Face Recognition Based on ResNet Architecture

Peilu Zhu
Yongkang No.1 High School, Chengnan Road, Yongkang, Zhejiang Province 321300, China

Keywords: ResNet, AdamW, Early Stopping, Deep Learning, Face Recognition.

Abstract: With deep learning developing, face recognition techniques are widely applied in various scenarios. Currently, most methods for achieving face recognition are based on Convolutional Neural Networks (CNN), although they have flaws. To solve the problem of long training time and improve accuracy, the Residual Network (ResNet), as a strong tool, was introduced as the main architecture in place of CNN to deal with complex data. Besides, Adam with Weight Decay (AdamW) was chosen as the optimizer and early stopping was implemented to improve model performance. 400 facial images from 40 different individuals were selected from the Olivetti faces dataset in the experiment. The experiment was conducted successfully. Losses, training time and accuracy of ResNet-18 and ResNet-50 in both 10 epochs and 30 epochs were calculated and then compared. The experimental results show that ResNet-18 performed better than ResNet-50 on such a 400 images dataset with overall lower losses and less training time. The training time is saved by 10%~15% after introducing the early stopping method.

1 INTRODUCTION

Face recognition is a kind of biological recognition technology. Nowadays face recognition technique is widely used in multiple situations, including security monitoring, password authentication and law enforcement (Taskiran, Kahraman, & Erdem, 2020). Though this technique makes life more convenient and improves the security of private data to a certain extent, the problems of slow recognition and uncertainty still exist when the dataset is large. It is vitally significant to accelerate the recognition process and improve the accuracy of recognition.

There are various methods to achieve face recognition (Changwei, Jun, Lingyun, Yali, Sheng, 2020) (Opanasenko, Fazilov, Mirzaev, Sa’dullo ugli Kakharov, 2024) (Opanasenko, Fazilov, Radjabov, 2024.). For example, Opanasenko, V. M., Fazilov, S. K., Mirzaev, O. N. and Sa’dullo ugli Kakharov, S. used a method for recognizing faces in mobile devices, based on an ensemble approach to solving the problem (Opanasenko, Fazilov, Mirzaev, Sa’dullo ugli Kakharov, 2024). Compared with others, methods based on deep learning seem to be the most simplified and efficient. Deng, N., Xu, Z., Li, X., Gao, C. and Wang, X. used a noise-applied spatial clustering algorithm based on density to cluster a large dataset to a self-constructed dataset (Deng, Xu, Li, Gao, Wang, 2024). They reduced the uncertainty of the model and saved space taken up by the dataset. Said, Y., Barr, M. and Ahmed, H. E. designed and evaluated a deep learning model based on CNN to detect facial information in a real-time environment (Said, Barr, Ahmed, 2020). They successfully adjusted parameters to improve the accuracy in standard datasets and real-time input. Khalifa, A., Abdelrahman, A.A. and Hempel, T. introduced a robust and efficient CNN-designed model for face recognition (Khalifa, Abdelrahman, Hempel, 2024). They succeeded in achieving a balance through incorporating multiple features and attention mechanisms. Xie, Z., Li, J. and Shi, H. focused on the influence of the increasing number of neural and feature maps (Xie, Li, Shi, 2019). They used Python with Keras methods to test CNN in face recognition and gained accuracy close to 100%. Liu, Y. and Qu, Y. combined multiple algorithms to build an improved multitask face recognition CNN (Liu, Qu, 2024). They reached an accuracy of 99.05% in feature matching.

Du, A., Zhou, Q., and Dai, Y. utilized Intersection over Union (IoU) to quantify the ratio of task-relevant features and evaluate the generalization ability of their Residual Networks (ResNet) model set (Du, Zhou, Dai, 2024). The results show that ResNet is a

Zhu and P.
Face Recognition Based on ResNet Architecture.
DOI: 10.5220/0013486600004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 18-22
ISBN: 978-989-758-754-2

stronger network than CNN, and it has the potential to be an improvement in face recognition.

In this paper, the assumption is that ResNet is equipped with high performance in face recognition, including less loss, less time taken and higher accuracy. ResNet models with different numbers of layers were chosen as the main architecture of face recognition. 400 face images from 40 individuals were selected from the Olivetti Faces dataset to test the performance. Experiment was taken and outcomes were shown in the following contexts to prove the assumption.

2 METHOD

2.1 ResNet

ResNet is an improvement to CNN (Wu, Shen, Van Den Hengel, 2019). It solves the problem that training is harder with increasing depths of the network. The core thought of ResNet is that the network should learn the difference between inputs and outputs, namely residual, instead of learning mapping relation directly. Residual blocks are the basic building units of ResNet. Residual blocks contain two paths: one is a stack of convolutional layers and another is identity connections. Compared with traditional CNN, identity connections are added to ResNet models. Identity connection allows to input of the outputs that skip some layers directly. Then add these outputs with the original inputs to form the final outputs. Figure 1 shows the local architecture of ResNet. Identity connection solves the problem of the disappearance of gradients and is beneficial to the flow of gradients. Resnet increases the number of layers to improve performance. It has very deep architecture, such as ResNet-18, ResNet-50 and ResNet-101, the number represents the number of residual blocks in the network (Koonce, 2021). Batch normalization (BN) is used after each convolution layer. It mainly solves the problem of internal covariate shifts in deep networks, accelerating the process of training and improving the stability of models. BN reduces the effect of overfitting and makes the gradient descent algorithm converge more rapidly. (1) is a mathematics expression of BN:

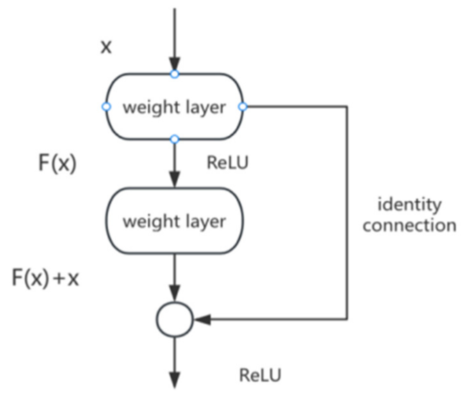

Figure 1: The architecture of ResNet model.

$$\mathrm{BN}(x) = \gamma\left(\frac{x-\mu}{\sigma}\right) + \beta \tag{1}$$

μ is the main value of mini-batch data, σ is the standard deviation, β and γ are parameters. Rectified linear unit (ReLU) is taken as the non-linear activation function. In the process of counterpropagation, the derivative of ReLU is 1 (when $x>0$) or 0 (when $x<0$), so gradients will not decrease significantly with the increase in depths of the network. To decrease the spatial dimensions of the feature maps, those convolutional layers and polling layers with more than one length of steps are downsampled. After the extraction of features, features are mapped to the final classified outcomes through the fully-connected layer.

3 EXPERIMENT

3.1 Dataset chosen and the overview of experiment

The dataset used in the experiment was downloaded from the website cs.nyu.edu. The Olivetti Faces dataset is available at: cs.nyu.edu (https://cs.nyu.edu/~roweis/data.html). The Olivetti Faces dataset is a widely used library of 400 grayscale images of the faces of 40 different people. The images were taken at varied angles and light conditions with varied facial expressions. It was first compiled by the AT&T Cambridge laboratory.

The experiment contains two contents: detecting faces and training models. In detecting faces, the Haar cascade classifier was used as the main tool to crop images of faces. In training models, ResNet-18 and ResNet-50 were chosen to recognize faces. Figure 2 shows the main steps of the whole experiment.

3.2 Face detection with Haar cascade classifier

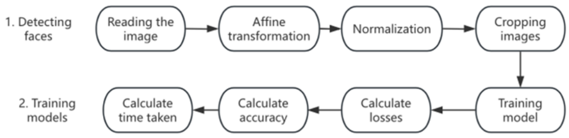

Figure 2: The main steps of the whole experiment.

Usually, in an image, faces do not take up the whole space of the image and there may be more than one face in an image. For the Olivetti faces dataset, 400 images were combined into a big image. Figure 3 shows the combined image of the Olivetti faces dataset. It is necessary to crop each face into a rectangle. Haar cascade classifier is an effective tool in object detection (Madan, 2021). This classifier is based on Haar features and the AdaBoost algorithm (Madan, 2021). Haar cascade classifier was used to detect faces and crop faces in this experiment.

For different sizes, images should be cropped to the specific sizes, which is up to the input requirement. By using the Haar cascade classifier, a list of positions of edges of faces was returned. After traversing these positions, rectangles were drawn on each face. Every detected face was cropped to the same size of rectangles by the Haar cascade classifier. It improves the accuracy of outcomes. Then the values of pixels were normalized into specific ranges. This operation helped the classifier recognize faces more easily, and it not only improved the efficiency of computation but also enhanced stability.

All the cropped images were put into a new folder and then labeled with numbers. Figure 4 shows a part of the cropped images with labels in the new folder. The folder could be converted to a dataset. All the grayscale images were changed into color images. Now the preprocessing of data has already been done.

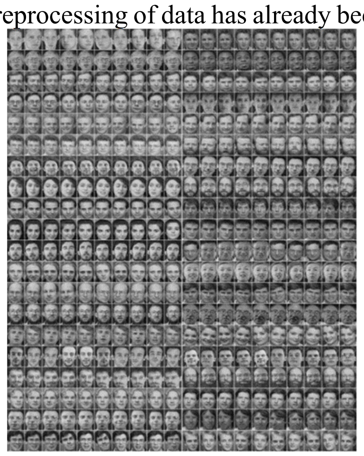

Figure 3: The original Olivetti Faces image downloaded.

Figure 4: The cropped faces in a new folder.

3.3 Training ResNet model

In the experiment, both ResNet-18 and ResNet-50 were chosen. The preprocessing dataset was put into a pre-training Resnet architecture model. The final full-connected layer was corrected to match the number of 400 classifications. The number of epochs was set to 10 and 30. Every sample would be dealt with. Firstly, data traveled through every layer with the ReLU activate function in forward propagation. With identity connection, output in each layer was added with the original inputs to form the final outputs. In every iteration, the model made predictions based on the current samples and calculated the loss function. The loss function used in the experiment was Mean Squared Error(MSE). Then in the process of counterpropagation, loss gradient was used to recursively calculate the gradient of each parameter in ResNet. AdamW was the optimizer to update weights and bias (Zhou, Xie, Lin, Yan, 2024). AdamW could adjust the learning rates of every parameter, accelerating the convergence rate. Compared to the Adam optimizer, AdamW separates weight decay from gradient update, which improves the efficiency of weight adjustment (Zhou, Xie, Lin, and Yan, 2024). That was the complete process of an epoch. Through 30 epochs, loss in each epoch was calculated. By constantly adjusting the weights and parameters, the loss at every epoch was reduced. Figure 5 shows loss in each epoch of the ResNet-18 model in 10 epochs. Figure 6 shows loss in each epoch of the ResNet-18 model in 30 epochs.

By comparing mean, maximum and minimum values, it was obvious that losses were decreased with an increasing number of epochs in the same model. Resnet-18 was replaced with Resnet-50 and losses were also calculated. Figure 7 shows loss in each epoch of the ResNet-50 model in 10 epochs.

ResNet-50 had a greater loss than ResNet-18 in the first epoch because it has deeper layers. It needs more time to adjust more parameters in the beginning. Table 1 shows the time taken in different epochs from the ResNet-18 and ResNet-50 models.

Though the ResNet-50 model has less loss, the total training time taken in 10 epochs was significantly longer than ResNet-18. With an increasing number of epochs, both losses in the start epoch and mean loss decreased. This is because the number of times of adjusting weights increased and the degree of fit increased gradually.

The most significant problem is consuming a long time. Actually, in the experiments, the AdamW already accelerated the rate of convergence and data enhancement already reduced the training time. So early stopping was introduced. As weights decay in the AdamW algorithm, early stopping is also a regularization technique. It adds a supervised process to the training process. When the loss in the test set did not decrease, overfitting might happen. In order to avoid further overfitting, the epoch was early stopped. By using this method in the experiment, the total time taken could be reduced by 10%~15%.

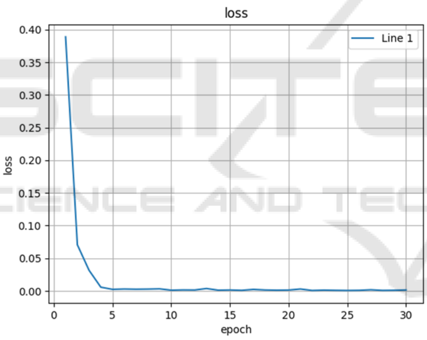

Figure 5: The line chart of losses in 10 epochs of ResNet-18 model.

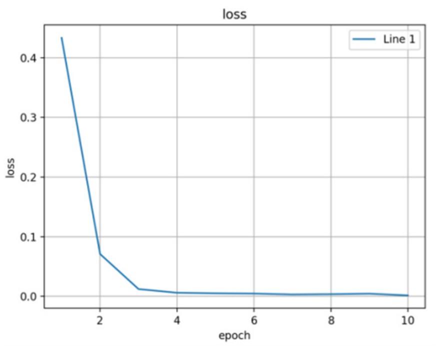

Figure 6: The line chart of losses in 30 epochs of ResNet-18 model.

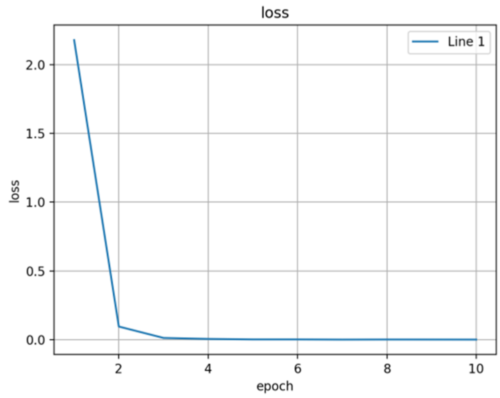

Figure 7: The line chart of losses in 10 epochs of ResNet-50 model.

Table 1: The total time taken in different models and numbers of epochs

Model	Total time taken in 10 epochs	Total time taken in 30 epochs
ResNet-18	12’16’’23°	39’54’’17°
ResNet-50	47’16’’17°	173’4’’23°

Table 2: The accuracy gained from data before and after transformation and enhancement.

Models	ResNet-18 in 10 epochs	ResNet-50 in 10 epochs	ResNet-18 in 30 epochs
Initial Accuracy	98.70%	95.50%	98.70%
Final Accuracy	100.00%	100.00%	100.00%

3.4 Evaluating ResNet model

In order to calculate the accuracy of the model after the experiment, the samples were not labeled, while the copies were labeled with the names of the owners of the faces. If the outputs were labeled correctly by the ResNet models, the number of correct counts would add one. The total number of correct classifications was divided by the total number of labels (400) and accuracy was gained. Table 2 shows the accuracy of different models with different numbers of epochs.

Both in 10 epochs, the accuracy of ResNet-50 was lower than the ResNet-18 model. By increasing the number of epochs in the same model, accuracy did not change. The final accuracy was gained from data introducing enhancement, and Adam optimizer was replaced by AdamW. The final accuracy was up to 100%.

4 CONCLUSIONS

In conclusion, by introducing AdamW and early stopping, the improved ResNet architecture was quite successful. Training time was saved by 10%~15%, and the accuracy was 100%. This paper offers a methodology for using ResNet to achieve face recognition and proves its high performance. Furthermore, for different quantities of data, it is important to choose suitable ResNet architectures. However, the research did not focus on large datasets and low-resolution images. In real-time situations, other problems may exist and still need to be improved.

REFERENCES

Changwei, L. U. O., Jun, Y. U., Lingyun, Y. U., Yali, L. I., & Sheng, W. A. N. G., 2020. Overview of research progress on 3-d face recognition. Journal of Tsinghua University (Science and Technology), 61(1), 77-88.

Deng, N., Xu, Z., Li, X., Gao, C., & Wang, X., 2024. Deep Learning and Face Recognition: Face Recognition Approach Based on the DS-CDCN Algorithm. Applied Sciences, 14(13), 5739.

Du, A., Zhou, Q., Dai, Y., 2024. Methodology for Evaluating the Generalization of ResNet. Appl. Sci., 14, 3951.

Khalifa, A., Abdelrahman, A.A., Hempel, T. et al., 2024. Towards efficient and robust face recognition through attention-integrated multi-level CNN. Multimed Tools Appl.

Koonce, B., 2021. ResNet 50. In: Convolutional Neural Networks with Swift for Tensorflow. Apress, Berkeley, CA.

Liu, Y., & Qu, Y., 2024. Construction of a smart face recognition model for university libraries based on FaceNet-MMAR algorithm. Plos one, 19(1), e0296656.

Madan, A., 2021. Face recognition using Haar cascade classifier. Int. J. Mod. Trends Sci. Technol, 7(01), 85-87.

Opanasenko, V. M., Fazilov, S. K., Mirzaev, O. N., & Sa’ dullo ugli Kakharov, S., 2024. An Ensemble Approach To Face Recognition In Access Control Systems. Journal of Mobile Multimedia, 749-768.

Opanasenko, V.M., Fazilov, S., Radjabov, S.S. et al., 2024. Multilevel Face Recognition System. Cybern Syst Anal 60, 146–151.

Said, Y., Barr, M., & Ahmed, H. E., 2020. Design of a face recognition system based on convolutional neural network (CNN). Engineering, Technology & Applied Science Research, 10(3), 5608-5612.

Taskiran, M., Kahraman, N., & Erdem, C. E., 2020. Face recognition: Past, present and future (a review). Digital Signal Processing, 106, 102809.

Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. Pattern recognition, 90, 119-133.

Xie, Z., Li, J., & Shi, H., 2019. A face recognition method based on CNN. In Journal of Physics: Conference Series (Vol. 1395, No. 1, p. 012006). IOP Publishing.

Zhou, P., Xie, X., Lin, Z., & Yan, S., 2024. Towards understanding convergence and generalization of AdamW. IEEE Transactions on Pattern Analysis and Machine Intelligence.