International Journal of Science and Engineering Applications
Volume 13-Issue 02, 12 - 16, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1302.1004

Design of Automatic Photo Archiving System Based on Deep Learning

Yongqi Gong
School of Electric Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract:In order to solve the problems encountered in the process of manually organizing photos nowadays, a deep learning-based automatic photo archiving system is designed and implemented. In the system, the deep learning-based face detection method MTCNN and face recognition method FaceNet are adopted, in which convolutional neural network is used to realize the detection, feature extraction and recognition of faces in photos. The automatic photo archiving system can provide an intuitive user interface and convenient operation, and users can realize the archiving management of photos through simple interaction. In the process of implementation, compared with the traditional face recognition method, the deep learning-based face recognition method will improve the accuracy of face recognition and help users organize photos more accurately.

Keywords: Deep Learning, Convolutional Neural Networks, MTCNN, FaceNet, Automatic Photo Archiving

1. INTRODUCTION

Traditional face recognition algorithms usually use local features to match and recognize faces, but this method has limitations, especially in the light, posture, expression and other changes encountered in practical applications, the recognition accuracy is easily reduced [1].

With the continuous development of digital technology, researchers have proposed a face recognition method based on deep learning. Compared with the traditional face recognition method, the deep learning based face recognition method will use the deep learning model, which can automatically learn the features, and use the multilayer network structure to represent the data in a high-dimensional abstraction, which is able to better extract the feature information of the face in the picture, and the method has a better adaptive ability and generalization ability in the process of recognition, and is able to more accurately recognize the face in different lighting, angles, expression and other complex situations of the face, so that better accuracy and robustness can be obtained [2].

Through the deep learning-based face recognition technology, it can automatically recognize the face in the photo and give the name of the person corresponding to the face, and the automatic photo archiving system will intelligently classify and organize the photos according to the results. This system aims to provide users with a convenient and personalized digital photo management solution.

2. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural network (CNN) is a deep learning algorithm specifically designed to handle tasks involving 2D or 3D data such as images and videos [3]. Its architecture is inspired by the human visual system, and it extracts and learns features from raw data in a hierarchical manner, allowing it to efficiently perform tasks such as image classification, target detection, or regression.

In recent years, deep learning has achieved wide application in the field of image recognition and classification. Convolutional neural networkwas first proposed to solve the image recognition problem, and the LeNet [4] network model in 1994 laid the foundation of convolutional neural networks.In 2012, Hinton and his student's AlexNet [5] won the ILSVRC Challenge, which popularized convolutional neural networks to a deeper and wider extent.GoogLeNet [6] in 2014 introduced the Inception module to utilize resources more efficiently.In 2015, ResNet [7] proposed by Kaiming He's team achieved the best result in the ILSVRC 2015 challenge. In the field of face recognition, deep learning-based techniques have made significant progress, and models such as FaceNet [8], DeepFace [9], and VGGFace [10] have achieved impressive results in face detection, recognition, and verification. Meanwhile, researchers focus on the robustness and generalization ability of deep learning models, and propose an adversarial learning-based approach to make the models more robust and able to handle complex situations such as occlusion and illumination [11].

3.METHODOLOGY

3.1 MTCNN

MTCNN adopts a cascade approach and consists of three networks, namely P-Net, R-Net and O-Net.Each network is responsible for a different task, including the generation of face candidate frames, the correction of candidate frames, and the localization of keypoints of the face.

P-Net is the first network of MTCNN, which is mainly responsible for generating a large number of candidate frames that may contain faces. It is a convolutional neural network that extracts image features through convolution and pooling operations and outputs whether the candidate box at each position contains a face and the corresponding corrected bounding box.

R-Net further filters the candidate frames on the basis of P-Net and corrects them to improve the accuracy of detection. Similar to P-Net, R-Net is also a convolutional neural network, which improves the detection accuracy and reduces the false detection rate through further convolutional network operations.

O-Net is the last network of MTCNN and is responsible for final face detection and keypoint localization. It further filters the candidate frames for finer corrections and outputs the key point locations of the face, such as eyes, nose and mouth.

The design of MTCNN allows the whole model to adapt to faces of different scales and angles, making it more robust in complex scenes [12]. The advantage of this cascade structure is that the output of each network is used as the input of the next network, which helps to gradually refine the detection results and improve the overall performance.

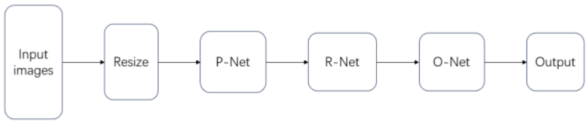

Figure. 1 MTCNN flowchart

3.2 FaceNet

FaceNet uses a deep neural network to map face images into a high-dimensional feature space and trains the network with a triplet loss function. Specifically, for the input of three images, two of them are different photos of the same person and the other is a photo of a different person. By training the network, the two photos of the same person are made to be closer in the feature space, while the distance of the different person is farther away, so that face recognition can be performed by comparing the distances of the face feature vectors in practical applications.

In addition, the feature extraction module of FaceNet adopts the Inception module in convolutional neural network and uses techniques such as Batch Normalization and Dropout to further improve the generalization ability of the model [13].

Deep Convolutional Neural Networks are used in Facenet, which means that face images are processed by deep convolutional neural networks to map the image to a high-dimensional feature space and learn to represent face features in this space. The model tunes the model parameters and trains the neural network through a ternary loss function. Figure 2 shows the flowchart of FaceNet algorithm model, first divide the training set and validation set to get the face training data, then predict the data from the training set by deep convolutional neural network to get the corresponding face feature vector, after L2 normalization, the parameters of the neural network are updated by using the ternary loss function, so that the value of the loss function decreases gradually, and finally use the validation set to verify the trained network using the validation set to observe the accuracy of the prediction results.

Figure 3 shows the training process of Triplet Loss. In the training set, one sample is randomly selected as Anchor (a), and then one sample is randomly selected from the similar samples as Positive (p), as well as one sample is randomly selected from the non-similar samples as Negative (n). Thus the (a, p, n) triad forms the basic unit in training.

International Journal of Science and Engineering Applications
Volume 13-Issue 02, 12 - 16, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1302.1004

The training goal of Triplet Loss is to make samples of the sameclass closer together in the feature space and samples of different classes farther apart by learning the model parameters. Specifically, the goal of the training process is to minimize thedistance between Anchor and Positive while maximizing the distance between Anchor and Negative.

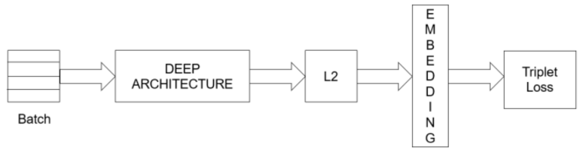

Figure. 2 FaceNetflowchart

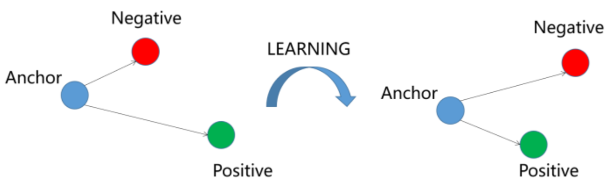

Figure. 3 Triplet Loss training

4. FACE RECOGNITION

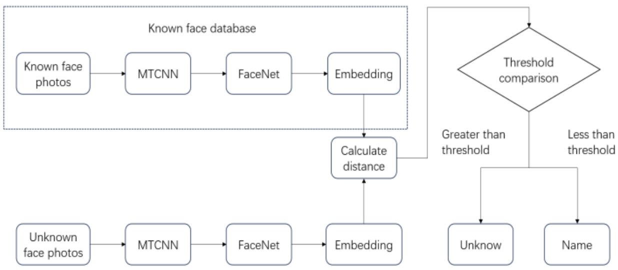

Figure. 4 Face recognitionflowchart

4.1 Face Recognition Process

Before face recognition, it is necessary to build a database of known faces, so that in face recognition can determine whether the face to be recognized is known.After completing the construction of the database of known faces, the face recognition task can be carried out. In this process, one or more photos of faces to be recognized are selected and input to the face recognition system one by one. Similar to the steps of constructing the known face database, the input photos are also processed by MTCNN and FaceNet to obtain the face feature vectors in the photos to be recognized.

Next, the obtained face feature vectors are computed with the feature vectors in the database of known faces to obtain the distance between the face to be recognized and each feature vector in the database. From these distances, a minimum value is selected and compared with a pre-set threshold value. If the minimum distance is less than the threshold, the system determines that the face to be recognized exists in the known face database, and then outputs the corresponding recognition results. On the contrary, if the minimum distance is greater than the threshold, the system determines that the face is unknown and outputs "unknown".

4.2 Face Recognition Results

Face recognition code part of the running results shown in Figure 5, to a photo to be detected as an example of the running results, the photo through the face detection to get three faces, the three lists in the figure represents the three faces to be detected and the distance of all the faces in the database, the number below each list represents the minimum distance value in the list, the minimum distance obtained and the threshold for comparison, the first two distance values less than the threshold, then The first two distance values are less than the threshold value, then the recognition result is output, and the third is greater than the threshold value output "unknow".

```
[0.9115743, 0.96622884, 0.94275856, 0.8353409
0.6004128
[0.75514907, 0.7784362, 0.7548879, 0.76242274
0.6048861
[0.8210844, 0.87317175, 0.8596094, 0.84547967
0.75344217
['Wang*Zhe', 'Zhang*Ming', 'unknow']
```

Figure. 5 Running results of face recognition (part)

5. AUTOMATIC PHOTO ARCHIVING SYSTEM

5.1 One-click Recognition

There is a default photo recognition directory and save path in this system, if you run the one-key recognition function directly without modifying any path, you can directly recognize the photos in the system's default recognition directory and save the photos to the default folder.

5.2 IdentifyDirectory Selection and Individual Photo Selection

These two functions can independently select the objects to be recognized according to the user's actual needs, either photos in a specified directory or one or more photos for face recognition, and if there is an error in the user's choice of path, it will be prompted at the bottom of the interface.

International Journal of Science and Engineering Applications
Volume 13-Issue 02, 12 - 16, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1302.1004

5.3 File saving location change

In the process of photo recognition, if users need to save photos to a specified location, they can use this function to change the photo saving path, and after changing the path and then performing face recognition, they can find the recognized photos in the specified saving location.

5.4 Create Face Directory and FileOrganization

The function of creating a face directory is to create a separate directory for each person in the recognition result based on the completion of face recognition, and the function of document organization is to file the corresponding face photos under the directory of the corresponding person's name. Creating a face directory is a prerequisite for document sorting, if the corresponding face directory is not created before document sorting, no file will appear.

5.5 Compression Path Selection and File Compression

Compression path selection allows users to choose the compression path according to their own needs, if there is no choice, the system will choose the default compression path, document compression is the result of document organization for packaging and compression, for each person to generate a separate compressed package, compressed package to the corresponding person's name, compressed package includes the corresponding person's photo.

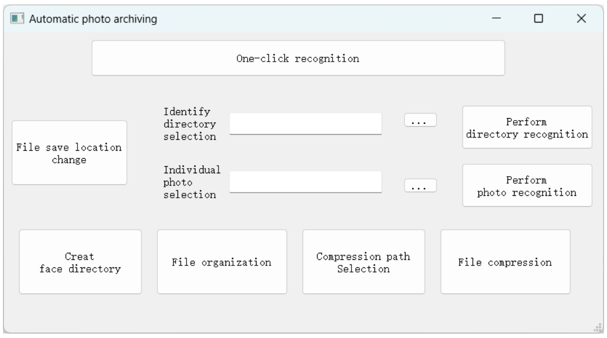

Figure. 6 Interface of the automatic photo archiving system

6. CONCLUSION

This paper implements a deep learning-based automatic photo archiving system, aiming at solving many challenges in digital photo management and providing users with an intelligent and efficient photo organizing solution.MTCNN implements face detection in photos, and FaceNet implements accurate recognition and feature extraction of faces in photos, providing users with a one-click intelligent recognition function that reduces the difficulty of photo organization. FaceNet is a powerful tool for recognizing faces in photos. Users can quickly and intuitively manage a huge digital photo library through the system, making photo management no longer a tedious task. However, the system still has some potential room for improvement, such as challenges in dealing with face recognition under occlusion, low-light conditions, etc. Future research can further optimize the algorithm to enhance the robustness and applicability of the system.

7. REFERENCES

[1] Zhu, Tiancai, & Zhou, Xiaobo. (2023). Research Overview of Face Recognition Methods Based on Deep Learning.Modern Computer, 29(17), 36-40.

[2] Lu, Hongtao, & Zhang, Qinchuan. (2016). A Comprehensive Review of the Application of Deep Convolutional Neural Networks in Computer Vision. Data Acquisition and Processing, 31(01), 1-17.

[3] Gai, Rongli, Cai, Jianrong, Wang, Shiyu, Cang, Yan, & Chen, Na. (2021). A Comprehensive Review of the Application of Convolutional Neural Networks in Image Recognition. Journal of Small and Microcomputers Systems, 42(09), 1980-1984.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 25.

[5] Girshick, R., Donahue, J., Darrell, T., et al. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580-587.

[6] He, K., Zhang, X., Ren, S., et al. (2016). Deep Residual Learning for Image Recognition. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[7] Elman, J. L. (1990). Finding Structure in Time. Cognitive Science, 14(2), 179-211.

[8] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition, 815-823.

[9] Taigman, Y., Yang, M., Ranzato, M., et al. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 1701-1708.

[10] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In Proceedings of the 26th BMVC, 1(3), 6.

[11] Sui, Chenhong, Wang, Ao, Zhou, Shengwen, et al. (2023). A Comprehensive Review of Adversarial Training Techniques for Robust Learning. Journal of Image and Graphics, China, 28(12), 3629-3650.

[12] Li, Zhihua, Zhang, Jianyu, & Wei, Zhongcheng. (2022). Design of a Face Recognition System Based on MTCNN and FaceNet. Modern Electronic Technology, 45(04), 139-143.

[13] Qi, Guanghua, & He, Mingxiang. (2020). Image Classification Method Using Convolutional Neural Network with Inception Module. Software Guide, 19(03), 79-82.