symmetry

MDPI

Article

NIR Reflection Augmentation for DeepLearning-Based NIR Face Recognition

Hoon Jo and Whoi-Yul Kim *

Department of Electronics and Computer Engineering, Hanyang University, Seoul 04763, Korea; hjo@vision.hanyang.ac.kr

* Correspondence: wykim@hanyang.ac.kr

Received: 22 August 2019; Accepted: 27 September 2019; Published: 3 October 2019

Abstract: Face recognition using a near-infrared (NIR) sensor is widely applied to practical applications such as mobile unlocking or access control. However, unlike RGB sensors, few deep learning approaches have studied NIR face recognition. We conducted comparative experiments for the application of deep learning to NIR face recognition. To accomplish this, we gathered five public databases and trained two deep learning architectures. In our experiments, we found that simple architecture could have a competitive performance on the NIR face databases that are mostly composed of frontal face images. Furthermore, we propose a data augmentation method to train the architectures to improve recognition of users who wear glasses. With this augmented training set, the recognition rate for users who wear glasses increased by up to 16%. This result implies that the recognition of those who wear glasses can be overcome using this simple method without constructing an additional training set. Furthermore, the model that uses augmented data has symmetry with those trained with real glasses-wearing data regarding the recognition of people who wear glasses.

Keywords: face recognition; deep learning; data augmentation; near-infrared image

1. Introduction

Recent studies in the field of computer vision have achieved many successes using deep learning. Deep-learning-based face recognition is one of the most actively researched domains. One representative method is DeepFace [1]. This approach estimates a 3D face model from images, and the estimated model helps the recognition. FaceNet [2] is another notable neural network for the task. FaceNet introduces a unified system which outputs an embedding that can be used in various applications, such as identification, verification, and clustering. Including the papers above, most of the research in deep-learning-based face recognition is concerned with visible light (RGB) images to demonstrate high performance under challenging conditions by making the architectures deeper.

However, active near-infrared (NIR)-light-based recognition systems are superior to visible-light-based systems in terms of reliability and security.

In terms of reliability, illumination-invariant images can be acquired under external lighting environments (e.g., natural light, indoor lighting, or low lighting) without any efforts when an active NIR light system is used. A change in lighting conditions makes recognition tasks harder. This increases the complexity of algorithms in order to deal with all possible lighting conditions. Given that active light systems provide constant lighting conditions, a good basis for face recognition is constructed without additional complications [3]. This advantage is especially useful for authentication systems on mobile devices, which should provide a consistent user experience in any environment.

Security is a more significant issue. An NIR face recognition system is an effective way to prevent spoofing attacks—unauthorized attempts to bypass the system with fake faces. The most common type of spoofing attack is the use of photos either printed on paper or displayed on a digital device.

An NIR face recognition system can easily block such attempts in the image acquisition stage because fake faces in the NIR spectrum are different from real faces, as shown in Figure 1.

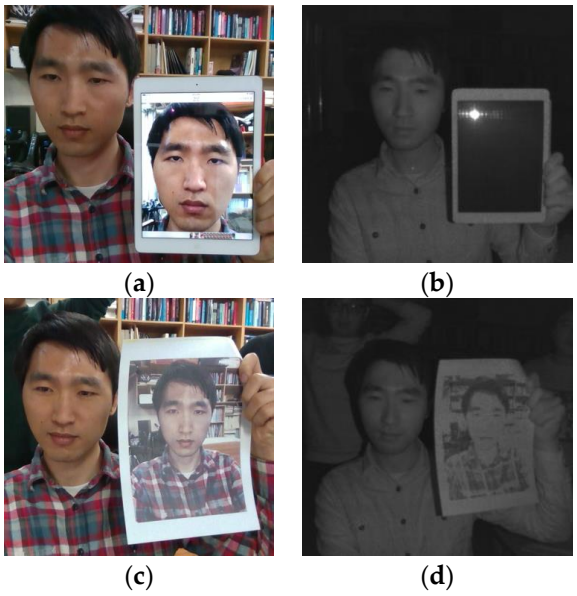


Figure 1. Spoofing attacks on RGB and near-infrared (NIR) images: (**a**) RGB image holding a face image in RGB on a digital device; (**b**) NIR image of (a); (**c**) RGB image holding a printed face image in RGB; (**d**) NIR image of (**c**). All images were taken by the Intel RealSense SR 300.

Although many studies [3,4] have considered these advantages with respect to NIR face recognition, few studies have applied deep learning to NIR face recognition. Some studies have used deep learning for cross-modality face recognition between RGB, NIR, and thermal. These provide an elegant solution to the complex problem by utilizing multiple networks [5–8]. However, there are very few deep-learning studies that focus solely on NIR face recognition. To the best of our knowledge, NIRFaceNet [9] is the only existing deep learning architecture designed for NIR face recognition. In this work, a compact architecture is introduced that is modified from GoogLeNet [10] for less computation. However, their experiments do not sufficiently validate the performance in real-world situations. Compared to FaceNet, which used a maximum of 260 million images for training, NIRFaceNet only used 591 images from the small-sized NIR face dataset.

Furthermore, NIRFaceNet [9] evaluated its performance on a closed set. However, in real-world applications, the faces for training the system and the faces in actual use are different (i.e., open set). More importantly, they did not include glasses wearers in their research. The recognition of glasses wearers is essential in this field because the reflections of NIR lighting on eyeglasses cause significant interference to the recognition, as shown in Figures 2 and 3a. People who wear glasses also account for about half of the total population.

In this paper, we tried to apply a deep-learning approach to NIR face recognition. First, several public NIR databases were collected to construct a massive NIR face database, as each database individually was too small for a deep learning approach. Thus, we used a larger NIR database than other NIR face recognition studies for training and evaluation. In order to determine how the size of the architecture affected NIR face recognition, two deep-learning architectures were compared using the same test conditions. Because face images in the NIR databases tend to have less head pose variation than images in the RGB databases, small architectures can recognize faces with adequate performance. Lastly, we propose a data augmentation method for improving the recognition of glasses wearers. We compared recognition rates between the networks trained with the original and augmented datasets. Open-set recognition was used throughout the experiments, assuming real-world use.

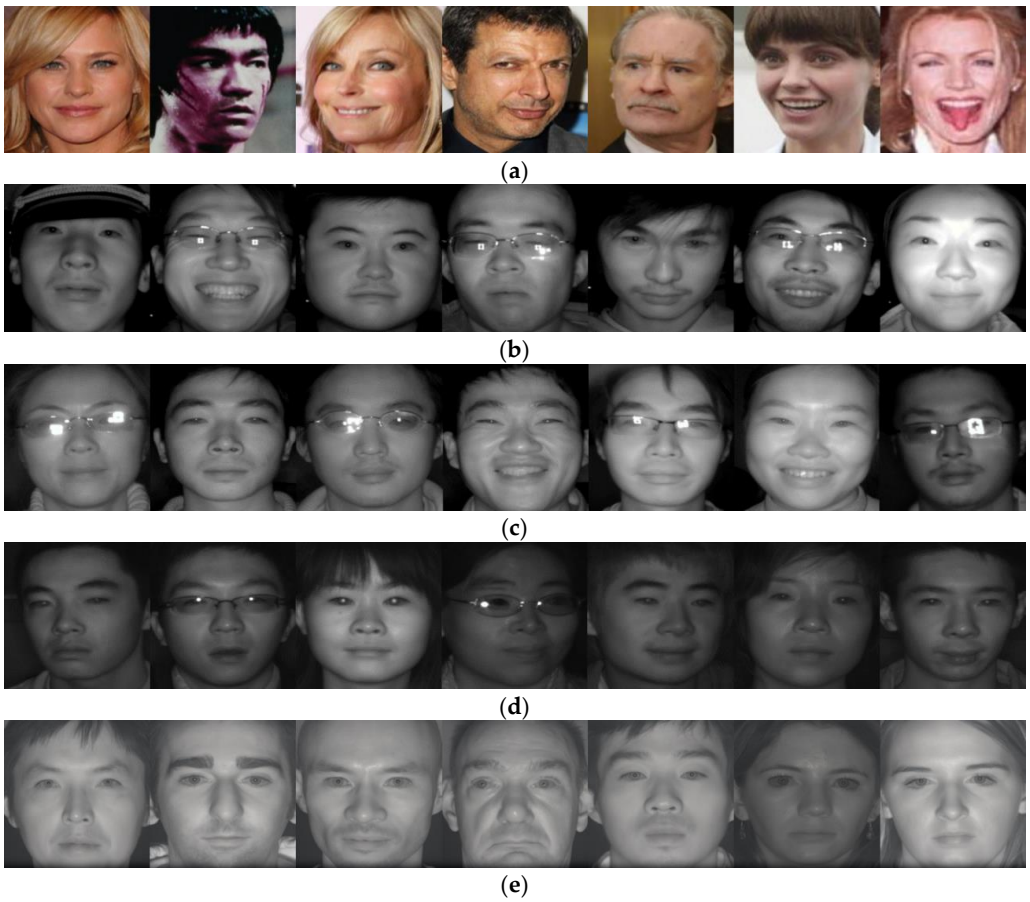


Figure 2. Sample images for each database. (**a**) CASIA web face; (**b**) CASIA NIR face data; (**c**) CASIA NIR-VIS face 2.0; (**d**) PolyU NIR face data; (**e**) ND-NIVL face data.

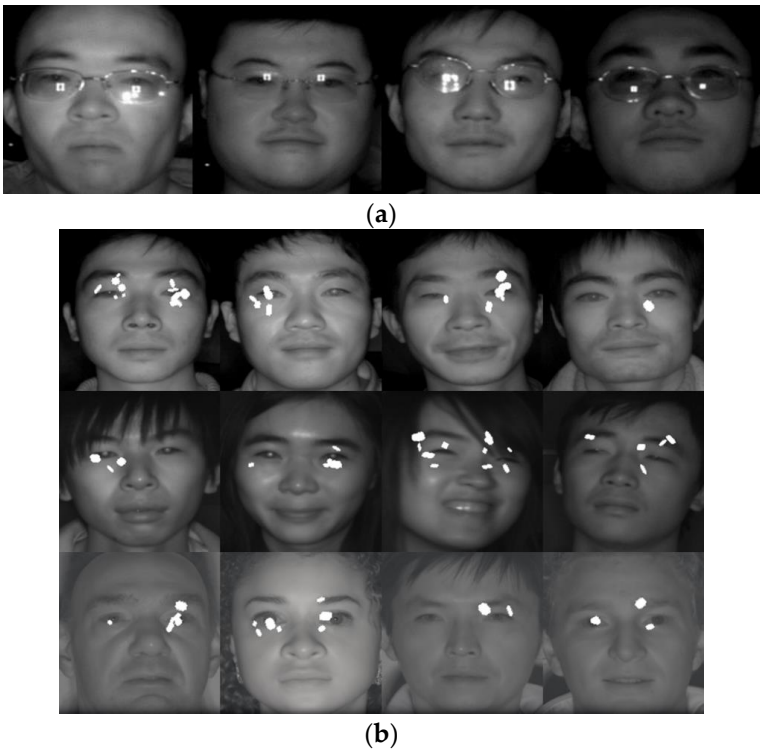


Figure 3. Sample images of (**a**) glasses-wearing faces and (**b**) augmented faces.

The rest of this paper provides a detailed description of the architectures, dataset, data augmentation, experiments, and conclusions.

2. Architectures

FaceNet [2] introduced a unified system for various face recognition applications. This versatility could be achieved because the system utilizes the output of the network not as a label, but rather as a 128-D vector. This is called an embedding and is unique to each individual. By measuring the proximity of two embeddings in Euclidean distance, we can decide whether these embeddings are from the same person or not. We adopted this method of utilizing an embedding for our face verification system. Our system used cross-entropy as the objective function for training, whereas the original FaceNet used a triplet loss. The training process becomes much faster and converges well in this approach if the training data is small [11]. The public databases used in this paper are much smaller (i.e., RGB with approximately 0.5 million and NIR with approximately 30,000 face images) than the private databases used in FaceNet (with over 100 million images).

Two deep-learning architectures were used in this paper: Inception-ResNet v1 [12] and the one used in NIRFaceNet [9]. Inception-ResNet is an outstanding deep-learning architecture that combines Inception [10] and Resnet [13]. This network showed one of the best performances in many image recognition challenges including ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, and many studies have employed this network. Inception-ResNet is based on the Inception network that provides much deeper layers with small-sized (3×3, 1×1) filters, and takes advantage of residual connection from ResNet. The performance of Inception-ResNet has been verified in many ways, and open sources are also provided in various deep-learning frameworks including TensorFlow. For this reason, we employed Inception-ResNet (version v1) for our face verification system.

NIRFaceNet is a deep neural network for NIR face recognition. NIRFaceNet has a compact architecture with less learning time required for recognition with respect to the Institute of Automation, Chinese Academy of Sciences (CASIA) NIR face database. NIRFaceNet claims that medium-sized networks can perform better if the dataset is small, which is the case with the CASIA NIR database. NIRFaceNet is a network that is motivated by a two-stage Convolutional Neural Network (CNN). The first stage is for low-level feature extraction and the second stage is for high-level global feature extraction. While the Inception-ResNet v1 has a total of 20 blocks (Inception-ResNet-A: 5, -B: 10, -C: 5 blocks), the NIRFaceNet architecture has only two blocks. The blocks used in Inception-ResNet and NIRFaceNet have a similar shape, which includes about three branches of two small, subsequent convolutional filters and are concatenated at the end of the block. We were inspired by the idea of using a small-sized network because the public NIR datasets contain small data as well as less-diverse face images compared to RGB datasets (NIR face datasets are composed of illumination-invariant and frontal face images). Thus, we employed NIRFaceNet in the experiments.

Although NIRFaceNet introduced a new compact architecture, their experiments need to be complemented considering real NIR face recognition scenarios. First, they evaluated the recognition rate on users without glasses. However, reflected NIR light on glasses significantly degraded recognition performance [3]. In addition, they trained the network with a small dataset, using only three images per person despite the whole training set having about 20 images per person (using only 705 images from the 3940 images). Reducing the size of the training dataset is undesirable because the amount of training data should be maximized to model the target entirely. Additionally, closed-set recognition was used, so the recognition rate described in the paper does not consider open-set recognition such as a mobile unlocking system.

3. Databases

Five public databases are used in this paper, four of which are NIR face databases and the remaining is an RGB face database. The databases are described in Table 1. Later in this paper, each database is expressed as the given symbol in the table for convenience. Sample images of each database

are shown in Figure 2. The CASIA NIR face data does not need a symbol because this data was only used as the test set in the experiments.

Table 1. Public databases used in this paper.

Database	Number of Images	Number of Identities	Symbol
CASIA-Web face [14]	453,415	10,575	***RGB***
CASIA NIR face data [3]	3940	197	N/A
CASIA NIR-VIS 2.0 Face Database [15]	12,485	725	***NIR1***
PolyU NIR Face Database [16]	24,698	335	***NIR2***
ND-NIVL Database [17]	22,261	655	***NIR3***

Compared to the RGB database, the sizes and variations of the NIR databases are limited. Given that the ***RGB*** database is composed of web-crawled celebrity data, it covers most of the challenges with face recognition, including lighting, age, hairstyle, severe head pose change, and image quality. On the other hand, NIR face images are collected as frontal head poses. Although this pose variation seems to be less general for face recognition in public places, it contains most of the common cases for unlocking smartphones and accessing control systems. This characteristic motivated us to find smaller networks that can have adequate performance.

4. NIR Reflection Augmentation

Practical and reliable face recognition systems must tolerate various changes to users' appearance. Among these facial variations, wearing eyeglasses makes recognition difficult for NIR face images. If people wear glasses, the active NIR light is reflected in the glasses and the reflected light covers the eyes, as shown in Figure 2. According to the literature [4], the area around the eyes is the most discriminative area for differentiating between faces. Thus, traditional face recognition methods which focus on the eye regions are hindered by artifacts covering the eye region.

In this case, removing or dispelling the reflected light can be the solution. Many studies have introduced ways to acquire invariant images using deep learning. However, the eye regions can be filled with incorrect information after removing glints (e.g., filled with skin texture around the eye region, or filled with the another person's eye in the training set). Also, making a simple framework with only one network could have practical applications. Thus, we focused on making NIR face recognition robust against corruption in one of the most informative regions (i.e., around the eye) by adding artificial noises.

Data augmentation is one of the techniques used in deep learning approaches when the size of training datasets is limited. Its purpose is to increase the size of the dataset by transforming the original images via rotation, scaling, cropping, or changing the color characteristics. This technique emulates various changes of the original images relating to geometric and color transformation, and it results in expanding the model of the target [18]. In this paper, we adopted a data augmentation technique to improve the recognition of glasses wearers. Instead of having a tolerance for geometric or color variations in faces, we wanted the trained networks to tolerate glasses-wearing situations. For this, we hoped that the networks would perceive the augmented face images as real glasses-wearing images in the training step.

For the augmentation method, we added artificial NIR-reflected lights onto face images as a simple way to make a virtual image of a face wearing glasses. There were then added back to the dataset. Several ellipses with random sizes and shapes were augmented around the eye region, as shown in Figure 3. To make realistic reflections, we considered the head pose, the lens properties, and active lighting configurations. However, we wanted to verify the feasibility of the augmentation method in advance. Thus, we added randomly generated contamination. From the different perspectives of machine learning, this process can be shown as lowering the importance of a specific region (in this

case, the region around the eyes) during the learning process by adding artificial contamination to this region.

5. Experiments

In this section, we evaluated the recognition performance of the two different architectures and the various training sets in which the data augmentation method was applied. All face images used in this paper were aligned by this method [19] in order to ensure the eyes were located at a similar position in the aligned images.

5.1. Evaluation Methods

Ten-fold cross-validation [2] was used as an evaluation method. The total number of pairs in the evaluation set was 6000 (3000 positive pairs and 3000 negative pairs). Each fold had 600 pairs. The face images in the CASIA NIR face data [3] were used for constructing the evaluation set. We used the validation rate (at false acceptance rate (FAR) = 0.1%) as an evaluation metric.

5.2. Architectures

The validation rate of FaceNet trained with ***RGB*** was 85.4%. The results were similar to the performance described in the original FaceNet paper, although our experiments were conducted for images using a different spectrum (NIR). We considered this result as the baseline and conducted comparative experiments. Figure 4 compares the performance of FaceNet and NIRFaceNet when trained with the same data. As shown in Figure 4, the validation rate became even lower when the NIR databases were used for training FaceNet. Interestingly, NIRFaceNet showed better performance when trained with NIR databases and was even higher than the baseline (FaceNet trained with RGB data). In other words, the recognition rate was increased by using a smaller network when the recognition was performed on NIR images with fewer head poses and facial variations.

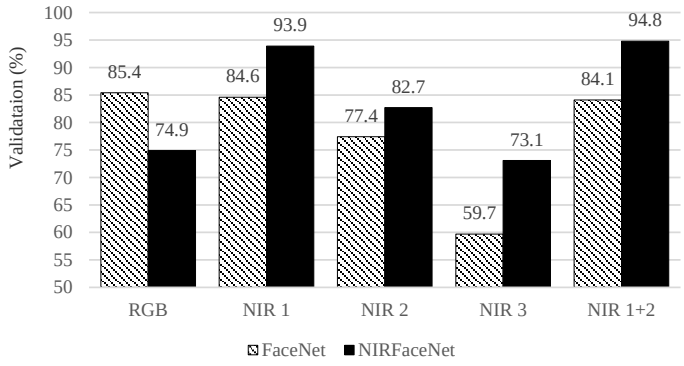


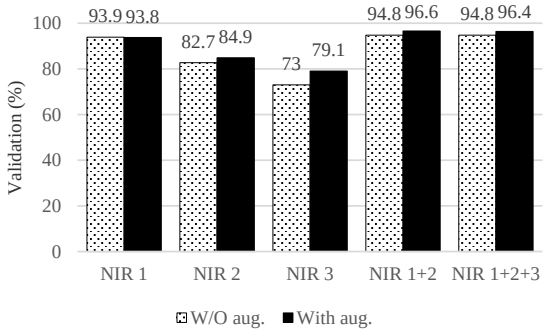
Figure 4. Comparisons between FaceNet and NIRFaceNet.

5.3. NIR Reflection Augmentation

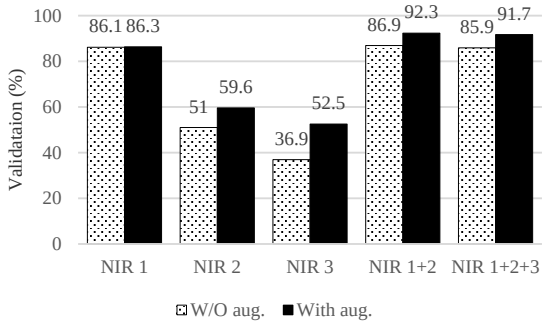
We also calculated the validation rate of glasses pairs—defined as a pair of face images that contain at least one face image that is wearing glasses. There were 1588 glasses pairs in the test set of 6000 pairs. Figure 5 shows the validation rate of all pairs and glasses pairs on the NIRFaceNet trained by each database. Note that the validation rate of the glasses pairs was lower than that of all the pairs because false negatives occurred more frequently for glasses pairs in these experiments. This is because the same person (a positive pair) was more often recognized as a different person (a negative pair) when wearing glasses in the NIR spectrum.

The overall validation rates were increased when the augmented data were used for training the networks. In particular, the performance after data augmentation increased more for the glasses pairs.

This shows that our augmentation method had a positive effect on the recognition of glasses wearers. Furthermore, the combined dataset (***NIR1*** + ***NIR2***) showed the best performance.



(**a**)



(**b**)

Figure 5. The validation rate of NIRFaceNet trained with (**a**) each database on all pairs and (**b**) glasses pairs.

Note that networks trained by the dataset, including ***NIR1***, generally had higher performances. One difference between ***NIR1*** and the others is the existence of glasses pairs. The images of wearing and not wearing glasses are not present for the same person in ***NIR2*** and ***NIR3***. This means the network does not have the opportunity to learn positive glasses pairs. On the other hand, the images in ***NIR1*** contain both types for each person. To demonstrate the effect of data augmentation in ***NIR1***, we trained NIRFaceNet with ***NIR1*** after excluding the glasses-wearing images (results are shown in Figure 6). The validation rate on the glasses pairs decreased to 59.6%. However, the validation rate improved to 76.1% when the network was trained by the training set (***NIR1*** without glasses-wearing images) after applying the augmentation method. From this performance improvement, the model can be seen as having symmetry with the model trained with real glasses-wearing images under the glasses pair situation. In conclusion, this augmentation method is especially useful when the training set does not have positive glasses pairs. Figure 7 shows the true-positive examples after augmentation.

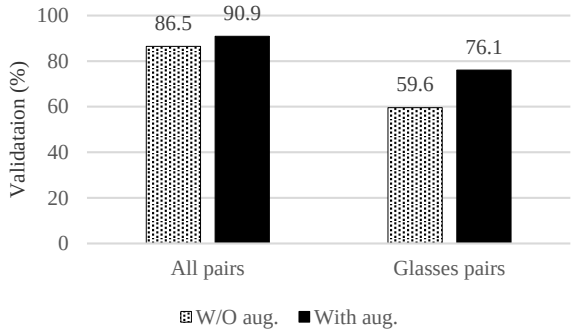


Figure 6. Validation rate of NIRFaceNet trained by ***NIR1*** excluding glasses-wearing images.

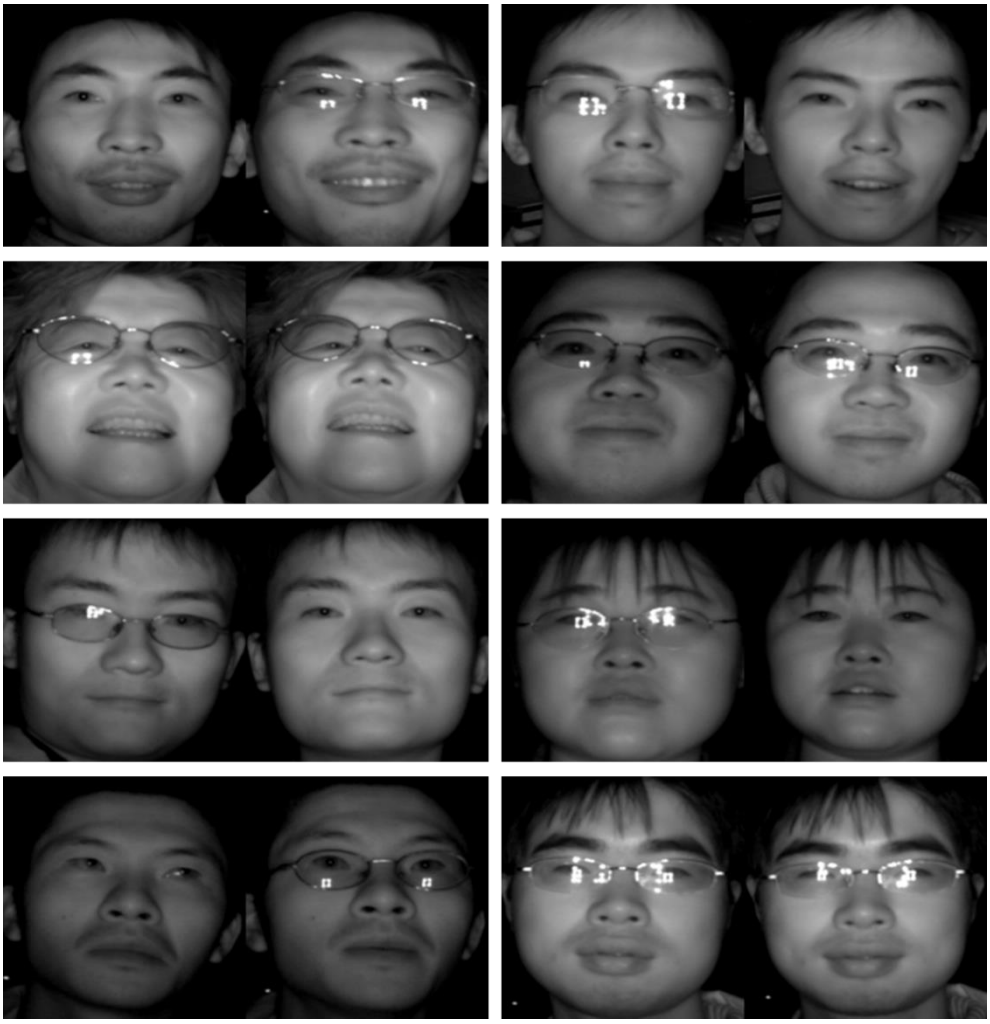


Figure 7. Examples of true-positive glasses pairs after NIR reflection augmentation.

6. Conclusions

In this paper, we applied a deep learning approach to NIR face recognition and studied how to improve the performance of this field compared to FaceNet. Several public NIR face databases were gathered to construct a sufficiently large training set, and the network trained by the integrated dataset showed the best performance. Furthermore, we found that a small architecture could have better performance for NIR face recognition. Our experiments showed that our data augmentation method could improve face recognition for glasses wearers. The method is simple, but it resolves one of the most significant issues regarding NIR face recognition. It also implies that the time and cost to add real glasses-wearing images to a training set can be reduced by applying our augmentation method.

In future studies, we plan to research the augmentation method further. This would allow us to emulate various characteristics of NIR face images in order to make an extensive NIR face training set. We will also evaluate the recognition performance at FAR = 0.001%, which is used in the field of fingerprint recognition.

Author Contributions: Conceptualization, H.J.; methodology, H.J.; software, H.J.; validation, H.J.; formal analysis, H.J.; investigation, H.J.; resources, H.J.; writing—original draft preparation, H.J.; writing—review and editing, H.J. and W.-Y.K.; supervision, W.-Y.K.; project administration, W.-Y.K.

Funding: This research was funded by Samsung Electronics (No. 201900000002726). And the APC was funded by Samsung Electronics' University R&D program.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 1701–1708.
2. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
3. Li, S.Z.; Chu, R.F.; Liao, S.C.; Zhang, L. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 627–639. [CrossRef] [PubMed]
4. Pan, K.; Liao, S.; Zhang, Z.; Li, S.Z.; Zhang, P. Part-based Face Recognition Using Near Infrared Images. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–6.
5. Wang, Z.; Wang, Z.; Zheng, Y.; Chuang, Y.-Y.; Satoh, S. Learning to Reduce Dual-Level Discrepancy for Infrared-Visible Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–21 June 2019.
6. Iranmanesh, S.M.; Dabouei, A.; Kazemi, H.; Nasrabadi, N.M. Deep cross polarimetric thermal-to-visible face recognition. In Proceedings of the 2018 International Conference on Biometrics (ICB 2018), Gold Coast, Australia, 20–23 February 2018; pp. 166–173.
7. He, R.; Cao, J.; Song, L.; Sun, Z.; Tan, T. Cross-spectral Face Completion for NIR-VIS Heterogeneous Face Recognition. *arXiv* **2019**, arXiv:1902.03565.
8. Lezama, J.; Qiu, Q.; Sapiro, G. Not afraid of the dark: NIR-VIS face recognition via cross-spectral hallucination and low-rank embedding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6807–6816.
9. Peng, M.; Wang, C.; Chen, T.; Liu, G. NIRFaceNet: A convolutional neural network for near-infrared face identification. *Information* **2016**, *7*, 61. [CrossRef]
10. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
11. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference 2015, British Machine Vision Association, Swansea, UK, 7–10 September 2015; pp. 41.1–41.12.
12. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Pattern Recognit. Lett.* **2016**, *42*, 11–24.
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; Volume 45, pp. 770–778.
14. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Learning Face Representation from Scratch. *J. Struct. Chem.* **2014**, *53*, 1062–1074.
15. Li, S.Z.; Yi, D.; Lei, Z.; Liao, S. The CASIA NIR-VIS 2.0 Face Database. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 348–353.
16. Zhang, B.; Zhang, L.; Zhang, D.; Shen, L. Directional binary code with application to PolyU near-infrared face database. *Pattern Recognit. Lett.* **2010**, *31*, 2337–2344. [CrossRef]
17. Bernhard, J.; Barr, J.; Bowyer, K.W.; Flynn, P. Near-IR to visible light face matching: Effectiveness of pre-processing options for commercial matchers. In Proceedings of the 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems, BTAS 2015, Arlington, VA, USA, 8–11 September 2015.
18. Perez, L.; Wang, J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv* **2017**, arXiv:1712.04621.
19. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]