J. KSIAM Vol.24, No.4, 363–374, 2020 http://dx.doi.org/10.12941/jksiam.2020.24.363

FINGERPRINT IMAGE DENOISING AND INPAINTING USING CONVOLUTIONAL NEURAL NETWORK

JUNGYOON BAE[1], HAN-SOO CHOI[2], SUJIN KIM[1], AND MYUNGJOO KANG[3†]

[1]DEPARTMENT OF COMPUTATIONAL SCIENCE AND TECHNOLOGY, SEOUL NATIONAL UNIVERSITY, SEOUL, 08826, REPUBLIC OF KOREA.

[2]DEPARTMENT OF MATHEMATICAL SCIENCES / RESEARCH INSTITUTE OF MATHEMATICS, SEOUL NATIONAL UNIVERSITY, SEOUL, 08826, REPUBLIC OF KOREA.

[3]DEPARTMENT OF MATHEMATICAL SCIENCES, SEOUL NATIONAL UNIVERSITY, SEOUL, 08826, REPUBLIC OF KOREA.

Email address: †mkang@snu.ac.kr

ABSTRACT. Fingerprint authentication identifies a user based on the individual's unique fingerprint features. Fingerprint authentication methods are used in various real-life devices because they are convenient and safe and there is no risk of leakage, loss, or oblivion. However, fingerprint authentication methods are often ineffective when there is contamination of the given image through wet, dirty, dry, or wounded fingers. In this paper, a method is proposed to remove noise from fingerprint images using a convolutional neural network. The proposed model was verified using the dataset from the ChaLearn LAP Inpainting Competition Track 3-Fingerprint Denoising and Inpainting, ECCV 2018. It was demonstrated that the model proposed in this paper obtains better results with respect to the methods that achieved high performances in the competition.

1. INTRODUCTION

Recently, biometric authentication has been used as a major authentication technique for smart devices. Biometric authentication utilizes personal biometric information, and leads to significantly less leakage or theft, compared to existing authentication methods such as personal identification number (PIN) or password [1]. Automated fingerprint identification system (AFIS) is a biometric authentication technology that has recently attracted much attention, because they use simple touch-based sensors [2].

In fingerprint recognition, there are cases in which various types of environmental noises are generated when the user's finger in some undesirable conditions makes contact with the touch screen, which severely corrupt the fingerprint image. This reduces the quality of the

Received by the editors August 25 2020; Revised December 17 2020; Accepted in revised form December 17 2020; Published online December 25 2020.

2000 *Mathematics Subject Classification.* 93B05.

Key words and phrases. fingerprint image, convolutional neural network, image denoising, FusionNet.

† Corresponding author.

image, making it difficult to authenticate the user. Therefore, in order to solve this problem, study of image restoration is preferentially necessary. Conventional fingerprint image processing methods include image filtering or using partial differential equations. In the early days of fingerprint image processing research, traditional image filtering methods using a directional median filter [3], a Wiener filter, and an anisotropic filter [4] were proposed. Some methods based on partial differential equations were also proposed for automatic fingerprint restoration [5].

Recently, image processing using a convolutional neural network (CNN) has attracted much attention. In particular, many developments using CNNs have been made in the areas of image denoising, image segmentation, image enhancement, *etc*. These techniques have recently been applied to fingerprint denoising [6–11].

In this paper, a CNN model is proposed, which improves artificially degraded fingerprint images using random transformations (blur, brightness, contrast, elastic transformation, occlusion, scratch, resolution, and rotation). The performance of the proposed model was verified through various experiments. It was confirmed that the proposed model obtained better results compared to the previously proposed CNN models.

This paper is organized as follows: Section 2 describes recent trends in research related to fingerprint image denoising and inpainting. In Section 3, the proposed fingerprint image denoising and inpainting model is explained. The feasibility of the proposed model is verified through experiments in Section 4. Section 5 presents our conclusions.

2. Related Work

For accurate authentication of fingerprint images, research has been conducted on denoising, enhancement, and reconstruction of fingerprint images [12–15]. These studies include techniques based on existing algorithms and CNN-based methods.

First, Hong *et al.* extracted accurate feature point information with an algorithm that used Gabor filters to remove noise and preserve the ridge structure clearly, so that the accuracy of matching feature points between fingerprint images was improved [12]. Chikkerur *et al.* employed the short-time Fourier Transform(STFT) to enhance low-quality fingerprint images [13]. Feng *et al.* proposed the orientation field estimation algorithm using the prior information embedded in the fingerprint structure and demonstrated an improved accuracy in fingerprint recognition compared to previous studies [14]. Cappelli *et al.* made an approach to restore the fingerprint image by extracting the information of the orientation field and ridge structure from given fingerprint images [15].

In addition, Tang *et al.* proposed FingerNet using a CNN [8]. This method extracted detailed fingerprints from noisy ridge patterns and complex backgrounds. The model first obtained detailed fingerprint information by segmenting the orientation field and reinforcing the latent fingerprint. Thereafter, Li *et al.* proposed a model to improve FingerNet [9], and Nguyen *et al.* proposed MinutiaeNet, consisting of a coarse and fine network, to perform fully automatic minutiae extraction [10]. Here, the coarse network improved the image based on domain knowledge, and provided candidate minutiae locations by extracting a segmentation map. The

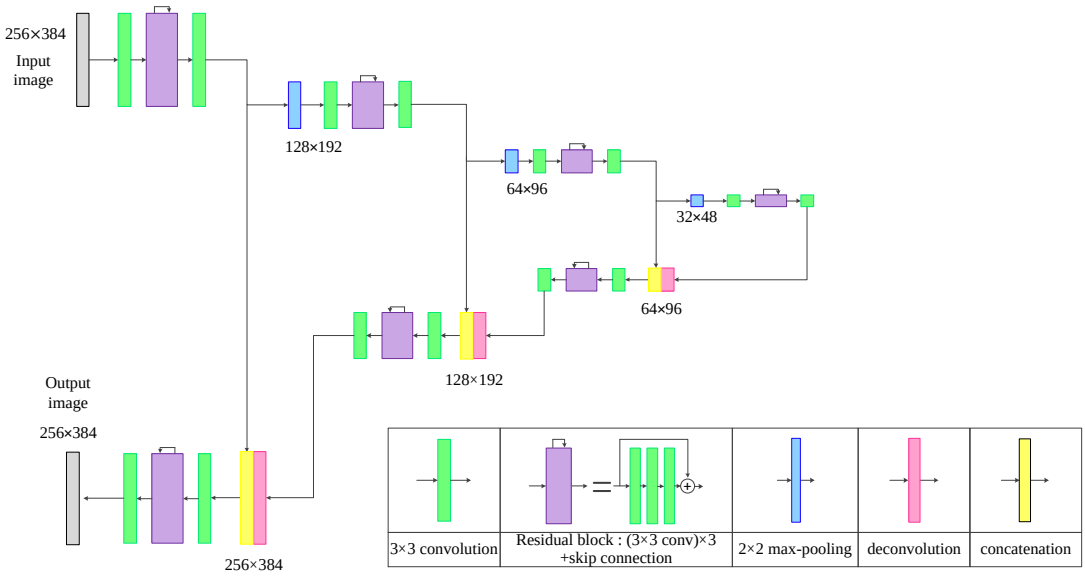

FIGURE 1. Proposed architecture in this paper.

fine network used a technique to refine these candidate minutiae locations. In addition, Svoboda *et al.* proposed a method for predicting and denoising the lost ridge region using a generative CNN [11].

Subsequently, based on these studies, the ChaLearn competition was held in 2018[1]. U-Finger [6], which achieved good results in the competition, performed fingerprint denoising using the DeepDenoising model [16]. The overall architecture of the model was designed using feature encoding and feature decoding that combines the convolution layer and residual block. In addition, FPD-M-Net [7] constructed a model based on M-Net [17] for brain segmentation. M-Net utilized the U-Net [18] to derive better segmentation results. FPD-M-Net was proposed by modifying the block arrangement order and loss function in the M-Net. Still, research related to fingerprint image denoising and inpainting has been ongoing [19, 20].

3. PROPOSED METHOD

3.1. **Proposed Architecture.** In this paper, we propose a new model for fingerprint denoising and inpainting, which was modified from FusionNet [21]. For the modifications made in the proposed model, refer to subsection 3.2. Unlike the existing FusionNet model, the size of the input/output image for training is set to 256×384, to match the size similar to the size of the input image of the dataset used in this study. Here, to perform image resizing, we used bicubic linear interpolation to use the lowest-order interpolation method to connect pixel values

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

at adjacent grid boundaries smoothly. A detailed description of the proposed architecture is as follows.

TABLE 1. Summary of proposed architecture. Here, conv means convolution, res means residual block, deconv means deconvolution and concat means concatenation.

Block type	Ingredient	Size of the 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	$256 \times 384 \times 1$

First, an input image in the dataset was resized to 256×384. The resized image was used as the input to the model and passed through a convolution block, a residual block, and a convolution block. Next, this feature map was used as an input for the next layer and the skip-connection.

Subsequently, the size of the feature map was reduced using stride-2 2×2 maxpooling. As shown previously, this downscaled feature map was passed sequentially through a convolution block, a residual block, and a convolution block. This feature map was used as an input for the next layer and another skip-connection. This method was repeated until the size of the feature map became 32×48.

Then, the size of the feature map was expanded to 64×96, by passing through a deconvolution block. After the skip-connection, which concatenates the feature map and the corresponding feature map of the encoder, the resulting feature map was passed sequentially through a convolution block, a residual block, and a convolution block.

As in the previous stage, the feature map was passed through a deconvolution block, concatenated with the corresponding feature map of the encoder through the skip-connection, and

sequentially passed through a convolution block, a residual block, and a convolution block. This process was repeated until the size of the feature map became 256×384. For the last convolution block of those repetitions, the output depth was set to 1.

Subsequently, the output feature map was normalized to obtain a grayscale image, with the minimum value of the feature map set to 0, and the maximum value set to 1. The filter size of all the convolution blocks in the model was 3×3, and the filter size of all the deconvolution blocks in the model was 2×2. Batch normalization and the ReLU activation function were used to pass through all the convolution blocks.

A more detailed description of the network is shown in Fig. 1 and 1. In Fig. 1, the green blocks represent convolution blocks, the purple blocks represent residual blocks, the red blocks represent deconvolution blocks, and the blue blocks depict 2×2 maxpooling. The residual block in the proposed model was composed of three convolution blocks with a short skip-connection, as in the FusionNet model.

3.2. **Comparison with FusionNet.** Some modifications were made to build the proposed model from FusionNet. First, the size of the input and the output image of the proposed model was set to 256×384, instead of 640×640 as in FusionNet. Also, the proposed network is shallower compared with FusionNet where the maximum depth of the proposed model is 192 rather than 1024, as in FusionNet. The number of downscaling and upscaling of the proposed model was reduced by one compared with FusionNet. Additionally, concatenations of feature maps were used in the skip-connections of the proposed model, whereas feature maps were added in the skip-connections of FusionNet. For an explanation of the above modifications, refer to the ablation study in subsection 4.3.

4. Experimental Results

4.1. **Implementation Details.** Experiments were performed to verify the architecture proposed in the previous section. This section outlines these experiments. The experimental environment is shown in 2. Adam optimizer was used to train the network. For the loss function, the loss function presented by Zhao *et al.* [22] $L = \alpha L_{ms-ssim} + (1 - \alpha)L_1$ was used, which utilizes both the L_1 loss and the MS-SSIM. In this study, we set the parameter $\alpha = 0.85$ as

TABLE 2. Data sheet of 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

TABLE 3. Cost time comparison of each method for inference: CPU-Only

Method	FPD-M-Net [7]	U-Finger [6]	Proposed method
Framework	Keras	Caffe	TensorFlow
Cost time (1 image/sec)	0.4116	3.0417	0.2963

in [7]. The number of epoch was set to 50, and the batch size was set to 32. The learning rate for the training was set to 10^{-4} in epoch 1, 10^{-5} from epoch 2 to 3, 5×10^{-6} from epoch 4 to 9, 10^{-6} from epoch 10 to 19, 5×10^{-7} from epoch 20 to 39, and 10^{-7} from epoch 40 to 50.

The dataset used in the experiment is a large-scale synthesized dataset of realistic artificial fingerprints released by the ECCV 2018 ChaLearn competition organizers. A more detailed description of the dataset is provided in subsection 4.2.

Moreover, FPD-M-Net [7] and U-Finger [6] were selected for comparison with the method proposed in this paper, because these models showed excellent performance in the 2018 ECCV challenge. And we also attached the baseline network provided in the competition which is a standard deep neural network[2] with residual blocks. The CPU-only cost time of the methods used in the comparison group is shown in 3.

In this study, peak signal-to-noise ratio (PSNR) and structural similarity metric (SSIM) were selected as the evaluation metrics used to verify the similarity between the ground-truth image and the predicted image. PSNR is defined as in Eq. (4.1) and indicates the ratio of the noise to the maximum amplitude that a signal can have. SSIM is defined as in Eq. (4.2) and is a method of measuring the similarity between the original image and the distorted image, where the distortions are caused by compression and transformation. SSIM is not a numerical error, but it is used to evaluate the similarity of the image qualities in the context of human visual perception [23].

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

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

4.2. **Experiment on Dataset.** The dataset used in the experiments was released in the ChaLearn fingerprint denoising competition. The dataset consists of pairs of ground-truth fingerprint images and degraded fingerprint images. The ground-truth fingerprint images are grayscale images of size 275×400, and the degraded fingerprint images are RGB images of size 275×400. An examples of the dataset are shown in 2.

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

FIGURE 2. Examples of the pair of images of the dataset.

All the ground-truth images in the dataset are synthetic images generated by the Anguli Synthetic Fingerprint Generator. Synthetic data are used, not only because it takes considerable time and effort to collect a large amount of real fingerprint images, but also because it may be problematic due to privacy laws – in some countries, the distribution of sensitive personal information is prohibited [24].

The degraded fingerprint images of the dataset were created by adding some random artifacts (blur, brightness, contrast, elastic transformation, occlusion, scratch, resolution, and rotation) and background to the ground-truth fingerprint images.

The dataset consists of training, validation, and test datasets; the number of image pairs in each category is shown in 4. The experiment using the dataset was as follows.

The columns in Fig. 3 show, from left to right: the degraded fingerprint images from the validation dataset, which were used as input images, and their corresponding ground-truth images; the output images of the FPD-M-Net, the U-Finger, and the proposed model, respectively.

As indicated by the red circles in the top-right corner of the output images in the first row of Fig. 3, the FPD-M-Net and the U-Finger predict the fingerprint inaccurately, owing to the

TABLE 4. Number of pairs of images of each dataset

Dataset	Training	Validation	Test
Number of images	75,600	8,400	8,400

FIGURE 3. Examples of the validation image results.

background of the degraded input image. However, the method proposed in this paper shows a better result for the corresponding part of the image.

TABLE 5. Comparison results with other networks 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 6. Comparison results with other networks 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**

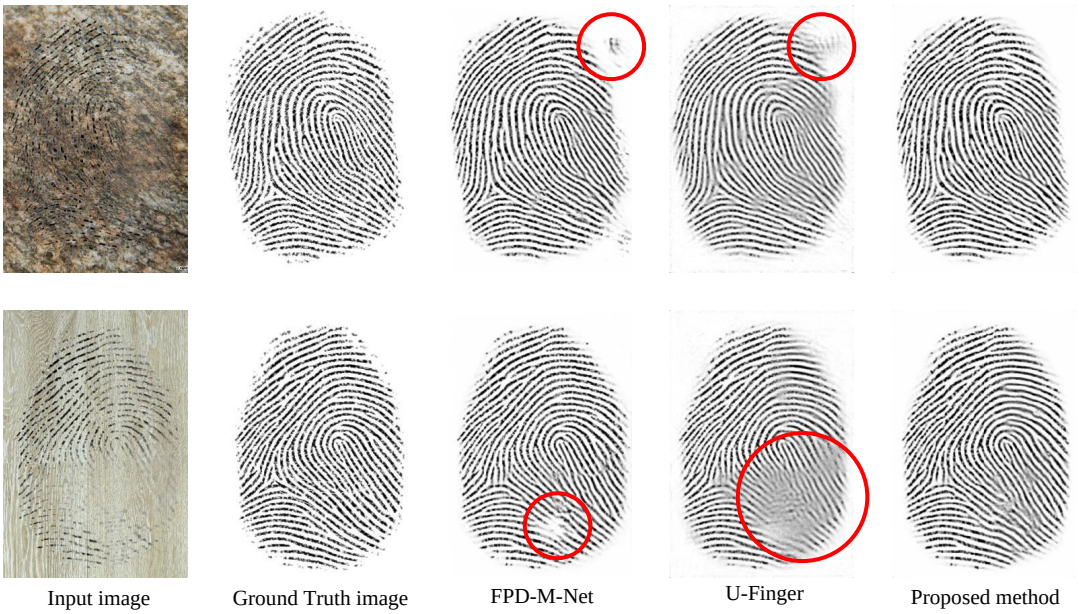

FIGURE 4. Examples of the test image results.

Similarly, as indicated by the red circles in the upper-left corner of the output images in the second row of Fig. 3, the FPD-M-Net and the U-Finger incorrectly predict the result of the fingerprint. However, the proposed model can predict the fingerprint well, in the corresponding area.

The average values of PSNR and SSIM of the proposed model using the validation dataset are shown in 5. The table indicates that the proposed method provides the best results in terms of both PSNR and SSIM among all the approaches in the experiment.

The columns in Fig. 4 show, from left to right: the degraded fingerprint images from the test dataset, which were used as input images, and their corresponding ground-truth images; the output images of the FPD-M-Net, the U-Finger, and the proposed model, respectively.

As indicated by the red circles at the top-right part of the output images in the first row of Fig. 4, the FPD-M-Net and the U-Finger inaccurately predict the outer part of the image as a fingerprint. However, the proposed network does not show such an incorrect prediction.

Additionally, the second row of Fig. 4 shows that the bottom center parts of the output images of the FPD-M-Net and the U-Finger are blurry, owing to the input image. However, the proposed model can predict the fingerprint well, in the corresponding area.

The average values of PSNR and SSIM of the proposed model using the test dataset are shown in 6. The table indicates that the proposed method provides the best results in terms of both PSNR and SSIM, among all the methods in the experiment.

4.3. **Ablation Study.** In this subsection, additional experiments performed using the proposed model are highlighted. As the proposed model was created based on FusionNet, an experiment was conducted to train FusionNet with the original input size, 640×640. Here, for fairness in the number of training parameters, the depths of the feature maps for training the FusionNet were set the same as for the proposed model.

We also experimented with a model of input size 256×384 that uses addition instead of concatenation in the skip-connections. In order to perform addition, the sizes of the feature maps of the encoder and the decoder of the skip-connection should be the same. Therefore, in the deconvolution block, the depths of the feature maps of the decoder were adjusted to be the same as those of the encoder.

The experimental results are shown in 7 and 8. The tables demonstrate that the method proposed in Section 3 provides the best results in terms of PSNR and SSIM, among all the approaches in the ablation study.

5. CONCLUSION

In this paper, a new CNN model for fingerprint denoising and inpainting was proposed. The proposed model is based on FusionNet, which is a CNN-based model that exhibits very good performance in semantic image segmentation. A few modifications were made to FusionNet to obtain a powerful fingerprint denoising and inpainting model. The performance of the proposed model was verified by comparing the experimental results with those of the existing CNN models for fingerprint denoising and inpainting. It was demonstrated that the proposed model obtained better results with respect to the other models that achieved substantial results in the ECCV 2018 workshop challenge.

The results of this study are expected to be applicable in real-life devices, such as fingerprint sensors.

TABLE 7. Ablation Study for proposed model : validation dataset

Method	PSNR	SSIM
FusionNet	17.0090	0.8037
Addition	17.1688	0.8472
Proposed	**17.1775**	**0.8480**

TABLE 8. Ablation Study for proposed model : test dataset

Method	PSNR	SSIM
FusionNet	17.0548	0.8044
Addition	17.1858	0.8471
Proposed	**17.1904**	**0.8478**

ACKNOWLEDGMENTS

Myungjoo Kang was supported by the National Research Foundation of Korea (2015R1A5A1009350) and the ICT R&D program of MSIT/IITP(No. 1711117093)

REFERENCES

[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] 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.

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

[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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Zhi-Feng Pang, Hui-Li Zhang, Shousheng Luo, and Tieyong Zeng. Image denoising based on the adaptive weighted tvp regularization. *Signal Processing*, 167:107325, 2020.

[20] Marc-André Blais, Andy Couturier, and Moulay A Akhloufi. Deep learning for partial fingerprint inpainting and recognition. In *International Conference on Image Analysis and Recognition*, pages 223–232. Springer, 2020.

[21] 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.

[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.