Enhancing Face Anti-spoofing Systems Through Synthetic Image Generation

César Vega(✉) and Ruben Manrique

Universidad de los Andes, Bogotá, Colombia
{c.vegaf,rf.manrique}@uniandes.edu.co

Abstract. This study introduces a strategy for synthetic image generation aimed at enhancing the detection capability of facial authentication systems (FAS). By employing various digital manipulation techniques, new synthetic fake images were generated using existing datasets. Through experiments and result analysis, the impact of using these new fake samples on improving the detection accuracy of FAS systems was evaluated. The findings demonstrated the effectiveness of synthetic image generation in augmenting the diversity and complexity of the training data. Fine-tuning using the enhanced datasets significantly improved the detection accuracy across the evaluated FAS systems. Nonetheless, the degree of improvement varied among systems, indicating varying susceptibility to specific types of attacks.

Keywords: Face Anti-Spoofing (FAS) · Presentation Attack Detection (PAD) · Digital Manipulation (MD) · Data Augmentation

1 Introduction

In the era of expanding intelligent workspace environments, the utilization of technologies like facial recognition has witnessed a notable surge across diverse domains, including but not limited to banking and security within restricted zones [10]. However, with the increasing popularity of these systems, there is a concurrent escalation in malicious activities aimed at compromising their integrity. Cyber attackers employ a range of tactics, encompassing presentation attacks utilizing printed photos or 3D facial masks [2,4], as well as digital manipulation attacks involving the creation of forgeries, such as DeepFakes [7,24].

Face Anti-Spoofing (FAS) systems have been developed to safeguard facial recognition systems. Early approaches in deep learning employed Convolutional Neural Networks (CNNs) to detect identity spoofing [7]. More recent advancements include anomaly detection methods [7] and liveness detection in images [23]. These approaches have been supplemented with widely adopted datasets. Notably, CASIA represents a comprehensive dataset for antispoofing model development, encompassing diverse ethnicities, modalities, and attack types [13,14,21]. However, the CASIA dataset lacks robust instances of digital manipulation, which is a prevalent type of attack in practice. According to [18] the

H. Florez and M. Leon (Eds.): ICAI 2023, CCIS 1874, pp. 16–30, 2024.
https://doi.org/10.1007/978-3-031-46813-1_2

generation of DeepFakes have grown by 16%, and the percentage of people who know how to create or use DeepFake applications has doubled since 2019, with 2% of the surveyed individuals in 2022 having this knowledge. Undoubtedly, these trends increase the risk of attempts at synthetic generation-based forgeries.

To overcome this limitation, data augmentation techniques have been proposed [5,22,25]. In this study, strategies for creating spoofing attack instances are presented, involving both simple face substitution and digital manipulation of images. Architectures based on probabilistic diffusion models for InPainting [15] were employed. These methods enable the generation of partially modified faces, which can be included as new examples of forgeries in existing datasets. This represents the primary contribution of this work.

To assess the effectiveness of FAS systems against artificially generated attack scenarios, three distinct FAS systems were chosen. Their performances were evaluated using metrics defined by the ISO/IEC 30107-3 standard. The selected FAS systems were evaluated both before and after undergoing Fine Tuning with the newly constructed datasets. The results of this evaluation, along with the corresponding discussion, constitute the second contribution of this study.

The structure of this paper is as follows: Sect. 2 provides an overview of the related work on FAS systems. Section 3 details the methodology employed for dataset construction and the selected FAS systems. Section 4 presents the experimental results. Subsequently, Sect. 5 and 6 offers a comprehensive discussion, future research directions, and concluding remarks, summarizing the key ideas and findings presented throughout the paper.

2 Related Work

Numerous investigations have been conducted to evaluate Facial Authentication Systems (FAS), with a particular emphasis on texture analysis to detect presentation attacks. Several approaches have focused on pixel-wise classification techniques, such as the automatic discovery of optimal pixel labels through pyramid supervision [24]. Additionally, there has been an exploration into effective methods grounded in the concept of Meta Patterns for enhancing the detection of deep forgeries [1]. Within the spectrum of deep texture-based features employed for robust forgery detection, notable techniques include Local Binary Patterns (LBP), Speeded-Up Robust Features (SURF), and Difference of Gaussians (DoG) [12].

The proliferation of digital manipulation attacks, predominantly facilitated by Generative Adversarial Network (GAN) architectures, encompasses the production of entirely or partially altered, photorealistic facial images through techniques such as expression swapping, attribute manipulation, and complete face synthesis. Confronting these attacks remains a formidable challenge for Facial Authentication Systems (FAS) [22], given that digitally manipulated content of this nature can be generated routinely via “no code” applications that effortlessly modify facial attributes through filter-based adjustments. The accessibility

to digital manipulation technologies is widespread and does not necessitate specialized expertise or skills.

This presents both a potential hazard and an opportunity to bolster existing systems through the generation of additional training data. The generation of synthetic images via digital manipulation to enhance training datasets has received substantial attention across diverse domains. Synthetic imagery has proven pivotal in supporting endeavors such as facial alignment, facial recognition, 3D facial pose estimation, pedestrian detection, action recognition, among others [5]. Furthermore, these data can be continuously collected from diverse systems, making them potentially suitable for semi-supervised learning [25], and, in combination with adversarial learning could serve as a viable option to increase model defense to a wider range of attack types [25].

To summarize, research indicates that utilizing data augmentation during model training can yield advantages. It is suggested that the inclusion of synthetic images depicting counterfeit samples can aid in the precise classification of genuine samples [4,8,24]. This paper contributes to this line of research.

3 Methodology

This section outlines the methodology employed for generating synthetic images used for substitution-based presentation attacks and digital manipulation attacks. Furthermore, it elaborates on the training of models for facial authentication systems (FAS) and the evaluation of these models.

3.1 Generation of Image Sets

Considering the difficulty and time-consuming process involved in creating and subsequently classifying examples of presentation attacks [2,5,19,22], the following image sets were developed with the aim of generating a larger quantity of forgery images covering various attack scenarios such as presentation attacks and digital manipulation attacks.

Substitution-Based Image Set. This image set aims to provide many images for substitution-based presentation attacks while maintaining the image conditions by not covering the surrounding environment of the subject to be impersonated. The construction process of this dataset begins with the recognition of the face of the person whose face is to be substituted (see Fig. 1 - Person X (a)), and in the same frame (where the face is recognized), the face of the person to be impersonated is placed (see Fig. 1 - Person y (b)), which was previously recognized, resulting in images with direct substitution (see Fig. 1 - (c)).

To generate the dataset, different combinations of x/y were created using 20 individuals from the CASIA dataset [13,14,21]. For each of these 20 individuals, the first 8 videos were taken, resulting in the creation of approximately 6,900,000

images, which allows for the creation of around 25,000 videos. This approach enables the preservation of various impersonation contexts originally present in the CASIA image set, such as mask printing with and without simulated eye or facial movement (Fig. 2a), as well as recordings on a mobile device with different camera angles (Fig. 2b).

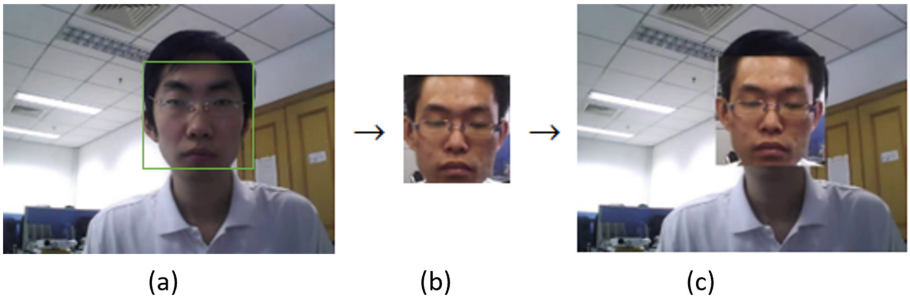

Fig. 1. Face substitution

Fig. 2. Examples of face substitution

Regeneration-Based Image Set (InPainting). This image set is generated using probabilistic diffusion models for noise removal (RePaint) [15]. It requires an input image of a person X (see Fig. 3 - Person X (a)) and a binary black and white mask (see Fig. 3 - Binary Mask (b)), which is used to determine the segments that need to be regenerated in the image. Once the repainting model is applied, a new image is obtained (see Fig. 3 - Resulting Image (c)). By

using appropriate masks with this method, we can preserve the original environment of the image and focus exclusively on facial elements. For generation, high-resolution images from CASIA were used, which had to be adjusted to allow the model to regenerate the image borders, resulting in a square resolution of 256×256.

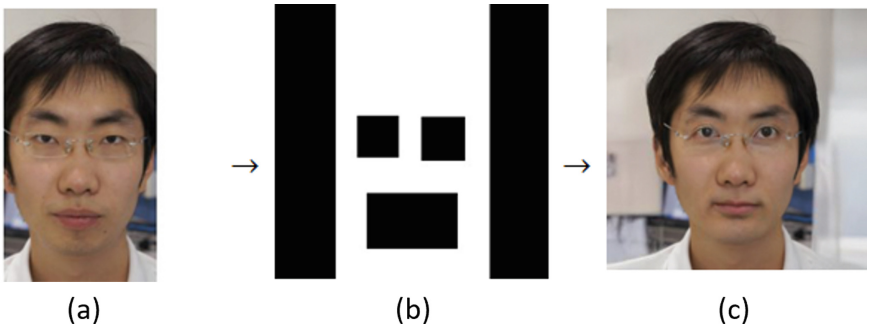

Fig. 3. Face InPainting

To generate the images, four different binary masks (Fig. 4) were used, each with different coverage of the face. Initially, masks were auto-detected using the OpenCV library[1]. However, due to the lack of accuracy, manual modifications were made to the mask as follows: coverage of eyes and mouth (a), coverage of eyes, nose, and mouth (b), coverage of eyes (c) and coverage of left eye (d).

From each mask, seven subsets were obtained, each consisting of approximately 400 images. As a result, a total of 8368 repainted images were generated.

The choice of these masks was based on the intention to replicate scenarios where facial authentication systems could be challenged by realistic presentation attacks. By creating masks that cover specific parts of the face, such as eyes, nose, and mouth, or even just one eye, the aim is to mimic the manipulations an attacker might perform to deceive the system.

In addition to the image sets generated in this research, the ffhq dataset from StyleGAN [11] was used. This dataset provides an additional set of images of digital manipulation attacks. In Table 1 a summary of the resulting datasets is shown. The total number of images in each set was randomly divided into three subsets: training, validation, and testing.

[1] https://github.com/opencv/opencv/tree/master.

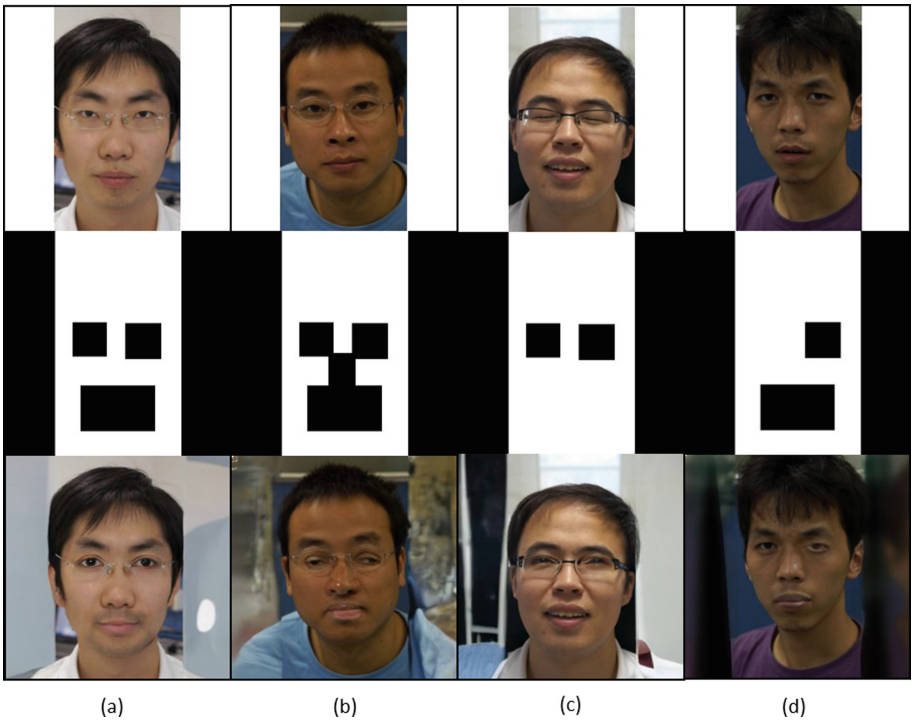

Fig. 4. Examples of images generated by the diffusion model

Table 1. Division of image sets

Image sets	Training	Validation	Testing	Total
CASIA	82536	16507	20635	119678
Substitution	4772300	954460	1193077	6919837
StyleGAN	160160	32032	40040	200200
Diffusion	5355	1339	1674	8368

3.2 Selected Models of FAS Systems

The selection of the FAS systems was based on the premise of addressing the issues present in the literature, particularly focusing on texture-based FAS approaches, which are widely used [5,19]. Within this category, the Silent FAS and Object FAS models were chosen. Additionally, FAS systems with face recognition capabilities, such as Liveness Detection FAS and Object FAS, were selected to determine the advantages and disadvantages of face detection in synthetically generated facial images, which are becoming increasingly sophisticated [3,22].

Silent Face Anti Spoofing. This model is primarily based on auxiliary supervision of the Fourier spectrum map in model training to perform liveness detection in an image in the frequency domain. The Silent FAS model adopts a silent or texture-based liveness detection approach. Firstly, the face is detected in the image, and then a scale of 2.7 and 4 is taken to increase the surrounding range. Prediction is performed by weighting two trained models, one with a scale of 2.7 and the other with a scale of 4. This model was trained using a derivation of CASIA and is available as open-source code[2].

Liveness Detection Face Anti Spoofing. This model is based on face detection and is used in a web application for biometric authentication[3]. The process begins with face detection, followed by validation using a deep neural network. This model was trained using the Replay-Attack image dataset[4] and is available as open-source code, implemented in Python using the TensorFlow and Keras libraries.

Objects Face Anti Spoofing. This model is based on the combination of two different color spaces: $CIE\ L^*u^*v^*$ and YC_bC_r, using histograms to classify an image as real or fake. It was trained using the Replay-Attack image dataset and a private database. This model utilizes an ensemble method with 10 decision trees and is implemented in Python using the ExtraTreesClassifier classifier from the sklearn library. The source code is openly available[5].

3.3 Datasets Preparation

The Sustitution, StyleGAN, and Diffusion datasets only contain examples of fake images (i.e., positive samples). To perform Fine Tuning of the FAS models, data balancing is necessary. A subset of real images from the CASIA* dataset was randomly selected for this purpose. Additionally, downsampling was applied to the generated fake images. The sizes of the resulting datasets after balancing are presented in Table 2.

Table 2. Size of datasets after balancing

Image Dataset	Fake	Real
Substitution	21000	20753*
StyleGAN	21000	20753*
Diffusion	6694	6694*

[2] https://github.com/minivision-ai/Silent-Face-Anti-Spoofing/.
[3] https://github.com/birdowl21/Face-Liveness-Detection-Anti-Spoofing-Web-App.
[4] https://www.idiap.ch/en/dataset/replayattack.
[5] https://github.com/ee09115/spoofing_detection.

Additionally, to perform the Fine Tuning process, the images need to be transformed according to the expected input of each FAS model, and in some cases, the output (i.e., labels). Table 3 provides a summary of the expected inputs/outputs for each model.

Table 3. Input and output formats for each FAS model

Silent FAS	Liveness Detection FAS	Object FAS
Input: 3D tensor (80, 80, 3)	Input: 3D tensor (32, 32, 3)	Input: Vector (1536)
Output: Binary	Output: One-hot encoding	Output: Binary

3.4 Model Evaluation

The evaluation of the models was performed using test datasets that included both substitution-based presentation attacks and digital manipulation attacks. The comparison will be made in terms of metrics outlined in ISO/IEC30107-3[6]. For this research, the standard labeling convention will be followed, where fake images are assigned the label 0 and real images are assigned the label 1. The metrics used for evaluation are detailed below:

- **Attack Presentation Classification Error Rate (APCER):** It is the proportion of presentation attacks incorrectly classified as genuine presentations (i.e., the error rate of fake images classified as real).

$$APCER = FP/(TN + FP) \tag{1}$$

- **Average Classification Error Rate (ACER):** It is the average of the two error rates mentioned above.

$$ACER = (APCER + NPCER)/2 \tag{2}$$

where, TP is the number of fake images classified as fake (true positives), TN is the number of real images classified as real (true negatives), FP is the number of real images classified as fake (false positives), and FN is the number of fake images classified as real (false negatives).

4 Experimental Setup

In this section, different experiments are conducted to evaluate the performance of the FAS models before and after Fine Tuning using the generated image sets. The experiments were conducted on three selected FAS models: Silent FAS, Liveness Detection FAS, and Object FAS.

[6] https://www.iso.org/standard/79520.html.

4.1 Experiment 1: Evaluation of the Performance of FAS Models Without Fine Tuning

In Experiment 1, the FAS models without Fine Tuning were evaluated. We aimed to observe the behavior of the models, without additional training rounds on our generated datasets. As shown in Table 4, the results demonstrate that all models achieved high error rates (APCER, ACER) in most of the evaluated image sets. Silent FAS achieved the best performance with an error of 0 in the CASIA dataset, but it clearly exhibited poor performance in the other datasets. This can be interpreted as a limitation in detecting digital manipulation attacks.

On average, Silent FAS exhibitedan average APCER of 0.6795 and an average ACER of 0.3326. Liveness Detection FAS and Object FAS demonstrated average APCER values of 0.8141 and 0.6122, and average ACER values of 0.4820 and 0.3545, respectively. These results underscore the importance of considering the performance of FAS models across diverse image sets, as each model exhibits strengths and weaknesses in different scenarios. These findings highlight the importance of evaluating and comparing multiple FAS models on different datasets to gain a more comprehensive understanding of their performance and select the most suitable model for a specific presentation attack detection application.

Table 4. Experiment 1: Results of FAS models without Fine Tuning. The test subset was used for each dataset.

Image Dataset	Silent FAS		Liveness Detection FAS		Object FAS	
	APCER	ACER	APCER	ACER	APCER	ACER
CASIA	0	0	0.4496	0.5247	0.61	0.4986
Substitution	0.7517	0.3758	0.9099	0.455	0.8227	0.4114
StyleGAN	0.9729	0.4577	0.9153	0.4577	0.4681	0.2341
Diffusion	0.9934	0.4967	0.9815	0.4907	0.5478	0.2739
Averages	**0.6795**	**0.3326**	**0.8141**	**0.4820**	**0.6122**	**0.3545**

4.2 Experiment 2: Models Fine Tuning

In Experiment 2, the Fine Tuning process of the models was carried out using different image sets. Each image set's corresponding training set was used to fine-tune the model. The results are presented for both training and validation. The results on the test datasets are reserved for Experiment 3. For the models based on neural architectures, the number of epochs corresponding to the best checkpoint obtained in validation is reported, without exceeding 50 epochs.

Silent FAS. The results are presented in Table 5. Overall, it can be observed that the model achieved high accuracy (Acc) values and low loss values both in the training and validation sets for all evaluated image datasets.

Specifically, in the CASIA image dataset, the model with a scale of 4 achieved a training accuracy of 0.9968 and a loss of 0.0071, while in the validation set, it achieved an accuracy of 0.9938 and a loss of 0.0105. On the other hand, the model with a scale of 2.7 achieved a training accuracy of 0.996 and a loss of 0.0098, and a validation accuracy of 0.9948 and a loss of 0.0142.

For the Sustitution, StyleGAN, and Diffusion image datasets, the model also demonstrated solid performance in terms of accuracy and loss. These results indicate that the Fine Tuning process improved the performance of the Silent FAS model in detecting presentation attacks, demonstrating the effectiveness of this approach in fine-tuning the model for specific image datasets.

Table 5. Experiment 2: Accuracy/Loss obtained in the Fine Tuning process of Silent FAS using each image dataset

Image Dataset	Model with 4-scale					Model with 2.7-scale				
	Epochs	Acc	Loss	Acc_val	Loss_val	Epochs	Acc	Loss	Acc_val	Loss_val
CASIA	4	0.9968	0.0071	0.9938	0.0105	4	0.996	0.0098	0.9948	0.0142
Substitution	10	0.976	0.0335	0.9666	0.0462	9	0.9782	0.0295	0.9703	0.0395
StyleGAN	4	0.9999	0.0026	0.9991	0.0033	3	0.9975	0.0066	0.9961	0.0103
Diffusion	4	0.9998	0.0031	0.9981	0.0036	3	0.9986	0.0051	0.9992	0.005

Liveness Detection FAS. The results obtained are shown in Table 6. It can be observed that the model achieved high accuracy (Acc) values and low loss (Loss) values both in the training and validation sets for all evaluated image datasets.

Specifically, in the CASIA image dataset, the model achieved a training accuracy of 0.999 and a loss of 0.0008, while in the validation set, it achieved an accuracy of 0.9979 and a loss of 0.0007. In the Sustitution image dataset, the model exhibited a training accuracy of 0.9862 and a loss of 0.0362, and a validation accuracy of 0.9904 and a loss of 0.0215. As expected, similar to the Silent FAS model, the fine-tuning process proves to be effective with all datasets, indicating its effectiveness in handling different types of attacks.

Object FAS. The results obtained are shown in Table 7. The CASIA image dataset achieved the highest level of accuracy with a value of 0.9976, followed by the Sustitution image dataset with 0.9971, the Diffusion image dataset with 0.997, and finally the StyleGAN image dataset with 0.9976. These results indicate that the Fine Tuning process was effective in improving the performance of the Object FAS model in detecting fake images. Additionally, the accuracy values in the validation set (Acc_val) are also high, indicating that the model generalizes well to previously unseen data.

Table 6. Experiment 2: Accuracy/Loss obtained in the Fine Tuning process of Liveness Detection FAS using each image dataset

Image Dataset	Liveness model				
	Epochs	Acc	Loss	Acc_val	Loss_val
CASIA	40	0.999	0.0008	0.9979	0.0007
Substitution	40	0.9862	0.0362	0.9904	0.0215
StyleGAN	50	0.9993	0.0025	1	0.0009
Diffusion	40	0.867	0.2867	0.8734	0.2873

Table 7. Experiment 2: Accuracy obtained in the Fine Tuning process of Object FAS using each image dataset

Image Dataset	Decision Trees	
	Acc	Acc_val
CASIA	0.9976	0.9967
Substitution	0.9971	0.9944
StyleGAN	0.9976	0.9887
Diffusion	0.997	0.991

4.3 Experiment 3: Evaluation of FAS Model Performance After Fine Tuning

In Experiment 3, the Fine Tuning process was performed using all image datasets on the models. In contrast to Experiment 2, where Fine Tuning was conducted on individual datasets, in this experiment, Fine Tuning was performed collectively using all datasets, and evaluation was done on the test sets. The results obtained are summarized in Table 8. In terms of the APCER metric (Attack Presentation Classification Error Rate), both the Silent FAS and Diffusion models achieved an APCER of 0, indicating no false positives were detected in the classification of attacks. On the other hand, the StyleGAN model obtained an APCER of 0.2629, and the Sustitution model achieved an APCER of 0.4857.

Regarding the ACER (Average Classification Error Rate), the averages obtained were 0.1086 for Silent FAS, 0.0050 for Liveness Detection FAS, and 0.4814 for Object FAS. This indicates that the Liveness Detection FAS model had the best performance in terms of the average error rate, while the Silent FAS model had the lowest performance. Overall, it is evident that the Fine Tuning process improved the performance of the models when compared to the results of Experiment 1 (without Fine Tuning).

Table 8. Experiment 3: Results on the test sets after Fine Tuning with all image datasets

Image Dataset	Silent FAS		Liveness Detection FAS		Object FAS	
	APCER	ACER	APCER	ACER	APCER	ACER
CASIA	0	0.0602	0	0.0003	0.7817	0.5444
StyleGAN	0.2629	0.1314	0	0	0.9201	0.46
Diffusion	0	0	0	0	0.9502	0.4751
Substitution	0.4857	0.2428	0.0396	0.0198	0.892	0.446
Averages	**0.1872**	**0.1086**	**0.0099**	**0.0050**	**0.8860**	**0.4814**

5 Results Analysis

In this section, the analysis of the results obtained from the conducted experiments will be presented. The most relevant findings will be discussed, and corresponding conclusions will be drawn.

Firstly, it was observed that the strategy of generating synthetic images using presentation attacks and digital manipulation was effective in improving the performance of the evaluated FAS systems. The generated datasets allowed for increased diversity and complexity in the training data, resulting in an enhancement in attack detection capability.

Regarding the individual experiments, it was found that Fine-Tuning using presentation attack datasets had a positive impact on the Silent FAS and Liveness Detection FAS systems. These systems exhibited a significant improvement in detection accuracy metrics, with an average reduction in APCER of 25% and 83%, respectively. On the other hand, the Object FAS system did not show a significant improvement after Fine-Tuning with presentation attack datasets.

In the case of Fine-Tuning using digital manipulation datasets, it was observed that the Silent FAS and Object FAS systems demonstrated an improvement in detection metrics, with an average reduction in APCER of 61% and 13%, respectively. However, the Liveness Detection FAS system did not show a significant improvement after Fine-Tuning with digital manipulation datasets.

Regarding the analysis of the combined results from all datasets, it was found that the Liveness Detection FAS system showed the most significant improvement, with an average reduction in APCER of 83%. The Silent FAS system also exhibited a considerable improvement, with an average reduction in APCER of 25%. On the other hand, the Object FAS system showed a limited improvement, with an average reduction in APCER of 10%.

In conclusion, the results obtained demonstrate the effectiveness of the strategy of generating synthetic images in improving attack detection capability in the evaluated FAS systems. However, it was observed that the type of image dataset used in Fine-Tuning can have a significant impact on the results. Further research is needed to determine the most appropriate type of image dataset for each specific FAS system.

6 Conclusions and Future Work

This study employed strategies to artificially simulate presentation attacks and digital manipulation, resulting in the generation of new fake samples to complement existing anti-spoof datasets. The explored strategies have demonstrated improvements in the detection performance of three selected FAS systems, and the results are anticipated to provide a valuable resource for future research in the field.

FAS models trained without fine-tuning on the generated data exhibited a bad performance in terms of APCER metrics, with an average rate of 70% across the three FAS systems. However, after applying fine-tuning techniques, a noteworthy enhancement in APCER was achieved, with an average rate of 53% for the three FAS systems. The implementation of data augmentation techniques utilizing synthetic image generation exhibited positive outcomes, particularly for the Silent FAS and Liveness Detection FAS models. The Average Presentation Classification Error Rate (APCER) demonstrated improvements of 25% and 83% respectively.

In future research, further exploration will be conducted to investigate the correlation between the optimal type of image dataset for augmentation and the prevailing landscape of FAS architectures documented in the existing literature. Drawing from the knowledge obtained through the literature review, it is planned to assess FAS architectures that utilize feature extraction techniques, such as Local Binary Patterns (LBP) [17], Speeded-Up Robust Features (SURF) [20], and Difference of Gaussians (DoG) [9], for image representation. These approaches effectively capture the intrinsic textures inherent in the images. Lastly, to explore novel possibilities in synthetic image generation, we plan to implement and investigate state-of-the-art architectures such as CUT [16], CycleGAN [26], and DCLGAN [6]. Building upon recent successes reported in the literature, these approaches offer the potential for advancing the field of synthetic image generation for FAS systems.

References

1. Cai, R., Li, Z., Wan, R., Li, H., Hu, Y., Kot, A.C.: Learning meta pattern for face anti-spoofing. CoRR abs/2110.06753 (2021). https://arxiv.org/abs/2110.06753
2. Feng, H., et al.: Learning generalized spoof cues for face anti-spoofing. CoRR abs/2005.03922 (2020). https://arxiv.org/abs/2005.03922
3. George, A., Marcel, S.: Deep pixel-wise binary supervision for face presentation attack detection. CoRR abs/1907.04047 (2019). http://arxiv.org/abs/1907.04047
4. George, A., Mostaani, Z., Geissenbuhler, D., Nikisins, O., Anjos, A., Marcel, S.: Biometric face presentation attack detection with multi-channel convolutional neural network. CoRR abs/1909.08848 (2019). http://arxiv.org/abs/1909.08848
5. Guo, J., Zhu, X., Xiao, J., Lei, Z., Wan, G., Li, S.Z.: Improving face anti-spoofing by 3D virtual synthesis. CoRR abs/1901.00488 (2019). http://arxiv.org/abs/1901.00488

6. Han, J., Shoeiby, M., Petersson, L., Armin, M.A.: Dual contrastive learning for unsupervised image-to-image translation. CoRR abs/2104.07689 (2021). https://arxiv.org/abs/2104.07689
7. Hao, H., Pei, M.: Face liveness detection based on client identity using Siamese network. CoRR abs/1903.05369 (2019). http://arxiv.org/abs/1903.05369
8. Hernandez-Ortega, J., Fiérrez, J., Morales, A., Galbally, J.: Introduction to presentation attack detection in face biometrics and recent advances. CoRR abs/2111.11794 (2021). https://arxiv.org/abs/2111.11794
9. Huang, C., Huang, J.: A fast HOG descriptor using lookup table and integral image. CoRR abs/1703.06256 (2017). http://arxiv.org/abs/1703.06256
10. Innovatrics: Liveness Detection for Remote Identity Verification Solutions (2022). https://www.innovatrics.com/digital-onboarding-toolkit/liveness-detection/
11. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. CoRR abs/1812.04948 (2018). http://arxiv.org/abs/1812.04948
12. Kortli, Y., Maher, J., Alfalou, A., Atri, M.: A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF for security and face recognition (2018). https://doi.org/10.1088/978-0-7503-1457-2ch13
13. Liu, A., Tan, Z., Wan, J., Escalera, S., Guo, G., Li, S.Z.: CASIA-SURF CeFA: a benchmark for multi-modal cross-ethnicity face anti-spoofing. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1178–1186. IEEE Computer Society, Los Alamitos (2021). https://doi.org/10.1109/WACV48630.2021.00122, https://doi.ieeecomputersociety.org/10.1109/WACV48630.2021.00122
14. Liu, A., et al.: Cross-ethnicity face anti-spoofing recognition challenge: a review. CoRR abs/2004.10998 (2020). https://arxiv.org/abs/2004.10998
15. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Gool, L.V.: RePaint: inpainting using denoising diffusion probabilistic models. CoRR abs/2201.09865 (2022). https://arxiv.org/abs/2201.09865
16. Park, T., Efros, A.A., Zhang, R., Zhu, J.: Contrastive learning for unpaired image-to-image translation. CoRR abs/2007.15651 (2020). https://arxiv.org/abs/2007.15651
17. Rahim, M.A., Azam, M.S., Hossain, N., Islam, M.R.: Face recognition using local binary patterns (LBP). Glob. J. Comput. Sci. Technol. **13**, 1–8 (2013)
18. sarah.merker@iproov.com: Deepfake Statistics & Solutions – Protect Against Deepfakes (2022). https://www.iproov.com/blog/deepfakes-statistics-solutions-biometric-protection
19. Uricár, M., Krízek, P., Hurych, D., Sobh, I., Yogamani, S.K., Denny, P.: Yes, we GAN: applying adversarial techniques for autonomous driving. CoRR abs/1902.03442 (2019). http://arxiv.org/abs/1902.03442
20. Verma, R., Kaur, M.R.: Enhanced character recognition using surf feature and neural network technique (2014)
21. Wan, J., Guo, G., Escalera, S., Escalante, H.J., Li, S.Z.: Multi-modal Face Presentation Attack Detection. Synthesis Lectures on Computer Vision. Morgan & Claypool Publishers (2020)
22. Wang, M., Deng, W.: Deep face recognition: a survey. CoRR abs/1804.06655 (2018). http://arxiv.org/abs/1804.06655
23. Yu, Z., Li, X., Niu, X., Shi, J., Zhao, G.: Face anti-spoofing with human material perception. CoRR abs/2007.02157 (2020). https://arxiv.org/abs/2007.02157
24. Yu, Z., Li, X., Shi, J., Xia, Z., Zhao, G.: Revisiting pixel-wise supervision for face anti-spoofing. CoRR abs/2011.12032 (2020). https://arxiv.org/abs/2011.12032

25. Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., Zhao, G.: Deep learning for face anti-spoofing: a survey. CoRR abs/2106.14948 (2021). https://arxiv.org/abs/2106.14948
26. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. CoRR abs/1703.10593 (2017). http://arxiv.org/abs/1703.10593