

FPD-M-net: Fingerprint Image Denoising and Inpainting Using M-Net Based Convolutional Neural Networks

Sukesh Adiga V and Jayanthi Sivaswamy

Center for Visual Information Technology (CVIT), IIIT-Hyderabad, India
sukesh.adigav@research.iiit.ac.in

Abstract. The fingerprint is a common biometric used for authentication and verification of an individual. These images are degraded when fingers are wet, dirty, dry or wounded and due to the failure of the sensors, etc. The extraction of the fingerprint from a degraded image requires denoising and inpainting. We propose to address these problems with an end-to-end trainable Convolutional Neural Network based architecture called *FPD-M-net*, by posing the fingerprint denoising and inpainting problem as a segmentation (foreground) task. Our architecture is based on the *M-net* with a change: structure similarity loss function, used for better extraction of the fingerprint from the noisy background. Our method outperforms the baseline method and achieves an overall 3rd rank in the *Chalearn LAP Inpainting Competition Track 3—Fingerprint Denoising and Inpainting, ECCV 2018*.

Keywords: Fingerprint image, Denoising, Inpainting, Deep learning.

把指纹的修复看做从噪声背景中分离下来的分割问题

1 Introduction

The fingerprint is an impression left by the friction ridges of a finger. Human fingerprints are detailed, nearly unique, difficult to alter, and durable over the life of an individual, making them suitable as long-term biometrics for identifying the uniqueness of an individual. It plays an increasingly important role in security to ensure privacy and identity verification. Fingerprint-based authentication has ubiquitous in the day to day life (Example: unlocking in smartphones, mobile payments, international travel, accessing the restricted area, etc.). Fingerprint retrieval is commonly used in forensic applications where the accuracy of verification is critical. However, the recovery of fingerprints deposited on surfaces such as glass or metal or polished stone remains challenging.

Fingerprints can be degraded when fingers are wet, dirty, skin dryness, etc. resulting in poor image quality. This problem can be viewed as denoising the fingerprint information from the background noise. In some cases, the image can have missing regions due to the failure of fingerprint sensors or wound in the finger. This requires a filling or inpainting from the neighboring region. Overall, the fingerprint image denoising and inpainting can be seen as a preprocessing

step to ease the authentication and verification carried out either by humans or existing software.

There are many methods for fingerprint extraction and analysis in literature. Early efforts were based on traditional image processing methods [2,1,19,9,4,13]. Recently, the Convolution Neural Networks (CNN) have been successful in many computer vision tasks such as segmentation, denoising, and inpainting. Some of the recent works based on CNN addressed for fingerprint extraction and analysis [11,16,8,5]. The success of deep learning in dealing with the inpainting and denoising [20] problems has led to the ChaLearn competition¹ which focuses on the development of a deep learning solution to restore fingerprint images from the degraded images. In our work, we pose the given problem as segmenting the fingerprint from the noisy background and hence propose a solution using an architecture developed for object segmentation.

2 Method

The distorted fingerprint images require denoising and inpainting for the restoration of accurate ridges which helps in reliable authentication and verification. The image consists of an object of interest (i.e., fingerprint) in a noisy or cluttered background. The problem can be solved using segmentation of object (fingerprint) from the noisy background. The M-net [7] does excellent segmentation, which forms motivation for our work.

Our aim is to denoise and inpaint the fingerprint images simultaneously using a segmentation approach, where the fingerprint information is the foreground of interest and other details are the background. The filling of any missing information should be possible with appropriate training, rather than explicit inpainting. The M-net was proposed for 3D brain structure segmentation, where an initial block converts 3D information into a 2D image on which segmentation is performed. Further, a categorical cross entropy loss function is used for segmentation. The 3D-to-2D conversion block is redundant, hence dropped. The loss function is also changed to suit the task at hand. The resulting architecture is called *FPD-M-net*. The details of network architecture, training and loss function are described next.

2.1 FPD-M-net architecture

The U-net [10] architecture is commonly used for tasks such as segmentation or restoration. The M-net is modified U-net for better segmentation using skip connections in the input and output ends. The architecture of our FPD-M-net is adapted from the M-net [7]. It consists of a Convolutional layer (CONV), maxpooling layer, upsampling layer, Dropout layer [14], Batch Normalisation layer (BN) [3], and rectified linear unit (ReLU) activation functions with encoder and decoder style of architecture as shown in Fig. 1. The encoding layer consists

¹ <https://competitions.codalab.org/competitions/18426>

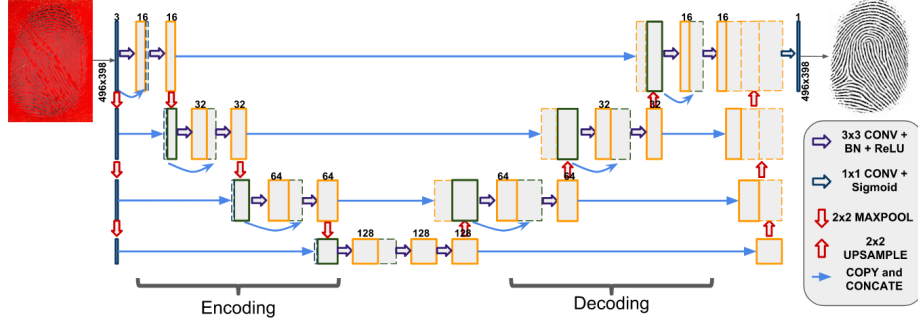


Fig. 1. The schematic representation of the FPD-M-net architecture. Solid yellow boxes represent the output of CONV-BN-ReLU block. Dashed boxes represent copied feature maps. The number of feature maps is denoted on the top of the box

of the repeated two blocks of 3×3 CONV, BN, and ReLU. Between two blocks of CONV-BN-ReLU layer, a dropout layer (with probability 0.2) is included. Dropout layer prevents over-fitting, BN layer enables faster and more stable training. The output of the two blocks CONV-BN-ReLU are concatenated and downsampled with a 2×2 maxpooling operation with stride 2. Decoder layer is similar to the encoder layer with one exception: maxpooling is replaced by the upsampling layer which helps to reconstruct an output image. The final layer is a 1×1 convolution layer with a sigmoid activation function which gives the reconstructed output image.

The skip connections used in FPD-M-net are shown (with blue arrows) in Fig. 1. The skip connection between adjacent convolution filters, enables the network to learn better features [15] and the skip connection from input-to-encoder, encoder-to-decoder, and decoder-to-output ends ensures that the network has sufficient information to drives fine grain details of the fingerprint image. The difference between FPD-M-net and M-net are: i) Conv-ReLU-BN blocks are replaced with the Conv-BN-ReLU blocks as in [3]; ii) a combination of a per-pixel loss and structure similarity loss are used for the loss function as the ground-truth fingerprint image is integer valued in the range $[0, 255]$.

2.2 Training details

The network is trained end-to-end with the pair of noisy/distorted and clean/ground-truth fingerprint image. The input and ground-truth images are padded with the edge values to suit the network and images are normalised to take values between $[0,1]$. The size of input and ground truth images are 275×400 pixels. After padding, the size of the images become 368×496 . Padding is done so that the output of the network effectively sees the input image size of 275×400 . In the testing phase, the distorted images are given to FPD-M-net, to get a clean fingerprint image as output. The output images are unpadded to match the original size and compared against the reference image.

2.3 Loss function

The mean squared error (MSE), a reference-based metric and Peak Signal-to-Noise Ratio (PSNR) are popular error measures for reconstruction problems. In deep learning, MSE is widely used as a loss function for many applications. However, neither MSE nor PSNR correlates well with human perception of image quality. Structure similarity index (SSIM) [17] is a reference-based metric that has been developed for this purpose. The SSIM is measured at a fixed scale and may only be appropriate for a certain range of image scales. A more advanced form of SSIM is the multi-scale structure similarity index (MS-SSIM) [18]. It preserves the structure and contrast in high-frequency regions better than the other loss functions [21]. In addition to choosing perceptually correlated metric, it is also of interest to preserve intensity as ground-truth fingerprint image has real value. So we choose a combination of per-pixel loss and MS-SSIM to define the loss function with weight δ , as shown:

$$L(\theta) = \delta \cdot L_{\text{MS-SSIM}}(\theta) + (1 - \delta) \cdot L_{l_1}(\theta) \quad (1)$$

where, $L_{l_1}(\theta)$ is l_1 loss and $L_{\text{MS-SSIM}}(\theta)$ is standard MS-SSIM loss. The weights are set to $\delta = 0.85$ as per [21] and MS-SSIM is computed over three scales.

3 Experiments and Results

3.1 Dataset and Parameters

The dataset used in our experiment is obtained with the *Anguli: Synthetic Fingerprint Generator* software, provided by the *Chalearn LAP Inpainting Competition Track 3*². The dataset consists of a pair of degraded/distorted and ground-truth fingerprint images. The distorted images are synthetically generated by first degrading fingerprints with a distortion model which introduces blur, brightness, contrast, elastic transformation, occlusion, scratch, resolution, rotation and then overlaying the fingerprints on top of various backgrounds. The dataset consists of training, validation and test sets with a pair of degraded and ground-truth fingerprint images. It is described in Table 1. The images are padded and normalised before training and testing. The test set has no ground-truth and evaluation requires uploading the images to the competition site to get the quantitative score.

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

Table 1. Fingerprint images dataset

² <https://competitions.codalab.org/competitions/18426#participate>

The FPD-M-net was trained for 75 epochs for a week. A stochastic gradient descent (SGD) optimiser was used to minimise the per-pixel loss and structure similarity loss. The training parameters were: learning rate of 0.1; Nesterov momentum was set to 0.75; decay rate was set at 0.00001; batch size was chosen as 8. After 50 epochs learning rate was reduced to 0.01; Nesterov momentum was increased to 0.95. The network parameters are presented in Table 2. The network was implemented on an NVIDIA GTX 1080 GPU, with 12GB of GPU RAM on a core i7 processor. The entire architecture was implemented in Keras library using Theano backend. The code of our method has been publicly released³.

Parameter	First 50 epoch	After 50 epoch
Learning Rate	0.1	0.01
Nesterov momentum	0.75	0.95
Decay rate	0.00001	0.00001
Batch size	8	8

Table 2. FPD-M-net training parameters

3.2 Results and performance evaluation

The results of FPD-M-net were evaluated both qualitatively and quantitatively. We first compared it with U-net architecture using metrics such as PSNR, MSE. The perceptual quality of the results was evaluated using structural similarity (SSIM). Next, we provide a performance comparison with other participants of *Chalearn LAP Inpainting Competition Track 3—Fingerprint Denoising and Inpainting, ECCV 2018*. Finally, sample qualitative results are presented.

Performance evaluation with U-net: The quantitative comparison of results of FPD-M-net is compared against U-net (trained with the same setting as FPD-M-net). The U-net was trained with the same loss function with only encoder-to-decoder skip connection [10]. The denoising and inpainting performance of fingerprint images are evaluated using PSNR, MSE and SSIM metric. These results are presented in Table 3 for both the validation and test sets. Our method outperforms the U-net in all the metrics, which indicates skip connections aid in achieving superior fingerprint restoration.

set	method	MSE	PSNR	SSIM
validation	U-net	0.0286	16.2122	0.8202
	FPD-M-net	0.0270	16.5149	0.8255
test	U-net	0.0284	16.2466	0.8207
	FPD-M-net	0.0268	16.5534	0.8261

Table 3. Quantitative comparison of results of FPD-M-net with U-net

³ <https://github.com/adigas/FDPMNet>

Ablation experiments with batch normalisation: In order to assess the effect of batch normalisation (BN) before and after the activation function, two FPD-M-net networks were trained: one with BN after ReLU activation (similar to M-net) and one with BN before ReLU activation. For convenience, the BN after and before ReLU activation network called as FPD-M-net-A and FPD-M-net-B, respectively. Both the networks were trained with the same settings as described in section 3.1. The quantitative results for validation and test set are presented in Table 4. Results indicate that the FPD-M-net-B is better in PSNR and MSE metric than FPD-M-net-A, whereas for SSIM the FPD-M-net-A has slightly better than FPD-M-net-B. Overall FPD-M-net-B has a fairly good results relative to FPD-M-net-A. Hence, BN before ReLU activation function is used in FPD-M-net.

set	method	MSE	PSNR	SSIM
validation	FPD-M-net-B	0.0270	16.5149	0.8255
	FPD-M-net-A	0.0277	16.4019	0.8265
test	FPD-M-net-B	0.0268	16.5534	0.8261
	FPD-M-net-A	0.0275	16.4336	0.8270

Table 4. Comparison of FPD-M-net with BN before and after activation function

Comparison with others in Challenge: The fingerprint denoising and inpainting challenge was organised by *Chalearn LAP Inpainting Competition, ECCV 2018*. The final quantitative results of the competition are presented in Table 5. The CVxTz and rgsl888 team also used a U-net [10] based architecture, whereas hcilab team used a hierarchical deep learning approach [12]. The baseline network provided in competition is a standard deep neural network⁴ with residual blocks. The rgsl888 team uses a dilated convolutions compared to CVxTz team. In our U-net implementation (section 3.2), a combination of l_1 and MS-SSIM loss function is used whereas CVxTz and rgsl888 used l_1 and l_2 loss function, respectively. The overall CVxTz team performs the best. It should be noted that the U-net network used by CVxTz team has almost double the network depth as compared to our FPD-M-net and also used additional data augmentation. Our method obtains 0.8261 (rank 2) in the SSIM metric, which shows the effectiveness of MS-SSIM in loss function.

Team	Rank	MSE	PSNR	SSIM
CVxTz	1.0000	0.0189 (1)	17.6968 (1)	0.8427 (1)
rgsl888	2.3333	0.0231 (2)	16.9688 (2)	0.8093 (3)
sukeshadigav (FPD-M-net)	3.3333	0.0268 (4)	16.5534 (4)	0.8261 (2)
hcilab	3.3333	0.0238 (3)	16.6465 (3)	0.8033 (4)
Baseline	4.6666	0.0241 (5)	16.4109 (5)	0.7965 (5)

Table 5. Performance of different methods in the Challenge

⁴ <http://chalearnlap.cvc.uab.es/challenge/26/track/32/baseline/>



Fig. 2. Illustration of fingerprint denoising and inpainting results for varying distorted images. From left to right: distorted fingerprints, corresponding ground-truth, results of U-net, our methods FPD-M-net-A and FPD-M-net-B

Qualitative results A qualitative comparison of fingerprint image denoising and inpainting can be done with sample images from the test set which are shown in Fig. 2. Two moderately distorted (Row 1 and 2) and two severely distorted fingerprint images (Row 3 and 4) and its corresponding results are shown. The weak fingerprints are successfully recovered as shown in Row 1. The networks are robust to even strong background clutter (Row 2). The automatic filling is seen

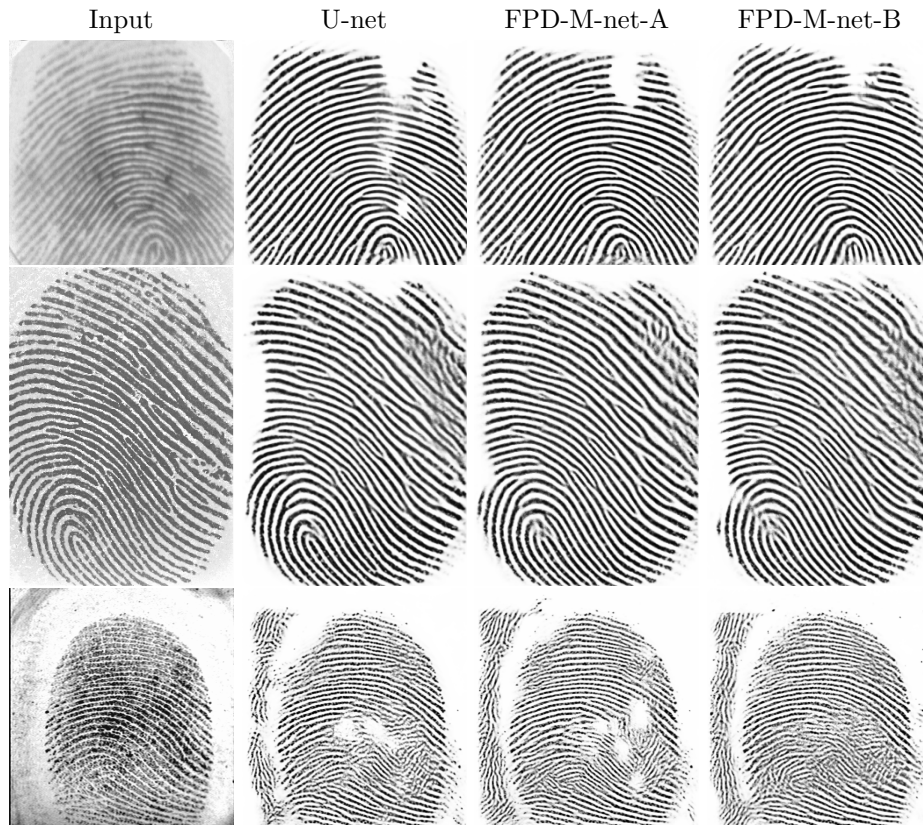


Fig. 3. Sample results of fingerprint denoising and inpainting on real images. From left to right: distorted fingerprints, results of U-net, our methods FPD-M-net-A and FPD-M-net-B

to be successful in images in Row 3 and 4. Our FPD-M-net method produces better results for severely distorted images (Row 4) compared to U-net.

Qualitative comparison with real fingerprint Since the images provided in the Challenge were synthetically generated it is of interest to test the proposed architecture on real images also. The qualitative performance of denoising and inpainting results on real images from three datasets: FVC2000 DB1, DB2 and DB3 [6] are shown in Fig. 3. These datasets are captured by different sensors having varying resolutions. DB1 images appear closer to the synthetic dataset. A sample image from DB1 (Row 1), DB2 (Row 2) and DB3 (Row 3) along with the outputs are shown in Fig. 3. The FPD-M-net methods produce the better result for DB1 image compared to U-net. In the case of a DB2 image, portions the fingerprint are missing in the top and left part of the image. Some artefact is also seen in all the results in the top right of the image. Apart from these defects, all the methods perform fairly well. In the case of a DB3 image, all

the results exhibit some loss of information, unlike FPD-M-net-B which however has some distortion (in the lower part). The difference in the results of testing on synthetic versus real images could be due to a number of factors including variation in acquisition (sensors and resolutions) which affect the width of ridges.

4 Conclusion

In this work, we presented an FPD-M-net model for fingerprint denoising and inpainting using a pair of synthetic data. The model, based on a segmentation architecture, was shown to handle both denoising and inpainting of fingerprint images, simultaneously. It outperforms the U-net, and the baseline model given in the competition. Our model is robust to strong background clutter, weak signal and performs automatic filling effectively. The good perceptual results for both qualitatively and quantitatively indicate the effectiveness of the MS-SSIM loss function. Results for images acquired with different sensors suggest the need for sensor-specific training for better results.

References

1. Greenberg, S., Aladjem, M., Kogan, D.: Fingerprint image enhancement using filtering techniques. *Real-time Imaging* 8(3), 227–236 (2002)
2. Hong, L., Wan, Y., Jain, A.: Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Transactions on pattern analysis and machine intelligence* 20(8), 777–789 (1998)
3. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015)
4. Khan, M.A.: Fingerprint image enhancement and minutiae extraction (2011)
5. Li, J., Feng, J., Kuo, C.C.J.: Deep convolutional neural network for latent fingerprint enhancement. *Image Communication Signal Processing* 60, 52–63 (2018)
6. Maio, D., Maltoni, D., Cappelli, R., Wayman, J.L., Jain, A.K.: Fvc2000: Fingerprint verification competition. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (3), 402–412 (2002)
7. Mehta, R., Sivaswamy, J.: M-net: A convolutional neural network for deep brain structure segmentation. In: *Proc. of 14th International Symposium on Biomedical Imaging (ISBI)*. pp. 437–440. IEEE (2017)
8. Nguyen, D.L., Cao, K., Jain, A.K.: Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge. *arXiv preprint arXiv:1712.09401* (2017)
9. Rahmes, M., Allen, J.D., Elharti, A., Tenali, G.B.: Fingerprint reconstruction method using partial differential equation and exemplar-based inpainting methods. In: *Biometrics Symposium*. pp. 1–6. IEEE (2007)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention (MICCAI)*. pp. 234–241. Springer (2015)
11. Sahasrabudhe, M., Namboodiri, A.M.: Fingerprint enhancement using unsupervised hierarchical feature learning. In: *Proceedings of Indian Conference on Computer Vision Graphics and Image Processing*. p. 2. ACM (2014)

12. Salakhutdinov, R., Tenenbaum, J.B., Torralba, A.: Learning with hierarchical-deep models. *IEEE transactions on pattern analysis and machine intelligence* 35(8), 1958–1971 (2013)
13. Singh, K., Kapoor, R., Nayar, R.: Fingerprint denoising using ridge orientation based clustered dictionaries. *Neurocomputing* 167, 418–423 (2015)
14. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15(1), 1929–1958 (2014)
15. Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. *arXiv preprint arXiv:1505.00387* (2015)
16. Tang, Y., Gao, F., Feng, J., Liu, Y.: Fingernet: An unified deep network for fingerprint minutiae extraction. In: *IEEE International Joint Conference on Biometrics (IJCB)*. pp. 108–116. IEEE (2017)
17. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612 (2004)
18. Wang, Z., Simoncelli, E., Bovik, A., et al.: Multi-scale structural similarity for image quality assessment. In: *ASILOMAR CONFERENCE ON SIGNALS SYSTEMS AND COMPUTERS*. vol. 2, pp. 1398–1402. IEEE (2003)
19. Wu, C., Shi, Z., Govindaraju, V.: Fingerprint image enhancement method using directional median filter. In: *Biometric Technology for Human Identification*. vol. 5404, pp. 66–76. International Society for Optics and Photonics (2004)
20. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: *Advances in neural information processing systems*. pp. 341–349 (2012)
21. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging* 3(1), 47–57 (2017)