Discover Artificial Intelligence

Discover

Research

Presentation attack detection: an analysis of spoofing in the wild (SiW) dataset using deep learning models

Niraj Thapa[1] · **Meenal Chaudhari**[2] · **Kaushik Roy**[1]

Received: 21 March 2023 / Accepted: 9 August 2023
Published online: 22 August 2023

Abstract
Presentation attacks are executed to attain illegitimate access to the system. They are categorized by their mode of action as a print attack, replay attack, and spoof attack, and by their media of action as iris, biometrics, fingerprint, and face. Though there has been a rise in computational algorithms and models to detect presentation attack, generalization across different datasets remain an essential aspect of performance measure. In this paper, we present presentation attack detection (PAD) and presentation attack types of classification (PATC) models based on convolutional neural networks (CNN). We utilize the different attacks presented on the Spoofing in the wild (SiW) dataset to build these models. The PAD-CNN model is developed with a minimal footprint to optimize training time. High-performing models such as Mobilenet and Inceptionv3 are also used in this research work. In this study, we perform an independent test on images extracted from videos of both seen and unseen subjects. Overall, PAD-CNN performed better or on par with Mobilenet and Inceptionv3, even with less training time. Furthermore, these models were also trained to classify the types of presentation attacks with good results. The benchmarking of these models were done on two different datasets, NUAA photo imposter database and Replay-attack database utilizing transfer learning. Together, these results suggest the robustness and effectiveness of the proposed presentation attack detection models based on CNN on the SiW dataset.

Keywords Presentation attack · Deep learning · Machine learning · CNN

Abbreviations

ML	Machine learning
DL	Deep learning
CNN	Convolutional neural network
PAD	Presentation attack detection
PATC	Presentation attack type classification

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s44163-023-00077-1.

✉ Kaushik Roy, kroy@ncat.edu | [1]Department of Computer Science, North Carolina A&T State University, Greensboro, NC 27411, USA. [2]Department of Biology, North Carolina A&T State University, Greensboro, NC 27411, USA.

1 Introduction

In order to secure systems, different types of biometric authentication systems have been employed, such as the face, iris, and fingerprint authenticators. Computational advancements have taken place to determine presentation attacks on such systems. A presentation attack can be defined as an act of breaching the system by the presence of an illegitimate subject, either by impersonation or hiding its identity. Such attacks are grossly categorized as print attack [1], where a printed photo is presented to the authenticator; replay attacks, where a video of the subject is shown; and masked attacks, where objects are used to hide the identity. Both static and dynamic methods have been employed in the detection of presentation attacks. A static attack is less computationally intensive, as it considers the information per video frame or images. The dynamic attack allows generalization and aims at the detection of liveness in the subject, such as eye movement and pulse; thus, is computationally intensive. In this research work, the static detection of presentation attacks has been used with prospect of dynamic detection in future using live videos.

Traditionally, the static detection methods look for handcrafted features such as distortion in the image [2], Moiré-effect [3], color distortion, and shape deformation [4]. Like in other fields [5, 6], deep learning methods drift from handcrafted features to computationally learned features. Feature-based classifiers have shown promise in cross-dataset evaluations [7], while deep learning-based models show better generalization.

Further, the research space has grown with publicly available datasets such as CASIA-SURF [8], CRMA[9], and continual face presentation attack detection challenges organized by conferences. Such a competitive environment provides a new direction for research. While the initial datasets focused on handcrafted features such as image quality, color features, and frequency patterns, the use of learned feature representations and ensemble techniques have overtaken mainly in the recent research work [10–12]. Also, the research question has evolved from assessing the presentation detection algorithm to improving the generalization of the algorithms. In our work, we use the Spoofing in the wild (SiW) dataset developed by Liu et al. [13] to build our presentation attack, detection model. To our best knowledge, performance metrics of any machine learning or deep learning models on the SiW dataset have not been published.

Previous research focus has been mainly on detecting the presentation attack. In this research work, we have trained different deep learning models to identify types of presentation attacks as well. Using SiW dataset, our models will be able to determine if its presentation attack or not; furthermore, they will also be able to classify if its print attack or replay attack.

In this research work, we have developed models for presentation attack detection (PAD) and presentation attack type classification (PATC) using SiW datasets. Finally, we implement transfer learning and perform benchmark tests on two datasets: NUAA photo imposter database [14] and Replay-attack database [15]. The contribution of this research can be summarized as follows: 1) comparative analysis using different deep learning models, 3) development of the PAD-CNN model for presentation attack detection and presentation attack type classification, and 3) different sets of independent sets including benchmark tests.

The rest of the paper is organized as follows. Section 2 includes a literature review. Section 3 covers materials and methods, which description of the dataset, preprocessing, and different DL models. Section 4 presents results and analysis, and Sect. 5 is a discussion and conclusions.

2 Related work

The research in presentation attack detection is supported by consistent, collaborative competition that summarizes the limitations and future directions in the field. The first competition [16] defined the evaluation algorithm with the case of detecting 2D face spoofing attacks, while the second competition [17] expanded the detection of attacks to display attacks and defined the evaluation algorithm. Later competitions [10, 12, 18, 19] took advantage of large-scale datasets and deep learning algorithms to improve their performance. A multiclass model [20] was proposed to detect 3D high-fidelity mask face presentation attack detection in ICCV2021. With increased ease of defeating face authentication with improved technology, there is a need for an improved algorithm that can detect attacks with precision and speed.

Likewise, the algorithms in initial competitions were focused on handcrafted features such as local binary pattern (LBP) [21], and the use of first-order Haar wavelet to distinguish noise in the images. Further, with the availability of

datasets in different modalities- RGB, depth, and NIR, multimodal and unimodal tracks of the algorithms were proposed. There have been many datasets developed [13–15, 22] and provided as open source. Some datasets include print attack [14], and others replay attacks [15]. The availability of different modalities has been exploited using the multi-model techniques [10–12], and fusion techniques [23, 24] has also been employed to improve the model's capability. Khade et al. [25] proposed an ensemble-based approach using the handcrafted features of Discrete Cosine Transform and Haar transform for iris-liveness detection. Though, the ensemble-based model is not a reliable technique to use in real-time situations, as it will take more resources to train in data scarce like mobile settings.

While there are few models that explore the feature fusion methods that integrate the traditional features with deep learning models, MFNET-Leu [24] utilizes Laplacian embedding to discover discriminative features with a multilevel fusion network to detect face presentation attacks. BioPAD [26] applies a biologically inspired method that uses spectral information to fuse at feature and score levels to detect face presentation attack detection.

Convolutional neural networks (CNN) [27] have been a predominant choice in prediction tasks involving videos and images. With the attention mechanisms, many models sprouted based on the CNN in transformers action [28, 29]. The feature-based methods learn the discriminative features of the images, such as the RGB frequency spectrum, to identify live images from fake images. Agarwal et al. [22] use both handcrafted features and CNN-based features to detect 2D video replay attacks with an SVM classifier, emphasizing the impact of brightness and transformations on the error rate of the detection algorithm.

Fang et al. [30] proposed the Attention-based Pixel wise Binary Supervision (A-PBS) network to address the generalization issue in Iris presentation attack detection. The same group did work in partial attack supervision [31]. DFCANet [32] applied channel attention to determine Iris level PAD and showed results on benchmark databases IIITD-CLI, IIIT-CSD, and NDCLD'13 to measure generalizability across the intra-domain and cross-domain scenarios. ViTransPAD [33] applies the attention mechanism on the vision transformers to obtain pixel-level discrimination for face PAD.

Other approaches to propose face PAD systems is involved with detection of liveness in the subject through the observation of gaze and other contributing factors. DeepEyeIdentification [34] live trains on the gaze patterns of an eye to detect the liveness of the subject, and is shown well to perform for replay attacks, is a CNN-based model. TransRPPG [35] is a pure-transformer framework to extract features to support liveness detection for mask attack detection. Chou [36] proposes an algorithm that detects subjects' liveness using a multimodal presentation attack detection method utilizing the score level fusion technique.

The generalizability and the need to measure the performance of PAD algorithms on images from people with not a single image used for training (hereafter termed as unseen subjects) is crucial. Thus, it concludes with the necessity of determining the spoof features that can be used to attain good performance on unseen attacks. A cross-domain prediction has been explored to improve performance on unseen datasets and thus improve generalization. A domain-based generalization technique [37] aims at capturing features by taking into account sample bias and irrelevant domain features, thus improving performance for unseen scenarios. Shao et al. [38] propose a generalized model, FedPAD, while preserving the privacy of face images using the domain-invariant and domain-specific features from the image. Federated Test time Adaptive Face presentation attack detection [39] is another model that maintains the privacy of images and improves generalizability by minimizing entropy on the test data and using a test-time adaptation. Mohammadi et al. [40] explored domain-guided pruning of CNN in a mobile setting and tested the generalization capabilities of the model on intra-domain and cross-domain evaluation. Thus, it concludes with the need to determine the spoof features that can be used to attain good performance on unseen attacks.

However, analysis and development of presentation attack detection systems are lacking for different types of attacks as well as identifying the type of attack itself. Furthermore, models based on CNN have been utilized underwhelmingly, even though CNN is the preferred algorithm for images and videos. In this research, we develop presentation attack detection (PAD) models based on CNN to apply to the Spoofing in the wild (SiW) dataset developed by Liu et al. [13], which contains both printed and replay attacks. SiW provides live and spoof videos from 165 subjects; each subject has 8 live and up to 20 spoof videos, in total 4,478 videos. These models will be further used to classify types of attacks as well. Benchmarking on two different datasets will also be performed to analyze the performance of other models.

3 Materials and methods

In this section, we present the development of a presentation attack detection on the Spoofing in the Wild (SiW) dataset. Overall approach is shown in Fig. 1.

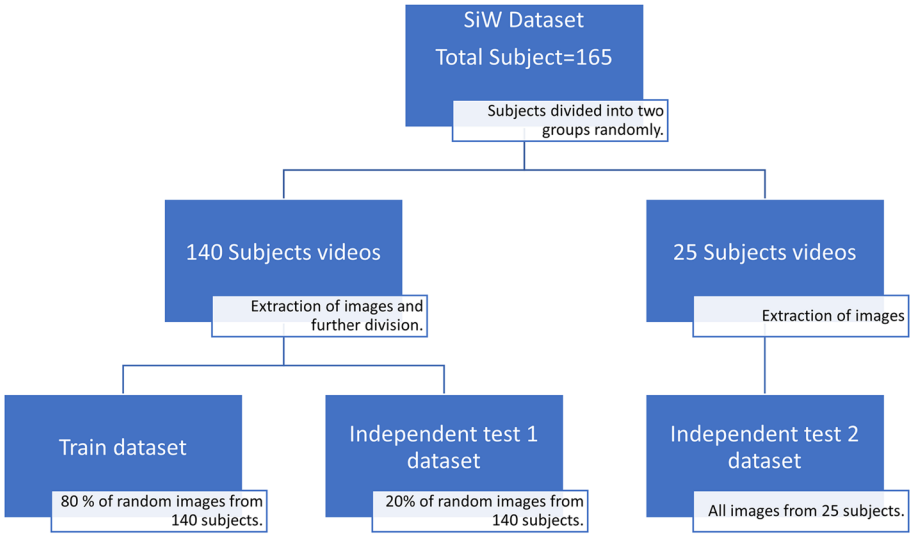

Fig.1 Dataset preprocessing and preparation

3.1 Spoofing in the wild dataset

Spoofing in the wild (SiW) dataset developed by Liu et al. [13] was used primarily for the analysis of the presentation attack detection system. It contains live and spoof videos from 165 subjects with a total of 4478 videos. These are 30 frames per second videos with an average length of 15 s at 1080p resolution. The live videos are collected with variations of distance, pose, illumination and expression. There are two types of spoof videos, printed paper, and replay.

3.1.1 Image extraction

Images were extracted from the videos in SiW dataset for the development of models. Ten frames per second from each video were extracted to build our training and testing dataset. These images were further classified/labeled in two ways. For presentation attack detection (PAD), binary classes, spoof or not, were created. Three classes; live, printed paper, and replay were used for multiclass classification: presentation attack type classification (PATC).

3.1.2 Training and testing dataset

For both binary and multiclass classification, images from 140 subjects were used for initial training and testing. Following the division of the dataset into 80% train and 20% test, it was balanced using undersampling. The initial independent test on this data will be termed Test 1. Images from the remaining 25 subjects will be used for the seconduppl independent test termed Test 2, where the model has none of the images fed from these subjects. Table 1 shows the total number of training and test data for binary classification and multiclass classification, PAD, and PATC, respectively.

3.2 Benchmark datasets

In this research work, we have used the NUAA photo imposter database [14] and Replay-attack database [15] as benchmark datasets. These benchmark datasets were used to analyze the performance of different models, trained and tested on the SiW database for generalizability. NUAA photo imposter database contains images from 15 subjects, whereas Replay-attack contains videos from 50 subjects.

Table 1 Training and test datasets for both PAD and PATC

Dataset	Train	Test 1	Test 2
PAD	169140	42268	56422
PATC	129384	32349	15549

3.3 DL Models

In this research, we have used the PAD-CNN model based on CNN [27], Mobilenets [41], and InceptionV3 [42]. InceptionV3 has the highest number of parameters and the most increased runtime, whereas PAD-CNN has the least number of parameters and the least runtime. These three models were applied as they represent the base model, medium, and high level, respectively. All these three models are used in SiW dataset for both binary and multiclass classification. Furthermore, these models are tested in benchmark datasets as well.

3.4 PAD-CNN

Firstly, the input image is preprocessed with resolution 224×224, batch size 32, and color mode RGB. The preprocessed image input is fed into a first 2D convolutional layer with a filter size of 3×3. The output from the first convolutional layer is fed into the max pooling layer. Similarly, the following output is fed into a set of three convolutional layers and a max pooling layer. The dropout layers are used in between to minimize overfitting and maximize generalization. Next, a flattening layer is used. Dense layers follow the flattening layer. Adam [43] was used as an optimizer for the PAD-CNN architecture; it uses adaptive learning rates to calculate individual learning rates for each parameter. TruncatedNormal is used as layer weight initializer. ModelCheckpoint function is used to save best model from different number of epochs. SoftMax is used as an activation function. It assigns probabilities to each class that sums up to one. PAD-CNN model architecture is shown in Fig. 2. More information on hyperparameter tuning has been added to Additional file 1.

3.4.1 Mobilenet

MobileNet [41] is based on a streamlined architecture that uses depth-wise(dw) separable convolutions to build lightweight deep neural networks. The Mobilenet architecture is shown in Table 2. It introduces two simple global hyper-parameters that efficiently tradeoff between latency and accuracy. The network consists of 28 convolutional layers and one fully connected layer followed by a SoftMax layer. Batch normalization and ReLU are applied after convolution layers. Similar performance with state-of-the-art approaches but with a much smaller network is achieved using MobileNet, favored by depth-wise separable convolution.

3.4.2 Inception v3

Inception v3 [42] has proved to be more computationally efficient, both in terms of the number of parameters generated by the network and the economic cost. It utilizes factorized convolutions that keep a check on the network efficiency by reducing the number of parameters involved in a network. It consists of smaller convolutions replacing bigger ones in its prior version. The convolutions are asymmetric in nature to reduce the number of parameters. Auxiliary classifier and grid size reduction are applied to improve the model.

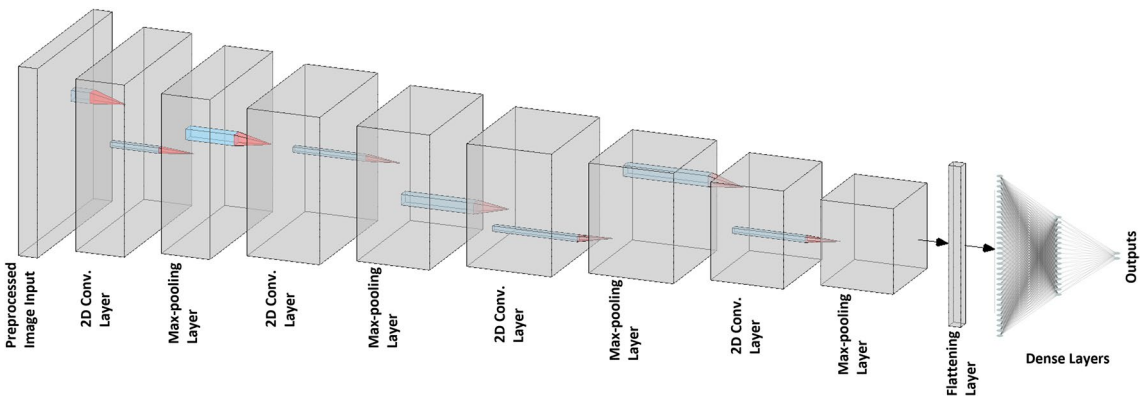

Fig.2 Presentation attack detection (PAD) CNN model architecture

Table 2 Mobilenet architecture with convolutional layers (Conv) and depth-wise(dw) [41]

Type/Stride	Filter Shape	Input Size
Conv/s2	3×3×3×32	224×224×3
Conv dw/s1	3×3×32 dw	112×112×32
Conv/s1	1×1×32×64	112×112×32
Conv dw/s2	3×3×64 dw	112×112×64
Conv/s1	1×1×64×128	56×56×64
Conv dw/s1	3×3×128 dw	56×56×128
Conv/s1	1×1×128×128	56×56×128
Conv dw/s2	3×3×128 dw	56×56×128
Conv/s1	1×1×128×256	28×28×128
Conv dw/s1	3×3×256 dw	28×28×256
Conv/s1	1×1×256×256	28×28×256
Conv dw/s2	3×3×256 dw	28×28×256
Conv/s1	1×1×256×512	14×14×256
5×(Conv dw/s1 and	3×3×512 dw	14×14×512
Conv / s1)	1×1×512×512	14×14×512
Conv dw/s2	3×3×512 dw	14×14×512
Conv/s1	1×1×512×1024	7×7×512
Conv dw/s2	3×3×1024 dw	7×7×1024
Conv/s1	1×1×1024×1024	7×7×1024
Avg Pool/s1	Pool 7×7	7×7×1024
FC/s1	1024×1000	1×1×1024
Softmax/s1	Classifier	1×1×1000

3.5 Performance metrics

In this research work, the first independent test was performed on 20% of the initially held out set. Furthermore, the second independent test was performed on unseen subjects' images.

Confusion matrix, precision, recall, and F1-score were used as performance metrics. For binary classification PAD, the dimension of the confusion matrix is 2×2, and for multiclass classification PATC with three classes, the dimension is 3×3. The diagonal of the matrix gives the counts of true predicted values. It consists of true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN). The following metrics are used, and their values are between 0 to 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{3}$$

$$F1Score = \frac{Precision \times Recall}{Precision + Recall} \times 2 \tag{4}$$

Table 3 Performance metrics for presentation attack detection (PAD)

Test on SiW dataset	Models	Accuracy	Precision	Recall	F1-score
Test 1 (Independent test)	PAD-CNN	1.00	1.00	1.00	1.00
	Mobilenet	1.00	1.00	1.00	1.00
	Inceptionv3	1.00	1.00	1.00	1.00
Test 2 (Independent test on unseen subjects)	PAD-CNN	0.91	0.92	0.91	0.91
	Mobilenet	0.86	0.89	0.86	0.86
	Inceptionv3	0.85	0.88	0.85	0.85

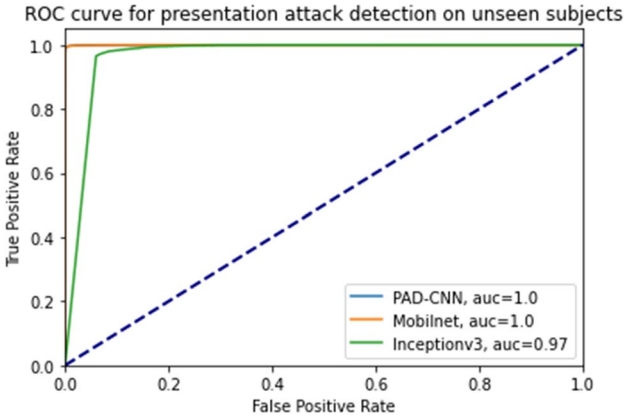

Fig.3 ROC curve for independent test on unseen subjects for PAD

4 Results

In this section, we present the results for presentation attack detection (PAD) and presentation attack type classification (PATC) on the SiW dataset by various implemented DL models. Model benchmarking on the NUAA photo imposter database and Replay-attack database with models trained on the SiW dataset is also included.

4.1 Presentation attack detection (PAD)

Different DL models were applied for the analysis of the SiW dataset for presentation attack detection, which was binary classification. The dataset was divided into training and test sets with an 80:20 ratio. First, an independent test on holds out 20% test set termed test 1 was done. The results are shown in Table 3. All three models achieved 100% accuracy, precision, recall, and F1 score in test 1. PAD- CNN had a shorter training time with respect to other DL models.

Furthermore, the second independent test was performed on 25 unseen subjects, termed test 2. The results are shown in Table 3 and Fig. 4. The ROC curve is shown in Fig. 3. PAD-CNN achieved the highest accuracy of 91%, with Mobilenet and Inceptionv3 lagging with 86% and 85%, respectively. Similarly, PAD-CNN has a precision of 92%, with MobiliNet and Inceptionv3 getting 89% and 88%, respectively. PAD-CNN has 91% recall, whereas Mobilenet and Inceptionv3 got 86% and 85%, respectively. Furthermore, PAD-CNN achieved 91% on the F1 score, followed by Mobilenet and Inceptionv3 with 86% and 85%, respectively. The lower complexity of the problem could have made the lighter model PAD-CNN perform better or on par with more complex models like Mobilenet and Inceptionv3.

4.2 Presentation attack type classification (PATC)

The next step in this analysis was the multiclass classification of attack types as well as live video images. In total, there were three classes printed paper, replay, and live video. Like presentation attack detection, three DL models were applied for independent tests on a 20% hold-out dataset: termed test 1. The results are shown in Table 4. All

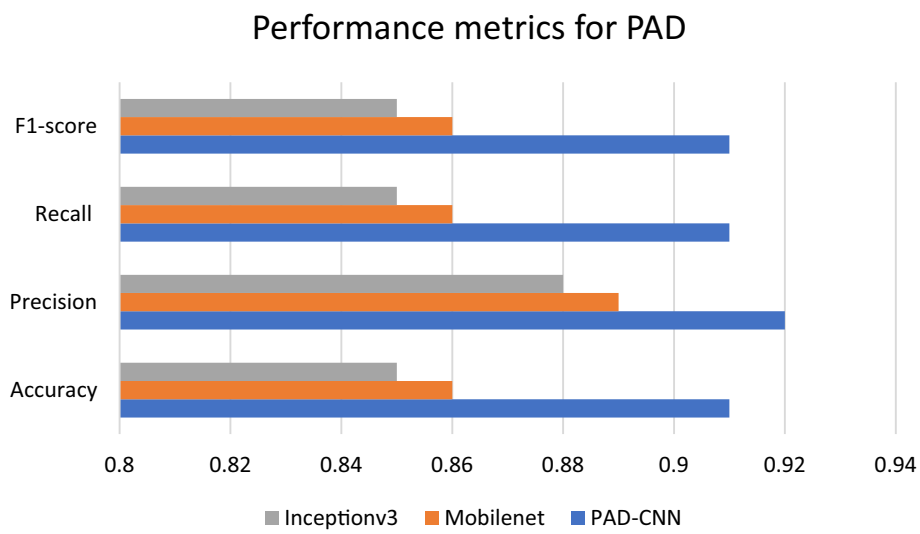

Fig.4 Performance metrics for presentation attack detection (PAD) for different models

Table 4 Performance metrics for presentation attack type classification (PATC)

Test on Siw dataset	Models	Accuracy	Precision	Recall	F1-score
Test 1 (Independent test)	PAD-CNN	1.00	1.00	1.00	1.00
	Mobilenet	1.00	1.00	1.00	1.00
	Inceptionv3	1.00	1.00	1.00	1.00
Test 2 (Independent test on unseen subjects)	PAD-CNN	0.89	0.90	0.89	0.89
	Mobilenet	0.92	0.93	0.92	0.92
	Inceptionv3	0.89	0.91	0.89	0.88

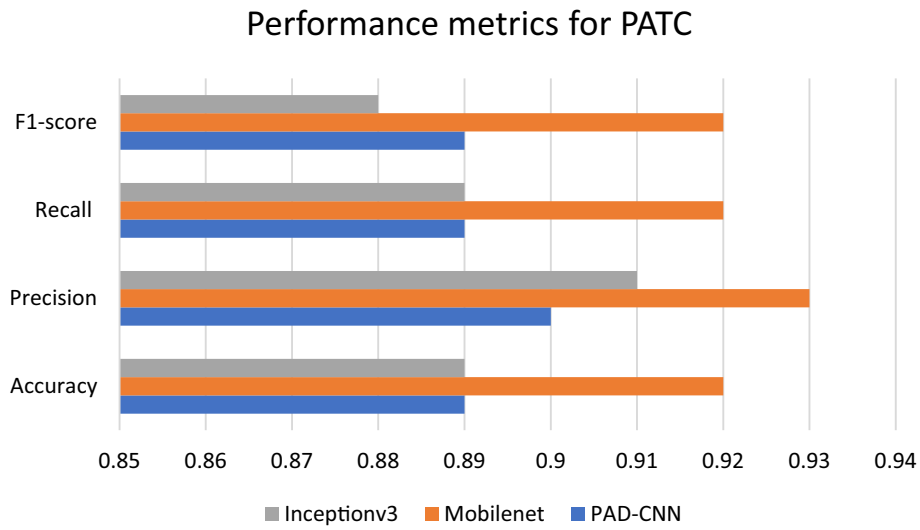

Fig.5 Performance metrics for presentation attack type classification (PATC) for different models

three models achieved 100% on all performance metrics: accuracy, precision, recall, and F1-score. Training times per epoch for PAD-CNN, Mobilenet, and Inceptionv3 were ~ 1 h, ~ 2 h, and ~ 6 h, respectively.

Furthermore, the second independent test was done on unseen 25 subjects' datasets like PAD analysis. The results are shown in Table 4 and Fig. 5. Mobilenet achieved the highest accuracy of 92%, with Inceptionv3 and PAD-CNN achieving 89% and 89%, respectively. Mobilenet gained 93% on precision, followed by Inceptionv3 and PAD-CNN with 91% and 90%, respectively. Similarly, Mobilenet has a recall of 92%, with Inceptionv3 and PAD-CNN getting 89% and 89%, respectively. Mobilenet has a 92% F1 score, whereas Inceptionv3 and PAD-CNN got 88% and 89%, respectively. By adding complexity to the analysis, Mobilenet performed well compared to PAD-CNN and Inceptionv3.

Table 5 Performance metrics for benchmarking on NUAA and Replay-attack dataset

Dataset	Models	Accuracy	Precision	Recall	F1-score
NUAA	PAD-CNN	1.00	1.00	1.00	1.00
	Mobilenet	1.00	1.00	1.00	1.00
	Inceptionv3	1.00	1.00	1.00	1.00
Replay-attack	PAD-CNN	0.98	0.98	0.98	0.96
	Mobilenet	1.00	1.00	1.00	1.00
	Inceptionv3	0.97	0.97	0.97	0.97

4.3 Model benchmarking

We implemented transfer learning by utilizing a trained model on the SiW dataset to train and test on benchmark datasets; NUAA photo imposter database [14] and Replay-attack database [15]**.** All three trained deep learning models were used for benchmarking. Independent tests on the NUAA dataset yielded 100% performance metrics (accuracy, precision, recall, and F1-score) for all three models. The results are shown in Table 5.

Mobilenet achieved 100% accuracy, precision, recall, and F1-score on independent tests on the Replay-attack dataset. Both PAD-CNN and Inceptionv3 performed highly in this test. All DL models performed well in the independent tests on both benchmark datasets, with high-performance metrics.

4.4 Conclusion and discussions

In this research work, we developed presentation attack detection (PAD) and presentation attack types of classification (PATC) models for the SiW dataset. We used the PAD-CNN model based on CNN, Mobilenet, and Inceptionv3 for both binary and multiclass classification.

Initially, for both binary and multiclass classification, images from 140 subjects were used for initial training and testing, following the division of the dataset into 80% train and 20% test. Images from the remaining 25 subjects were used for the second independent test, where the model has none of the images fed from these subjects.

For PAD, all models achieved the highest performance metrics on the independent test on the 20% hold-out set. PAD-CNN achieved the highest performance metrics for independent tests on images from 25 unseen subjects. Mobilenet and Inceptionv3 were slightly lower in performance than PAD-CNN, whose accuracy was 91%.

With the added complexity of multiclass classification to identify types of attack, Mobilenet performed slightly better than PAD-CNN and Inceptionv3 on independent tests on unseen subjects. Mobilenet was able to achieve 92% accuracy.

Furthermore, transfer learning was implemented by using trained models on the SiW dataset to train and test on benchmark datasets (NUAA photo imposter database and Replay-attack database). Independent tests on the NUAA dataset yielded 100% performance metrics (accuracy, precision, recall, and F1-score) for all three models. Mobilenet achieved 100% accuracy, precision, recall, and F1-score on independent tests on the Replay-attack dataset. Both PAD-CNN and Inceptionv3 performed highly in this test. Training times per epoch for PAD-CNN, Mobilenet, and Inceptionv3 were ~ 1 h, ~ 2 h, and ~ 6 h, respectively. Mobilenet and Incepetionv3 performance are slightly better than PAD-CNN and current state of the art models are expected to have better performance as well. However, at high accuracy slight performance boost does come with high computational cost.

For images and videos, models based on CNN from low to high complexity have been widely preferred and used. In this research work we have presented PAD-CNN, the model based on CNN with lower complexity than Mobilnet and Inceptionv3; both CNN based models. PAD-CNN was able to perform in par or even better in some scenarios than these recent models with half the runtime. With the addition of more data overtime, this runtime difference will only grow larger. Therefore, in terms of performance cost and high tuning capabilities, PAD-CNN is a viable option for practicality and general use for presentation attack detection.

For future work, new datasets with added attacks need to be developed. With improvements in datasets, further optimization of models will be required to compensate for the added complexities. Furthermore, to deal with the black box nature of the deep learning models, research on the interpretability side of it is required. With these

developments, presentation attacks can be nullified by the dynamic nature of the presentation attack detection systems.

Acknowledgements This research is partially supported by National Science Foundation (NSF) and Siemens. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF and Siemens.

Author contribution NT developed PAD models, and wrote the draft, MC revised the draft, KR conceived and designed the experiment, and revised the manuscript.

Availability of data and materials The datasets used and analyzed during the current study are available from SiW: Spoofing in the Wild Database [http://cvlab.cse.msu.edu/siw-spoof-in-the-wild-database.html].

Declarations

Competing interests The authors declare no competing interests.

References

1. George A, Mostaani Z, Geissenbuhler D, Nikisins O, Anjos A, Marcel S. Biometric face presentation attack detection with multi-channel convolutional neural network. IEEE Trans Inform Forensic Secur. 2020. https://doi.org/10.1109/TIFS.2019.2916652.
2. Wen D, Han H, Jain AK. Face spoof detection with image distortion analysis. IEEE Trans Inf Forensics Secur. 2015;10(4):746–61. https://doi.org/10.1109/TIFS.2015.2400395.
3. Patel K, Han H, Jain AK, Ott G. Live face video vs. spoof face video: use of moiré patterns to detect replay video attacks. Int Conf Biom. 2015. https://doi.org/10.1109/ICB.2015.7139082.
4. Abdullakutty F, Elyan E, Johnston P. A review of state-of-the-art in face presentation attack detection: from early development to advanced deep learning and multi-modal fusion methods. Inform Fusion. 2021;75:55–69. https://doi.org/10.1016/j.inffus.2021.04.015.
5. Thapa N, et al. A deep learning based approach for prediction of Chlamydomonas reinhardtii phosphorylation sites. Sci Reports. 2021;11(1):12550. https://doi.org/10.1038/s41598-021-91840-w.
6. Thapa N, Liu Z, Shaver A, Esterline A, Gokaraju B, Roy K. Secure cyber defense: an analysis of network intrusion-based dataset CCD-IDSv1 with machine learning and deep learning models. Electronics. 2021;10(15):1747. https://doi.org/10.3390/electronics10151747.
7. Arashloo SR, Kittler J, Christmas W. An Anomaly detection approach to face spoofing detection: a new formulation and evaluation protocol. IEEE Access. 2017;5:13868–82. https://doi.org/10.1109/ACCESS.2017.2729161.
8. Zhang S, et al. CASIA-SURF: a large-scale multi-modal benchmark for face anti-spoofing. arXiv. 2019. https://doi.org/10.48550/arXiv.1908.10654.
9. Fang M, Damer N, Kirchbuchner F, Kuijper A. Real masks and spoof faces: on the masked face presentation attack detection. Pattern Recognit. 2022;123:108398. https://doi.org/10.1016/j.patcog.2021.108398.
10. Purnapatra S, et al. "Face liveness detection competition (LivDet-Face). IEEE Int Joint Conference Biom (IJCB). 2021. https://doi.org/10.1109/IJCB52358.2021.9484359.
11. Boutros F, et al. MFR 2021: masked face recognition competition. arXiv. 2021. https://doi.org/10.48550/arXiv.2106.15288.
12. Liu A, et al. Cross-ethnicity face anti-spoofing recognition challenge: a review. arXiv. 2020. https://doi.org/10.48550/arXiv.2004.10998.
13. Y Liu, A Jourabloo, X Liu, "Learning deep models for face anti-spoofing: binary or auxiliary supervision," in 2018 IEEE/CVF Conference on computer vision and pattern recognition. 2018;389–398. https://doi.org/10.1109/CVPR.2018.00048.
14. Tan X, Li Y, Liu J, Jiang L. Face Liveness Detection from a single image with sparse low rank bilinear discriminative model. In: Daniilidis K, Maragos P, Paragios N, editors. Computer Vision—ECCV 2010: 11th European conference on computer vision, Heraklion, Crete, Greece, September 5–11, 2010, proceedings, part VI. Berlin, Heidelberg: Springer, Berlin Heidelberg; 2010.
15. I Chingovska, A Anjos, S Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in 2012 BIOSIG—Proceedings of the international conference of biometrics special interest group (BIOSIG). 1–7.
16. MM Chakka et al. "Competition on counter measures to 2-D facial spoofing attacks," in 2011 International Joint Conference on Biometrics (IJCB). 2011;1–6. https://doi.org/10.1109/IJCB.2011.6117509.
17. I Chingovska et al. 2013 "The 2nd competition on counter measures to 2D face spoofing attacks," in 2013 International Conference on Biometrics (ICB). 2013;1–6. https://doi.org/10.1109/ICB.2013.6613026.
18. Liu A, et al. 3D high-fidelity mask face presentation attack detection challenge. arXiv. 2021. https://doi.org/10.48550/arXiv.2108.06968.
19. Zhang Y, et al. "CelebA-spoof challenge 2020 on face anti-spoofing: methods and results." arXiv. 2020. https://doi.org/10.48550/arXiv.2102.12642.

20. S Chen et al. "A dual-stream framework for 3D mask face presentation attack detection," in 2021 IEEE/CVF International conference on computer vision workshops (ICCVW). 2021. https://doi.org/10.1109/ICCVW54120.2021.00098.
21. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell. 2002;24(7):971–87. https://doi.org/10.1109/TPAMI.2002.1017623.
22. Agarwal A, Singh R, Vatsa M, Noore A. "Boosting face presentation attack detection in multi-spectral videos through score fusion of wavelet partition images," (in eng). Front Big Data. 2022;5:836749. https://doi.org/10.3389/fdata.2022.836749.
23. Abdullakutty F, Johnston P, Elyan E. Fusion methods for face presentation attack detection. Sensors. 2022. https://doi.org/10.3390/s22145196.
24. Niu S, Qu X, Chen J, Gao X, Wang T, Dong J. MFNet-LE: multilevel fusion network with Laplacian embedding for face presentation attacks detection. IET Image Proc. 2021;15(14):3608–22. https://doi.org/10.1049/ipr2.12308.
25. Khade S, Gite S, Thepade S, Pradhan B, Alamri A. Detection of iris presentation attacks using hybridization of discrete cosine transform and Haar transform with machine learning classifiers and ensembles. IEEE Access. 2021. https://doi.org/10.1109/ACCESS.2021.3138455.
26. Tsitiridis A, Conde C, Gomez Ayllon B, Cabello E. Bio-inspired presentation attack detection for face biometrics. Front Comput Neurosci. 2019;13:34. https://doi.org/10.3389/fncom.2019.00034.
27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.
28. Mazzia V, Angarano S, Salvetti F, Angelini F, Chiaberge M. Action transformer: a self-attention model for short-time pose-based human action recognition. Pattern Recognition. 2022;124:108487. https://doi.org/10.1016/j.patcog.2021.108487.
29. Vaswani A, et al. Attention is all you need. arXiv. 2017. https://doi.org/10.48550/arXiv.1706.03762.
30. Fang M, Damer N, Boutros F, Kirchbuchner F, Kuijper A. Iris presentation attack detection by attention-based and deep pixel-wise binary supervision network. arXiv. 2021. https://doi.org/10.48550/arXiv.2106.14845.
31. M Fang, F Boutros, A Kuijper, N Damer, "Partial attack supervision and regional weighted inference for masked face presentation attack detection," in 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021). 2021. https://doi.org/10.1109/FG52635.2021.9667051.
32. Jaswal G, Verma A, Dutta Roy S, Ramachandra R. DFCANet: dense feature calibration-attention guided network for cross domain iris presentation attack detection. arXiv. 2021. https://doi.org/10.48550/arXiv.2111.00919.
33. Ming Z, Yu Z, Al-Ghadi M, Visani M, MuzzamilLuqman M, Burie J-C. ViTransPAD: video transformer using convolution and self-attention for face presentation attack detection. arXiv. 2022. https://doi.org/10.48550/arXiv.2203.01562.
34. Makowski S, Prasse P, Reich DR, Krakowczyk D, Jäger LA, Scheffer T. DeepEyedentificationLive: oculomotoric biometric identification and presentation-attack detection using deep neural networks. IEEE Trans Biom Behavior Ident Sci. 2021;3(4):506–18. https://doi.org/10.1109/TBIOM.2021.3116875.
35. Yu Z, Li X, Wang P, Zhao G. TransRPPG: remote photoplethysmography transformer for 3D Mask face presentation attack detection. IEEE Signal Process Lett. 2021;28:1290–4. https://doi.org/10.1109/LSP.2021.3089908.
36. Chou C-L. Presentation attack detection based on score level fusion and challenge-response technique. J Supercomput. 2021;77(5):4681–97. https://doi.org/10.1007/s11227-020-03461-1.
37. Liu S, et al. Dual reweighting domain generalization for face presentation attack detection. arXiv. 2021. https://doi.org/10.48550/arXiv.2106.16128.
38. Shao R, Perera P, Yuen PC, Patel VM. Federated generalized face presentation attack detection. IEEE Trans Neural Networks Learning Syst. 2022. https://doi.org/10.1109/TNNLS.2022.3172316.
39. R Shao, B Zhang, PC Yuen, VM Patel, "Federated test-time adaptive face presentation attack detection with dual-phase privacy preservation," in 2021 16th IEEE international conference on automatic face and gesture recognition (FG 2021). 2021. https://doi.org/10.1109/FG52635.2021.9666952.
40. A. Mohammadi, S. Bhattacharjee, and S. Marcel, "Domain adaptation for generalization of face presentation attack detection in mobile settengs with minimal information," in ICASSP 2020–2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2020. https://doi.org/10.1109/ICASSP40776.2020.9053685.
41. Howard AG, et al. Mobile Nets: efficient convolutional neural networks for mobile vision applications. arXiv. 2017. https://doi.org/10.48550/arXiv.1704.04861.
42. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. arXiv. 2015. https://doi.org/10.48550/arXiv.1512.00567.
43. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv. 2014. https://doi.org/10.48550/arXiv.1412.6980.