Optimizations in Applied Machine Learning
Research Article | Volume 3 | Issue 3 | Dec 2023
Received: 2 Nov 2023 | Revised: 30 Nov 2023
Accepted: 10 Dec 2023 | Published Online: 15 Dec 2023

Mitigating FGSM-Based White-Box Attacks Using Convolutional Autoencoders for Face Recognition

Jiahuai Ma[1], Alan Wilson[2*]
[1]University of Florida, Gainesville, FL 32611, USA
[2*] Corresponding, Intact Financial Corporation, Toronto, Ontario M5H 1H1, Canada
Email: alan.wilson@intact.net

Abstract: Facial recognition technology has become a crucial biometric tool in various applications, from security systems to personalized user experiences. However, its susceptibility to adversarial attacks, such as FGSM-based white-box attacks, raises significant concerns about its reliability and robustness. This paper proposes a novel framework that leverages a convolutional autoencoder to mitigate the effects of adversarial perturbations. The FGSM method generates imperceptible perturbations to input images, which, while invisible to the human eye, significantly degrade model performance. The autoencoder reconstructs perturbed images to reduce the impact of adversarial noise, improving the system's resilience. MobileNetV2 serves as the backbone model for facial recognition, with cosine similarity used for face matching. Experimental results demonstrate that the equal error rate (EER) increases under FGSM attacks but improves after reconstruction, reducing EER from 0.36 to 0.32 (FGSM-0.1) and from 0.37 to 0.31 (FGSM-1). While the proposed approach enhances robustness, further work is needed to address stronger adversarial attacks and evaluate performance on larger datasets.

Keywords: *Face recognition; White-box attack; FGSM; Autoencoder.*

1. Introduction

Facial recognition technology has emerged as one of the most prominent and widely adopted tools in the field of biometrics [1] [2] [3] . It refers to the automated process of identifying or verifying individuals by analyzing their facial features from images or video streams. This technology plays a critical role in numerous applications, including security systems, access control, surveillance, personalized user experiences, and financial transactions [4] [5] [6] [7] . For instance, facial recognition has been widely deployed in smartphones, border control systems, and payment platforms, revolutionizing how individuals interact with digital systems. Its ability to provide fast, contactless, and accurate identification has positioned it as a fundamental component in the modern digital landscape. However, alongside its increasing importance, concerns regarding the robustness and security of facial recognition models have gained significant attention.

Traditional methods [8] [9] [10] for facial recognition were primarily based on handcrafted features such as Local Binary Patterns (LBP) [11] , Principal Component Analysis (PCA) [12] , and Histogram of Oriented Gradients (HOG) [13] . These methods rely heavily on feature

extraction and are sensitive to variations in lighting, pose, and occlusions, limiting their effectiveness in real-world scenarios. With the advent of Artificial Intelligence (AI) [14] [15] [16] , particularly Deep Learning [17] [18] , facial recognition has undergone a paradigm shift. Deep neural networks, such as Convolutional Neural Networks (CNNs) [19] [20] [21] , have demonstrated remarkable performance in extracting high-level features from facial images, significantly improving recognition accuracy. Models like VGGFace, FaceNet, and ArcFace have set new benchmarks for performance, enabling robust and scalable face recognition systems.

While AI has significantly enhanced the capabilities of facial recognition, it has also introduced new challenges, particularly concerning security and adversarial robustness [22] [23] [24] . For instance, Xiong et al. highlighted this issue in their study on a distributed data parallel acceleration-based generative adversarial network for fingerprint generation [22] . One of the most pressing issues is the vulnerability of AI models to adversarial attacks, which involve adding imperceptible perturbations to input images to deceive the model. Among these, white-box attacks, such as the Fast Gradient Sign Method (FGSM) [25] [26] , pose a severe threat. In a white-box scenario, the attacker has full access to the model architecture, parameters, and gradients, allowing them to generate highly effective adversarial examples. For facial recognition systems, this can result in catastrophic consequences, including identity impersonation, unauthorized access, and reduced model reliability.

The FGSM-based white-box attack works by perturbing the input image in the direction of the gradient of the loss function, forcing the model to misclassify the input. Although these perturbations are often imperceptible to the human eye, they are sufficient to manipulate the output of the AI model. This highlights a critical security flaw in AI-based facial recognition systems: their susceptibility to adversarial attacks. Traditional defenses, such as input preprocessing, adversarial training, or model regularization, have shown limited success in mitigating such attacks, particularly for real-world applications that require high levels of accuracy and robustness.

To address this challenge, this paper proposes a novel framework shown in Figure 1. As illustrated in the figure, the framework leverages a Convolutional Autoencoder to reduce the effect of adversarial perturbations on input images. The autoencoder acts as a defense mechanism by reconstructing adversarial inputs into clean, noise-free images, thereby improving the robustness of the facial recognition system. The process begins with an original image that undergoes a white-box FGSM-based attack, resulting in an adversarial image. This adversarial image is passed through a convolutional autoencoder, which extracts latent variables and reconstructs a clean version of the image, effectively reducing the effect of the attack. The reconstructed image is then fed into a pre-trained backbone model (e.g., a deep convolutional neural network) to extract feature vectors. The similarity between the reconstructed image and the image stored in the database is computed using cosine similarity. If the similarity exceeds a certain threshold, the images are classified as belonging to the same identity; otherwise, they are deemed different.

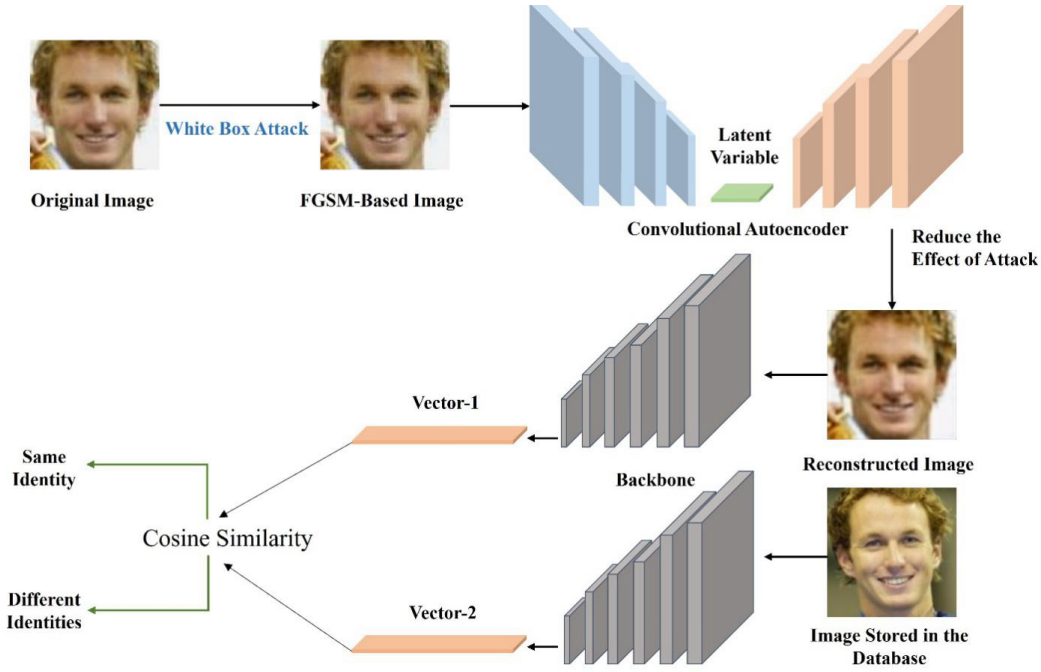


Figure 1. The architecture of the proposed framework.

2. **Literature Review**

2.1 Face recognition

Facial recognition has been an active research area for several decades [27] [28] [29] , evolving from traditional approaches to modern deep learning-based techniques due to the advancements of these technologies in many domains [30] [31] [32] [33] . Early methods focused on handcrafted feature extraction, which, despite their simplicity, laid the foundation for subsequent advancements. Turk and Pentland pioneered the use of Principal Component Analysis (PCA) for facial recognition, introducing the concept of Eigenfaces to reduce dimensionality and extract discriminative features [34] . However, PCA-based methods were highly sensitive to variations in lighting, pose, and facial expressions. To address these challenges, researchers introduced Linear Discriminant Analysis (LDA) [35] and Local Binary Patterns (LBP) [36] , which improved robustness to certain variations but were still limited by their reliance on manually designed features.

The advent of deep learning revolutionized facial recognition systems, enabling automatic feature extraction from raw images. Taigman et al. introduced DeepFace [37] , one of the first deep learning models for face recognition, which leveraged a deep neural network to achieve near-human performance. Schroff et al. further improved upon this with FaceNet [38] , a model that utilized a triplet loss function to learn a compact embedding space, enabling both recognition and clustering. The use of deep embeddings significantly enhanced the scalability and accuracy of facial recognition systems. To address the challenge of intra-class variations and inter-class similarities, subsequent studies introduced margin-based loss functions. Liu et al. [39] proposed the SphereFace model, which incorporated an angular softmax loss to increase the discriminative power of deep embeddings. Building on this, Deng et al. presented ArcFace [40] , which introduced an additive angular margin loss, further improving face verification accuracy by enforcing a clearer

decision boundary in the embedding space. These models set new benchmarks for performance on public datasets such as Labeled Faces in the Wild (LFW) and MegaFace.

3. **Method**

3.1 Dataset preparation

The dataset used in this study is sourced from Kaggle and consists of a total of 1,105 images representing 248 individuals, including both male and female subjects. The images are in RGB format with varying numbers of impressions per individual, but no individual has more than 10 impressions. The original resolution of each image is 128x128 pixels. For the purpose of training and evaluation, 70% of the dataset is used as the training set, while the remaining 30% is reserved for testing. Figure 2 illustrates a subset of sample images from the dataset.

Figure 2. The samples of original datasets.

3.2 FGSM-based white-box attack

The FGSM-based white-box attack [41] [42] is a well-known adversarial attack method designed to deceive deep learning models by introducing subtle perturbations to input images. In a white-box scenario, the attacker has complete access to the model's architecture, parameters, and gradients, which allows for the generation of adversarial examples that can significantly alter the model's predictions. FGSM works by perturbing the input image in a way that maximizes the model's loss, pushing the model to misclassify the image while keeping the perturbations nearly imperceptible to the human eye.

In this study, FGSM was employed to generate adversarial images with two levels of perturbation strength, 0.1 and 1. Despite the perturbations, it is visually difficult to distinguish these adversarial images from the original ones. This is because the FGSM attack introduces changes at a pixel level, often too subtle for the human visual system to detect, especially at lower perturbation strengths. However, these small changes are sufficient to mislead deep learning models, exposing their vulnerability to adversarial attacks. Figure 3 presents examples of images after applying FGSM.

Figure 3. The samples of datasets preprocessed by FGSM.

3.3 Convolutional autoencoders for mitigating FGSM-based white-box attacks

Convolutional Autoencoders (CAEs) [43] [44] [45] are a type of neural network designed to reconstruct input images by compressing and restoring them through an encoder-decoder architecture. They are particularly effective in tasks like image denoising and reconstruction, which makes them well-suited for mitigating adversarial perturbations introduced by FGSM-based white-box attacks. In this study, the CAE architecture is carefully designed to process perturbed images and restore them to a clean version, allowing the facial recognition system to perform accurately despite adversarial interference.

The architecture of the convolutional autoencoder used in this work is composed of an encoder-decoder structure designed to process RGB images of size 128x128. The encoder begins with a 3x3 convolutional layer with 32 filters, applied using a ReLU activation function and same padding, to extract low-level spatial features from the input image. To reduce spatial dimensions while retaining critical features, a 2x2 max pooling layer with stride 2 is employed. The process is repeated with another 3x3 convolutional layer, this time with 64 filters, followed by another max pooling operation. By successively reducing the spatial resolution, the encoder compresses the input image into a compact latent representation, which contains the most salient features necessary for reconstruction. The decoder mirrors the encoding process and restores the image to its original dimensions by gradually increasing the spatial resolution. It begins with a 3x3 convolutional layer with 64 filters, followed by an upsampling layer that doubles the spatial size of the latent representation. A subsequent 3x3 convolutional layer with 32 filters refines the reconstruction, and another upsampling operation brings the image closer to its original dimensions. Finally, a 3x3 convolutional layer with a sigmoid activation function outputs the reconstructed image, ensuring that pixel values remain within the range of 0 to 1. Through this process, the decoder effectively removes the adversarial noise while preserving the visual integrity of the input.

3.4 MobileNet-based model for face recognition

MobileNet is a lightweight Convolutional Neural Network (CNN) designed to deliver efficient performance on resource-constrained devices [46] [47] . Unlike traditional deep networks, MobileNet reduces computational complexity by using depthwise separable convolutions, which decompose standard convolutions into depthwise and pointwise operations. This approach significantly reduces the number of parameters and computations while maintaining high accuracy, making it ideal for real-time facial recognition tasks.

In this study, we utilize MobileNetV2 as the backbone of our CNN model for face recognition. MobileNetV2 is an improved version of MobileNet, which introduces inverted residual blocks and linear bottlenecks to enhance the network's efficiency and accuracy. The model is initialized with pre-trained weights from the ImageNet dataset, allowing it to leverage features learned on large-scale image data. The base model excludes the fully connected top layers, and its weights are frozen during training to prevent updates and retain the pre-trained features. On top of the MobileNetV2 backbone, a new classification head is built. The structure begins with the output of MobileNetV2 being passed through a global average pooling layer to reduce feature maps to a lower-dimensional vector. A dropout layer with a rate of 0.2 is added to prevent overfitting, followed by a dense layer that outputs a 128-dimensional feature vector. This vector serves as a compact and discriminative representation of the input face image. Finally, a softmax classification layer maps these features to the corresponding class labels during the training phase.

During training, the entire network is treated as a classification model, where each individual face is assigned a unique class label. After training is complete, the model can be used for face matching by extracting the 128-dimensional feature vector from the dense layer. To evaluate the similarity between two input face images, their corresponding feature vectors are compared using cosine similarity. This metric measures the angular distance between the two vectors, where higher similarity scores indicate that the images belong to the same identity, while lower scores suggest different identities.

3.5 Implementation details

In this study, the models were implemented using TensorFlow. The convolutional autoencoder and the MobileNet-based CNN model were trained separately for 20 epochs and 30 epochs, respectively. Both models were optimized using the Adam optimizer to ensure efficient convergence. For the autoencoder, mean absolute error (MAE) was used as the loss function to minimize the reconstruction error and effectively remove adversarial perturbations. For the MobileNet-based model, categorical cross-entropy was employed as the loss function during training, treating the task as a classification problem. During the evaluation stage, equal error rate (EER) was used as the performance metric for face matching, providing a reliable measure of the system's accuracy and robustness.

4. **Experimental results and discussion**

4.1 The performance of the model

Figure 4 illustrates the training curves of the MobileNet model used for face recognition. The left graph shows the training accuracy over 30 epochs, where the accuracy increases steadily as training progresses. Initially, the model starts with a low accuracy, but it improves consistently, reaching over 90% by the final epoch, indicating that the model is learning effectively. The right graph depicts the training loss, which decreases significantly as the number of epochs increases. At the beginning of training, the loss is relatively high, but it drops sharply within the first few epochs. As training continues, the loss decreases gradually, approaching convergence near the final epochs. This demonstrates that the MobileNet model is optimizing its weights effectively and minimizing the classification error during training.

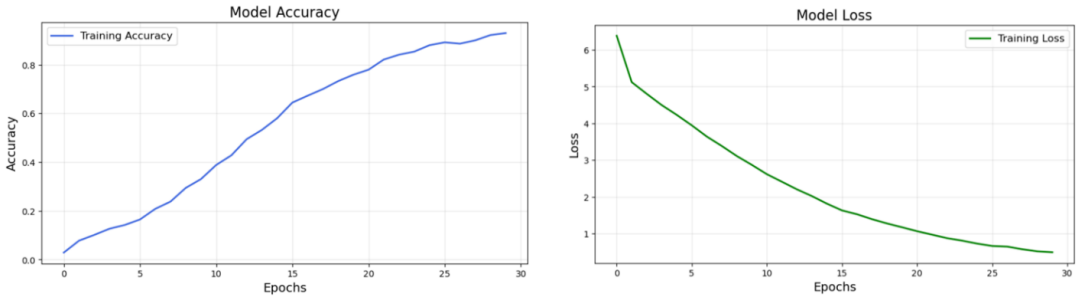
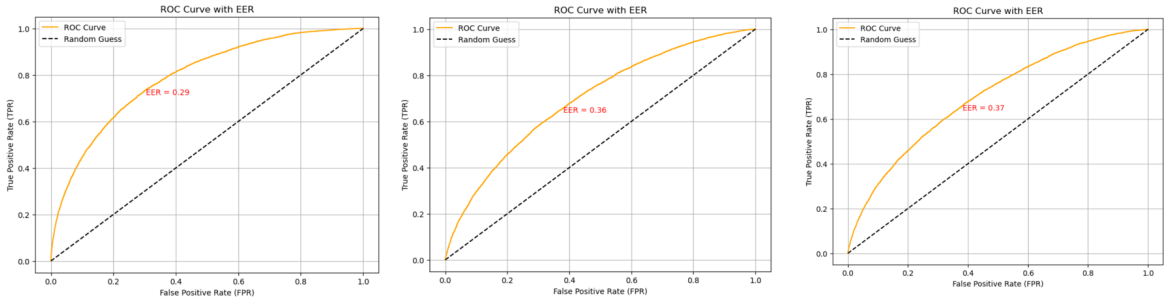


Figure 4. The training curves of the MobileNet model.

The results of the model's performance under different conditions are shown in Figure 5 and Table 1. The equal error rate (EER) is evaluated on the test set under three scenarios: without FGSM perturbation, with FGSM perturbation of strength 0.1, and with FGSM perturbation of strength 1. Without FGSM, the model achieves an EER of 0.29, indicating strong recognition performance in

clean conditions. When FGSM perturbations with strengths of 0.1 and 1 are applied, the EER increases to 0.36 and 0.37, respectively. These results clearly demonstrate that even subtle perturbations generated by FGSM can significantly impact the model's ability to distinguish between identities. While the perturbed images appear visually identical to the original images to the human eye, the model perceives these adversarial perturbations, leading to a decline in recognition accuracy.



Without FGSM FGSM-0.1 FGSM-1

Figure 5. The EER performance based on different conditions.

Table 1. The numerical EER comparison under different conditions.

Model name	EER
Without FGSM	0.29
FGSM-0.1	0.36
FGSM-1	0.37

Figure 6 illustrates the impact of FGSM-based perturbations on the cosine similarity scores between pairs of images belonging to the same or different identities. In the same identity row, where both images belong to the same person, the similarity score without FGSM is 0.94, indicating a high degree of matching. When FGSM perturbations with strengths of 0.1 and 1 are applied, the similarity scores slightly decrease to 0.92 and 0.91, respectively. In the different identities row, where the images belong to two different individuals, the similarity score without FGSM is 0.77, showing that the model can effectively distinguish between different identities. However, when FGSM perturbations with strengths of 0.1 and 1 are applied, the similarity scores increase to 0.79 and 0.81, respectively. These results indicates that the perturbations make it harder for the model to distinguish between different identities.

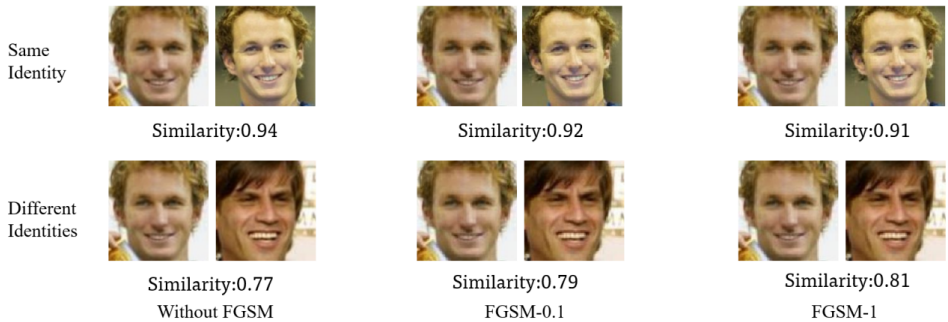


Figure 6. The prediction samples based on different conditions.

Figure 7 shows a visualization comparison using Grad-CAM under different conditions. Grad-CAM highlights the regions where the model focuses most when making predictions. In the original image, the model's attention is concentrated on key facial features, such as the eyes, nose, and mouth, which are crucial for identity recognition. When FGSM perturbations are applied, although the images appear visually identical to the human eye, the model's attention shifts. For the image perturbed with FGSM-0.1, the attention becomes more dispersed, with less focus on critical facial regions. In the FGSM-1 case, the attention further deteriorates, shifting its focus primarily to the mouth and weakening the emphasis on other key facial features, such as the eyes and nose. This demonstrates that adversarial noise, even if imperceptible to humans, disrupts the model's focus, leading to potential misclassifications. The comparison highlights the vulnerability of the model to adversarial attacks, where subtle pixel-level perturbations can significantly alter the model's perception of the input image.

Figure 7. The visualization comparison using Grad-CAM [48] [49] based on different conditions.

Figure 8 shows the training loss curve of the autoencoder. The loss decreases sharply during the initial epochs, indicating that the model quickly learns to minimize reconstruction errors. As the training progresses, the loss gradually stabilizes, converging to a low value after 20 epochs, which demonstrates that the autoencoder successfully reconstructs the input images while mitigating adversarial noise. Figure 9 presents a visual comparison between FGSM-based images and their reconstructed counterparts using the trained autoencoder. The top row shows images perturbed by FGSM, where the adversarial noise is imperceptible to the human eye. The bottom row shows the corresponding reconstructed images produced by the autoencoder.

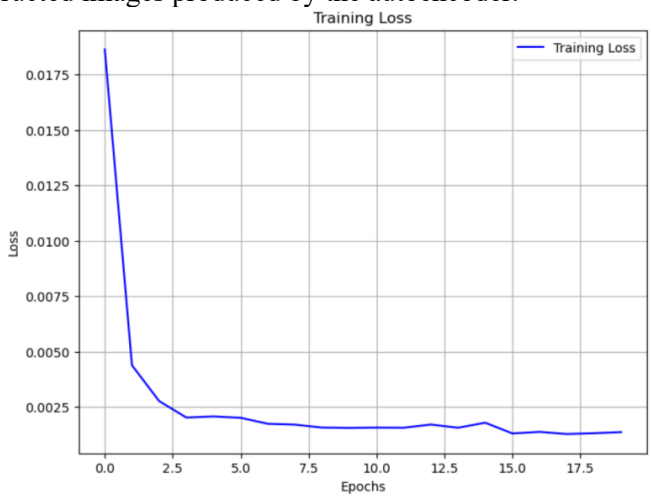


Figure 8. The training curves of the autoencoder.

Figure 9. The comparison between FGSM-based and reconstructed images.

Figure 10 and Table 2 present the comparison of EER before and after autoencoder reconstruction under FGSM perturbations with strengths of 0.1 and 1. Before reconstruction, the EER values for FGSM-0.1 and FGSM-1 are 0.36 and 0.37, respectively, showing that the model's performance degrades under adversarial attacks. After applying the autoencoder to reconstruct the images, the EER values decrease to 0.32 for FGSM-0.1 and 0.31 for FGSM-1. This reduction indicates that the autoencoder mitigates the effect of FGSM perturbations and improves the recognition performance. Although the EER values do not completely return to their original state, the results demonstrate that the autoencoder enhances the model's robustness against adversarial noise. The ROC curves in Figure 10 visually confirm the improvement, where the reconstructed images yield better performance compared to the FGSM-perturbed inputs. Figure 11 provides samples for further observation of similarity scores. After reconstruction, most samples of the same identity show an increase in facial similarity, while the similarity for faces belonging to different identities decreases. These results highlight the effectiveness of the autoencoder in reducing the impact of adversarial attacks and improving face recognition accuracy.

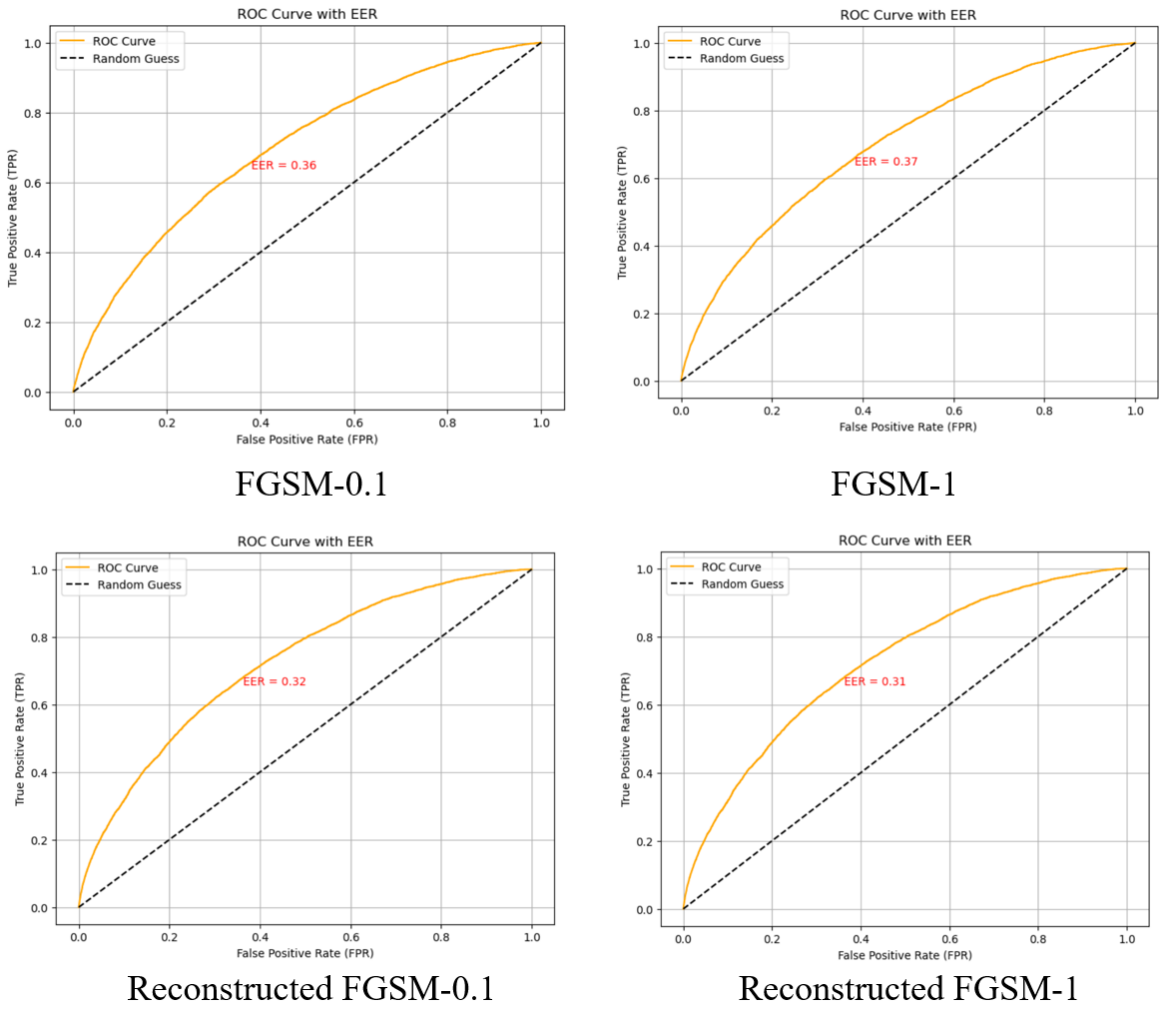


Figure 10. The comparison of EER before and after autoencoder reconstruction.

Table 2. The numerical EER comparison before and after autoencoder reconstruction.

Model name	EER
FGSM-0.1	0.36
FGSM-1	0.37
Reconstructed FGSM-0.1	**0.32**
Reconstructed FGSM-1	**0.31**

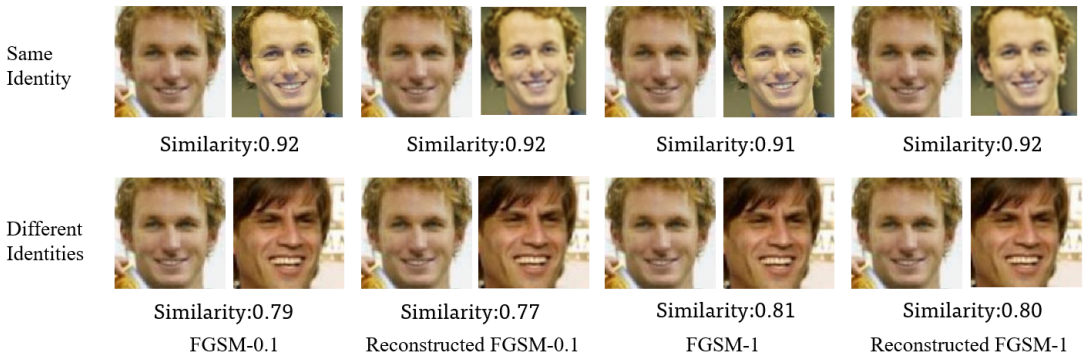


Figure 11. Impact of FGSM reconstruction on facial similarity across identities.

4.2 Discussion

Despite the promising results achieved in mitigating FGSM-based attacks using the convolutional autoencoder, there are still several limitations in this study that warrant further investigation. First, while the autoencoder effectively reduces the equal error rate (EER) under adversarial perturbations, the reconstructed images do not fully restore the model's performance to the clean condition. This suggests that the current autoencoder may not completely eliminate adversarial noise, especially for stronger perturbations like FGSM-1. Further improvements in the network architecture or the inclusion of adversarial training techniques could enhance the reconstruction quality and overall robustness. Second, the experiments in this study were limited to FGSM-based white-box attacks with fixed perturbation strengths of 0.1 and 1. Although FGSM is a widely studied adversarial attack, more advanced and complex attacks, such as Projected Gradient Descent (PGD) or Carlini & Wagner (C&W) attacks, should be considered in future work to comprehensively evaluate the proposed framework. Additionally, the study did not account for black-box attack scenarios, where the attacker lacks direct access to the target model's parameters. Addressing these limitations would provide a more holistic understanding of the model's robustness. Another limitation lies in the dataset size and diversity. The experiments were conducted on a relatively small dataset containing 1,105 images with 248 identities, which may not generalize well to larger or more complex datasets. Future work will focus on evaluating the proposed framework on larger, more diverse datasets to validate its effectiveness across varying conditions.

Moving forward, we plan to explore the integration of adversarial defense methods such as adversarial training and denoising networks. We will also investigate hybrid approaches that combine autoencoders with ensemble methods or advanced deep learning architectures to further improve model resilience. Additionally, attention-based mechanisms could be incorporated to enhance the focus on critical facial regions, potentially improving recognition accuracy under adversarial conditions.

5. **Conclusion**

This study optimizes the vulnerability of facial recognition models to FGSM-based white-box attacks by introducing a convolutional autoencoder framework. The results show that the autoencoder effectively reconstructs perturbed images, reducing the EER and improving model robustness. However, the reconstructed images do not entirely restore the model's performance to clean conditions, especially for stronger perturbations. Future work will focus on enhancing the

reconstruction process, integrating advanced adversarial defense strategies, and evaluating the framework against black-box attacks and more complex adversarial methods. Larger and more diverse datasets will also be explored to ensure scalability and generalization. The proposed approach represents a significant step toward building resilient facial recognition systems in adversarial environments.

Funding
Not applicable

Author Contributions
Jiahuai Ma contributed to conceptualization, methodology, and investigation. Alan Wilson supervised the project, conducted formal analysis, and reviewed the manuscript. Both authors participated in writing and approved the final manuscript.

Institutional Reviewer Board Statement
Not applicable

Informed Consent Statement
Not applicable

Data Availability Statement
Not applicable

Conflict of Interest
The authors declare no conflict of interest.

References

[1] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: A literature survey. ACM computing surveys (CSUR). 2003 Dec 1;35(4):399-458.

[2] Tolba AS, El-Baz AH, El-Harby AA. Face recognition: A literature review. International Journal of Signal Processing. 2006 Feb;2(2):88-103.

[3] Li L, Mu X, Li S, Peng H. A review of face recognition technology. IEEE access. 2020 Jul 21;8:139110-20.

[4] Zhou Z, Wu J, Cao Z, She Z, Ma J, Zu X. On-Demand Trajectory Prediction Based on Adaptive Interaction Car Following Model with Decreasing Tolerance. In2021 International Conference on Computers and Automation (CompAuto) 2021 Sep 7 (pp. 67-72). IEEE.

[5] Zhang G, Zhou T, Cai Y. CORAL-based Domain Adaptation Algorithm for Improving the Applicability of Machine Learning Models in Detecting Motor Bearing Failures. Journal of Computational Methods in Engineering Applications. 2023 Nov 3:1-7.

[6] Li C, Tang Y. The Factors of Brand Reputation in Chinese Luxury Fashion Brands. Journal of Integrated Social Sciences and Humanities. 2023 Nov 20:1-4.

[7] Gan Y, Ma J, Xu K. Enhanced E-Commerce Sales Forecasting Using EEMD-Integrated LSTM Deep Learning Model. Journal of Computational Methods in Engineering Applications. 2023 Nov 11:1-1.

[8] Chen X, Zhang H. Performance Enhancement of AlGaN-based Deep Ultraviolet Light-emitting Diodes with AlxGa1-xN Linear Descending Layers. Innovations in Applied Engineering and Technology. 2023 Oct 31:1-0.

[9] Wang H, Li J, Xiong S. Efficient join algorithms for distributed information integration based on XML. International Journal of Business Process Integration and Management. 2008 Jan 1;3(4):271-81.

[10] Xiong S, Li J. Optimizing many-to-many data aggregation in wireless sensor networks. InAsia-Pacific Web Conference 2009 Apr 2 (pp. 550-555). Berlin, Heidelberg: Springer Berlin Heidelberg.

[11] Pietikäinen M. Local binary patterns. Scholarpedia. 2010 Mar 3;5(3):9775.

[12] Abdi H, Williams LJ. Principal component analysis. Wiley interdisciplinary reviews: computational statistics. 2010 Jul;2(4):433-59.

[13] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) 2005 Jun 20 (Vol. 1, pp. 886-893). Ieee.

[14] Wenjun D, Fatahizadeh M, Touchaei HG, Moayedi H, Foong LK. Application of six neural network-based solutions on bearing capacity of shallow footing on double-layer soils. Steel and Composite Structures. 2023;49(2):231-44.

[15] Dai W. Design of traffic improvement plan for line 1 Baijiahu station of Nanjing metro. Innovations in Applied Engineering and Technology. 2023 Dec 21;10.

[16] Dai W. Evaluation and improvement of carrying capacity of a traffic system. Innovations in Applied Engineering and Technology. 2022 Nov 22:1-9.

[17] Dai W. Safety evaluation of traffic system with historical data based on Markov process and deep-reinforcement learning. Journal of Computational Methods in Engineering Applications. 2021 Oct 21:1-4.

[18] Hao Y, Chen Z, Jin J, Sun X. Joint operation planning of drivers and trucks for semi-autonomous truck platooning. Transportmetrica A: Transport Science. 2023 Oct 7:1-37.

[19] Lei J, Nisar A. Investigating the Influence of Green Technology Innovations on Energy Consumption and Corporate Value: Empirical Evidence from Chemical Industries of China. Innovations in Applied Engineering and Technology. 2023 Nov 27:1-6.

[20] Xiong S, Zhang H, Wang M, Zhou N. Distributed Data Parallel Acceleration-Based Generative Adversarial Network for Fingerprint Generation. Innovations in Applied Engineering and Technology. 2022:1-2.

[21] Xiong S, Chen X, Zhang H. Deep Learning-Based Multifunctional End-to-End Model for Optical Character Classification and Denoising. Journal of Computational Methods in Engineering Applications. 2023 Nov 15:1-3.

[22] Xiong S, Li J. An efficient algorithm for cut vertex detection in wireless sensor networks. In2010 IEEE 30th International Conference on Distributed Computing Systems 2010 Jun 21 (pp. 368-377). IEEE.

[23] Li J, Xiong S. Efficient Pr-skyline query processing and optimization in wireless sensor networks. Wireless Sensor Network. 2010 Nov 19;2(11):838.

[24] Yu L, Li J, Cheng S, Xiong S. Secure continuous aggregation via sampling-based verification in wireless sensor networks. In2011 Proceedings IEEE INFOCOM 2011 Apr 10 (pp. 1763-1771). IEEE.

[25] Liu Y, Mao S, Mei X, Yang T, Zhao X. Sensitivity of adversarial perturbation in fast gradient sign method. In2019 IEEE symposium series on computational intelligence (SSCI) 2019 Dec 6 (pp. 433-436). IEEE.

[26] Naqvi SM, Shabaz M, Khan MA, Hassan SI. Adversarial attacks on visual objects using the fast gradient sign method. Journal of Grid Computing. 2023 Dec;21(4):52.

[27] Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition. IEEE transactions on pattern analysis and machine intelligence. 2010 Jul 8;32(11):2106-12.

[28] Nixon M. Eye spacing measurement for facial recognition. InApplications of digital image processing VIII 1985 Dec 19 (Vol. 575, pp. 279-285). SPIE.

[29] Gray M. Urban surveillance and panopticism: will we recognize the facial recognition society?. Surveillance & Society. 2003;1(3):314-30.

[30] Xiong S, Zhang H, Wang M. Ensemble Model of Attention Mechanism-Based DCGAN and Autoencoder for Noised OCR Classification. Journal of Electronic & Information Systems. 2022;4(1):33-41.

[31] Yu L, Li J, Cheng S, Xiong S, Shen H. Secure continuous aggregation in wireless sensor networks. IEEE Transactions on Parallel and Distributed Systems. 2013 Mar 7;25(3):762-74.

[32] Xiong S, Yu L, Shen H, Wang C, Lu W. Efficient algorithms for sensor deployment and routing in sensor networks for network-structured environment monitoring. In2012 Proceedings IEEE INFOCOM 2012 Mar 25 (pp. 1008-1016). IEEE.

[33] Feng Z, Xiong S, Cao D, Deng X, Wang X, Yang Y, Zhou X, Huang Y, Wu G. Hrs: A hybrid framework for malware detection. InProceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics 2015 Mar 4 (pp. 19-26).

[34] Turk M, Pentland A. Eigenfaces for recognition. Journal of cognitive neuroscience. 1991 Jan 1;3(1):71-86.

[35] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on pattern analysis and machine intelligence. 1997 Jul;19(7):711-20.

[36] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence. 2006 Oct 30;28(12):2037-41.

[37] Taigman Y, Yang M, Ranzato MA, Wolf L. Deepface: Closing the gap to human-level performance in face verification. InProceedings of the IEEE conference on computer vision and pattern recognition 2014 (pp. 1701-1708).

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

[39] Liu W, Wen Y, Yu Z, Li M, Raj B, Song L. Sphereface: Deep hypersphere embedding for face recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 212-220).

[40] Deng J, Guo J, Xue N, Zafeiriou S. Arcface: Additive angular margin loss for deep face recognition. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 4690-4699).

[41] Lupart S, Clinchant S. A study on FGSM adversarial training for neural retrieval. InEuropean Conference on Information Retrieval 2023 Mar 17 (pp. 484-492). Cham: Springer Nature Switzerland.

[42] Sen J, Dasgupta S. Adversarial attacks on Image classification models: FGSM and patch attacks and their impact. arXiv preprint arXiv:2307.02055. 2023 Jul 5.

[43] Zhang Y. A better autoencoder for image: Convolutional autoencoder. InICONIP17-DCEC. Available online: http://users. cecs. anu. edu. au/Tom. Gedeon/conf/ABCs2018/paper/ABCs2018_paper_58. pdf (accessed on 23 March 2017) 2018 Mar.

[44] Guo X, Liu X, Zhu E, Yin J. Deep clustering with convolutional autoencoders. InNeural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24 2017 (pp. 373-382). Springer International Publishing.

[45] Holden D, Saito J, Komura T, Joyce T. Learning motion manifolds with convolutional autoencoders. InSIGGRAPH Asia 2015 technical briefs 2015 Nov 2 (pp. 1-4).

[46] Qin Z, Zhang Z, Chen X, Wang C, Peng Y. Fd-mobilenet: Improved mobilenet with a fast downsampling strategy. In2018 25th IEEE International Conference on Image Processing (ICIP) 2018 Oct 7 (pp. 1363-1367). IEEE.

[47] Sinha D, El-Sharkawy M. Thin mobilenet: An enhanced mobilenet architecture. In2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON) 2019 Oct 10 (pp. 0280-0285). IEEE.

[48] Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: Why did you say that?. arXiv preprint arXiv:1611.07450. 2016 Nov 22.

[49] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. InProceedings of the IEEE international conference on computer vision 2017 (pp. 618-626).