Image Processing, Electronics and Computers
L. Trajkovic et al. (Eds.)

doi:10.3233/ATDE240470

Research on the Application of CNN Face Recognition Technology in the Airport

Libin XIAO[a], Xiaoyu LI[a,1], Yang LUO[a], Lei NI[a] and Jiaqi WANG[b]
[a] *Chengdu Technological University, Chengdu Sichuan, China, 611730*
[b] *Advanced Micro/Nano Device Laboratory, Waterloo, Canada*

Abstract. In today's interconnected world, ensuring the safety of airports assumes paramount importance, given their pivotal role in facilitating global transportation. The rapid evolution of face recognition technology emerges as a pivotal tool in fortifying airport security measures. This study meticulously examines three prominent convolutional neural network (CNN) models: VGGNET, GoogLeNet, and ResNet, evaluating their effectiveness in recognizing faces across diverse and uncontrolled scenarios, leveraging the extensive LFW dataset. The findings unequivocally demonstrate ResNet's superiority in unconstrained face recognition scenarios compared to its counterparts. Following rigorous training on the LFW dataset, the ResNet-50 model achieves a remarkable accuracy rate of 71.9%. Consequently, the study infers that the ResNet-50 model exhibits exceptional suitability for deployment within airport environments, offering heightened security protocols. This research underscores the pivotal role of CNN-based face recognition technology in enhancing the robustness of airport security measures, thereby contributing significantly to the safety and efficiency of global transportation hubs.

Keywords. Face recognition technology, CNN-based models, VGG-Net, GoogLeNet, ResNet

1. Introduction

In the context of globalization, airports serve as pivotal nodes for international travel and trade, facilitating crucial tasks such as the movement of people and goods. Facial recognition technology has emerged as a key avenue for enhancing the efficiency and level of airport security management. The rise of deep learning technology, particularly convolutional neural networks (CNN), has significantly advanced the development and application of facial recognition technology.

The application of facial recognition in airport settings has been discussed in previous research. Chen's[1] team mentioned the practical use of facial recognition at airports in their paper "An Analysis of the Application of Facial Recognition Technology in Airport Settings" published in the "Communications World" journal. They provided specific examples of its application in security checks and boarding processes, but did not provide experimental data or specify which CNN model is most suitable for airport environments. Similarly, She's[2] paper "A Brief Analysis of the Application of Facial

[1] Corresponding Author: Xiaoyu LI, yuter1211@163.com.

Recognition Systems in Airport Security Checks" in the "Electronic World" journal highlights the significant value of facial recognition in airport security checks but lacks specific model applications and data comparisons.

This study aims to address this gap by investigating CNN algorithms to identify the most suitable model for airport settings. Three classical CNN algorithmic models were employed to simulate facial detection in airport environments. The LFW dataset, known for its unconstrained facial recognition data, was used to simulate the complex scenarios and environments present in airports.

The results of the experimental simulations revealed the accuracy rates of the three models, with the ResNet-50 model achieving an impressive accuracy rate of 71.9%. Among the three classical CNN models studied, ResNet emerged as the most suitable model for facial detection in airport environments.

In conclusion, this study provides valuable insights into the application of CNN-based facial recognition technology in airport security management. By identifying ResNet as the optimal model for airport environments, this research contributes to the development of more effective and reliable facial recognition systems in airports. These advancements have the potential to enhance overall airport security measures, ensuring safer and more secure travel experiences for passengers worldwide.

2. Overview of Face Recognition Technology

As a deep learning model, Convolutional Neural Network (CNN) has performed very well in the field of image recognition. The main advantage is that it can automatically learn the features in the image, without the manual design of the feature extractor, simplifying the design process of the face recognition system. Through multi-level convolution and pooling operation, CNN can gradually abstract the high-level features of images, thus improving the accuracy and robustness of identification. For example, by introducing more convolution layers and pooling layers, the deep CNN model achieves excellent performance in face recognition tasks[3]. The CNN's fundamental architecture comprises five layers: input layer, convolutional layer, pooling layer, fully connected layer, and output layer. This configuration is depicted in Figure 1 below.

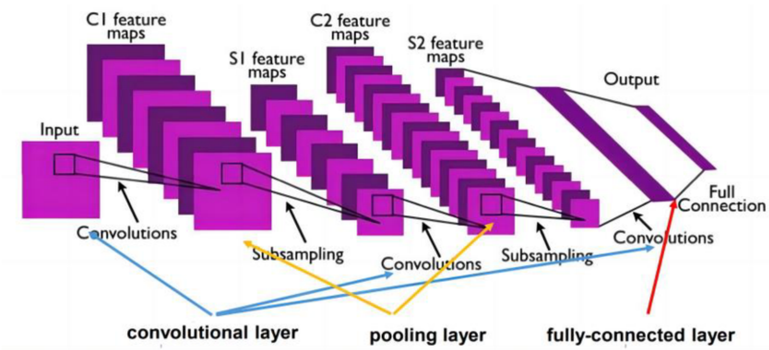

Figure 1. The CNN network structure.

Input layer: input image information

Convolution layer[4]: Extract graphic features

Pooling layer[5]: The extracted feature map is extracted again

Fully connected layer[6]: summarize the image of the convolutional layer and the pooling layer

Output layer: judge and output the results according to the full connection

In the whole process, the input layer converts the image into a two-dimensional matrix, and then enters the convolution layer consists of three channels representing the three primary color RGB channels, and then convolution them. The convolution here is the extraction of the feature graph, and the extraction process is shown in Figure 2.

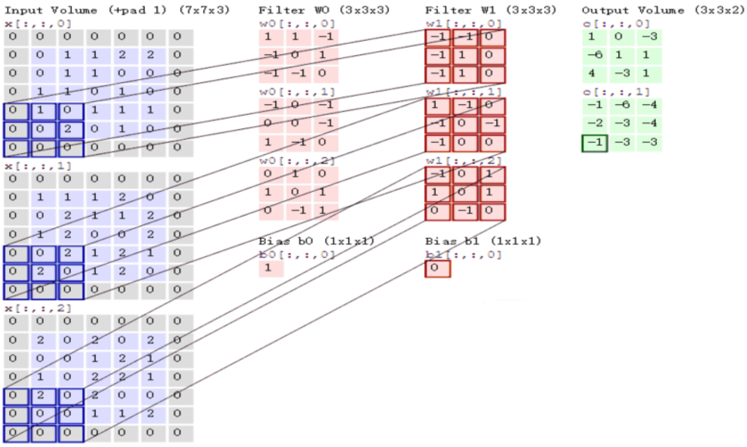

Figure 2. Processing process of CNN.

The pooling layer will extract the feature map obtained by the convolution layer again, and select the largest number in the input data to form a new feature map, as shown in Figure 3, which can reduce the role of overfitting[7] and reducing the dimension.

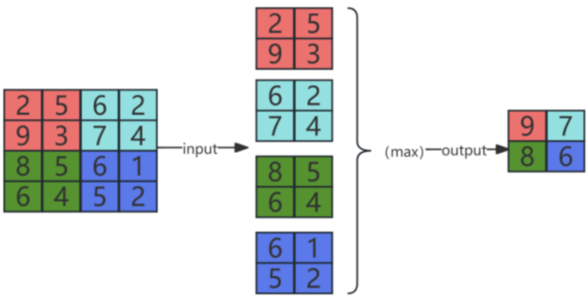

Figure 3. Pooling layer processing.

At this time, we only need to turn all the extracted feature maps into a string of one-dimensional matrices. By calculating this string of data, we can get a probability value, which is the probability value of whether the input picture is human or not. However, if our identification result is classification, we need to select a result with the largest probability in these classifications through the output layer to output.

3. Model Design

In the model design, the classic frame structure VGGNet16, GoogLeNet and ResNet-50 were selected for the effect verification of the model.

3.1. VGGNet Model

For the research of VGGNet16 model in face, Yan and Jiang[8] team used faceScrub data and achieved high accuracy. However, with the characteristics of the data set, the image is collected as photos of celebrities and the image quality is good, which can not be used as a reference for face recognition model in the airport environment.

The VGGNet deep convolutional neural network comprises 6 network structures with 11 to 16 layers, with the most commonly used being VGG 16, which consists of 13 convolutional layers and 3 fully connected layers. This is illustrated in Figure 4.

<table>
<tr><th colspan="6">ConvNet Configuration</th></tr>
<tr><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th></tr>
<tr><td>11 weight layers</td><td>11 weight layers</td><td>13 weight layers</td><td>16 weight layers</td><td>16 weight layers</td><td>19 weight layers</td></tr>
<tr><td colspan="6">input(244 × 244 RGB image)</td></tr>
<tr><td>conv3-64</td><td>conv3-64
LRN</td><td>conv3-64
conv3-64</td><td>conv3-64
conv3-64</td><td>conv3-64
conv3-64</td><td>conv3-64
conv3-64</td></tr>
<tr><td colspan="6">maxpool</td></tr>
<tr><td>conv3-128</td><td>conv3-128</td><td>conv3-128
conv3-128</td><td>conv3-128
conv3-128</td><td>conv3-128
conv3-128</td><td>conv3-128
conv3-128</td></tr>
<tr><td colspan="6">maxpool</td></tr>
<tr><td>conv3-256
conv3-256</td><td>conv3-256
conv3-256</td><td>conv3-256
conv3-256</td><td>conv3-256
conv3-256
conv3-256</td><td>conv3-256
conv3-256
conv3-256</td><td>conv3-256
conv3-256
conv3-256
conv3-256</td></tr>
<tr><td colspan="6">maxpool</td></tr>
<tr><td>conv3-512
conv3-512</td><td>conv3-512
conv3-512</td><td>conv3-512
conv3-512</td><td>conv3-512
conv3-512
conv3-512</td><td>conv3-512
conv3-512
conv3-512</td><td>conv3-512
conv3-512
conv3-512
conv3-512</td></tr>
<tr><td colspan="6">maxpool</td></tr>
<tr><td>conv3-512
conv3-512</td><td>conv3-512
conv3-512</td><td>conv3-512
conv3-512</td><td>conv3-512
conv3-512
conv3-512</td><td>conv3-512
conv3-512
conv3-512</td><td>conv3-512
conv3-512
conv3-512
conv3-512</td></tr>
<tr><td colspan="6">maxpool</td></tr>
<tr><td colspan="6">FC-4096</td></tr>
<tr><td colspan="6">FC-4096</td></tr>
<tr><td colspan="6">FC-1000</td></tr>
<tr><td colspan="6">soft-max</td></tr>
</table>

Figure 4 .VGGNet16 Network structure.

Its structural characteristics are unified with the convolution kernel and pooling size. The convolution operation 33 and 22 are carried out in the network, and the strategy of convolutional layer stacking[9] is adopted to form multiple continuous convolution layers into a convolution layer group.

Compared with the single convolutional layer, the convolution group can expand the sensory range of the visual field and enhance the learning ability and feature expression ability of the network. Compared with the convolution layers with larger cores, the series method with multiple convolution layers with small convolution kernels can reduce the network parameters. Each layer of convolution will perform ReLU nonlinear operation[10], which can further improve the learning ability of defeatures.

3.2. GoogLeNet Model

GoogLeNet[11] model is applied in the field of human face, mainly used for the application of human expression recognition[12]. The GoogLeNet network is characterized by the Inception[13] block in the medium-convolution block, whose structure is shown in Figure 5. It adopts four parallel paths to obtain the perception data information of the unused area through the first three convolution cores of 11, 33 and 55, and finally extracts the data point information by pooling the 33 maximum kernel.

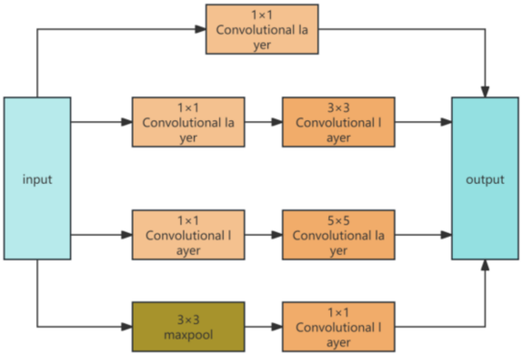

Figure 5. Inception block.

The advantage is that the addition of 11 convolution kernel, greatly reduces the number of parameters, and also can improve the expression ability of the network. And the entire model uses a modular structure, so the setting and adjustment is relatively simple, and can use the flat pooling layer as the full connection layer in the entire network architecture. This structure provides an accuracy of about 0.6%.

Since the main composition of GoogLeNet consists of the input layer, 5 convolutional modules and the data output module, which contains 22 data layers and 5 pooling layers. The convolution module of the first two groups also presents the broad convolution layer and the maximum pooling layer. The convolutional module structure of the third, fourth and fifth layers is composed of the initial module structure and the largest one, which is more unique in that the input layer is composed of the average layer, Dropout, and fully connected layer[14].

3.3. ResNet Model

ResNet[15] is widely used, but its most suitable application in the airport is to use ResNet to predict the airport visibility[16]. The discovery that simply increasing the network depth by stacking convolution layers does not improve the effect of the model, and will even make the model worse and more difficult to train, as shown in Figure 6. The ResNet model has been born to solve this problem. The reason for this problem is that due to the increase of network depth, the gradient slowdown and gradient phenomenon will become very serious. The gradient backates to the earlier layer, and repeated multiplication may make the gradient infinite, and its performance saturation and even deteriorate rapidly with the deepening of the network[17] .

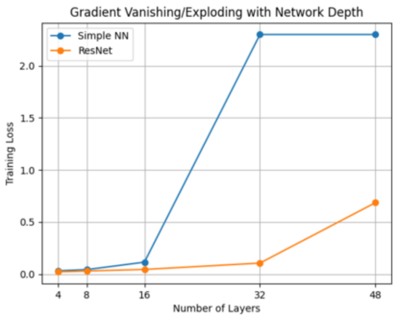

Figure 6. Plot of network layers and error rate.

Because the deep networks are difficult to train. The network is designed as H (x) = F (x) + x, and its structure is shown in Figure 7 below. It can be converted to learn a residual function, F (x) = H (x) -x. As long as F (x) =0, it constitutes an identity map H (x) = x. Moreover, fitting the residuals is certainly easier.

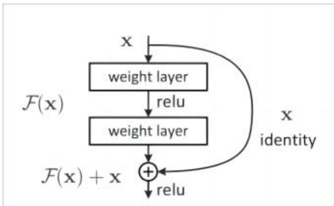

Figure 7. The residual structure.

ResNet by introducing the residual connection, increasing the network depth, and through the pre-training weight and other techniques, the training problem of the traditional deep network mentioned above is solved. To compare with the other two networks, ResNet can obtain more features of the image by increasing the number of network layers, so as to extract more image features in more complex situations, while GoogLeNet and VGGNet can no longer extract more feature information of the image by increasing the network structure. In a complex environment, more features can be extracted, which means a higher accuracy rate. Applied it to face detection in airport environment, ResNet has innate advantages.

4. Model Accuracy Comparison

4.1. LFW Data Set

In order to verify that the model is suitable for application in the airport environment, this paper needs to verify it as a specific data set similar to the airport environment. Therefore, the LFW[18] (Labeled Faces in the Wild) data set, a facial photo database is designed to study unconstrained face recognition problems. This dataset contains a network of over 13,000 face images collected from the following locations. Each face is marked with a human name map. The 1,680 people in the photos had two or more different photos in the dataset. Unconstrained face recognition is similar to the actual environment of the airport, so the LFW data set is used. Some pictures of the data set are shown in Figure 8.

Figure 8. Pictures of part of the data set.

4.2. Methods of Model Evaluation

Training on the above three models, after each training cycle, using the ` model. The eval ()` method sets the model to an evaluation mode. In the ` Torch. Inside the no_grad ()` Context Manager, the gradient calculation is disabled for evaluation. The model traverses the test data loader (` test _ loader`) loads the image and its corresponding labels. These images and labels are then moved to the available device CPU. Images undergo a model processing to obtain the output of the model. Using the ` torswitch. The max ()` function finds the maximum value and its corresponding index along the first dimension of the output tensor (representing the batch size), representing the prediction category. Using the ` (predicted == labels).sum(). The iitem ()` calculates the number of correctly predicted samples in the batch. In the test dataset, the cumulative total sample number and the number of correctly predicted samples. Finally, the accuracy is calculated, and once all the samples in the test dataset are evaluated, the overall accuracy is calculated. Accuracy was calculated by dividing the total number of correctly predicted by the total number of samples and multiplying by 100%.

The pseudo-code was assessed below in Figure 9.

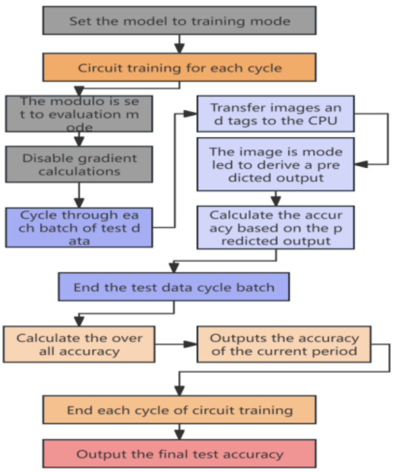

Figure 9. Flow chart of model evaluation.

At the end of each training cycle, the final test accuracy, reflecting the model performance on the entire test dataset.

4.3. Analysis of the Experimental Results

The effect diagram of identifying all ResNet-50 models for prediction classification in FLW is shown in Figure 10.

Figure 10. CNN classification identification.

The three models were imported into the dataset for model evaluation and part of the code is shown below in Figure 11. The resulting accuracy is performed as shown in Figure 12 below.

```
# Define ResNet-50 model
class ResNet50FaceRecognition(nn.Module):
    def __init__(self, num_classes):
        super(ResNet50FaceRecognition, self).__init__()
        self.resnet50 = torchvision.models.resnet50(pretrained=True)
        num_ftrs = self.resnet50.fc.in_features
        self.resnet50.fc = nn.Linear(num_ftrs, num_classes)

    def forward(self, x):
        return self.resnet50(x)
# Define VGGNet-16 model
class CGGNet16FaceRecognition(nn.Module):
    def __init__(self, num_classes):
        super(CGGNet16FaceRecognition, self).__init__()
        self.cggnet16 = torchvision.models.vgg16(pretrained=True)
        self.cggnet16.fc = nn.Linear(512 * 7 * 7, num_classes)

    def forward(self, x):
        return self.cggnet16(x)
# Load GoogLeNet model pre-trained on ImageNet
base_model_inception = InceptionV3(weights='imagenet')
inception_model = Model(inputs=base_model_inception.input, outputs=base_model_inception.layers[-2].output)  # Remove the last softmax layer
# Path to the dataset directory
dataset_dir = r'C:\Users\x\Desktop\Project_cnn_face\lfw'
```

Figure 11. Experimental code (in part).

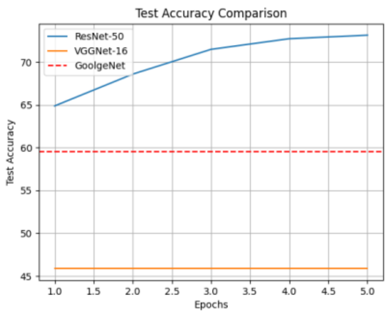

Figure 12. Comparison of the three models.

As can be seen from Figure 11 above, ResNet-50 performed the best on the LFW data set used, with an accuracy of 71.90% in its model evaluation. However, the GoogLeNet had an accuracy rate of 59.50%. VGGNet16 performed the worst, with an accuracy of 45.87%. In the LFW unconstrained scenario, the performance of various models is analyzed. VGGNet has the lowest accuracy due to the small number of network structure layers, but blindly improving the number of network structure layers will lead to the reduction of its accuracy. The Inception block in GoolgNet can indeed effectively improve the accuracy of the model, but under the unconstrained environment conditions, the accuracy of the model is still limited. And ResNet-50 network due of its unique residual structure, improve the accuracy of the model in the unconstrained environment, effectively solve the model from the increase of the network structure layer, ensure its accuracy in complex environment, suitable for application to the face recognition in the airport environment.

5. Summary

This study extensively investigates the widespread utilization of CNN-based facial recognition technology in airport settings, conducting a comprehensive comparative analysis of three classical CNN models. Trained on the LFW dataset, these models underwent rigorous evaluation. By simulating scenarios mirroring airport conditions, their effectiveness in facial recognition was meticulously examined. The findings unequivocally highlight ResNet-50 as the top performer, demonstrating significantly superior accuracy compared to its counterparts. Consequently, ResNet emerges as the optimal solution for addressing airport detection tasks due to its heightened precision and reliability. This research provides invaluable insights into the seamless integration of CNN-based facial recognition within airport security management. Moreover, it offers a robust theoretical framework for the judicious selection of the most suitable technology for practical deployment. By leveraging these insights, airport authorities can enhance security measures while ensuring the efficiency and reliability of facial recognition processes. Consequently, this not only enhances the overall security posture of airports but also contributes to safer and more secure travel experiences for passengers worldwide.

In an era characterized by heightened security concerns and increasing passenger volumes, the adoption of advanced technologies such as CNN-based facial recognition holds immense promise in bolstering airport security protocols. Through the implementation of ResNet-50 and similar cutting-edge models, airports can remain at the forefront of security innovation, thereby safeguarding the well-being and confidence of travelers globally.

Acknowledgments

This work was supported by the [Innovation and Entrepreneurship training Program for college students] under Grant [202311116095].

References

[1] Chen RW. An Analysis of the Application of Facial Recognition Technology in the Airport Domain. Communication World, 2019, 26(09): 10-11.2023, 24(12):146-148.DOI:10.16009/j.cnki.cn13-1295/tq.2023.12.039.

[2] She ST. The application of face recognition system is analysed in airport security. Journal of electronic world, 2018, (15): 201. DOI: 10.19353 / j.carol carroll nki DZSJ. 2018.15.116

[3] Zhang SL. Summary of image recognition based on convolutional neural networks. Journal of Xi'an Aeronautical University, 2023, 41 (01): 74-81. DOI:10.20096/j.xhxb.1008-9233.2023.01.013.

[4] Yi X. Asymptotic Spectral Representation of Linear Convolutional Layers. IEEE Transactions on Signal Processing, 2022: 566-581. http://dx.doi.org/10.1109/tsp.2022.3140718. DOI:10.1109/tsp.2022.3140718.

[5] Singh P, Raj P, Namboodiri V P. EDS pooling layer. Image and Vision Computing, 2020, 98: 103923. http://dx.doi.org/10.1016/j.imavis.2020.103923. DOI:10.1016/j.imavis.2020.103923.

[6] Ma W, Lu J. An Equivalence of Fully Connected Layer and Convolutional Layer.Cornell University - arXiv, Cornell University - arXiv, 2017.

[7] Ding HW, Wan L, Zhou K, et al.. Research on intrusion detection based on deep convolutional neural networks. Computer Science, 2019, 46 (10): 173-179.

[8] Yan XB, Jiang ZF. Research on face recognition method based on VGGNet deep convolutional neural network. Computer Knowledge and Technology, 2023, 19 (25): 34-37. DOI:10.14004/j.cnki.ckt.2023.1370.

[9] Maia H, Souza M, Santos A, et al. Learnable Visual Rhythms Based on the Stacking of Convolutional Neural Networks for Action Recognition//2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA.2019. http://dx.doi.org/10.1109/icmla.2019.00290. DOI:10.1109/icmla.2019.00290.

[10] Liu X, Wang D. Universal consistency of deep ReLU neural networks. Science in China: Information Science, 1-15 [2024-03-20]. http://kns.cnki.net/kcms/detail/11.5846.TP.20240311.1850.004.html.

[11] Ren M. Overview of the image visual recognition algorithm. Smart Computer and Application, 2019, 9 (03): 294-297.

[12] Zhang HL, Bai XY. Using the face expression recognition method for optimized pruning GoogLeNet. Computer Engineering and Application, 2021, 57 (19): 179-188.

[13] Szegedy C, Wei Liu, Yangqing JIA, et al. Going Deeper with Convolutions//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA.2015. http://dx.doi.org/10.1109/cvpr.2015.7298594. DOI:10.1109/cvpr.2015.7298594.

[14] Xu MZ, Zhang J, Peng HB. Face and facial expression recognition method based on an improved Goog Le Net convolutional network. Computers and Information Technology, 2024, 32 (01): 32-36. DOI:10.19414/j.cnki.1005-1228.2024.01.028.

[15] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA.2016. http://dx.doi.org/10.1109/cvpr.2016.90. DOI:10.1109/cvpr.2016.90.

[16] Qi Y, Fan H, Zhang RH, et al. Airport visibility prediction based on the ResNet convolutional neural network. Automation application, 2023, 64 (24): 206-209.

[17] Li BZ, Liu K, Gu JJ, etc. Review of Convolutional neural network research. Computer Age, 2021, (04): 8-12 + 17. DOI:10.16644/j.cnki.cn33-1094/tp.2021.04.003.

[18] Knoche M, Hormann S, Rigoll G. Cross-Quality LFW: A Database for Analyzing Cross-Resolution Image Face Recognition in Unconstrained Environments//2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), Jodhpur, India. 2021. http://dx.doi.org/10.1109/fg52635.2021.9666960.DOI:10.1109/fg52635.2021.9666960.