Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2024, 11(10):54-58

Research Article ISSN: 2394 - 658X

Recognition of Faces in Images Using MT-CNN and Facenet Networks on Various Image Datasets

Phong Bui Hai[1]

Hanoi Architectural University Hanoi, Vietnam
Email: phongbh@hau.edu.vn

ABSTRACT

The recognition of face plays an important role in our communications every day. In recent years, with the advancements of image processing and artificial intelligence (AI), the automatic recognition of faces has attracted more and more researches. The paper presents the recognition of facial emotion using the MT-CNN and Facenet networks. Firstly, The MT-CNN network is applied to detect faces in images. Then, the Facenet is applied to recognize faces in images. We have evaluated the proposed method on public datasets of face images that are Yale and Vietnamese student Face Datasets. The obtained recognition accuracies of 90% and 94% on Yale and Vietnamese student Face Datasets have shown the effectiveness of the proposed method.

Keywords: Face recognition, Deep Neural Network, CNN

INTRODUCTION

In today's technological era, facial recognition technology has become indispensable in many fields including security, entertainment, healthcare, and business. The automatic recognition of faces aims to recognize faces from input images based on the analysis of collected data without applying human knowledge. The task has been well researched in recent years. The machine learning and deep learning techniques are applied to solve the task [1]. The recognition accuracy of the task is improved however, there are still errors of the recognition. Therefore, the task needs to be further investigated.

There exist many difficulties of the recognition of faces in images:

(1) The face images may be blur or shadow that caused the difficulties of the recognition.
(2) The similarity between faces may cause the errors of the recognition.
(3) The complex background of images may cause errors of the recognition.

Figure 1 demonstrates various kind of faces. One of the efficient solutions to improve the recognition accuracy of the faces is the using of multiple deep learning models. To address these challenges and enhance the security of facial recognition systems, a multi-layer security method requiring users to change their facial expressions and then recognize and verify these expressions using facial landmarks has been proposed. Facial landmarks are distinctive points on the face such as the eyes, nose, and mouth. Recognizing and matching these expressions can help improve the system's security. Combined with MT-CNN, this method promises to deliver higher accuracy and performance in facial recognition under diverse environmental conditions, ensuring that users must be physically present and perform specific authentication actions to be verified, rather than just using static images or "masks" to deceive the system.

Multi-task Cascaded Convolutional Networks (MT-CNN) [2] are recognized and widely used as one of the most advanced methods in facial recognition. MT-CNN and Facenet improve the performance of facial recognition processes and autonomously learn and extract features from images, providing convenience and flexibility in applications. Nevertheless, there are still some vulnerabilities that the system has to face. The MT-CNN plays an important role to detect faces from images. The Facenet [3] is a powerful network to recognize detected faces. The proposed method has been evaluated on various datasets. We collect and test the method on public and Vietnamese face datasets.

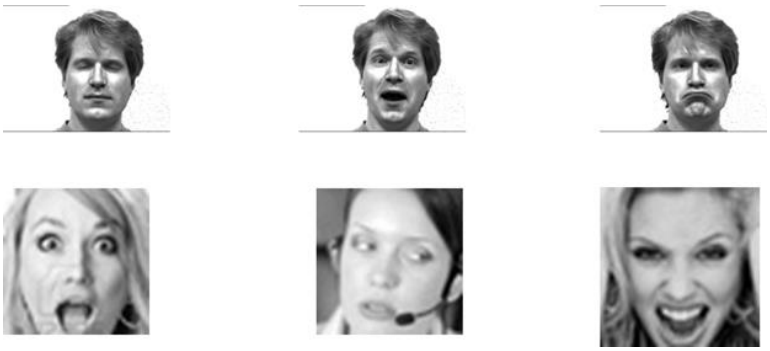


Fig.1 *Examples of various kind of face in images.*

RELATED WORK

This section reviews existed approaches for the recognition of faces in images. Traditional methods used the extracted features and machine learning classifiers to classify leaf images [4]. Modern methods applied the convolutional neural networks (CNN) [10] to perform the task.

Recognition of face images using handcrafted feature extraction and machine learning classifiers

Existed methods of face recognition attempted to extracted low level visual features of images. The research in [5] proposes the wavelet transformation and local binary patterns for the classification of images. The work in [6] extracts text features to recognize face images. The work in [7] applies image processing to recognize face in gray-scale images. The work focuses on the application of local binary patterns and nonparametric discrimination to recognize images. The work in [8] proposes a modeling of the local binary patterns operator for texture recognition. The work in [9] proposes the hybrid method that use the convolutional network and Restricted Boltzmann Machine models to recognize faces. The traditional methods obtain the low complexity when performing the face recognition. However, the accuracy of the methods is not high.

Recognition of face images using deep neural networks (DNNs).

The DNNs have been applied and fine-tuned to improve the accuracy of the face recognition. The work in [10] uses deep neural networks to learn effective feature representations to solve the face recognition. The work in [11] designs a deep learning models to recognize faces in large datasets. The work obtained the accuracy of 99.15%. The recent work in [12] applies the data augmentation and the CNN to recognize faces in real environment. The work in [13] and [14] propose the data synthesization and augmentation to obtain large number images of faces to train and test DNNs efficiently.

PROPOSED METHOD

The overall steps of the face recognition is described in Fig.2. Firstly, we collect and prepare face images from various datasets. Then, the input images are processed and normalized using image processing techniques. After that, face regions in the images are detected using the MT-CNN network. The Facenet network is fine-tuned to improve the performance of the face recognition.

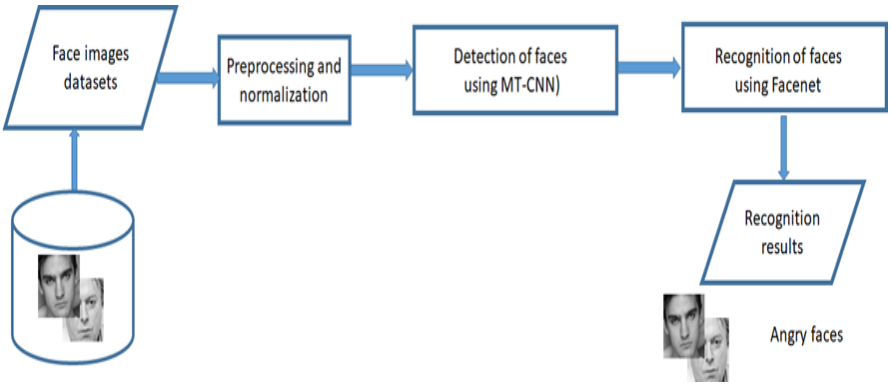


Fig. 2 *Overall steps of the proposed method for the classification of leaf images.*

Image normalization based on image processing

One of the difficulties of the face recognition is the diversity of image datasets. Therefore, to solve the issue, we have collected data from public datasets. Moreover, we have prepared our private datasets from Vietnamese students. Fig. 3 illustrates various face images from public and private datasets. The collected images are normalized to the size of 224x224x3 for the training and testing of DNNs.

Fig. 3 *Examples of face images in public (a) and Vietnamese (b) datasets.*

The detection of faces using the MT-CNN network

In the step, we applied the MT-CNN network to detect faces in images. Fig. 4 shows the structure of the network. Basically, the network consists of the following sub-networks: Proposal Network (P-Net), Refine Network (R-Net), and O-Net [2].

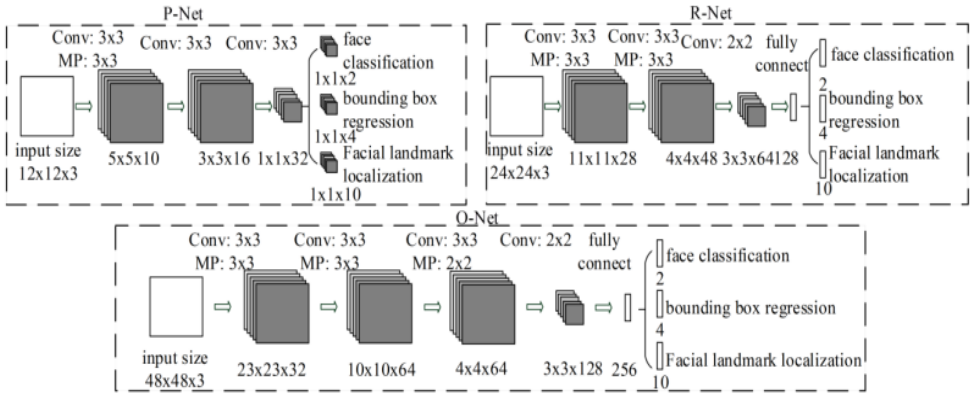


Fig. 4 *Structure of MT-CNN network. "MP" is the max pooling and "Conv" is convolution layer [2].*

The recognition of faces using the Facenet network

In the step, the detected faces from input images are recognized using the facenet network [3]. Fig. 5 shows the structure of the facenet network. The network consists of the following components:

- A batch input layer
- A deep CNN
- L2 normalization layer
- Face embedding.
- Triplet loss

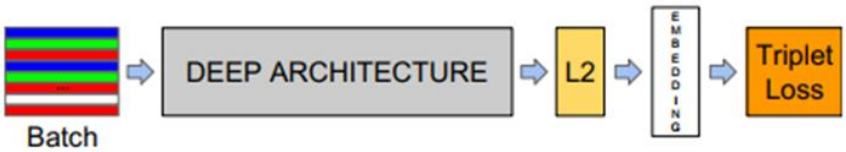


Fig. 5 *Structure of the facenet network [3]*

The method is implemented on the environment of a computer with 8 GB RAM, GPU and tensorflow library in Python 3.7 language.

EXPERIMENTAL RESULTS

Datasets and evaluation metrics

We have evaluated our method on two datasets. The first one is the Yale face dataset. The dataset consists of 165 images. The images are labeled: happy, angry, sad, surprising and normal. The second one is the Vietnamese student dataset. The dataset consists of 770 images of students.

To obtain the clear performance evaluation of the proposed method, we applied the Precision, Recall and F1 score metrics [15]. The metrics are widely adopted for the classification task.

Table 1: Performance comparison of the face recognition in Yale and Vietnamese datasets.

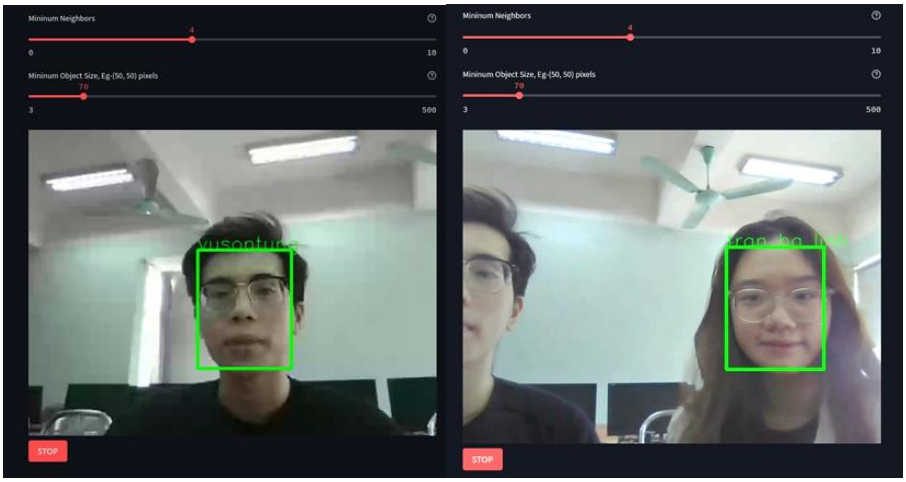
Dataset	P	R	F1 score
Yale	90%	88%	88.99%
Vietnamese dataset	94%	92%	92.99%

Table 2: Performance comparison of the face recognition in Yale dataset

Dataset	P	R	F1 score
Yee et al. [16]	85%	84%	84.50%
Ravi et al. [17]	75%	74%	74.50%
Our method	90%	88%	88.99%

Performance evaluation

Performance comparison of the recognition of faces in images on the two datasets is shown in Table 1. The performance of the proposed method on Vietnamese student dataset is higher than that on the Yale dataset. The Vietnamese student dataset is larger and clearer compared to the Yale dataset. We compare the performance of the recognition of faces with the existed methods [16] and [17]. The work in [16] proposes the extraction of A Laplacian Completed Local Ternary Pattern features of images to recognize faces. The work in [17] applies the Local binary patterns feature extraction and CNN to recognize face images. The performance comparison of the proposed and other methods are shown in Table 2. The use of the DNNs helps to improve the performance of the proposed method compared to existed approaches.



(a) (b)

Fig. 6 *Examples of the recognition of face images (a) Recognition of face in images, (b) Recognition of faces in images from complex background*

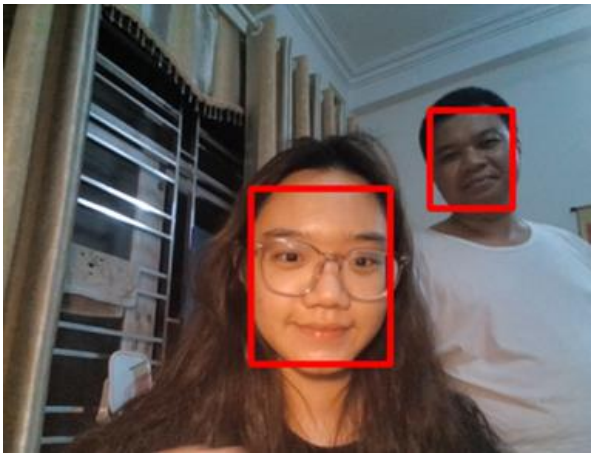


Fig. 7 *Examples of the detection of many face images*

Fig. 6 and 7 demonstrate the recognition and detection of the faces in images. The demonstration shows that our proposed method works correctly for various images. The proposed method can detect and recognize faces in different of captured images in reality.

CONCLUSION

The paper has presented the use of MT-CNN and Facenet to recognize faces in images. The MT-CNN allows to detect face regions efficiently. After that the Facenet recognize detected faces with high performance. The training and testing of the proposed method on various datasets have shown the effectiveness of the proposed method. Obtained recognition accuracies of 90 % and 94% on public and Vietnamese datasets show the high performance of the proposed method. The comparison shows that the DNNs obtain better results compared to traditional methods using feature extraction. In the future, the results can be applied to develop the security systems to support end users in real applications.

REFERENCES

[1]. L.Li et al. A Review of Face Recognition Technology. IEEE Access, 2016, vol.4.
[2]. Khan, S. MTCNN++: A CNN-based face detection algorithm inspired by MTCNN. The Visual Computer, 2023, vol. 40, doi: 10.1007/s00371-023-02822-0.
[3]. Florian Schroff et al.FaceNet: A Unified Embedding for Face Recognition. https://arxiv.org/abs/1503.03832, 2015.
[4]. Lyons MJ (1999). Automatic classification of single facial images. IEEE Trans Pattern Anal Mach Intell 21(12):1357–1362.
[5]. Doost HE, Amirani M (2013) Texture classification with local binary pattern based on continues wavelet transformation. Int J Adv Res Electr Electron Instrum Eng 2(10):4651–4656
[6]. Tan X, Triggs B (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650
[7]. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987.
[8]. Guo Z, Zhang L, Zhang D (2010) A completed modeling of local binary pattern operator for texture classification. IEEE Trans Image Process 19(6):1657–1663.
[9]. Rassem TH, Makbol NM, Yee SY (2017) Face recognition using completed local ternary pattern (CLTP) texture descriptor. Int J Electr Comput Eng 7(3):1594–1601.
[10]. Sun, Y.; Wang, X.; and Tang, X. (2013). Hybrid Deep Learning for Face Verification. In IEEE International Conference on Computer Vision.
[11]. Sun, Y.; Wang, X.; and Tang, X. 2014. Deep Learning Face Representation by Joint Identification-Verification. Advances in neural information processing systems 27.
[12]. Ben Fredj, H., Bouguezzi, S., Souani, C.: Face recognition in unconstrained environment with CNN. Visual Comput. 37(2), 217–226 (2021).
[13]. Masi, I., Trần, A.T., Hassner, T., Leksut, J.T., Medioni, G.: Do we really need to collect millions of faces for effective face recognition? In: European Conference on Computer Vision, pp. 579–596. Springer, Cham (2016)
[14]. Leng, B., Yu, K., Jingyan, Q.I.N.: Data augmentation for unbalanced face recognition training sets. Neurocomputing 235, 10–14 (2017).
[15]. Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. arXiv:1409.4842, 2014.
[16]. Yee, S. Y., Rassem, T. H., Mohammed, M. F., & Awang, S. (2020). Face recognition using Laplacian completed local ternary pattern (LapCLTP). In Advances in electronics engineering, 2020.
[17]. Ravi, R., & Yadhukrishna, S. V. (2020, March). A face expression recognition using CNN & LBP. Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020.