Original Research Article:
full paper

(2023), «EUREKA: Physics and Engineering»
Number 4

COMPARATIVE STUDY ON THE PERFORMANCE OF FACE RECOGNITION ALGORITHMS

Truong Van Nguyen✉
Department of Mechatronics Engineering[1]
nguyenvantruong@haui.edu.vn

Tuan Duc Chu
Department of Mechatronics Engineering[1]

[1]*Hanoi University of Industry*
298 Cau Dien str., Bac Tu Liem dist., Hanoi, Vietnam, 100000

✉**Corresponding author**

Abstract

Facial and object recognition are more and more applied in our life. Therefore, this field has become important to both academicians and practitioners. Face recognition systems are complex systems using features of the face to recognize. Current face recognition systems may be used to increase work efficiency in various methods, including smart homes, online banking, traffic, sports, robots, and others. With various applications like this, the number of facial recognition methods has been increasing in recent years. However, the performance of face recognition systems can be significantly affected by various factors such as lighting conditions, and different types of masks (sunglasses, scarves, hats, etc.). In this paper, a detailed comparison between face recognition techniques is exposed by listing the structure of each model, the advantages and disadvantages as well as performing experiments to demonstrate the robustness, accuracy, and complexity of each algorithm. To be detailed, let's give a performance comparison of three methods for measuring the efficacy of face recognition systems including a support vector machine (SVM), a visual geometry group with 16 layers (VGG-16), and a residual network with 50 layers (ResNet-50) in real-life settings. The efficiency of algorithms is evaluated in various environments such as normal light indoors, backlit indoors, low light indoors, natural light outdoors, and backlit outdoors. In addition, this paper also evaluates faces with hats and glasses to examine the accuracy of the methods. The experimental results indicate that the ResNet-50 has the highest accuracy to identify faces. The time to recognize is ranging from 1.1 s to 1.2 s in the normal environment.

Keywords: face recognition, support vector machine, visual geometry group, residual network, machine learning.

DOI: 10.21303/2461-4262.2023.002831

1. Introduction

Face recognition is one of the hot issues of computer vision with huge applications applied in many fields such as timekeeping, identity authentication, security monitoring, etc. It is being used more and more in recent years due to its high stability, high precision, and ease to use [1]. The issue with face recognition is system should recognize different faces with the highest speed and accuracy [2]. Besides, the face recognition system should be able to respond to a variety of variations in face photos [3]. These issues have been attracting a lot of scientists to solve them.

Face recognition has two tasks including face identification and face verification [4]. The task of matching a given face picture to one in a database of faces is known as face identification. The face identification process works after the face detection process. The next task is face verification, which is known as the task of comparing a face to another and determining if they are identical. In both face identification and face verification, face characteristics such as the eyes, nose, mouth, and chin are recognized using geometric feature-based algorithms [5]. Other factors such as areas, distances, and angles are used as face characteristics.

In the past, face recognition methods usually use handcraft methods to extract features such as histogram of oriented gradients (HOG) [6], scale-invariant feature transform (SIFT) [7], speeded-up robust features (SURF) [8], etc. However, for these handcraft methods, the extracted features neglect to train, because the algorithm cannot update the parameters. In addition, the extracted features are not linked with the classifier. In recent decades, with the development of deep learning and rapid improvement of hardware, face recognition systems are also much more efficient. The calculation time and accuracy are better than traditional methods. The current deep learning methods are built based on a convolutional neural network (CNN) [9]. At the low level, the CNN model extracts the features, which are the same as features of traditional methods extract. Since the success of AlexNet [10], there are many CNN models other have been proposed, such as VGG Net [11], GoogleNet [12], ResNet [13], DenseNet [14], and so on. These models solve many tasks such as object detection, face recognition, semantic segmentation, object tracking, etc. The training process of these models needs a lot of training data to achieve high accuracy and avoid overfitting. Despite the above neural network achieving remarkable results, how to improve the performance of these models is still a challenge.

Current face recognition systems have good results in controlled environments [15]. However, the traditional and deep learning methods are affected by viewpoint, indoor and outdoor lighting, and occlusions. Besides, traditional methods are limited by their algorithms, the number of data, etc. While deep learning methods need large and high-quality data to improve performance.

In traditional machine learning methods, to achieve good performance, the input features must be carefully selected before the algorithm is trained on the data. For example, SVM [16] is used for classification and regression analysis [16, 17]. SVM is an efficient algorithm when applying kernels to solve complex situations, allowing it to capture complex relationships between facial features. This method also reduces overfitting and is suiTable for a large number of features. Current deep learning methods are often using convolution layers to extract features and using fully connected layers for classification such as LeNet-5 [18], AlexNet [19], VGG-16 [11], GoogleNet [12], ResNet-50 [13], and DenseNet [14]. Among these deep learning methods, VGG-16 and ResNet-50 have a remarkable change when compared with other methods. The VGG-16 model is a very deep convolutional network and is the first model to apply a convolutional block [11]. In addition, VGG-16 is use kernel 3×3, which reduces the number of parameters and increases the accuracy. While ResNet-50 is also the first model that applied batch normalization [13]. The main advantage of ResNet-50 is the residual block, which reduces the parameters and avoids overfitting. Both VGG-16 and ResNet-50 need a lot of datasets and time to train the model.

In face recognition, various methods are employed, each with distinct capabilities that can handle different conditions such as variations in illumination, facial expressions, and poses. In this paper, for our purpose, which is the comparison of the performance of face recognition algorithms in different light conditions, let's focus on three methods including SVM, VGG-16, and ResNet-50. Let's perform experiments and find out a good performance algorithm among these face recognition.

2. Materials and methods

2. 1. SVM Algorithm

The SVM is a supervised algorithm and is often used in classification or regression [17]. The idea of SVM is to build a hyperplane to separate the type of data point [20]. This hyperplane divides space into different parts wherein each class lay in each separate part. There are a lot of hyperplanes, so SVM needs to find the optimal hyperplane. The optimal hyperplane needs to choose is the plane with the largest margin [21].

In face recognition systems, the features are often extracted by handcraft methods such as PCA, HOG, SIFT, etc., and then these features are passed to the SVM algorithm. The difference value of distance measurement is calculated via Euclidean distance-based methods [22]. If the two vectors are close, the features of the two face images are similar. Euclidean distance equation is shown as:

$$E_j = \sqrt{\sum_{i=1}^{k}\left(a_{ij} - b_{ij}\right)^2}, \tag{1}$$

where a_{ij} is the i-th input eigenvector, b_{ij} is the i-th eigenvector in the database, k is the dimension of the eigenvector, and j is the j-th image in the database.

The Euclidean distance method requires a lot of computation time in case of a huge database because the target face image is compared with the other face images in the database. To reduce computation time for face recognition [23], the SVM creates the hyperplanes to classify the Euclidean distance as shown in **Fig. 1**.

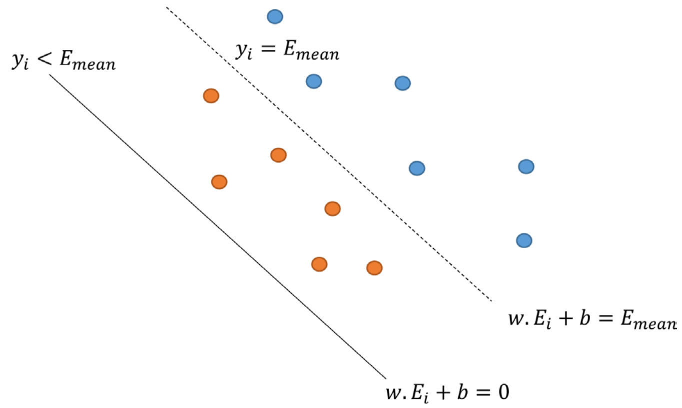


Fig. 1. Hyperplane of the Euclidean distance

In **Fig. 1**, E_i is the Euclidean distance between the target image and i-th image of the database. E_{mean} is the mean value distance between the target image and all images in the database and there are M images in the database.

The SVM is trained to learn a boundary that separates different classes of face images based on their features:

$$E_{mean} = \frac{\sum_{i=1}^{j} E_i}{j}, \tag{2}$$

where $i = 1,2,\ldots, M$ and $y_i \in E_{mean}$. Then, consider the SVM hyperplane:

$$w \cdot E_i + b \geq E_{mean}. \tag{3}$$

To find the hyperplane, the problem of quadratic optimization is solved as:

$$y_i(w \cdot E_i) - b \geq E_{mean}, \tag{4}$$

where $y_i = E_{mean}$, Lagrange is applied to obtain:

$$L(\omega, b, \alpha) = \frac{1}{2}w^2 - \sum_{i=1}^{M} \alpha_i \left[y_i (w \cdot E_i + b) - E_{mean} \right], \alpha_i \geq 0, \tag{5}$$

where α_i is Lagrange multipliers and $i = 1,2,\ldots, M$.

As the SVM algorithm does not determine the optimal solution, the optimization problem is solved [23] by the below equation. The equation is also called dual:

$$L_D = \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j E_i E_j. \tag{6}$$

If $\alpha_i \geq 0$ this means the data is the support vector and it is found on the outside boundary of the hyperplane:

$$f(d) = sgn\left[\sum_{i=1}^{j} \alpha_i y_i (E_i E) + b \right]. \tag{7}$$

These data are utilized to compute the classification and rerun it in the same method. Before terminating the program, this step would be repeated until all of the data had been plotted. The Euclidean distance closest to the target image is represented by this data.

Using SVM for face recognition, there are some advantages such as effectiveness in high-dimensional spaces. Also, when the number of dimensions is greater than the number of samples, SVM is still successful. In addition, SVM only requires small storage memory. In contrast, SVM has some disadvantages such as: When the number of features is greater than the number of samples, the performances are ineffective. The SVM achieves efficient results for small training samples. Besides, how to select the kernel is still a difficult issue for users.

2. 2. VGG-16 Algorithm

VGG-16 developed by Karen Simonyan and Andrew Zisserman is a deep convolutional network with 13 convolutional layers and 3 fully connected layers [11]. VGGNet is a modification based on AlexNet [19]. The structure of VGG-16 is shown in **Fig. 2**.

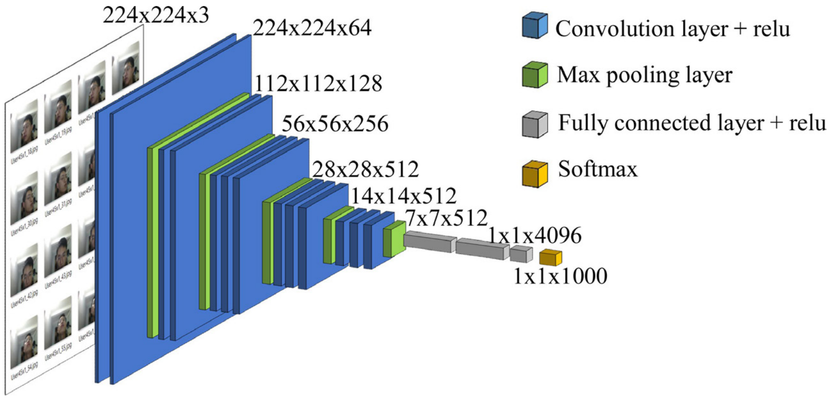


Fig. 2. VGG-16 Architecture

The inputs of VGG-16 are the images are changed to a size of 224×224. The input image is passed into 2 convolutional layers with a filter size of 3×3 and a depth of 64. Subsequently, the output is propagated to max pooling with a stride equal to 2. Then the resulting continually is propagated to convolutional layers with kernel size 3×3. These layers use 128, 256, and 512 feature maps respectively. Following these layers are max pooling with a stride of 2. The final layers are fully connected layers of 4096 nodes, followed by a softmax activation with 1000 nodes. The loss function of VGG16 is shown as:

$$Loss = \sum_{i=1}^{N}\sum_{c=1}^{M} y_{i,c} \log f_{i,c}, \tag{8}$$

where M is the number of classes, N is the number of images, and $y_{i,c}$ is the indicator that example i have label c and 0 otherwise. A smaller result means the estimated value is closer to the actual value.

The VGG-16 algorithm has the following gain: deep convolutional networks of VGG-16 improve the model of accuracy. Besides, the stacking convolutional layers improve extract features better than a layer convolutional. Using a small filter 3×3 reduced the number of parameters for the model and increase calculation speed [24]. However, VGG-16 has some disadvantages as experiments on the model VGG16 indicate that the first fully connected layer generates a great number of parameters, which increases the amount of calculation [25]. Furthermore, the small and medium-sized data samples do not perform well in the deep network due to the size limits of the dataset. The limited data scale causes an overfitting problem, which results in the unable of the model to generalize.

2. 3. ResNet-50 algorithm

This paper proposed ResNet-50 architecture that has good performance when compared to other models such as GoogLeNet, DenseNet, etc. ResNet-50 is a deep residual network, with 50 deep layers [13]. In a deep convolutional neural network, the number of stacked layers improves the model accuracy [26]. However, when the number of layers of the neural network reaches a certain threshold, the accuracy becomes saturated. It is caused by the problem of vanishing or exploding gradients. Thus, ResNet is created to solve this problem. The main different point of Resnet, when compared with other deep convolution neural networks, is the concept of «skip connection», which is the core of the residual blocks. Thanks to the skip connection, the effectiveness of ResNet is still high [27] while the number of parameters decreasing. **Fig. 3** shows the residual block.

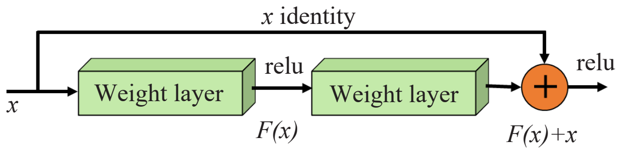


Fig. 3. Residual block

In essence, a skip connection is an identity mapping in which the preceding input of layer input is directly appended to the output of the next layer. ResNet-50 had two types of residual blocks. There are the convolution block and the identity block. **Fig. 4** shows two types of residual blocks.

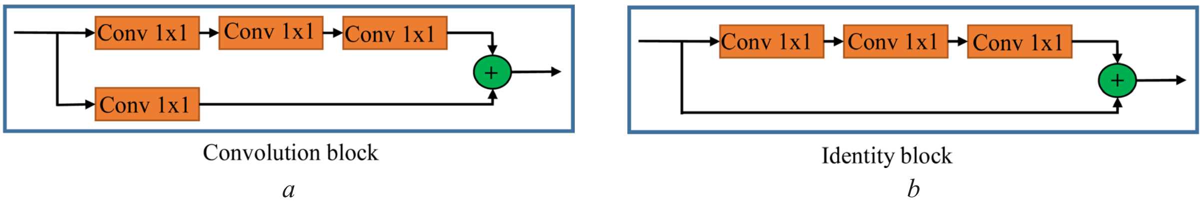


Fig. 4. Two types of residual blocks in ResNet-50: *a* – Convolution block; *b* – Identity block

Three convolution layers and a summation are the main components of the residual block [28]. The connection conducts three convolutions on x to get $F(x)$. Then, the other input of the summation is x, which is provided by a shortcut connection. The output is a combination of $F(x)$ and x:

$$H(x)=F(x)+x. \tag{9}$$

When $H(x)=x$, identity mapping is used to remove the convolution layers and reduce the depth while keeping accuracy. **Fig. 5** shows the detailed structure of ResNet-50. The loss function of ResNet-50 is given as:

$$L_{CE}=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i+b_{y_i}}}{\sum_{j=1}^{n}e^{W_j^T x_i+b_j}}, \tag{10}$$

where W is the weight matrix, b is the bias term, x_i is the training sample, y_i is the class label of the training sample, and N is the number of samples.

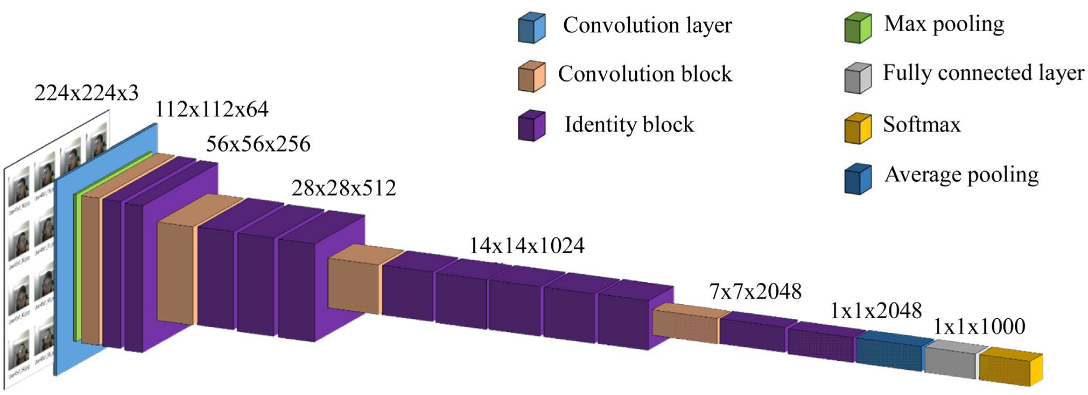


Fig. 5. ResNet-50 architecture

In a very deep neural network, increasing the number of layers for improved accuracy, but if the number of layers is above 20 layers, the model is unable to converge. This is the main reason for the vanishing gradient problem as well as the learning rate becoming so less [27]. To overcome this problem, residual learning is applied to solve the complexity of a deep neural network by using identity mapping. Besides, batch normalization also helps reduce the explosion of gradients problem. The approach of residual learning with batch normalization significantly saves training time and increases accuracy. Training networks with a high number of layers are simple and do not increase the training error percentage. However, ResNet-50 also has some disadvantages such as the complexity of architecture being increased and the implementation of batch normalization layers effect to the performance of the model since ResNet heavily depends on it.

2. 4. The structure of face recognition

In the face recognition system, both VGG-16 and ResNet-50 algorithms work as follows. First of all, face detection algorithms are applied to detect faces in input images. Next, these faces are passed into VGG-16 or ResNet-50. The output of VGG-16 or ResNet-50 is vectors, that are converted from face images. A database containing the face images is registered before. The face images in the database also are converted to vectors. Finally, similarity measure algorithms are used such as cosine similarity [29], correlation coefficient [30], and so on, to compare input vectors to vectors in the database. The output of similarity measure algorithms is called distance, if the distance is smaller, the faces are similar. **Fig. 6** shows the face recognition system architecture.

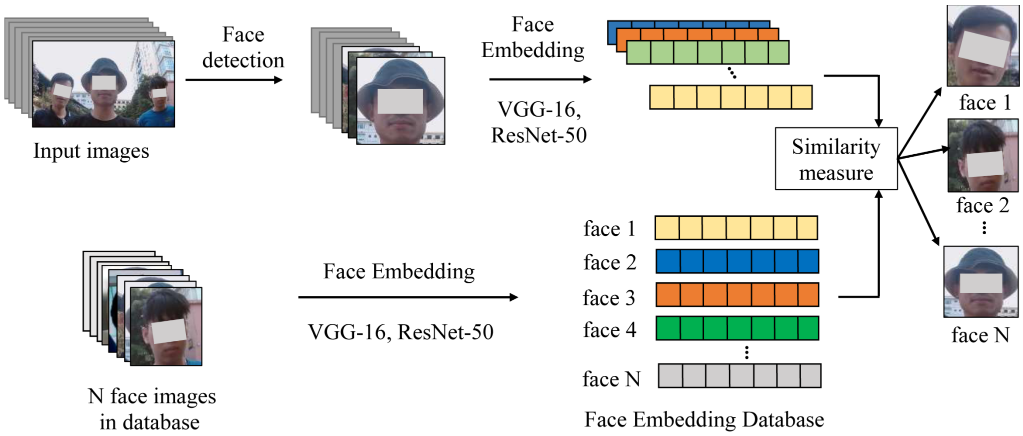


Fig. 6. Face recognition system using VGG-16 and Resnet-50

In this paper, the cosine similarity algorithm is used to measure the similarity between faces. Specifically, the cosine similarity algorithm compares the direction or orientation of the vectors, and the cosine of the angle between two vectors is used to evaluate their similarity [31]. If two vectors are compared with each other and have smaller angles between vectors, it produces larger cosine values. In other words, the larger cosine values indicating the compared vectors are more similar. Cosine similarity is defined by the dot product of vectors divided by the magnitude of the vector. The formula for cosine similarity is given below [31]:

$$simi(V,T)=\cos(\theta)=\frac{V\cdot T}{\|V\|\cdot\|T\|}, \tag{11}$$

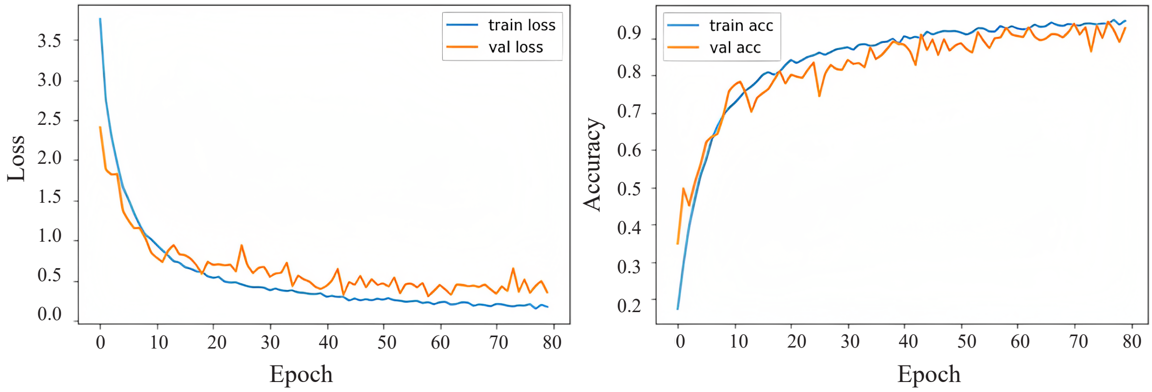
where θ is the angle between the vectors, $V\cdot T$ is the dot product V and T, and $\|V\|$, $\|T\|$ are the $L2$ norm and calculated by $V=\sqrt{V_1^2+V_2^2+\ldots+V_n^2}$.

3. Results and discussion

The SVM, VGG-16, and ResNet-50 are implemented to measure and compare the accuracy. The proposed methods are executed on a computer with Intel Core i7-4800 MQ CPU @ 2.7 GHzx8, 16GB RAM. The mini-batch stochastic gradient descent with momentum optimizer is utilized.

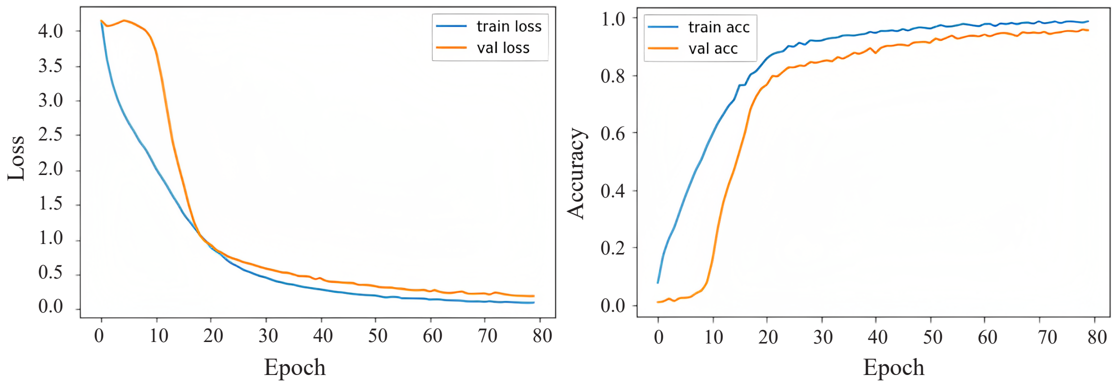
The learning rate on VGG-16 and ResNet-50 is set to 0.0001. The model is trained for 80 epochs. In the cosine similarity algorithm, the threshold value is set to 0.2.

To train the model, let's collect 2900 face pictures from 52 people. For each face, take a picture of 55 photos including changing viewpoint and illumination. In VGG-16 and ResNet-50 models, to increase the number of the dataset and save time, the data augment method is used. This research used 2300 photos for training and 600 photos for the test. This study also uses a pre-trained model to save time. Training results show that the accuracy on the validation set of SVM, VGG-16, and ResNet-50 is 88 %, 94 %, and 95.6 % respectively. **Fig. 7, 8** indicate the training and validation accuracy of VGG-16 and ResNet-50. **Table 1** shows the detailed comparison results using SVM, VGG-16, and ResNet-50.



a *b*

Fig. 7. Loss and accuracy of VGG-16:
a – training and validation loss; *b* – training and validation accuracy



a *b*

Fig. 8. Loss and accuracy of ResNet-50:
a – training and validation loss; *b* – training and validation accuracy

Table 1
The comparison of train results

Method	SVM	VGG-16	Resnet-50
Number of classes	52	52	52
Input image size	224×224	224×224	224×224
Number of trains	2300	2300	2300
Number of tests	600	600	600
Batch size	–	32	32
Epochs	–	80	80
Time training	15 m	57 m	1 h 2 m
Accuracy	88 %	94 %	95.6 %

The efficiency of algorithms is evaluated in various environments such as normal light indoors, backlit indoors, low light indoors, natural light outdoors, and backlit outdoors. In addition, this paper also evaluates faces with hats and glasses to examine the accuracy of the methods. The results show that the proposed method is effective as follows:

Case 1. Normal light indoor (500 lux)

The results show that the algorithms recognize exactly the target face despite wearing a hat. **Table 2** shows the result with one person.

In the case of a group of people. Let's perform a test on the faces with marks (hats, glasses) and no marks. The SVM algorithm is unstable. Sometimes, the SVM recognizes wrong all people or does not recognize the target faces. VGG-16 has some frame recognition of «unknown person». **Table 3** shows the result of the algorithms with a group of people.

Table 2

Evaluate algorithms with a person in normal light condition

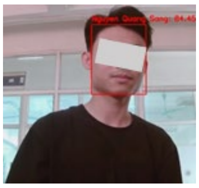
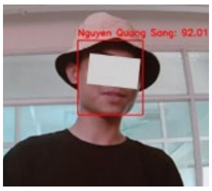
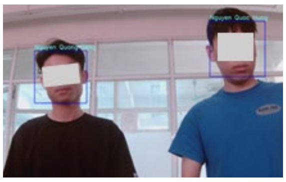
Case	SVM	VGG-16	ResNet-50
No hat			
With hat			

Table 3

Evaluate algorithms with a group of people in normal light condition

Case	SVM	VGG-16	ResNet-50
No hat, no glasses			
With a hat and glasses			

Case 2. Backlit indoor (1000 lux)

In the case of one person, the SVM algorithm still recognized the target face but when wearing a hat, the SVM algorithm has low accuracy or does not recognize the target face. The VGG-16 and ResNet-50 algorithms work with speed and accuracy lower case 1. However, both algorithms still recognized the target face. **Table 4** shows the result of the proposed method in backlit conditions.

In the case of a group of people, the accuracy goes down. Particularly, the SVM algorithm does not recognize all the people. VGG-16 and ResNet-50 still recognized the target faces, but sometimes some frames are missed. **Table 5** shows the result of a group of people.

Table 4
Evaluate algorithms in backlit conditions with one person

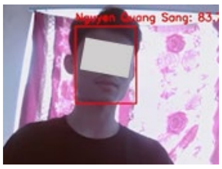
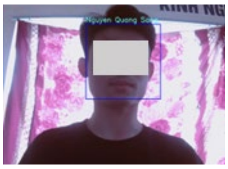
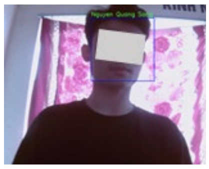
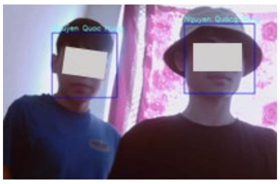
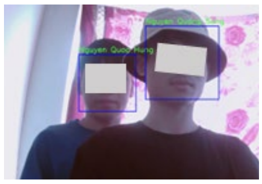
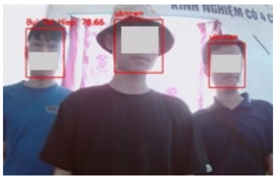
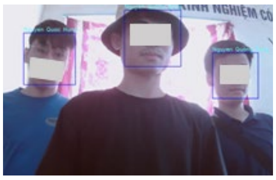
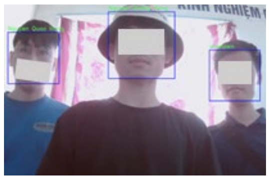
Case	SVM	VGG-16	ResNet-50
No hat			
With hat			

Table 5
Evaluate algorithms with a group of people

Case	SVM	VGG-16	ResNet-50
No glasses			
With a hat and glasses			

Case 3. Low light indoor (150 lux)

In the one-person case, SVM has low accuracy when no hat. In the case of a hat, SVM does not recognize the target face while VGG-16 and ResNet-50 recognize the target face. **Table 6** shows the result when performing low light conditions.

In the case of a group of people, the recognition speed is decreased. When using the hat or glasses, the SVM mistakes target faces or do not recognize faces. VGG-16 and ResNet-50 also have some frames do not be recognized. **Table 7** shows the face recognition results in low-light conditions.

Table 6
Evaluate algorithms with a person in low light condition

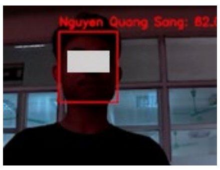
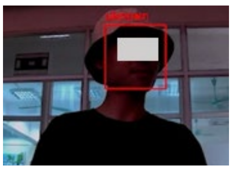
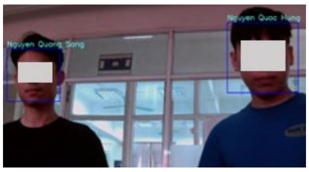
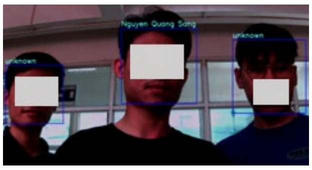
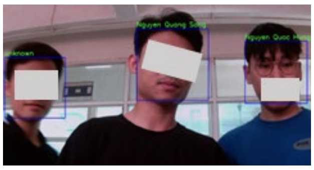
Case	SVM	VGG-16	ResNet-50
No hat			
With hat			

Table 7
Evaluate algorithms with a group of people in low light condition

Case	SVM	VGG-16	ResNet-50
No glasses, no hat			
With a hat and glasses			

Case 4. Natural light outdoor (10752 lux)

In this case, all three algorithms recognize target faces with high accuracy. **Table 8** shows the results outdoors with one person.

In the case of a group of people, the proposed methods work efficiently, and recognize all of the target faces. **Table 9** shows the results with a group of people.

Table 8
Evaluate algorithms with one person in natural light outdoor

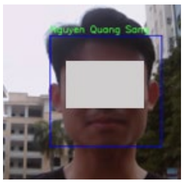
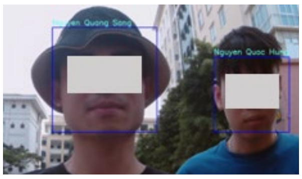
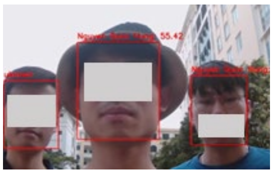
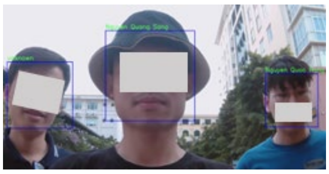
Case	SVM	VGG-16	ResNet-50
No hat			
With hat			

Table 9
Evaluate algorithms with a group of people in natural light outdoor

Case	SVM	VGG-16	ResNet-50
No glasses			
With a hat and glasses			

Case 5. Backlit outdoor (28000 lux)

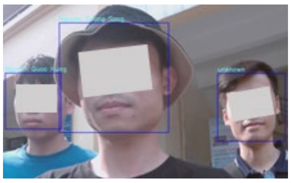
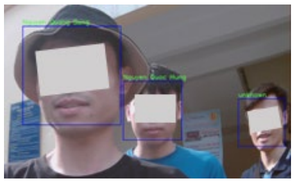
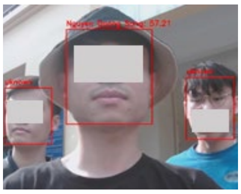
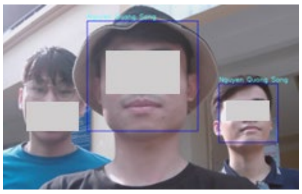
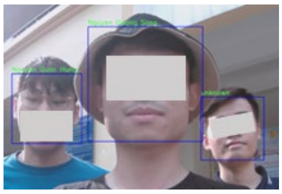
In the case of one person, the accuracy is lower than in case 4. Sometimes, the proposed algorithms do not detect the target faces. When wearing the hat, the algorithms still recognize target faces. **Table 10** shows the results in the backlit outdoors with one person.

In the case of a group of people, sometimes the target face is not detected. In addition, the target faces are recognized with the name of other faces. **Table 11** shows the results in the backlit outdoors with a group of people.

Table 10
Evaluate algorithms with one person in backlit outdoor condition

Case	SVM	VGG-16	ResNet-50
No hat			
With hat			

Table 11
Evaluate algorithms with a group of people in backlit outdoor condition

Case	SVM	VGG-16	ResNet-50
No glasses			
With a hat and glasses			

In all five cases, the results show that the proposed algorithms in the case of one person have higher accuracy when compared with the case of a group of people. The recognition accuracy of the SVM algorithm is lowest at 56.5 % in backlit indoor conditions and highest is 88 % in normal light. The SVM algorithm has a recognition speed higher than the others. However, the SVM is sensitive to noise and the accuracy depends on the feature extraction process before. In complex light conditions, the accuracy of SVM is decreased significantly. The VGG-16 and ResNet-50 have the highest accuracy of up to 93.7 % and 95.4 % respectively. The lowest accuracy of VGG-16 and ResNet-50 is 87.6 % and 88.2 % respectively. The results also indicate that the environmental conditions are important because each evaluation environment has a different similarity threshold. In addition, if faces can be not detected, face recognition algorithms do not work efficiently. All three algorithms need a fixed camera to ensure the algorithms work efficiently.

4. Conclusions

In this paper, three types of facial recognition methods including SVM, VGG-16, and ResNet-50 are discussed and compared. From the result, let's found the pros and cons of the proposed methods. SVM is an efficient algorithm to face recognition. However, if the number of support vectors is too large, the time to find the hyperplane is too long. Both VGG-16 and ResNet-50 are the algorithms using CNN to extract features and achieve effectiveness in terms of speed and accuracy. Yet, the quality and the number of data training greatly affect the training process. Besides, for each condition, similarity measure algorithms need a different threshold of similarity value to optimize face recognition. Experiment results show that ResNet-50 is the most accurate model with the number of training data being the largest. In the future, ResNet-50 will improve to detect and recognize multi faces at the same time. In addition, the face tracking and facial expression classification model can be added to increase the accuracy and efficiency of the model.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Manuscript has no associated data.

Acknowledgements

The authors acknowledge the members of the Intelligent Robotics Laboratory of the Hanoi University of Industry for participating in this research.

References

[1] Mingtsung, C., Wei, Q., Jiaqi, H., Zhuomin, Z. (2020). Research on the application of face recognition system. Journal of Physics: Conference Series, 1684 (1), 012126. doi: https://doi.org/10.1088/1742-6596/1684/1/012126

[2] Oloyede, M. O., Hancke, G. P., Myburgh, H. C. (2020). A review on face recognition systems: recent approaches and challenges. Multimedia Tools and Applications, 79 (37-38), 27891–27922. doi: https://doi.org/10.1007/s11042-020-09261-2

[3] Guo, G., Zhang, N. (2019). A survey on deep learning based face recognition. Computer Vision and Image Understanding, 189, 102805. doi: https://doi.org/10.1016/j.cviu.2019.102805

[4] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J. et al. (2018). CosFace: Large Margin Cosine Loss for Deep Face Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi: https://doi.org/10.1109/cvpr.2018.00552

[5] Kim, J.-H., Kim, B.-G., Roy, P. P., Jeong, D.-M. (2019). Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure. IEEE Access, 7, 41273–41285. doi: https://doi.org/10.1109/access.2019.2907327

[6] Shu, C., Ding, X., Fang, C. (2011). Histogram of the oriented gradient for face recognition. Tsinghua Science and Technology, 16 (2), 216–224. doi: https://doi.org/10.1016/s1007-0214(11)70032-3

[7] Nguyen, N.-Q., Su, S.-F., Tran, Q.-V., Nguyen, V.-T., Jeng, J.-T. (2017). Real time human tracking using improved CAM-shift. 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS). doi: https://doi.org/10.1109/ifsa-scis.2017.8023295

[8] Paul, M., Karsh, R. K., Ahmed Talukdar, F. (2019). Image Hashing based on Shape Context and Speeded Up Robust Features (SURF). 2019 International Conference on Automation, Computational and Technology Management (ICACTM). doi: https://doi.org/10.1109/icactm.2019.8776713

[9] Nguyen, V.-T., Nguyen, A.-T., Nguyen, V.-T., Bui, H.-A. (2021). A Real-Time Human Tracking System Using Convolutional Neural Network and Particle Filter. Lecture Notes in Networks and Systems, 411–417. doi: https://doi.org/10.1007/978-981-16-2094-2_50

[10] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60 (6), 84–90. doi: https://doi.org/10.1145/3065386

[11] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv. doi: https://doi.org/10.48550/arXiv.1409.1556

[12] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D. et al. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: https://doi.org/10.1109/cvpr.2015.7298594

[13] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: https://doi.org/10.1109/cvpr.2016.90

[14] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: https://doi.org/10.1109/cvpr.2017.243

[15] Bah, S. M., Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. Array, 5, 100014. doi: https://doi.org/10.1016/j.array.2019.100014

[16] Ghosh, S., Dasgupta, A., Swetapadma, A. (2019). A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. 2019 International Conference on Intelligent Sustainable Systems (ICISS). doi: https://doi.org/10.1109/iss1.2019.8908018

[17] Chandra, M. A., Bedi, S. S. (2018). Survey on SVM and their application in image classification. International Journal of Information Technology, 13 (5), 1–11. doi: https://doi.org/10.1007/s41870-017-0080-1

[18] Wang, G., Gong, J. (2019). Facial Expression Recognition Based on Improved LeNet-5 CNN. 2019 Chinese Control And Decision Conference (CCDC). doi: https://doi.org/10.1109/ccdc.2019.8832535

[19] Vedalankar, A. V., Gupta, S. S., Manthalkar, R. R. (2021). Addressing architectural distortion in mammogram using AlexNet and support vector machine. Informatics in Medicine Unlocked, 23, 100551. doi: https://doi.org/10.1016/j.imu.2021.100551

[20] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215. doi: https://doi.org/10.1016/j.neucom.2019.10.118

[21] Agarap, A. F. (2017). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. arXiv. doi: https://doi.org/10.48550/arXiv.1712.03541

[22] Kherchaoui, S., Houacine, A. (2014). Facial expression identification system with Euclidean distance of facial edges. 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR). doi: https://doi.org/10.1109/socpar.2014.7007973

[23] Shieh, M.-Y., Chiou, J.-S., Hu, Y.-C., Wang, K.-Y. (2014). Applications of PCA and SVM-PSO Based Real-Time Face Recognition System. Mathematical Problems in Engineering, 2014, 1–12. doi: https://doi.org/10.1155/2014/530251

[24] Agarwal, T., Mittal, H. (2019). Performance Comparison of Deep Neural Networks on Image Datasets. 2019 Twelfth International Conference on Contemporary Computing (IC3). doi: https://doi.org/10.1109/ic3.2019.8844924

[25] Wu, S. (2021). Expression Recognition Method Using Improved VGG16 Network Model in Robot Interaction. Journal of Robotics, 2021, 1–9. doi: https://doi.org/10.1155/2021/9326695

[26] Mandal, B., Okeukwu, A., Theis, Y. (2021). Masked Face Recognition using ResNet-50. arXiv. doi: https://doi.org/10.48550/arXiv.2104.08997

[27] Alnuaim, A., Zakariah, M., Hatamleh, W. A., Tarazi, H., Tripathi, V., Amoatey, E. T. (2022). Human-Computer Interaction with Hand Gesture Recognition Using ResNet and MobileNet. Computational Intelligence and Neuroscience, 2022, 1–16. doi: https://doi.org/10.1155/2022/8777355

[28] Zhang, M., Yu, Z., Wang, H., Qin, H., Zhao, W., Liu, Y. (2019). Automatic Digital Modulation Classification Based on Curriculum Learning. Applied Sciences, 9 (10), 2171. doi: https://doi.org/10.3390/app9102171

[29] Elmahmudi, A., Ugail, H. (2019). Deep face recognition using imperfect facial data. Future Generation Computer Systems, 99, 213–225. doi: https://doi.org/10.1016/j.future.2019.04.025

[30] Xu, Y., Cheng, J. (2020). Face Recognition Algorithm Based on Correlation Coefficient and Ensemble-Augmented Sparsity. IEEE Access, 8, 183972–183982. doi: https://doi.org/10.1109/access.2020.3028905

[31] Lahitani, A. R., Permanasari, A. E., Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. 2016 4th International Conference on Cyber and IT Service Management. doi: https://doi.org/10.1109/citsm.2016.7577578

Received date 27.03.2023
Accepted date 23.06.2023
Published date 27.07.2023

How to cite: *Nguyen, T. V., Chu, T. D. (2023). Comparative study on the performance of face recognition algorithms. EUREKA: Physics and Engineering, 4, 120–132. doi: https://doi.org/10.21303/2461-4262.2023.002831*