Crafting A Panoptic Face Presentation Attack Detector

Suril Mehta, Anannya Uberoi, Akshay Agarwal, Mayank Vatsa, and Richa Singh
{suril15104, anannya15014, akshaya, mayank, rsingh}@iiitd.ac.in
IIIT-Delhi, India

Abstract

With the advancements in technology and growing popularity of facial photo editing in the social media landscape, tools such as face swapping and face morphing have become increasingly accessible to the general public. It opens up the possibilities for different kinds of face presentation attacks, which can be taken advantage of by impostors to gain unauthorized access of a biometric system. Moreover, the wide availability of 3D printers has caused a shift from print attacks to 3D mask attacks. With increasing types of attacks, it is necessary to come up with a generic and ubiquitous algorithm with a panoptic view of these attacks, and can detect a spoofed image irrespective of the method used. The key contribution of this paper is designing a deep learning based panoptic algorithm for detection of both digital and physical presentation attacks using Cross Asymmetric Loss Function (CALF). The performance is evaluated for digital and physical attacks in three scenarios: ubiquitous environment, individual databases, and cross-attack/cross-database. Experimental results showcase the superior performance of the proposed presentation attack detection algorithm.

1. Introduction

With growing interest in the field of contactless identification, face recognition is one of most popular biometric modality. A market research conducted by Counterpoint Research estimated that over a billion smartphones will have a facial unlocking feature by 2020. The increasing usage of face recognition systems necessitates that their security should also be ensured. In November 2017, Vietnamese security firm Bkav exposed a critical weakness of the iPhone X's Face ID by using a simple $150 3D printed silicone mask within a month of its market release. This demonstrates that face recognition based security measures are susceptible to presentation attacks and require further research.

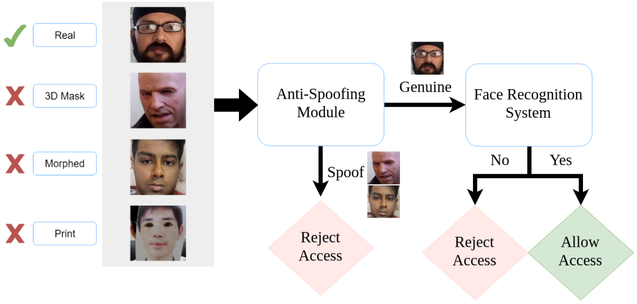

Figure 1: A brief pipeline for a biometrics system with anti-spoof module.

As shown in Figure 1, presentation attacks can be physical or digital. These attacks can be either used to elude a system, wherein the system cannot recognize the person, or duplicate an identity, wherein the system matches the probe image (presentation attacked) with a target identity. To protect face recognition systems against such attacks, several anti-spoofing or presentation attack detection algorithms have been developed [12, 21]. Majority of the research efforts in the literature are focused on either same-database, or cross-database experiments between similar attack categories (such as mask attacks, 3D print attacks, or replay attacks – all of which belong to the broad category of physical attacks). They are able to detect either physical attacks or digital attacks with high confidence. However, in a real life scenario, one may encounter either of these attacks through any image input to the biometric system. Therefore, it is desirable that the successful presentation attack detection algorithms should be agnostic to the kind of attacks.

This research is centered on developing a panoptic solution for detecting multiple digital and physical presentation attacks. The key contributions of this research are: **(i)** a novel learning approach for digital and physical presentation attack detection using a new loss function to train the deep convolutional neural network (CNN), **(ii)** experiments corresponding to both inter and intra spoof attack scenarios. Intra experiments are defined as the experiments where the training and testing samples belong to the same attack category (i.e., physical or digital), whereas in inter exper-

www.counterpointresearch.com/one-billion-smartphones-feature-face-recognition-2020/

iments, training (such as silicone mask attack) and testing (such as digital swapping) samples belong to different attack categories, and **(iii)** detailed experimental analysis on amalgamation of digital and physical attacks, which has not been performed before, to show the generalizability of the proposed detection algorithm.

2. Related Work

Existing spoof detection approaches can be broadly classified into two groups: static and dynamic [12, 21]. Static anti-spoofing techniques rely on a single image source to classify whether it is genuine or spoofed. Majority of the existing anti-spoof methods involve extraction of discriminating features to analyze the face texture, such as Haralick texture features, local binary pattern (LBP), partial least square (PLS), and difference of Gaussian (DoG) [1, 2, 3, 20, 30]. Features such as LBP, the co-occurrence of LBP, binarized statistical image features (BSIF), and the scale-invariant descriptor (SID), can be extracted along different color spaces to preserve the chroma component of images [7]. Patel et al. [19] studied features reliant on surface reflection by the spoof medium such as Moiré patterns, color distortions, and shape deformations along different intensity and use them as measure for spoof detection. Further, there have been shifts towards convolutional neural network (CNN) based approaches for classification of real and spoofed images. Liu et al. [4] proposed fusion of separately trained patch-based/depth-based and locally specialized CNN architectures for photo and video-based attack detection.

Dynamic anti-spoofing techniques mostly target blinking [13, 18], motion magnification [6], or liveness detection [8, 28], given a sequence of frames. CNNs have also been employed for dynamic anti-spoofing. Feng et al. [11] proposed a neural network which fuses both image quality and motion cues for liveness detection. Due to large amounts of available data and pre-trained CNN models, these algorithms can exploit both spatial and temporal information in videos.

The aforementioned algorithms are mainly developed on physical presentation attack databases. In the digital attack landscape, Agarwal et al. [2] proposed a variation of the illumination invariant local binary pattern (LBP) feature descriptor followed by SVM for classification of real and digital spoofed images. Robertson et al. [22] performed an elaborate study, which accumulates responses of three experiments: human matching of two faces, human matching of two faces along with an option of fraudulent (morphed) image, and testing on an automated face recognition device. It suggests that a pre-detection tool is indispensable to the creation of robust biometrics systems. The details of the existing face PAD algorithms related to physical and digital attacks can be found in the survey papers [21, 23].

3. Proposed Algorithm

The decision of a face recognition system can be manipulated either at the sensor level i.e. through a physical presentation attack, or at the data level where face images can be manipulation using morphing/retouching. Synthetic face images could also be produced using machine learning algorithms to populate the system with fake data. As a result, it is difficult to track down a singular cause of presentation attacks. Handcrafted features often are not sufficient to capture the vast variations that can be seen in different presentation attacks [10]. Similarly, high quality facial texture of a silicone mask renders attack detection a complex task to accomplish using only a predefined set of features. Hence, deep learning based methods seem promising in such a complex scenario.

3.1. Loss Function

In this paper we propose a new loss function termed as Cross Asymmetric Loss Function (CALF or CA loss) through the combination of two loss functions – cross entropy (CE) and focal loss (FL) [16]. The motivation behind the proposed loss function is that images belonging to the spoof class form a loosely held cluster due to a wide variety of attacks. They cannot be mapped to a single, tight cluster. On the other hand, it is possible to map all the real samples to a single, compact cluster. Therefore, we focus on building a classifier for correctly detecting real images. The techniques for facial spoofing would continue to evolve, so we leverage the idea that real images will always belong to a particular cluster which indeed should be universal. Therefore, in the proposed loss function, we penalize the incorrect classification of real images.

CA loss can be written as $Loss_{ca} = Loss_{ce} + Loss_f$, where, $Loss_{ce}$ and $Loss_f$ represent the cross entropy and focal loss respectively. $Loss_f$ is represented as:

$$Loss_f = -(1 - p_r)^\gamma log(p_r) \tag{1}$$

$Loss_{ce}$, for binary classification, is represented as:

$$Loss_{ce} = -ylog(p_r) - (1 - y)log(1 - p_r) \tag{2}$$

where y is the class label for real images. For $\gamma = 2$,

$$\begin{aligned} Loss_{ca} = -[y + (1 - p_r)^2]log(p_r) - (1 - y) \\ log(1 - p_r) \end{aligned} \tag{3}$$

where p_r is the probability of an image belonging to the real class. Training the AlexNet architecture with the proposed loss function improves the efficacy of AlexNet and facilitates a faster training phase.

3.2. Feature Extraction and Classification

In the proposed algorithm, features from a fine-tuned CNN model, AlexNet [14], are extracted. AlexNet is a shallow network and contains five convolution layers followed

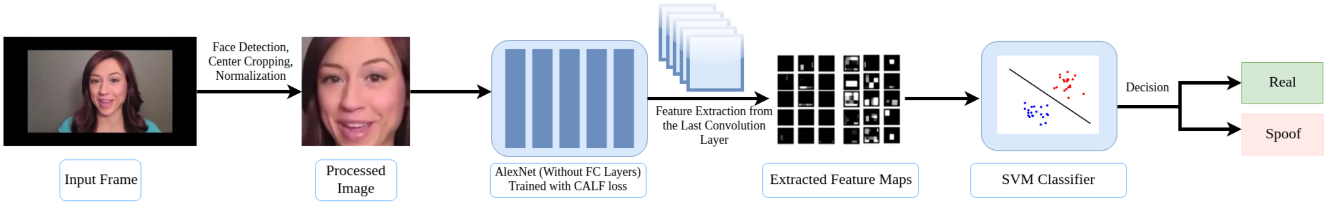
Face Detection, Center Cropping, Normalization
Feature Extraction from the Last Convolution Layer
Decision
Real
Spoof
Input Frame
Processed Image
AlexNet (Without FC Layers) Trained with CALF loss
Extracted Feature Maps
SVM Classifier

Figure 2: The overall pipeline of the proposed algorithm with Cross Asymmetric Loss Function.

by three fully connected (FC) layers. ReLU is applied after every convolution and fully connected layer. Dropout is used after the first and second FC layers. Feature maps are extracted from each of the convolution blocks for initial analysis. In the proposed architecture, the FC layer is replaced by a two-class SVM classifier [9] to classify the input image as real or spoof. The final classification block from each of the models is removed. For AlexNet, layers following the last max pooling layer are removed.

Support vector machine (SVM) classifier is used to perform the classification (*real* vs *spoof*). The parameters of the SVM are optimized using grid-search (with linear kernel, $c = 0.1$, l_2 penalty and hinge loss observed as the best parameters).

3.3. Implementation Details

Images are cropped using the Viola-Jones face detection algorithm [29] and resized to 224×224 pixels. The images are then normalized to zero mean and unit variance. To train the proposed presentation attack detector, we use the AlexNet architecture with cyclic learning rates [27] and stochastic gradient descent (SGD) optimization with warm restarts and differential learning rates across the layers. AlexNet returns a feature map of dimension $256 \times 6 \times 6$. Learning rate (lr) is a hyper-parameter that controls how speedily the weights of the network are adjusted. To tackle the problem of uncertainty about the optimal learning rate, we bring in the concept of cyclic learning – changing the learning rate after each epoch via a cyclic function, thereby preventing getting stuck in local minima or saddle points. In this paper, we employ the cosine annealing cyclic function [26].

In the experiments, we have observed that it is better to have different learning rates across layers; the motivation being that updates to the learning rate are greater in the classification layer than in the feature extraction layers. In other words, the first few layers would typically contain granular details of the data, such as edges. We would not want to alter this information much, rather retain this information. As such, there is no need to change their weights by a substantial amount, as opposed to the final dense layers which perform classification in deep CNN models. To achieve this, differential fine-tuning is implemented to perform this task.

4. Databases and Protocol

In this section, we provide the details of the databases used in this research. We cover both kind of attacks: (i) physical (silicone mask, photo, and video) and (ii) digital (swapped). The physical attack databases used in this research are: (i) CASIA-FASD and (ii) silicone mask attack database (SMAD). Similarly, SWAPPED digital attack database is used to perform the experiments. The details of each databases are given below:

1. **CASIA-FASD - Physical Attack:** CASIA-FASD, proposed by Zhang et al. [30] is a standard physical attack database consisting of 600 video samples from the warped photo, cut photo, and video attacks in three qualities: low, normal, and high. According to the overall test protocol defined in the paper, all kinds of spoofed samples are used for evaluation.

2. **SMAD - Physical Attack:** SMAD, collected by Manjani et al. [17], is a first-of-its-kind silicone mask attack database containing 130 real and mask attack videos obtained from the web. The experiments are performed according to the frame and video-based protocol presented in the paper.

3. **SWAPPED - Digital Attack:** SWAPPED database proposed by Agarwal et al. [2] is a curation of more than 600 videos created through face swap or face switch feature of Snapchat, and more than 120 real videos. In our experiments, a total of $20,336$ images ($9,804$ real and $10,532$ spoofed) were extracted from the videos for training. Class imbalance was taken care of by using an approximately equal number of real and spoofed images in the training phase.

5. Results and Evaluation Study

This section summarizes the experiments performed and the results obtained to demonstrate the efficacy of the proposed algorithm. We first report the results on the combined attack setup. To compare our performance with existing algorithms, we also evaluate its performance on attack-specific experimental setups i.e. independently on SMAD, SWAPPED, and CASIA-FASD databases in both intra-attack and inter (cross) attack scenarios. The performance of the proposed presentation attack detection model is reported using the Equal Error Rate (EER).

Combined Attack Detection: In this experimental setup, training sets from each database is used to form the train set and combination of test sets from three databases are used for testing. The proposed model is trained on this train set (which is an amalgamation of all attack databases) and tested on the combined test set. The proposed algorithm yields an EER of 5.63% on the combined dataset. This low EER value suggest that the proposed approach can handle multiple attacks together. We have observed that the proposed loss function indeed learns the real class representation such that it can differentiate real samples with different kinds of spoofed samples.

Attack-Specific Detection: We next demonstrate the comparison of the proposed algorithm with existing state-of-the-art algorithms. Video-based detection refers to the classification of an entire video, whereas frame-based detection refers to the classification of each single entity (frame) of the video as real or spoof. Figure 3 shows the ROC curves on CASIA-FASD, SMAD, and SWAPPED, respectively. The results corresponding to *intra* database experiments are reported in Tables 1 and 2. In the literature, it is found that 2D print attacks are readily detectable and have been solved to a large extent, while silicone mask attacks exhibit complex data distributions that cannot be captured by handcrafted features. Similarly, digital presentation attacks are not well explored in the literature. From the results, it can be observed that the proposed algorithm provides a low EER on the challenging SWAPPED, SMAD, and CASIA-FASD datasets which are representatives of digital and physical attacks.

The proposed CNN features yield EER values of 3.68% and 1.44% on SWAPPED for frame and video-based attack detection, respectively. The algorithm yields average EER values of 6.21% (Table 1) and 6.16% (Table 2) on SMAD for frame and video-based attack detection, respectively. On the other hand, existing state-of-the-art algorithms [17, 24] yield EER of 14.9% and 12.3%, respectively using the same experimental protocol. This is an improvement of at least 6.1% in terms of EER from the best-performing algorithm [17]. It also yields 0.04% and 0.00% EER on CASIA-FASD for frame and video-based attacks, respectively. If we compare with existing state-of-the-art algorithms [1] [7] on CASIA-FASD, the best reported results are 2.1% EER for frame based detection and 1.1% for video based detection.

Cross-Attack Detection: Cross attack experiments are defined as those where one particular kind of attack images (such as digital) are used in training the detector and unseen attack images (such as physical) are used at the time of testing the network. For cross attack network training, the real classes of all databases are combined to form a single real set. For attack set, one attack at a time is used for training and other attacks from different databases are used for testing. Cross database experiments are reported in Table 4. When SWAPPED digital attack database is used for training the classifier, the proposed algorithm yields EER values of 24.8% and 35.0% on SMAD and CASIA-FASD, respectively. Similarly, when CASIA-FASD physical attack database is used to train the detector, the proposed algorithm yields EER of 31.6% and 33.5% on digital and silicone mask attack databases, respectively.

Table 1: Face PAD performance (EER %) of the proposed algorithm for frame-based ‘attack-specific’ experiments.

Algorithm	**SWAPPED**	**SMAD**	**CASIA-FASD**
VGG-Face	11.69	9.45 ± 1.80	12.77
AlexNet	6.43	10.14 ± 2.25	2.29
AlexNet+CE	6.43	6.36 ± 0.89	0.38
AlexNet+FL	9.39	6.98 ± 1.95	1.41
Proposed	**3.68**	**6.21 ± 2.52**	**0.04**

Table 2: Face PAD performance (EER %) of the proposed algorithm for ‘attack-specific’ video-based experiments.

Algorithm	**SWAPPED**	**SMAD**	**CASIA-FASD**
VGG-Face	10.14	2.56 ± 0.81	13.02
AlexNet	6.06	3.32 ± 0.72	2.53
Proposed	**1.44**	**6.16 ±5.15**	**0.00**

Table 3: Comparison (EER %) of the proposed algorithm with recent algorithms on video-based physical presentation attack databases.

Algorithm	**CASIA-FASD**	**SMAD**
Haralick Texture [1]	**1.1**	–
Deep Belief Network [5]	–	16.9
Videolets [25]	3.1	–
Deep Dictionary [17]	1.3	**12.3**
Depth and Patch CNN [4]	2.7	–
3D CNN [15]	1.4	–
Deep Dynamic Texture [24]	–	14.9
Proposed	**0.0**	**6.2**

Table 4: Cross database and cross attack presentation attack detection performance of the proposed algorithm.

Cross Type	**Training On**	**Testing on**	**EER%**
Attack	SWAPPED	SMAD	24.8
		CASIA-FASD	35.0
	SMAD	SWAPPED	54.1
	CASIA-FASD		31.6
Database	SMAD	CASIA-FASD	47.5
	CASIA-FASD	SMAD	33.5
Combined	ALL	ALL	5.6

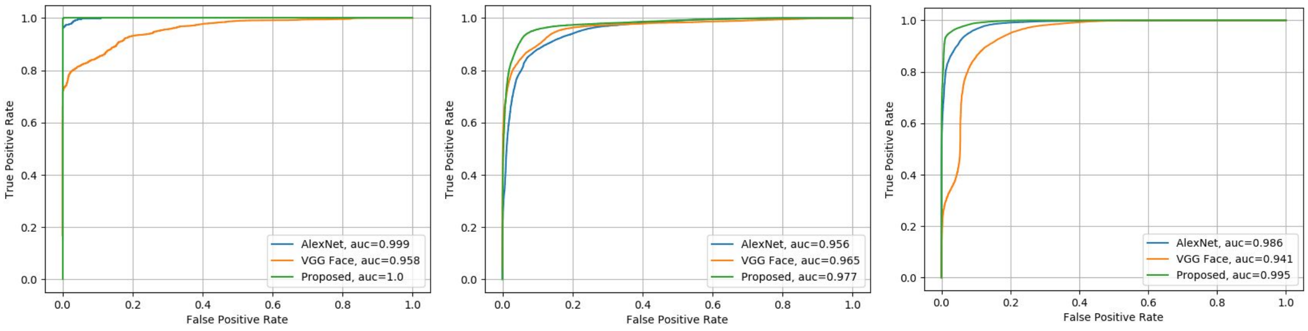

Figure 3: ROC curves for frame-based PAD on CASIA-FASD, SMAD, and SWAPPED (left to right).

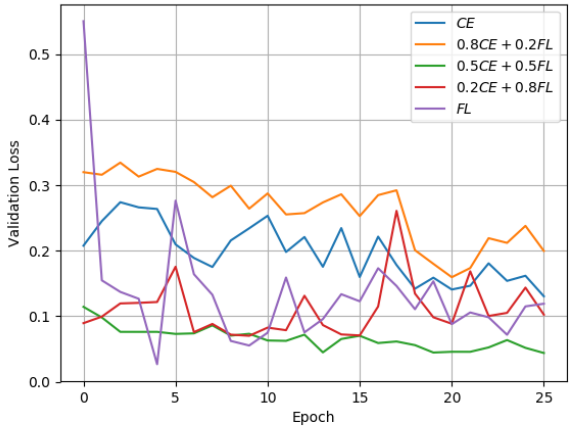

Figure 4: Loss values on the validation set with different loss functions. In this research, we have given equal weights (green curve) to both CE and FL loss, i.e. 0.5 each.

5.1. Component-Wise Analysis

In this subsection, we discuss the components of the proposed algorithm and a performance-centered analysis of other alternatives to these components: weights of the Cross Asymmetric Loss Function, choice of CNN architecture for feature extraction, and the choice of classifier.

- The proposed loss function can be generalized as $Loss_{ca} = (1 - \alpha)Loss_{ce} + \alpha Loss_f$. Therefore, different α values can be selected for CE and FL losses. Figure 4 shows experimental results corresponding to different weighted combinations of the loss functions. It suggests that the proposed loss (CALF) with equally weighted $Loss_{ce}$ and $Loss_f$ converges faster in comparison to other combinations. AlexNet with proposed loss function yields at least 2.75% and 0.34% lower EER (Table 1) than AlexNet+*cross entropy* and AlexNet+*focal* loss on SWAPPED and CASIA-FASD, respectively.
- Further, from Tables 1 and 2, we can observe that VGG-Face (a deeper architecture) generally yields higher EER than AlexNet (shallow network). The observation could be attributed to the fact that attacks add anomalies at a micro level which can be viewed as noise to a CNN based architectures. Going deeper into the pipeline brings unnecessary subtle features into account which are not relevant to the problem of learning spoofed images.
- The comparative results shown in Table 3 illustrate the feasibility of the proposed algorithm based on CNN features as opposed to handcrafted features for generalization over the presentation attacks. The proposed algorithm outperforms existing algorithms on the most challenging silicone mask based physical presentation attack. Moreover, the proposed algorithm achieves state-of-the-art performances in comparison to existing deep CNN based algorithms and handcrafted features based algorithms on standard physical attack database i.e., CASIA-FASD.
- The final experiment is performed with changing the classifier from SVM to neural network (NN). On the SWAPPED database, the proposed architecture with NN classifier yields at least 2.56% higher EER in comparison to SVM based approach. SVM outperforms NN based detector on physical attack databases as well. The proposed algorithm of training AlexNet with Cross Asymmetric Loss Function and an SVM classifier yields lowest error rates across all the experiments.

6. Conclusion

Photo, replay, silicone masks, and digital swapping attacks have been known to spoof the recognition process of biometric systems. Therefore it becomes crucial to have a presentation attack (or spoofing) detection step in the face recognition pipeline to filter out spoofed images or videos before any further action. In this paper, a *panoptic* deep learning based algorithm with Cross Asymmetric Loss Function is presented for face presentation attack detection against both digital and physical attacks. Experiments on SMAD, CASIA-FASD, and SWAPPED databases show that the proposed algorithm achieves state-of-the-art performance compared to the existing presentation attack detection algorithms. As an extension of this research, we plan to improve cross-attack performance and move towards a one-class learning approach for handling face spoofing.

7. Acknowledgement

A. Agarwal is partly supported by Visvesvaraya PhD Fellowship, and M. Vatsa and R. Singh are partly supported through Infosys Center for AI at IIIT-Delhi. M. Vatsa is also partially supported through Department of Science and Technology, Government of India through Swarnajayanti Fellowship.

References

[1] A. Agarwal, R. Singh, and M. Vatsa. Face anti-spoofing using Haralick features. In *IEEE BTAS*, pages 1–6, 2016.

[2] A. Agarwal, R. Singh, M. Vatsa, and A. Noore. Swapped! digital face presentation attack detection via weighted local magnitude pattern. In *IEEE IJCB*, pages 659–665, 2017.

[3] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore. Face presentation attack with latex masks in multispectral videos. In *IEEE CVPRW*, pages 275–283, 2017.

[4] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu. Face anti-spoofing using patch and depth-based CNNs. In *IEEE IJCB*, pages 319–328, 2017.

[5] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. In *NIPS*, pages 153–160, 2007.

[6] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion magnification. In *IEEE CVPRW*, pages 105–110, 2013.

[7] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face spoofing detection using colour texture analysis. *IEEE TIFS*, 11(8):1818–1830, 2016.

[8] P. P. K. Chan, W. Liu, D. Chen, D. S. Yeung, F. Zhang, X. Wang, and C. C. Hsu. Face liveness detection using a flash against 2d spoofing attack. *IEEE TIFS*, 13(2):521–534, 2018.

[9] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[10] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *IEEE ICB*, pages 1–8, 2013.

[11] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. Cheung, and K.-W. Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *JVCIR*, 38:451–460, 2016.

[12] J. Galbally, S. Marcel, and J. Fiérrez. Biometric antispoofing methods: A survey in face recognition. *IEEE Access*, 2:1530–1552, 2014.

[13] K. Kollreider, H. Fronthaler, and J. Bigun. Verifying liveness by multiple experts in face biometrics. In *IEEE CVPRW*, pages 1–6, 2008.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105. 2012.

[15] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot. Learning generalized deep feature representation for face anti-spoofing. *IEEE TIFS*, 2018.

[16] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. *CoRR*, abs/1708.02002, 2017.

[17] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar. Detecting silicone mask-based presentation attack via deep dictionary learning. *IEEE TIFS*, 12(7):1713–1723, 2017.

[18] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *IEEE ICCV*, pages 1–8, 2007.

[19] K. Patel, H. Han, and A. K. Jain. Secure face unlock: Spoof detection on smartphones. *IEEE TIFS*, 11(10):2268–2283, Oct 2016.

[20] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha. Face spoofing detection through visual codebooks of spectral temporal cubes. *IEEE TIP*, 24(12):4726–4740, 2015.

[21] R. Ramachandra and C. Busch. Presentation attack detection methods for face recognition systems: a comprehensive survey. *ACM CSUR*, 50(1):8, 2017.

[22] D. J. Robertson, R. S. Kramer, and A. M. Burton. Fraudulent id using face morphs: Experiments on human and automatic recognition. *PloS One*, 12(3):e0173319, 2017.

[23] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019.

[24] R. Shao, X. Lan, and P. C. Yuen. Joint discriminative learning of deep dynamic textures for 3d mask face anti-spoofing. *IEEE TIFS*, 14(4):923–938, 2019.

[25] T. A. Siddiqui, S. Bharadwaj, T. I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha. Face anti-spoofing with multifeature videolet aggregation. In *ICPR*, pages 1035–1040, 2016.

[26] L. N. Smith. No more pesky learning rate guessing games. *CoRR*, abs/1506.01186, 2015.

[27] L. N. Smith. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*, 2018.

[28] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho. Detection of face spoofing using visual dynamics. *IEEE TIFS*, 10(4):762–777, 2015.

[29] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[30] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. *IAPR ICB*, pages 26–31, 2012.