

Face recognition technology based on ResNet-50

Jingyu Liu

University of Macau. Macau, China.999078

DC22836@umac.mo

Abstract. Face recognition technology is progressively finding its place across diverse domains. In pursuit of enhancing the efficacy of face recognition systems, this study employs a ResNet-50 deep convolutional neural network. The dataset is meticulously gathered and processed via OpenCV, thus amplifying the precision and utility of face recognition. ResNet, an advanced convolutional neural network, incorporates the concept of residual connections, bridging convolutional layers through shortcut connections. These connections facilitate the addition of input to output, forming residual blocks. Consequently, ResNet-50 efficiently tackles the vanishing gradient issue, enabling the training of exceptionally deep networks. With 49 convolutional layers and a fully connected layer, ResNet-50 boasts a robust architecture. To emulate varying brightness conditions, post-collection image adjustments are applied randomly. This strategy curbs the impact of divergent lighting scenarios on recognition accuracy, bolstering the model's practical applicability. Notably, experimental outcomes underscore the commendable performance of the trained ResNet-50 model in face recognition trials. This substantiates the broad-spectrum viability of face recognition technology in domains such as security surveillance, human-machine interaction, identity verification, and beyond.

Keywords: ResNet-50, Face Recognition, Deep Convolutional Neural Networks, Deep Learning.

1. Introduction

Face recognition has garnered substantial attention in the realm of computer science, particularly as a biometric technology. With the continuous advancement of deep learning and artificial intelligence research, there has been remarkable progress in the performance of AI technology. This progress isn't limited to the security domain; instead, face recognition technology finds widespread applications in diverse areas including mobile device unlocking, electronic payments, network account access, and other scenarios. This proliferation offers users a convenient means of biometric verification. Furthermore, its significance extends to domains like individual tracking, social media, and real-name authentication in travel. However, despite its myriad benefits, face recognition technology encounters several hurdles. Factors such as lighting conditions, facial expressions, and viewing angles can significantly impact its accuracy. Addressing these challenges necessitates intensive research and the development of innovative solutions.

In the field of computer vision, the convolutional neural network (CNN) has an indispensable position. Among a series of deep convolutional neural network models released in recent years, ResNet-50 (Residual Network-50) has been widely recognized in image recognition and classification, target

detection, and other visual tasks due to its excellent performance and architectural innovation. Stand out from the crowd of models. ResNet-50 was proposed by Kaiming He et al. In 2015, the common problem of gradient disappearance when training deep networks was solved. A key feature of ResNet-50 is the residual structure it introduces. It allows the network to directly pass the input to the output, and to make cross-layer connections by introducing residual blocks between different layers. This connection method allows the network to build deeper layers, effectively overcoming the problems of deep network training such as gradient disappearance and limited expression. Make the network easier to train while enabling identity mapping. As a result, it has achieved excellent performance in test tasks such as image classification and detection, object recognition, and face feature recognition.

The core of this article is to deeply explore the application of ResNet-50 in the field of face recognition. The article uses OpenCV to obtain the face dataset and performs random brightness adjustments on the dataset. Subsequently, ResNet-50 models were trained and evaluated using these modified datasets. The overall goal is to gain insight into the capabilities of ResNet-50 in learning facial image features. In the following sections, we detail our approach and experimental results, demonstrating the effectiveness and potential of ResNet-50 on face recognition tasks.

2. Related work

Before ResNet, the development of face recognition technology had a long history. In 1991, Matthew Turk and others introduced Eigenfaces [1]. Eigenfaces is a face recognition method based on principal component analysis. It calculates the face image's main feature vector, learns many images for the feature image's weighted reconstruction, and projects the face image into a low-dimensional space to realize real-time face positioning and recognition.

Compared to Eigenfaces, Fisherfaces [2] focuses on maximizing the difference between categories and minimizing the difference within categories. Fisherfaces is a dimensionality reduction algorithm based on Fisher's discriminant analysis (FDA). In the application of face recognition, Fisherfaces can find the best projection direction by calculating the Within-Class Scatter and Between-Class Scatter and judging the Fisher discriminant criterion, that is, the ratio of between-class scatter and intra-class scatter. The difference between is expanded in the space after dimension reduction, and the difference in the face sample is reduced in the space after dimension reduction.

In 2014, Saining Xie and others launched Aggregated Residual Transformations for Deep Neural Networks [3] and proposed that deep learning to recognize faces needs to use two supervisory signals of face recognition and verification signals at the same time to learn faces, from each face area and resolutions of the face. After PCA dimensionality reduction concatenated to form the final DeepID2 feature. Afterward, it is done to verify that the image belongs to that identity. The proposal of this method lays the foundation for the application of deep learning in face recognition.

Resnet was proposed by Kaiming He et al. [4] in 2015. In this paper, a 152-layer residual network is used. By introducing a residual block, learning the residual from input to output to train a very deep neural network, and gradually adjusting the network by learning the residual. weight to solve the problem of gradient disappearance.

3. Method

The ResNet50 framework employs the ResNet50 algorithm as its foundational structure. This architecture encompasses several essential components, including the initial input layer, convolutional layers, distinctive residual blocks, pooling layers, fully connected layers, and an activation function. ResNet50 draws inspiration from VGGNet but introduces a groundbreaking concept known as a "residual block" [5]. These blocks consist of dual convolutional layers, which introduce skip connections. These connections enable specific convolutional layers in the primary pathway to be bypassed. By doing so, the residual path captures crucial features that could otherwise be lost in the main pathway. Each residual block is composed of multiple residual units, with each unit housing two convolutional layers. The outcome of the second convolutional layer is combined with the input of the residual block. This mechanism effectively preserves the original information, counteracts the vanishing

gradient problem encountered during training, and bolsters the network's depth. It's important to note that ResNet50 boasts a relatively lower complexity and employs fewer filters when compared to the VGG architecture [6].

ResNet50 is a profound neural network not only containing 49 convolutional layers but also a fully connected layer, structured into five stages, and each stage encompasses numerous residual blocks [7]. The process of convolution involves the movement of a convolutional kernel across the input data. This action generates an output feature map through the computation of a weighted sum involving the kernel and input information. The kernel size initially measures 7x7, followed by a 1x1 kernel in the subsequent convolutional layer. Multiple convolutional layers make use of a 3x3 kernel.

The output of the residual block merges with the input and undergoes normalization through batch normalization, enhancing training stability. Commencing with input image conversion, the process involves matrix manipulation, channeling this matrix into the ResNet50 network. Convolution operations transpire within convolutional layers, allowing the network to learn object features embedded within the image. Outcomes from convolutional layers flow into the pooling layer, where downsampling diminishes feature map dimensions.

Progressing through five stages, the image arrives at the fully connected layer, yielding 1000 outputs, correlating to classes in the ImageNet dataset. Downsampling bolsters output robustness by decreasing dimensions.

In essence, ResNet50 leverages residual blocks to facilitate effective feature learning and skip connections, ameliorating issues such as vanishing gradients. The architecture's strategic convolutional layers, pooling operations, and fully connected layers collectively amplify precise and robust image [8] recognition. The pooling operation's outcome is subsequently fed into the fully connected layer, wherein it's converted into a vector. The vector's elements denote the probabilities of images belonging to diverse classes. The classifier then uses this output for image classification. The cross-entropy loss function and Adam optimizer play an important role in training. The Adam optimizer, with its adaptive learning rate adjustment capability, significantly promotes the improvement of convergence speed and training efficiency. The cross-entropy loss quantifies how well the model's prediction accuracy compares to the true labels, with lower losses reflecting smaller differences between predictions and true values. During training, optimization algorithms like gradient descent adjust model parameters to minimize cross-entropy loss, bolstering classification performance. Activation functions, nonlinear within neural networks, introduce complex expressive abilities. The training spanned 50 epochs, with Adam Optimizer dynamically adjusting learning rates [9] to expedite training and evaluating the gap between predicted and actual values.

4. Results and analysis

This part showcases the outcomes of the face recognition system using the ResNet-50 model on our dataset, encompassing model accuracy on the test set and other pertinent performance metrics. The experimentation process utilizes OpenCV for image acquisition, specifically to capture face images and resize them to 224x224 pixels. In each program execution, 200 photographs of a person are captured. For portrait acquisition, the Haar feature classifier is employed, adept at detecting human facial features like eyes, lips, and their geometric relationships. This classifier interprets the image as a pixel matrix, detecting regions with distinct gray values beyond a predefined threshold. In this context, it identifies the facial region of interest.

Post-image acquisition, random alterations are applied to the photo's brightness, along with horizontal mirroring, enriching image complexity [10]. Once the images are gathered, they are organized into folders with corresponding names. For dataset creation, the 200 images per person are divided into three segments: randomly choose 120 images of 200 images to train, choose 40 random images to validate, and 40 to test.

The test set's model performance yields the subsequent results:

Model accuracy refers to the ratio between the number of samples correctly predicted by the model and the total number of predicted pictures, which can be used to evaluate the ability of the ResNet50

model to distinguish different categories in face recognition tasks. Table 1 shows the loss value and test accuracy value under different parameters. When the model is trained under the conditions of epoch 50 and batch size 16, the test accuracy is the highest when the learning rate is 0.002 and 0.001, the value is 98.75%, and the loss is 0.0415 and 0.1137 respectively. When the batch size is 32 and the learning rate is 0.001, the minimum test accuracy is 88.75%, and the maximum loss is 0.4050. In the data in this figure, when the epoch is 50 and the batch size is constant, the analysis and training of the model are less sensitive to the learning rate. The model shows the best performance when trained with a batch size of 16 and a learning rate of 0.001 or 0.002.

Table 1. The loss value and test accuracy value under different parameters.

Epoch	Batch size	Learning rate	Test Accuracy	Loss
50	10	0.001	0.9625	0.3904
50	16	0.001	0.9875	0.1137
50	32	0.001	0.8875	0.4050
50	10	0.002	0.9792	0.0481
50	16	0.002	0.9875	0.0415
50	32	0.002	0.8925	0.4019

When the epoch is 50, the validation accuracy is not stable in the tests with batch sizes of 10, 20, and 32, and learning rates of 0.001 and 0.002, and there are large fluctuations, but the final test accuracy can be maintained. Above 88.00%. This demonstrates the powerful capability of ResNet-50 on face image classification tasks.

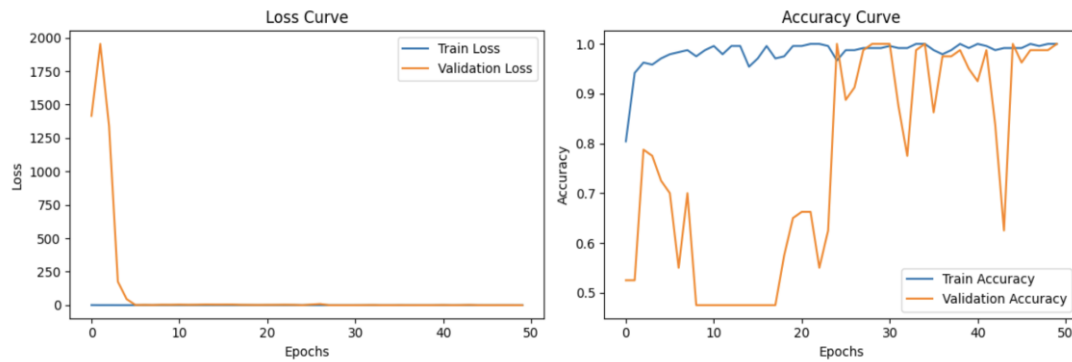


Figure 1. The loss curve and accuracy curve of learning rate=0.001, epoch=50, batch size=16.

As shown in Figure 1, the validation loss curve rises first because the model is trying to adjust the parameters to improve the face recognition ability of the model. Then the validation loss drops sharply, indicating that the model starts to gradually learn the patterns and characteristics of the data from a randomly initialized state.

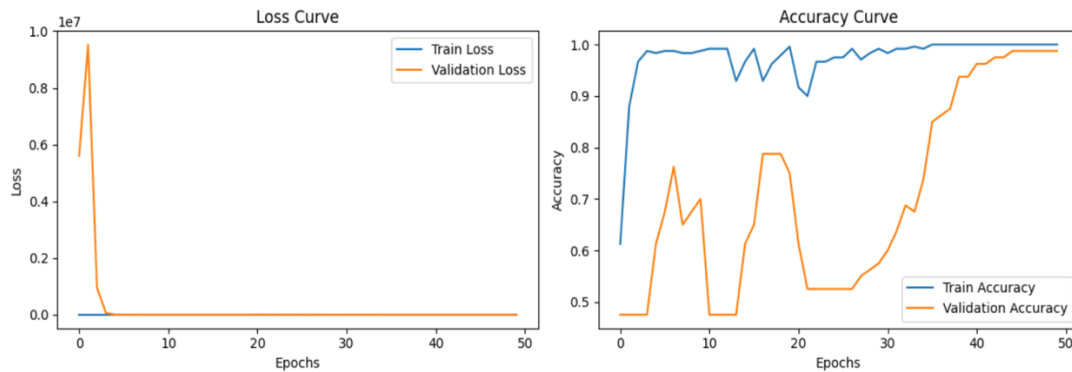


Figure 2. The loss curve and accuracy curve of learning rate=0.002, epochs = 50, batch size=16.

As shown in Figure 2, the loss curve shows a horizontal straight line after the extreme drop stagnates, and tends to overfit, but the validation accuracy curve shows a stable upward trend again after a large fluctuation, and at the end Achieved an accuracy rate of 98.75%. This further demonstrates the powerful capability of ResNet-50 on face image classification tasks.

5. Conclusion

The research embarked on designing a deep learning model for face recognition, anchored on the ResNet-50 deep convolutional neural network architecture. Leveraging OpenCV, the dataset was acquired, followed by model training and evaluation, unraveling the architectural potential in the realm of face recognition. Findings underscore ResNet-50's commendable performance, marked by high accuracy and precise facial recognition. Through the fusion of deep convolutional networks and residual connections, the model adeptly acquires intricate, abstract feature representations. This feature augmentation yields substantial performance enhancements in recognizing facial images across varying brightness conditions. Notably, the model's high accuracy showcases its resilience and trustworthiness in identifying facial images. In summation, the ResNet-50-based face recognition model stands out in this investigation, exhibiting promising outcomes applicable to real-world face recognition requirements. As the demand for face recognition escalates, future endeavors should delve into augmenting the model's efficiency, adaptability, and capacity to address practical challenges that may arise.

References

- [1] Alex P and Matthew T 1991 Eigenfaces for Recognition. *J Cogn Neurosci.* 3 (1) (1991) 71–86.
- [2] Peter N B, Joao P H and David J K 1997 Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* vol. 19, no. 7(1997) 711-720
- [3] Saining X, Ross G, Piotr D, Zhuowen T and Kaiming H 2017 Aggregated Residual Transformations for Deep Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* p1492-1500
- [4] Kaiming H, Xiangyu Z, Shaoqing R and Jian S 2016 Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* p770-778
- [5] Haque M F, Lim H Y and Kang D S 2019 Object Detection Based on VGG with ResNet Network. *2019 International Conference on Electronics, Information, and Communication (ICEIC)* p1-3
- [6] Quach L D, Pham-Quoc N, Tran D C and Fadzil H M 2020 Identification of Chicken Diseases Using VGGNet and ResNet Models. *Networks and Intelligent Systems. Industrial Networks and Intelligent Systems 2020(INISCOM)* p259–269
- [7] Alex K, Ilya S and Geoffrey E H 2017 ImageNet classification with deep convolutional neural networks. *Communications of the ACM.* 60.6 (2017) 84-90.

- [8] Hu G, Yang Y, Yi D, et al. 2015 When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. Proceedings of the IEEE International Conference on Computer Vision workshops. (2015) 142-150.
- [9] Zhang Z 2018 Improved Adam Optimizer for Deep Neural Networks. 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)p1-2
- [10] Zhao W, Chellappa R, Phillips P J, et al. 2003 Face recognition: A literature survey. ACM computing surveys (CSUR). 35(4) (2003) 399-458.