International Journal of Applied Science and Engineering

Masked face recognition based on facenet and genetic algorithm

Wen-Chang Cheng[1], Hung-Chou Hsiao[2*], You-Quan Hong[1], De-Yu Wang[1]

[1] *Department of Computer Science & Information Engineering, Chaoyang University of Technology, Taichung City, Taiwan*
[2] *Department of Information Management, Chaoyang University of Technology, Taichung City, Taiwan*

ABSTRACT

This study proposes a face recognition method based on FaceNet for face masks. FaceNet converts faces into 128-dimensional feature vectors using a network of deep learning models. The smaller the Euclidean distance between different face images, the closer the person is to the same person; conversely, the larger the Euclidean distance, the more different the person is. Wearing a mask affects some feature vector dimensions of FaceNet conversion, resulting in reduced recognition ability. This study uses a Genetic Algorithm (GA) to select and remove the feature vectors affected by mask wearing. The experimental results show that the validation rate value of 55 features removed by the GA is increased from 0.550 to 0.650 for the original unremoved features.

Keywords: Deep learning, Genetic algorithm, Optimization algorithm, Face recognition, Euclidean distance.

OPEN ACCESS

Received: March 24, 2023
Revised: May 8, 2023
Accepted: May 22, 2023

Corresponding Author:
Hung-Chou Hsiao
s10814902@gm.cyut.edu.tw

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

The World Health Organization (WHO, 2023) and the Centres for Disease Control and Prevention (CDC, 2023) recommend wearing a mask and keeping a social distance from others to avoid the spread of COVID-19 virus or infection. Therefore, governments around the world require people to wear masks in public places. Face recognition (FR) requires the use of a full-face image in the recognition process. Masked Face Recognition (MFR) can be a major challenge for face verification or identification of people wearing masks. From the spread of the COVID-19 epidemic to the present, research efforts in the field of MFR have continued to increase. Biometrics research units such as NEC (NEC) and Thales (Thales Group, 2020) have also been forced to adapt their original algorithms to improve the accuracy of FR systems for people wearing masks. The National Institute of Standards and Technology (NIST) (Ngan et al., 2020) developed a tuned FR algorithm pre-COVID-19, which concluded that the performance of most FR algorithms declined when faces were obscured after the virus pandemic.

In recent years, with the dramatic increase in the performance of hardware devices and the popularization of Graphic Processing Unit (GPU). Deep learning is prevalent and has applications in many fields, e.g., musical instruments (Dewi et al., 2023), image captioning (Dewi et al., 2023), etc. Traditional FR methods (Kelly, 1970; Sirovich and Kirby, 1987; Takacs, 1988; Liu and Wechsler, 2022) have been replaced by deep learning methods (Huang et al., 2021; Kai et al., 2021). Most FR systems have switched to using models constructed by deep learning. Research related to face masking has been proposed using deep learning methods before the COVID-19 pandemic. Zeng et al. (2022) proposed a sparse representation combining L1 and L2 regularization exhibiting good performance. The paper analyses how the proposed method helps in MFR and illustrates the basic principles of its anti-noise. Song et al. (2019) proposed a mask learning method to repair damaged feature elements in human faces to develop the Pairwise Differential Siamese Network (PDSN). Masking dictionary using the difference between masked and unmasked face pairs of convolutional features. Compositing and realistic masking of the face dataset to get good results in the end. Chen et al. (2022) proposed a novel deep panoramic segmentation method based on a bidirectional learning pipeline. An occlusion processing algorithm was developed to handle the occlusion between different targets. Bian and Li (2021) proposed a Conditional Adversarial Consistent Identity Auto Encoder (CACIAE) learn more details about the face and generate more natural images, which can fix the occluded face before recognition. Dewi and Chen (2022) enhanced the Yolov4 model by adding Cross Stage Partial (CSP) and Spatial Pyramid Pooling (SPP) mechanisms to detect medical masks. Dewi et al. (2023) use Yolov7 to detect whether users are wearing the mask correctly and classify the results into three categories (none, bad and good). In response to the COVID-19 outbreak, Wang et al. (2023) created three datasets, Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Synthetic Masked Face Recognition Dataset (SMFRD), and tested using off-the-shelf object detection. Chong et al. (2023) developed a histogram-based recurrent neural network (HRNN) model using histograms of oriented gradients (HOG) as a feature in combination with a HRNN. Shukla et al. (2023) used VGG16, VGG19, ResNet50 and ResNet101 pre-trained models combined with MobileNetV2 to determine the presence or absence of masks. A study of MFR based on an Attention mechanism called Convolutional Visual Self-Attention Network (CVSAN) was proposed by Ge et al. (2023).

Among the deep learning solutions for FR problems, FaceNet is a good and practical approach (Schroff et al., 2015; Cheng et al., 2021). The FaceNet concept uses a neural network model to obtain the feature vectors from the image of a human face. By comparing the Euclidean distance between the feature vectors of two face images, we can determine whether they are the same person. According to a previous study in our lab (Cheng and Hong, 2022), it was found that some dimensions of the FaceNet transformed 128-dimensional feature vector are affected by the masks. Fig. 1 shows the 128-dimensional standard deviation of multiple face image feature vectors of the same person. The dashed line in the figure is the 128-dimensional standard deviation of the image feature vector of the unmasked face, and it can be found that the standard deviation is very small. The solid line is the 128-dimensional standard deviation of face image feature vectors with and without a mask. The difference in the standard deviation of some dimensions of the feature vectors becomes larger with the mask on after trying to remove the top 10 dimensions with the largest standard deviations from the 128 dimensions. The accuracy of FaceNet's MFR recognition has increased from 0.58 to 0.93. Therefore, removing some of the feature vectors affected by the mask improves the accuracy. However, selecting feature vector dimensions for removal using statistical standard deviation is too subjective. In this study, we try to choose the removable dimensions by the optimization algorithm so that the feature vectors can still be applied to the MFR problem after the removal of dimensions.

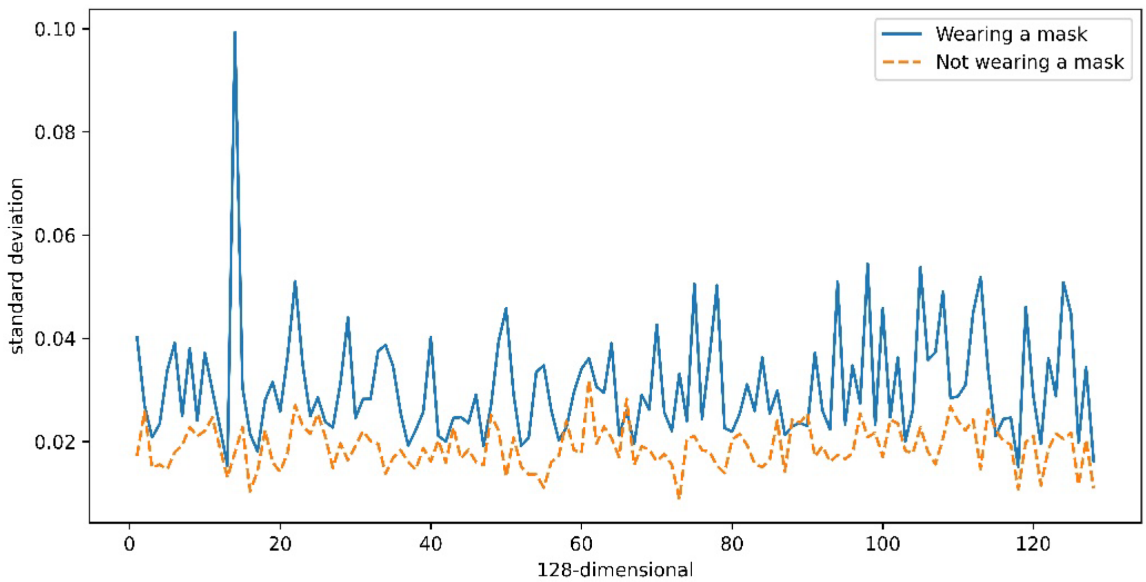

Fig. 1. Standard deviation in 128 dimensions (Cheng and Hong, 2022)

There are many optimization algorithms, such as Genetic Algorithm (GA) (Holland, 1992), Particle Swarm Optimization (PSO) (Eberhart and Kennedy, 1995), Firefly Algorithm (FA) (Yang, 2009), Ant Colony Optimization (ACO) (Dorigo et al., 1996), and Sparrow Search Algorithm (SSA) (Xue and Shen, 2020), among others. The GA can solve most optimization problems, including optimization parameter design, scheduling problems, classification systems, control system engineering, travel agent problems, etc. GA has the following advantages (1) high practicality, GA can adjust the chromosome according to various needs, (2) the whole group search, using a randomized multi-point search method, the solution obtained is not easy to occur in the region of the best solution problem, (3) easy to understand the concept. GA is more simple than other algorithms. Therefore, this study uses GA to select and remove the dimensions affected by masks from the 128-dimensional feature vectors of face images after FaceNet computation, which proves that some of the dimensional features in face images are affected by masks and that the MFR performance can be effectively improved by dimensional reduction.

2. MATERIALS AND METHODS

This section describes the research methods in this study, FaceNet, and Genetic Algorithm, described below.

2.1 FaceNet

FaceNet is one of the methods based on deep learning and is widely used in FR, which was proposed by the Google team in 2015 (Schroff et al., 2015). The concept of FaceNet is to map the pixels of a face in an image into feature space and convert them into feature vectors. FaceNet is trained by a loss function called Triplet Loss (Schroff et al., 2015). In the process of model learning, the aim is to make the distance between the facial feature vectors of the same person in different images as small as possible and the distance between the facial feature vectors of different people as large as possible. In the FaceNet training process, the input image consists of two different face images of the same person (A and A' in Fig. 2) and one face image of a different person (B in Fig. 2). 128-dimensional feature vectors x_i^A, $\mathrm{x}_i^{A'}$, and x_i^B are extracted from the face images by a deep learning model. The loss is then calculated using the Triplet Loss function, which is shown in Equation 1:

$$L = \sum_i^P \left[||\boldsymbol{x}_i^A - \boldsymbol{x}_i^{A'}||_2^2 - ||\boldsymbol{x}_i^A - \boldsymbol{x}_i^B||_2^2 + \alpha \right] \quad (1)$$

Where α is the margin between positive and negative samples, and P is the number of ternary training sample pairs.

The concept of Triple Loss is shown in Fig. 3. In Triplet Loss, A and A' are the same person and B is a different person. If the distance between A and A' feature vector is $\overline{d1}$, and the distance between A and B feature vector is $\overline{d2}$. The distance between the two before training is $\overline{d1} < \overline{d2}$, as shown in the left half of Fig. 3 By minimizing the face feature vector distance of the same person and maximizing the face feature vector distance of different people. The distance between A and A' is brought closer, and then A and B are brought further apart. That is, $\overline{d1}' > \overline{d2}'$, as shown in the right half of Fig. 3 After the above training, the FaceNet training method can effectively extract the face feature vector. The face feature vector distance can be a similarity discriminator in face matching. For a detailed description, please refer to the paper by William et al. (2019).

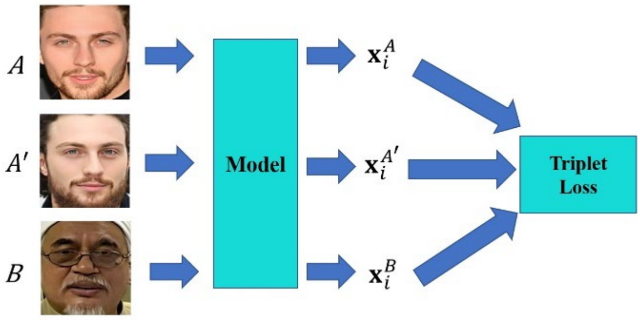

Fig. 2. Triplet loss function training method (Schroff et al., 2015) (Face Image Source (ZENBOT99, 2022))

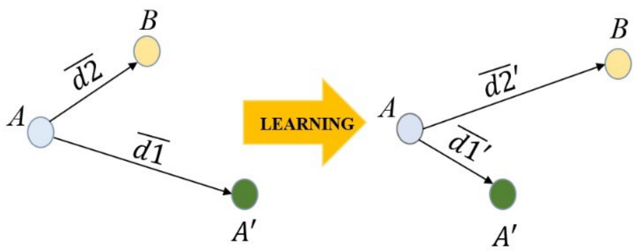

Fig. 3. Triple loss concept (Schroff et al., 2015)

2.2 Genetic Algorithm

Genetic Algorithm (GA) (Holland, 1992; Negnevitsky, 2019), which follows the evolutionary rule of "survival of the fittest and elimination of the unfit" as proposed in evolutionary theory (Darwin and Wallace, 1858). The core concept is to simulate the reproduction of natural organisms by mating and mutating genes in chromosomes to produce the genes of offspring. We represent a solution or parameter of the target problem to be solved as a set of chromosomes. The chromosomes are then converted into numerical form, and the values represent the genes in the chromosomes. Then, through mutation and mating, the next generation will have different potential solutions. The good chromosomes are then retained for the next round of mating mutations to produce a better solution until the set stopping condition is reached.

The GA starts by defining the optimization problem, which is often a function that requires the maximum or minimum value. A Chromosome in GA usually represents a Potential Solution. Chromosomes are made up of genes, which are part of the solution. Therefore, before starting the algorithm, the chromosome must be designed according to the problem attributes, and the fixed chromosome length represents the domain of the problem variables. There are many ways to encode chromosomes. The common encoding methods are binary encoding and Gray code. In other words, the chromosome is converted into an array of 1 s and 0 s, and this coding method is often used to solve for numerical forms. The general way to generate the initial group is randomly generated with random numbers. In a study by Wu et al. (2021), it was mentioned that the initial population could be set using a greedy method to increase the rate of astringency. This study still uses the random generation method to maintain species diversity. The important steps of the GA process are as follows.

Step 1: Generate the initial population, and generate a group of chromosomes as the initial population, that is, the initial solution.

Step 2: To calculate the fitness, we calculate the adaptability of each chromosome f_i according to the definition of the fitness function, and then calculate the adaptability F_i.

$$F_i = \frac{f_i}{\sum_{j=1}^{N} f_j}, i = 1, 2, \dots, N \tag{2}$$

N in Equation 2 denotes the number of chromosomes, and F_i is the adaptation of the ith chromosome. The higher the fitness, the greater the chance the chromosome will be retained for reproduction in the next step of chromosome selection. Conversely, the smaller the adaptation, the greater the chance of being eliminated.

Step 3: Selection, select chromosomes of better quality and keep them to form new populations. There are two commonly used options. (a) Elite method: Select chromosomes with higher adaptive F_i as the parental generation. It retains excellent chromosomes, but tends to fall into local optimal solutions. (b) Roulette wheel, a roulette wheel was designed according to the adaptive F_i level of each chromosome group. Chromosomes with larger areas will have a higher chance of being selected. This method will be used in this study.

Step 4: Crossover; before mating, we will decide whether to select two chromosomes for mating according to the set crossover rate. In the process, some genes of the two chromosomes are exchanged and recombined to produce new chromosomes.

Step 5: Mutation, decide whether a chromosome should be mutated according to the set mutation rate. It can increase the diversity of solutions while avoiding getting trapped in locally optimal solutions. Genes at certain positions in the chromosome are randomly selected for swapping when mutations occur.

The process of selection, mating, and mutation produces a new generation of populations and then calculates the adaptation of the new populations to F_i. The new cohort is merged with the existing cohort, and those with low fitness F_i are eliminated according to the set initial cohort size. Let the overall adaptability improve, and continue to iterate until the termination conditions are met.

This study uses GA to select which FaceNet transformed feature vector dimensions will be removed. Binary coding of chromosomes was used in the experiments. FaceNet has 128-dimensional feature vectors, and the 7th power of 2 can represent 0 to 127 in the decimal system, so the chromosome length is set to 7 bits. Assuming that the number of removed feature vectors is M, a chromosome has 7 × M bits, and the initial population is N, as shown in Fig. 4.

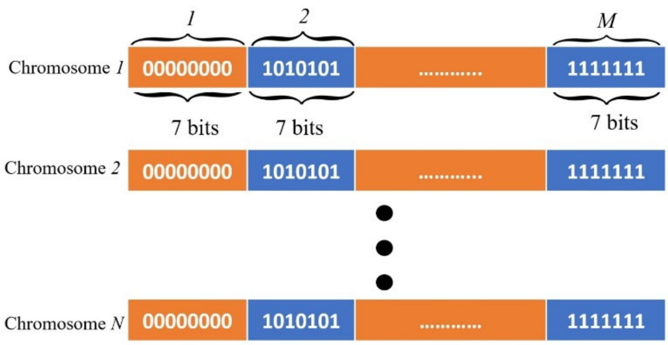

Fig. 4. *N* chromosomes in the *M*-group range 0-127

3. RESULTS AND DISCUSSION

This section describes the dataset used in the experiment, the evaluation criteria, and the experiment results using the GA to select M dimensions with a different number of feature vectors removed. Finally, the experimental results are discussed.

3.1 Datasets and Evaluation Criteria

3.1.1 Datasets

The experiments use VGGFace2_HQ_CROP (ZENBOT99, 2022) and LFW simulated masked face dataset (LFW-SMFRD) (X-zhangyang, 2020). The VGGFace2_HQ_CROP dataset is modified from the VGGFace2 dataset (Cao et al., 2018). VGGFace2 has 3.31 million images, but each VGGFace2 image has a different resolution and size. Therefore, (Liu N., 2021) to VGGFace2 High Quality as VGGFace2_HQ. VGGFace2_HQ_CROP is to crop the face in VGGFace2_HQ to 160 × 160 pixels. However, to use VGGFace2_HQ_CROP, it is necessary to synthesize the masks from the images of human faces. For masks, LFW-SMFRD was synthesized using images from the Labeled Faces in the Wild (LFW) dataset (Huang et al., 2007).

5713 faces are classified in the LFW-SMFRD dataset, and the number of images is 13117, and the faces have been captured, aligned, and cropped to 128 × 128 pixels from LFW. VGGFace2_HQ_CROP and LFW-SMFRD have completed the basic pre-processing action, so no other pre-processing action was performed in the experiment.

Since VGGFace2_HQ_CROP does not have images of faces wearing masks, the experimental experiment uses MaskTheFace (Anwar and Raychowdhury, 2020) to synthesize masks from VGGFace2_HQ_CROP by referring to Cheng et al. (2023). The experiment randomly selected 40 faces from the VGGFace2_HQ_CROP dataset, each with 10 face images, and synthesized 10 images corresponding to the masks, so there will be 800 images in total. In the experiment, V_MASK is used as the name of this dataset, and Fig. 5 shows the image of V_MASK.

Fig. 5. A part of the V_MASK experimental image. The above are the original images (ZENBOT99, 2022), and the following are synthesized images (Referring to the synthetic approach of Cheng et al., 2023)

The LFW-SMFRD is already a mask synthesis dataset, so there is no need to perform mask synthesis again. In the experiment, 40 faces were selected randomly from the LFW-SMFRD dataset, 10 images of each face wearing a mask were selected, and the corresponding 10 images of the original faces were found from the LFW dataset. Therefore, there will be 800 images. For the convenience of explanation, the experiment is named L_MASK for this dataset. Fig. 6 shows some of the images used in the experiment of L_MASK.

Fig. 6. A part of the L_MASK experimental image. The above are the LFW images (Huang et al., 2007), and the following are LFW-SMFRD images (X-zhangyang, 2020)

V_MASK and L_MASK are divided into two parts: the training set and the test set. In the training set, 40 faces are classified into 10 images for each face (5 are the original face images, and 5 are the corresponding mask images), so there are 400 images. The test set also has the same 40 face categories as the training set, with 10 images for each face not used in the training set (5 of which are the original face images and 5 are the corresponding mask images).

3.1.2 Evaluation Criteria

In this study, the evaluation is performed with reference to the model evaluation approach in the original FaceNet paper (Schroff et al., 2015), which defines the true accepts $TA(d)$ set as Equation 3:

$$TA(d) = \{(A, B) \in P_{same},\ with\ D(\mathrm{x}^A, \mathrm{x}^B) \le d\} \tag{3}$$

Where (A, B) are two paired face images, P_{same} is the set of two paired face images of the same person, $D(\mathrm{x}^A, \mathrm{x}^B)$ is the feature vector distance, and d is the minimum distance identified as the same person. Thus $TA(d)$ is the set of two face images of the same pair of people in the set that matches $D(\mathrm{x}^A, \mathrm{x}^B)$ less than d.

Then, the validation rate $VAL(d)$ is defined as an indicator to evaluate the model. $VAL(d)$ refers to the proportion of the set of two face images of the same person pairs smaller than d in the set of two face images of the same person pairs. The equation of $VAL(d)$ is as in (4):

$$VAL(d) = \frac{|TA(d)|}{|P_{same}|} \tag{4}$$

where $|TA(d)|$ is the number of $TA(d)$ sets and $|P_{same}|$ is the number of P_{same} sets. Since $VAL(d)$ only considers two face images of the same person pairs. Since $VAL(d)$ only considers the two face images of the same person pair, it does not consider the two face images of the different person pair, so it defines the false accepts $FA(d)$ set separately, $FA(d)$ is the set of two face images of different people paired in the set that matches $D(\mathrm{x}^A, \mathrm{x}^B)$ less than d. As in Equation 5.

$$FA(d) = \{(A, B) \in P_{diff},\ with\ D(\mathrm{x}^A, \mathrm{x}^B) \le d\} \tag{5}$$

Where P_{diff} is the set of two face images of different people. Since the larger the $|TA(d)|$ the better, and conversely, the smaller the $|FA(d)|$ the better. The adaptation value of the GA can be proportional to the ratio of these two values:

$$fitness \approx \frac{|TA(d)|}{|FA(d)|} \tag{6}$$

Before starting the experiment, the $F1$ fraction is used to find the suitable threshold (d) by referring to the paper by Cheng et al. (2023). $F1$ is the summed average of *Precision* and *Recall*, as in Equation 7:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

$Recall$ the correct prediction for the sample that is actually true, as in Equation 8. $Precision$ is predicted correctly in positive samples, as in Equation 9.

$$Recall = \frac{tp}{tp+fn} \quad (8)$$

$$Precision = \frac{tp}{tp+fp} \quad (9)$$

Where True Positive (tp) is actually true, and the judgment is true, False Negative (fn) is actually true, but the judgment is false, and False Positive (fp) is actually false, but the judgment is true.

Since two different datasets are used, d needs to be defined individually. The $F1$ score is calculated from the normal face images in the training set of V_MASK and L_MASK according to different threshold values, and the threshold range is set to [0.3-1.8]. The $F1$ score is calculated from the threshold value of 0.3, and each time is increased by 0.01 until the end of 1.8. Fig. 7 shows the result of the threshold calculation for the V_MASK data set. The $F1$ is highest when the threshold value is 0.83, so d is set to 0.83. Fig. 8 shows the result of the threshold calculation for the L_MASK data set. The $F1$ is highest when the threshold value is 0.90, so d is set to 0.90 when testing L_MASK.

3.2 Experiment Method

The concept diagram of the experiment combining FaceNet and GA is shown in Fig. 9. After sending the image to the CNN model, 128 dimensions are obtained. Then, according to the number of dimensions to remove, M and the GA automatically select which dimensions to remove. Finally, (128-M) dimensions are obtained. Table 1 shows the parameter settings of the GA.

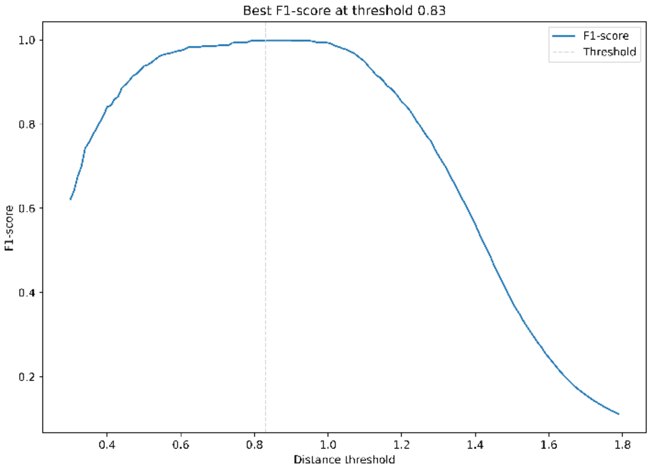

Fig. 7. F1 for different thresholds in V_MASK (from 0.3 to 1.8)

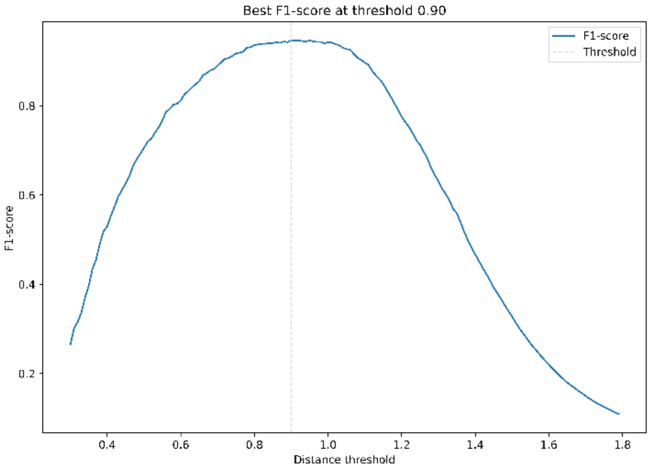

Fig. 8. F1 for different thresholds in L_MASK (from 0.3 to 1.8)

The threshold d for determining whether the individual is the same is adjusted linearly according to Equation 10.

$$d_{new} = (128 - M) \times \frac{d}{128} \quad (10)$$

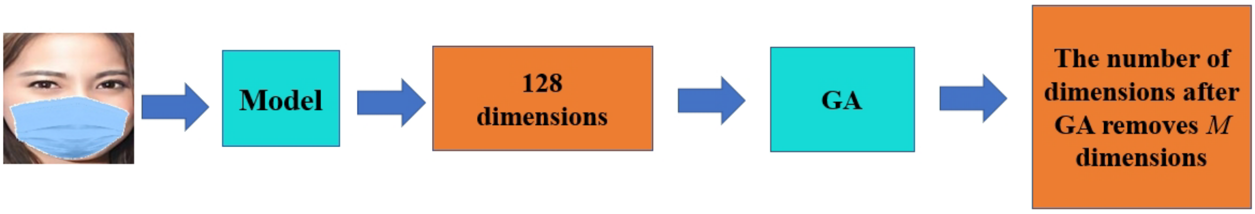

Fig. 9. Experimental concept diagram

Table 1. Genetic algorithm parameters

Remove feature number (M)	Initial population size N	Logarithm of progeny	Mating rate	Mutation rate	Number of iterations
5, 10, 15, …, 60	M × 2	M/2	0.9	0.1	1500

where the original d without the number of dimensions removed is 0.83 and 0.90, calculated in subsection 3.1, and M represents the number of dimensions to be removed. Fig. 10 and 11 show the images of V_MASK and L_MASK using the training set, respectively, with GA convergence curves at 5, 10, 15 and 20 dimensions removed. It can be concluded that the GA tends to converge.

3.3 Experiment Results

Fig. 12 shows the $VAL(d)$ results of FaceNet after removing M = 5, 10, 15, ..., 60 dimensions for both V_MASK and L_MASK training sets, where the average value is taken 10 times each time. The horizontal axis indicates the number of features removed (M), and the vertical axis shows the corresponding $VAL(d)$ after removing the features. V_MASK's $VAL(d)$ is only 0.624 when no feature is removed at the beginning (M = 0). The highest $VAL(d)$ of 0.774 is obtained when M = 55, which is 0.150 more elevated than the $VAL(d)$ without feature removal, and the $VAL(d)$ stops increasing and starts decreasing when M = 60. The $VAL(d)$ of L_MASK is only 0.174 when M = 0. As M grows, the $VAL(d)$ also shows an increasing trend. When M = 45, M = 50, and M = 55, the $VAL(d)$ are 0.440, 0.438 and 0.442, respectively, and the results are similar in these three cases, while when M = 60, the $VAL(d)$ have started to decrease. Therefore, the experiments with these two datasets only try to remove 60 features.

In V_MASK, it is concluded that M = 55 is the best $VAL(d)$, while in L_MASK, when M = 45, M = 50, and M = 55, the $VAL(d)$ is about the same, but the purpose of this study is to find the maximum number of M that can be removed, so the result of M = 55 still prevails. The test set images were used again in the experiment to verify the effect. Table 2 shows the $VAL(d)$ test results of removing 0 and 55 features. The V_MASK test set has a $VAL(d)$ of 0.550 when no feature is removed, and the $VAL(d)$ of removing M = 55 is 0.650. the L_MASK test set has a $VAL(d)$ value of 0.154 when no feature is removed, and the $VAL(d)$ of removing M = 55 is 0.404.

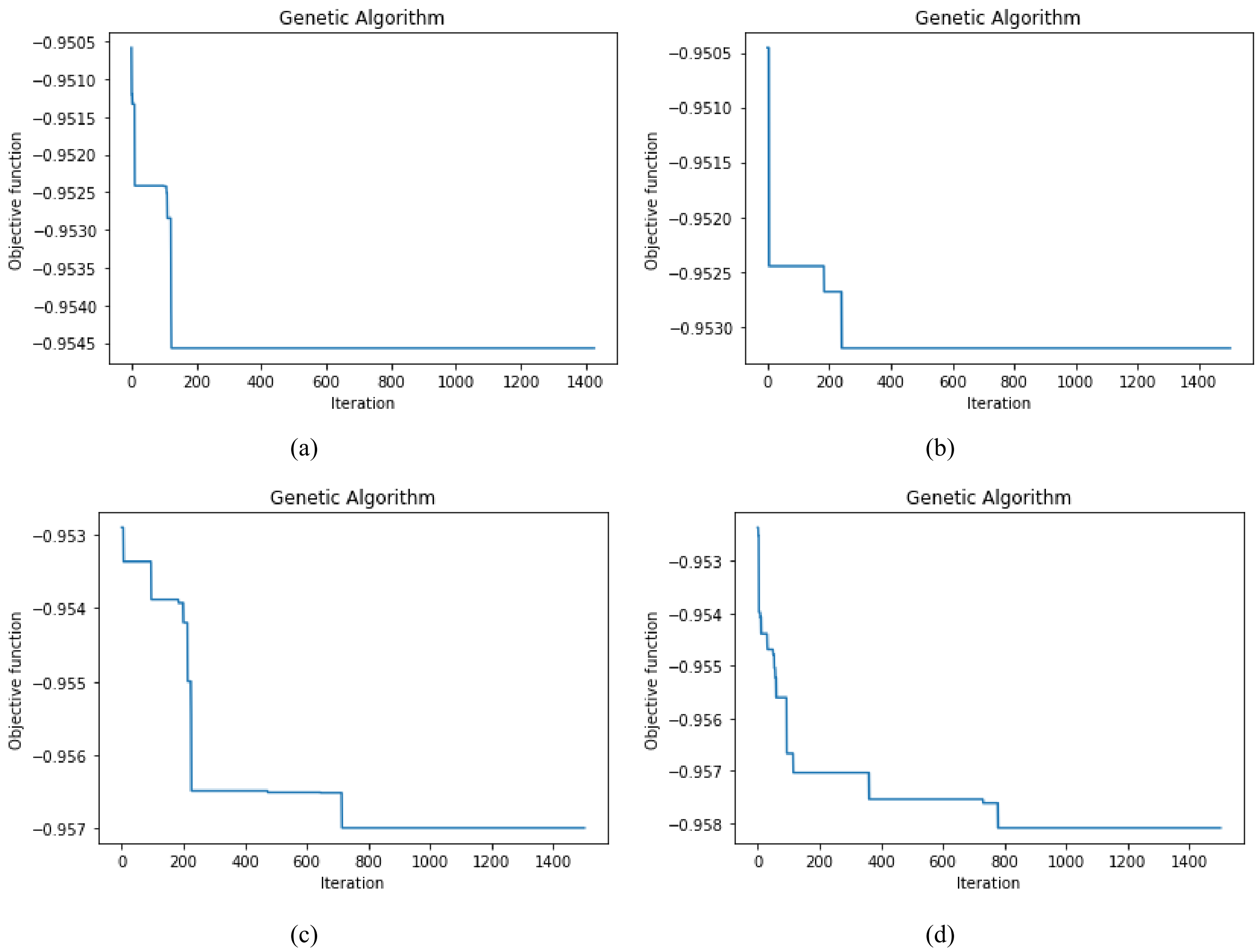

Fig. 10. V_MASK training set. convergence curve of the genetic algorithm (a) 5 dimensions are removed, (b) 10 dimensions are removed, (c) 15 dimensions are removed, (d) 20 dimensions are removed

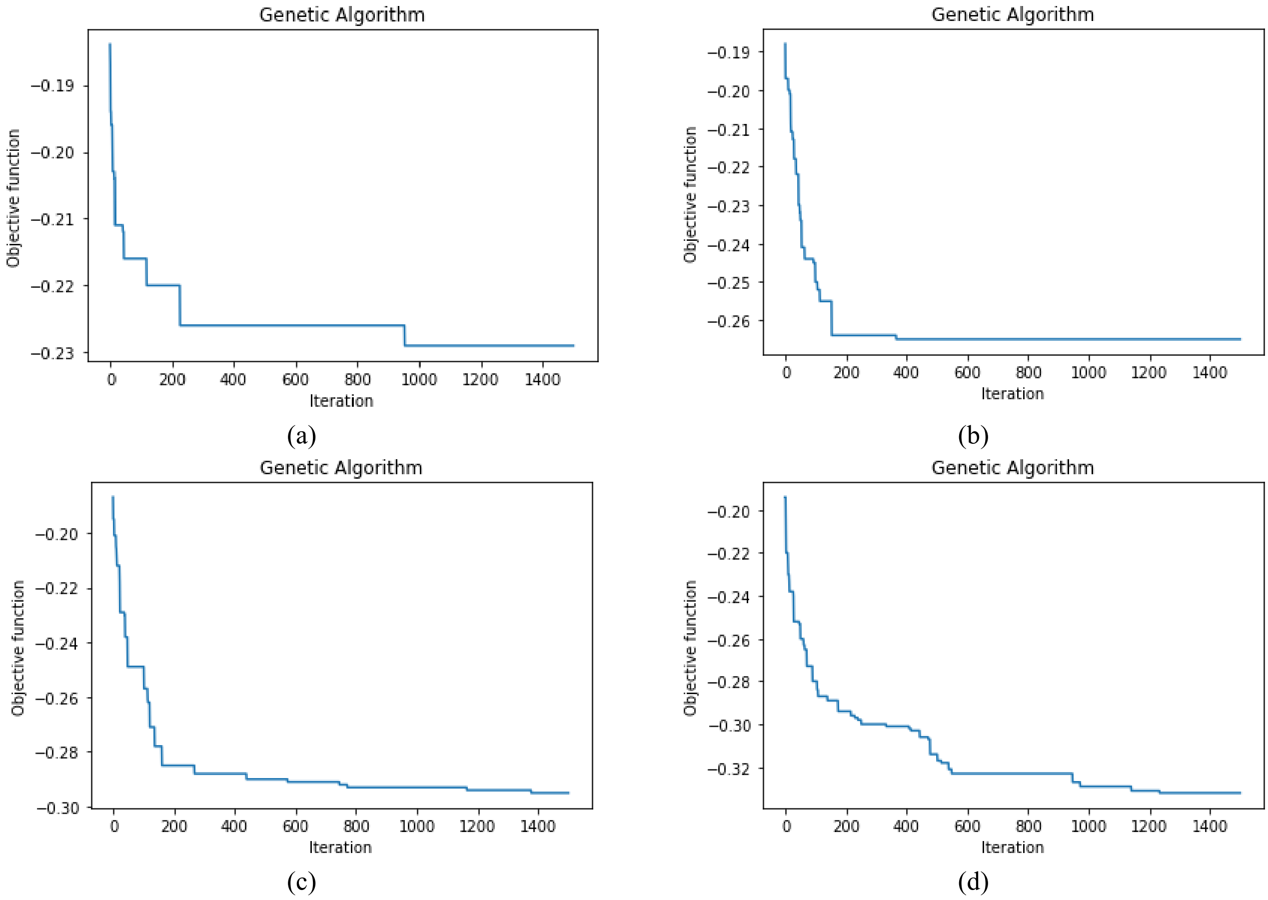

Fig. 11. L_MASK training set. convergence curve of the genetic algorithm (a) 5 dimensions are removed, (b) 10 dimensions are removed, (c) 15 dimensions are removed, (d) 20 dimensions are removed

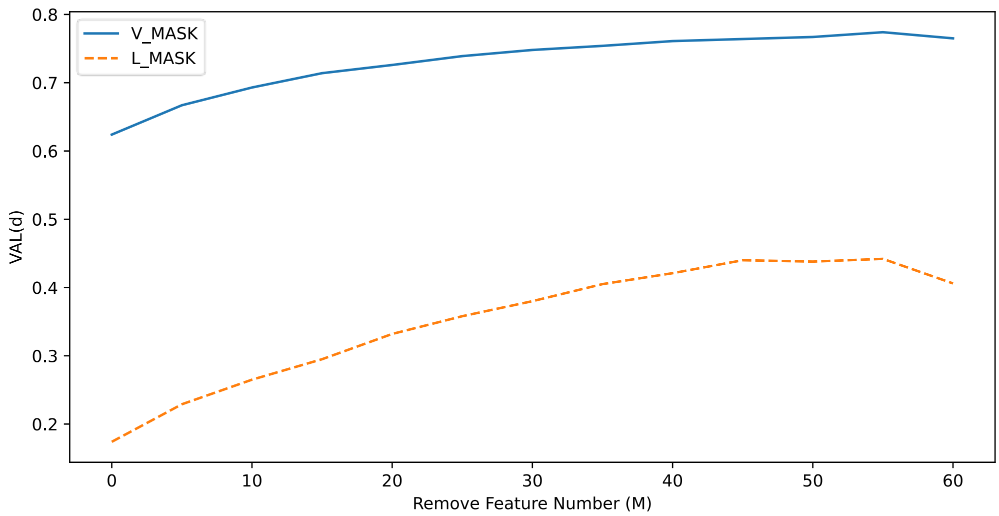

Fig. 12. Remove $M = 0, 5, 10, 15, \ldots, 60$ features of $VAL(d)$

Table 2. The result of the VAL(d) test with M features removed

Remove feature number (M)	V_MASK	L_MASK
0	0.550	0.154
55	0.650	0.404

3.4 Discussion

The experimental results in 3.3 show that when using GA to select the number of M to be removed, the $VAL(d)$ value increases as the number of M increases. The $VAL(d)$ is optimal until M = 55. When M = 60, $VAL(d)$ starts to decline because too many features are removed, and too little information is available for comparison, resulting in the decline of $VAL(d)$. From this situation, it can be seen that there are as many as 55 affected dimensions in the original FaceNet 128 dimensions. Therefore, it is theoretically possible to ignore 55 of these dimensions. In the other test results, the $VAL(d)$ = 0.550 for M = 0 in V_MASK increased to $VAL(d)$ = 0.650 for M = 55, while the $VAL(d)$= 0.154 for L_MASK increased to $VAL(d)$= 0.404 for M = 55, which has potential for improvement. (1) The number of images is too small. Although attempts have been made to increase the number of images, GA takes much time to select the dimensions, and as the number of images increases, it takes more time. Adding images to enhance $VAL(d)$ would require significant energy costs. (2) The image quality affects the performance. The performance of L_MASK is not as good as V_MASK because the image resolution of LFW and LFW-SMFRD, the original data set of L_MASK, is worse than that of VGGFace2_HQ_Crop, the original data set of V_MASK. Therefore, in future studies, we will adjust our strategy to increase the number of experimental images in the dataset and also retrain the model to process face images of people wearing masks, aiming to improve the model's performance.

4. CONCLUSION

This study uses FaceNet and GA to demonstrate that wearing a mask on a face image affects some dimensional features, and deleting dimensions can effectively improve the performance of MFR. The 128-dimensional feature vector generated by FaceNet is then used to select which dimensions are affected by mask wearing and remove them by GA. It is proved that after using GA to select the features affected by the mask and remove them, the $VAL(d)$ of the V_MASK training set increases from 0.624 to 0.774 during the experiment, and the best $VAL(d)$ reaches 0.650 in the test. The $VAL(d)$ of the L_MASK training set is improved from 0.174 to 0.442 during the experiment, and the best $VAL(d)$ of 0.404 is achieved in the test. It is proved that the performance of MFR can be improved by dimensional reduction. Therefore, future research will change the idea. We will re-train the model for masked face images and increase the dataset of experimental images. To improve the accuracy of the model and achieve the practical purpose.

ACKNOWLEDGMENT

This paper is funded by the National Science and Technology Council (Restructuring of the former Ministry of Science and Technology), Taiwan. The No. is MOST-111-2637-E-324-001, Taiwan.

REFERENCES

Anwar, A., Raychowdhury, A. 2020. Masked face recognition for secure authentication. arXiv, arXiv: 2008. 11104.

Bain, X., Li, J. 2021. Conditional adversarial consistent identity autoencoder for cross-age face synthesis. Multimedia Tools and Applications, 80, 14231–14253.

Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A. 2018. VGGFace2: A dataset for recognizing faces across pose and age. 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 67–74.

CDC. 2023. COVID-19 prevention actions. Retrieved 2023-06-06 from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/index.html

Cheng, W.C., Hong, Y.Q. 2022. FaceNet-based feature vector statistics for masked face recognition. Workshop on Consumer Electronics (WCE2022).

Cheng, W.C., Hsiao, H.C., Lee, D.W. 2021. Face recognition system with feature normalization. International Journal of Applied Science and Engineering, 18, 1–9.

Cheng, W.C., Hsiao, H.C. Li, L.H. 2023. Deep learning mask face recognition with annealing mechanism. Applied Sciences, 13, 732.

Chen, Y., Lin, G., Li, S., Omar, B., Wu, Y., Wang, F., Feng, J., Xu, M., Li, X. 2020. Bidirectional aggregation network with occlusion handling for panoptic segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3793–3802.

Chong, W.-J.L., Chong, S.-C., Ong, T.-S. 2023. Masked face recognition using histogram-based recurrent neural network. Journal of Imaging, 9, 38.

Darwin, C., Wallace, A.R. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Zoological Journal of the Linnean Society, 3, 45–62.

Dewi, C., Chen, R.-C. 2022. Automatic medical face mask detection based on cross-stage partial network to combat COVID-19. Big Data and Cognitive Computing. 6, 106.

Dewi, C., Chen, A.P.S., Christanto, H.J. 2023. YOLOv7 for face mask identification based on deep learning. 15th International Conference on Computer and Automation Engineering (ICCAE), 193–197.

Dewi, C., Chen, A.P.S., Christanto, H.J. 2023. Recognizing similar musical instruments with YOLO models. Big Data and Cognitive Computing. 7, 94.

Dewi, C., Chen, R.-C., Yu, H., Jiang, X. 2023. XAI for image captioning using SHAP. Journal of Information Science and Engineering. 39, 711–724.

Dorigo, M., Maniezzo, V., Colorni, A. 1996. Ant system: Optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 26, 29–41.

Eberhart, R., Kennedy, J. 1995. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 39–43.

Ge, Y., Liu, H., Du, J., Li, Z., Wei, Y. 2023. Masked face recognition with convolutional visual self-attention network. Neurocomputing. 518, 496–506.

Holland, J.H. 1992. Adaptation in natural and artificial systems. Ann Arbor, MI: University of Michigan Press.

Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E. 2007. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Technical Report, 7-49.

Huang, B., Wang, Z., Wang, G., Jiang, K., He, Z., Zou, H., Zou, Q. 2021. Masked face recognition datasets and validation. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 1487–1491.

Kai, W., Shuo, W., Jianfei, Y., Xiaobo, W., Baigui, S., Hao L., Yang, Y. 2021. Mask aware network for masked face recognition in the wild. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 1456–1461.

Kelly, M.D. 1970. Visual identification of people by computer. Techical Report AI-130.

Liu, C., Wechsler, H. 2002. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image processing, 11, 467–476.

Liu N., VGGFace2-HQ, 2021. Retrieved 2023-06-06 from https://github.com/NNNNAI/VGGFace2-HQ

NEC. Face recognition: Biometric authentication. Retrieved 2023-06-06 from https://www.nec.com/en/global/solutions/biometrics/face/

Negnevitsky, M., Translated by Hsieh, C.H., Liao, H.C., Lee, L.W. 2019. Artificial intelligence: A guide to intelligent systems, 3/e. Pearson Education Taiwan Ltd. 7-1, 7–44.

Ngan, M.L., Grother, P.J., Hanaoka, K.K. 2020. Ongoing face recognition vendor test (FRVT) Part 6B: Face recognition accuracy with face masks using post-Covid-19 algorithms. NIST Interagency/Internal Report (NISTIR).

Schroff, F., Kalenichenko, D., Philbin, J. 2015. FaceNet: A unified embedding for face recognition and clustering. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 815–823.

Shukla, R.K., Tiwari, A.K. 2023. Masked face recognition using mobilenet v2 with transfer learning. Computer Systems Science and Engineering, 45, 293–309.

Sirovich, L., Kirby, M. 1987. Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A Optics and image science, 4, 519–524.

Song, L., Gong, D., Li, Z., Liu C., Liu, W. 2019. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. arXiv, arXiv: 1908. 06290.

Takacs, B. 1998. Comparing face images using the modified Hausdorff distance. Pattern Recognition, 31, 1873–1881.

Thales Group. 2020. Biometric technology to control COVID-19. Retrieved 2023-06-06 from https://www.thalesgroup.com/en/spain/magazine/biometric-technology-control-covid-19

Wang, Z., Huang, B., Wang, G., Yi, P., Jiang, K. 2023. Masked face recognition dataset and application. IEEE Transactions on Biometrics, Behavior, and Identity Science. 5, 298–304.

WHO 2023. Coronavirus disease (COVID-19): Masks. Retrieved 2023-06-06 from https://www.who.int/news-room/ questions-and-answers/item/coronavirus-disease-covid-19-masks

William, I., Ignatius Moses Setiadi, D.R., Rachmawanto, E.H., Santoso, H.A., Sari, C.A. 2019. Face recognition using facenet (survey, performance test, and comparison). In 2019 Fourth International Conference on Informatics and Computing (ICIC), 1-6.

Wu, C., Fu, X., Pei, J., Dong, Z. 2021. A novel sparrow search algorithm for the traveling salesman problem. IEEE Access, 9, 153456–153471.

Xue, J., Shen, B., 2020. A novel swarm intelligence optimization approach: Sparrow search algorithm. Systems Science & Control Engineering, 8, 22–34.

X-zhangyang, Real-world-masked-face-dataset, 2020. Retrieved 2023-06-06 from https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset

Yang, X.S., 2009. Firefly algorithms for multimodal optimization. SAGA 2009, Lecture Notes in Computer Science, 5792.

ZENBOT99, VGGface2_HQ_cropped, 2022. Retrieved 2023-06-06 from https://www.kaggle.com/datasets/zenbot99/vggface2-hq-cropped

Zeng, S., Gou, J., Deng, L. 2017. An antinoise sparse representation method for robust face recognition. Expert Systems with Applications: An International Journal, 82, 1–9.