

Deep End-to-end Fingerprint Denoising and Inpainting

Youness MANSAR

Abstract This work describes our winning solution for the Chalearn LAP Inpainting Competition Track 3 - Fingerprint Denoising and In-painting. The objective of this competition is to reduce noise, remove the background pattern and replace missing parts of fingerprint images in order to simplify the verification made by humans or third-party software. In this paper, we use a U-Net like CNN model that performs all those steps end-to-end after being trained on the competition data in a fully supervised way. This architecture and training procedure achieved the best results on all three metrics of the competition [6].

1 Background

Fingerprints play an important role in privacy and identity verification but can also be used in forensic operations. This means that having the ability to accurately process and match fingerprints can be a valuable asset. This is what motivates this work where the objective is to retrieve a cleaned image of a fingerprint out of a noisy, distorted version.

Generally, images contain noise and perturbations that may be due to the acquisition device, compression method or post-processing done. This motivates the research on tools and methods like denoising and inpainting to alleviate this problem. They are used as a pre-processing step in order to simplify the subsequent tasks and improve the target performance. In our case, the end goal is to improve the fingerprint false acceptance rate or the false rejection rate.

One approach to denoising is the TV method [2] which is based on the principle that noisy images have a high total variation, the aim of the TV approach is to thus reduce the regularized total variation of the input image.

[5] reviews multiples methods to denoising like the Gaussian smoothing model or

translation invariant wavelet thresholding, among others.

A more recent direction to denoising and inpainting is based on deep neural networks where a sequence of convolution layers are optimized to learn a mapping from a noisy image to a "clean" version of that image. [7] studies the same problem as the Chalearn competition and uses and proposes encoder-decoder architecture to solve it. [1] shows that using skip connections helps avoid the issues related to training deep neural networks like the vanishing gradient problem.

Similar to [1], [3] introduces an architecture called U-net that is also an encoder-decoder type with skip connections that is used primarily for image segmentation. U-Net showed impressive results when used along with data augmentation even when the size of the dataset is small. In this work, we are going to use an architecture that is similar to U-Net and show that it can be applied successfully even outside pure segmentation tasks.

2 Data

The dataset provided by the organizers consisted of 84000 (200, 400) fingerprint images generated using Anguli: Synthetic Fingerprint Generator. Those images were then artificially degraded by adding a background and random transformations (blur, brightness, contrast, elastic transformation, occlusion, scratch, resolution, rotation). The objective is to retrieve the clean fingerprint image from the degraded version. We use the set of parallel data (Degraded image, Clean image) as the (Input, Ground Truth) of our model training.

3 Proposed solution

3.1 Model

The architecture used is described in Figure 1 and is similar to the one introduced in [1], except that we pad the input with zeros instead of mirroring the edges. The major advantage of this architecture is its ability to take into account a wider context when making a prediction for a pixel. This is thanks to the large number of channels used in the up-sampling operation.

3.2 Image processing

Input image processing : We apply this sequence of processing steps before feeding it to the CNN.

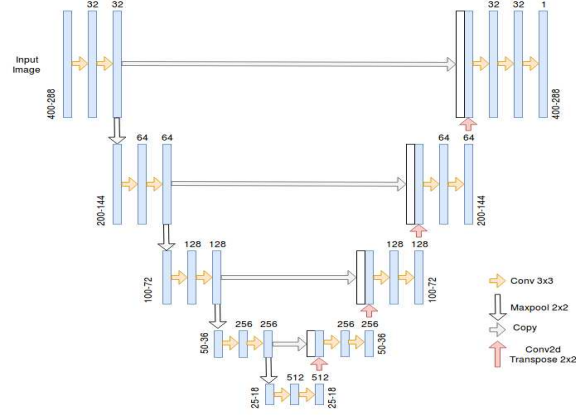


Fig. 1: U-net architecture used

- Normalization : we divide pixel intensities by 255 so they are in the 0-1 range.
- Re-sizing : The network expects each dimension of the input image to be divisible by 2^4 because of the pooling operations.
- Data augmentation : Random flip (Horizontal or vertical or both), Random Shear, Random translation (Horizontal or vertical or both), Random Zoom, Random Contrast change, Random Saturation change, Random Rotation. Performed during training only.

Output image processing : We apply this sequence of processing steps before submitting the results.

- Min-Max scaling : We min-max scale the output to the 0-255 range.
- Re-sizing : We re-size the size of the output to the original size of the input.

3.3 Training Procedure

We use Adam optimizer with an initial learning rate of $1e^{-4}$ that is reduced by a factor of 0.5 each time the validation loss plateaued for more than 3 epochs and the learning is stopped if the validation loss does not improve for the last 5 epochs. Implementation was done using Keras [4] with Tensorflow backend on a 1070 GTX card.

Team	MAE ↓	PSNR ↑	SSIM ↑
CVxTz (This work)	0.0189	17.6968	0.8427
rgsl888	0.0231	16.9688	0.8093
hcilab	0.0238	16.6465	0.8033
sukeshadigav	0.0268	16.5534	0.8261

Table 1: Best Test results in bold

4 Results

Our approach gets the best results on all three metrics. Even though we only used the MAE in our loss function, it seems to have acted a good proxy for the other two metrics.

As a comparison rgsl888 used a similar architecture to ours but added dilated convolutions to expand the receptive field of the network. hcilab used a hierarchical approach and sureshadigav used an M-Net Based Convolutional Neural Network.

5 Advantages and Limitations

Our approach has the merit of being end-to-end, requires minimal pre-processing to the input and uses a single model. All of this simplifies the use of the approach in a real-life scenario.

This approach also comes with few limitations like the fact that the train and test sets are both synthetic, which means that we do not know if the same performance will be preserved if the trained model is applied to real data. Another issue is that since the model is trained in a fully supervised way, then it is unlikely to generalize beyond the perturbations that it was trained on. This reaffirms the need to train on real fingerprint data.

6 Conclusion and future work

In this paper, we describe the approach we used to achieve 1st place on the Chalearn LAP In-painting Competition Track 3 - Fingerprint Denoising and In-painting. We describe the pre-processing steps needed, data augmentation, training procedure and network architecture used.

In our future work, we will experiment with transferring representations from higher level supervised tasks or by using a semi-supervised approach like adding an adversarial loss.

References

1. Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using convolutional auto-encoders with symmetric skip connections. *CoRR*, abs/1606.08921, 2016.
2. Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *MULTISCALE MODEL. SIMUL.*, 4(2):460–489, 2005.
3. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
4. Chollet, François and others Keras <https://keras.io>
5. A. Buades and B. Coll and J. M. Morel A review of image denoising algorithms, with a new one *SIMUL*, volume 4 490–530, 2005.
6. 2018 Looking at People ECCV Satellite Challenge - Track 3 - fingerprint denoising - <http://chalearnlap.cvc.uab.es/challenge/26/track/32/result/>
7. Jan Svoboda and Federico Monti and Michael M. Bronstein Generative Convolutional Networks for Latent Fingerprint Reconstruction *CoRR*, abs/1705.01707, 2017.