이학석사 학위논문

Deep Convolutional Neural Network for Fingerprint Image Denoising and Inpainting

지문 영상 잡음 제거 및 복원을 위한 심층 합성곱 신경망

2021년 2월

서울대학교 대학원
협동과정 계산과학전공
배 정 윤

이학석사 학위논문

Deep Convolutional Neural Network for Fingerprint Image Denoising and Inpainting

지문 영상 잡음 제거 및 복원을 위한 심층 합성곱 신경망

2021년 2월

서울대학교 대학원
협동과정 계산과학전공
배 정 윤

Deep Convolutional Neural Network for Fingerprint Image Denoising and Inpainting

지문 영상 잡음 제거 및 복원을 위한 심층 합성곱 신경망

지도교수 강 명 주

이 논문을 이학석사 학위논문으로 제출함

2020년 10월

서울대학교 대학원
협동과정 계산과학전공
배 정 윤

배정윤의 이학석사 학위논문을 인준함

2020년 11월

위 원 장: 국 응 (인)
부위원장: 강 명 주 (인)
위 원: 박 지 훈 (인)

Abstract

Biometric authentication using fingerprints requires a method for image denoising and inpainting to extract fingerprints from degraded fingerprint images. A few deep learning models for fingerprint image denoising and inpainting were proposed in *ChaLearn LAP Inpainting Competition - Track 3, ECCV 2018*. In this thesis, a new deep learning model for fingerprint image denoising is proposed. The proposed model is adapted from FusionNet, which is a convolutional neural network based deep learning model for image segmentation. The performance of the proposed model was demonstrated using the dataset from the ECCV 2018 ChaLearn Competition. It was shown that the proposed model obtains better results compared with the models that achieved high performances in the competition.

Key words: Deep learning, Fingerprint image, Image denoising, Convolutional neural network, FusionNet
Student Number: 2018-28868

Contents

Abstract (In Korean)

Chapter 1

Introduction

Fingerprint is an impression of friction ridges of a finger left on a surface. Because of its uniqueness and durability over the life of an individual, fingerprint authentication methods are used in various real-life devices. In recent years, biometric authentication has been used as a major authentication technique for smart devices. Biometric authentication utilizes personal biometric information to significantly reduce leakage or theft, compared with traditional authentication methods such as personal identification number (PIN) or password[1]. In biometric identification, the matching of two fingerprints is one of the most widely used and reliable technique. Automated fingerprint identification system (AFIS) is a biometric authentication technology that has recently attracted much attention, because of its convenience in using simple touch-based sensors[2].

In fingerprint recognition, various kinds of environmental noise are generated when a user's finger touches the sensor in some undesirable conditions, which can severely damage the fingerprint image. This degrades the quality of the image, which makes user authentication difficult. Therefore, in order to solve this problem, image restoration studies are preferentially necessary.

Several studies have been conducted on noise reduction, enhancement, and reconstruction of fingerprint images based on existing algorithms. Traditional fingerprint image processing methods include image filtering or the use of partial differential equations. Traditional image filtering methods using a Wiener filter, an anisotropic filter[3] and a directional median filter[4] were proposed. A method based on partial differential equations was also proposed for automatic fingerprint restoration[5]. Several methods using orientation information have been proposed. Hong *et al.* used local ridge orientation and frequency information to improve the clarity of ridge and valley structures of a fingerprint image[6]. The method used Gabor filters as bandpass filters to remove noise and enhance the ridge and valley structures. Chikkerur *et al.* improved low-quality fingerprint images by estimating intrinsic properties of the fingerprints

including local ridge orientation with an algorithm based on Short-time Fourier Transform (STFT)[7]. Feng *et al.* proposed an orientation field estimation algorithm that uses the prior information embedded in the fingerprint structure[8]. Cappelli *et al.* took an approach to reconstruct fingerprint images from standard templates by extracting the information of the orientation field and ridge structure from a given fingerprint image[9].

Recently, image processing using a convolutional neural network (CNN) has attracted much attention. CNN models are able to extract hierarchical image features from input data, which showed great performances in image recognition. Using convolutional neural networks, many developments have been made in the areas of image denoising, image segmentation and image enhancement.

In this thesis, a new CNN-based model for fingerprint image denoising and inpainting is proposed, which improves artificially degraded fingerprint images with random transformations. Several experiments were conducted to verify the performance of the proposed model. It was confirmed that the proposed model shows better results compared with the previously proposed CNN models.

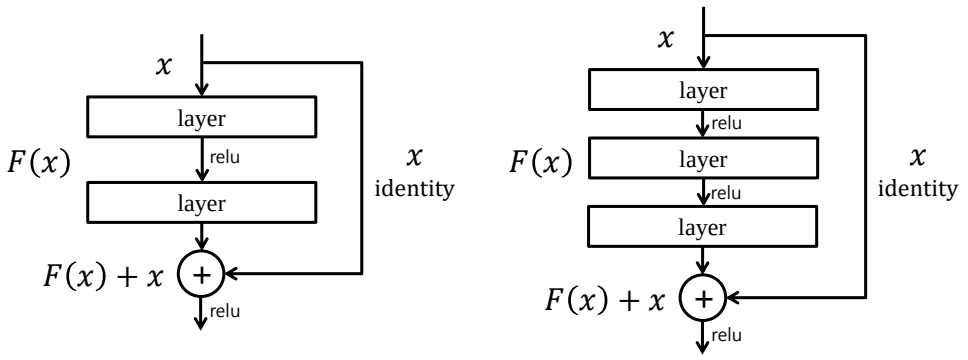
The remainder of the paper is organized as follows: In Section 2, a brief introduction to the neural networks that are related to the study is given. Also, recent research related to the topic is described. In Section 3, a new CNN model for fingerprint image denoising and inpainting is proposed. The feasibility of the proposed model is verified through the experiments in Section 4. The conclusions are presented in Section 5.

Chapter 2

Related Work

2.1 Residual Neural Network

Residual Neural Network (ResNet)[10] is one of the best known deep neural networks for image recognition that showed great performance in ILSVRC 2015 challenge. The model is composed of a stack of residual blocks, where each residual block contains a skip connection between few layers. The skip connection adds the input to the output in each residual block as shown in Figure 2.1. The input x is added to the output after the layers, so that the output of the residual block before the ReLU activation function is $H(x) = F(x) + x$. The network is called residual neural network since the network learns the residual mapping $F(x) = H(x) - x$. The skip connection allows gradient information to pass through the layers, which makes it possible to build a much deeper network without vanishing gradient problem. Residual blocks and skip connections are widely adopted in many deep neural networks.



(a) A residual block with a stack of 2 layers

(b) A residual block with a stack of 3 layers

Figure 2.1: Examples of the residual blocks.

2.2 Convolutional Neural Networks for Semantic Segmentation

2.2.1 U-Net

Ronneberger *et al.* proposed U-Net[11] for image segmentation in 2015. The model is an end-to-end deep learning model for semantic image segmentation which stems from the fully convolutional network (FCN)[12]. The U-Net consists of a contraction path (also known as an encoder) and an expanding path (also known as a decoder). The encoder includes a stack of convolution and maxpooling layers to capture the context of an image, while the decoder is symmetric to the encoder to propagate context information of the image to higher resolution layers. The decoder enable precise localization using transposed convolutions. Long skip connections are used in the decoder to combine the feature information with the spatial information. In the skip connections, the feature maps of the decoder are concatenated with the feature maps from the same level of the encoder. Figure 2.2 shows the architecture of U-Net where the green, the blue, the red and the yellow boxes in the figure represent the convolution layer, 2×2 maxpooling, the deconvolution layer and the concatenation, respectively.

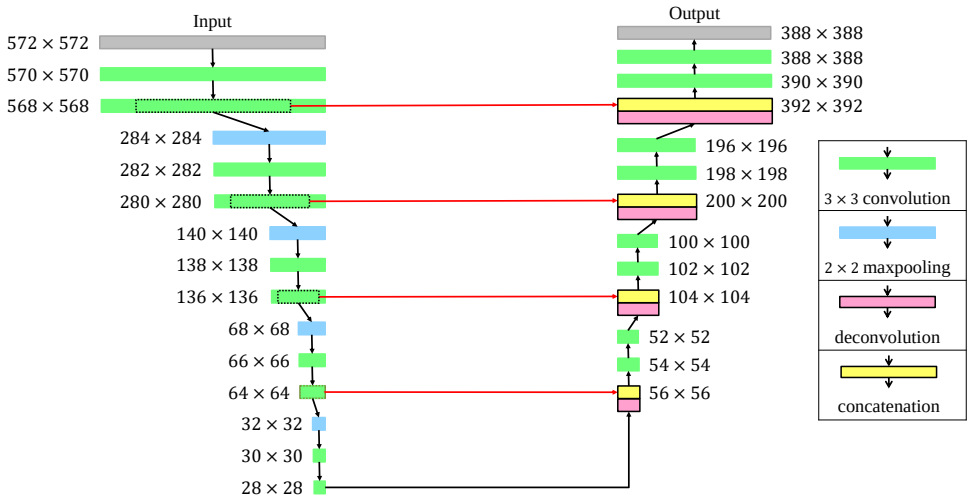


Figure 2.2: U-Net Architecture.

2.2.2 FusionNet

FusionNet[13] is a model for image segmentation that improves U-Net by increasing the depth of a network using residual blocks and summation-based skip connections. Like U-Net, the network consists of an encoder, a decoder and long skip connections. FusionNet utilizes residual blocks in the encoder and the decoder parts in order to build a deeper network. A residual block in the model consists of three convolutional blocks and a short skip connection. Every residual block is surrounded by two convolution layers before and after the block. Also, a pixel-wise addition is used in the long skip connections in order to make the network fully residual network. Therefore, the network is nested with short and long skip connections to pass information through the network. A part of FusionNet architecture is shown in Figure 2.3.

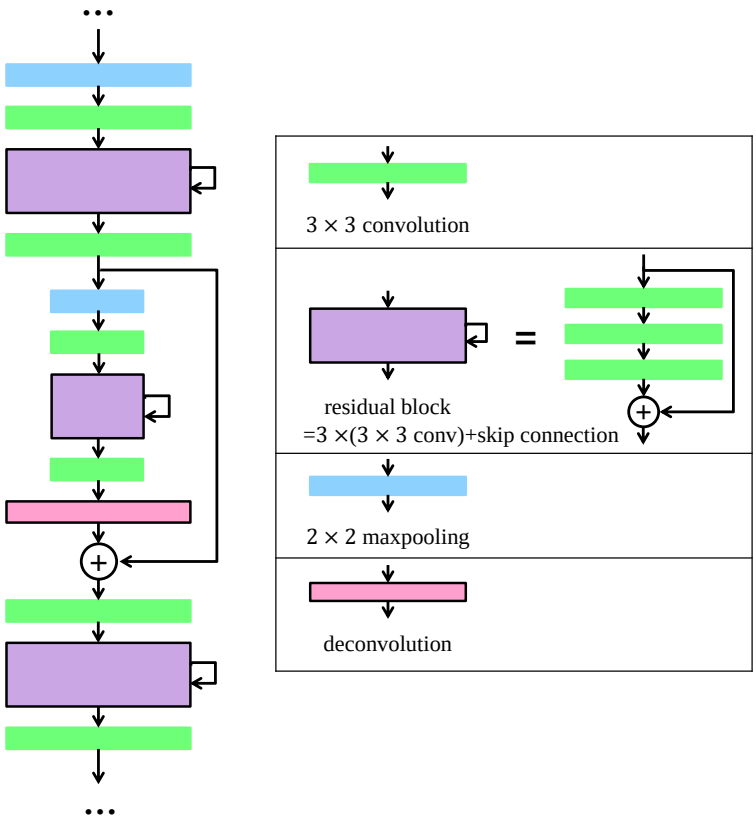


Figure 2.3: A part of FusionNet architecture

2.3 Recent Trends in Fingerprint Image Denoising

Several convolutional neural network based models have been proposed for fingerprint image denoising and enhancement. FingerNet was proposed by Tang *et al.* which is a CNN-based model to extract latent fingerprint minutiae from noisy ridge patterns and complex backgrounds[14]. The method extracted fingerprint details by integrating orientation estimation, segmentation, Gabor enhancement and extraction as a convolutional network. Also, Li *et al.* improved FingerNet to propose a latent fingerprint enhancement model[15]. Nguyen *et al.* proposed MinutiaeNet that performs automatic minutiae extraction using a coarse and fine network[16]. The coarse network improved the image based on domain knowledge and extracted a segmentation map to provide candidate minutiae locations. The fine network used a technique to refine these candidate minutiae locations and approximate minutiae orientation. Also, Svoboda *et al.* proposed a model that uses generative neural network to reconstruct latent fingerprint images that predicts the missing parts of the ridge pattern[17].

Based on these studies, ChaLearn LAP Inpainting Competition Track 3 for fingerprint image denoising and inpainting[1] was held. The ECCV 2018 Chalearn competition motivated researchers to develop deep learning models to restore degraded fingerprint images. Several CNN-based architectures were proposed in the competition. U-Finger[18], which adopted the U-shaped deep denoiser[19] to perform fingerprint image denoising, achieved good results in the competition. The overall architecture of U-Finger includes feature encoding and feature decoding that consist of convolution layers and residual blocks. In addition, FPD-M-Net[20] was proposed in the competition. The FPD-M-Net architecture was adapted from M-Net[21], where it modified the block arrangement order and loss function of M-Net. M-Net is a model for 3D brain structure segmentation and utilizes U-Net[11] to derive better segmentation results. The method approached fingerprint image denoising and inpainting with a deep neural network for image segmentation. Image segmentation was applied to the degraded fingerprint image to segment fingerprint ridges from the noisy background, while the proper training of the deep neural networks solved image inpainting problem.

Overall, these related studies inspired us to propose a new convolutional neural network for fingerprint image denoising and inpainting.

[1] http://chalearnlap.cvc.uab.es/challenge/26/track/32/description/

Chapter 3

Proposed Model

3.1 Model Architecture

In this section, a new model for fingerprint image denoising and inpainting is described. The proposed model is adapted from FusionNet[13], which utilizes U-shaped structure for image segmentation. The proposed model is based on the architecture of a convolutional autoencoder, which consists of an encoder, a symmetric decoder and long skip connections. The residual block which consists of three convolution layers and a residual skip connection is used in the proposed model as in the FusionNet model.

There are some differences between FusionNet and the proposed model. First, the size of the input and the output of the proposed model is set to 256×384 in order to match the size similar to the size of the image from the dataset used in the experiments. Therefore the number of downsampling and upsampling of the proposed network were reduced by one compared with FusionNet. Next, concatenations of feature maps were used in the skip connections of the proposed model whereas summation-based skip connections were used in FusionNet. Also, the number of channels (depth) is reduced in the proposed model compared with FusionNet. Subsection 4.4.1 provides an explanation for the above modifications.

The overall architecture of the proposed model is shown in Figure 3.1. There are five types of basic components that are used to build the proposed network: a convolution layer with filter size 3×3, a deconvolution layer with filter size 2×2, a residual block, a concatenation, and a 2×2 maxpooling. The green, purple, blue, red and yellow boxes in the figure represent the convolution layer, the residual block, the 2×2 maxpooling, the deconvolution layer and the concatenation, respectively. The batch normalization and ReLU are always performed after the convolution blocks. Table 3.1 shows the detail description of the proposed network.

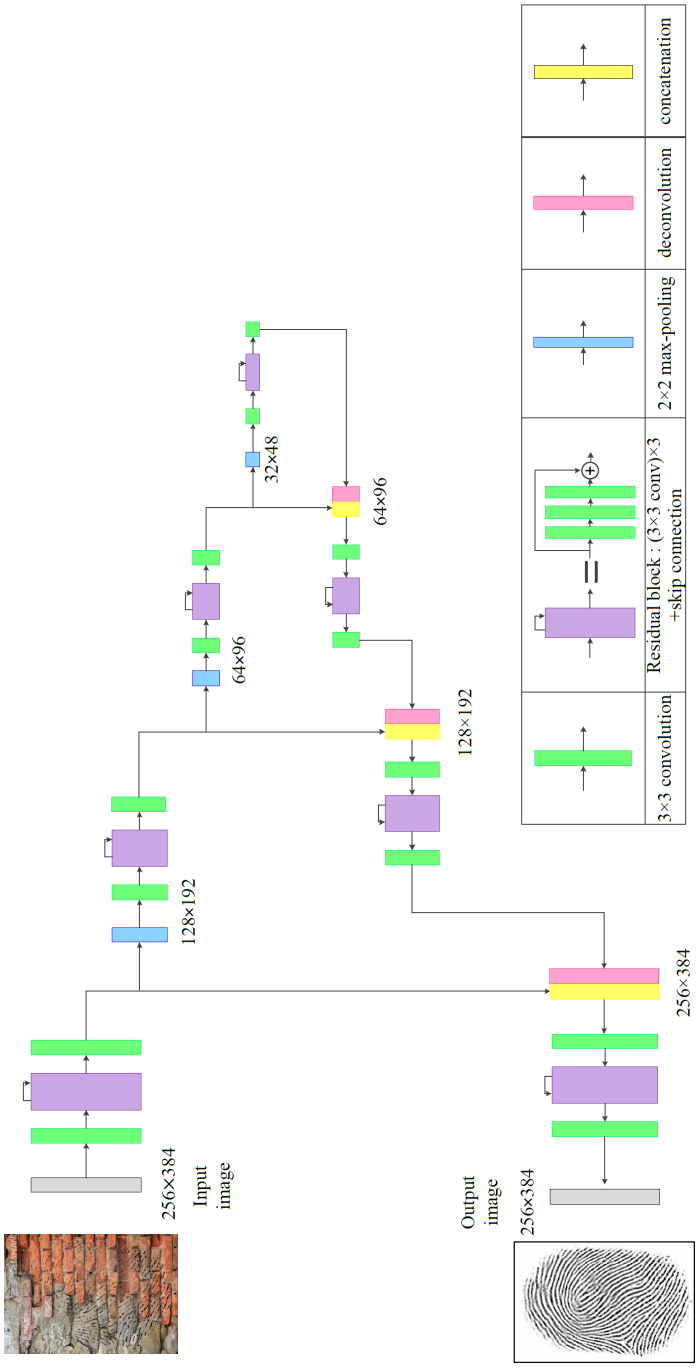


Figure 3.1: Proposed network architecture.

Table 3.1: Summary of proposed network architecture

Block type	Ingredients	Size of feature map
Inputs		$256 \times 384 \times 3$
Downscaling1	conv+res+conv	$256 \times 384 \times 16$
	+maxpool(2×2)	$128 \times 192 \times 16$
Downscaling2	conv+res+conv	$128 \times 192 \times 32$
	+maxpool(2×2)	$64 \times 96 \times 32$
Downscaling3	conv+res+conv	$64 \times 96 \times 64$
	+maxpool(2×2)	$32 \times 48 \times 64$
Bridge	conv+res+conv	$32 \times 48 \times 128$
Upscaling3	deconv	$64 \times 96 \times 128$
	+concat	$64 \times 96 \times 192$
	+conv+res+conv	$64 \times 96 \times 64$
Upscaling2	deconv	$128 \times 192 \times 64$
	+concat	$128 \times 192 \times 96$
	+conv+res+conv	$128 \times 192 \times 32$
Upscaling1	deconv	$256 \times 384 \times 32$
	+concat	$256 \times 384 \times 48$
	+conv+res	$256 \times 384 \times 16$
Output	conv and normalization	$256 \times 384 \times 1$

3.2 Architecture Detail

3.2.1 Residual Block

The proposed network utilizes residual blocks in the encoder and the decoder to build a deeper network. Each residual block consists of three convolution blocks and a residual skip connection as shown in Figure 3.2. The purple block in the figure represent the residual block, and the green blocks represent the convolution layers with filter size 3×3.

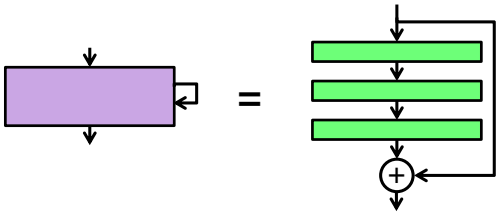


Figure 3.2: A residual block in the proposed model.

3.2.2 Encoder

The encoder consists of three downscaling blocks. Each downscaling block consists of a residual block, two convolution blocks that surround the residual block and a 2×2 maxpooling. For each downscaling block, a feature map is sequentially passed through a convolution block, a residual block, and a convolution block. The resulting feature map is then used as an input for the 2×2 maxpooling and the long skip connection. The feature map is downsampled using stride-2 2×2 maxpooling. Figure 3.3 depicts a downscaling block in the encoder. In the figure, the green blocks represent the convolution layers, the blue block represents the 2×2 maxpooling and the purple block represents the residual block. The red arrow represents the long skip connection to the decoder.

In the proposed model, a degraded fingerprint image of size 275×400 is resized to an image of size 256×384 to be used as an input. The resized image is then passed through three downscaling blocks of the encoder. The feature map of size $32 \times 48 \times 64$ is the output of the encoder to be passed to the bridge.

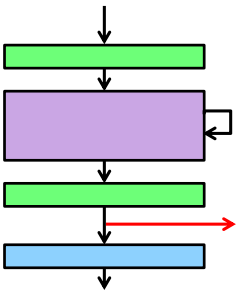


Figure 3.3: A downscaling block in the encoder.

3.2.3 Bridge

In the bridge, the feature map of size $32 \times 48 \times 64$ is sequentially passed through a convolution block, a residual block and a convolution block to output a feature map of size $32 \times 48 \times 128$. The bridge part is depicted in Figure 3.4. In the figure, the green blocks represent the convolution layers, and the purple block represents the residual block.

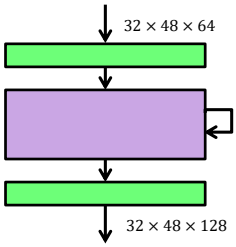


Figure 3.4: Bridge in the proposed model.

3.2.4 Decoder

The decoder consists of three upscaling blocks. Each upscaling block consists of a deconvolution block, a concatenation, a residual block and two convolution blocks that surrounds the residual block. For each block, the feature map is passed through a deconvolution block to expand the size of the feature map. Next, the feature map and the corresponding feature map of the encoder is concatenated using a long skip connection. Then the resulting feature map is passed through a convolution block, a residual block and a convolution block, sequentially. Figure 3.5 depicts an upscaling block in the decoder. In the figure, the red block represents the deconvolution block with filter size 2×2, the green blocks represent the convolution layers, the purple block represents the residual block and the yellow block represents the concatenation. The red arrow represents the long skip connection from the encoder.

After the last upscaling block, the feature map is passed through the last convolution block with an output depth of 1. Then the feature map of size $256 \times 384 \times 1$ is normalized to obtain a grayscale image. The normalization is done by setting the minimum value of the feature map to 0 and the maximum value to 1. The size of the resulting feature map from the decoder is $256 \times 384 \times 1$.

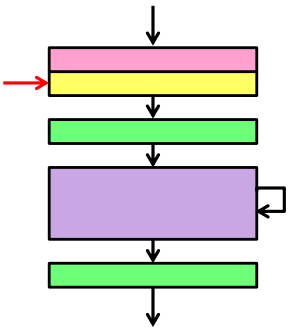


Figure 3.5: An upscaling block in the decoder.

3.3 Loss Function

The loss function used in the training is the loss function presented by Zhao *et al.*[22], which was proposed for neural networks for image restoration. The loss function is defined as in Equation 3.1.

$$L = \alpha L_{ms-ssim} + (1 - \alpha) L_1 \tag{3.1}$$

The goal of the loss function is to output an image that is appealing to a human observer. Experimental results showed that L_1 loss preserves colors and luminance while the MS-SSIM loss better preserves the contrast in high-frequency regions[22]. To capture the best characteristics of both loss functions, the loss function utilizes both the L_1 loss and the MS-SSIM using a weighted sum. α was set to 0.85 in this study.

Chapter 4

Experiments

4.1 Experimental Setup

Experiments were performed to verify the proposed architecture. Table 4.1 shows the experimental environment used in this study.

Table 4.1: Experimental environment

Hardware Specification	
CPU	Intel Core i5-6500
GPU	Tesla K 80
Memory Capacity	16GB
Software Specification	
Python	Version : 2.7.12
TensorFlow	Version : 1.12.0
Operation System	Linux Ubuntu 16.04

The proposed network was trained for 50 epochs with the batch size of 32. In training the network, Adam optimizer was used for the training. The learning rate schedule was a step decay of 10^{-4} in epoch 1, 10^{-5} from epochs 2 to 3, 5×10^{-6} from epochs 4 to 9, 10^{-6} from epochs 10 to 19, 5×10^{-7} from epochs 20 to 39, and 10^{-7} from epochs 40 to 50.

4.2 Evaluation Metrics

We used peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as the evaluation metrics to verify the similarity between the ground-truth image and the predicted image. PSNR is defined as in Equation 4.1 and it indicates the ratio of the power of corrupting noise to the maximum possible power of a signal. SSIM is the measure of similarity between the distorted image and the original image, where the distortions are caused by compression and transformation. SSIM is not a numerical error, but it is more correlated to assessing the similarity of the image qualities perceived by humans[23]. SSIM is defined as in Equation 4.2.

$$\begin{aligned} \text{PSNR} &= 10 \cdot \log_{10}\Big(\frac{MAX_I^2}{MSE}\Big) \\ &= 20 \cdot \log_{10}\Big(\frac{MAX_I}{\sqrt{MSE}}\Big) \end{aligned} \tag{4.1}$$

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4.2}$$

4.3 Dataset

The dataset used in this study is a large-scale synthesized dataset of fingerprint images released in the Chalearn LAP Inpainting Competition Track 3. The dataset consists of pairs of ground-truth fingerprint images and degraded fingerprint images, where the ground-truth images are grayscale images of size 275×400 and the degraded images are RGB images of size 275×400. There are training, validation, and test sets in the dataset. Table 4.2 shows the number of image pairs in each category.

Table 4.2: Number of images in the dataset

Dataset	Training	Validation	Test
Number of image pairs	75,600	8,400	8,400

The ground-truth images in the dataset are synthetic data generated by the Anguli Synthetic Fingerprint Generator. Synthetic data are used not only because collecting a large amount of real fingerprint images takes considerable time and effort, but also due to the privacy legislation in some countries where the distribution of such personal information is prohibited[24, 25].

The degraded fingerprint images in the dataset were created by adding some random artifacts and background to the ground-truth images. Such random artifacts include rotation, blur, contrast, brightness, occlusion, scratch, resolution and elastic transformation. Figure 4.1 shows the examples of the dataset. In the figure, the first and the third rows show the degraded fingerprint images of the dataset, while the second and the fourth rows show the corresponding ground-truth images of the dataset.

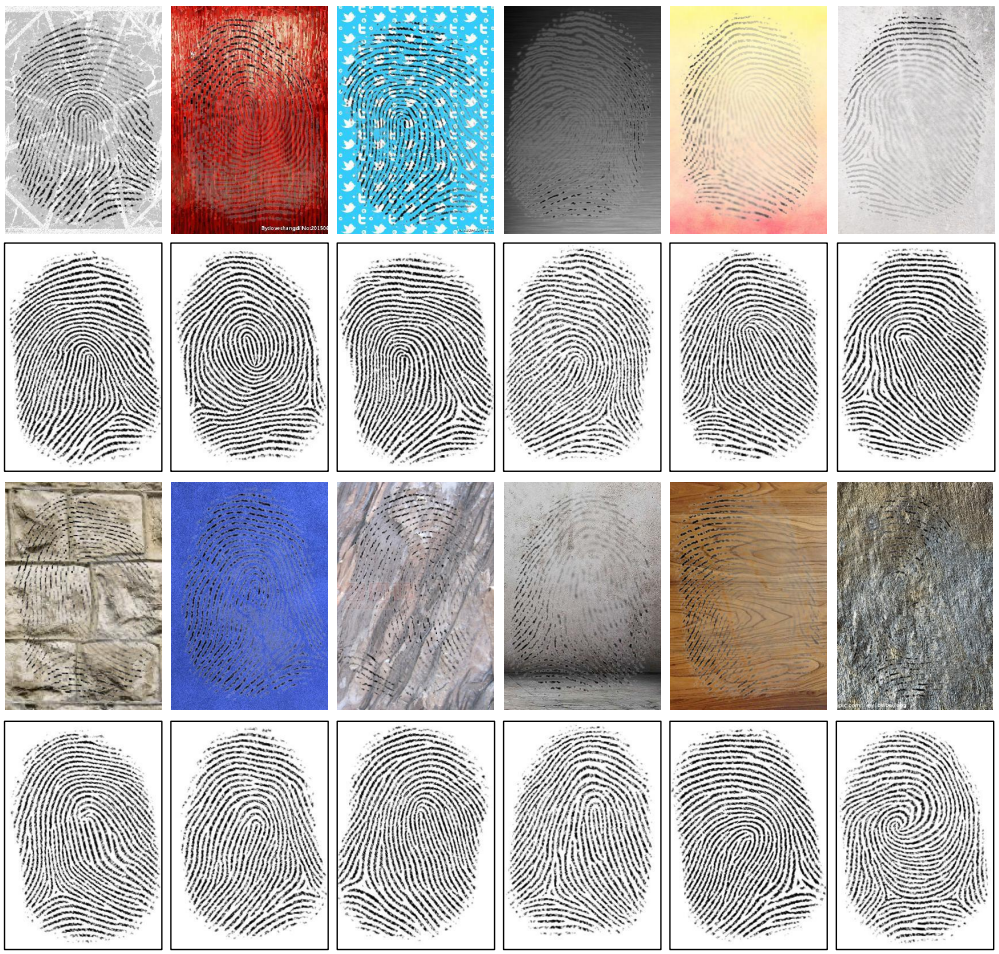


Figure 4.1: The examples of image pairs from the dataset.

4.4 Experimental Results

4.4.1 Ablation Study

As the proposed model was created based on FusionNet, the performance of the proposed model was compared with those of other models that modified FusionNet. Two different modifications of FusionNet were trained for comparison with the proposed model.

First, a modified FusionNet model with a decreased number of feature maps was trained with the original input size, 640×640. The number of feature maps for the modified model was set the same as for the proposed model for fairness in the number of training parameters. Also, a model of input size 256×384 that used addition instead of concatenation in skip connections was trained. In order to perform the addition in the skip connections, the depths of the feature maps of the decoder were adjusted in the deconvolution blocks to match those of the encoder.

For a more detailed description of the modified FusionNet models used in the study, refer to Table 4.3.

Table 4.3: Description of modified models used in the ablation study

	Modified(1)	Modified(2)	Proposed
Input size	$640 \times 640 \times 3$	$256 \times 384 \times 3$	$256 \times 384 \times 3$
Output size	$640 \times 640 \times 1$	$256 \times 384 \times 1$	$256 \times 384 \times 1$
Number of downscaling/upscaling	4	3	3
Channels in the downscaling block	16, 32, 64, 128	16, 32, 64	16, 32, 64
Merge in skip connection	addition	addition	concatenation

The experimental results are shown in Table 4.4 and Table 4.5. The tables show that the method proposed in Section 3 provides the best results among the approaches in the modified models.

Table 4.4: Experimental results on the validation dataset

Method	PSNR	SSIM
Modified(1)	17.0090	0.8037
Modified(2)	17.1688	0.8472
Proposed	**17.1775**	**0.8480**

Table 4.5: Experimental results on the test dataset

Method	PSNR	SSIM
Modified(1)	17.0548	0.8044
Modified(2)	17.1858	0.8471
Proposed	**17.1904**	**0.8478**

4.4.2 Comparison with Other Models

U-Finger[18] and FPD-M-Net[20] were selected for comparison with the proposed model. The models are two CNN-based models that showed great performances in the ECCV 2018 challenge. The proposed model, U-Finger and FPD-M-Net all have u-shaped architecture with encoder, decoder and long skip connections. FPD-M-Net uses the same loss function that is used in the proposed model while U-Finger uses mean-square-error (MSE) for the loss function. Also, U-Finger and the proposed model both uses residual blocks, while FPD-M-Net does not. Unlike the other two models, FPD-M-Net includes two side paths called a left leg and a right leg which make input-to-encoder and decoder-to-output skip connections.

Table 4.6: Quantitative results for the validation dataset

Method	PSNR	SSIM
Base-model	16.4782	0.7889
FPD-M-Net	16.5149	0.8265
U-Finger	16.8623	0.8040
Proposed	**17.1775**	**0.8480**

Table 4.7: Quantitative results for the test dataset

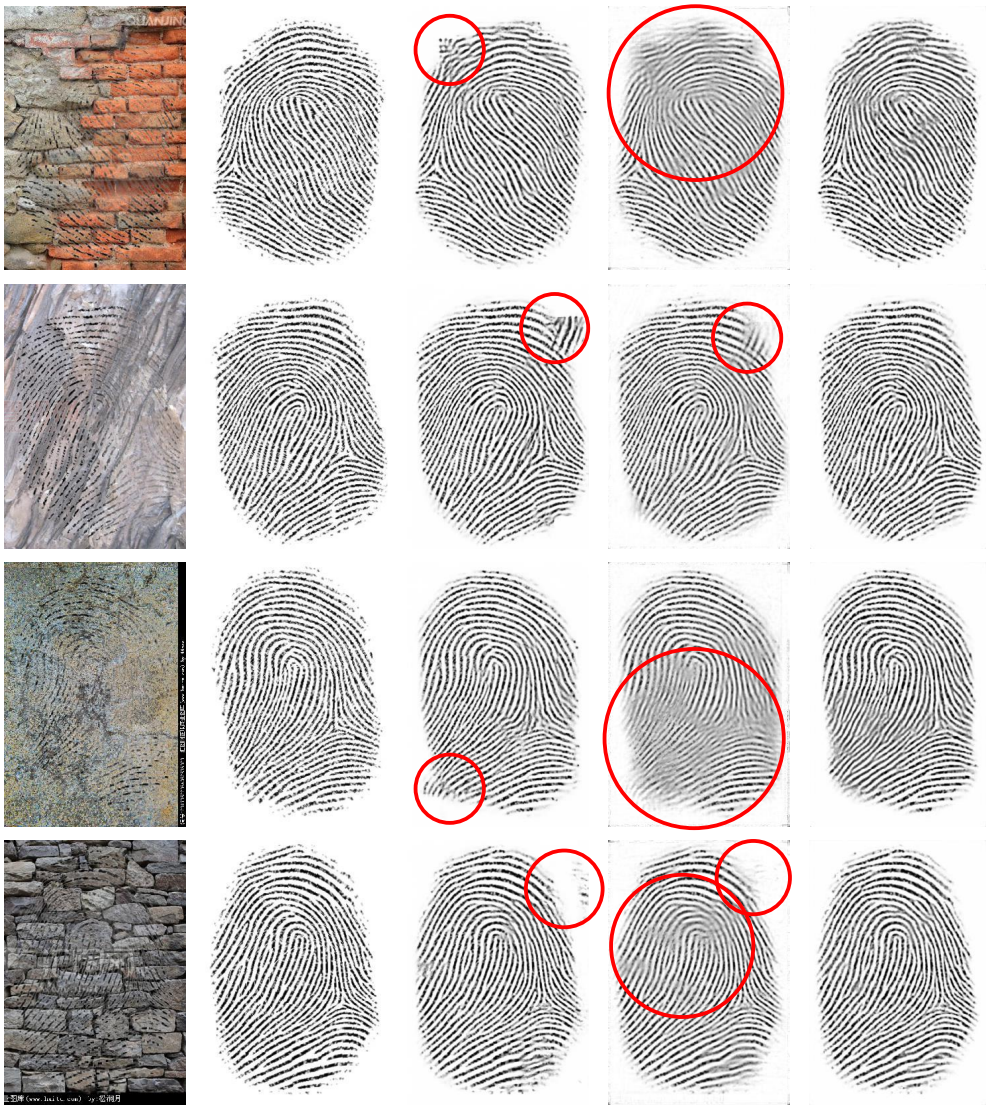
Method	PSNR	SSIM
Base-model	16.4109	0.7965
FPD-M-Net	16.5534	0.8261
U-Finger	16.9688	0.8093
Proposed	**17.1904**	**0.8478**

Table 4.6 and Table 4.7 show the average values of PSNR and SSIM of the models using the validation and the test dataset. The results from the baseline network provided in the competition were attached for comparison[1]. The tables indicate that the proposed method provides the best results in both PSNR and SSIM among all the approaches in the experiment.

The qualitative results from the validation dataset are shown in Figure 4.2. From left to right, the input and the ground-truth images from the validation dataset, the output images of FPD-M-Net, U-Finger and the proposed model are shown. The red circles in the figure indicate the areas that are inaccurately predicted in the output images. As indicated by the red circles in the first and second rows, FPD-M-Net and U-Finger incorrectly predict the fingerprints due to the background noise of the input images. In addition, the large red circles in the first, third and fourth rows of column (d) indicate that the output images of U-Finger often include the blurry parts. Also, the small red circles in the third and fourth rows of column (c) and (d) show that FPD-M-Net and U-Finger inaccurately predict the outer parts of the images as fingerprints. However, the proposed model predicted the fingerprints well in the corresponding areas of the images.

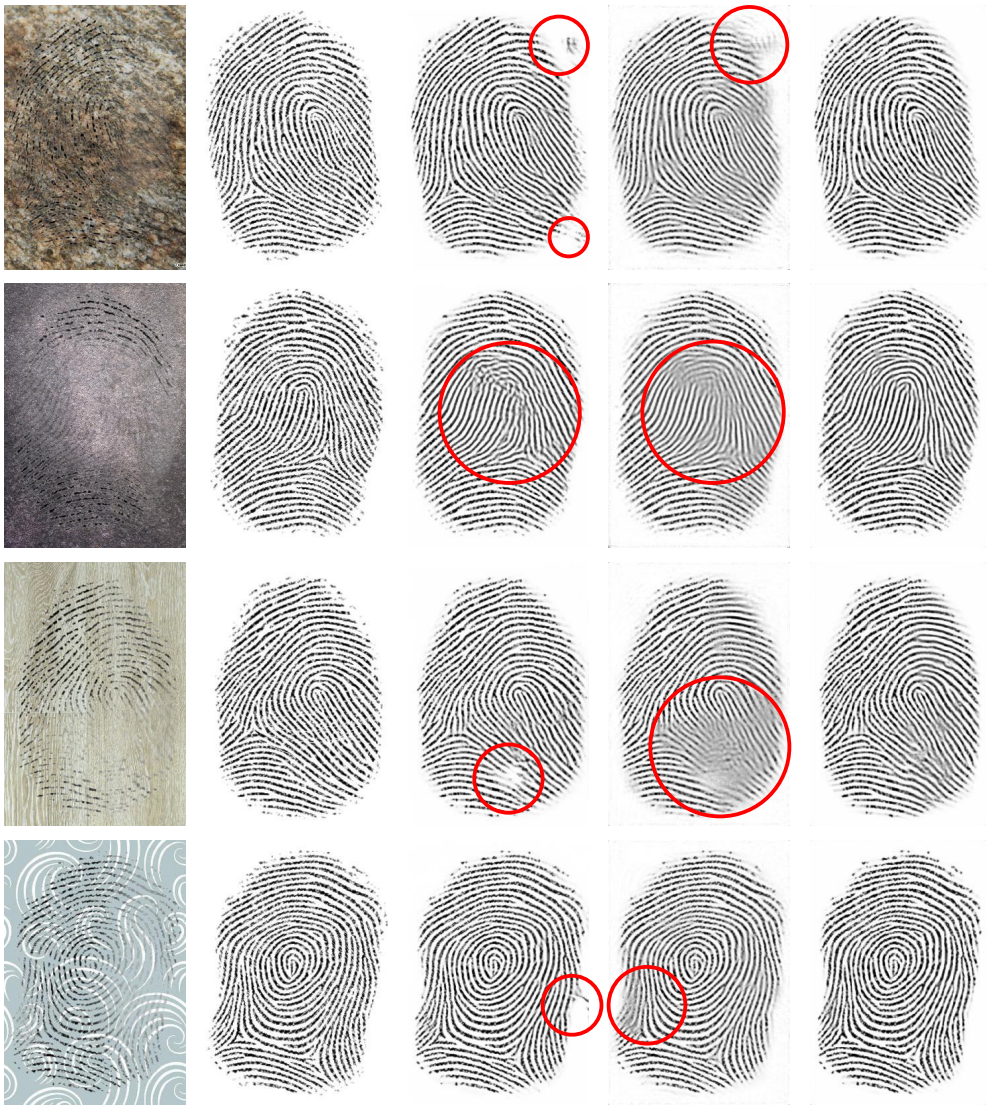
The qualitative results from the test dataset are shown in Figure 4.3. From left to right, the input and the ground-truth images from the test dataset, the output images of FPD-M-Net, U-Finger and the proposed model are shown. As indicated by the red circles in the the first row of column (c) and (d) and the fourth row of column (c), FPD-M-Net and U-Finger inaccurately predict the outer parts of the images as fingerprints. Also, the red circles in the second, third and fourth rows of column (d) indicate that some of U-Finger's output images include blurry areas. In addition, the red circles in the second and third rows of column (c) indicate that FPD-M-Net incorrectly predicts the fingerprints due to the non-uniform parts of the input fingerprint images. However, the proposed network predicts the fingerprints well in the corresponding areas.

[1] http://chalearnlap.cvc.uab.es/challenge/26/track/32/baseline/



(a) Input (b) GT (c) FPD-M-Net (d) U-Finger (e) Proposed

Figure 4.2: Qualitative results from the validation dataset.



(a) Input (b) GT (c) FPD-M-Net (d) U-Finger (e) Proposed

Figure 4.3: Qualitative results from the test dataset.

Chapter 5

Conclusion

In this thesis, a new CNN-based deep learning model for fingerprint image denoising and inpainting was proposed. The proposed model is based on FusionNet, which is a CNN-based image segmentation model with a very good performance. A few modifications were made to the FusionNet model to build a powerful fingerprint denoising and inpainting model. To verify the performance of the proposed model, the experimental results of the proposed model were compared with those of the existing CNN models for fingerprint denoising and inpainting. It was shown that the proposed model outperforms other models that achieved good results in the ECCV 2018 workshop challenge. The results are expected to be applied to real-life devices such as fingerprint sensors.

Bibliography

[1] Hoyeon Lee and Taekyoung Kwon. Fingerprint smudge attacks based on fingerprint image reconstruction on smart devices. *Journal of the Korea Institute of Information Security & Cryptology*, 27(2):233–240, 2017.

[2] Anil K Jain, Karthik Nandakumar, and Arun Ross. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern recognition letters*, 79:80–105, 2016.

[3] Shlomo Greenberg, Mayer Aladjem, and Daniel Kogan. Fingerprint image enhancement using filtering techniques. *Real-Time Imaging*, 8(3):227–236, 2002.

[4] Chaohong Wu, Zhixin Shi, and Venu Govindaraju. Fingerprint image enhancement method using directional median filter. In *Biometric Technology for Human Identification*, volume 5404, pages 66–75. International Society for Optics and Photonics, 2004.

[5] Mark Rahmes, Josef DeVaughn Allen, Abdelmoula Elharti, and Gnana Bhaskar Tenali. Fingerprint reconstruction method using partial differential equation and exemplar-based inpainting methods. In *2007 Biometrics Symposium*, pages 1–6. IEEE, 2007.

[6] Lin Hong, Yifei Wan, and Anil Jain. Fingerprint image enhancement: algorithm and performance evaluation. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):777–789, 1998.

[7] Sharat Chikkerur, Alexander N Cartwright, and Venu Govindaraju. Fingerprint enhancement using stft analysis. *Pattern recognition*, 40(1):198–211, 2007.

[8] Jianjiang Feng, Jie Zhou, and Anil K Jain. Orientation field estimation for latent fingerprint enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 35(4):925–940, 2012.

[9] Raffaele Cappelli, Dario Maio, Alessandra Lumini, and Davide Maltoni. Fingerprint image reconstruction from standard templates. *IEEE transactions on pattern analysis and machine intelligence*, 29(9):1489–1503, 2007.

[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 (CVPR)*, June 2016.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[12] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[13] Tran Minh Quan, David GC Hildebrand, and Won-Ki Jeong. Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics. *arXiv preprint arXiv:1612.05360*, 2016.

[14] Yao Tang, Fei Gao, Jufu Feng, and Yuhang Liu. Fingernet: An unified deep network for fingerprint minutiae extraction. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 108–116. IEEE, 2017.

[15] Jian Li, Jianjiang Feng, and C-C Jay Kuo. Deep convolutional neural network for latent fingerprint enhancement. *Signal Processing: Image Communication*, 60:52–63, 2018.

[16] Dinh-Luan Nguyen, Kai Cao, and Anil K Jain. Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge. In *2018 International Conference on Biometrics (ICB)*, pages 9–16. IEEE, 2018.

[17] Jan Svoboda, Federico Monti, and Michael M Bronstein. Generative convolutional networks for latent fingerprint reconstruction. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 429–436. IEEE, 2017.

[18] Ramakrishna Prabhu, Xiaojing Yu, Zhangyang Wang, Ding Liu, and Anxiao Andrew Jiang. U-finger: Multi-scale dilated convolutional network for fingerprint image denoising and inpainting. In *Inpainting and Denoising Challenges*, pages 45–50. Springer, 2019.

[19] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S Huang. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*, 2017.

[20] Sukesh Adiga and Jayanthi Sivaswamy. Fpd-m-net: Fingerprint image denoising and inpainting using m-net based convolutional neural networks. In *Inpainting and Denoising Challenges*, pages 51–61. Springer, 2019.

[21] Raghav Mehta and Jayanthi Sivaswamy. M-net: A convolutional neural network for deep brain structure segmentation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 437–440. IEEE, 2017.

[22] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.

[23] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[24] Steven M Bellovin, Preetam K Dutta, and Nathan Reitinger. Privacy and synthetic datasets. *Stan. Tech. L. Rev.*, 22:1, 2019.

[25] Sergey I Nikolenko. Synthetic data for deep learning. *arXiv preprint arXiv:1909.11512*, 2019.

국문초록

지문을 사용한 생체 인식 인증은 품질이 저하된 지문 영상에서 지문을 추출하기 위한 영상 잡음 제거 및 복원 방법을 필요로 한다. 지문 영상 잡음 제거 및 복원을 위한 몇 가지 딥러닝 모델이 *ChaLearn LAP Inpainting Competition - Track 3, ECCV 2018*에서 제안되었다. 본 논문에서는 지문 영상 잡음 제거를 위한 새로운 딥러닝 모델을 제안한다. 제안된 모델은 영상 분할을 위한 합성곱 신경망 기반 딥러닝 모델인 FusionNet을 수정하여 작성하였다. 제안된 모델의 성능은 ChaLearn Competition의 데이터셋을 사용하여 검증되었다. 이를 통해 제안된 모델이 대회에서 높은 성능을 획득한 다른 모델들에 비하여 더 나은 결과를 얻음을 확인하였다.

주요어휘: 딥러닝, 지문 영상, 영상 잡음 제거, 합성곱 신경망, FusionNet
학번: 2018-28868