Deep Learning based Face Liveness Detection in Videos

Yaman AKBULUT
Informatics Dept.
Fırat University
Elazig, Turkey
yamanakbulut@firat.edu.tr

Abdulkadir ŞENGÜR
Technology Faculty,
Electrical & Electronics Eng.
Fırat University
Elazig, Turkey
ksengur@firat.edu.tr

Ümit BUDAK
Engineering Faculty,
Electrical-Electronics Eng.
Bitlis Eren University
Bitlis, Turkey
ubudak@beu.edu.tr

Sami EKİCİ
Technology Faculty,
Energy Systems Eng.
Fırat University
Elazig, Turkey
sekici@firat.edu.tr

***Abstract*—The human face is an important biometric quantity which can be used to access a user-based system. As human face images can easily be obtained via mobile cameras and social networks, user-based access systems should be robust against spoof face attacks. In other words, a reliable face-based access system can determine both the identity and the liveness of the input face. To this end, various feature-based spoof face detection methods have been proposed. These methods generally apply a series of processes against the input image(s) in order to detect the liveness of the face. In this paper, a deep-learning-based spoof face detection is proposed. Two different deep learning models are used to achieve this, namely local receptive fields (LRF)-ELM and CNN. LRF-ELM is a recently developed model which contains a convolution and a pooling layer before a fully connected layer that makes the model fast. CNN, however, contains a series of convolution and pooling layers. In addition, the CNN model may have more fully connected layers. A series of experiments were conducted on two popular spoof face detection databases, namely NUAA and CASIA. The obtained results were then compared, and the LRF-ELM method yielded better results against both databases.**

***Index Terms*—Face recognition, face spoof detection, deep learning, CNN, LRF-ELM**

I. INTRODUCTION

Face recognition plays a critical role in the authentication of users, and is essential for many user-based systems [1]. The past decade has seen the rapid development of face recognition in many fields [2]. Face recognition systems are faced with various types of face spoof attacks such as print-attack, replay-attack, and 3D mask attack [3].

Patel et al. studied face spoof detection on mobile phones. They used mobile face spoof databases to develop a prototype that runs on the Android mobile operating system. The authors also built a spoof face database called MSU MSF, which contains more than 1,000 subjects [3]. Wen et al. proposed an efficient face spoofing detection algorithm. The authors goal was to design a system with good generalization ability in addition to a quick response. Image distortion analysis is the key role in the algorithm to extracting the feature vector. Features consist of specular reflection, blurriness, chromatic moment, and color diversity. Printed photo attack and replay video attack are used as a face spoof attack in order to determine between a live and a spoof face. Multiple support vector machine (SVM) classifiers were employed for the classification task [4]. Tirunagari et al. developed an algorithm for facial anti-spoofing detection. They exploited the content of videos by using an algorithm called dynamic mode decomposition (DMD) to capture liveness cues such as blinking eyes, moving lips, and other facial dynamics. To show the effectiveness of the algorithm, experimental studies were performed on three public databases [5]. In the literature, Komulainen et al. were the first to have investigated the dynamic texture of the face for the purposes of face spoof detection. An approach was introduced to learn the structure of facial texture by using a local binary pattern (LBP) algorithm. Experiments on two public databases showed experimental results beyond the state-of-the-art in 2013 [6]. Tan et al. presented a real-time and non-intrusive method for face spoofing detection. Their method's involved analysis of the Lambertian model. To realize this method, a large face spoof database was collected with 15 subjects under various conditions of illumination. Over 50,000 photograph images were captured by a standard webcam. Evaluation of the proposed method provided promising performance for spoof detection [7]. Zhang et al. released a face anti-spoofing database with 50 subjects. The database covered three types of attacks and consisted of three imaging qualities, and is described in detail in Section III. SVM was used in order to reach a final decision in the classification process. The authors hoped that the database could serve future works on face spoofing [8].

In this current paper, a deep-learning-based spoof face detection is proposed. In order to achieve this, two different deep learning models are used, namely LRF-ELM and CNN. The LRF-ELM model contains a convolution layer, a pooling layer, and a fully connected layer. In addition, the CNN model has five convolutional layers and three fully connected layers. Rectification linear unit (RELU) and local response normalization layers follow the first and second convolution layers. There are also five Max-pooling layers in the model which follow some of the convolution layers. There are two dropout layers, which come after the first and second fully

978-1-5386-1880-6/17/$31.00 ©2017 IEEE

connected layers, with a probability of 0.5. Finally, a loss layer is used as the last layer. Face spoof detection has been analyzed in terms of print-attack and replay-attack. A series of experiments were conducted on two popular spoof face detection databases, namely NUAA and CASIA.

The organization of this paper is as follows: in Section II, the components of a deep learning model are briefly introduced. The core of the work is in Section III, where the databases, deep learning models and experimental results are given. In addition, all the experimental results and related comparisons are within Section III. The concluding remarks and future works plan are then given in Section IV.

II. Face Spoof Detection Method

In the literature, the authors generally use a framework for face spoof detection where firstly a feature extraction stage is handled and then a classification stage follows the previous stage. In this current work, the aim is to use a compact structure where both feature extraction and classification stages are combined. To this end, the more recently popular deep CNN and LRF-ELM methods are considered. Details about the deep models are provided as follows. The flowchart of the proposed method is shown in Fig. 1.

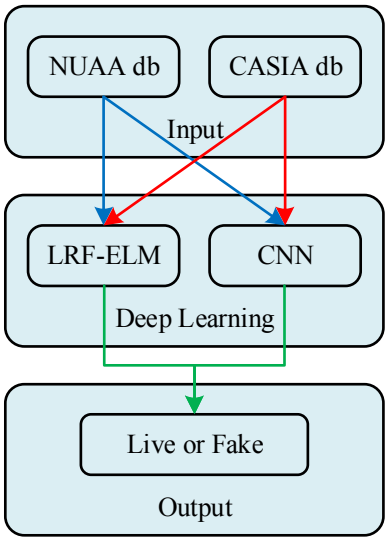

Fig. 1. The flowchart of the proposed method.

A. A Brief Theory for Deep Models

This section briefly reviews the theory of the deep model. See [9, 10] for a more detailed explanation of LRF-ELM and CNN models. A general deep learning model is composed of convolution layer, pooling layer, and a fully connected layer.

1) *Convolution layer*: This layer is known as the core layer of the CNN architecture. There is a collection of learnable filters in this layer. During training of the CNN, each filter is convolved across the width and height of the input volume in the forward pass. After convolution operation, 2-Dimensional activation maps of the filters are constructed. As a result, the network learns filters that activate when they see some specific type of feature at some spatial position in the input.

2) *Pooling Layer*: Another important concept of the CNN architecture is pooling. It forms a non-linear down-sampling. Pooling operation can be handled with several non-linear functions. Max pooling seems the most common, whereby the input image is partitioned into a set of non-overlapping rectangle sub-regions. For each sub-region, the maximum value is used as the output. The pooling operation reduces the spatial size of the input, which also reduces the amount of parameters and computation in the network.

3) *Fully Connected Layer*: After several convolutional and pooling layers, the classification process is handled in a fully connected layer. Neurons in a fully connected layer have full connections to all activations in the previous layers. Their activations can be computed with a matrix multiplication followed by a bias offset.

III. Experimental Works

As mentioned earlier, two deep models were considered, namely CNN and LRF-ELM. The LRF-ELM model contains a convolution layer, a pooling layer and a fully connected layer. In addition, the CNN model has five convolutional layers and three fully connected layers. Rectification linear unit (RELU) and local response normalization layers follow the first and second convolution layers. There are also five Max-pooling layers in the model which follow some of the convolution layers. There are two dropout layers, which come after the first and second fully connected layers, with a probability of 0.5. Finally, a loss layer is used as the last layer. It is worth mentioning that the all input images are resized to 32×32 pixels for the LRF-ELM model and 224×224 pixels for the CNN model.

In order to evaluate the performance of the proposed methods, experiments were conducted on two public face spoof databases. Comparison on NUAA and CASIA databases are tabulated in Table I. The related information about the databases can be seen in sub-sections A and B.

TABLE I. Comparison of Databases

	NUAA [7]	***CASIA*** [8]
Subject (#)	15	50
Data type	Still image	Video
Data (#)	23,641	600
Attack types	Printed	Printed Cut Replay
Quality	1	3

A. NUAA Database

The NUAA database was constructed to distinguish a real face from a photograph by using a generic webcam. It was collected in different illumination conditions and places. There

are fifteen subjects in this work. The authors captured two types of images; live subject images called *Client* and their photographs called *Imposter* [7]. Samples of *Client normalized* and *Imposter normalized* images are shown in Fig. 2.

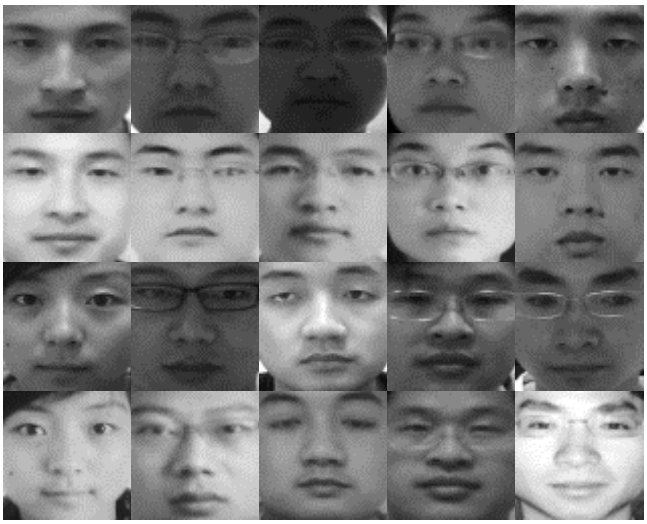

Fig. 2. Sample images of NUAA database. Rows 1 and 3: Client normalized. Rows 2 and 4: Imposter normalized.

In this experimental work, a database of geometrically normalized gray-scale face images was utilized. The normalized database contains 3,362 live subjects for client testing and 1,743 live subjects for client training. In addition, it has 5,761 photographs for imposter testing and 1,748 for imposter training. Each of the images in the database has 8-bit gray-scale with 64×64 pixels, and the total number of images in the normalized database is 12,614.

B. CASIA Database

The CASIA face spoof database, consisting of 50 subjects, was built by Zhang et al. in order to determine live faces from fake face attacks [8]. Three kinds of attacks were designed for this purpose, which are printed photo attack, cut photo attack, and video replay attack. Attack types from video images are shown in the second, third, and fourth rows of Fig. 3.

Three different cameras were used to capture three different imaging quality videos for the database (low-resolution, normal-resolution, high-resolution). The low-resolution videos are sized at 480×640 pixels, and normal-resolution at 640×480 pixels. However, although the high-resolution videos have an original size of 1920×1080 pixels, to save computing costs, the authors cropped them to 1280×720 pixels. The video quality is shown in the first, second, and third columns of Fig. 3.

While arranging the database, each subject has a set of 12 videos (three live, nine fake), as shown in Fig. 3. The test part of the database has 30 subjects and therefore 360 videos. For the training part, there are 240 videos recorded for 20 subjects. In total, the database has 600 videos as can be seen in Table I.

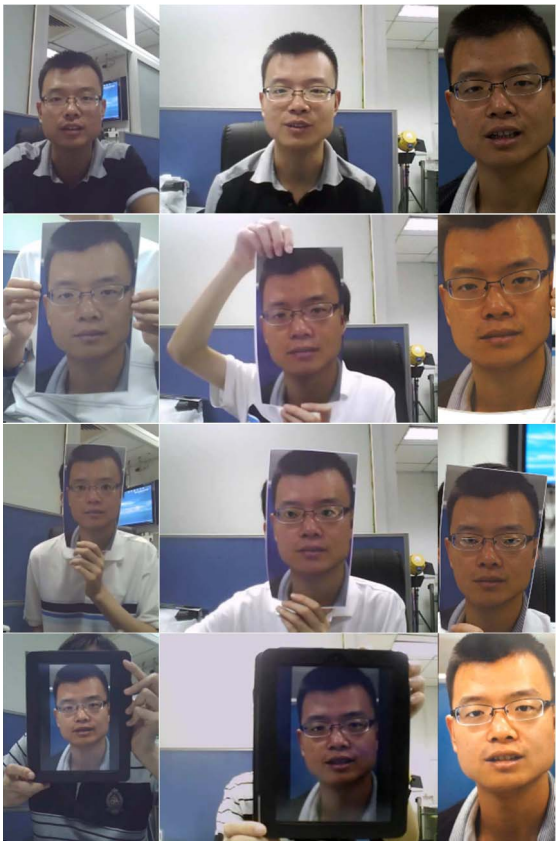

Fig. 3. Sample images of CASIA database. Row 1: Live, Row 2: Printed photo attack, Row 3: Cut photo attack, Row 4: Video replay attack. Column 1: Low-resolution, Column 2: Normal-resolution, Column 3: High-resolution.

C. Performance Evaluation

Experiments were conducted on both databases with two deep models. The convolutional layer of LRF-ELM model contains 40 filters of size 5×5. The regularization parameter (C) of the LRF-ELM method was chosen as 0.2. The batch size for LRF-ELM model was assigned as 500. In addition, CNN model's three convolutional layers contain 64 filters of size 11×11, 256 of size 5×5, and 256 of size 3×3. The learning parameter of the CNN model was fixed to 0.001 and the batch size was chosen as 25.

TABLE II. OBTAINED RESULTS

	NUAA [7]	***CASIA*** [8]
CNN	76.31%	82.03%
LRF-ELM	84.04%	88.75%

The obtained results are given in Table II. The LRF-ELM model yields higher accuracy values for both databases. For the NUAA database, while the LRF-ELM model obtains a correct classification rate of 84.04%, the CNN model achieves 76.31%. In other words, the LRF-ELM model produces almost 8% more accurate results. Similar performances are seen for the CASIA database. The LRF-ELM model yields almost 6% more accurate results than for the CNN model.

IV. Conclusions

In this paper, a comparative study is achieved on the detection of face liveness. Face liveness detection is a hot topic in digital forensic environments, where the reliability of face-based access systems is in demand. With the development of deep learning tools, more real world applications are being proposed. In this work, the authors of this paper developed a deep-learning-based face spoof detection system. Popular deep learning methods (LRF-ELM and CNN) are used for face liveness detection purposes. Two widely used face liveness detection databases are utilized in this research. The obtained results show that the LRF-ELM method produced more accurate results for both databases. In addition, the period of training time for the LRF-ELM method is shorter than for the CNN model. In future works, the authors plan to enhance the performance of CNN by using different deep models. In addition, the plan is to use various sizes of face images in order to improve the quality of the CNN model.

References

[1] E. Hjelmås and B. K. Low, "Face Detection: A Survey," *Comput. Vis. Image Underst.*, vol. 83, pp. 236–274, 2001.

[2] H. Zhou, A. Mian, L. Wei, D. Creighton, M. Hossny, and S. Nahavandi, "Recent advances on singlemodal and multimodal face recognition: A survey," *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 6, pp. 701–716, 2014.

[3] K. Patel, H. Han, and A. K. Jain, "Secure Face Unlock: Spoof Detection on Smartphones," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 10, pp. 2268–2283, 2016.

[4] Di Wen, Hu Han, and A. K. Jain, "Face Spoof Detection With Image Distortion Analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 4, pp. 746–761, 2015.

[5] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 4, pp. 762–777, 2015.

[6] J. Komulainen, A. Hadid, and M. Pietikäinen, "Face spoofing detection using dynamic texture," *Comput. Vis. - ACCV 2012 Work.*, pp. 146–157, 2013.

[7] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Computer Vision–ECCV 2010*, 2010, pp. 504–517.

[8] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proceedings - 2012 5th IAPR International Conference on Biometrics, ICB 2012*, 2012, pp. 26–31.

[9] A. Krizhevsky, I. Sutskever, and H. Geoffrey E., "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[10] G.-B. Huang, Z. Bai, L. L. C. Kasun, and C. M. Vong, "Local Receptive Fields Based Extreme Learning Machine," *IEEE Comput. Intell. Mag.*, vol. 10, no. 2, pp. 18–29, 2015.