Face Anti-Spoofing via Sample Learning Based Recurrent Neural Network (RNN)

Usman Muhammad
Muhammad.Usman@oulu.fi
Tuomas Holmberg
tuomas.holmberg@oulu.fi
Wheidima Carneiro de Melo
wheidima.melo@oulu.fi
Abdenour Hadid
hadid.abdenour@oulu.fi

Center for Machine Vision and Signal
Analysis, University of Oulu
Oulu, Finland

Abstract

Face biometric systems are vulnerable to spoofing attacks because of criminals who are developing different techniques such as print attack, replay attack, 3D mask attack, etc. to easily fool the face recognition systems. To improve the security measures of biometric systems, we propose a simple and effective architecture called sample learning based recurrent neural network (SLRNN). The proposed sample learning is based on sparse filtering which is applied for augmenting the features by leveraging Residual Networks (ResNet). The augmented features form as a sequence, which are fed into a Long Short-Term Memory (LSTM) network for constructing the final representation. We show that for face anti-spoofing task, incorporating sample learning into recurrent structures learn more meaningful representations to LSTM with much fewer model parameters. Experimental studies on MSU and CASIA dataset demonstrate that the proposed SLRNN has a superior performance than state-of-the-art methods used now.

1 Introduction

With applications in speech recognition, fingerprint identification, mobile device authentication, and access control, governments, businesses and organizations can use biometric systems to get more information about individuals. Among different biometrics, a facial recognition system is a technology capable of identifying or verifying a person from a digital image by comparing and analyzing facial patterns. Due to advancement in technology, people are expecting secure and convenient ways to access their personal information. On the other hand, criminals are active for spoofing by masquerade or concealing one's identity to gain illegitimate access and advantages. In this regard, a high security requirement for face authentication is needed.

Most early face-anti-spoofing methods are proposed with hand-crafted or low-level features such as SURF, LBP [3, 5], color-texture [38], lips movement [14], etc. However, these methods heavily depend on the human experience to extract detailed information. As an alternative, the Bag of Visual Words (BoVW) or Fisher Vector encoding methods have been commonly adopted [6], in which the input is a set of low-level features and the output is a set of learned features. These types of algorithms are computationally fast and easy to implement, while their accuracy has drawback due to the lack of prior information with respect to the given training samples. These unsupervised methods can develop a high-dimensional vector where the spatial information is not fully exploited. The traditional way of coping with the problem of dimensionality is to use feature learning algorithms such as Independent Component Analysis (ICA) [7], Principal Component Analysis (PCA) [13], Canonical Correlation Analysis (CCA) [31], etc. However, direct feature learning change the original feature space and we cannot specify the spoofed faces in the new feature space. Therefore, the focus is shifting to learn neural network features rather than acquiring hand-crafted (low-level) or clustering based features.

Recently, deep learning models have gained great popularity to learn better features to distinguish the real faces from the spoofing ones. The most commonly used models include deep belief networks (DBNs) [11], stacked auto-encoders (SAEs) [36], and convolutional neural networks (CNNs) [16]. Despite the great success, training DBN, SAE and CNN models remains challenging, since a large number of hyperparameters need to be tuned. By using only one hyperparameter, sparse filtering is introduced that focuses only on optimizing the sparsity of the learned representations [22]. In our problem, we want to propose the idea of sample learning that can map a high-dimensional feature space of the original data to a space of fewer dimensions and maintains the distance information between samples. Most unsupervised face anti-spoofing methods did not consider greedy criteria-based learning with backpropagation. To address these concerns, the proposed sample learning extracts high-dimensional features set from the average pooling layer of the ResNet and ensures that the features can be better represented by the transformed space so that we can specify the exact characteristics from the transformed sample space. By combining recurrent neural network and sample learning, a simple and effective architecture called sample learning based recurrent neural network (SLRNN), is proposed for face anti-spoofing. The overall architecture is optimized in an end-to-end manner.

To summarize, this paper makes the following contributions: (1) We introduce sample learning mechanism into the recurrent structure, which is optimized by data-driven feature learning, whereas the number of model parameters is greatly reduced. (2) We present in-depth analysis of the strengths and limitations of sample learning and recurrent neural network by leveraging residual learning. (3) The augmented features are adaptively predicted from a LSTM in respect of sequence learning.

2 Related work

To solve face anti-spoofing problem, existing approaches can be roughly grouped into two main categories: methods using (a) hand-crafted feature based approaches and (b) deep representation methods especially those deep learning based methods.

Hand-crafted based methods: The texture-based features have been widely analyzed in the early works to detect presentation attacks. For instance, HSV, YCbCr and gray scale texture information is used to train the Local Phase Quantization (LPQ) descriptor [5]. Boulkenafet *et al.* [6] focus on SURF features and apply feature vector (FV) encoding on different color space features. Their work claims that the color features play an important role in face anti-spoofing. Due to different resolutions and illumination conditions of face images, a multiscale filtering is used and features are concatenated from each scale space to form the final

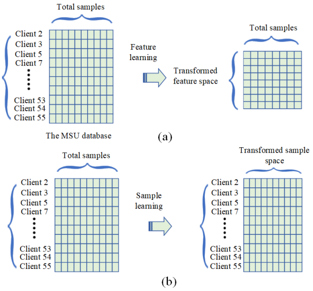

Figure 1: The representation of feature learning and sample learning. (a) A feature learning model for the MSU database. (b) A sample learning model for the the MSU database.

representation [4]. The uniform LBP histograms are extracted from different patch sizes to explore micro-texture analysis. The support vector machine (SVM) classifier is used to distinguish between the real faces and attacking ones [21]. In addition to texture based methods, different variants of Image Distortion Analysis (IDA) features (color moments, blurriness, specular reflection, and color diversity) are developed to capture the image distortion in the spoof face images. The fusion is performed by simply concatenating them [38]. These hand-crafted features have low computational cost which is suitable for real-time application. However, the discriminative ability of them is limited. *Deep learning based methods:* Researchers have explored several ways to use convolutional neural network (CNN) that can be used to enhance the ability of handcrafted features. Nguyen *et al.* [23] form hybrid features by extracting CNN features and the multi-level local binary pattern (MLBP) descriptor to classify the image features into real or presentation attack class. Two streams: patch-based CNN, and depth-based CNN are proposed to learn holistic deep features for anti-spoofing [1]. The off-the-shelf CNN features are proposed and the principle component analysis (PCA) was applied to decrease the dimensions of the features. By applying PCA, an improved detection to distinguish the real and spoofed faces was achieved [18]. Two kind of features such as static and dynamics were separately extracted by using the CNN network and finally fusion is performed to fuse facial features [24]. The temporal structure from video is used to propose an LSTM-CNN architecture because temporal information is found to be useful for face anti-spoofing [39]. By combining a LSTM layer with convolutional neural networks, a deep learning architecture was introduced in [37]. The pre-trained VGGNet is used to extract deep features and the eulerian motion magnification is used with LSTM to get the final detection [35]. Although [20] has some exploration of extracting multi-scale information with LSTM, the above methods leaving room for exploring the characteristics of augmented features with LSTM.

3 The sample learning based recurrent neural network (SLRNN)

We first define a sample learning procedure which will be used to learn the underlying structure of the Long Short-Term Memory (LSTM). Generally, an appropriate parameter initialization is inevitable for LSTM networks to have a satisfactory performance. For instance,

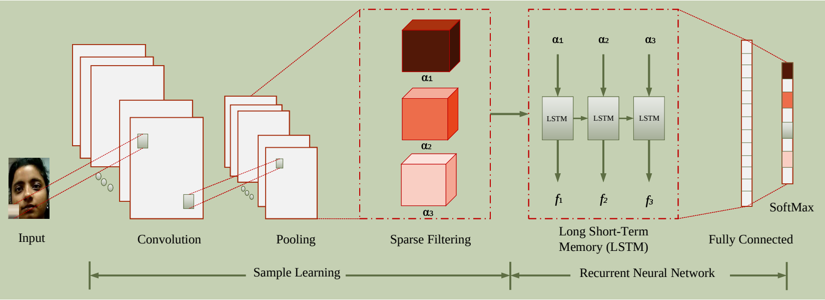

Figure 2: An overview of the proposed sample learning based recurrent neural network, referred to as SLRNN. SLRNN is designed in an end-to-end manner. Specifically, ResNet-50 is applied to extract features and sparse filtering is used to normalize and augmenting the features. Then the augmented features are treated as a sequence which is fed into a Long Short-Term Memory (LSTM) network for the final representation.

If we apply feature learning algorithms to normalize or process the spoofed image data, the original feature space will be transformed and we cannot specify the exact spoofed faces in the new feature space. In order to illustrate this problem intuitively, Fig.1(a) shows the traditional feature learning model. The MSU Mobile Face Spoof database, is taken as an example, where each row expresses number of clients (clients are with spoofed images) and each column denotes a sample. Direct feature learning may also loose the distance information between the samples in new feature space. The proposed sample learning transforms the sample space based on sparse filtering [22] which optimize the sparsity of the learned representations as shown in Fig.1(b). In this way, the features of each spoofed face can be better expressed by the transformed sample space. An LSTM is exploited to learn the representations and maintains the distance information between the samples. Therefore, we end up sample learning with recurrent neural network (SLRNN). We first describe the procedure of sample learning. Then, the LSTM architecture is implemented and discussed.

Sample learning: We utilize ResNet-50 and the last average pooling layer is used to convert the raw images into feature vector. Let's assume that the input dataset as $A \in R^{r \times q}$. To convert into transformed subspace, normalization is applied into B which is used to perform sample learning. Concretely, a sample distribution matrix over B as $M \in R^{p \times q}$, where $M_j^{(i)}$ represent j^{th} rows (feature) value for the i^{th} columns. For making each feature to be equally active, we normalize rows of M by dividing each feature by its l_2-norm:

$$\tilde{M}_j = M_j / \parallel M_j \parallel_2 \tag{1}$$

We then simply normalize by columns, so that they lie on l_2-ball, by calculating:

$$\overset{*}{M}^{(i)} = \tilde{M}^{(i)} / \parallel \tilde{M}^{(i)} \parallel_2 \tag{2}$$

Hence, using the l_1- norm, the normalized elements are optimized for sparseness. For q features in the dataset A, the sparse learning objective can be written as [22]:

$$minimize \quad \sum_{i=1}^{q} \|\overset{*}{M}^{(i)}\|_1 = \sum_{i=1}^{q} \left\| \frac{\tilde{M}^{(i)}}{\|\tilde{M}^{(i)}\|_2} \right\|_1 . \tag{3}$$

The term $\|\overset{*}{M}^{(i)}\|_1 = \left\| \frac{\tilde{M}^{(i)}}{\|\tilde{M}^{(i)}\|_2} \right\|_1$ estimates the population sparsity of the features on the i^{th} space. Hence, following three distributions are achieved.1) Population sparsity: each component in the matrix (or each element) is illustrated by only few (non-zero) features, 2) high dispersal: for activation of each feature, mean squared activation of each feature are similar or to be roughly equal, 3) lifetime sparsity: since these features are highly dispersed, it allow us to distinguish the features because every feature must have a high number of zero entries and be lifetime sparse for achieving a good feature distribution.

In particular, after training average pooling layer of ResNet-50 with sparse filtering, one can compute different normalized features by measuring the relative change in parameter value over each iteration of the algorithm. We use three different parameter values to map visual features into three transformed space where augmentation takes place. L-BFGS method [28] is used to optimize the sparse filtering objective until convergence. As illustrated in Fig.2, the augmented features are treated as the sequence input, which is extremely important for spatial information in the proposed LSTM architecture. A further explanation on function values will be discussed in section 4.

Recurrent Neural Network (RNN): The variations between the augmented features encode additional useful information for the identification of spoofed and real faces. We propose to use LSTM unit [9] which operates augmented features based on the memory cell (M_t), and responsible for keeping track of the dependencies between the elements in the input sequence. The three gates: the input gate (q_t), output gate (k_t) and forget gate (p_t) regulate the flow of information into and out of the cell. Specifically, the input gate regulates the extent to which a new value flows into the cell by multiplying the cell's non-linear transformation of inputs n_t. The output gate decides which the value in the cell is used to transfer to the unit output. The forget gate modulates the extent to which a value remains in the cell. The LSTM unit updates for time step t are:

$$p^t = \sigma(W_p u^t + F_p a^{t-1} + e_p) \tag{4}$$

$$q^t = \sigma(W_q u^t + F_q a^{t-1} + e_q) \tag{5}$$

$$n^t = \phi(W_n u^t + F_n a^{t-1} + e_n) \tag{6}$$

$$M^t = n^t \odot q^t + M^{t-1} \odot p^t \tag{7}$$

$$k^t = \sigma(W_k u^t + F_k a^{t-1} + e_k) \tag{8}$$

$$i^t = \phi(M^t) \odot k^t \tag{9}$$

For time step t, input and output are denoted as u^t and i^t, respectively. W is the input weight matrix. F represents the recurrent weight matrix, and e is the bias vector. $\sigma(u) = \frac{1}{1+e^{-u}}$ and $\phi(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$ are denoted as the element-wise non-linear activation functions, guiding real values to $(0,1)$ and $(-1,-1)$, respectively.

Pre-trained CNN model	Accuracy(%)	Feature Size
VGG-19	85.10±0.20	4096
AlexNet	86.45±0.30	4096
GoogLeNet	84.78±0.40	1000
VGG-16	85.80±0.40	4096
ResNet-18	87.69±0.20	1000
ResNet-50	91.63±0.40	1000
ResNet-101	90.20±0.40	1000
DenseNet-201	89.80±0.30	1000
Inception-ResNet-v2	84.80±0.50	1000
ResNet-50 with SVM	92.43±0.30	2000
ResNet-50 with Sparse filtering	94.10±0.60	800
SLRNN (The proposed)	**98.66±1.00**	**800**

Table 1: Results of different feature extractors and fusion strategies for MSU MFSD dataset using sequence based evaluation metric

During building the LSTM, we aim to encode the hidden sequences into a fixed-length vector $v = (a_1, a_2, ..., a_N)$:

$$v = \sum_{i=1}^{N} z_i c_i \quad i = 1, 2, ..., N \tag{10}$$

Where c_i represents the hidden state at time t_i. The weight z_i denotes the corresponding weight mapping c_i to vector v. It is possible to compute z_i at each step by

$$z_i = \frac{exp(J_i^\top c_{i-1})}{\sum_{x=1}^{N} exp(J_x^\top c_{i-1})} \tag{11}$$

where z_i is regarded as the probability which emphasize the importance of the hidden state c_i. The softmax layer is used for the class probabilities by giving a fixed-length vector v.

4 Experiments

The first dataset, MSU Mobile Face Spoof Database [38], was collected from 280 videos of fake and real faces. The videos were recorded from 35 subjects using two kinds of cameras. The duration of each video is about nine seconds. Different types of attacks such as high definition replay attacks, printed attacks and mobile replay attacks were used to generate attack faces. For the performance evaluation, 15 subjects are used in the training set and the rest for the test.

The second dataset used in the experiments is the CASIA Face Anti-Spoofing database [41]. The videos were collected from 50 subjects, where the images from fake faces were taken in the high resolution recordings of the original faces. Three kinds of fake face attacks were made: cut photo attacks, warped photo attacks and video attacks. In addition, three kinds of image qualities have been used for recording such as low, normal and high. To perform experiments, the dataset is splitted into two subsets for training and testing (20 and 30, respectively).

	CASIA	MSU
Method	EER (%)	EER (%)
Texture analysis [5]	2.1	4.9
Motion mag+LBP [2]	14.4	-
DMD [34]	21.7	-
CNN [40]	7.4	-
IQA [8]	32.4	-
LGBP [26]	4.52	5.10
HSV + YCbCr fusion [6]	2.8	2.2
Generalized Deep Feature [17]	1.4	0.0
Multiscale Fusion [4]	4.2	6.9
LDP-TOP [27]	8.94	6.54
LBP + CM [25]	5.88	8.41
IDA [38]	-	8.5
Colour LBP [3]	7.1	10.6
SPMT + SSD [30]	0.04	-
SLRNN (The proposed)	**0.01**	**0.02**

Table 2: Comparisons between the proposed approach and state-of-the-art methods on CASIA and MSU MFSD dataset using sequence based evaluation metric.

4.1 Experimental Setup

The strategy of adopting a pretrained CNN model as feature extractor is simple, since no data augmentation or fine-tuning is necessary. Moreover, one just needs to choose a CNN model and the features can be extracted from any layer, such as convolutional, pooling or fully-connected. In this paper, several pretrained CNN models, GoogLeNet [32], AlexNet [15], ResNet-18, ResNet-50, ResNet-101 [10], VGG-16, VGG-19 [29], Inception-ResNet-v2 [33] and DenseNet-201 [12], are applied as off-the-shelf CNN features extractors. Table 1 shows the experimental results for the MSU MFSD dataset. After that, we choose ResNet-50 because it gives better performance than others. Since the CNNs learn more generic features on the bottom of the network but the feature distributions of the last fully connected layer does not fully consider the spatial information which is important to enhance the face anti-spoofing detection. It could be observed from the Table 1 that the pooling layer provides better performance in ResNet-50 with support vector machine (SVM) classifier. All the face images are scaled to 224×224 for the pre-defined size requirement of the network.

We use 50, 80 and 120 parameter values to optimize the features in sparse filtering and use these augmented features as input to LSTM. To update network weights of LSTM, an optimizing algorithm based on Adam is used. To prevent from overfitting, we add a dropout layer where small dropout value of 20% is applied. Additionally, a learning rate with a size of 5 is used.

4.2 Results and Analysis

As shown in Table 1, the performance of the proposed SLRNN demonstrates that feature augmentation play an important role for identifying real faces from the spoofing ones. It can be observed that direct adaption of sparse filtering can also benefit the performance of ResNet. The two observations enlighten us naturally that it would be a better solution to

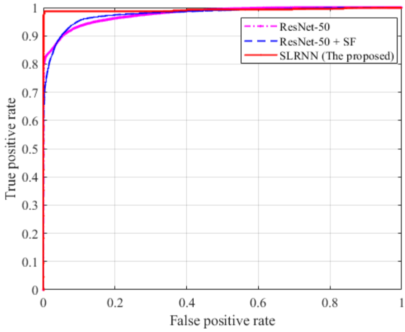

Figure 3: ROC curve of the MSU MFSD database with SLRNN.

use sample learning to adaptively determine the performance of LSTM. We also compare the performance on CASIA FASD database, the proposed method achieves $(99.40 \pm 0.50\%)$ accuracy which is comparable to the approach in [19].

Table 2 provides a comparison between the results of our proposed method with several state-of-the-art approaches on MSU MFSD and CASIA FASD database. The Spatial-temporal information is proposed in [17] to build a 3D convolutional neural network architecture with augmented facial samples. The generalized features are obtained by manipulating their feature distribution distances. A guided scale space is introduced to reduce the redundancy of the original facial texture. To improve the feature information, scale based local binary pattern (GS-LBP) and local guided binary pattern (LGBP) are proposed [26]. The work reported in [5], emphasizes the importance of different colour representations and fuse LPQ and CoALBP descriptors by concatenating their resulting histograms. The high-order Local Derivative Pattern from Three Orthogonal Planes (LDP-TOP) [27] is developed to encode both spatial and temporal information in different directions of subtle face movements. Another fusing scheme is introduced based on SURF and Fisher vector encoding, where fusion is performed using HSV and YCbCr color spaces [6]. A new multiscale space is introduced to expand images using different filtering schemes. Then, feature histograms are concatenated into final feature vector [4]. Different image distortions such as regions (detected face), intensity channels are analyzed to implement on an Android smartphone in [25]. The LBP based descriptor is proposed to investigate the most suitable color space for face spoofing in [3]. In addition to texture based methods, different variants of Image Distortion Analysis (IDA) features (color moments, blurriness, specular reflection, and color diversity) are developed to capture the image distortion in the spoof face images. The fusion is performed by simply concatenating them [25, 38]. Other approaches including, detection with motion magnification [2], face spoofing using visual dynamics [34], detection based on general image quality assessment [8], CNN for face anti-spoofing [40] and discriminative representation combinations for accurate spoofing detection [30].

Making a comparison with above methods, the proposed fusion framework achieves the best equal error rate (0.01) for CASIA and (0.02) for MSU database. It proves that the performance of our proposed approach competes the state-of-the-art methods by a fair margin. For further analysis, the ROC (Receiver Operating Characteristics) curves of MSU MFSD and CASIA are shown in Fig.3, and Fig.4, respectively. The true positive rate (TPR) deter-

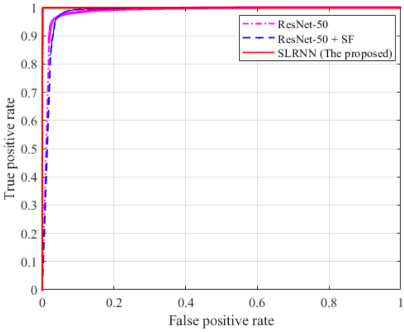

Figure 4: ROC curve of the CASIA database with SLRNN.

mines how many samples are classified as true positives among all positive samples during the test. In contrast to TPR, the false positive rate (FPR) illustrates how many test samples are classified as false positives among all negative samples. As can be seen, our proposed method yields the highest point in the upper left corner.

5 Conclusions

Deep learning has become a new trend for face anti-spoofing due to it's supremacy in terms of performance. However, several fully connected layers are always inserted to the end of CNN models but no clear consensus is available in the literature about which one performs the best. With the advent of deep learning, the pretrained CNNs models with linear SVMs (off-the-shelf representation) have been proved effective as feature extractors for face anti-spoofing in biometric domain, but the potential characteristics and how powerful the CNN features off-the-shelf are not fully understood. In this regard, this paper proposes an effective sample learning based recurrent neural network named SLRNN to make full use of the merits of these three models: CNN, sparse filtering and LSTM. We keep focus on a more generalized features, and show that direct adaptation of ResNet with sparse filtering performs also well. In addition, when combined the augmented features with the gating mechanism in LSTM, the experimental studies confirm that this dynamic changes contained in the spatial features are quite useful to enhance the generalization ability of the proposed method.

Acknowledgment

The financial support of the Academy of Finland is acknowledged.

References

[1] Yousef Atoum, Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Face anti-spoofing using patch and depth-based cnns. In *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pages 319–328. IEEE, 2017.

[2] Samarth Bharadwaj, Tejas I Dhamecha, Mayank Vatsa, and Richa Singh. Computationally efficient face spoofing detection with motion magnification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 105–110, 2013.

[3] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face anti-spoofing based on color texture analysis. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2636–2640. IEEE, 2015.

[4] Zinelabidine Boulkenafet, Jukka Komulainen, Xiaoyi Feng, and Abdenour Hadid. Scale space texture analysis for face anti-spoofing. In *Biometrics (ICB), 2016 International Conference on*, pages 1–6. IEEE, 2016.

[5] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016.

[6] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, 24(2):141–145, 2017.

[7] Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

[8] Javier Galbally and Sébastien Marcel. Face anti-spoofing based on general image quality assessment. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 1173–1178. IEEE, 2014.

[9] Felix A Gers, Nicol N Schraudolph, and Jürgen Schmidhuber. Learning precise timing with lstm recurrent networks. *Journal of machine learning research*, 3(Aug):115–143, 2002.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[12] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.

[13] Ian T Jolliffe. Mathematical and statistical properties of population principal components. *Principal component analysis*, pages 10–28, 2002.

[14] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun. Real-time face detection and motion analysis with application in liveness assessment. *IEEE Transactions on Information Forensics and Security*, 2(3):548–558, 2007.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[16] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[17] Haoliang Li, Peisong He, Shiqi Wang, Anderson Rocha, Xinghao Jiang, and Alex C Kot. Learning generalized deep feature representation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(10):2639–2652, 2018.

[18] Lei Li, Xiaoyi Feng, Zinelabidine Boulkenafet, Zhaoqiang Xia, Mingming Li, and Abdenour Hadid. An original face anti-spoofing approach using partial convolutional neural network. In *Image processing theory tools and applications (IPTA), 2016 6th international conference on*, pages 1–6. IEEE, 2016.

[19] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu. Deep tree learning for zero-shot face anti-spoofing. *arXiv preprint arXiv:1904.02860*, 2019.

[20] Shiying Luo, Meina Kan, Shuzhe Wu, Xilin Chen, and Shiguang Shan. Face anti-spoofing with multi-scale information. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3402–3407. IEEE, 2018.

[21] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *2011 international joint conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011.

[22] Jiquan Ngiam, Zhenghao Chen, Sonia A Bhaskar, Pang W Koh, and Andrew Y Ng. Sparse filtering. In *Advances in neural information processing systems*, pages 1125–1133, 2011.

[23] Dat Tien Nguyen, Tuyen Danh Pham, Na Rae Baek, and Kang Ryoung Park. Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors. *Sensors*, 18(3):699, 2018.

[24] Keyurkumar Patel, Hu Han, and Anil K Jain. Cross-database face antispoofing with robust feature representation. In *Chinese Conference on Biometric Recognition*, pages 611–619. Springer, 2016.

[25] Keyurkumar Patel, Hu Han, and Anil K Jain. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security*, 11(10): 2268–2283, 2016.

[26] Fei Peng, Le Qin, and Min Long. Face presentation attack detection using guided scale texture. *Multimedia Tools and Applications*, pages 1–27, 2018.

[27] Quoc-Tin Phan, Duc-Tien Dang-Nguyen, Giulia Boato, and Francesco GB De Natale. Face spoofing detection using ldp-top. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 404–408. IEEE, 2016.

[28] M Schmidt. Minfunc (a matlab function for unconstrained optimization of differentiable real-valued multivariate functions using line-search methods)(2005). *URL http://www. cs. ubc. ca/˜ schmidtm/Software/minFunc. html.*

[29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[30] Xiao Song, Xu Zhao, Liangji Fang, and Tianwei Lin. Discriminative representation combinations for accurate face spoofing detection. *Pattern Recognition*, 85:220–231, 2019.

[31] Quan-Sen Sun, Sheng-Gen Zeng, Yan Liu, Pheng-Ann Heng, and De-Shen Xia. A new method of feature fusion and its application in image recognition. *Pattern Recognition*, 38(12):2437–2448, 2005.

[32] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[33] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[34] Santosh Tirunagari, Norman Poh, David Windridge, Aamo Iorliam, Nik Suki, and Anthony TS Ho. Detection of face spoofing using visual dynamics. *IEEE transactions on information forensics and security*, 10(4):762–777, 2015.

[35] Xiaoguang Tu, Hengsheng Zhang, Mei Xie, Yao Luo, Yuefei Zhang, and Zheng Ma. Enhance the motion cues for face anti-spoofing using cnn-lstm architecture. *arXiv preprint arXiv:1901.05635*, 2019.

[36] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103. ACM, 2008.

[37] Yiren Wang and Fei Tian. Recurrent residual learning for sequence classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 938–943, 2016.

[38] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.

[39] Zhenqi Xu, Shan Li, and Weihong Deng. Learning temporal features using lstm cnn architecture for face anti-spoofing. pages 141–145, 2015.

[40] Jianwei Yang, Zhen Lei, and Stan Z Li. Learn convolutional neural network for face anti-spoofing. *arXiv preprint arXiv:1408.5601*, 2014.

[41] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *Biometrics (ICB), 2012 5th IAPR international conference on*, pages 26–31. IEEE, 2012.