A Self Supervised Defending Mechanism Against Adversarial Iris Attacks based on Wavelet Transform

Meenakshi K[1]
Department of Computer Science and Engineering
School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

G. Maragatham[2]
Department of Information Technology
School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

***Abstract*—In biometric applications, deep neural networks have presented significant improvements. However, when presenting carefully designed input training data known as adversarial examples, their output is severely reduced. These types of attacks are termed as adversarial attacks, and any biometric security system is greatly affected by these attacks. In the proposed work, an effective defensive mechanism has been developed against adversarial attacks which are introduced in iris images. The proposed defensive mechanism is following the concept of wavelet domain processing and it investigates the mid and high frequency components of wavelet domain components. Based on this, the model reproduces the various denoised copies of input iris images. The proposed strategies are intended to denoise each sub-band of the wavelet domain and assess the sub-bands most likely to be affected by the adversary using the reconstruction error measured for each sub-band. We test the effectiveness of the proposed adversarial protection mechanism against various attack methods and analyzed the results with other state of the art defense approaches.**

Keywords—Iris classification; deep neural networks; adversarial attack; defense method; wavelet processing; biometrics

I. Introduction

In various classification applications such as biometric spam filtering, autonomous vehicle system, and speech recognition, etc. machine learning and deep learning-based classifiers have now achieved outstanding performance [1, 2]. Besides that, the classifiers are more prone to adversarial attacks that make the classification models behave more confidently in the wrong direction, i.e. the model misclassified the sample. Adversarial examples have been created by these attacks that are classified as data samples built to manipulate the Classifier model [3] The adversary has thus used these adversarial examples and these compromised examples to target the security system to give access to unauthorised users in order to modify the identity of the actual subject. So the adversarial examples are considered as security risks which are structured to affect the performance of the ML based classifier [4]. The adversarial attacks are classified as two types: 1. Black Box and 2. White box attacks. The attacker has a detailed understanding of the layout of the classification model in the case of a white box attack, such as parameters and algorithms used, etc. [5], whereas the adversary has no knowledge about the classification model in the black box method [6]. The techniques for coping with adversarial threats are called defensive methods. The defence mechanisms focus on making the classification model safer and more stable, and few methods seek to recognise the adversarial data, i.e. manipulated image [7]. To identify the person uniquely, various biometric characteristics such as fingerprint, face, iris, signature, voice, retina etc. are used. In [8], the important features of iris are captured which makes it more significant and secure biometric trait for the unique identification of an individual with a high degree of confidence. But the attacker introduces adversarial examples (manipulated iris images) to fool the recognition system and it is a big challenge to the security system. Protecting the iris recognition system from these types of attacks is important and it is a significant research direction to define the necessary countermeasures used to effectively detect adversarial attacks.

The Wavelet Decomposition technique is used in the proposed paper to classify the manipulated adversarial data. Kim et al. [9] have already shown that wavelet components of iris image with low and low-mid frequencies have high data to detect the subject, and these components are reliable and difficult to inject noise. To build the manipulated samples, the adversaries add the high frequency sections to the iris images. On the basis of this fact, we have proposed a defensive mechanism that effectively recognises adversarial attacks.

The following contributions are presented in this paper: a) An efficient defensive mechanism has been implemented which is applied before the iris recognition process. b) The proposed work analyzes the wavelet components, to identify the adversarial data and it is accurate and stable against adversarial attacks. C) The proposed methodology is compared with other state of the art defensive mechanisms in terms of accuracy.

The paper is organized in the following way. Related works are presented in Section II. Section III explains in depth the proposed approach. The experimental results are listed in Section IV. The conclusion and possible future developments are drawn in Section V.

II. Related Works

A. Adversarial Attacks

Recently, deep learning models have performed tremendously in a large range of applications like biometrics [10, 11], security [12, 13], autonomous vehicle control systems and Spam Filtering. However these models are more susceptible to manipulated input data which is called adversarial examples. Synthetic information is described as the small disturbances are added to the input image, often referred to as poisoning data. It has been shown that a minor change in input data causes a substantial decrease in model accuracy [14, 15]. To exploit the biometric protection framework, the intruder will use these adversarial examples, resulting in either an unauthorized user having access to the system or an authorized user being unable to access the system.

Szegedy et al. implemented the first adversarial attack, it is called as L-BFGS [16] and it is a costly method of computation. Goodfellow et al. have addressed the shortcomings of the previous system. Another method called FGSM, the Fast Gradient Sign Method, has been implemented, which introduces the degree of disturbance by considering the gradient sign [17]. Goswami et al. suggested an adversarial blackbox attack, which introduces the distortions in the face image and it leads the poor performance face recognition system [18]. The evolutionary algorithm was used by Dong et al. to build adversarial examples [19] and it follows the white box attack strategy. Lu et al. are suggesting FGSM-based attacks, which cause a disruption in all frames of a video. Milton et al have suggested a momentum-based FGSM attack and the CNN model is affected by that attack [20]. Generative Adversarial Networks (GAN) are used in [21] to construct distorted images with regard to samples of face images. In order to create adversarial instances, Rozsa et al. have enhanced the efficacy of the FGSM approach by considering the gradient value, whereas the previous method uses the gradient sign [22]. The DeepFool method has also been used to generate adversarial samples to classify the Lp disturbance that converts the input samples into adversarial data [23].

B. Defensive Mechanism

Two kinds of defensive techniques are used to handle the adversarial attacks, 1. Reactive defensive strategy 2.Proactive defensive strategy [24]. In the reactive defensive mechanism, after the deep learning models are designed, the designer tries to classify the adversarial examples. Whereas the designer aims to build the models more stable until the attacker implements the manipulated samples in the constructive defensive strategy. Few types of proactive defensive methodologies are developing robust classifier, adversarial training, and network distillation. Classifier Models are used as filters to remove the crafted data from the training data which act as preprocessing step. So that the robustness of the model is increased effectively [25].

i) Adversarial example recognition ii) network verification iii) input reconstruction are examples of reactive defensive mechanisms. The binary classifier was considered for the identification of the manipulated samples [26]. The adversarial examples were transformed to approximate original examples in the input reconstruction strategy. In order to recreate the adversarial samples into actual samples by eliminating the perturbations, a denoising auto encoder is used [27]. Network verification, which investigates the input data and tests whether the input violates the characteristics of the deep neural network, is the last technique [28]. In [29], to filter the adversarial instances, the authors used the appropriate dropouts in hidden layers. Agarwal et al. [30, 35, 36] have used the Principal component analysis (PCA) and the Support Vector Machine (SVM) to consider the presence of adversarial attacks.

III. Research Methodology

The Proposed method aims to identify the adversarial examples by removing the perturbations without changing the classifier model. Initially the classifier model is trained with the actual iris images i.e. unperturbed images. In the input examples, the adversarial Iris examples are generated by adding perturbations. To counter this, by using an encoder from a model trained to denoise the perturbations, we aim to eliminate the denoise in the Iris examples. We subsequently decompose the iris example image input into wavelet sub-bands by using wavelet transformation. This defensive mechanism utilizes the convolution layers that are trained to recreate the benign iris images by removing the adversarial noise and it analyzes the mid and high frequencies of wavelet components. In this approach Robust Normalization is used, which has a connection between the removal of outliers in activations and robustness. Dropouts are used to decrease the inter neuron dependencies. Therefore, the neural network is restricted from depending heavily on neuron weights, which could be model vulnerability. To enhance robustness, we suggest using average pooling layers that introduce less loss of information than max pooling layers.

A. Encoder – U-Net Architecture

The goal of this methodology is to extract the perturbations from the manipuated Images and the features of generated Iris examples are retained. For this, a deep convolutional neural network is used which follows an U-net architecture. The U-Net architecture has skip connections that have an effect on problems with gradient vanishing and can transfer image information from convolution layers to deconvolution layers that play a role in reconstructing noisy input. The U-net architecture could learn to denoise and get simple denoised outputs in a stable manner. The explanation why U-net-based denoising models are effective in denoising may be linked to the relations between the contracting path and the expansive path. The U-Net architecture has three sections, a contracting path, a bottleneck and an expanding path. In this architecture, the contracting path utilizes many convolutional operations followed by average-pooling operations. Then the input flows into the expanding layer with corresponding layers of convolution. The contracting and expanding path is linked by the bottom layer. The same is illustrated in Fig. 1.

The necessary preprocessing operations are carried out in the following way: adversarial input image are normalized and reshaped. These images are given as input to the U-net architecture. The encoder layer consists mainly a convolutional layer followed by strong normalization and a dropout layer followed again by a convolutional and robust normalisation layer. Then the corresponding output is applied on Average pooling layer. With the exception of the Convulational Transpose Layer, the decoder portion of the U-net architecture is identical to the encoder part. The representation of the image is fed from the U-net model's earlier layers. That is, from the encoder to the decoder layer. The output is then moved to a convolutional layer to recreate the image without the adversarial noise.

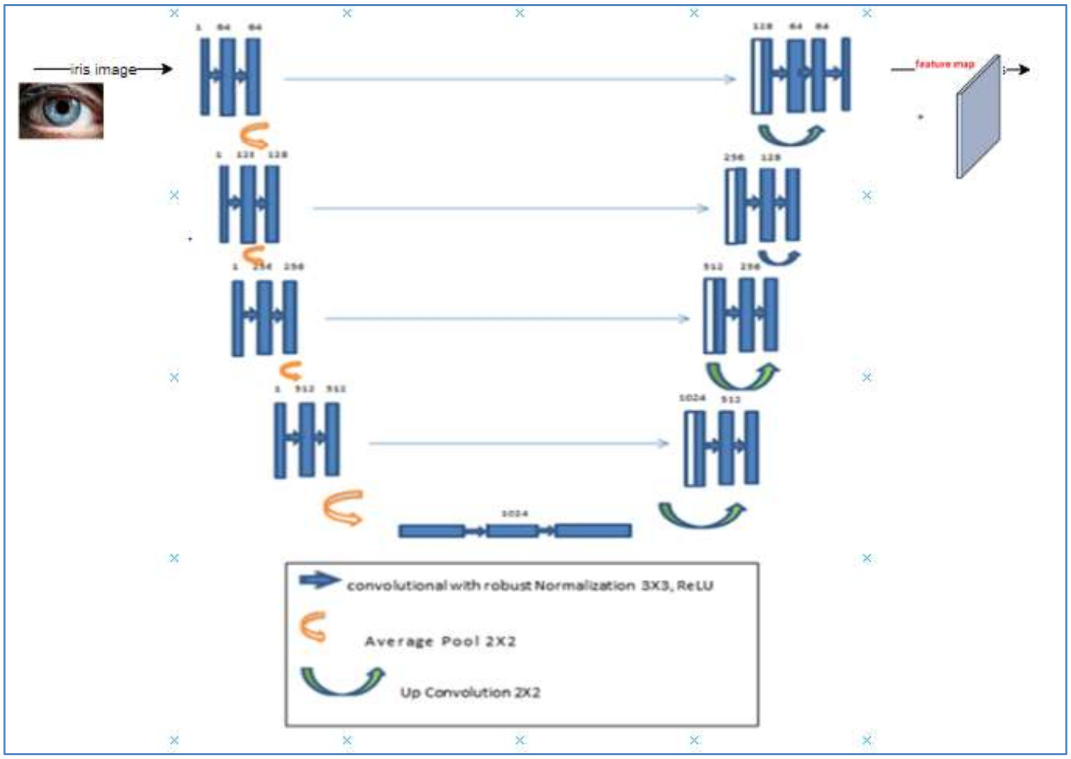


Fig. 1. Architecture of U-Net Denoiser.

The U-net denoising model uses Robust Normalization that outperforms BatchNorm on many datasets for adversarial accuracy while retaining other Normalization advantages. The model is trained to reduce reconstruction errors in order to eliminate the adversarial noise from the adversarial example, so it aims to transform the adversarial examples into their respective benign examples. From equation (1) the reconstruction error is calculated for every batch.

$$reconstruction_{err} = \left|\left|\tilde{x}_i^{(i)} - x_i^{(\iota)}\right|\right|_2 \qquad (1)$$

Where,

$\tilde{x}_i^{(i)}$- reconstructed input

$x_i^{(\iota)}$- actual input

Without the adversarial noise, the encoder learns the best characteristic representations necessary for reconstruction of the input image. Fig. 2 shows the single instance of the encoder layer.

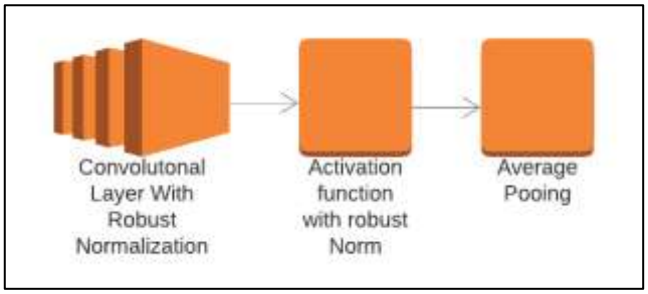

Convolutonal Layer With Robust Normalization
Activation function with robust Norm
Average Pooing

Fig. 2. Single Instance of the Encoder Layer.

B. Wavelet Decomposition

The input image is decomposed in to identical sub bands by using wavelet decomposition technique. This uniform decomposition offers more flexibility for our proposed system to select mid and high-frequency sub-bands. Wavelet image transformation is a very efficient and stable technique and has many benefits. For example, in a digital image, the wavelet analysis preserves the high-frequency edge information and prevents the image from being fuzzy. The method of wavelet analysis is a time-frequency analysis method that selects the appropriate frequency band adaptively based on signal characteristics. In denoising, this property is incredibly helpful as it reduces the loss of data during denoising. In order to achieve optimal reconstruction of the original signal, the wavelet transformation process relies on the best mapping of signals from the actual space to the function space of the wavelet. The proposed solution uses the multi-level discrete wavelet decomposition. This wavelet transformation decomposes the signal into a wavelet range that is mutually orthogonal, and this particular decomposition of the wavelet more finely decomposes sub-bands of low passes. Fig. 3 illustrates the wavelet decomposition stages. The wavelet transform can be expressed by using equation (2).

$$R(a,b) = \int_{-\infty}^{\infty} r(x)\, \phi_{(a,b)}^{*}(x)dx \qquad (2)$$

Where,

* - conjugate symbol

ϕ – Any function, chosen arbitrarily, should follow certain rules

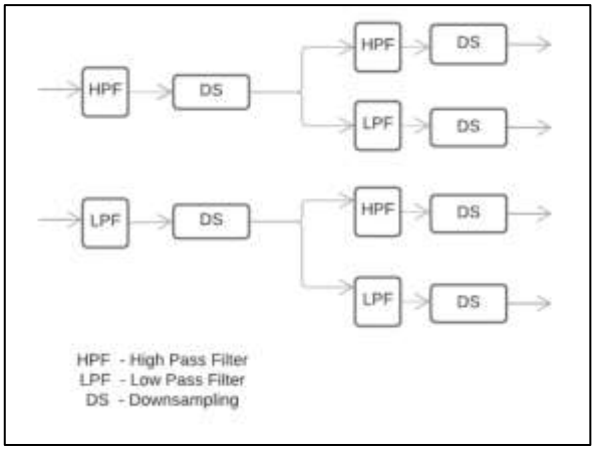


Fig. 3. Wavelet Transform Decomposition Stages.

The encoder's output is fed to the Wavelet layer and the wavelet transform is applied to the image, which splits it into four sub-bands hierarchically. In order to implement the wavelet transforms, we perform a series of operations on each axis to construct partitions. After investigating the directions of low and high pass filters, Multi-level Discrete Wavelet Transformations are determined. For downsampling of the images, even index columns are chosen. The resulting image is then transmitted again to the low pass and high pass filters where the convolved image is generated as an output. The inputs are now down-sampled by rows. This process results in four sub-bands. In these four sub-bands, there are diagonal, horizontal and vertical descriptions of the images.

$$y_{high}[n] = \sum_{i=-\infty}^{\infty} x\,[i]h[2n-i] \qquad (3)$$

Where,

x - input signal

h - high pass filter

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x\,[i]l[2n-i] \qquad (4)$$

Where,

x - input signal

l - low pass filter

Equation (3) and (4) show the functioning of Low pass and High pass filters with down sampling. All the sub bands are functioning efficiently. The convolution layers which are present in encoder part of the denoising models with Robust normalization are trained to filter the targeted adversarial attack. One more sub band is trained to remove random noise by decreasing the reconstruction error. The deep U-Net architecture is subsequently concatenated by all these sub-bands.

The U-Net Model's output is applied to the convolutional layers with robust normalization, then it has been passed to the global average pooling layer. The purpose of Global average pooling layer is to reduce the number of model parameters drastically and it prevents the overfitting, it results the increase in performance. The average pooling layer of Convolutional Neural Network model doesn't preserve the low level feature sets, but it restores the high level feature map. The results from the previous layers are given into dense layers with dropouts undergoing Robust Normalization. For classification, we have used the softmax activation function. Sparse categorical entropy is the loss metric and Adam is the optimizer.

We integrate regularization in the form of L1 regularizer to avoid overfitting the model. The explanation for preferring L1 rather than L2 is that L1 tends to minimise the coefficients to zero, while L2 reduces them equally. This enables L1 more acceptable for the selection of features, since it helps us to drop any variable with coefficients moving to zero. We observe an increase in the validation accuracy of our classifier by adding L1 regularization.

IV. Results and Discussions

In this part, we discuss the dataset used and how the adversarial perturbations and noise were applied to produce adversarial examples. We conclude the section by describing the findings of the proposed systems and analyzing their efficiency with the previous state-of-the-art mechanisms.

A. Dataset

In the proposed method, by integrating different forms of noise in the clean examples, we produce the adversarial Iris examples. The adversarial dataset is generated by the algorithms FGSM, iGSM and deepfool which are most popular algorithms to produce adversarial examples. Table I gives the descriptions of datasets used in the proposed work. From one model to another, the adversarial noises have remarkable transferability. The perturbations are added in the CASIA Iris V4 Dataset then the Deep CNN U-Net model is trained to remove the noises. The key features required to reconstruct the denoised version of the image from adversarial image are preserved by minimizing the reconstruction error. This can be achieved by using the encoder – decoder layers of the framework.

The wavelet domain decomposition layer belongs to a DCNN denoising model is trained to remove the adversarial noise. The encoder part of this denoiser works as on the wavelet sub-bands. As a whole, using the wavelet transformation function, a single adversarial image is decomposed into four wavelet sub-bands here. Of these four sub-bands, three are trained to remove adversarial noise, while the fourth is trained to eliminate random noise. All these four sub-bands are then concatenated at various layers and eventually transferred into the global average pooling layer. On our final evaluation we observed a rise in the validation accuracy of our classifier. Table II indicates the accuracy of the model before the attack and after the attack. The Deep CNN model is applied on CASIA Iris V4 dataset for classification and the accuracy before FGSM attack is 98.01% whereas after the attack it reduces into 90.24%. The same table indicates the accuracy before and after the attack in case of iGSM, Depfool. The comparison of classification accuracy for the proposed model with existing state of art model is tabulated in Table III.

It is observed that the proposed method is outperformed and the accuracy is good competed to other state of art models. The graphical representation of the comparison is shown in Fig. 4(a) and 4(b).

TABLE I. DESCRIPTION ABOUT THE DATASET

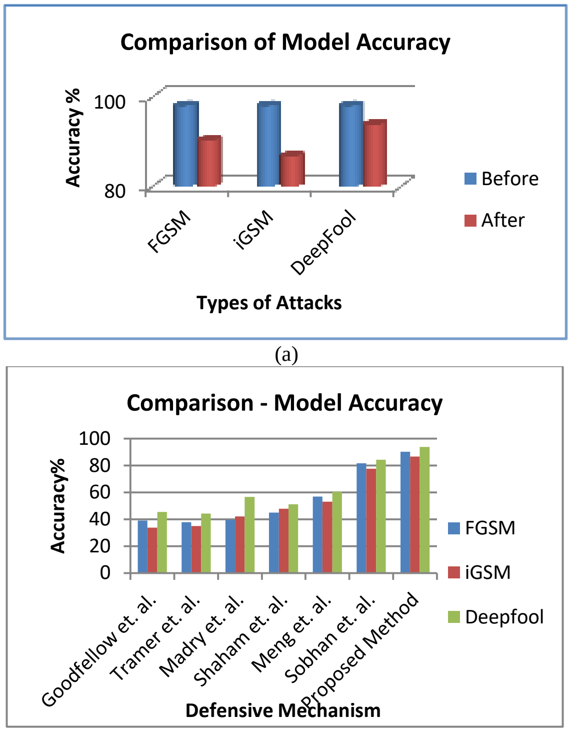
Dataset	**Images**	**Classes**
Casia-IrisV4	20000	1000
Casia-IrisV1	1080	108
IITD Database	2240	224
FGSM Database	50000	1000
iGSM Database	20000	1000
Deepfool Database	21080	1000
Noise Dataset	10000	1000

TABLE II. MODEL ACCURACY BEFORE AND AFTER ADVERSARIAL ATTACKS

Accuracy Vs Attack	**FGSM**	**iGSM**	**Deepfool**
Before	98.01	98.01	98.01
After	90.24	86.70	93.83

TABLE III. COMPRARISON OF PROPOSED MODEL WITH STATE-OF-THE-ART MODELS IN TERMS OF ACCURACY

Defensive MechanismVs Attacks	**FGSM**	**iGSM**	**Deepfool**
	Accuracy		
Goodfellow et. al. [17]	38.98	33.78	45.47
Tramer et. al. [31]	37.87	34.97	44.41
Madry et. al. [32]	39.51	42.18	56.78
Shaham et. al. [33]	45.15	47.89	51.24
Meng et. al. [27]	57.08	53.26	60.54
Sobhan et. al. [34]	81.65	77.59	84.36
Proposed Method	92.24	86	94.8



(a)

(b)

Fig. 4. (a): Comparison of Model Accuracy before and after the Attack.Three Types of Attacks are Compared- FGSM, iGSM and Deepfool, (b): Model Accuracy for Proposed Methods on Adversarial Attacks.

V. CONCLUSION

Defending adversarial attacks is a crucial move towards reliable implementation of biometrics authentication solutions driven by deep learning. In this proposed work, a novel defending framework has been developed to defend the adversarial attack targeted on Deep Convolutional Neural Networks. Iris recognition system is considered as one of the popular biometric systems which uses the Deep Neural Network for recognition. The proposed strategy is able to detect and reconstruct the adversarial examples consistently. Using an encoder architecture and wavelet decomposition, a framework has been built that takes adversarial input examples and analyzes the wavelet sub bands. Based on the reconstruction error, the framework identifies the attack.

From the Experimental results, it was observed that the proposed strategy was very effective. The proposed framework is compared with other state of art defending strategies and it achieves 92% accuracy during classification of iris images. Further this work can be extended to consider other attack strategies. In this work the wavelet decomposition is applied to detect the adversarial image. In future other equivalent transformation functions like curvelet transform can be applied and study further for other biometric traits.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[2] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, AbdelRahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, and Brian Kingsbury. Deep Neural Networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6):82–97, 2012.

[3] N. Carlini and D. Wagner. Adversarial examples are not easily detected: Bypassing ten detection methods. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, pages 3–14. ACM, 2017.

[4] J. Bruna, C. Szegedy, I. Sutskever, I. Goodfellow, W. Zaremba, R. Fergus, and D. Erhan. Intriguing properties of neural networks. International Conference on Learning Representations, 2014.

[5] Nicholas Carlini and David Wagner. Towards evaluating the robustness of Neural Networks. arXiv preprint arXiv:1608.04644, 2016.

[6] Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In Proceedings of the ACM Asia Conference on Computer and Communications Security, pages 506–519, 2017.

[7] K. Meenakshi and G. Maragatham, "A review on security attacks and protective strategies of machine learning", International Conference on Emerging Current Trends in Computing and Expert Technology, pp. 1076-1087, 2019.

[8] A. K. Jain, K. Nandakumar and A. Ross, "50 years of biometric research: Accomplishments challenges and opportunities", Pattern Recognition Letters, vol. 79, pp. 80-105, 2016.

[9] J. Kim, S. Cho, J. Choi, and R. J. Marks. Iris recognition using wavelet features. Journal of VLSI signal processing systems for signal, image and video technology, 38(2):147– 156, 2004.

[10] S. Soleymani, A. Dabouei, S. M. Iranmanesh, H. Kazemi, J. Dawson, and N. M. Nasrabadi. Prosodic-enhanced siamese convolutional neural networks for cross-device text-independent speaker verification. arXiv preprint arXiv:1808.01026, 2018.

[11] S. Soleymani, A. Torfi, J. Dawson, and N. M. Nasrabadi. Generalized bilinear deep convolutional neural networks for multimodal biometric identification. In 25th IEEE International Conference on Image Processing, pages 763–767, 2018.

[12] F. Taherkhani, N. M. Nasrabadi, and J. Dawson. A deep face identification network enhanced by facial attributes prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 553–560, 2018.

[13] V. Talreja, M. C. Valenti, and N. M. Nasrabadi. Multibiometric secure system based on deep learning. In 2017 IEEE Global conference on signal and information processing (globalSIP), pages 298–302, 2017.

[14] A. Kurakin, I. Goodfellow, and S. Bengio. Adversarial examples in the physical world. International Conference on Learning Representations-Workshop, 2017.

[15] Akshay Agarwal, Akarsha Sehwag, Richa Singh, and Mayank Vatsa. Deceiving face presentation attack detection via image transforms. In IEEE International Conference on Multimedia Big Data, pages 373–382, 2019.

[16] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. arXiv preprint, 2013.

[17] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572, 2014.

[18] Gaurav Goswami, Akshay Agarwal, Nalini Ratha, Richa Singh, and Mayank Vatsa. Detecting and mitigating adversarial perturbations for robust face recognition. International Journal of Computer Vision, 127(6-7):719–742, 2019.

[19] Yinpeng Dong, Hang Su, Baoyuan Wu, Zhifeng Li, Wei Liu, Tong Zhang, and Jun Zhu. Efficient decision-based blackbox adversarial attacks on face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7714–7722, 2019.

[20] Md Ashraful Alam Milton. Evaluation of momentum diverse input iterative fast gradient sign method (M-DI2- FGSM) based attack method on MCS 2018 adversarial attacks on black box face recognition system. arXiv preprint arXiv:1806.08970, 2018.

[21] Debayan Deb, Jianbang Zhang, and Anil K Jain. Advfaces: Adversarial face synthesis. arXiv preprint arXiv:1908.05008, 2019.

[22] A. Rozsa, E. M. Rudd, and T. E. Boult. Adversarial diversity and hard positive generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 25–32, 2016.

[23] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard. Deepfool: a simple and accurate method to fool deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2574–2582, 2016.

[24] X. Yuan, P. He, Q. Zhu, and X. Li. Adversarial examples: Attacks and defenses for deep learning. IEEE transactions on neural networks and learning systems, 2019.

[25] J. Bradshaw, A. G. d. G. Matthews, and Z. Ghahramani. Adversarial examples, uncertainty, and transfer testing robustness in gaussian process hybrid deep networks. arXiv preprint arXiv:1707.02476, 2017.

[26] Z. Gong, W. Wang, and W.-S. Ku. Adversarial and clean data are not twins. arXiv preprint arXiv:1704.04960, 2017.

[27] D. Meng and H. Chen. Magnet: a two-pronged defense against adversarial examples. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 135–147. ACM, 2017.

[28] G. Katz, C. Barrett, D. L. Dill, K. Julian, and M. J. Kochenderfer. Reluplex: An efficient SMT solver for verifying deep neural networks. In International Conference on Computer Aided Verification, pages 97–117, 2017.

[29] Kathrin Grosse, Praveen Manoharan, Nicolas Papernot, Michael Backes, and Patrick McDaniel. On the (statistical) detection of adversarial examples. arXiv preprint arXiv:1702.06280, 2017.

[30] Akshay Agarwal, Richa Singh, Mayank Vatsa, and Nalini Ratha. Are image-agnostic universal adversarial perturbations for face recognition difficult to detect? In IEEE International Conference on Biometrics Theory, Applications and Systems, pages 1–7, 2018.

[31] F. Tramer, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel. Ensemble adversarial training: Attacks and defenses. arXiv preprint arXiv:1705.07204,2017.

[32] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083, 2017.

[33] U. Shaham, J. Garritano, Y. Yamada, E. Weinberger,A. Cloninger, X. Cheng, K. Stanton, and Y. Kluger. Defending against adversarial images using basis functions transformations. arXiv preprint arXiv:1803.10840, 2018.

[34] Sobhan Soleymani, Ali Dabouei, Jeremy Dawson, and Nasser M. Nasrabadi, Defending Against Adversarial Iris Examples Using Wavelet Decomposition. arXiv:1908.03176,2019.

[35] Saranya, G., & Pravin, A. A comprehensive study on disease risk predictions in machine learning. International Journal of Electrical and Computer Engineering (IJECE), 10(4), 4217, 2020.

[36] M. K and G. Maragatham, "A Comprehensive survey on Iris Presentation attacks and Detection based on Generative Adversarial Network," 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2020, pp. 1-9, doi: 10.1109/ICPECTS49113.2020.9336966.