtain the parameters of the deep network feature extractor in Figure 1. When experimenting with the common face recognition database, it is divided into two parts: training set and test set. All samples passed the preprocessing module, and unified image sets of 64×64 to large and gray value $[0, 1]$. When training the convolution kernel used in the deep network feature extractor, the input is set to 8×8 , the lambda is set to $1e-4$, and the number of levels is set to level 1, and various types of 8×8 convolution kernels can be obtained. The pooling layer window size is set to 3×3 , and the pooling type is non-overlapping pooling.

3.2. Experimental results

3.2.1. Experimental results on the face database. The CMU-PIE face database consists of 68 people, a total of 41,368, grayscale images, and 11,554 of them were selected to complete the experiment. The face database contains multiple poses, lighting and expression images, and the pose and illumination changes are also acquired under strictly controlled conditions. In the experiment, 7,000 of them were selected as training samples, and the rest were used as test samples.

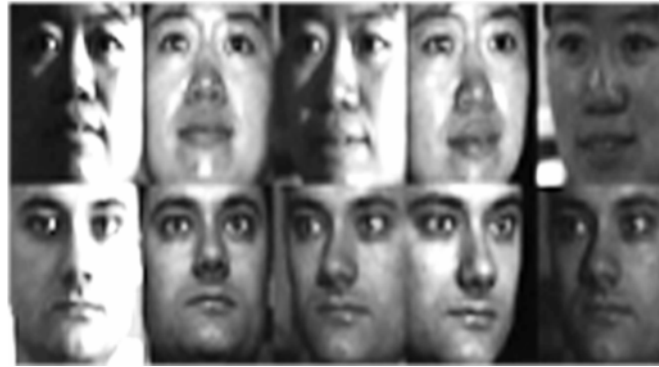


Figure 4. CMU-PIE part of the face database

Figure 4 is a partial face of the database, it can be seen that this database contains rich gestures and illumination changes.

The experimental results are shown in Table 2. It reflects that the algorithm has good feature expression ability for multi-pose, illumination and expression, and the effect is improved compared with other methods. Comprehensive analysis of experimental data shows that compared with hand-crafted features and traditional deep learning algorithms, the improved deep network learning and extraction features have strong adaptability to illumination, expression, attitude, etc., and average accuracy. High, the system has strong generalization ability and robustness.

3.2.2. Experimental results. The simulation results of the experiment are shown in the figure below. Figure (a) is the original image of 256kbit, and Figure (b) is the image after scrambling. It can be seen from the figure that the density distribution of the pixels after scrambling has changed greatly, and the picture is similar to white noise, showing a good and scrambling effect. When the image is scrambled, if the number of fixed points is smaller, the better the scrambling effect is, and the higher the confidentiality. The final encryption result is shown in (c), which shows that the original image has been completely hidden, and the outline of the original image is no longer visible.

Table 2. Correct recognition rate for different algorithms on the cmu-pie library

Method	Recognition rate
PCA	80.14%
LGBPHS	80.06%
SVM	81.51%
LBP	90.54%
Block LBP	92.93%
SAE	94.20%
Algorithm	96.17%

It can be seen from Fig. d and Fig. e that the encrypted histogram is flat and the gray value is evenly distributed compared with the original histogram with uneven distribution. This indicates that the ciphertext's pixel values are equal in the range of [0-255], that is, the uniform distribution of the ciphertext space, which indicates that the improved algorithm proposed in this paper can effectively prevent attacks.

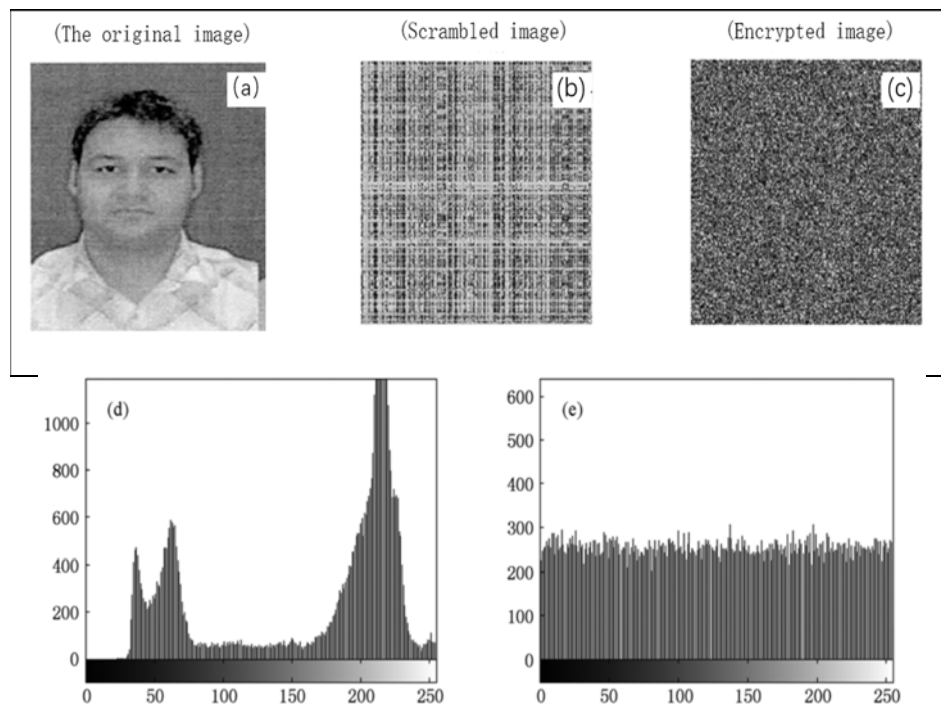


Figure 5. Original image and encrypted image simulation (a) original image; (b) scrambled image; (c) encrypted image (d) a histogram of the grayscale distribution of the original image; (e) an encrypted map

4. Summary

In cryptography, the key to success is the process of encryption and decryption. The security of the information depends on the security of the key. Traditional image encryption algorithms are not immune to malicious key sharing and rejection attacks. If the key is too long, it is easy to lose and hard to remember. Biometric encryption technology came into being, trying to solve the problem of poor key security. The key is generated based on the biometric characteristics of the individual and then applied to the corresponding image encryption algorithm to achieve information encryption. Encrypted biological features should be unique, stable, and non-aggressive. The face not only meets the above

Requirements, but also has rich feature information, strong anti-attack capability and excellent encryption potential. Face image encryption has become an important branch of image encryption and plays an important role in image encryption.

This paper proposes an improved automatic learning and extraction feature based on deep network model, and improves the extracted features to improve the HIE algorithm encryption, thus effectively defending against attacks. This paper will use the large sample set for unsupervised learning, pre-training to obtain the parameters of the deep network feature extractor, and combine with the training set using the experimental data set to fine-tune and extract the discriminative features. In this paper, the improved deep network is used as a feature extractor to capture deep features with strong expressive ability, which is helpful for accurate classification. The combination and improvement of preprocessing, convolution, pooling and multi-layer sparse autoencoders enable the system to extract facial features well. Based on this, in the improved HIE algorithm, each plaintext byte can still be maintained. One encryption operation, without repeated iterations to increase complexity, so the encryption and decryption speed is fast, suitable for real-time communication. At the same time, the experimental results show that the algorithm has good statistical characteristics and key sensitivity and other cryptographic features, which makes the algorithm have strong generalization ability and excellent performance. Its test results on the public data set also prove the effectiveness of the proposed method.

References

- [1] Yang Fei, Su Jianbo. Fusion construction method and recognition of facial dominant features [J]. Journal of Electronics, 2012, 40 (3): 466-471. Yang Fei, Su Jian-bo. Face recognition based on explicit facial features by fusion construction method [J]. Acta Electronica Sinica, 2012, 40 (3): 466-471. (in Chinese)
- [2] Deng W, Hu J, Lu J, et al. Transform-Invariant PCA: A unified approach to fully automatic face alignment, representation and recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2014, 36 (6): 1275-1284.
- [3] Braverman M. Poly-logarithmic independence fools bounded-depth boolean circuits [J]. Communications of the ACM, 2011, 54 (4): 108-115.
- [4] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2 (1): 1-12.
- [5] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection [A]. Conference on Computer Vision and Pattern Recognition [C]. Portland, OR, USA: IEEE, 2013. 3476-3483.
- [6] Yao H X, Li M 2009 Inter.J. Nonlin. Sci. 7 379
- [7] Gao T G, Chen Z Q 2008 Phys. Lett. A 372 394
- [8] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification [A]. Conference on Computer Vision and Pattern Recognition [C]. Columbus, OH, USA: IEEE, 2014. 1701-1708.
- [9] Deng J, Zhang Z, Marchi E, et al. Sparse autoencoder based feature transfer learning for speech emotion recognition [A]. Humaine Association Conference on Affective Computing and Intelligent Interaction [C]. Geneva, Switzerland: IEEE, 2013. 511-516.
- [10] Zhang C, Zhang Z. Improving multiview face detection with multi-task deep convolutional neural networks [A]. IEEE Winter Conference on Applications of Computer Vision [C]. Steamboat Springs, CO, USA: IEEE, 2014. 1036-1041.
- [11] Rao Y, Ni J. A deep learning approach to detection of splicing and copy-move forgeries in images [C] // IEEE International Workshop on Information Forensics and Security. IEEE, 2017: 1-6.
- [12] Ye J, Ni J, Yi Y. Deep Learning Hierarchical Representations for Image Steganalysis [J]. IEEE Transactions on Information Forensics & Security, 2017, 12 (11): 2545-2557.
- [13] Aminanto M E, Kim K. Detecting Impersonation Attack in WiFi Networks Using Deep Learning Approach [C] // International Workshop on Information Security Applications. Springer, Cham, 2016: 136-147.
- [14] Le T P, Aono Y, Hayashi T, et al. Privacy-Preserving Deep Learning: Revisited and Enhanced [C] // International Conference on Applications and Techniques in Information Security. Springer, Singapore, 2017: 100-110.
- [15] Chen Z, Information D O. Face Deep Learning Technology in the Design and Implementation of the Security in Colleges and Universities [J]. Journal of Anyang Institute of Technology, 2017.