Fingerprint image enhancement using fully convolutional deep autoencoders

Destaque de imagens de impressão digital utilizando autoencoders profundos totalmente convolucionais

DOI:10.34117/bjdv8n5-474

Recebimento dos originais: 21/03/2022
Aceitação para publicação: 29/04/2022

Sandoval Veríssimo de Sousa Neto
Mestrando em Informática
Instituição: Universidade Federal da Paraíba (UFPB)
Endereço: Centro de Informática. Rua dos Escoteiros, s/n - Mangabeira, João Pessoa PB, CEP: 58055-000 - Brasil
E-mail: sandovalverissimo@eng.ci.ufpb.br

Leonardo Vidal Batista
Doutor em Engenharia Elétrica
Instituição: Universidade Federal da Paraíba (UFPB)
Endereço: Centro de Informática. Rua dos Escoteiros, s/n - Mangabeira, João Pessoa PB, CEP: 58055-000 - Brasil
E-mail: leonardo@ci.ufpb.br

Pedro Ivo Aragão Guimarães
Mestrando em Engenharia Elétrica
Instituição: Instituto Federal da Paraíba (IFPB)
Endereço: Instituto Federal da Paraíba. Av. Primeiro de Maio, 720 - Jaguaribe, João Pessoa - PB, CEP: 58015-435 - Brasil
E-mail: pedro.aragao@academico.ifpb.edu.br

Túlio Emanuel Santana de Souza
Graduando em Ciências da Computação
Instituição: Universidade Federal da Paraíba (UFPB)
Endereço: Centro de Informática. Rua dos Escoteiros, s/n - Mangabeira, João Pessoa PB, CEP: 58055-000 - Brasil
E-mail: tuliosouza@cc.ci.ufpb.br

ABSTRACT
Image quality for fingerprint samples is critical for the matching process. Novel methods introduce deep learning matching techniques based on convolutions neural networks to enhance degraded fingerprint images. However, due to the nature of the enhanced image problem, these methods tend to rely on processing small image patches to achieve their goal. Such an approach may often yield satisfactory results while having high computational costs due to overlapping in patches. In this paper, we propose a fast and accurate fully convolutional neural network based on an auto-encoder architecture to enhance the quality of fingerprint images. We do not use the patch processing method and instead train a model to enhance the image as a whole. After exhaustive testing, we achieve a model that can quickly perform the desired task, while achieving an average of

97.956% and 83.748% per pixel accuracy on the easiest and hardest dataset respectively. The models were trained on the publicly available Fingerprint Verification Competition datasets. We then highlight the most general model that can best enhance the quality of all datasets.

Keywords: fingerprint, image enhancement, convolutional neural networks, fvc fingerprint dataset.

RESUMO
A qualidade da imagem para amostras de impressões digitais é crítica para o processo de correspondência. Novos métodos introduzem técnicas de correspondência de aprendizagem profunda baseadas em redes neurais de convoluções para melhorar as imagens de impressões digitais degradadas. No entanto, devido à natureza do problema da imagem melhorada, estes métodos tendem a depender do processamento de pequenas manchas de imagem para alcançar o seu objectivo. Tal abordagem pode muitas vezes produzir resultados satisfatórios, ao mesmo tempo que tem elevados custos computacionais devido à sobreposição de manchas. Neste artigo, propomos uma rede neural rápida e precisa totalmente convolutiva, baseada numa arquitectura de autocodificador para melhorar a qualidade das imagens das impressões digitais. Não utilizamos o método de processamento de remendos e, em vez disso, treinamos um modelo para melhorar a imagem como um todo. Após testes exaustivos, conseguimos um modelo que pode executar rapidamente a tarefa desejada, alcançando uma média de 97,956% e 83,748% por pixel de precisão no conjunto de dados mais fácil e mais difícil, respectivamente. Os modelos foram treinados nos conjuntos de dados do Concurso de Verificação de Impressões Digitais disponíveis ao público. Destacamos então o modelo mais geral que melhor pode melhorar a qualidade de todos os conjuntos de dados.

Palavras-chave: impressão digital, melhoramento de imagem, redes neurais convolutivas, conjunto de dados de impressões digitais fvc.

1 INTRODUCTION

Nowadays the identification of a person given a single biological feature (e.g., face, fingerprint, or voice) is commonly used in many scenarios and applications due to its high security. It is used in smartphones, when performing bank transactions, or even in addition to other biometric inputs in high-security facilities. The use of fingerprint as a recognition tool is over a century old, still, due to, significant advancements in sensors technology, fingerprint's uniqueness, and facility to be obtained, the fingerprint as an authentication method has experienced a quick growth in recent years.

Once an image is acquired, the process of fingerprint recognition can be divided into three steps. First, it is necessary to perform a quality enhancement on the image, in most of the real applications sensors are subjected to noise like fragments left by a previous user, or moist, this step will unite separated ridges or separate the jointed ones.

After that information is extracted from the fingerprint, these are features such as bifurcations and ridge endings. Finally, a matching process will compare the acquired features from the ones on the system database.

The main step of recognition based on a fingerprint is the matching step at the end of the process. Nevertheless, the previous tasks contribute heavily to the performance of the final step. When building a fingerprint verification system many features that can be acquired in the extraction process can downright eliminate the needing for the full matching step, this happens because aside from bifurcations and ridge endings several other features can be extracted from the image i.e. a fingerprint may be from an arch or a loop, knowing these characteristics the matching process can be optimized to not perform an unnecessary comparison.

Even though it is possible to fraud someone's identity using silicon fingers (GALBALLY-HERRERO, 2006), (SCHUCKERS, 2002), systems are updated regularly and there are techniques to detect these attacks (MARASCO, 2014), (SAJJAD, 2019), (ABHYANKAR, 2006). Fingerprint recognition is still one of the most used and secured methods of identification since there are not two equal fingerprints in the world. No matter the biometric trait is used, good image quality is always desirable, since recognition is based on what features are extracted, any noise can hinder the matching process performance.

The image enhancement process will be the main focus of this paper. This will be achieved by using a convolutional neural network (CNN). Since its first appearance in 1989, CNNs have achieved excellent advances in image processing and computer vision problems. These include, but are not limited by image classification (KRIZHEVSKY, 2012), image segmentation (RONNEBERGER, 2015) and image denoising (ZHANG, 2017). The latter is of most interest to the fingerprint recognition problem since it is not uncommon for images to have some degree of noise. Architectures used in this type of problem usually consist of autoencoders. The Stacked Denoising Autoencoders paper (VINCENT, 2010) is one of the earliest works on the matter.

In this paper, a simple and fast CNN architecture for the image denoising problem on fingerprint images is proposed. Results for models trained using different FVC (Fingerprint Recognition Competition) datasets. As a primary metric, the per-pixel accuracy of the image was chosen. Results vary depending on the FVC dataset used for training and testing, yet, they are mostly positive.

The remainder work is divided as follows: Section 2 will present work related to the main topic of this paper. Section 3 presents the network architecture and explains the training process. The result of the experiments and brief commentary of all datasets will be presented in section 4. Section 5 contains some unsolved problems and provides possible solutions for future implementations.

2 RELATED WORK

Throughout the years, many solutions for the image-enhancing problem have been proposed. Significant improvements made in the image processing field have generated quite a bit of scientific production. Specifically, when we narrow it down to fingerprint image enhancement, recent solutions are closing the gap with the new machine learning techniques.

The first works focusing on fingerprint image enhancement used common filtering techniques (HONG, 1998), Gabor filters performed particularly well in this task. These classic methods were predominant until Yang. *et. al.* (YANG, 2003), who proposed a modified Gabor filter to solve some issues of the common Gabor approaches.

Many recent works still make use of Gabor filters (RAMOS, 2018), though nowadays its use is often paired with some other technique, such as Volterra filters (ARIF, 2018), or Fast Fourier Transforms (ZAHEDI, 2015). These recent works also reassure that despite their age the FVC datasets are still relevant for the fingerprint enhancement problem.

Inspired by previously successful image enhancement convolutional neural network techniques, Svoboda *et. al.* (SVOBODA, 2017) applied a CNN for the fingerprint enhancement task on latent fingerprint images. This work achieved fairly good results. Nevertheless, it does suffer from the false minutiae reconstruction problem due to the sub-sampling and up-sampling processes. As explained in (ALSMIRAT, 2019), the compression ratio of an image severely affects the final accuracy result. The work also does not provide the number of images processed per second (IPS), which can be fairly long due to in an autoencoder depending on the architecture and dimensions of the input image, as re-sizing the image for a more suitable input size could imply in loss of quality.

Later, Tang *et. al.* (TANG, 2017) also employed a CNN in the fingerprint image field, but their time to extract minutia from images on an end-to-end approach with no middle process being required. His work emphasizes the importance of the image enhancement step as a way to boost the performance of the minutia extraction process, as mentioned in the future works section.

Other recent methods make use of the new deep learning techniques to solve the image-enhancing problem. Yet, those mainly focus on the use of convolutional neural networks for latent fingerprint enhancement (SVOBODA, 2017), (TANG, 2017).

Regardless, CNN-based approaches have a common issue; the minutiae are frequently lost due to the image downsampling process (ALSMIRAT, 2019). Also, due to the high resolution of the fingerprint image, the computational cost during train and evaluation can be expensive. The most common way to reduce the occurrence of this issue is to train the convolutional neural network with patches of the image. However, an overlap of these patches is needed, thus increasing the computational cost even further.

3 METHODOLOGY

After exhaustive tests, the architecture employed in this paper consists of a common hourglass autoencoder. The encoder has three convolutional layers, with 64 filters per layer, a stride of 1, and kernels size of 3x3, and make use of the zero paddings. Between each of these convolutional layers, a MaxPooling layer halves the dimensions of the input.

The decoding process is symmetric to the encoding one. It consists of the same three convolutional layers, with the same characteristics as the ones used in the encoding process. Except, the MaxPool layers in this process are replaced by UpSampling layers. At the end of the process, a convolutional layer generates an image of 1 channel. This layer has a kernel size of 1x1. Figure 1 provides a diagram of the network architecture.

Our network is different from common autoencoder architectures since it does not change the number of filters per layer. Instead, our results have proven to be better when we maintained a fixed amount of filters through the whole encoding-decoding process. When we diverged from this approach our network would only blur the fingerprint region.

This network is intended to receive square images as inputs, so before the training process, all images had their smallest dimensions padded to equal their highest dimension.

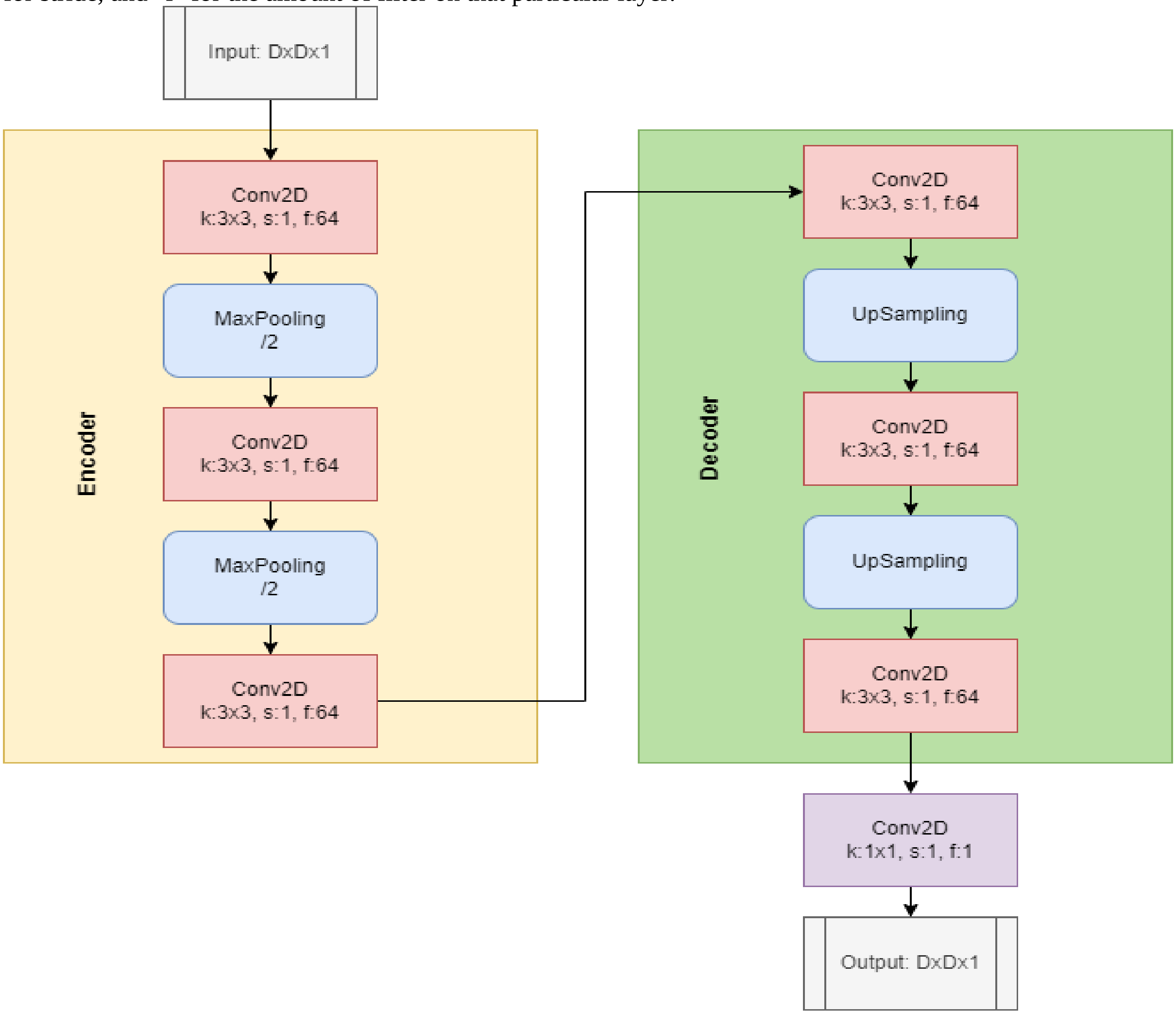
A process of data augmentation was made, including horizontal and vertical mirroring, rotations, and shifts. The network output is not binarized, so after the network provides an enhanced image we apply a simple threshold filter where values inferior to 0.5 are forced to 0 and values equal or higher are forced to 1.

It is worth noting that to train our network with multiple datasets of varying image dimensions, we do not specify any input size. Other hyperparameters chosen for our architecture are; Adam optimizer (KINGMA, 2014), binary cross-entropy loss function, and a batch size that varies between 8 to 32 depending on the input image size. The ReLU activation function was used in all layers. (NAIR, 2010).

Most of the FVC datasets have about 800 images. We randomly chose 500 of them for training, 150 for validation, and 150 for testing. We also did a 450, 143, 143 split for the dataset with 736 images and a 1000, 340, 340 split for the datasets with 1680 images, for train, validation, and test respectively). Once the data augmentation was done the 500 train images resulted in between 10000 to 30000 fingerprint images for training, this variation happened due to the different sizes of the image, and 150 for validation and 150 for testing. The networks were trained for 30 epochs, the point where the network validation accuracy stopped improving. Learning rate decay was also employed, reducing the learning rate by a factor of 10 after the accuracy stopped improving for a few epochs.

We trained one model on each of the FVC datasets, and each of these models was then tested on each of the datasets. When tested on the same trained dataset, a model would only process its test split. When evaluating other datasets, all images were taken into consideration since none of them were seen during the training process. Our ground truth images were provided by NeuroTechnology's third-party tool, since, to the best of our knowledge, there is not an official ground truth set of images for these datasets.

Figure.1 - Diagram highlighting our proposed network architecture, where "k" stands for kernel size, "s" for stride, and "f" for the amount of filter on that particular layer.

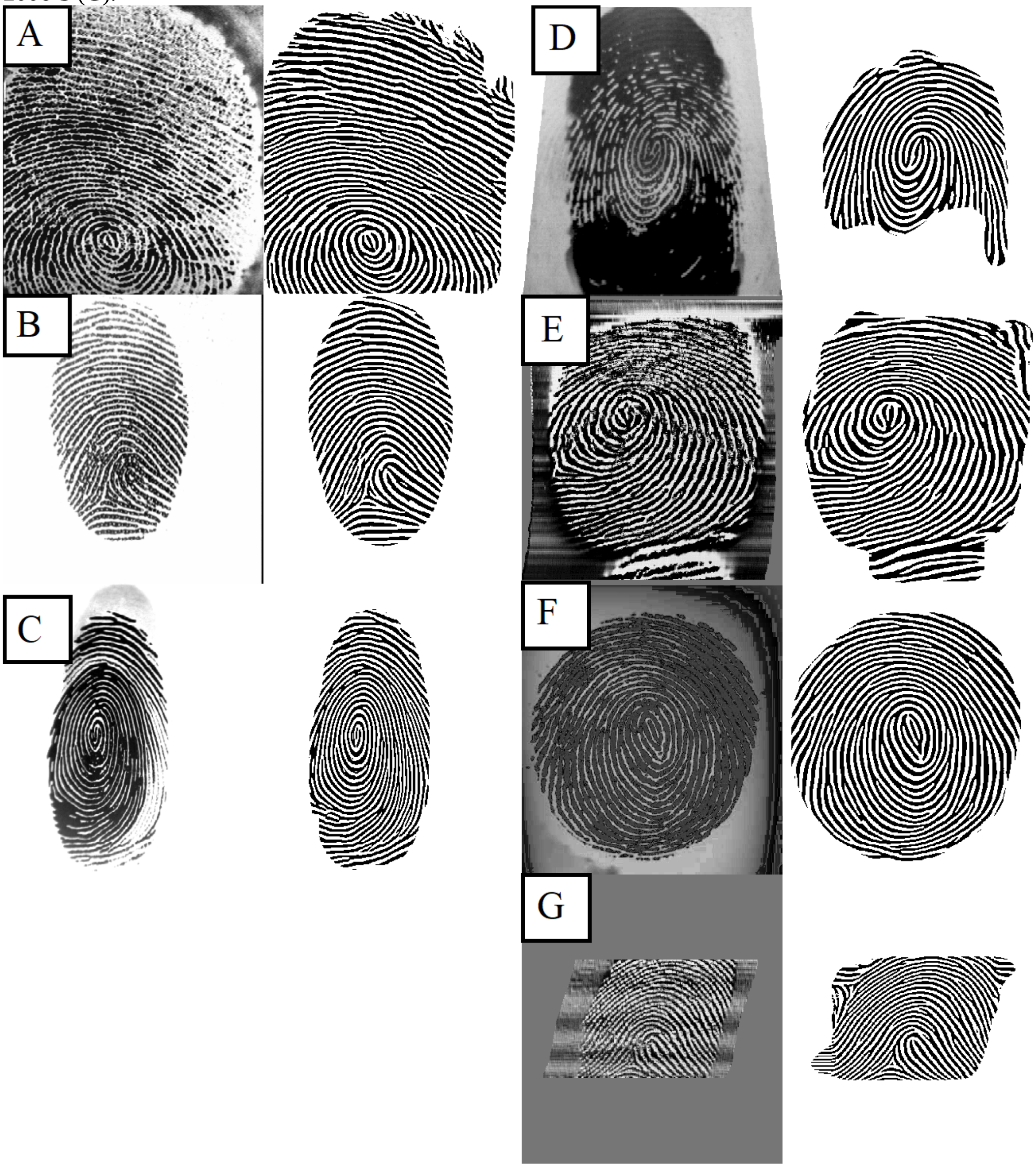


Source: Authors

4 RESULTS AND COMMENTARIES

The experiments were made using all publicly available datasets from the FVC competition, namely the 2000-3, 2002-1, 2004-1, 2004-2, 2004-3, 2006-2, 2006-3. The datasets are well-known for covering a good range of different types of noisy images. As mentioned before, images used as ground truth for the network were generated by using the results of enhancement from NeuroTechnology API, although the images are not perfectly enhanced. Figure 2 shows a side-by-side comparison of the input image and its enhanced ground truth counterpart. None of the test images were seen during the training process. The main metric used in the experiments was pixel-by-pixel accuracy.

Figure.2 - A side by side comparison of an input image (left) and its enhanced output (right). A single sample is shown for each dataset: 2000-3 (A), 2002-1 (B), 2004-1 (C), 2004-2 (D), 2004-3 (E), 2006-2 (F), 2006-3 (G).

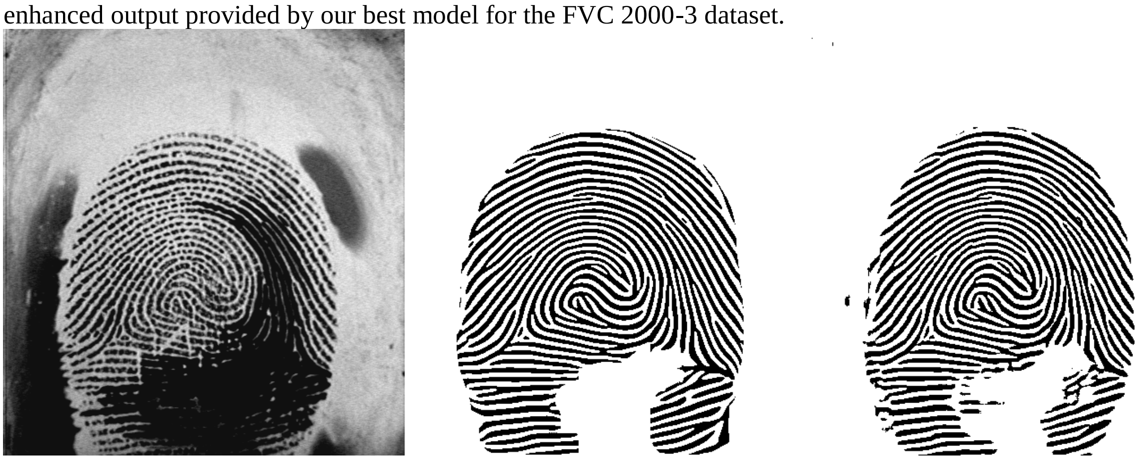


Source: Authors

The same network architecture was trained using each of the seven datasets. The models generated from this training process were tested against all FVC datasets. The tests were carried out using the as detailed in the previous section, it is worth noting that the training, validation, and test sets are disjoint. A brief comment about the variations on each of the datasets and the results achieved will now be presented.

2000-3 Dataset. The 2000-3 dataset has shown the worst results from our tests in all trained models. Figure 3 shows an example of the input image as well as an enhanced goal for this particular dataset and output when trained on the 2000-3 model. We can identify some areas where the ridges are lost, and all that remains is a very noisy black area. Another factor that severely affected our results on this particular dataset is that the accuracy is not computed over the Region of Interest (RoI). The artifacts around the actual fingerprint are also taken into consideration after the enhancement. We may notice that even the ground truth cannot fully restore the fingerprint. Our best result for such a test was 88.158% of per-pixel accuracy on the model trained using the FVC 2000-3 dataset itself.

Figure.3 - From left to right: An example of an input image, the image used as ground truth, and the enhanced output provided by our best model for the FVC 2000-3 dataset.



Source: Authors.

2002-1 and 2004-1 Datasets. Both these datasets have a pretty good image quality overall as shown in Figure 4. Some samples can be anomalies and have slightly degraded areas, although these datasets are quite different from the others. The evaluation on the 2002-1 and 2004-1 datasets showed the best accuracy per-pixel accuracy of 98.261% and 98.624% respectively. The models that provided these results were trained on 2002-1 and 2004-2 datasets.

Figure.4 - From left to right: An example of an input image, the image used as ground truth, and the enhanced output provided by our best model for the FVC 2002-1 dataset (top) and 2004-2 dataset (bottom).

Source: Authors.

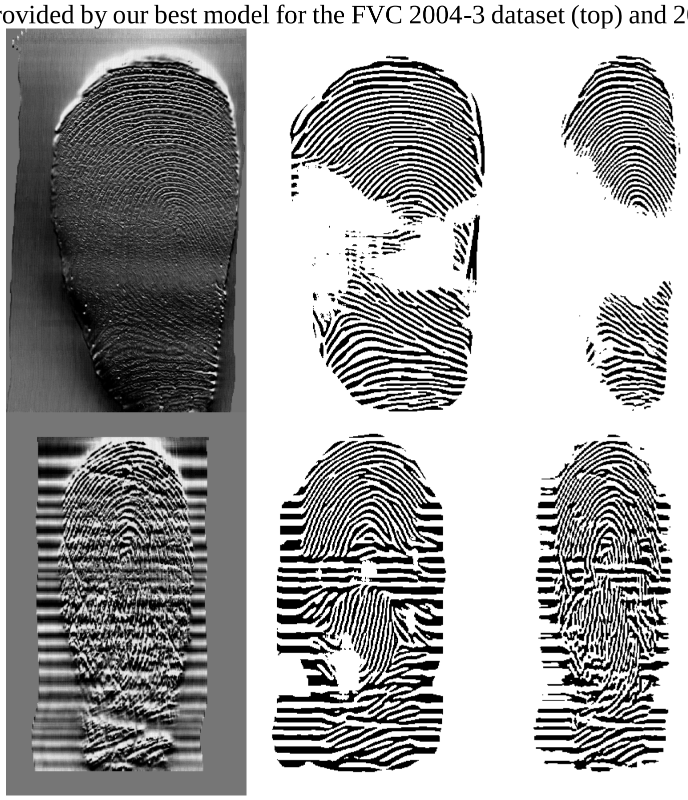
2004-2 and 2006-2 Datasets. Figure 5 shows samples from the 2004-2 and 2006-2 datasets. Both of the image sets have a gray background and some unique features. For the 2004-2 dataset, there are many areas where the ridge is interrupted and the presence of a darker square-like region in the middle of almost all images. The best result in this dataset was 95.852% per pixel accuracy with the model trained on this own dataset. We may observe our network was able to reconstruct areas where the ground truth was not. The 2006-2 images also have areas where the fingerprint patterns are missing; though, this is much more due to a very faint capture. The best result on this dataset was 98.168%, on the model trained by the 2006-2 dataset.

2004-3 and 2006-3 Datasets. One of the common problems of any biometric verification method is sensor error, i.e., a damaged capture device will hinder the image quality or insert noise in it. Both these datasets have images with different degrees of sensor noise in it, and as shown in Figure 6, in some cases our ground truth enhanced an image is either not able to restore the noised areas, or it considers the noise as the pattern of the ridges. The model trained on the 2004-3 dataset proved to be more generic and achieved the best results for both of these datasets, 97.782% and 97.991% for the 2004-3 and 2006-3 respectively. Just like in the 2004-2 dataset, our model was able to better differentiate fingerprint from sensor noise on the image of the 2006-3 dataset.

Figure.5 - From left to right: An example of an input image, the image used as ground truth, and the enhanced output provided by our best model for the FVC 2004-2 dataset (top) and 2006-2 dataset (bottom).

Source: Authors

Figure.6 - From left to right: An example of an input image, the image used as ground truth, and the enhanced output provided by our best model for the FVC 2004-3 dataset (top) and 2006-3 dataset (bottom).



Source: Authors

Overall report. We summarized the accuracy results in Table 1. The results show that a large majority of the datasets achieved over 90% per pixel accuracy in every dataset, and many results reached over 97% per-pixel accuracy. Table 1 also shows that the 2000-3 dataset has proven to be the most complex for our model since the best model on this dataset only achieved 88.158% per-pixel accuracy. Figure 7 shows that once the enhancement was not applied in a region of interest models usually took background as relevant fingerprint information. Therefore, artifacts were created outside the fingerprint. Such effect happens in some datasets, especially in the 2000-3 image set.

The average accuracy and performance results achieved by each model are summarized in Table 2 and Table 3, respectively. The average results showed that the model trained on the 2000-3 dataset generalized better for others. As previously discussed, we believe this happened because the model was trained with the most challenging dataset, so easier datasets were trivial for the models. The IPS results are based on a computer with an Intel i7-8700 3.20Ghz and a Titan X GPU.

Table.1 - Accuracy results of all trained models evaluating each of the datasets.

		Test Set						
		2000-3	**2002-1**	**2004-1**	**2004-2**	**2004-3**	**2006-2**	**2006-3**
Train Set	**2000-3**	88.15%	97.37%	98.58%	94.13%	96.49%	97.28%	97.29%
	2002-1	84.06%	98.26%	98.26%	94.39%	90.29%	93.71%	95.52%
	2004-1	76.08%	97.18%	98.56%	92.29%	90.95%	93.28%	93.07%
	2004-2	82.04%	97.29%	98.62%	95.85%	94.11%	96.74%	96.44%
	2004-3	84.35%	94.36%	96.98%	82.41%	97.78%	95.22%	97.99%
	2006-2	85.87%	93.85%	97.20%	93.89%	96.66%	98.16%	97.38%
	2006-3	85.65%	94.96%	97.46%	87.60%	97.29%	95.67%	97.64%

Table.2 - Table showing for each model, average per pixel accuracy results on all datasets.

Train Set	2000-3	2002-1	2004-1	2004-2	2004-3	2006-2	2006-3
Avg. Acc.	95.61%	93.50%	91.63%	94.44%	92.73%	94.72%	93.75%

Table.3 - Table showing average accuracy per pixel results of all models for each of the datasets tested, image dimensions, and processing time (images per second).

Dataset	2000-3	2002-1	2004-1	2004-2	2004-3	2006-2	2006-3
Avg. Acc.	83.74%	96.18%	97.95%	91.51%	94.80%	95.72%	96.48%
Input Size	478x478	388x388	640x640	364x364	480x480	492x492	500x500
IPS	142	200	38	250	142	125	125

5 CONCLUSIONS

It is a fact that nowadays convolutional neural networks are being used to solve a large array of problems, especially when narrowing to the imaging processing field. Using the most suitable architecture for the image enhancement task a fully convolutional autoencoder architecture was employed to restore the quality of noisy fingerprint images on a per-pixel level of accuracy.

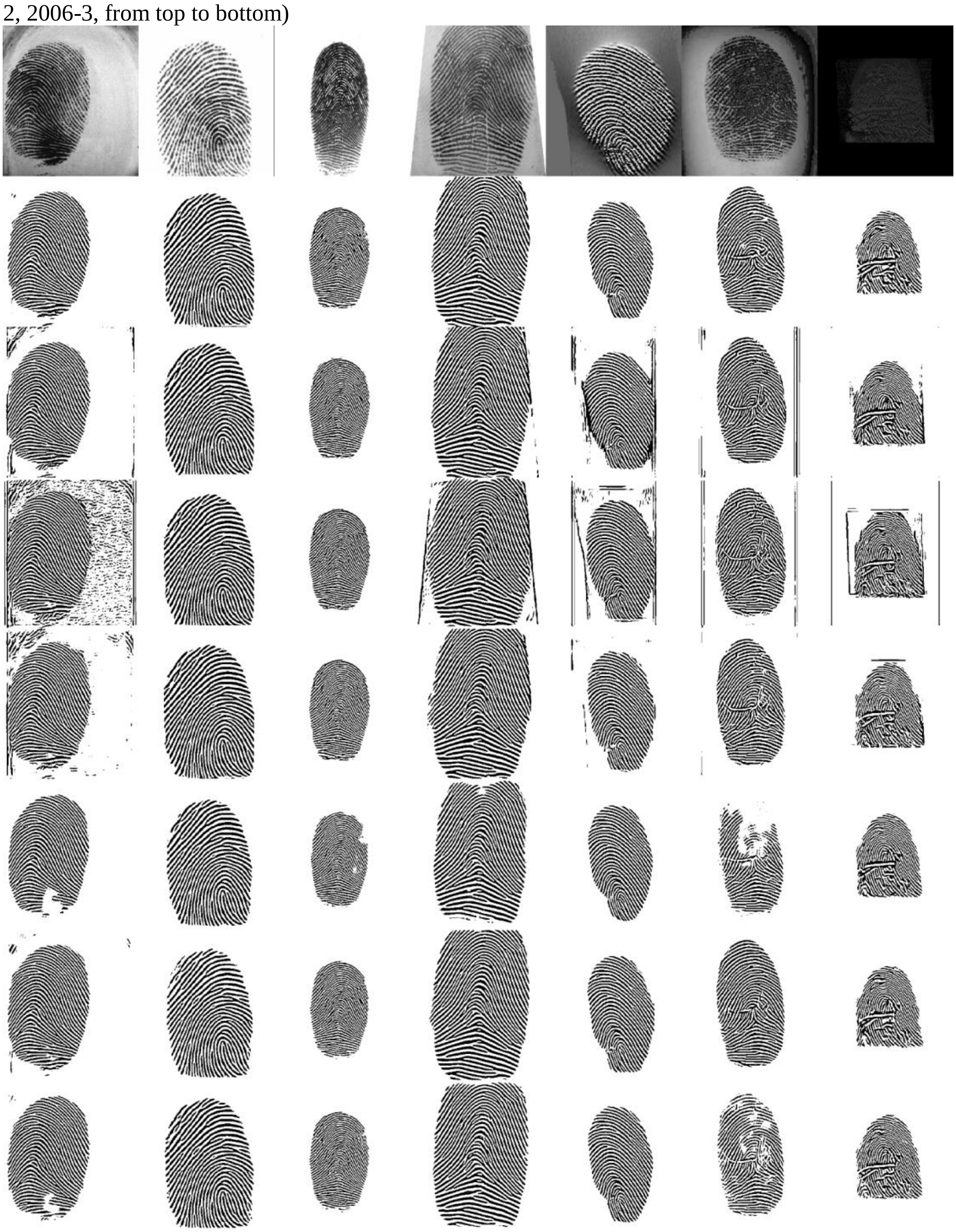
The attained results show that it is possible to greatly enhance the quality of fingerprints taking the full image as input of the network however it is possible to see some lost minutiae. That happens because the images have a high resolution, hence when trained the neural network tends to lose some of the minutiae as a trade-off to better enhance the quality of the image as a whole.

To solve these issues some alternatives may give a better result, such as feeding patches of an image to the network rather than the full fingerprint, creating a custom and more complex loss function that takes into consideration the loss in the vicinity of minutiae and not only a pixel by pixel computation of the loss.

In case a train by patches approach is chosen an overlap of the patches is needed to reduce image blocking effects. Such a method would greatly increase the computational cost and processing time since a larger number of pixels would be processed. The custom loss function that is calculated in the minutiae vicinity could increase the training time while maintaining the same evaluation time.

To sum up, the proposed architecture has shown good results in most of the datasets tested, as the goal of this paper is to provide a good enhancing architecture that can deliver results faster than if the network was fed by patches while not losing so much on the minutiae regions. Going forward our main focus will be the implementation of the custom loss function calculated in the minutiae vicinity previously mentioned.

Figure.7 - In the first row each image is an original input from the FVC datasets (2000-3, 2002-1, 2004-1, 2004-2, 2004-3, 2006-2, 2006-3, from left to right respectively). Each subsequent row presents our network enhanced output for the input image for each model trained (2000-3, 2002-1, 2004-1, 2004-2, 2004-3, 2006-2, 2006-3, from top to bottom)



Source: Authors.

REFERENCES

ABHYANKAR, Aditya; SCHUCKERS, Stephanie. Fingerprint liveness detection using local ridge frequencies and multiresolution texture analysis techniques. In: **2006 international conference on image processing**. IEEE, 2006. p. 321-324.

ALSMIRAT, Mohammad A. et al. Impact of digital fingerprint image quality on the fingerprint recognition accuracy. **Multimedia Tools and Applications**, v. 78, n. 3, p. 3649-3688, 2019.

ARIF, Annatoma; LI, Tuo; CHENG, Chi-Hao. Blurred fingerprint image enhancement: algorithm analysis and performance evaluation. **Signal, Image and Video Processing**, v. 12, n. 4, p. 767-774, 2018.

GALBALLY-HERRERO, J. et al. On the vulnerability of fingerprint verification systems to fake fingerprints attacks. In: **Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology**. IEEE, 2006. p. 130-136.

HONG, Lin; WAN, Yifei; JAIN, Anil. Fingerprint image enhancement: algorithm and performance evaluation. **IEEE transactions on pattern analysis and machine intelligence**, v. 20, n. 8, p. 777-789, 1998.

KINGMA, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, 2012.

MARASCO, Emanuela; ROSS, Arun. A survey on antispoofing schemes for fingerprint recognition systems. **ACM Computing Surveys (CSUR)**, v. 47, n. 2, p. 1-36, 2014.

NAIR, Vinod; HINTON, Geoffrey E. Rectified linear units improve restricted boltzmann machines. In: **Icml**. 2010.

RAMOS, Ramon et al. Analysis and improvements of fingerprint enhancement from gabor iterative filtering. In: **2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. IEEE, 2018. p. 266-273.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: **International Conference on Medical image computing and computer-assisted intervention**. Springer, Cham, 2015. p. 234-241.

SAJJAD, Muhammad et al. CNN-based anti-spoofing two-tier multi-factor authentication system. **Pattern Recognition Letters**, v. 126, p. 123-131, 2019.

SCHUCKERS, Stephanie AC. Spoofing and anti-spoofing measures. **Information Security technical report**, v. 7, n. 4, p. 56-62, 2002.

SVOBODA, Jan; MONTI, Federico; BRONSTEIN, Michael M. Generative convolutional networks for latent fingerprint reconstruction. In: **2017 IEEE International Joint Conference on Biometrics (IJCB)**. IEEE, 2017. p. 429-436.

TANG, Yao; GAO, Fei; FENG, Jufu. Latent fingerprint minutia extraction using fully convolutional network. In: **2017 IEEE International Joint Conference on Biometrics (IJCB)**. IEEE, 2017. p. 117-123.

VINCENT, Pascal et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. **Journal of machine learning research**, v. 11, n. 12, 2010.

YANG, Jianwei et al. A modified Gabor filter design method for fingerprint image enhancement. **Pattern Recognition Letters**, v. 24, n. 12, p. 1805-1817, 2003.

ZAHEDI, Morteza; GHADI, Ozra Rostami. Combining Gabor filter and FFT for fingerprint enhancement based on a regional adaption method and automatic segmentation. **Signal, Image and Video Processing**, v. 9, n. 2, p. 267-275, 2015.

ZHANG, Kai et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. **IEEE transactions on image processing**, v. 26, n. 7, p. 3142-3155, 2017.