Effect of Facial Shape Information Reflected on Learned Features in Face Spoofing Detection

Su-Gyeong Yu[1], So-Eui Kim[1], Kun Ha Suh[1], Eui Chul Lee[2*]

[1] Dept. of AI & Informatics, Graduate School, Sangmyung University, Republic of Korea
[2] Department of Human-Centered AI, Sangmyung University, Republic of Korea
tnrud7495@gmail.com, soeui291@gmail.com, tjrjsgk@naver.com, eclee@smu.ac.kr

Abstract

Face recognition is a convenient and non-contact biometric method used widely for secure personal authentication. However, the face is an exposed body part, and face spoofing attacks, which compromise the security of systems that use face recognition for authentication, are frequently reported. Previous face spoofing attack detection studies proposed texture-analysis-based methods using handcrafted features or learned features to prevent spoofing attacks. However, it is unclear whether spoofing attack images reflect the face distortion resulting from failing to reflect the three-dimensional structure of a real face. To resolve this problem, we compared and analyzed the face spoofing attack detection performances of two typical convolutional neural network models, namely ResNet-18 and DenseNet-121. CASIA-FASD, Replay-Attack, and PR-FSAD were used as the training data. The classification performance of the model was evaluated based on four protocols. DenseNet-121 exhibited better performance in most scenarios. DenseNet-121 reflected facial shape information well by uniformly applying the learned features of both the initial and final layers during training. It is expected that this study will support the realization of spoofing technology with enhanced security.

Keywords: Face recognition, Spoofing attacks, Convolutional neural network, Facial shape information

1 Introduction

Biometric information is difficult to manipulate, share, and lose because it contains the unique characteristics of an individual [1]. Therefore, biometrics are utilized in many internet-based systems. Biometrics typically include fingerprint, iris, vein, and face recognition. Among these, face recognition is a more convenient method with less rejection rates [2-3]. However, several potential vulnerabilities in biometric systems have been reported in recent years, particularly related to spoofing attacks [4]. For example, in an unmanned store where authentication is done through face recognition, a minor can purchase items such as alcohol or cigarettes through a spoofing attack using a parent's photo. Spoofing attacks collectively refer to attacks that use fake data to deceive people or systems [5]. Fake faces used in face spoofing attacks are created using printed photos and 2D or 3D masks [6]. Consequently, anti-spoofing has been garnering attention recently to ensure the reliability of facial recognition [7-10].

A previous study on face spoofing used texture analysis to distinguish between real and fake faces. This method directly extracts the desired features of the face image using techniques such as local binary pattern, Gabor wavelets, histogram of oriented gradient, and difference of Gaussians, and distinguishes between real and fake faces using a linear or nonlinear support vector machine [10-11]. However, since it requires a long time to process a large amount of data, the classification accuracy for complex data is low. Also, there is an inconvenience that humans have to extract the features of the data themselves. Therefore, recent studies on spoofing detection have used convolutional neural network (CNN) models, which automatically extract and learn the features of the input data for classification [12]. This method exhibits high performance in image recognition and classification. In the convolutional layer of a CNN, a feature map is created by extracting features from images [13]. Shape information, such as edges and blobs, are extracted from initial convolutional layers, and features, such as textures and objects, are learned by combining them with features extracted from previous layers as the propagation continues. In a fully connected layer, a CNN classifies the input data based on the extracted features. However, traditional CNN models can face degradation problems as the layers deepen. Additionally, the feature information of an image extracted from the initial layer is gradually blurred. Therefore, ResNet and DenseNet, which are neural network models with new structures, were proposed to resolve this problem. ResNet was designed to enable normal learning up to deeper layers; thus, it learns features from deep layers effectively [14]. DenseNet resolves the degradation problem because its structure preserves the image features from the initial layer to deeper layers relatively well [15].

However, when these existing models are applied to a real situation, they may show low accuracy depending on the light reflection and angle of the face. This is because various environmental factors and situations have not been applied to commonly used public face databases. Therefore, in this study, PR-FSAD [16] containing various lighting, angle, and distance information was used as training data, and CASIA-FASD [7] and Replay-Attack [8], which are public face databases, were additionally used for model performance evaluation. In addition, based on comparison of the face spoofing detection performance of ResNet and DenseNet, we investigated whether maintaining the facial shape information, which is the learned feature of the initial layer, as well as learned features of the deep layer, affects the performance of

*Corresponding Author: Eui Chul Lee; E-mail: eclee@smu.ac.kr

DOI: 10.53106/160792642022052303010

these neural network models in distinguishing a real face from a fake face. The results of this study can contribute to the development of a face recognition security system with higher accuracy in real situations.

The remainder of this paper is structured as follows. Section 2 describes the database, the structure of the two neural network models used in this study, and four protocols for evaluating the performance of the two neural network models. Section 3 presents the experimental results for each protocol, and Section 4 analyzes the results. Finally, Section 5 concludes the study and discusses the future research direction.

2 Methods

2.1 Database

We used two public databases, CASIA-FASD and Replay-Attack, and PR-FSAD, a database developed by the laboratory, as training data for the fake face detection model.

CASIA-FASD comprises three types of spoofing attacks, namely printed photographs, photographs with the eyes cut out, and replays of recorded videos. Further, it includes various factors affecting the detection of face spoofing attacks, such as low-, normal-, and high-quality images [7]. Replay-Attack DB is a database built at the Idiap research facility. It includes printed photo and video playback attacks. These attacks are performed by displaying a printed photo or recording a video for at least 9 s under two different lighting conditions [17]. Figure 1 and Figure 2 show examples of the images in CASIA-FASD and Replay-Attack DB. However, most face anti-spoofing databases do not contain face data captured at various distances and angles. Instead, most databases contain faces with a fixed background and artificial lighting. In real life, face recognition systems are mainly used in various angles, distances, backgrounds, and lighting. Therefore, systems trained with databases that do not reflect these characteristics may not work effectively in real life. Therefore, in addition to CASIA-FASD and Replay-Attack DB, we used PR-FSAD, which comprises face data captured from various distances and angles, as the learning data.

PR-FSAD contains face data captured from three angles and distances, along with conventional spoofing attacks, namely printed photographs and video playbacks [16]. In PR-FSAD, the angles are top, middle, and bottom. The images corresponding to the top and bottom angles were taken by offsetting the camera to approximately ±30° from the front. Distance is divided into near, halfway, and distant. Since the physical structure of each subject is different, the relative ratio of the face occupying the photographing device's screen was used rather than the absolute value. Figure 3 and Figure 4 show faces images obtained from each angle and distance, respectively. Furthermore, the face images from PR-FSAD have varying background and lighting environments. Existing face recognition systems do not account for the background in face images. However, a face recognition system in real life operates in a varying environment, and may be affected by factors inside the camera such as automatic brightness or white balance. For this reason, PR-FSAD contains face images in natural environments without a fixed background.

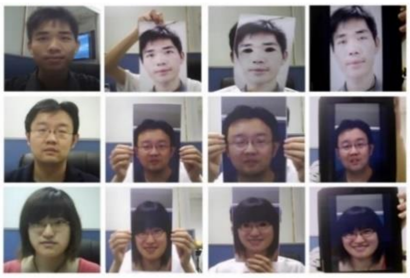


Figure 1. Three types of spoofing attacks in CASIA-FASD (left to right: original face, printed photograph, photograph with eyes cut out, video playback)

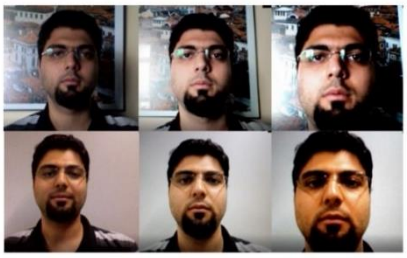
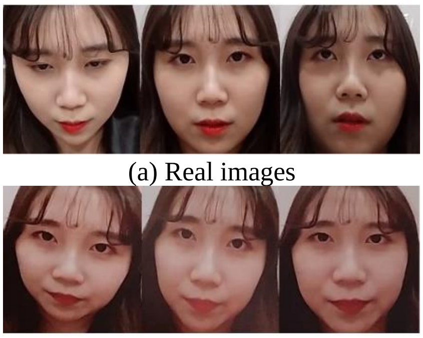


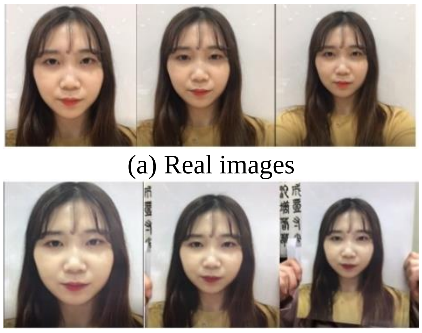
Figure 2. Two types of spoofing attacks in Replay-Attack DB (left to right: original face, printed photograph, video playback)



(a) Real images

(b) Fake images

Figure 3. PR-FSAD face data based on angle (left to right: top, middle, and bottom angles)



(a) Real images

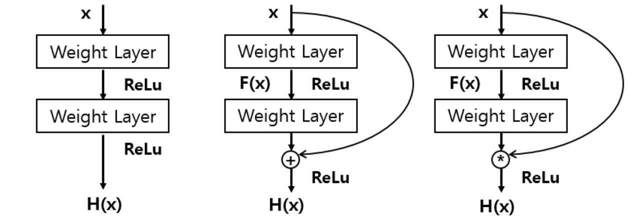
(b) Fake images

Figure 4. PR-FSAD face data based on distance (left to right: near, halfway, distant)

2.2 Experimental Method

In this study, two neural network models, namely, ResNet-18 and DenseNet-121, were used. ResNet has a "shortcut connection" structure that connects the input and output of the layer [14]. The existing CNN's use the output data of the layers as the input of the next layer, and their purpose is to obtain the optimal condition of F(X)=H(X). On the other hand, ResNet aims to obtain F(x)=H(x)-x using the input data again for the output of the subsequent output layer through a shortcut connection. This can be written as H(x)=F(x)+x, which indicates that the previously learned information is preserved and used for additional learning. Therefore, ResNet exhibits high performance in learning features of images by sensitively detecting subtle changes in the input by learning the residuals of the network through shortcut connections. Using this method, ResNet can be optimized relatively easily even if the layer is deeper, and better accuracy can be obtained as the number of layers increases.

DenseNet is similar to ResNet but has a different structure of shortcut connection. Since the shortcut connection of ResNet proceeds through the addition operation of the feature map, it extracts and learns image features by reflecting and combining the information of the initial layer. Unlike this, DenseNet has a dense connection that connects the input and output of a layer using a concatenation operation between the feature maps [15]. In the dense connection, the learning features of the initial layer are reflected for the first image and the previous feature maps are merged in the forward direction until the learning process is complete. In other words, this structure not only preserves the facial shape information of the image, which is a learned feature of the initial layer, relatively better than other neural networks but also alleviates the loss due to the gradient problem, enhances function propagation, encourages function reuse, and reduces the number of parameters. Figure 5 depicts the basic structure of the three models.



(a) Existing CNN (b) ResNet (c) DenseNet

Figure 5. Structure of CNN models

Therefore, ResNet-18 and DenseNet-121, which have the aforementioned characteristics, were selected as the neural network models to be used for face spoofing attack detection in this study. Figure 6 and Figure 7 show the detailed structures of the two models. As the input data for the model, a face image of resolution 224 × 224 was used. In the learning process, the stride was set to 2, binary-cross entropy for the loss function was used, and the probability gradient descent (SGD) was used as the optimization function.

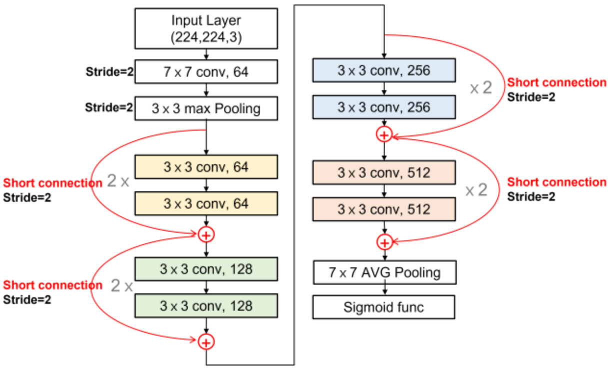


Figure 6. Structure of ResNet-18

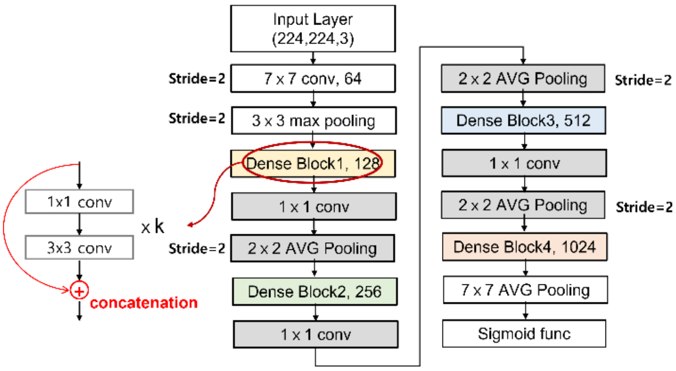


Figure 7. Structure of DenseNet-121

In addition, we trained each model by setting the learning rate to 0.01, 0.05, and 0.001; the batch size was set to 8, 16, 32, etc. The value resulting in the highest performance was set as the hyperparameter value of the optimization function. Through this process, the learning rate of the two models was set to 0.001, the batch size of ResNet-18 was set to 16, and the batch size of DenseNet-121 was set to 8. The order of the image files was randomly imported to prevent the order of the data from being used as a learning pattern when the training data were provided as input to the neural network model. The epoch, which is the maximum number of repetitions of learning, was designated as 100 for PR-FSAD and Replay-Attack DB, and 17 was set as the optimal epoch obtained by the K-fold method for CASIA-FASD. At this time, an early stopping technique was used to prevent overfitting caused by repeated learning. When the value calculated by the loss function increases by more than 10 times, learning is automatically stopped, and only the models with the best learning results are saved during the iteration. Furthermore, validation was performed for each epoch. The training process of DenseNet-121 can be seen in Figure 8.

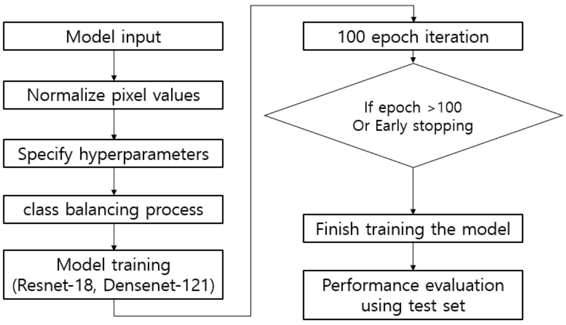

Model input
Normalize pixel values
Specify hyperparameters
class balancing process
Model training (Resnet-18, Densenet-121)
100 epoch iteration
If epoch >100 Or Early stopping
Finish training the model
Performance evaluation using test set

Figure 8. Training process

2.3 Experimental Protocol

The experiment using the previously constructed ResNet-18 and DenseNet-121 proceeded as follows. First, a protocol consisting of four scenarios was designed to evaluate the performance of PR-FSAD and measure the validation value. In Protocol 1, the data containing the information for each of the three angles were used; in Protocol 2, the data containing two types of distance information (near and distant) were used. This is a process of learning based on each angle and distance, followed by testing using the dataset of the corresponding characteristic. By contrast, Protocol 3 performed learning and testing using both protocols 1 and 2. Finally, Protocol 4 was performed to evaluate the performances of two public databases, namely CASIA-FASD and Replay-Attack DB. The experimental process was the same as in Protocol 3. After learning with the training set for the entire data, testing was performed with the test set. To execute the designed protocol, the data were divided into training, validation, and test sets, with the exception of CASIA-FASD, which was divided into training and test sets, without a separate validation set. Therefore, CASIA-FASD's training set was divided into four folds using the k-fold method, of which three were used for learning and one for verification. The number of subjects was used as the criterion for dividing the dataset. Table 1 shows the detailed configuration for each protocol.

Table 1. Data configuration for each protocol

Protocol			Training	Validation	Test	Total
PR-FSAD	1	Top	13,680	10,080	18,720	42,480
		Middle	13,680	10,080	18,720	42,480
		Bottom	13,680	10,080	18,720	42,480
	2	Near	13,680	10,080	18,720	42,480
		Distant	13,680	10,080	18,720	42,480
	3	Total	41,040	30,240	56,160	127,440
Public-Database	4	CASIA-FASD	27,901	-	40,389	68,290
		Replay-Attack	92,934,	92,975	123,790	309,699

3 Results

In this study, an experiment was conducted to determine how the structure that preserves facial shape information of an image affects the performance of the neural network models in distinguishing between real and fake faces. To this end, the performances of ResNet-18 and DenseNet-121, which have excellent binary classification performance among neural network models that automatically extract and learn features, were compared. First, we used PR-FSAD, considering the angle, distance, and varying background, to learn various possible situations. Precision, recall, and accuracy were used as indicators to verify its performance on face spoofing attack detection. Precision is the proportion of true positives among data classified as True. Recall is the probability of classifying data that are actually True as True. Accuracy was determined using the results of normal classification and misclassification. Precision, recall, and accuracy are calculated as

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \textbf{(1)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \textbf{(2)}$$

$$\text{Accuracy} = \left(\frac{\text{TN+TP}}{\text{TN+FP+FN+TP}}\right) * 100.0(\%) \quad \textbf{(3)}$$

Here, a true positive (TP) returned as a classification result is the correct classification of a real face, and a true negative (TN) is the correct classification of a fake face. In addition, a false positive (FP) means that a fake face is incorrectly classified as a real face, and a false negative (FN) means that a real face is incorrectly classified as a fake face. Additionally, we used a confusion matrix, one of the simplest and most intuitive methods used to evaluate the performance of binary classification models, for calculating the normal and misclassified images. The confusion matrices obtained for ResNet-18 and DenseNet-121 are shown in Table 2 and Table 3, respectively. At this time, the denominators of genuine positive and genuine negative indicate a real positive (real) image and real negative (fake) image, respectively. ResNet-18 misclassified ~3.24% (1214/37440) of fake face data and ~3.247% (608/18720) of real face data. DenseNet-121 misclassified ~2.337% (875/37440) of fake face data and ~2.339% (438/18720) of real face data. Considering that the amount of fake data using printed photographs and video replays in both models was approximately twice that of real face data, it can be considered that the classification performances for fake and real faces were equal. Furthermore, DenseNet-121 performed better in face spoofing detection than ResNet-18.

Table 2. Confusion matrices of ResNet-18 (TNR: true negative rate, FPR: false positive rate, FNR: false negative rate, TPR: true positive rate)

Confusion matrix	Predicted fake	Predicted real
Actual fake	36,044 / 37,440, 96.27% (TN / GN, TNR)	1,396 / 37,440, 3.73% (FP / GN, FPR)
Actual real	700 / 18,720, 3.74% (FN / GP, FNR)	18, 020 / 18,720, 96.26% (TP / GP, TPR)

Table 3. Confusion matrices of DenseNet-121

Confusion matrix	Predicted fake	Predicted real
Actual fake	36, 565 / 37, 440, 97.67% (TN / GN, TNR)	875 / 37, 440, 2.34% (FP / GN, FPR)
Actual real	438 / 18, 720, 2.34% (FN / GP, FNR)	18, 282 / 18, 720, 97.66% (TP / GP, TPR)

Table 4. Results of precision, recall, and accuracy for each protocol

Protocol		ResNet-18			DenseNet-121		
		Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)
Protocol 1	Top	94.92	97.39	97.39	96.01	97.96	97.96
	Middle	90.52	95.03	95.03	93.25	96.51	96.51
	Bottom	93.85	96.79	96.82	93.31	96.54	96.54
Protocol 2	Near	93.49	96.63	96.63	95.04	97.44	97.45
	Distant	91.63	95.61	95.62	92.13	95.93	95.91
Protocol 3	Total	92.81	96.26	96.27	97.66	96.43	97.66
Protocol 4	CASIA-FASD	90.30	96.66	96.62	93.16	97.65	97.65
	Replay-Attack	99.69	99.90	99.90	99.90	99.97	99.97

Precision, recall, and accuracy for each of the protocols are summarized in Table 4. As mentioned in section 2.2, because CASIA-FASD has no separate validation data, verification was performed using the k-fold method. Comparing the performances of ResNet-18 and DenseNet-121, the precision, recall, and accuracy on PR-FSAD were approximately 92.81%, 96.26%, and 96.27 for ResNet-18, and approximately 97.66%, 96.43%, and 97.66% for DenseNet-121. The test execution time was approximately 0.016 s and 0.029 s per image for ResNet-18 and DenseNet-121, respectively. Precision, recall, and accuracy for CASIA-FASD were 90.30%, 96.66%, and 96.62% for ResNet-18, and 93.16%, 97.65%, and 97.65% for DenseNet-121, respectively. With the Replay-Attack DB, the corresponding values were 99.69%, 99.90%, and 99.90% for ResNet-18, and 99.90%, 99.97%, and 99.97% for DenseNet-121. These results show that DenseNet-121 performed better than ResNet-18 in detecting face spoofing. In addition, when PR-FSAD was used as the training data, it was possible to obtain a relatively superior face spoofing detection performance than when public DBs were used for training.

To visually analyze the performance, the receiver operating characteristic (ROC) curve was used. A larger area under the curve (AUC) (the base of the ROC curve) implies a better model performance. The equal error rate (EER) line is drawn on the ROC curve. EER is the point where the line intersects the curve; it refers to the rate at which the false recognition rate and false rejection rate become equal. Figure 9 to Figure 11 show the ROC curves for the face spoofing detection results for each database. In all the databases, the ROC curve for DenseNet-121 is closer to the upper left; therefore, the AUC is wide. Figure 12 shows the ROC curve for Protocol 1. Among the three angles, DenseNet-121 showed a better performance than ResNet-18 for the top and middle angles. As mentioned earlier, DenseNet retained the facial shape features of images better than ResNet due to its structural characteristics. Therefore, the maintenance of facial shape features affects the improvement of face spoofing detection performance. However, for the bottom angles, ResNet-18 performed slightly better. In the discussion section, the impact of facial shape features is elaborated and error cases are analyzed for each model.

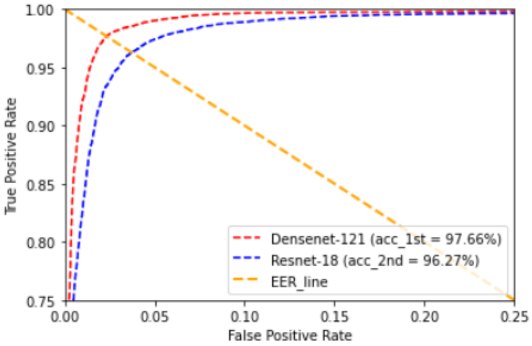


Figure 9. ROC curves for ResNet-18 and DenseNet-121 on PR-FSAD

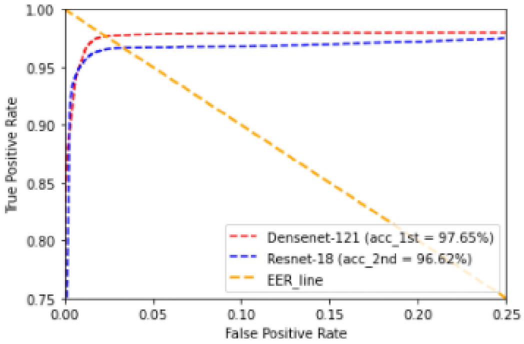


Figure 10. ROC curves for ResNet-18 and DenseNet-121 on CASIA-FASD

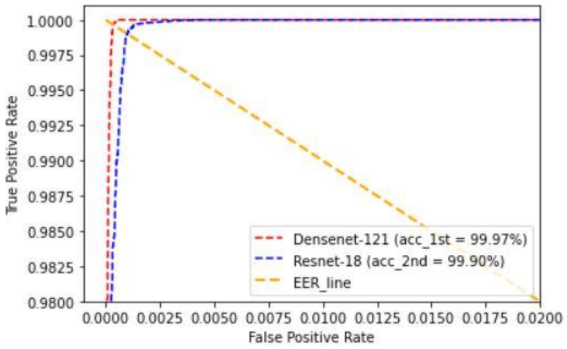
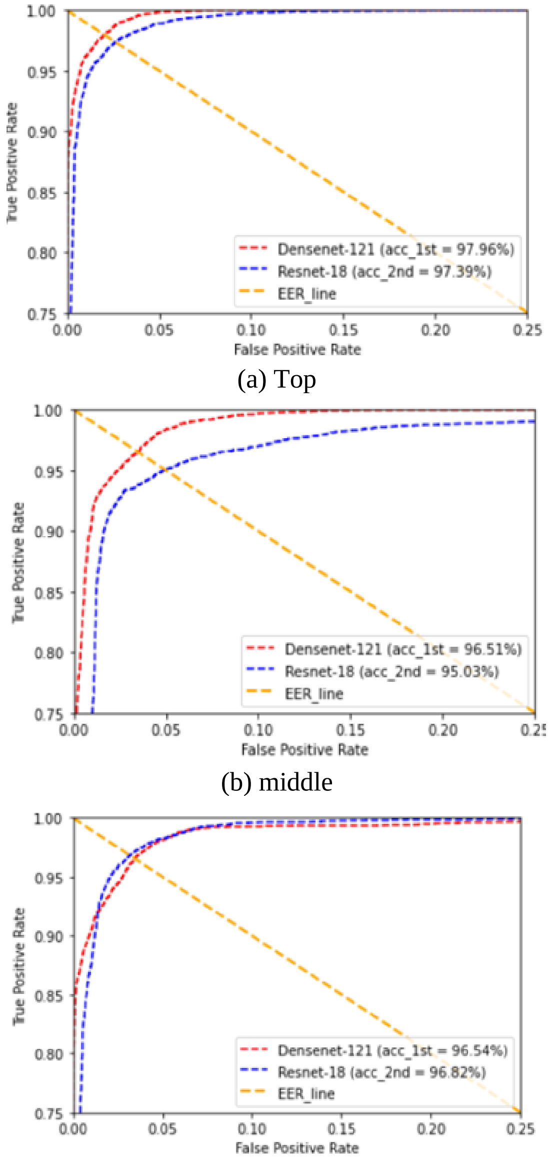


Figure 11. ROC curves for ResNet-18 and DenseNet-121 on Replay-Attack DB



(a) Top

(b) middle

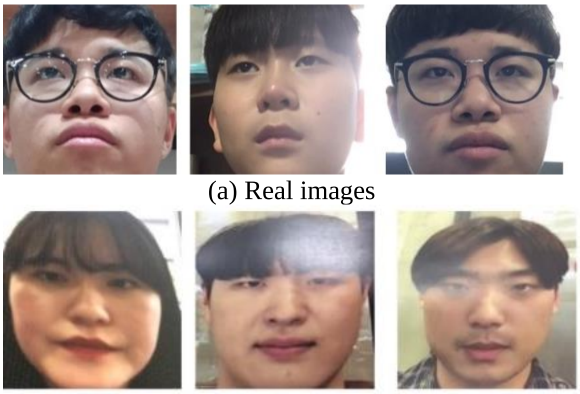
(c) bottom angles

Figure 12. ROC curves for ResNet-18 and DenseNet-121 on Protocol 1

4 Discussion

PR-FSAD uses the characteristic of angle, unlike the public database. Therefore, it can be said that PR-FSAD can better reflect the actual scenarios that would occur. Previously, we evaluated the classification performance based on the angle in Protocol 1. As shown in Table 4, at the top angle, the accuracy of DenseNet-121 was ~0.57% (97.39% vs. 97.96%) higher, and for the middle angle, it was ~1.48% (95.03% vs. 96.51%) higher. Exceptionally, for bottom angles, the accuracy of ResNet-18 was higher than that of DenseNet-121 by 0.28% (96.82% vs. 96.54%). Therefore, in this section, error cases of each model for Protocols 1 and 3 are analyzed in detail.

First, we extracted images that were misclassified in DenseNet-121 but well-classified in ResNet-18 for the "bottom" angle in Protocol 1. Figure 13 shows an example of the extracted error image. From Figure 13(a), which shows the real images among the extracted error case, it is seen that light is reflected on the spectacles and specular reflection occurs, or sunlight is captured in the background in the photograph. In Figure 13(b), which shows the fake images in the extracted error case, a periodic pattern generated by the paper texture can be observed, or a specific area of the image is highly saturated by specular reflection caused by an indoor lighting source. It can be concluded that all these features (periodic or rapid changes in terms of spatial frequency) are confusing factors affecting the classification performance of DenseNet-121. However, ResNet-18, which learns more information about the terminal layer than DenseNet-121, seems to detect these features well. Therefore, ResNet-18 has superior performance over DenseNet-121 in learning and testing with "bottom" angle images.



(a) Real images

(b) Fake images

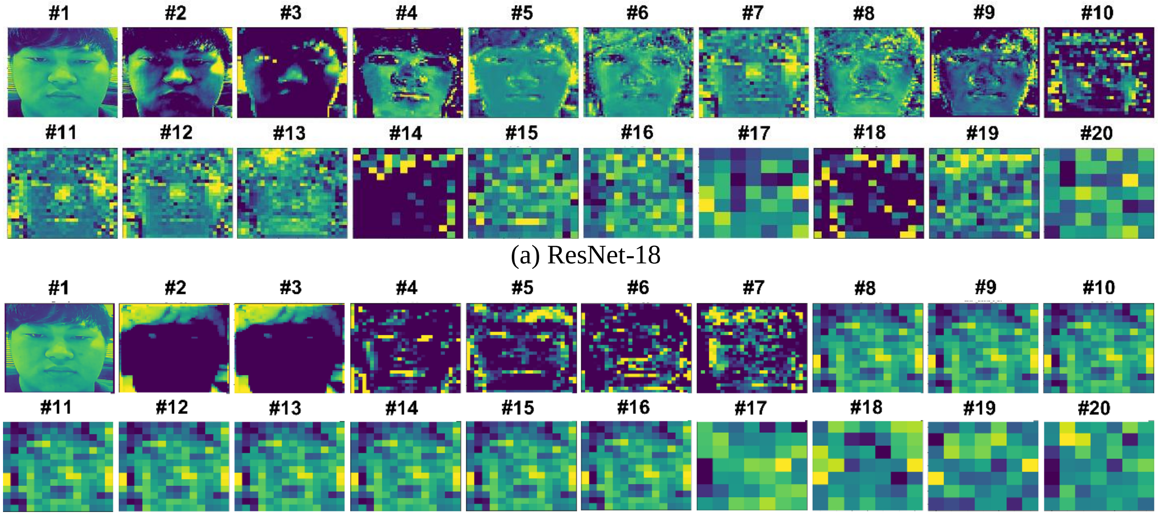
Figure 13. Examples of error cases in bottom angles of images (correctly classified by ResNet-18 but misclassified by DenseNet-121)

Next, for Protocol 3 tested after training all PR-FSAD images, we analyzed images classified well in DenseNet-121 but misclassified in ResNet-18. Table 5 shows the extracted misclassification images divided by three angle labels. Of the error cases, fake images accounted for ~0.62% (fake error images/total error images=896/1429), and real images accounted for ~0.38% (real error images=533/1429). In addition, misclassification results ordered by each angle were in the order of bottom, top, and middle. In particular, the bottom angle has the highest distribution at 53.88%. The "bottom" angle images misclassified by DenseNet-121 constituted ~1.22% ("bottom" angle error images/total "bottom" angle images=520/42,480), and that by ResNet-18 constituted ~2.51% ("bottom" angle error images/total "bottom" angle images=1067/42,480). This observation shows that although ResNet-18 had slightly better performance than DenseNet-121 in classifying "bottom" angle images in Protocol 1, DenseNet-121 had better performance than ResNet-18 in classifying “bottom” angle images in Protocol 3. DenseNet-121 can detect distortion and boundary information that ResNet-18 cannot through a dense connection structure using a concatenation operation. Therefore, DenseNet-121 is better at detecting face spoofing attacks than ResNet-18 because it reflects the difference in distortion and boundary information between fake and real faces well during training.

Finally, feature maps were visualized to check the differences in the features used by each model for training. Figure 14 is a visualization of the functional maps of ResNet-18 and DenseNet-121 used to classify the bottom angle images. Considering the difference in the number of layers between ResNet-18 and DenseNet-121, 20 feature maps were extracted at equal intervals for each model. Colors closer to yellow represent characteristics mainly reflected in the learning. The feature map of ResNet-18 in Figure 14(a) shows that the training was focused on the skin area of the face and the brightness change caused by light reflection. It seems that the texture was used as a feature in #10 – #13, and the face shape information was preserved up to #13, but it cannot be confirmed for the following feature maps. Conversely, as shown in Figure 14(b), DenseNet-121 learned the boundaries of the face, hair, and background in feature maps #2 and #3. In feature maps #4 – #7, eye, nose, mouth, and face shapes were used for training. In addition, the face shape information was preserved up to #16 (extracted from the deep layer). Shape information is a learned feature extracted from the initial layer. In the case of face data, 3D information of an image can vary according to the angle. Considering this, the above results show that DenseNet-121 is superior to ResNet-18 in detecting face shape and distortion information because it better preserves the information from lower layers to the higher layers.

Table 5. Proportion of real and fake images misclassified only by ResNet-18, for all three angles

Class	Top	Middle	Bottom	Total
Real	1.61% (23/1429)	13.08% (187/1429)	22.60% (323/1429)	37.29% (533/1429)
Fake	17.07% (244/1429)	14.35% (205/1429)	31.28% (447/1429)	62.70% (896/1429)
Total	18.68% (267/1429)	27.43% (392/1429)	53.88% (770/1429)	100% (1429/1429)



(a) ResNet-18

(b) DenseNet-121

Figure 14. Visualization result of feature maps extracted at 20 equal intervals from the used backbone network (Black pixel: not reflected; yellow: most reflected)

5 Conclusion

In this study, the face spoofing detection performances of two neural network models were evaluated using PR-FSAD and the public databases CASIA-FASD and Replay-Attack DB. ResNet-18 and DenseNet-121 were used for classification, and precision, recall, and accuracy were used as indicators to evaluate the classification performance. The face spoofing attack detection performance of DenseNet-121 was better than that of ResNet-18 for all three databases. We further analyzed this result in terms of the angles highlighted by the features of PR-FSAD. By analyzing the error cases of each model, DenseNet-121 was found to detect high-frequency and low-frequency information such as face shape, distortion, and texture better than ResNet-18. This can be attributed to the fact that, unlike the skip connections of ResNet, the dense connection structure of DenseNet-121 concatenates the information of previous layers. This inference could be confirmed through the feature map visualization of each model.

In our future study, we will perform a cross-validation for the three databases used in this study. In addition, we will conduct spoofing attack detection studies based on various forged data by utilizing generative adversarial networks to artificially generate forged data. This is expected to contribute to the development of Internet-based software with higher face recognition security.

Acknowledgement

This research was supported by a 2020 Research Grant from Sangmyung University.

Su-Gyeong Yu and So-Eui Kim contributed equally to this work.

References

[1] A. Ross, A. K. Jain, Information fusion in biometrics, *Pattern recognition letters*, Vol. 24, No. 13, pp. 2115-2125, September, 2003.

[2] A. K. Das, M. Wazid, N. Kumar, A. V. Vasilakos, J. J. Rodrigues, Biometrics-based privacy-preserving user authentication scheme for cloud-based industrial Internet of Things deployment, *IEEE Internet of Things Journal*, Vol. 5, No. 6, pp. 4900-4913, December, 2018.

[3] A. A. Elngar, M. Kayed, Vehicle Security Systems using Face Recognition based on Internet of Things, *Open Computer Science*, Vol. 10, No. 1, pp. 17-29, March, 2020.

[4] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, F. Roli, Security evaluation of biometric authentication systems under real spoofing attacks, *IET Biometrics*, Vol. 1, No. 1, pp. 11-24, March, 2012.

[5] A. Hadid, Face biometrics under spoofing attacks: Vulnerabilities, countermeasures, open issues, and research directions, *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, Ohio, USA, 2014, pp. 113-118.

[6] N. Erdogmus, S. Marcel, Spoofing face recognition with 3D masks, *IEEE transactions on information forensics and security*, Vol. 9, No. 7, pp. 1084-1097, July, 2014.

[7] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S. Z. Li, A face antispoofing database with diverse attacks, *2012 5th IAPR international conference on Biometrics (ICB)*, New Delhi, India, 2012, pp. 26-31.

[8] N. Ratha, J. Connell, R. Bolle, Enhancing security and privacy in biometrics-based authentication systems, *IBM systems Journal*, Vol. 40, No. 3, pp. 614-634, 2001.

[9] L. Li, P. L. Correia, A. Hadid, Face recognition under spoofing attacks: countermeasures and research directions, *IET Biometrics*, Vol. 7, No. 1, pp. 3-14, January, 2018.

[10] A. George, S. Marcel, Deep pixel-wise binary supervision for face presentation attack detection, *Proc. 2019 International Conference on Biometrics (ICB)*, Crete, Greece, 2019, pp. 1-8.

[11] X. Shu, H. Tang, S. Huang, Face spoofing detection based on chromatic ED-LBP texture feature, *Multimedia Systems*, Vol. 27, No. 2, pp. 161-176, April, 2021.

[12] G. B. De Souza, D. F. Da S. Santos, R. G. Pires, A. N. Marana, J. P. Papa, Deep texture features for robust face spoofing detection, *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 64, No. 12, pp. 1397-1401, December, 2017.

[13] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, Vol. 521, No. 7553, pp. 436-444, May, 2015.

[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 2016, pp. 770-778.

[15] G. Huang, Z. Liu, K. Q. Weinberger, L. Maaten, Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA, 2017, pp. 4700-4708.

[16] J. Y. Bok, K. H. Suh, E. C. Lee, Verifying the Effectiveness of New Face Spoofing DB with Capture Angle and Distance, *Electronics*, Vol. 9, No. 4, Article No. 661, April, 2020.

[17] Z. Boulkenafet, J. Komulainen, A. Hadid, Face anti-spoofing based on color texture analysis, *2015 IEEE International Conference on Image Processing (ICIP)*, Quebec City, QC, Canada, 2015, pp. 2636-2640.

Biographies

Su-Gyeong Yu received the B.S. degree in Human-Centered Artificial Intelligence from Sangmyung University, Seoul, South Korea, in 2020. She received the M.S degree in the Department of AI & Informatics from Sangmyung University, Seoul, South Korea, in 2022. Her current research interests include remote biometrics and AI.

So-Eui Kim received the B.S. degree in Human-Centered Artificial Intelligence from Sangmyung University, Seoul, South Korea, in 2020. She received the M.S degree in the Department of AI & Informatics from Sangmyung University, Seoul, South Korea, in 2022. Her current research interests include remote biometrics and AI.

Kun Ha Suh received a B.S. degree in Computer Science from Sangmyung University, Seoul, South Korea, in 2015. He received M.S. and Ph.D. degrees in Computer Science from Sangmyung University, in 2017 and 2021, respectively. Currently, he is a researcher at R&D team in Zena Inc., working on vision AI. His research interests include computer vision, image processing, pattern recognition, and video based health monitoring.

Eui Chul Lee received a B.S. degree in Software from Sangmyung University, Seoul, South Korea, in 2005. He received M.S. and Ph.D. degrees in Computer Science from Sangmyung University, in 2007 and 2010, respectively. He was Researcher in Division of Fusion and Convergence of Mathematical Sciences at the National Institute for Mathematical Sciences from March 2010 to February 2012. Since March 2012, he has been

Assistant Professor in the Department of Intelligent Engineering Informatics for Human at Sangmyung University, Seoul, Korea. His research interests include computer vision, image processing, pattern recognition, and human computer interface (HCI).