Journal of Industrial Information Technology and Application *ISSN(Online): 2586-0852*
Vol.8 No.2

Learning deep models for face anti-spoofing by pixel-wise supervision with depth labels

Myoung-Kyu Sohn[1,*], Sang-Heon Lee[1], Hyunduk Kim[1], and Junkwang Kim[1]
[1] Division of Automotive Technology, DGIST, Republic of Korea.

Abstract. With the development of face recognition technology, vision-based face recognition systems are widely used. At the same time, various methods of attacking these face recognition systems have also begun to emerge. In this paper, we implement two types of anti-spoofing systems that detect such presentation attacks in a face recognition system using a deep learning network. The performance of the two systems was compared. A simple binary classifier using the entire face image and a depth information-based classifier that estimates depth information on a pixel basis is implemented. The performance of the implemented network was evaluated using the CelebA-Spoof database and the results of the two networks were compared.

Keywords; face recognition, anti-spoofing, deep learning

1. Introduction

In recent, face recognition technology has received a lot of attention and is being applied to various fields such as security, authentication, surveillance, and payment systems [1]. However, these face recognition systems are exposed to attacks using reproduced face photos. And it is assumed that attacks on face recognition systems through such spoofing will increase. Face presentation attacks mainly use photos, video playbacks, or masks rather than the actual face of the user in the field. These fake faces can trick recognition systems into recognizing them as normal users, causing security and reliability issues. Therefore, the development of technologies for face anti-spoofing is expected to play a key role in increasing the reliability of face recognition systems [2, 3].

* Corresponding author: smk@dgist.ac.kr
Received: Feb. 10, 2024; Accepted: May. 9, 2024; Published: Jun. 30, 2024

In this paper, we propose a method of face anti-spoofing technique based on depth information. Depth information provides 3D spatial information of a face obtained from a 2D image, and its use can help distinguish fake faces from real faces. In addition, this study compares the results of simple binary classification compared to face presentation attack detection using depth information.

2. Design and Implementation

There are several deep learning-based anti-spoofing methods to distinguish fake faces in face recognition, but in this paper, we design and implement the two most efficient methods, simple binary classification and depth information-based classification [4,5]. Then we compare the simulation results. Both classifiers have a basic CNN-based structure. The simple binary classification is a method that directly classifies fake faces and real faces through binary cross-entropy loss. However, such binary classification can easily learn inconsistent patterns, such as screen bezels or photo borders. In contrast, pixel-level learning can learn internal features in more detail. Figure 1 shows the structure of the architecture used in each deep learning models.

The network used in binary classification stacked several convolutional layers and used a fully connected layer at the final stage. The final output is a vector that can immediately distinguish between a fake face and a real face. The depth information-based classifier has a U-net [6] type network structure and uses down-sampling using convolution followed by up- sampling using transpose convolution. Instead of using a fully connected layer, the output from up-sampling is used directly in final classification, and the output is trained to have depth information in training stage.

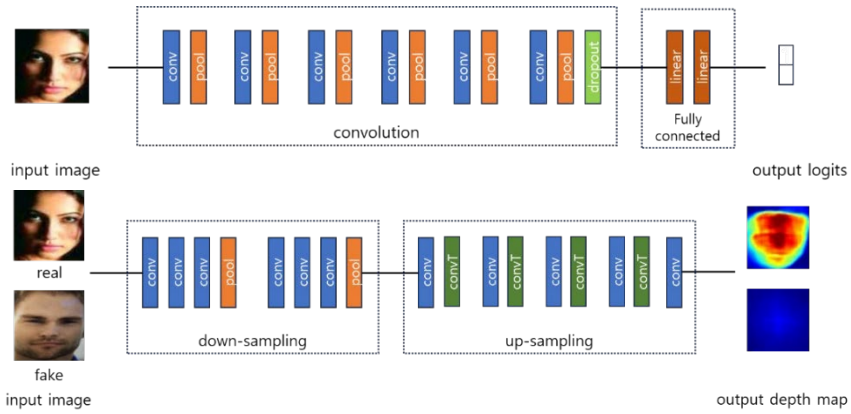

Figure 1. Overall architecture of deep learning network.

a) Binary classification (upper), b) Pixel-wise depth estimation (lower) To compare the above methods, the following experiment was conducted. CelebA-Spoof

[7] was used as the face database. The database contains approximately 600,000 face photos from a total of 10,177 people. In addition, many real-life situations were assumed as the photos were taken indoors/outdoors, in various lighting environments, and with various devices. In the case of binary classification, the network output value was directly used, and in the depth information-based classifier, the average value of the output depth was used for classification by a simple threshold value.

Table I. RECOGNITION RATE FOR EACH METHOD

	apcer (%)	**bpcer (%)**	**acer (%)**	**acc (%)**
binary	0.33	33.04	16.69	85.14
depth	0.04	16.94	8.49	93.89

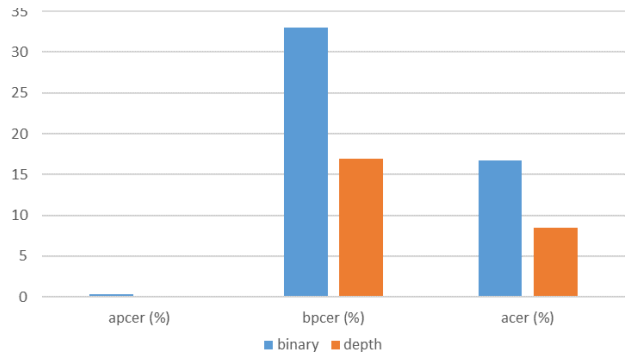

Figure 2. Recognition rate for each method

Table 1 and Figure 2 show the classification performance of each classifier. Each measure APCER(attack presentation classification error rate), BPCER(bona fide presentation classification error rate), and ACER(average classification error rate) have the following meanings. APCER refers to a case where a fake face is incorrectly determined to be a real face, and BPCER refers to a case where a real face is incorrectly determined to be a fake face. ACER is the average value of APCER and BPCER. ACC refers to the recognition rate in the binary classification of fake face and real face. A common characteristic of each classification result is that APCER has a significantly smaller value than BPCER. This means that fake faces are detected well, while real faces are often recognized as fake. Comparing the two classifiers, the classifier based on depth information showed better results in all measures. In the case of ACER, the error rate was reduced by about 2 times and the recognition rate was about 93.9% in pixel-wise

recognition. Figure 3 shows the input and corresponding output results of the depth information-based classifier.

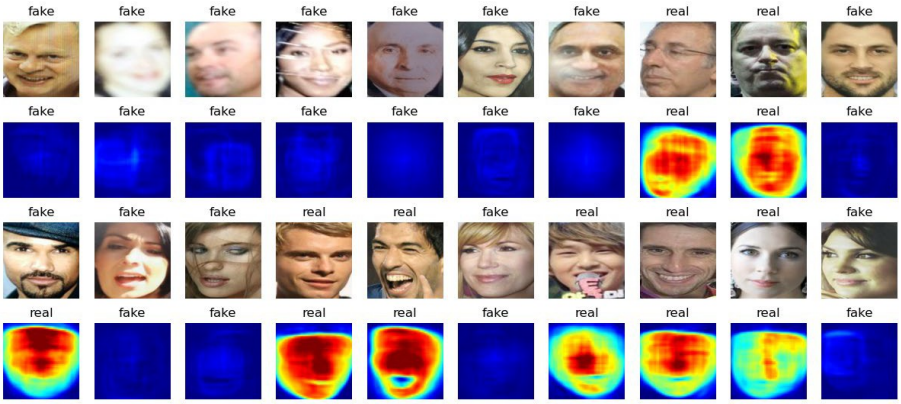

Figure 3. Simuation examples (1st and 3rd rows : input image 2nd and 4th rows : estimated depth and final result (failure case: 1st image of the 3rd row))

3. Conclusion

In this paper, we implemented two deep learning methods for face anti-spoofing and compared the performances of the two methods. It detects presentation attacks that can disable the face recognition systems. Compared to a simple binary classifier, a classifier based on estimating depth information on a pixel-by-pixel basis showed much better performance. In the future, we plan to research to further increase the recognition rate by simultaneously using various information such as binary classification and depth information.

Acknowledgment

This work was supported by the DGIST R&D Program of the Ministry of Science and ICT (23-IT-02). It was also supported by the Technology Development Program of MSS (S3237206)

References

[1] Fuad, Md Tahmid Hasan, et al. "Recent advances in deep learning techniques for face recognition." IEEE Access 9 (2021): 99112-99142.

[2] Deng, Jiankang, et al. "Retinaface: Single-shot multi-level face localisation in the wild." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[3] Li, Xiaobai, et al. "Generalized face anti-spoofing by detecting pulse from face videos." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016.

[4] Yu, Zitong, et al. "Deep learning for face anti-spoofing: A survey." arXiv preprint arXiv:2106.14948 (2021).

[5] Liu, Yaojie, Amin Jourabloo, and Xiaoming Liu. "Learning deep models for face anti-spoofing: Binary or auxiliary supervision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[6] Ronneberger, et al. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015

[7] Zhang, Yuanhan, et al. "Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations." European Conference on Computer Vision. Springer, Cham, 2020.