Impact of Resolution and Resize Process on Face Recognition Accuracy with GAN-Generated Images

Yu Xing

{21022100042@stu.xidian.edu.cn}

School of Electronic Engineering, Xidian University, Xi'an, 710126, China

Abstract. This paper discusses the challenges facial recognition systems face in recognizing images generated by generative adversarial networks (GANs) and then proposes solutions by simple image processing methods. The study highlights that facial recognition models behave completely differently when dealing with low-resolution, super-resolution, or resized facial images and that the resizing method specifically affects the success rate of GANs. Changing the resolution also seems to affect the attack's success rate slightly.

Keywords: Generative Adversarial Networks, FaceNet, Image Post-processing

1 Introduction

Face recognition has evolved rapidly over the last few decades due to the aggressive development of algorithms, the availability of large databases of face images, and advances in methods for evaluating the performance of face recognition algorithms [1][2]. Many face recognition models have been developed to recognise high-quality faces and have been shown to achieve human performance levels on tasks [3].

However, facial recognition systems still have many obstacles to overcome. For example, these systems do not perform well in recognising blurred faces, obstructed faces or faces in complex lighting conditions. Coincidentally, the security domain, which relies heavily on the performance of facial recognition systems, often captures images of faces through surveillance cameras that are unclear, obscured, or taken under challenging lighting conditions [4]. What makes it even worse is the advent of Generative Adversarial Networks (GANs), which can create completely artificial images from scratch [5]. It has been shown that GANs can create fake face images to deceive humans, allowing one to modify the context and semantics of an image in a very realistic way [6]. Therefore, the misuse of GANs can negatively affect the trustworthiness of social media users and pose a serious threat to the security domain, leading to social concerns and property damage. It is important to come up with ways to combat GANs.

Scholars have done much-related work to address these challenges, but all of it is relatively complex. For example, they have studied the co-occurrence matrix of colors and investigated the use of deep-learning neural networks to recognize GAN-generated images. However, only a limited number of studies investigate the effect of post-processing on the success rate of face recognition systems in recognising GAN-generated images.

CONF-MLA 2024, November 21, Adana, Turkey

DOI 10.4108/eai.21-11-2024.2354613

This study examines how face recognition models can be better protected against GANs. It assumes that the resolution and proportion of the face to the image will directly affect the recognition model's accuracy in discriminating between true and false images. The dataset used in the study is CelebA, which contains 202,599 face images of 10,177 people [7][8], and FaceNet is the face recognition model we used.

2 Related Work

Two aspects are relevant to our work: attack detection and face recognition models.

2.1 Detection of attacks: artifacts and colour effect of GAN

A common approach to defending against adversarial attacks is to use deep learning neural networks [9] to detect generated images. Since GAN-generated images have specific artifacts, a defender can train a detection model to capture these artifacts to prevent adversarial attacks. Common detection networks include Convolutional Neural Network (CNN) layers, pooling layers, feature embedding layers, and Long-Short Term Memory (LSTM) or Multi-Layer Perceptron (MLP) layers. These models have shown high accuracy in detecting generated images [10].

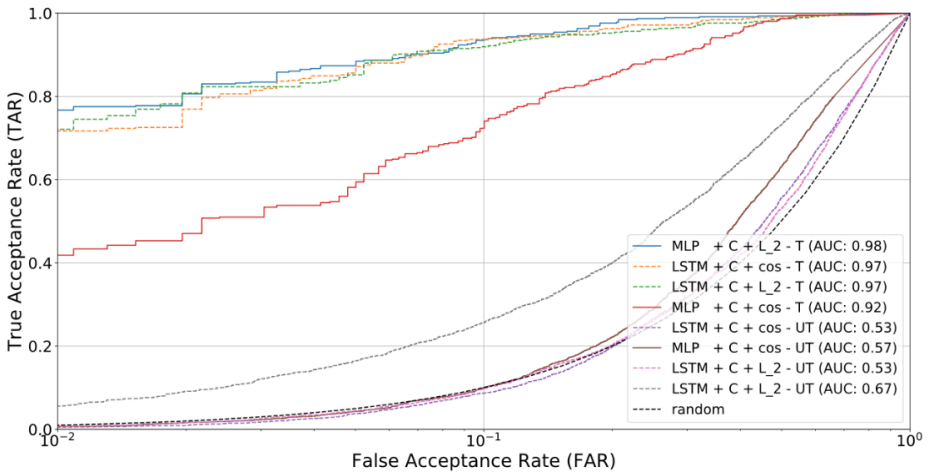


Fig. 1. ROC curves of detection models

However, since artifacts are highly correlated with the type of GAN used, these models will overfit the training data and will not work well if an attacker uses some new type of GAN. So, in conclusion, these models are not reliable.

By looking at recent GAN models, it was found that it can create extremely high-quality images with almost imperceptible spatial errors, so reconstructing relationships between colors may be more effective. More specifically, CNN detection methods detect cross-band and gray-scale co-occurrences, which are estimated separately on individual color bands. The experimental results in [11] rely on the well-known StyleGAN model, which creates higher-quality images than the

ProGAN model in an attempt to make the detection task more difficult. Nevertheless, the CNN detectors obtained from this training method achieve almost optimal detection performance [11].

2.2 Face recognition models: FaceNet

Many studies have introduced face recognition models which embed images into hyperspace. One of the widely used models is FaceNet developed by Google.

FaceNet uses a multi-task cascaded convolutional network (MTCNN) that combines three sub-networks in a cascade for efficient and accurate face detection and alignment, enabling recognition of facial features in different lighting and angles, capturing feature points from the face and embedding these features into a hypersphere using a deep network. The learning process uses a ternary loss, which describes the relationship between two images of a person (anchor point and positive) and an image of another person (negative). Through this learning process, the model may eventually have the most appropriate embedding capability [12].

Fig. 2. Triplet loss of FaceNet

$$\sum_{i}^{N}\left[\left\|f(x_i^a)-f\left(x_i^p\right)\right\|_2^2-\left\|f(x_i^a)-f(x_i^n)\right\|_2^2+\alpha\right] \tag{1}$$

3 Methodology

3.1 Generate fake images dataset using StarGAN-v2

To test the robustness of the face recognition system against GAN-generated images, we chose StarGAN-v2 as our generator. StarGAN-v2 abstracts the features of two images and generates a new image with all these features [13], with the ability to generate realistic faces. We chose the CelebA dataset as the source for generating faces, which is a high-quality face image dataset, and most of the identities in this dataset have more than one face image. They can be handed over to GAN learning using photos of the same face from different angles to generate more realistic fake face images. So, this dataset is very suitable for face generation. For each identity in this dataset, we randomly select two images and hand them over to the GAN to generate a fake image, which is finally collected as our fake image dataset.

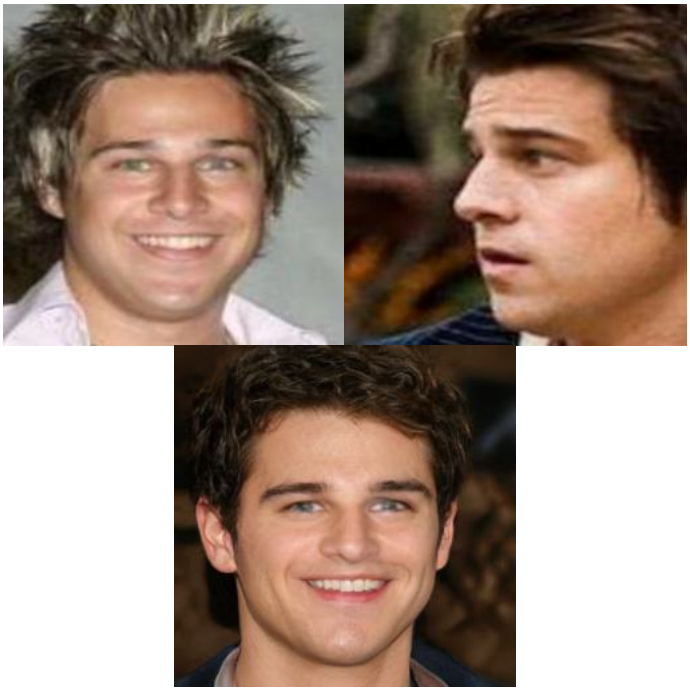


Fig. 3. Generate face images (image on the right side is generated)

3.2 Datasets Processing

After obtaining the dataset of dummy face images generated by StarGAN-v2, we apply four different algorithms to obtain various datasets for our experiments.

We first take the resolution of the initial image as a (high, wide) tuple, then set a Gaussian blurring kernel with a width of 1, and then blur the image's resolution to 2, 4, and 6 times the original. In addition, we use Real-ESRGAN [14], a deep learning-based image super-resolution model designed to upgrade low-resolution images to higher resolution and improve image quality and details to super-resolve the initial image dataset and the blurred dataset. After this step blurred group, high-resolution group, and blur, then super-resolution group are obtained.

In addition, we resized the initial image to change the proportion of faces in the image. For example, we cropped the image by 5% and 10% from each side, reducing the image area by 19% and 36% and increasing the proportion of faces in the image. In addition, we mirror the image symmetrically on all sides to expand each side by 5% and 10%, respectively. This increases the area by 21% and 44% and decreases the proportion of faces in the image. After this step, we obtain the cropped group and the expanded group.

3.3 Recognize fake images with FaceNet

Now, for each individual, there are three images: two are real, and one is a fake image generated by GAN learning these two real images. We take the three images for each individual in the dataset and randomly select two images to form an experimental set, which we mark as True if

both images are real and False if one is fake and the other is real. We then use FaceNet to embed all these images and compute the difference between the two images for each identity. These different data will be further used for model evaluation. We use FaceNet pre-trained on VGGFace2.

4 Experiment Result

4.1 Performance Criteria

Confusion matrix is a widely used tool in machine learning and statistics, mainly for evaluating the performance of classification models, especially in classification tasks with supervised learning. It visualizes the comparison between model predictions and actual labels in a matrix. This visualization not only clearly shows the number of correctly classified samples but also reveals the specifics of classification errors, thus helping us to gain a deeper understanding of the model's performance. In our experiments, the confusion matrix generated is shown in Table 1:

Table 1. Confusion matrix for face recognition: The table represents the model's performance regarding True Positives, False Positives, False Negatives, and True Negatives for predicting whether two images are of the same face or different faces.

	Actual	
Predicted	**Different Face**	**Same Face**
Different Face	True Positive	False Positive
Same Face	False Negative	True Negative

For example, a False Positive means that the two faces were taken from the same person, but the model regarded them as different people. In our experiment, the person was GAN-generated.

Regarding selecting the threshold in the confusion matrix, we adopt the 1.1 threshold commonly used in FaceNet, which is more representative in real life. On the other hand, we pay more attention to the concern of model attack performance. Specifically, we are particularly concerned about the performance of GAN-generated fake images when recognized by the face recognition system. We define an important metric called Attack Success Rate (ASR), which indicates how many generated fake images are not successfully recognized by the face recognition system. To calculate this rate, we compare the total number of fake images to the number of unrecognized ones with the formula (2). In this way, we can quantify the model's vulnerability in the face of potential attacks, which in turn guides further optimization and enhancement of the robustness of the model.

$$\text{AttackSuccessRate} = \frac{FN}{TP + FN} \tag{2}$$

In addition, we will evaluate the overall recognition accuracy of the model. To fully understand the model's performance, we will perform a more in-depth analysis using the receiver operating characteristic (ROC) curve in addition to using the confusion matrix. The ROC curve is an effective tool that helps us visualize how the model performs under different thresholds. A significant advantage is that ROC curves perform well when dealing with unbalanced data. Even

if the ratio of positive to negative samples changes, the ROC curve still provides a stable and reliable performance assessment, enabling us to understand the model's effectiveness in real-world applications more accurately. In addition, the area under the ROC curve (AUC) can also be used as a quantitative metric to help us compare the performance of different models. Therefore, by combining the confusion matrix analysis and the ROC curve, we can comprehensively assess the recognition ability of the model and ensure its validity and reliability in practical applications.

4.2 Resolution analysis

As shown in Figure 4, the performance of FaceNet does not change significantly when processing the four groups of Images: High-resolution blurred x2 and blurred x4. This indicates that the model can maintain a relatively stable recognition ability even though the image is affected by blurring to a certain extent. However, when the degree of image blurring reaches x6, it is obvious from the change of the dark red line in the figure that the performance of the model shows a significant decrease, indicating that at higher blurring degrees, the loss of image information makes the model's recognition ability seriously affected.

This observation is further supported by the data on attack success rates in Table 2. The data shows that in the Raw, High, x2, and x4 groups, the attack success rate of the model fluctuates within a range of 63%, with a difference of only two percentage points, demonstrating the stability of the model under these conditions. However, once the image blurring level is raised to x6, the attack success rate drops significantly, a change that reaffirms the model's vulnerability in the face of highly blurred images. It is also worth noting that the model's recognition accuracy slightly improves compared to the x4 group, which may be related to the change in the blur level and the model's ability to extract certain features.

Subsequently, we increased the resolution of the blurred images and showed the corresponding ROC curves in Figure 5. By comparison, we found that the model's performance hardly changed after increasing the resolution. This suggests that despite the improved resolution, FaceNet's recognition ability is still limited and fails to improve significantly when dealing with highly blurred images.

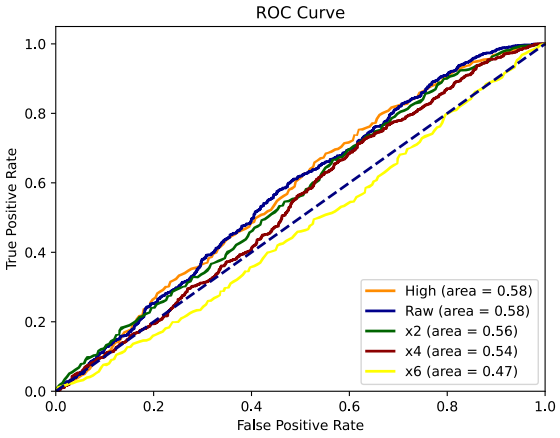


Fig. 4. ROC curve of different resolution groups

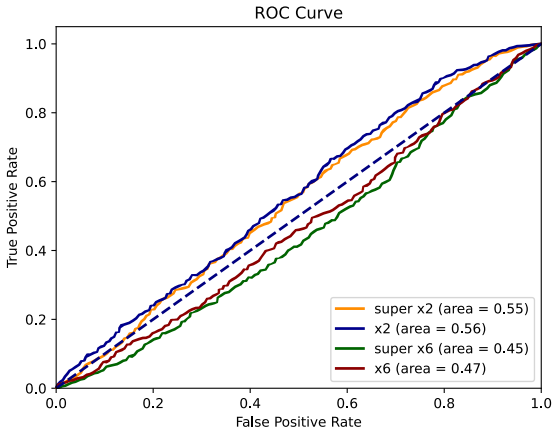


Fig. 5. ROC curve compared between raising the resolution of the blurred images group and blurred image group

Table 2. Accuracy and Attack success rate in different resolution

	Accuracy	**Attack Success Rate**
high	47.18%	65.35%
raw	48.67%	61.72%
x2	46.68%	63.57%
x4	45.56%	62.43%
x6	47.18%	54.73%

4.3 Resizing analysis

In addition to resolution, the proportion of faces in the image is an important factor to consider carefully. To study this in-depth, we performed two main operations on the dataset: first, removing non-face boundary regions, and second, expanding part of the background. These two operations aim to enhance the clarity and saliency of faces in the image, thus potentially improving the recognition performance of the model. After completing these processes, a new dataset was formed, and experiments were conducted to evaluate the face recognition effect.

In Figure 6, the results are visible. For the cropped group, the recognition performance is significantly better than the other group because the face occupies a larger proportion of the image. This indicates that when faces occupy a larger proportion of the image, the FaceNet model can extract features more efficiently and improve recognition accuracy. In addition, among the cropped groups, the group with more cropping, indicated by the orange line, exhibits better performance than the group with less cropping, indicated by the dark blue line, further emphasizing the effect of the face proportion on the model performance.

The group with extended backgrounds also exhibits a similar trend. It can be observed that the model's performance performs relatively poorly in the group with more extensions. This may be because the over-extension of the background distracts the model from the face features,

which reduces the recognition effect. These experimental results suggest that the relative size and proportion of faces in an image play a key role in influencing the recognition performance.

Since FaceNet's output under different processing conditions significantly differs, the accuracy and attack success of analyzing these datasets under the same thresholds become non-comparable. This means that relying solely on these metrics for model performance comparisons is insufficient, and multiple factors, such as the actual proportion of the face occupying the image and so on, must be considered comprehensively for their impact on model performance.

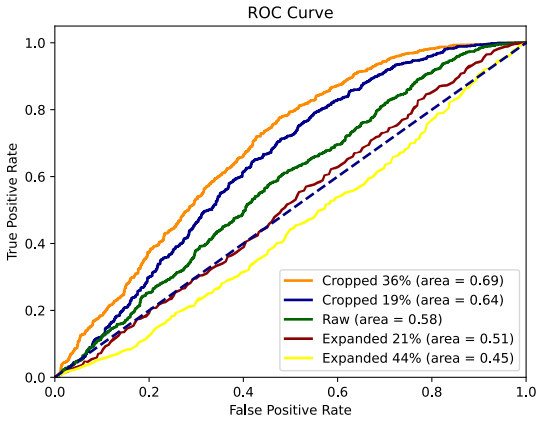
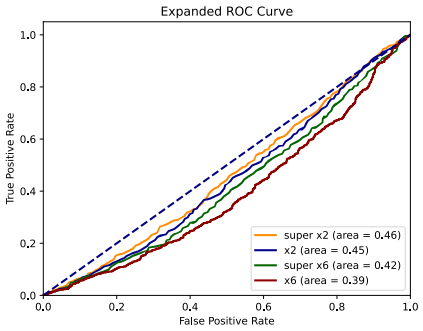


Fig. 6. ROC curve of different processing groups

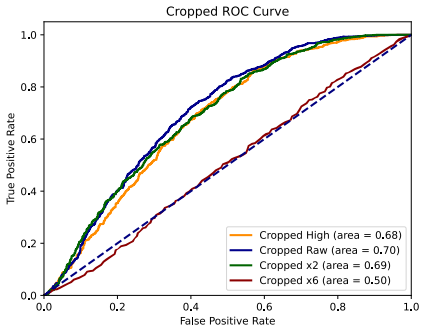
4.4 Within-group analysis

When we turn to within-group comparisons, it is clear from Figure 7 that the resolution conclusions within the same group are consistent with our previous conclusions. This suggests that the performance of the models maintains some stability across resolutions. Specifically, in the High, Raw, x2, and x4 groups, the performance of each model shows only negligible fluctuations. In the x6 group, a significant performance drop is observed. This degradation is not only reflected in the overall performance of the model, but also in the significant decrease in the attack success rate. Similar to the previous findings, a slight increase in accuracy is observed in the x6 group compared to the x4 group.

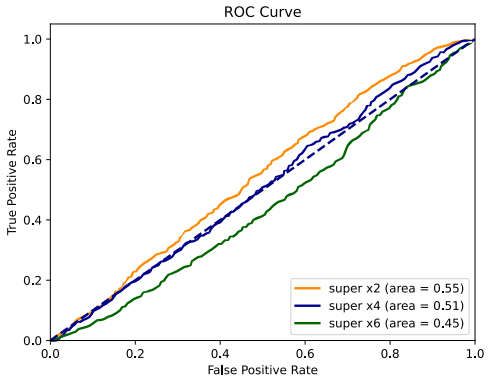
The results of the within-group comparisons cross-corroborate the validity of the conclusions in 4.2.



(a)ROC curve within expanded group



(b)ROC curve within the cropped group



(c)ROC curve within raising the resolution of blurred images group

Fig. 7. Different within-group ROC curve

Table 3. Accuracy and Attack success rate within expanded group and cropped group

	Accuracy	**Attack Success Rate**
Expanded		
raw	42.45%	65.32%
x2	42.99%	62.29%
x4	44.77%	55.44%
x6	46.06%	48.17%
Cropped		
high	47.01%	69.93%
raw	49.92%	67.23%
x2	50.29%	65.83%
x6	51.16%	47.74%

5 Conclusion

Based on the results presented above, the conclusions are obtained as below:

(1) Different resolutions impact the performance of face recognition systems when recognizing GAN-generated images. At the group blurred x6, the performance of the face recognition system shows a significant decline. Meanwhile, the Attack Success Rate decreases noticeably. Therefore, we believe that in security-sensitive areas, reducing the resolution of input images could be useful to enhance the defensive abilities of face recognition systems, even though this may lead to an overall decline in model performance.

(2) blurring an image and then raising its clarity has almost no effect on the model.

(3) It is concluded that resizing has a substantial and clearly observable effect on Facenet's performance.

6 Future Work

Since our resizing group only performed simple background cropping and expanding of the images and did not analyze the proportion of the face in the image after processing, we cannot determine the relationship between the face recognition system's performance and the proportion of the face in the image. This will be a direction for our further research.

Secondly, the face recognition system we used in our experiments has been FaceNet, which may not be universal. At the same time, AdaFace is better at handling low-resolution images[15] and may bring different results, which will be the direction of our further work.

In addition to post-processing images to affect the model's performance, analyzing image color bands has also been proven effective[11]. By observing recent GAN models, it is evident that they can generate extremely high-quality images with spatial errors that are almost invisible. In comparison, establishing relationships between color bands may be more effective. The established results can then be fed into a CNN, as it can be used to analyze some feature points in the color bands that are not obvious to the human eyes. This can be an important direction for future research.

References

[1] Tolba, A. S., El-Baz, A. H., & El-Harby, A. A. (2006). Face recognition: A literature review. *International Journal of Signal Processing*, *2*(2), 88-103.

[2] Flach, P. (2019, July). Performance evaluation in machine learning: the good, the bad, the ugly, and the way forward. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 9808-9814).

[3] O'Toole, A. J., & Castillo, C. D. (2021). Face recognition by humans and machines: three fundamental advances from deep learning. *Annual Review of Vision Science*, *7*(1), 543-570.

[4] Verschae, R., Ruiz-del-Solar, J., & Correa, M. (2008). Face recognition in unconstrained environments: A comparative study. In *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*.

[5] Durgadevi, M. (2021, July). Generative Adversarial Network (GAN): A general review of different variants of GAN and applications. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1-8). IEEE.

[6] Marra, F., Gragnaniello, D., Cozzolino, D., & Verdoliva, L. (2018, April). Detection of gan-generated fake images over social networks. In *2018 IEEE conference on multimedia information processing and retrieval (MIPR)* (pp. 384-389). IEEE.

[7] Liu, Z., Luo, P., Wang, X., & Tang, X. (2018). Large-scale celeb faces attributes (celeb) dataset. *Retrieved August*, *15*(2018), 11.

[8] Lingenfelter, B., Davis, S. R., & Hand, E. M. (2022, October). A quantitative analysis of labeling issues in the celeba dataset. In *International Symposium on Visual Computing* (pp. 129-141). Cham: Springer International Publishing.

[9] Guo, H., Hu, S., Wang, X., Chang, M. C., & Lyu, S. (2022). Robust attentive deep neural network for detecting gan-generated faces. *IEEE Access*, *10*, 32574-32583.

[10] Massoli, F. V., Carrara, F., Amato, G., & Falchi, F. (2021). Detection of face recognition adversarial attacks. *Computer Vision and Image Understanding*, *202*, 103103.

[11] Nowroozi, E., & Mekdad, Y. (2023). Detecting high-quality GAN-generated face images using neural networks. In *Big Data Analytics and Intelligent Systems for Cyber Threat Intelligence* (pp. 235-252). River Publishers.

[12] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).

[13] Choi, Y., Uh, Y., Yoo, J., & Ha, J. W. (2020). Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8188-8197).

[14] Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-organ: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1905-1914).

[15] Kim, M., Jain, A. K., & Liu, X. (2022). Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18750-18759).