Hindawi
Journal of Applied Mathematics
Volume 2023, Article ID 4448861, 8 pages
https://doi.org/10.1155/2023/4448861

Research Article

Lip Print Recognition Algorithm Based on Convolutional Network

Hongcheng Zhou

School of Electronic and Information Engineering, Jinling Institute of Technology, Nanjing 211169, China

Correspondence should be addressed to Hongcheng Zhou; zhouhcnj@163.com

Received 1 November 2022; Revised 9 January 2023; Accepted 10 January 2023; Published 10 February 2023

Academic Editor: Wei-Chiang Hong

Identity information security is faced with various challenges, and the traditional identification technology cannot meet the needs of public security. Therefore, it is necessary to further explore and study new identification technologies. In order to solve the complex image preprocessing problems, difficult feature extraction by artificial design algorithm, and low accuracy of lip print recognition, a method based on the convolutional neural network is proposed, by building a convolutional neural network called LPRNet (Lip Print Recognition Network). The obtained lip print image is inputted into the training recognition model of the network to simplify the lip print image preprocessing. By extracting feature information and sampling operation, the model training parameters are reduced, which overcomes the difficulty of designing a complex algorithm to extract features. By analyzing and comparing the experimental results, a higher recognition rate is obtained, and the validity of the method is verified.

1. Introduction

Lip print recognition originated from the field of criminal investigation and forensic practice, which is used as a tool to identify the suspect or the victim in a criminal investigation and provides help for the criminal investigation. Lip print can be used not only as an identification tool and court evidence but also as a source of criminal investigation and criminal information [1]. The features of lip print are rich in information, including linear, curvilinear, bifurcated, reticular, and irregular texture features, and lip print concealment is good; not easy to be copied and imitated, with uniqueness, permanence, and stability of the characteristics; an important biological feature of human identity. As a new biometric technology, lip print recognition has many advantages compared with other biometric technologies, such as high recognition rate, short recognition time, high user acceptance, and noncontact acquisition [2].

In the background of the continuous innovation and development of science and technology, lip print recognition technology has been rapidly developed, and domestic and foreign scholars have proposed a lip print recognition algorithm. In 2015, Wrobel et al. proposed a recognition method based on lip print cross-analysis, which achieved 77% recognition accuracy on 120 lip print data sets [3]. In 2021, Sandhya et al. implemented a lip print recognition method based on a machine learning algorithm and experimented with support vector machine (SVM), *k*-nearest neighbor (KNN), integrated classifier, neural network, and so on; the integrated classifier achieved 97% recognition rate on 150 lip print data sets [4]. Although these recognition algorithms have achieved good recognition results, the lip print image preprocessing in the algorithm is complex, and the contact acquisition method of the lip print image is not well accepted by users. The artificial extraction algorithm is designed to extract the feature information of the lip print image, which leads to a long recognition period and no real time, and the recognition rate needs to be improved [5]. With the development of science and technology, deep learning has been the focus of many researchers, and the convolutional neural network has been developed rapidly and applied successfully in the fields of computer vision and natural language processing. Therefore, in order to solve the problems existing in the traditional recognition algorithms, this paper proposes a convolutional neural network lip pattern recognition method, builds the lip pattern data

set, designs the convolutional neural network structure, and injects the lip pattern data set for the network model training experiment, finally achieving the expected recognition rate.

In recent years, convolutional neural network model-based image recognition methods have achieved great success and rapid development in various biometric fields, such as face recognition [6], palmprint recognition [7], fingerprint recognition [8], and gait recognition [9]. Therefore, the combination of convolutional neural network and lip print recognition will be a hot research topic in the future.

2. Convolutional Neural Networks

A convolutional neural network (CNN) is a class of feedforward neural network with deep structure and convolution operation [10]. The network structure includes the convolution layer, pooled layer, and fully connected layer. As one of the deep learning representative algorithms, the convolutional neural network has the representation learning ability, which can classify input images by translation invariance according to their hierarchical structure. It is successfully applied to image feature extraction for the image classification task, and there is no need to design a complicated algorithm to extract features. The image feature is extracted by convolution operation, the data dimension is reduced by sampling in the pool layer, the training parameters are reduced, and the full-connection layer is used to solve the classification problem.

2.1. Convolution Layer. The structure of the common neural network includes the input layer, hidden layer, and output layer. The hidden layer is fully connected, which leads to too many parameters and a poor training effect. The difference between a convolutional neural network and a normal neural network is that in the convolutional neural network's convolutional layer, a neuron only connects to a subset of its neighbors. The convolution layer has the characteristics of local connection and shared weights, which greatly reduces the network training parameters. In the training process of the network, the convolution kernel will convolve with different regions of the input image to obtain reasonable weights and extract different image features. Shared weights reduce the connectivity between layers of the network and reduce the risk of overfitting the network model.

The core of the convolution layer is to use different-sized convolution kernels to convolute the image and extract the feature information of the image. The number of convolution kernels is more abundant, which means more features are used for classification and recognition. According to the step length sliding window, the image features are extracted by convolution with the pixels in the image. The commonly used convolution kernel sizes include 3×3, 5×5, and 7×7. In this experiment, the RGB three-channel color image is trained by the input network, and the multichannel convolution is shown in Figure 1. The input image size is 6×6, and the number of channels is 3. The convolution core is 3×3, and the number of channels is 3 (consistent with the number of input image channels). The convolution is a 4×4 feature map which is generated by multiplying the pixel values of the corresponding positions of three channels from left to right and from top to bottom in the form of sliding windows.

2.2. Pool Layer. The image features are extracted by the convolution layer, and the next step is to use these features' information for classification and recognition. In theory, the extracted features can be directly used to train the classifier, but it will face a lot of parameter calculation challenges, and the network model is prone to overfitting. In order to solve this problem, a pool layer is used in the next layer of the convolution layer to process the output results of the convolution layer, and the feature information of different positions is aggregated; that is, the maximum value (or average value) of a particular feature in a region of the image is calculated, also known as a downsampling operation. Unlike the convolution layer, the pool layer does not participate in the weight update, compressing the width and height of the image, but it does not change the number of channels. It not only reduces the data dimension but also expands the range of the convolution kernel and avoids overfitting effectively.

The common pool operations include the average pool, maximum pool, and overlap pool. Because of the good performance of the maximum pooling operation, it is often used in the depth convolutional neural network. The calculation process is to select the maximum pixel value in an image region as the pooled value of that region; the size of the image area is determined by the size of the lower sampling window, which is usually 2×2 and 3×3.

2.3. Fully Connected Layer. By using several convolution layers and pooling layers alternately, the training parameters of the convolutional neural network are greatly reduced, which not only reduces the computation of the parameters but also shortens the training time and improves the robustness of the extracted features [11]. Each neuron in the fully connected layer connects to all the neurons in the upper layer. After extracting the features from the convolution layer and the pool layer, the output image features are integrated and mapped into a fixed-length one-dimensional feature vector, which contains all the feature combination information of the input image; then, the vector is outputted to the classifier for image classification. From the point of view of image classification, the computer only needs to judge the feature information, calculate the probability of the number of the categories to which the input image belongs, and output the most possible categories to complete the task of classification. Therefore, the function of the fully connected layer is to connect all features and output one-dimensional vectors to the classifier for classification recognition.

3. Lip Print Recognition Method Based on LPRNet

3.1. Batch Normalization and Activation Function. Batch normalization is a method of optimizing network training, batch refers to the number of pictures set during model

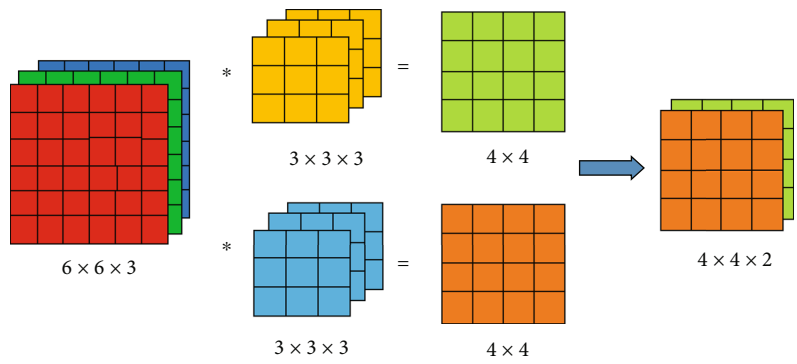

Figure 1: Multichannel convolution.

training, and data normalization is the normalization of the input data [12]. During the convolutional neural network training, we added a BN layer to make the network input training samples and test samples distributed in the same way, aiming at the problem that each iteration has to adapt different data distributions which leads to the slow training speed; training different data distributions can reduce the generalization ability of the network. It is added to the convolution layer activation function of the neural network and is fused with the convolution layer to process the output data after the convolution operation. The mathematical expression for the calculation is as follows:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i, \tag{1}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu)^2, \tag{2}$$

$$\widehat{x}_l = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \tag{3}$$

$$y_i = \gamma \widehat{x}_l + \beta. \tag{4}$$

Firstly, the mean μ and variance σ^2 of the input data are calculated, and then, it is standardized by formula (3). Finally, the data are translated and scaled by introducing the scaling factor γ and the offset value β, and the result is used as the input data of the activation function. In order to improve the nonlinear expression of the model, improve the model robustness, and reduce the gradient loss, the activation function can be used to map the convolution output. The activation functions should be nonlinear, continuous differentiable, monotonic, and nearly linear at the origin of coordinates. The common activation functions include Sigmoid, Tanh, ReLU, Leaky ReLU, ELU, and Maxout.

In the experiment, the ReLU function is chosen as the activation function. Because the Sigmoid function can make the gradient easily disappear and the output value is not zero, the network cannot update the weight parameter. The mathematical expression is as follows.

$$f(x) = \max(0, x). \tag{5}$$

When $x > 0$, the gradient constant is 1, which effectively solves the gradient vanishing problem and converges quickly, and increases the network sparsity. When $x < 0$, the output of the layer is 0, after the model training. The extracted features are representative, the generalization ability of the model is strong, and the computation is small. The disadvantage is that when the learning rate is too high, a large number of neurons will be inactivated during training. That is, the neuron may not be activated, resulting in the corresponding neuron weight not being updated. Therefore, it is necessary to choose the appropriate learning rate when training the network model.

3.2. Softmax Classifier. This experiment involves a multiclassification problem, so we choose Softmax as the classifier and place it in the next layer of the convolutional neural network as part of the network structure [13]. The working principle of the Softmax layer is to calculate the probability value, which belongs to class j, and then normalize it to ensure that the sum of the probability value is 1; the output is the maximum probability that x belongs to a certain class. The Softmax function outputs the maximum probability value of the category which the input image belongs to, as shown in the following formula.

$$f_j(x) = \frac{e^{xj}}{\sum_j^k e^{xj}}. \tag{6}$$

In the formula, x_j represents the classifier output of the upper layer output unit, j represents the number of categories, the total number of categories is k, and $f_j(x)$ represents the ratio of the current element index and the index sum of all elements. By using this formula, the multiclass output value can be converted into relative probability, the probability value of one-dimensional vector x belonging to class j can be calculated, and the probability value between [0,1] can be normalized in the exponential domain; the maximum probability output is the category which it belongs to. Therefore, the function is often used to achieve multiple image classification tasks.

3.3. Lip Print Recognition Model Based on LPRNet. The traditional lip print recognition flow is shown in Figure 2, which includes preprocessing, feature extraction, feature matching, and classification. All kinds of image-processing

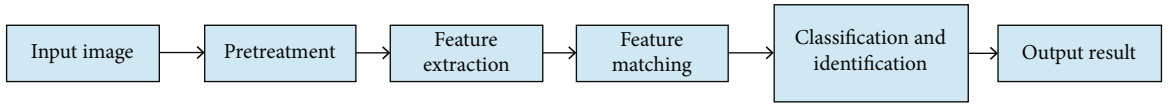

Figure 2: The traditional lip print recognition flow.

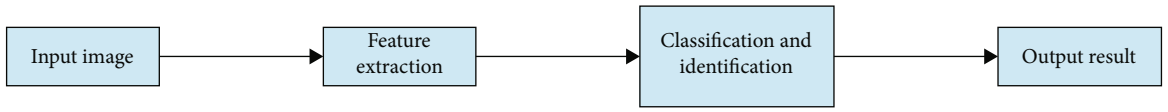

Figure 3: Lip print recognition flow based on LPRNet.

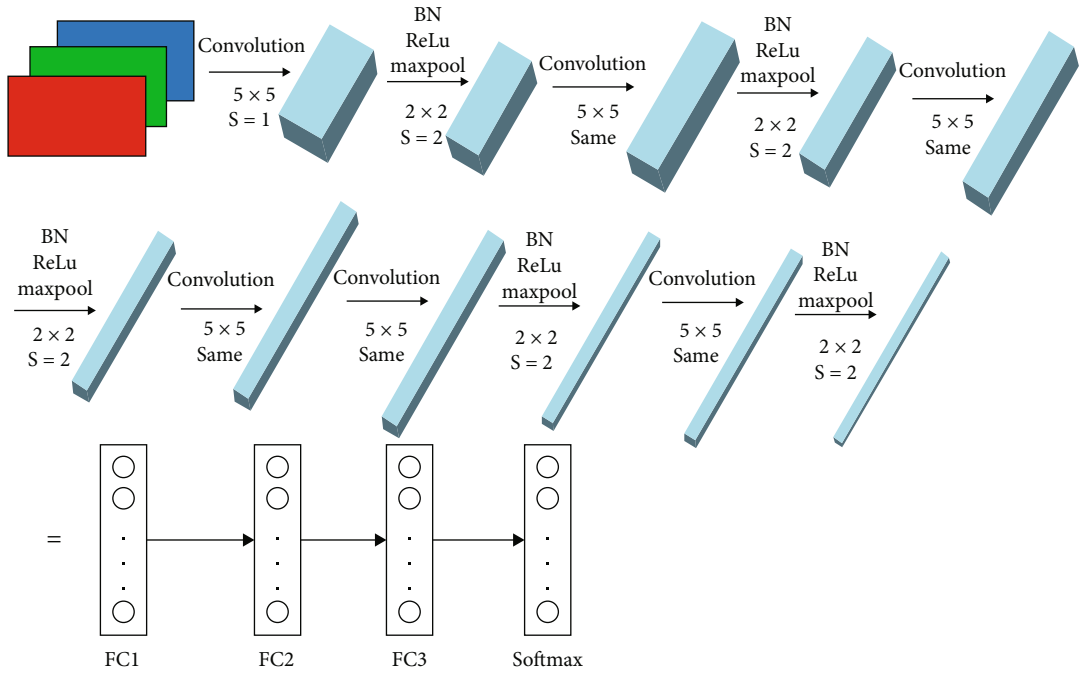

Figure 4: LPRNet network architecture.

methods are used to highlight the feature information in the image, and the quality of the lip ridge image is required to be high. Feature extraction requires manual design of an extraction algorithm, so the whole recognition process is long, the preprocessing is complex, and the feature extraction is difficult. Therefore, the recognition accuracy is low. To solve these problems, this paper presents a lip print recognition method based on the convolutional neural network, and its recognition flow is shown in Figure 3. Compared with the traditional recognition process, the lip print image preprocessing is simplified, the complex feature extraction algorithm is avoided, and the original image can be inputted directly. A feature extractor composed of several convolution layers and pooled layers is used to extract image features automatically, and the Softmax classifier is used to process feature information and output classification results.

In 2012, AlexNet won the first prize in the ImageNet competition, which attracted a lot of attention for its deep learning [14]. In this paper, we build a convolutional neural network called LPRNet (Lip Print Recognition Network). It consists of 6 convolution layers, 5 pooled layers, and 3 fully connected layers and uses the Softmax classifier for multiclassification tasks to output the classification results. A batch normalization layer is added between the convolution layer and the pool layer, which distributes the training and test data in the same way, and improves the generalization ability and stability of the model. The activation function uses the ReLU function to solve the problems of gradient vanishing and gradient explosion, to speed up the convergence of the network, and to extract the features and classify the lip print data set; the network structure is shown in Figure 4.

The detailed parameters of each layer of the network structure are shown in Table 1. The feature extractor is composed of the convolution layer and pool layer alternately. The size of the sampling window is set to 2×2, and the number of convolution cores and output characteristic graphs is the same.

The lip print recognition method based on the LPRNet is described as follows: the size of the input image is fixed at 244×244, and the size of the input image is $244 \times 244 \times 3$ because the acquired images are all color images with three

Table 1: Information of all network layers.

Network layer	Name	Dimensions	Channel number	Step length	Activation function
C1	Convolution layer	5 × 5	8	1	ReLU
M1	Pool layer	2 × 2	8	2	——
C2	Convolution layer	5 × 5	32	1	ReLU
M2	Pool layer	2 × 2	32	2	——
C3	Convolution layer	5 × 5	64	1	ReLU
M3	Pool layer	2 × 2	64	2	——
C4	Convolution layer	5 × 5	128	1	ReLU
C5	Convolution layer	5 × 5	128	1	ReLU
M4	Pool layer	2 × 2	128	2	——
C6	Convolution layer	5 × 5	64	1	ReLU
M5	Pool layer	2 × 2	64	2	——
FC1	Fully connected layer		3136 nodes		ReLU
FC2	Fully connected layer		2048 nodes		ReLU
FC3	Fully connected layer		2048 nodes		ReLU
Softmax	Output layer		40 nodes		——

channels of RGB. Firstly, the convolution layer C1 uses 8 convolution kernels of 5×5, with a step length of 1 sliding convolution kernel window; the convolution operation is carried out on the input image; and the feature graph of $240 \times 240 \times 8$ is obtained. Then, the M1 pool layer samples the image feature information by using the maximum pool operation to reduce the network training parameters. The size of the sample window is 2×2, with the step length of 2. It slides the window from left to right and from top to bottom; the result is a feature map of $120 \times 120 \times 8$, in which 32 convolution cores of 5×5 size are used and the edge filling mode is set the same way; after the convolution operation, the $120 \times 120 \times 32$ feature map is outputted, and the sampling operation under the pooled layer M2 is the same as that under the pooled layer M1. The results of the convolution layer are mapped to the $60 \times 60 \times 32$ feature map. The C3 layer uses 64 convolution cores of 5×5 size and performs the convolution operation according to step size 1 and the edge filling mode is the same and outputs a feature map of $60 \times 60 \times 64$; the feature extracted from the convolution layer is mapped to a feature map of $30 \times 30 \times 64$.

The full connection layer FC1 tiles the two-dimensional feature matrix of the upper layer output into a one-dimensional feature vector of $7 \times 7 \times 64$. There are 3136 neurons in the entire connected layer. In order to avoid overfitting, the dropout method is used to randomly discard and deactivate some neurons. The dropout value is set to 0.2, so the number of nodes in the FC2 layer is 2048. The output from the FC2 layer is processed with a dropout value of 0.2, so that the number of nodes in the FC3 layer is 2048. Finally, the Softmax classifier is used as the output layer, and the 2048 nodes are mapped to the probability value corresponding to the 40 categories.

This method allows the direct input of the lip data set, simplifies the preprocessing of the lip print image, and automatically extracts the lip print feature information through the network training. It avoids spending much time on the design of the feature extraction algorithm and shortens the whole lip print recognition period, getting a higher recognition accuracy.

4. Experimental Procedure and Results

4.1. Experimental Environment and Data Set. Since there is no professional equipment for collecting the lip print, the traditional contact method is to apply a special material on the lips, then press the lips on the white paper, and use a scanner to convert the image appearing on the white paper into a digital image [15]. Because of the limitation application and the low acceptance, the quality of the acquired image depends on the pressure and direction; it is easy to be influenced by human factors. The lip print data set used in this experiment adopts the noncontact acquisition method with high user acceptance, and each image is taken under natural light conditions.

The data set is from 40 volunteers; 30 lip print images were collected from each volunteer and saved to a folder. In order to avoid the overfitting phenomenon during network model training, the data set is expanded by simple data enhancement methods, such as rotation, adding noise, mirroring, blurring, and changing image brightness; the 1200 lip print images were expanded into an 8000-image data set, so that each volunteer had 200 images, and the images were randomly divided into a training set and a test set in an 8 : 2 ratio. The training set will be used to train the lip print recognition model, and the test set will be used to detect the accuracy of the model. A partial image of the data set is shown in Figure 5.

Lip lines are the entire texture of the red part of the lips in the image, including lines, curves, bifurcations, lip grooves, and other texture features. These images are taken by mobile phones, so the original image is different in size

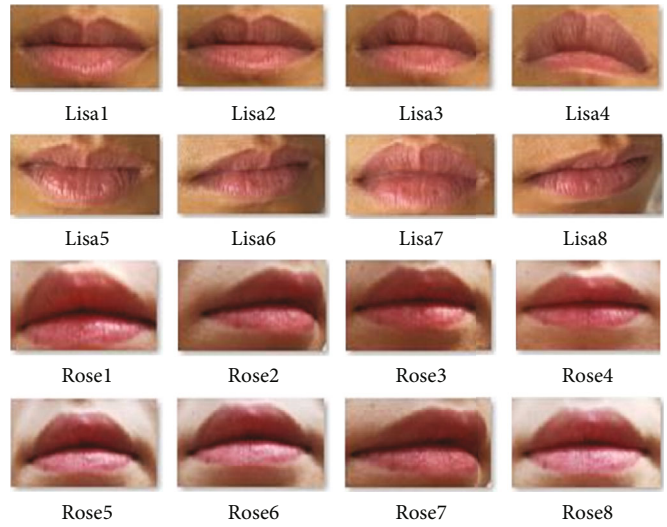

Figure 5: Volunteers' lip print images.

and resolution. However, the size of the lip print image affects the selection of the size of the convolutional neural network convolution nucleus, as well as the definition and extraction of lip print features. Therefore, the size of each image is very important; an image can cause the neural network to compute a large number of parameters and increase the training time of the data. If the image is very small, not only is the lip print not clear enough but also the key feature information will be lost if the resolution is too low. Before entering the data set into the convolutional neural network training, the uniform lip print image was reduced to a size of 244×244 (width $\times$ height).

4.2. Training Experiment of Network Model. Firstly, we set up the parameters of the network. Then, we fed the lip print data set into the convolutional neural network training. We randomly divided a data set of 8000 images into a training set and a test set on an 8 : 2 ratio; the input image size is set to 244×244. The network optimizer chooses the Adam optimizer and sets different learning rates to train and test the network model. Too small and too large learning rates will lead to a poor recognition effect, so it is necessary to adjust the training model of the network parameters several times. The cross-information entropy is used to calculate the loss value of the training set, and the visual function is used to draw the model recognition accuracy distribution graph, to observe and analyze the experimental results in order to achieve the expected recognition accuracy, and to obtain an optimal lip print recognition model.

By discarding the noise connection between the convolved source and core, the influence of noise on CNN is reduced. Based on the pixel value of CNN to improve the classification accuracy, the noise connection is discarded. Discarding the noise connection prevents the input noise pixels from going to the next layer. This method can be used for different convolutional kernel sizes.

When the gradient descent algorithm is used to optimize the network training, the learning rate is an important parameter. A parameter setting that is too small will not only lead to a long training time but may also fall into the local optimum. Because the batch normalization (BN) layer is added in the design of the neural network, which allows feature extraction and classification with a higher learning rate, other parameters will be fixed in the experiment, to set a different learning rate for training, to analyze the learning rate on recognition performance, and finally to get the best recognition model. Firstly, the epoch is set to 50 and the batch is set to 80, meaning that 100 images will be entered for each step for training and other parameters will be maintained. Then, a different learning rate is chosen to train the network model, and the results are shown in Table 2.

Table 2: The recognition results of different learning rates.

Learning rate	Training time (s)	Identification time (s)	Recognition rate
0.01	546 s	7 s	92.00%
0.005	503 s	6 s	95.31%
0.003	651 s	8 s	98.00%
0.001	639 s	8 s	97.00%
0.0005	531 s	7 s	97.25%
0.0003	780 s	10s	99.06%
0.0001	589 s	7 s	98.75%

Table 2 shows that the recognition rate is the best when the learning rate is 0.0003, and the recognition accuracy is 99.06% in the test set data. Compared with other learning rate recognition results, the recognition rate is given priority although the training and recognition time is longer. When a higher or lower learning rate is used, the network model is unstable, the time of training and recognition is shortened, the speed of training is improved, but the recognition rate is decreased.

The model of the learning rate of 0.0001 was trained to predict the recognition results for the data sets with different proportions as shown in Figure 6. The horizontal coordinate represents the times of the test set input model prediction,

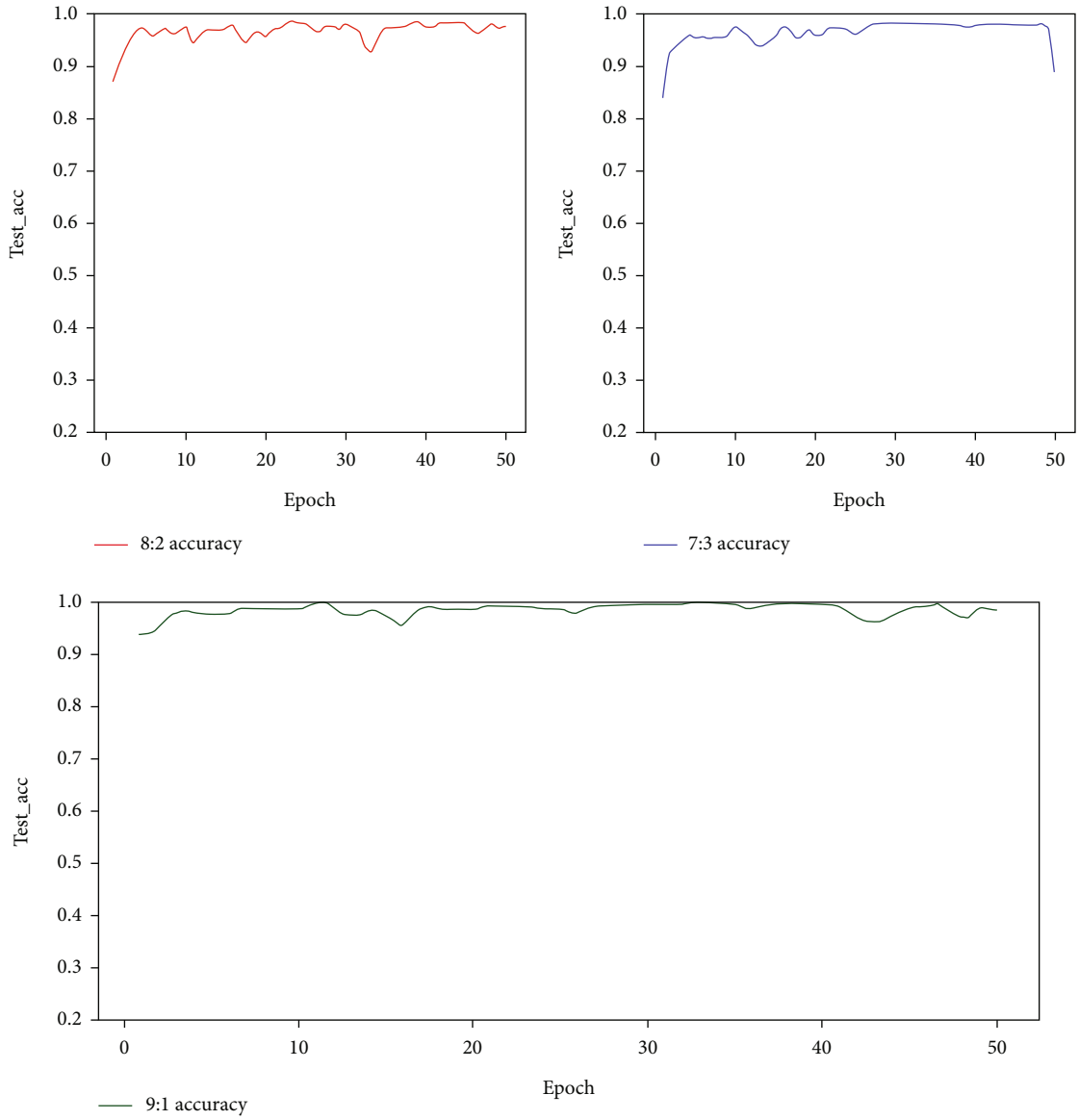

Figure 6: The recognition rate of test set.

and the vertical coordinate represents the recognition rate of the test set. By comparing the recognition results of three kinds of scale partition data sets in experiment 2, we can see that the data set with 9 : 1 scale partition has the best recognition effect, and the recognition rate of the other two methods has decreased.

Classical depth convolutional neural networks such as AlexNet, VGG, GoogLeNet, and ResNet were used to compare the performance of different networks for lip print data set recognition. The recognition model was trained using a 9 : 1 ratio of the data set, with the learning rate of 0.0001, the number of training of 50, the batch size of 80, and the remaining parameters kept unchanged. The input data set is trained for the network model, the model is loaded, and the test set is inputted for recognition. The recognition rate is the average of the last 10 predictions. The results of model recognition in this experiment are shown in Table 3.

Table 3: Prediction and identification results of different networks.

Network structure	Training time (s)	Identification time (s)	Recognition rate
AlexNet	1147 s	15 s	96.75%
VGG	1046 s	13 s	96.87%
GoogLeNet	995 s	12 s	97.05%
ResNet	983 s	12 s	96.98%
LNet-6	589 s	7 s	97.97%

4.3. Analysis of Experimental Results. From Tables 1 and 2 of the experimental results, it can be seen that different learning rates have a greater impact on model recognition. When the data set is partitioned 9 : 1 and the learning rate is 0.0001, the recognition effect is the best, and the average recognition rate is 97.97% in the test set. From Table 3, the LNet-6-based network has a shorter training and recognition time,

a smaller model file, and a higher average recognition rate compared with the convolutional neural network model recognition results. By comparing the recognition effects of different network models, the depth convolutional neural network has a relatively low recognition rate for the lip print data set and consumes a lot of computing resources, making it difficult to apply the model to actual terminal devices. Through the analysis of the experimental results, the recognition model based on LNet-6 has a good effect on lip print image recognition, which not only simplifies the preprocessing of the lip print image but also avoids the design of a complex feature extraction algorithm and combines the feature extraction and classification process and automatic extraction of lip features and classification.

5. Conclusion

In this paper, a new method of lip print recognition based on the convolutional neural network is proposed, which has the advantage of directly inputting the original lip print image and simplifying the image preprocessing. Combined with the advantages of a convolutional layer local connection and shared weight value, the feature extractor, which is composed of the alternate connection of the convolutional layer and pool layer, can extract lip features automatically and reduce the training parameters of the network greatly; it overcomes the problem of feature extraction by the artificial design algorithm in the traditional recognition method. The BN layer and ReLU activation function are used to speed up the convergence of the network model, effectively solve the problem of gradient disappearance and gradient explosion, and avoid the overfitting phenomenon of the model. The experiment sets up different learning rates to train the network, analyze and compare the effect of the learning rate on recognition performance, and finally achieve a recognition rate of 99.06% on the test set.

Further research will be carried out in the following areas: (1) collection of lip print images to build a larger data set, (2) further optimization of the network structure to reduce the training time, and (3) combined with the advantages of transfer learning, the pretraining model will be used to classify and identify the data set, and the time of training and recognition will be is shortened.

Data Availability

The data relating to the paper has been examined.

Conflicts of Interest

The author declares no conflict of interest in this article.

Acknowledgments

This work was supported by the Cooperative Project of Jiangsu Province Production, Teaching, and Research (No. BY2021381) and the Jin Ling Institute of Technology Ph.D. Startup Fund.

References

[1] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.

[2] Y. Yu, S. W. Liang, B. J. Samali et al., "Torsional capacity evaluation of RC beams using an improved bird swarm algorithm optimised 2D convolutional neural network," *Engineering Structures*, vol. 273, article 115066, 2022.

[3] K. Wrobel, R. Doroz, P. Porwik, and M. Bernas, "Personal identification utilizing lip print furrow based patterns. A new approach," *Pattern Recognition the Journal of the Pattern Recognition Society*, vol. 81, pp. 585–600, 2018.

[4] S. Sandhya, R. Fernandes, S. Sapna, and A. P. Rodrigues, "Comparative analysis of machine learning algorithms for lip print based person identification," *Evolutionary Intelligence*, vol. 15, no. 1, pp. 743–757, 2022.

[5] Y. Yu, M. Rashidi, B. Samali, M. Mohammadi, T. N. Nguyen, and X. Zhou, "Crack detection of concrete structures using deep convolutional neural networks optimized by enhanced chicken swarm algorithm," *Structural Health Monitoring*, vol. 21, no. 5, pp. 2244–2263, 2022.

[6] Y. Yu, B. J. Samali, M. Rashidi, M. Mohammadi, T. N. Nguyen, and G. Zhang, "Vision-based concrete crack detection using a hybrid framework considering noise effect," *Journal of Building Engineering*, vol. 61, article 105246, 2022.

[7] Y. R. Pandeya and J. W. Lee, "Deep learning-based late fusion of multimodal information for emotion classification of music video," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2887–2905, 2021.

[8] L. Wang, C. Z. Wu, L. B. Tang et al., "Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method," *Acta Geotechnica*, vol. 15, no. 11, pp. 3135–3150, 2020.

[9] M. Rabiei and A. J. Choobbasti, "Innovative piled raft foundations design using artificial neural network," *Frontiers of Structural and Civil Engineering*, vol. 14, no. 1, pp. 138–146, 2020.

[10] J. W. Liu, X. Yang, S. Lau et al., "Automated pavement crack detection and segmentation based on two-step convolutional neural network," *Computer-Aided Civil Infrastructure Engineering*, vol. 35, no. 11, pp. 1291–1305, 2020.

[11] S. Güney and M. Erkuş, "A real-time approach to recognition of Turkish sign language by using convolutional neural networks," *Neural Computing and Applications*, vol. 34, pp. 4069–4079, 2022.

[12] F. H. Huang, Y. Yu, and T. H. Feng, "Automatic building change image quality assessment in high resolution remote sensing based on deep learning," *Journal of Visual Communication and Image Representation*, vol. 63, article 102585, 2019.

[13] Z. Y. Lv, T. F. Liu, C. Shi, J. A. Benediktsson, and H. du, "Novel land cover change detection method based on *k*-means clustering and adaptive majority voting using bitemporal remote sensing images," *IEEE Access*, vol. 7, pp. 34425–34437, 2019.

[14] C. X. Zhang, P. Yue, D. Tapete et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.

[15] H. W. Jiang, X. Y. Hu, K. Li, J. Zhang, J. Gong, and M. Zhang, "PGA-SiamNet: pyramid feature-based attention-guided Siamese network for remote sensing orthoimagery building change detection," *Remote Sensing*, vol. 12, no. 3, p. 484, 2020.