Presentation Attack Detection Using Wavelet Transform and Deep Residual Neural Net

Prosenjit Chatterjee[1(✉)], Alex Yalchin[2(✉)], Joseph Shelton[3(✉)],
Kaushik Roy[1(✉)], Xiaohong Yuan[1(✉)], and Kossi D. Edoh[4(✉)]

[1] Department of Computer Science, North Carolina A&T State University,
Greensbor, USA
pchatterjee@aggies.ncat.edu, {kroy,xhyuan}@ncat.edu

[2] Department of Computer Science, Elon University, Elon, NC, USA
ayalcin@elon.edu

[3] Department of Engineering and Computer Science, Virginia State University,
Petersburg, VA, USA
jshelton@vsu.edu

[4] Department of Mathematics, North Carolina A&T State University,
Greensbor, USA
kdedoh@ncat.edu

Abstract. Biometric authentication is becoming more prevalent for secured authentication systems. However, the biometric systems can be deceived by the imposters in several ways. Among other imposter attacks, print attacks, mask-attacks, and replay-attacks fall under the presentation attack category. The biometric images, especially iris and face, are vulnerable to different presentation attacks. This research applies deep learning approaches to mitigate the presentation attacks in a biometric access control system. Our contribution in this paper is two-fold: first, we applied the wavelet transform to extract the features from the biometric images. Second, we modified the deep residual neural net and applied it on the spoof datasets in an attempt to detect the presentation attacks. This research applied deep learning technique on three biometric spoof datasets: ATVS, CASIA two class, and CASIA cropped image sets. The datasets used in this research contain images that are captured both in a controlled and uncontrolled environment along with different resolution and sizes. We obtained the best accuracy of 93% on the ATVS Iris dataset. For CASIA two class and CASIA cropped datasets, we achieved test accuracies of 91% and 82%, respectively.

Keywords: Biometrics · Wavelet transform · Deep residual neural network · Presentation attack detection

1 Introduction

Biometric authentication uses an individual's identity for access control and has been widely implemented for controlling the secured gateway of the member's login [1, 2]. Several organizations validate their members' access through biometric-enabled

G. Wang et al. (Eds.): SpaCCS 2019 Workshops, LNCS 11637, pp. 86–94, 2019.
https://doi.org/10.1007/978-3-030-24900-7_7

surveillance and security systems. The earlier biometric authentication techniques utilized physiological traits such as fingerprint, face, iris, periocular region, voice, heart rate, and body mass. Biometric authentication systems have evolved to make use of individual behavioral patterns such as touch pattern, keystroke dynamics, etc., to distinguish real and fake identity. It is evident that the human iris and facial biometric image samples are vulnerable to different types of presentation attacks [1, 2]. Though the presentation attack falls under the ‘hacking attack’ category, it is also referred to as a replay attack. There are several other vulnerable points that the hackers exploit to compromise a biometric-based authentication system. Fake human faces can be created through 3D printing devices, by the use of 3D Mask, or presenting an identical twin, or similar looking individual, to deceive an authentication system. Additionally, the human iris can be copied using textured contact lenses to deceive an iris-based authentication system. A key research focus is to build a robust classification technique that can identify the smallest deviation on real and fake iris and facial images in order to detect presentation attacks.

Recently, deep convolutional neural networks have been used to mitigate presentation attacks [3–5]; however, most of them require huge computational time to train and classify real and fake image samples. To counter the presentation attacks, a high-resolution image set is required. However, high-quality images increase computational complexity during training, validation, and classification. To address this issue, we apply a deep learning approach that can detect spoofing attacks with less computational effort and time.

In this paper, we apply the Wavelet Transform [6] on image datasets to extract features. Once the feature extraction is done, we feed the extracted features to a modified Residual Neural Net, denoted as ‘modified-ResNet’ inspired by the Residual Neural Net (ResNet) reported in [7], for accurate classification. The ResNet has superior features such as batch normalization, parameter optimization, and reduced error through skip layer connection techniques, with surprisingly less computational complexity. We observe a significant improvement in classification time with high accuracy, to distinguish real and fake images.

The rest of the paper is organized as follows. Section 2 discusses our related work. Section 3 describes the proposed methodology. Section 4 discusses the datasets used for this research effort, Sect. 5 shows the experimental results and discussion, and Sect. 6 provides us with conclusions.

2 Related Work

A reliable biometric recognition system has long been a prominent goal to mitigate presentation attacks in a wide range. Recently, deep learning techniques have been used for presentation attack detection (PAD) and have become instrumental to many secured organizations where biometric authentication is mandatory [4]. Some specific well-known spoofing instruments, like silicon masks, are widely used by attackers. Yang et al. [3] used numerous testing protocols and implemented facial localization, spatial augmentation, and temporal augmentation, to diverse feature learning for the deep convolutional neural network (deep CNN). They also experimented with texture based, motion-based, 3D Shape-based detection techniques to identify genuine and fake individuals [6]. Menotti et al. [4] performed research on deep representation for detecting presentation attacks in different biometric substances. Manjani et al. [5] proposed a multilevel deep dictionary learning-based PAD algorithm that can discriminate among different types of attacks. However, the biometric recognition system using the deep CNN technique is completely dependent on recognizing the pattern of test objects with the previously learned training objects. A successful match includes accurate pattern matching of the feature-sets from the test object to the already learned training object. The vital biometric information of an individual, such as faces and iris, are essential for the training. For “mug shot” images of faces taken at least one year apart, even the best current algorithms can have error rates of 43%–50% [8–12]. Henceforth, face and iris images of an individual, when implemented altogether on a biometric recognition system, is proven to be authentic amongst other biometric modalities and many researchers agreed on that.

Several researchers worked on the face and iris recognition system and proposed diverse methods to extract features in detail [13]. Feature extraction through Wavelet involves losses at the edges after the Label 1 decomposition. We investigated the Biorthogonal wavelet transform [6], the Discrete Wavelet Transform (DWT) [6] and 2D-Gabor wavelets [14] and their inverse form. From the experiment, we found that DWT [6] works better with images in the matrix format (n x m) and can convolve in either direction, which facilitates feature augmentation including spatial and temporal augmentation during the deep learning training phase.

In our previous work [15, 16], we applied convolutional neural network (CNN) to mitigate the spoofing attacks and obtained a reasonable performance. After extensive experimental analysis, we found that Residual Neural Network (RNN), especially ResNet [7], is the most effective deep learning approach for training and validation process. RNN utilizes single layer skip connection techniques during learning on the internal convolutional layers and avoids the vanishing gradients issues and optimize the huge parameters efficiently. Additionally, the effective single layer skipping makes the network less complex during the initial training phase, and towards the end of the training, all layers get expanded for detailed level learning.

3 Proposed Methodology

In this research, we used the DWT [6] and its inverse form to elicit features from the face and iris images. We then implemented a ‘modified ResNet’, inspired by the ResNet [7], in an attempt to mitigate the presentation attacks. We trained, validated and tested the ResNet model for the images captured under controlled and uncontrolled environment. The DWT [6] and its inverse form (IDWT) [6] were used to extract the features from the face and iris images. The extracted features are then fed into ‘modified ResNet’ for accurate classification. The high-level flow diagram of our methodology is shown in Fig. 1. In this effort, we used DWT for feature extraction and decomposition, and then applied the inverse form of DWT for reconstruction.

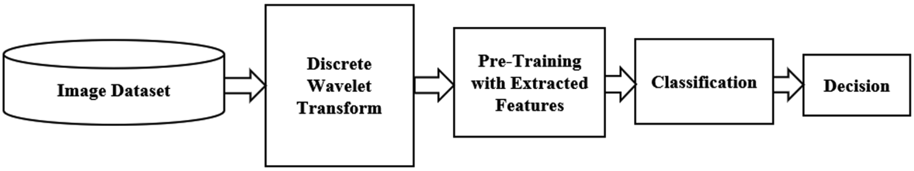


Image Dataset
Discrete Wavelet Transform
Pre-Training with Extracted Features
Classification
Decision

Fig. 1. High level architecture of the proposed methodology

The discrete function is represented as a weighted sum in the space spanned by the bases φ and ψ:

$$x|m| = \frac{1}{\sqrt{N}}\sum_{k} W_{\varphi}[j_{\mathrm{o}},k]\varphi_{j\mathrm{o},k}[m] + \frac{1}{N}\sum_{j=j\mathrm{o}}^{\infty}\sum_{k} W_{\psi}[j,k]\psi_{j,k}[m], \quad (m = 0,\cdots,N-1)$$
$$\ldots.. \tag{1}$$

The inverse wavelet transform, where the summation over j is used for different scale levels and the sum over k is used for different conversions in each scale level:

$$W_{\varphi}[j_{\mathrm{o}},k] = (\mathbf{x},\ \varphi_{j\mathrm{o}},\ k) = \frac{1}{\sqrt{N}}\sum_{m=0}^{N-1} x[m]\varphi_{j\mathrm{o},k} \quad -\forall(k) \quad \ldots. \tag{2}$$

$$W_{\psi}[j,k] = (\mathbf{x},\ \psi_{j,k}) = \frac{1}{\sqrt{N}}\sum_{m=0}^{N-1} x[m]\psi_{j,k}[m] \quad -\forall(k)\&\forall(j>j_0) \quad \ldots. \tag{3}$$

where W_{φ} *[j, k]* is called the *approximation coefficient* and W_{ψ} *[j, k]* is called the *detail coefficient.*

We limit our wavelet decomposition to label-1, in order to prevent data loss at edges. The feature extraction, decomposition, and reconstruction mechanism, which we implemented for our research, is shown in Fig. 2.

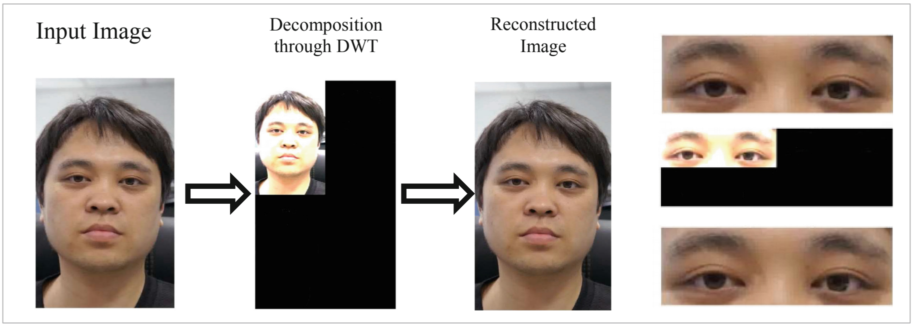
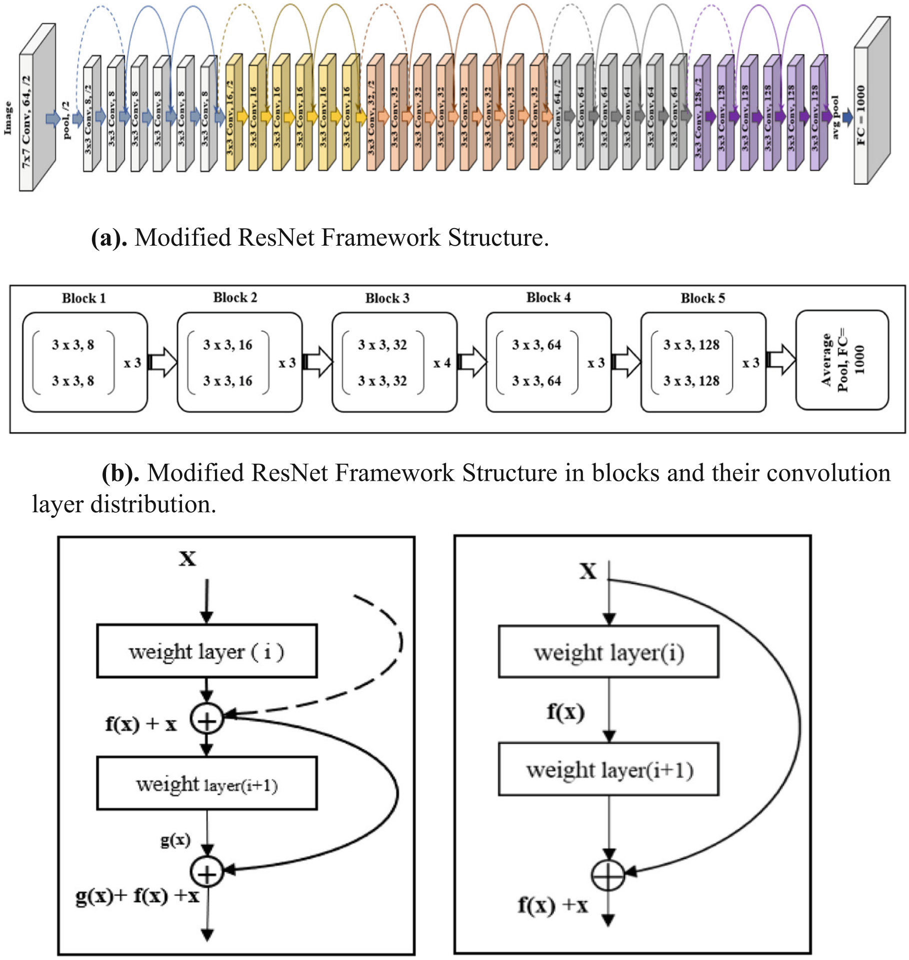


Fig. 2. DWT applied on CASIA [8, 9] sample face antispoofing image.

To prevent presentation attack in a wide range and to save training and execution time, we designed a 'modified ResNet' inspired by the ResNet [7]. He et al. [7] proposed and implemented an 18 layer and 34-layer ResNet, respectively. They experimented skip connection techniques and handled images through the deep CNN structure and showed the advantages they achieved based on computational complexities and low error rate.



(a). Modified ResNet Framework Structure.

(b). Modified ResNet Framework Structure in blocks and their convolution layer distribution.

(c.1) *(c.2)*

(c). Residual Neural Net single layer skip sequential function implemented.
(c.1) – represents single layer skip connection techniques implemented.
(c.2) – represents double layer skip connection techniques implemented.

Fig. 3. (a) Modified ResNet framework structure. (b). Modified ResNet framework structure in blocks and their convolution layer distribution. (c). Residual neural net single layer skip sequential function implemented. *(c.1) – represents single layer skip connection techniques implemented. (c.2) – represents double layer skip connection techniques implemented.*

In comparison, our ‘modified ResNet’ is quite lighter than the ResNet [7], as our ‘modified ResNet’ has a total of 32 discrete convolution 2D layers, 2 Max Pooling 2D layers and 1 Fully Connected (FC) dense layer, shown in Fig. 3(a) and (b). Our ‘modified ResNet’ has 5 building blocks of convolution layers and their distribution maintaining the single layer skip-connection techniques as shown in Fig. 3(c).

4 Datasets Used

Our research methodology is evaluated based on the three different categories of image datasets: ATVS [1, 2], CASIA [8, 9], and CASIA-cropped [8, 9]. ATVS [1, 2] contains the real and fake images of periocular regions. The datasets contain fifty subjects. Subjects had both eyes photographed four times per session, with two different sessions. Each image was then undertaken through gray-scaled printing and successive scanning for fake image generation. Each user contains 32 images, 800 per class (real and fake), and a total of 1600 with a uniform resolution of 640 × 480.

CASIA [8, 9] dataset contains both the high resolution still images and video clips. There are fifty subjects and four classes within the dataset. Every class contains one landscape-style video and one portrait-style video. The four classes are real subjects, “cut photo” attacks (printed photo of subject with eyeholes cut out, real user positioned behind photo to fool blinking detection systems), “wrap photo” (printed photo of subject held up to the camera, photo is moved back and forth to fool liveness detection systems), and video replay attacks (tablet or screen held up to the camera while playing a video of subject).

CASIA cropped is the customized image datasets that were created from the original CASIA images [8, 9] by using OpenCV Haar cascades [17]. Furthermore, we created custom resized CASIA image sets with different resolutions and lower dimensions. In our work, we considered high resolution image as 720 × 1280 (portrait style), and standard-resolution image as 640 × 480 (landscape style) and 480 × 640 (portrait style). We also experimented on custom resized image datasets. For custom resized image datasets, we modified images on different resolutions and dimensions, such as 240 × 320, and 225 × 225. Overall, there are approximately 127,000 images that were used for this work.

5 Results and Discussions

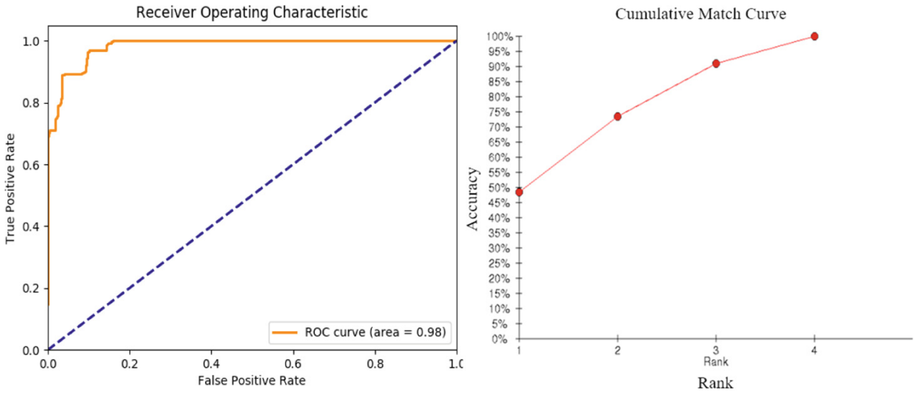
The training pattern of the CNNs varies depending on time and memory availability of CPU/GPUs. To achieve the realistic performances, we ran different implementations of CNNs, including the ‘modified ResNet’, on all the datasets separately for multiple times and observed their performances and minute variations. In Table 1, we compare the test accuracies of the ‘modified ResNet’ with the DWT and ‘modified ResNet’ on ATVS Iris [1, 2], CASIA Two Class [8, 9], and CASIA cropped [8, 9] datasets.

In addition, we compared the test accuracies on our previously implemented 'modified VGGNet' [16], and with DWT and the 'modified VGG Net' [16]. We achieved highest accuracies when we use the DWT with the 'modified ResNet'. The proposed approach takes the lowest computational time compared to the other approaches as given in Table 1.

Table 1. Classification accuracies and average execution time of different classification techniques implemented

Classification techniques used	Datasets used			Average execution time for ATVS Iris (in seconds)
	ATVS Iris	CASIA Two Class	CASIA Cropped	
Modified ResNet [34 layers]	94.40	91.0	85.70	1,962
DWT + 'Modified ResNet' [34 layers]	92.57	90.80	82.4	1311
Modified VGG Net' [19 layers] [16]	97.00	96.00	95.00	1459
DWT + 'Modified VGG Net' [19 layers] [16]	89.00	86.00	78.00	3045

The Receiver Operating Characteristic curve (ROC) for the ATVS Iris datasets shows a True Positive Rate (TPR) of 98%, as shown in Fig. 4(a). The average error rate for our proposed model is in the range of (3.6%–5.2%).



(a) (b)

Fig. 4. (a). ROC curve and (b). Cumulative match characteristics (CMC) curve on binary classification on ATVS dataset using DWT + 'Modified ResNet' classification technique.

The test accuracy reported of the ‘modified ResNet’ was significant, irrespective of its deep network architecture. The training and validation process is reasonably faster due to the skip connection techniques at initial stages of training and the 32 convolution layers towards the end of training for fast training. Our ‘modified ResNet’ uses optimum CPU/GPU memory. The ‘modified ResNet’ efficiently handled standard black and white ATVS iris image and CASIA two-class, cropped, and custom resized face images datasets with precision.

The ROC curve in Fig. 4(a) plots the TPR versus the False Positive Rate (FPR) for our proposed ‘modified ResNet’ on different image datasets. In the best-case scenario, TPR should be as close as possible to 1.0, meaning that during training and validation none of the images were rejected by mistake.

The FPR should be as close to 0.0 as possible, meaning that all presentation attacks were rejected, and none were mistakenly accepted as real users. In all cases, the Area Under the Curve (AUC) remains in the range of 0.96 to 1.00, which is close to the ideal value of 1.0. However, the TPR is generally higher for the controlled grayscale ATVS image datasets compared to less controlled CASIA color image datasets and CASIA cropped datasets.

We achieved the 90.1% rank 3 identification and recognition accuracy with precision as shown in Fig. 4(b) in the Cumulative Match Characteristics (CMC) curve.

6 Conclusion

In this paper, we applied the ‘modified ResNet’ in combination with DWT to mitigate the presanction attacks on iris and face images. The modified ResNet architecture used here takes less computational time compared to other deep nets, without compromising the accuracy. Our future work will focus on testing our framework with other popular biometric datasets and observing stability and performance. Our future plan includes an extensive focus on the test accuracy and improving the overall performance of our proposed architecture.

Acknowledgment. This research is based upon work supported by the Science & Technology Center: Bio/Computational Evolution in Action Consortium (BEACON) and the Army Research Office (Contract No. W911NF-15-1-0524).

References

1. Fierrez, J., Ortega-Garcia, J., Torre-Toledano, D., Gonzalez-Rodriguez, J.: BioSec baseline corpus: a multimodal biometric database. Pattern Recognit. **40**(4), 1389–1392 (2007)
2. Galbally, J., Ortiz-Lopez, J., Fierrez, J., Ortega-Garcia, J.: Iris liveness detection based on quality related features. In Proceedings of the International Conference on Biometrics, New Delhi, India, ICB, pp. 271–276, March 2012
3. Yang, J., Lei, Z., Li, S.: Learn convolutional neural network for face anti-spoofing. arXiv:1408.5601v2 [cs.CV], August 2014

4. Menotti, D., et al.: Deep representations for iris, face, and fingerprint spoofing detection, arXiv:1410.1980v3 [cs.CV], Pre-print of article that will appear in IEEE Transactions on Information Forensics and Security (T.IFS), 29 January 2015
5. Manjani, I., Tariyal, S., Vatsa, M., Singh, R., Majumdar, A.: Detecting silicone mask-based presentation attack via deep dictionary learning. IEEE Trans. Inf. Forensics Secur. **12**(7), 1713–1723 (2017)
6. Rao, R.: Wavelet Transforms. Encyclopedia of Imaging Science and Technology. Wiley, January 2002. https://doi.org/10.1002/0471443395.img112
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 [cs.CV] 10 December 2015
8. Zhiwei, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: Proceedings of IAPR International Conference on Biometrics (ICB), Beijing, China, pp. 26–31 (2012)
9. Chinese Academy of Sciences (CASIA), Institute of Automation, Face antispoofing dataset. http://www.cbsr.ia.ac.cn/english/FASDB_Agreement/Agreement.pdf
10. Pentland, A., Choudhury, T.: Face recognition for smart environments. Computer **33**(2), 50–55 (2000)
11. Phillips, P.J., Martin, A., Wilson, C.L., Przybocki, M.: An introduction to evaluating biometric systems. Computer **33**(2), 56–63 (2000)
12. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)
13. Daugman, J.: How iris recognition works. IEEE Trans. Circuits Syst. Video Technol. **14**(1), 21–30 (2004)
14. Lee, T.S.: Image representation using 2D Gabor wavelets. IEEE Trans. Pattern Anal. Mach. Intell. **18**(10), 959–971 (1996)
15. Spencer, J., Lawrence, D., Roy, K., Chatterjee, P., Esterline, A., Kim, J.: Presentation attack detection using convolutional neural networks and local binary patterns. In: First International Conference on Pattern Recognition and Artificial Intelligence, Montreal, Canada, 14–17 May 2018, pp. 529–534 (2018)
16. Chatterjee, P., Roy, K.: Anti-spoofing approach using deep convolutional neural network. In: Recent Trends and Future Technology in Applied Intelligence, January 2018. https://doi.org/10.1007/978-3-319-92058-0_72
17. Mordvintsev, A., Abid, K.: Face Detection using Haar Cascades. [online] OpenCV Tutorial (2013). https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html. Accessed 2018