Universidade de Pernambuco - UPE
Escola Politécnica de Pernambuco - POLI
Mestrado em Engenharia da Computação

A Deep Learning Approach to Generate Offline Handwritten Signatures Based on Online Samples

Victor Kléber Santos Leite Melo

vkslm@ecomp.poli.br

Advisor: Prof. Dr. Byron Leite Dantas Bezerra
Co-Advisor: Prof. Dr. Giuseppe Pirlo

August 23, 2017

Introduction

Biometric technology is used in several security applications for personal authentication.

Introduction

Biometric technology is used in several security applications for personal authentication.

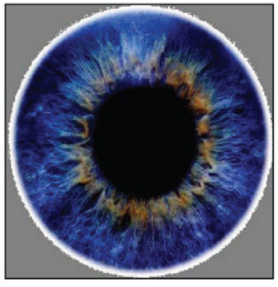
(a) Fingerprint (b) Iris (c) Handwritten Signature (d) Voice

Figure 1: Some biometric traits used for person authentication.

Introduction

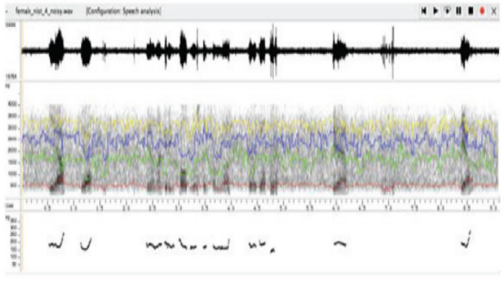
Biometric technology is used in several security applications for personal authentication.

(a) Fingerprint



(b) Iris

(c) Handwritten Signature



(d) Voice

Figure 1: Some biometric traits used for person authentication.

Biometry Categories

- Physiological - biological characteristics such as fingerprint, palm print, iris, face.
- Behavioral - individual acquired traits such as voice pattern and handwritten signature.

Handwritten Signature

Figure 2: Handwritten Signatures are widespread biometry.

Handwritten Signature (2)

The Handwritten Signature widespread can be attributed to[1]:

- Signature acquisition is easy and non-invasive;
- Most individuals are familiar with its use in their daily life;
- Signatures can be employed as a sign of confirmation in a wide variety of documents, namely, bank checks, identification documents and a variety of business certificates and contracts.

[1] Donato Impedovo and Giuseppe Pirlo (2008). "Automatic signature verification: the state of the art". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38.5, pp. 609–635.

Overview of a Handwritten Signature Verification System

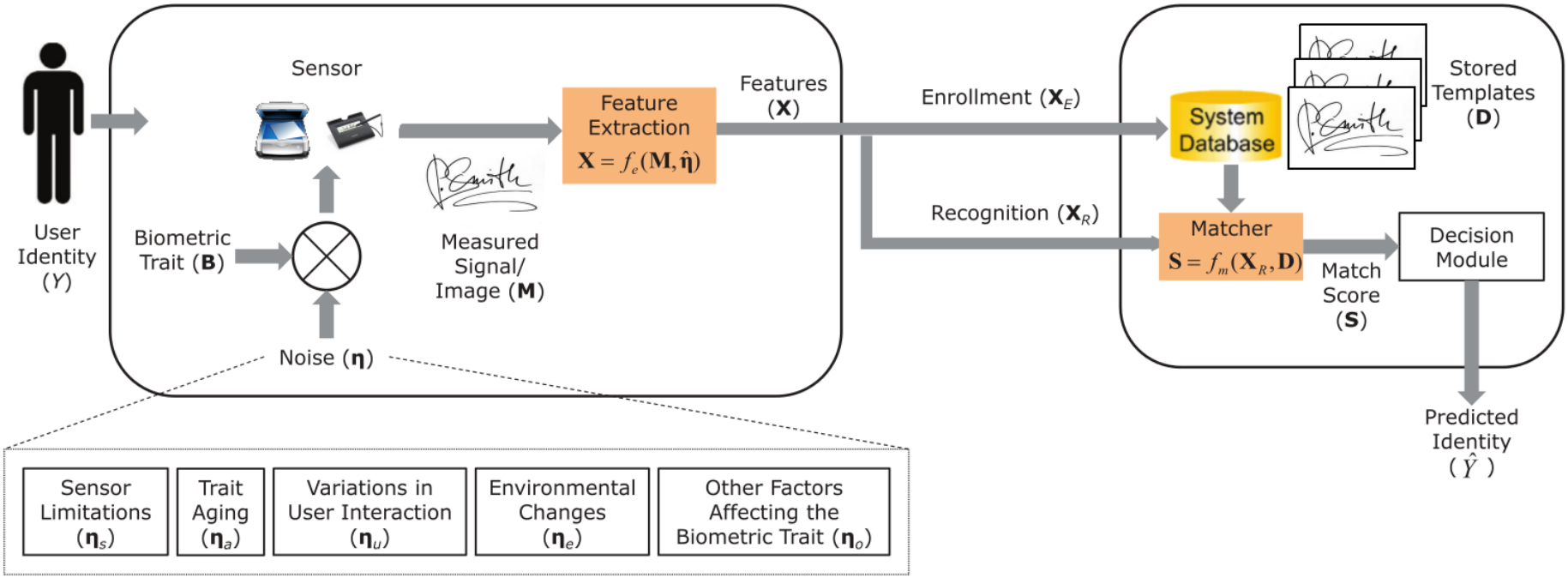
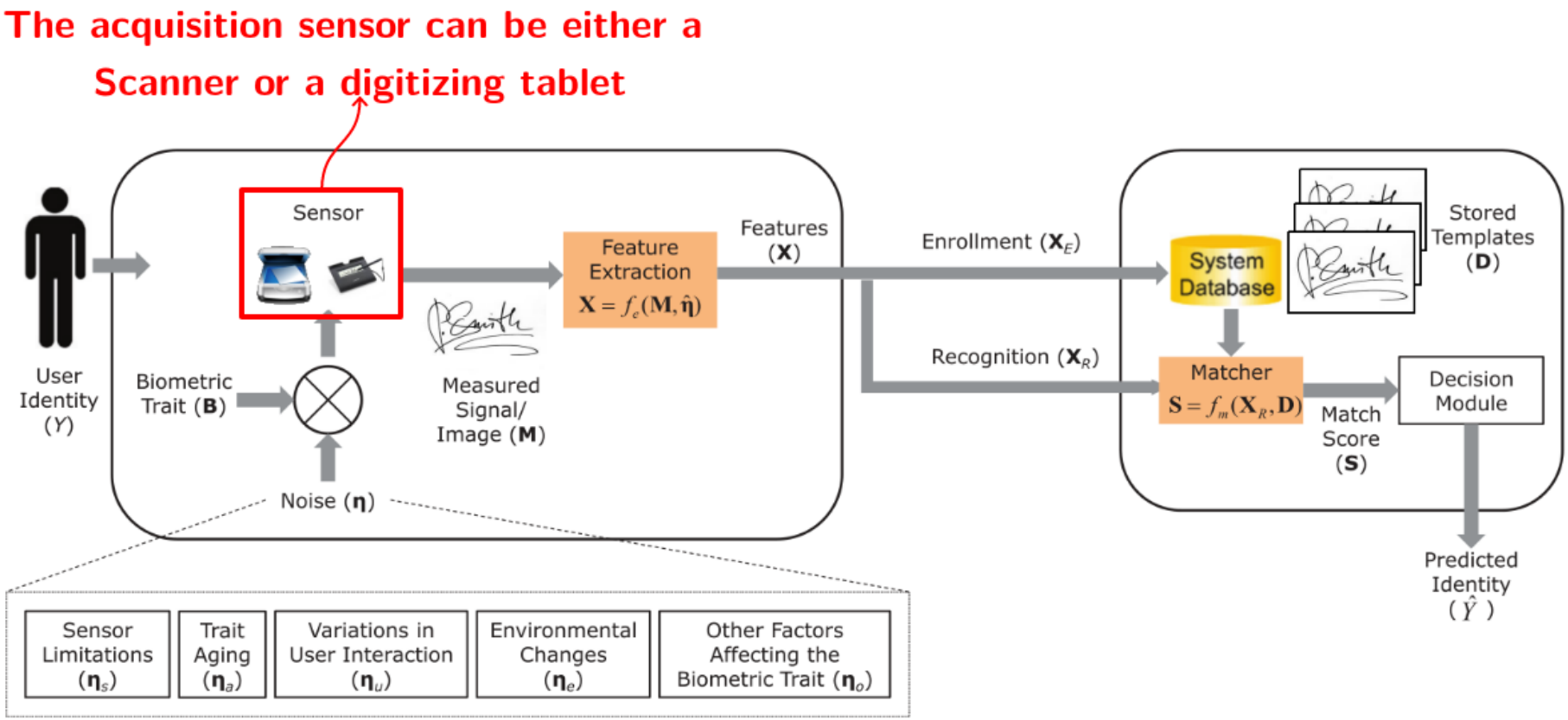


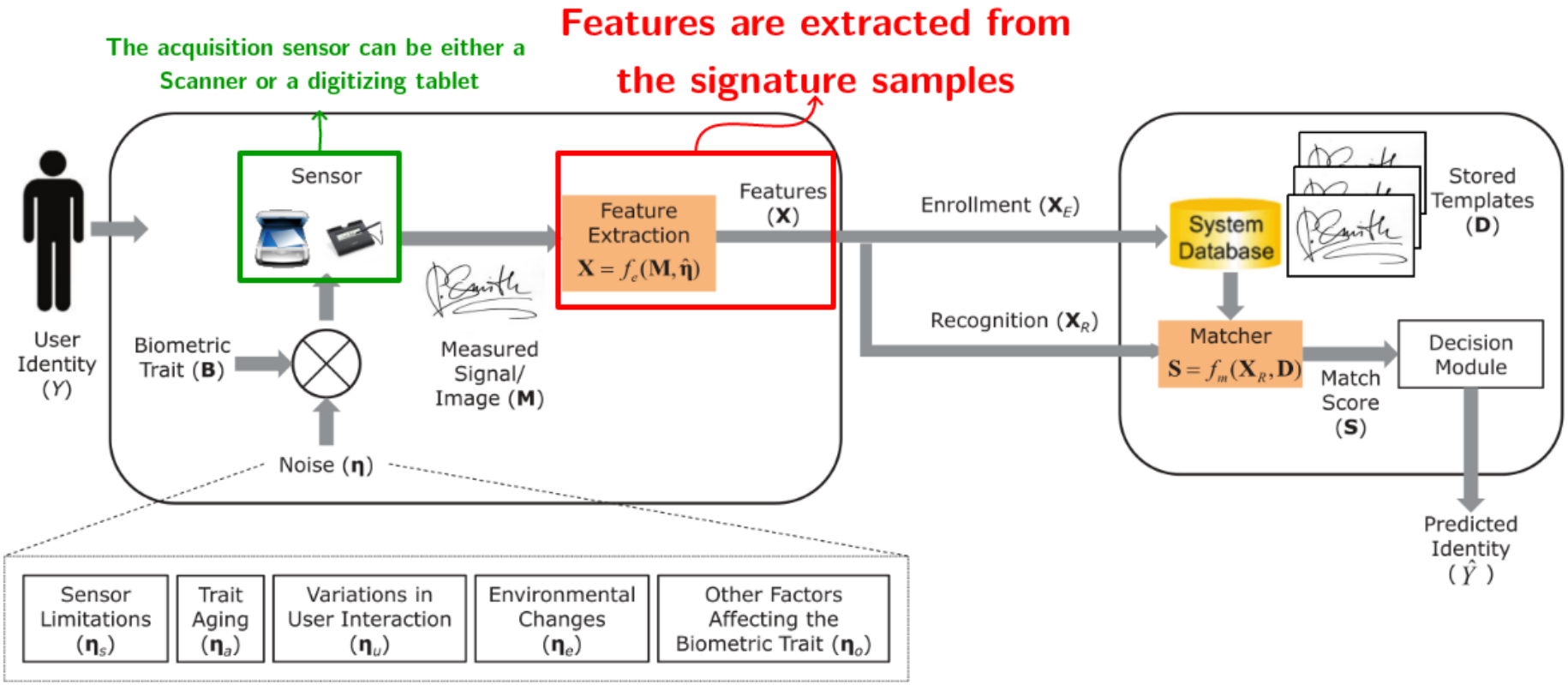
Figure 3: Overview of a typical handwritten signature based system. Figure adapted from (Jain, Nandakumar, and Ross, 2016)[2].

[2] Anil K Jain, Karthik Nandakumar, and Arun Ross (2016). "50 years of biometric research: Accomplishments, challenges, and opportunities". In: *Pattern Recognition Letters* 79, pp. 80–105.

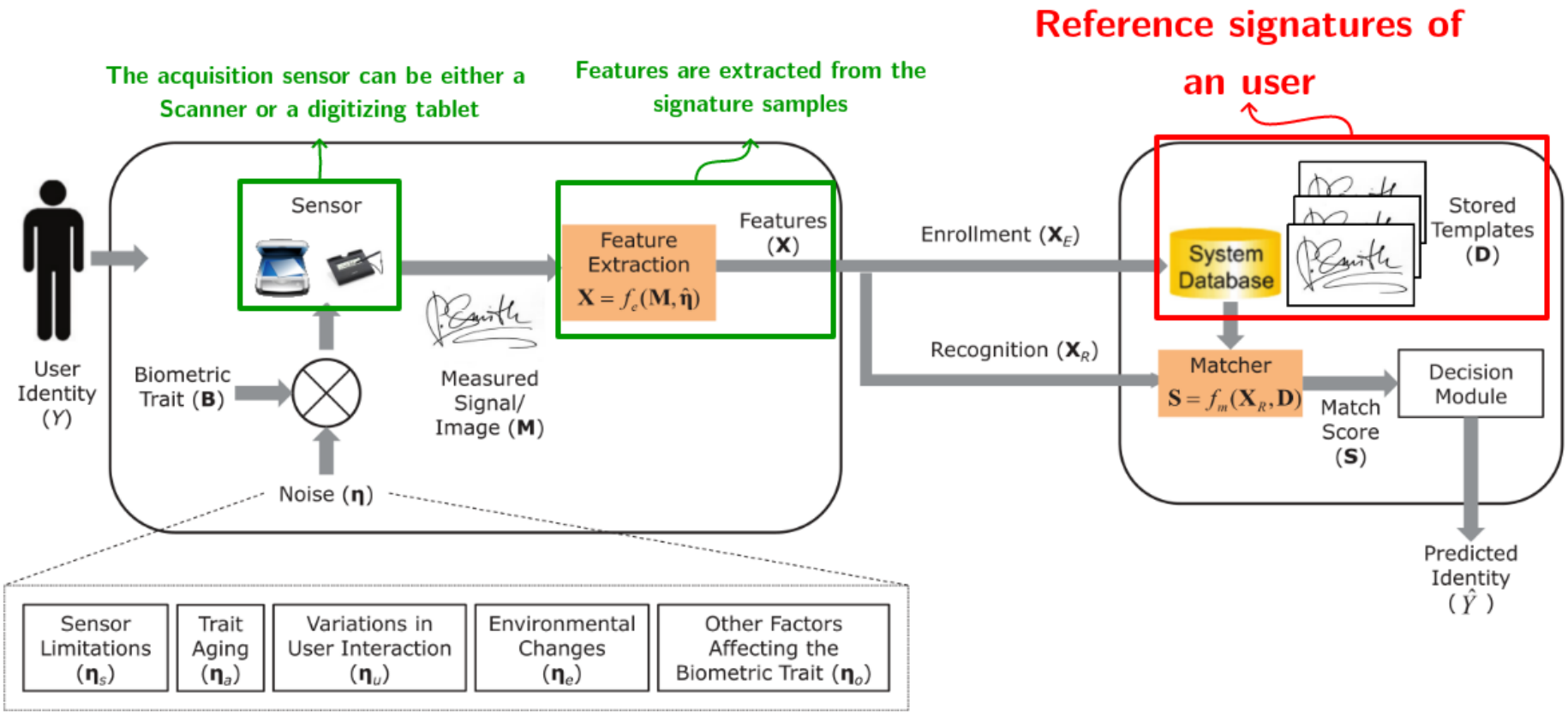
Overview of a Handwritten Signature Verification System



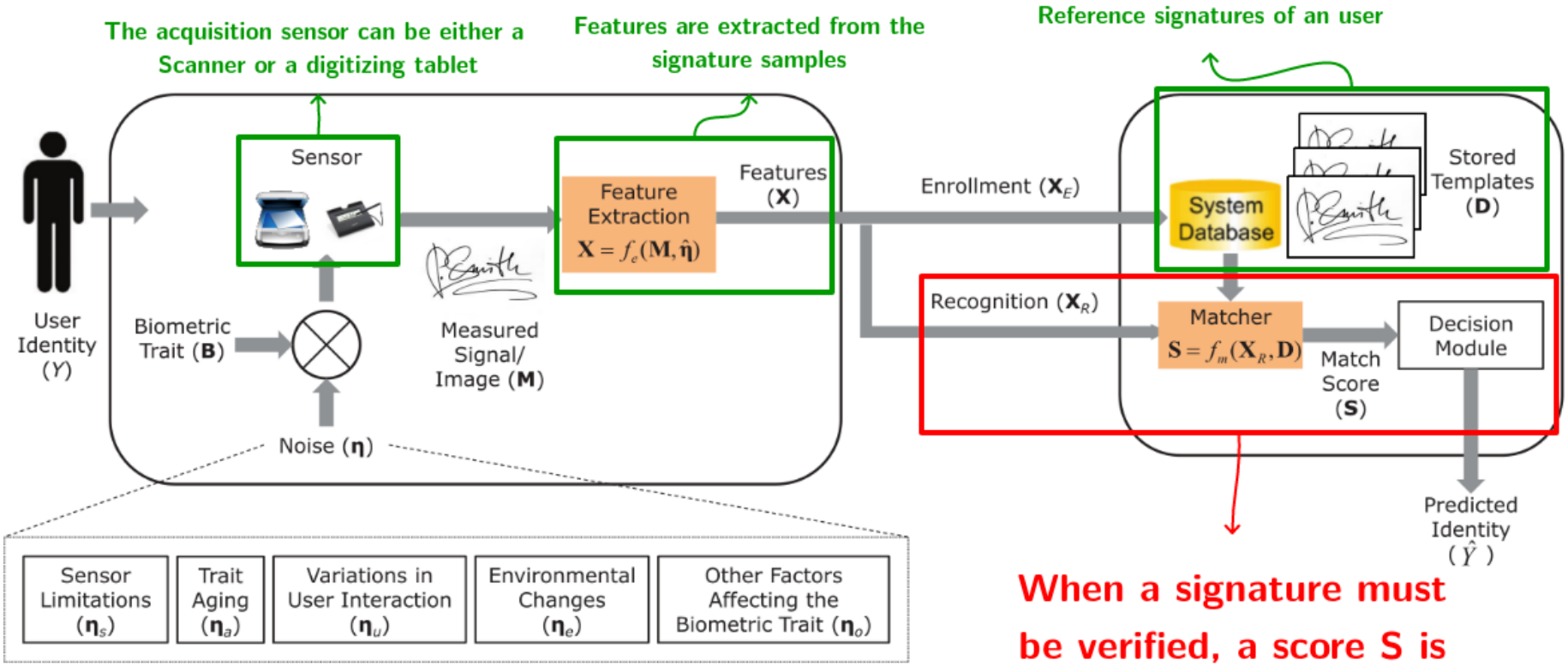
Overview of a Handwritten Signature Verification System



Overview of a Handwritten Signature Verification System



Overview of a Handwritten Signature Verification System



When a signature must be verified, a score S is obtained according to the similarity of the questioned sample to the claimed user references.

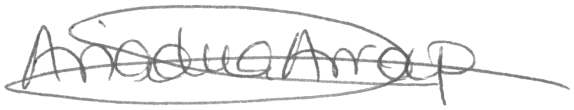
Online vs Offline Handwritten Signatures

OFFLINE -> STATIC

An optical scanner is used to obtain the signature directly from the pen on the paper, and only the digital image of the signature is available.

ONLINE -> DYNAMIC

Data is stored during the writing process and consists of a temporal sequence of the two-dimensional coordinates (x, y) of consecutive points.



(a) A signature scanned from paper

(b) Digitizing tablet Wacom STU-500

Figure 4: Different signature acquisition methods.

Online vs Offline Handwritten Signatures (2)

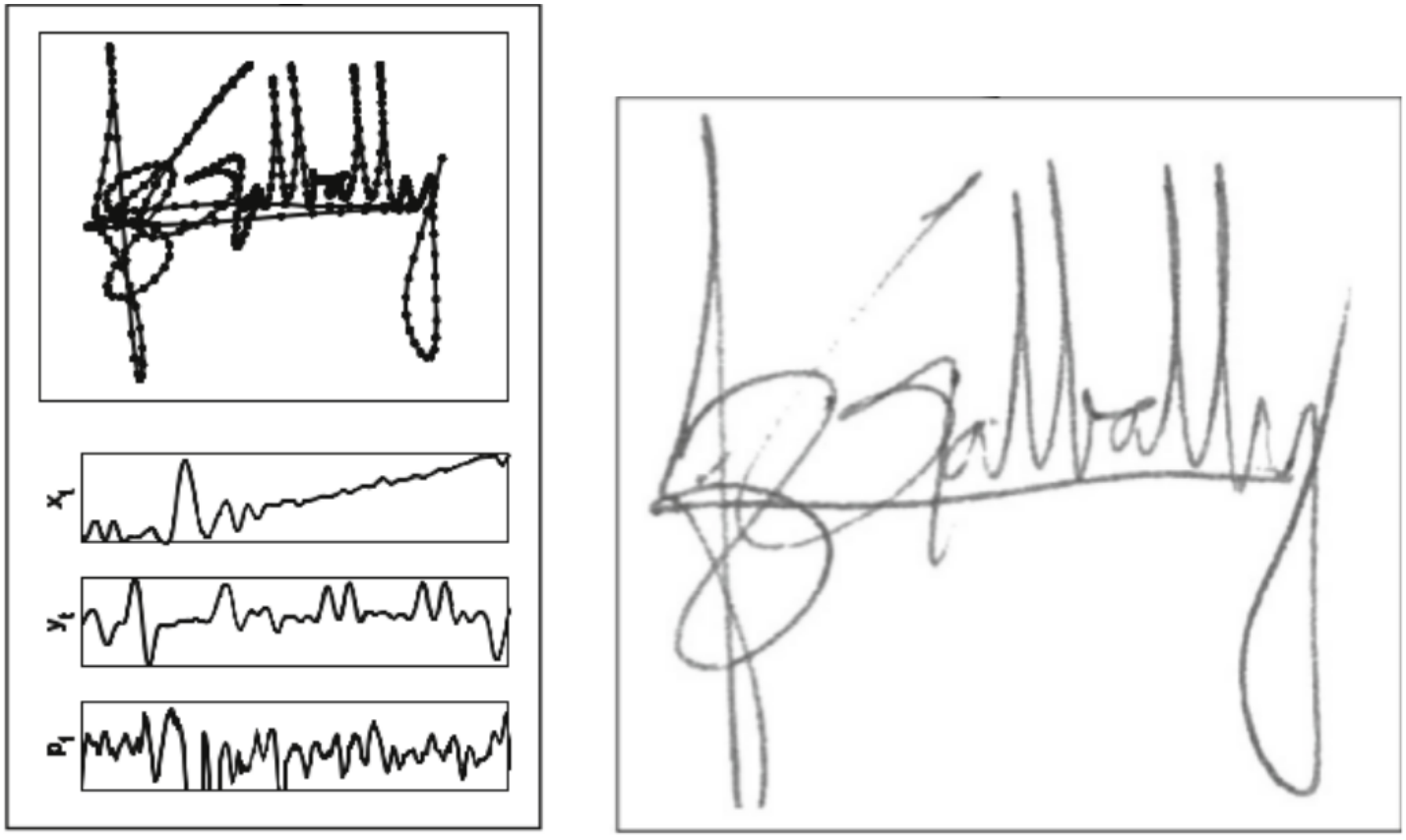


Figure 5: An online and the respective offline signature sample. Extracted from (Galbally et al., 2015)

Online vs Offline Handwritten Signatures (3)

ONLINE MODALITY

Does not convey information about:

- The overall shape of the signature;
- The width of the strokes;
- The texture of the ink on the paper.

Online vs Offline Handwritten Signatures (3)

ONLINE MODALITY

Does not convey information about:

- The overall shape of the signature;
- The width of the strokes;
- The texture of the ink on the paper.

OFFLINE MODALITY

- Lost all dynamic information about the signature;
- Dynamic features can only be inferred from a static image (Nel, Du Preez, and Herbst, 2005).

Handwritten Signature Intra-personal variability

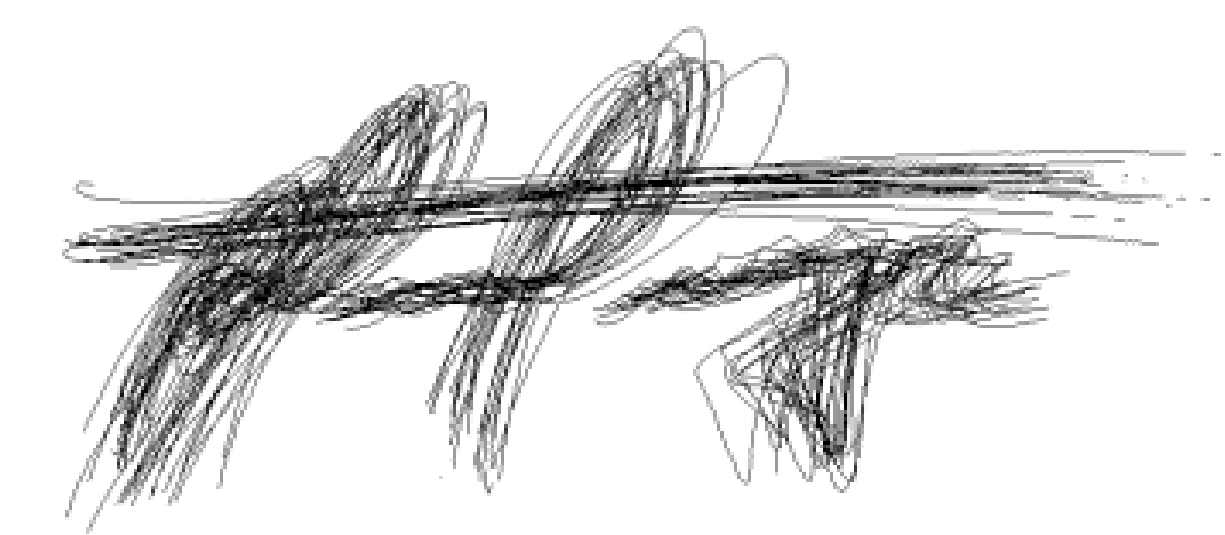


Figure 6: Superimposed genuine signatures of the same writer. A high intra-personal variability can be noticed. Extracted from (Hafemann, Sabourin, and Oliveira, 2015).

Handwritten Signature Inter-personal variability

The first column of signatures are genuine references, the following three samples are questioned signatures [3]. Signature images extracted from (Ortega-Garcia et al., 2003).

[3] From left to right, top to bottom (F means Forgery and G means Genuine): FGF FFG GFF

Types of Forgeries

In the field of signature verification, forgeries are usually classified into two types (Impedovo and Pirlo, 2008).

- The first one is the random forgery which is created in a situation which an impostor who has no information about the person or the shape of the original signature tries to verify the identity of a signer using his signature instead of the signature to be tested. The forger does not attempt to simulate or trace a genuine signature.
- The second type is the skilled forgery, in which the forger tries and practices imitating as closely as possible the static and dynamic information of the genuine signature model. The forger has access for both the user's name and signature and tries to reproduce it with a similar intra-class variability.

Data in Signature Verification

The performance of signature verification systems can be improved by increasing the number of samples in the training dataset. The amount of data available for each user is often insufficient in real applications. During the enrollment phase, users are often required to supply only a few samples of their signatures. In other words, even if there is a significant number of users enrolled in the system, a classifier needs to perform well for a new user, for whom only a small set of samples are available. Since the acquisition and distribution of real signatures arise legal and privacy concerns (Diaz-Cabrera et al., 2014), the use of realistic synthetic signatures could be regarded as a good alternative.

Synthetic Samples

The use of realistic synthetic signatures could be regarded as a good alternative. As a consequence, over the last years, several works on both online (Galbally Herrero et al., 2009; Galbally et al., 2012) and offline (Ferrer, Diaz-Cabrera, and Morales, 2013; Ferrer et al., 2013) signature synthesis have been carried out. These synthetically generated signatures show a similar behavior to real ones, thus enabling to enlarge existing databases and offering new possibilities for offline recognition.

Online to Offline

Some efforts have been performed on the generation of synthetic static data taking into account dynamic features during the synthesis process (Diaz-Cabrera et al., 2014). Among others, this type of synthesis approach presents some practical applications:

- generation of synthetic static samples to be fused with the original online signatures in order to improve the performance in an online verification scenario;
- enlargement of existing offline signature databases;
- development of systems capable of integrating both online and offline samples interchangeably, towards a unified signature biometry (Melo et al., 2017).

Goal

In contrast to the models proposed in the literature (Ferrer, Diaz-Cabrera, and Morales, 2013; Ferrer et al., 2013; Diaz-Cabrera et al., 2014), the approach presented in this work is designed under the perspective of supervised training.

"The goal of this work is to design an approach to generate synthetic offline handwriting signatures based on online data, modeling this problem as a supervised machine learning task, through a Deep Convolutional Neural Network, in order to enlarge offline signature datasets to improve offline signature verification systems recognition rates."

This statement is developed through the following actions:

- Creation and training of a Deep Neural Network model able to translate dynamic handwritten information into an offline manuscript
- Generation of an offline synthetic dataset based on a publicly available online signature dataset
- Comparison of the proposed approach's performance with a state-of-the-art method to evaluate the closeness of synthetic signatures with respect to real signatures.

Contents

Off-line Signature Synthesis Using On-line Samples

According to (Guest, Hurtado, and Henniger, 2013), although online samples can be stored as a time-series for use in any form of an AHSVS, there are situations where it might be needed to represent the data as an image reproducing the original static signature.

Neural Networks

The human brain is a complex, non-linear and parallel "computer" consisting of **millions of connected neurons** (Haykin et al., 2009).

Perceptron

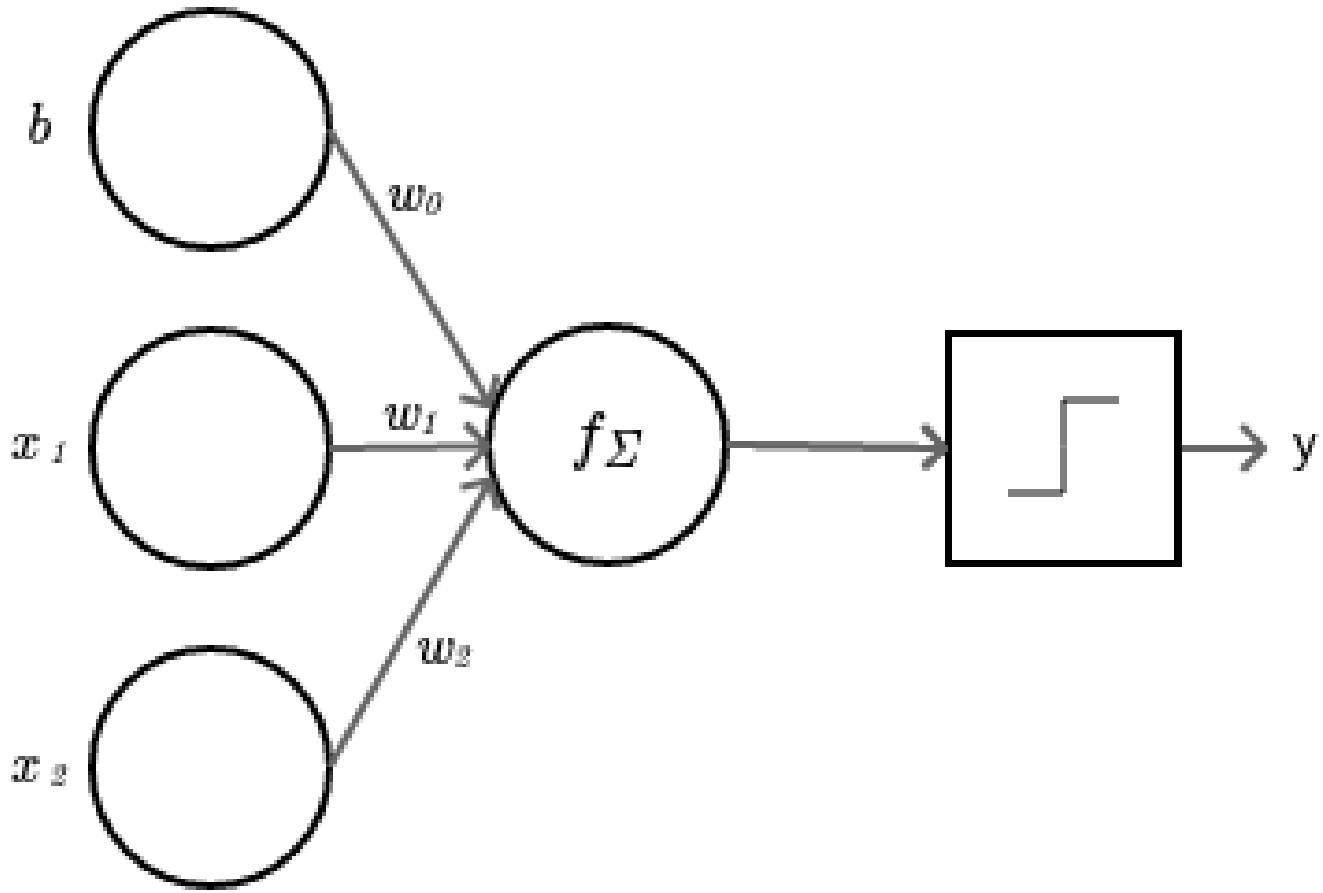


Figure 7: Perceptron representation. x_1 and x_2 represent the input signal, b the bias term, w_0, w_1, w_2 the weights, $f_{\sum}$ is the activation function (in this case, a step function) and the output signal is given by y.

Multilayer Perceptron

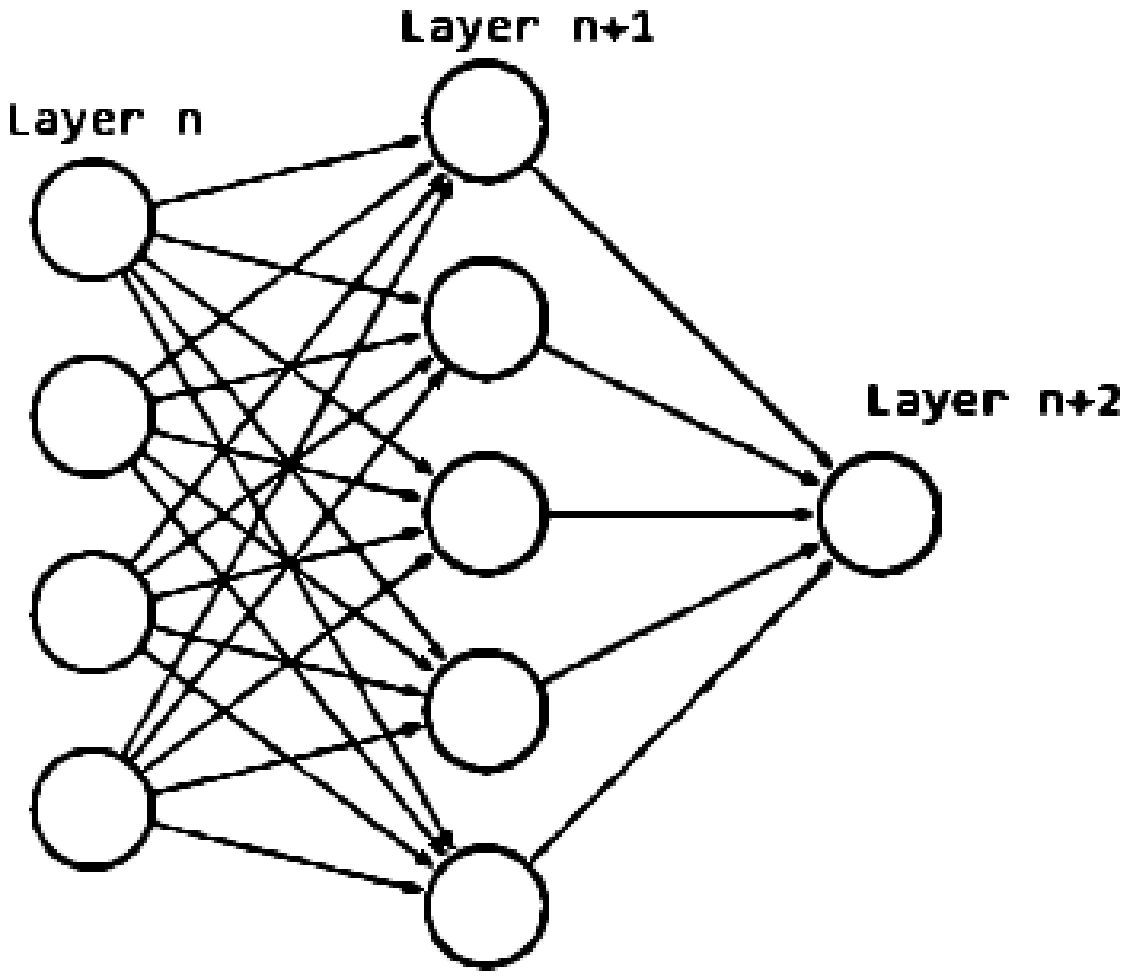


Figure 8: Each layer contains several perceptron units, which are then connected to units in the subsequent layer.

Multilayer Perceptron

An MLP can be thought of as a function that maps from input to output vectors, parameterized by the neuron connection weights. The output of a layer is calculated by applying the neuron activation function for all neurons on the layer, as noted below

$$y^{(l)} = f(W^{(l)} y^{(l-1)} + b^{(l)}) \tag{1}$$

where $W^{(l)}$ is a matrix of weights assigned to each pair of neurons from layer l and $l-1$, and $b^{(l)}$ is a vector of bias terms for each neuron in layer l. Calculating the output starting from the first hidden layer, up to the output layer is also referred to as the *forward propagation* phase.

Outline of the Proposed Method

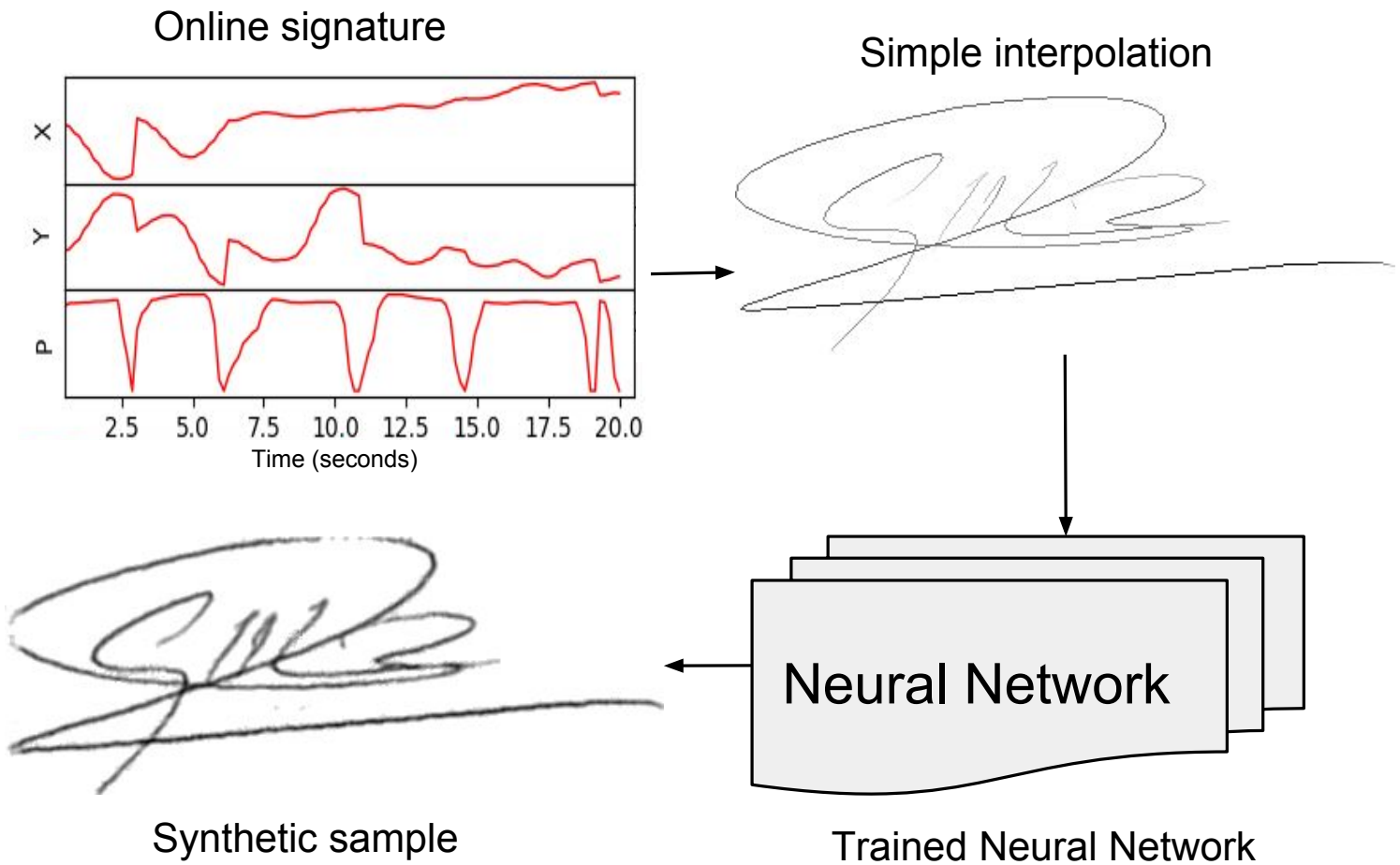


Figure 9: The proposed approach diagram and an example of the synthetic signature generation.

Training Data

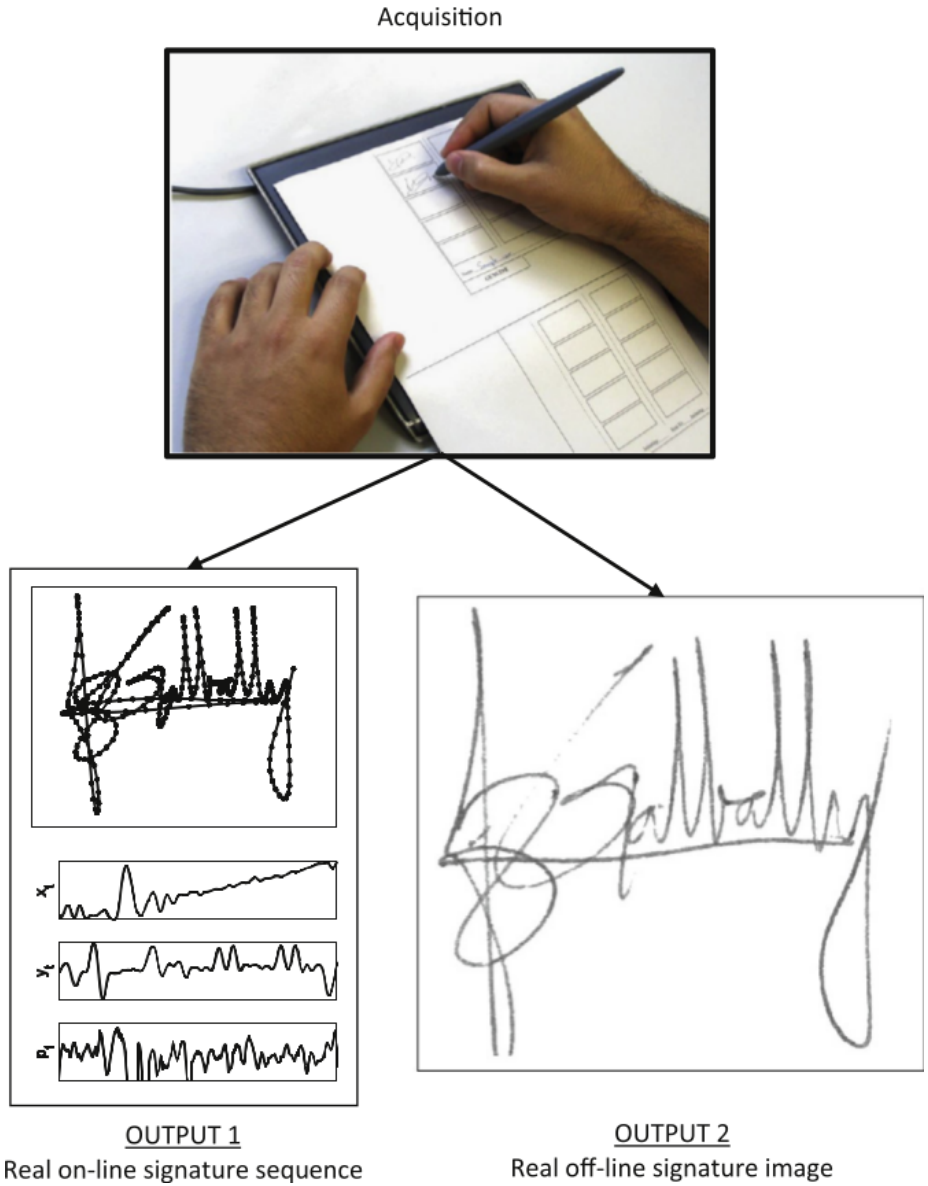


Figure 10: Figure extracted from (Galbally et al., 2015)

Training Data

Dual modal signature datasets (BiosecurID, Biomet, MyIdea, Sigcomp2009, Sigma, SigwiComp2013, SigWiComp2015) had not this characteristic satisfied. Both of the representations of the signatures do not match if we plot it in a single image.

Figure 11: A sample from the BiosecurID dataset. Here we can see that the online signature (interpolated in red) can not be projected in the respective offline version.

Training Data

The dual domain IRONOFF (Viard-Gaudin et al., 1999) handwriting dataset was thus used to train our model. Besides acquiring both domains of the handwriting manuscript, the online data is mapped to the same coordinate system of the offline data.
Using this dataset, we make the fair assumption that a handwritten signature is a manuscript. With that in mind, we expect that even if the network was trained on handwriting manuscripts, it would work for online signatures to generate its static version.
The IRONOFF dataset contains a total of 23000 mapped online and offline samples of the manuscripts. The offline handwriting signals have been sampled with a spatial resolution of 300 dots per inch (DPI), with 8 bits per pixel (256 gray level).

Preprocessing

(a) input

(b) ground truth

Figure 13: Preprocessed data used during the training phase. (a) is the interpolated online sample, used as input and (b) is the expected prediction from the neural network, the ground truth.

Training results

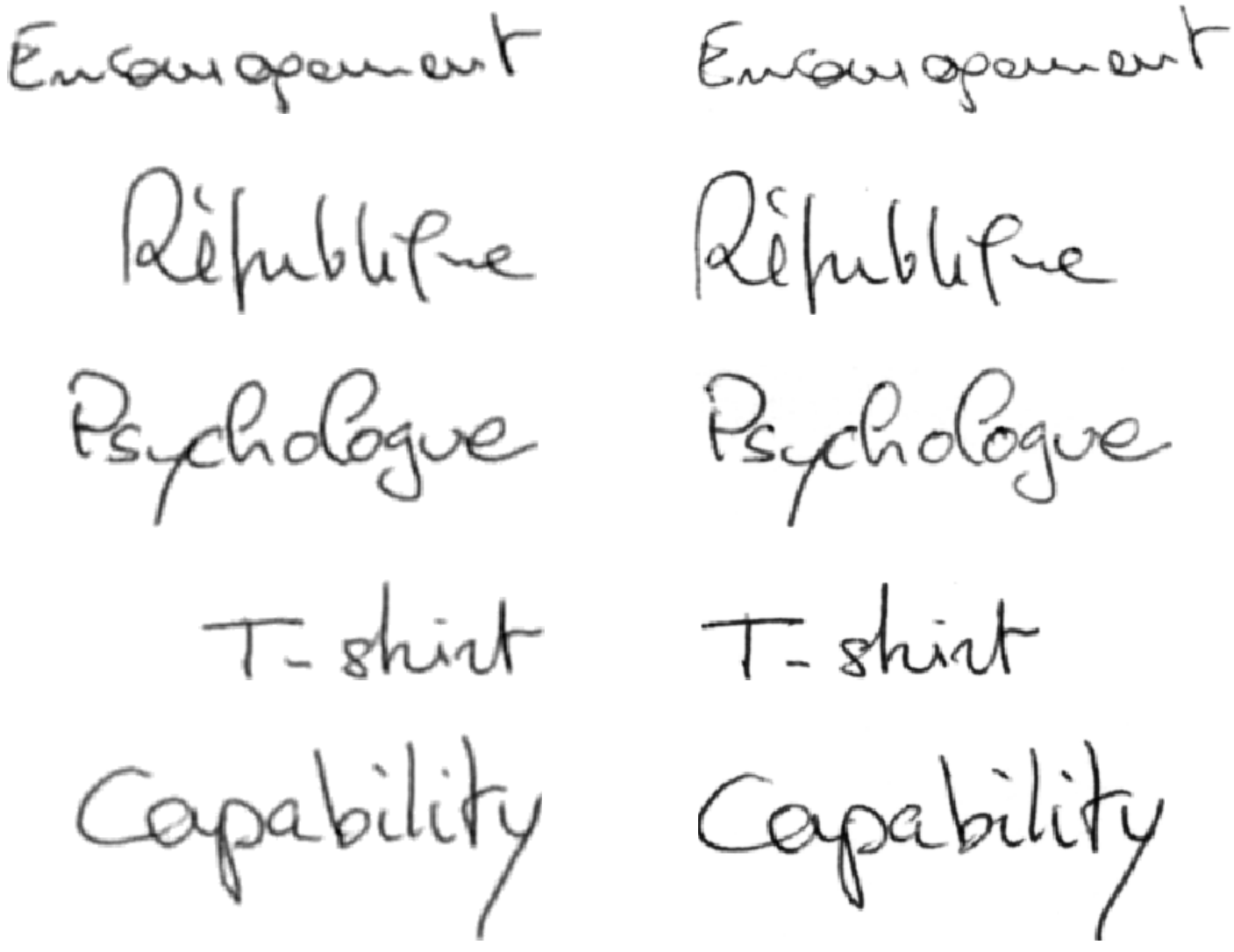


Figure 14: A not cherry-picked selection of synthetic manuscripts produced using our proposed method (left) alongside the expected output (right).

Evaluation Setup

In order to evaluate the quality of the synthetic signatures generated by our system we follow the same protocol presented on the work of Diaz *et al.* Diaz-Cabrera et al., 2014. Namely, we use a state-of-the-art offline verification system and a dataset comprising both online and offline signatures in order to train the verification system to evaluate the synthetic signatures.
The goal of the experiments is to measure the quality of the synthetic signatures taking into account an offline verification system performance.
The questions raised are

- is the synthetic signatures system performance similar to the offered by real offline signatures?
- is it feasible to increase the number of samples at the enrollment stage with our proposed method synthetic signatures?

Offline signature verification system

The system used for the evaluation of the real and synthetic signatures is a Linear SVM classifier and with a state-of-the-art feature extraction approach Hafemann, Sabourin, and Oliveira, 2017. The feature extraction system [4] uses ideas from transfer learning and multi-task learning to learn features using Convolutional Neural Networks (CNN). As discussed in Chapter before, one of the advantages of using deep learning techniques is that some models, such as the CNN, can learn filters that can be used as feature extractors. The offline system's feature extractor takes advantage of this concept.

[4] https://www.etsmtl.ca/Unites-de-recherche/LIVIA/Recherche-et-innovation/Projets/Signature-Verification

Database

The evaluation experiments were carried out on the BiosecurID database Fierrez et al., 2010. This multimodal database was made publicly available containing signatures of 132 subjects. Signatures were captured using a special digital inking pen on a paper placed over a digitizing tablet, exactly as shown in Figure 11. Consequently, both versions, online and offline, of the same real signature were acquired at the same time. This characteristic, therefore, makes BiosecurID the ideal benchmark for the experimental evaluation conducted in this work.

The signatures samples were captured in 4 different sessions, distributed over four months. Each subject signed 4 times and forged 3 signatures per session, thus leading to each subject having 4 genuine signatures x 4 sessions = 16 genuine samples and 3 signature forgeries x 4 sessions = 12 skilled forgeries.

Since the offline signature verification system is trained with both positive and negative samples, to ensure an unbiased result towards the dataset selection, similar to the protocol followed by Diaz-Cabrera et al., 2014 we used the MCYT database Ortega-Garcia et al., 2003 as the negative dataset samples. The MCYT dataset includes 75 signers each with 15

Experiments Protocol

We follow the same experiment protocol proposed by Diaz-Cabrera et al., 2014. Two different experiments are carried out.

- Experiment 1 focus on evaluating the synthetic signatures performance in comparison to real signatures
- Experiment 2 evaluates the feasibility of synthetically increasing the number of samples available in a dataset.

For both experiments, the BiosecurID dataset is split into two subsets. The first 90 users are separated as the enrollment set, used to compute the genuine and skilled impostor scores. The remaining 42 authors are considered as the test set and are used to compute the random impostor scores. The performance is evaluated regarding the equal error rate (EER), which is the point in the Detection Error Tradeoff curve (DET) where the false acceptance rate equals the false rejection rate.

Experiments Protocol

We follow the same experiment protocol proposed by Diaz-Cabrera et al., 2014. Two different experiments are carried out.

- Experiment 1 focus on evaluating the synthetic signatures performance in comparison to real signatures
- Experiment 2 evaluates the feasibility of synthetically increasing the number of samples available in a dataset.

Experiment 1

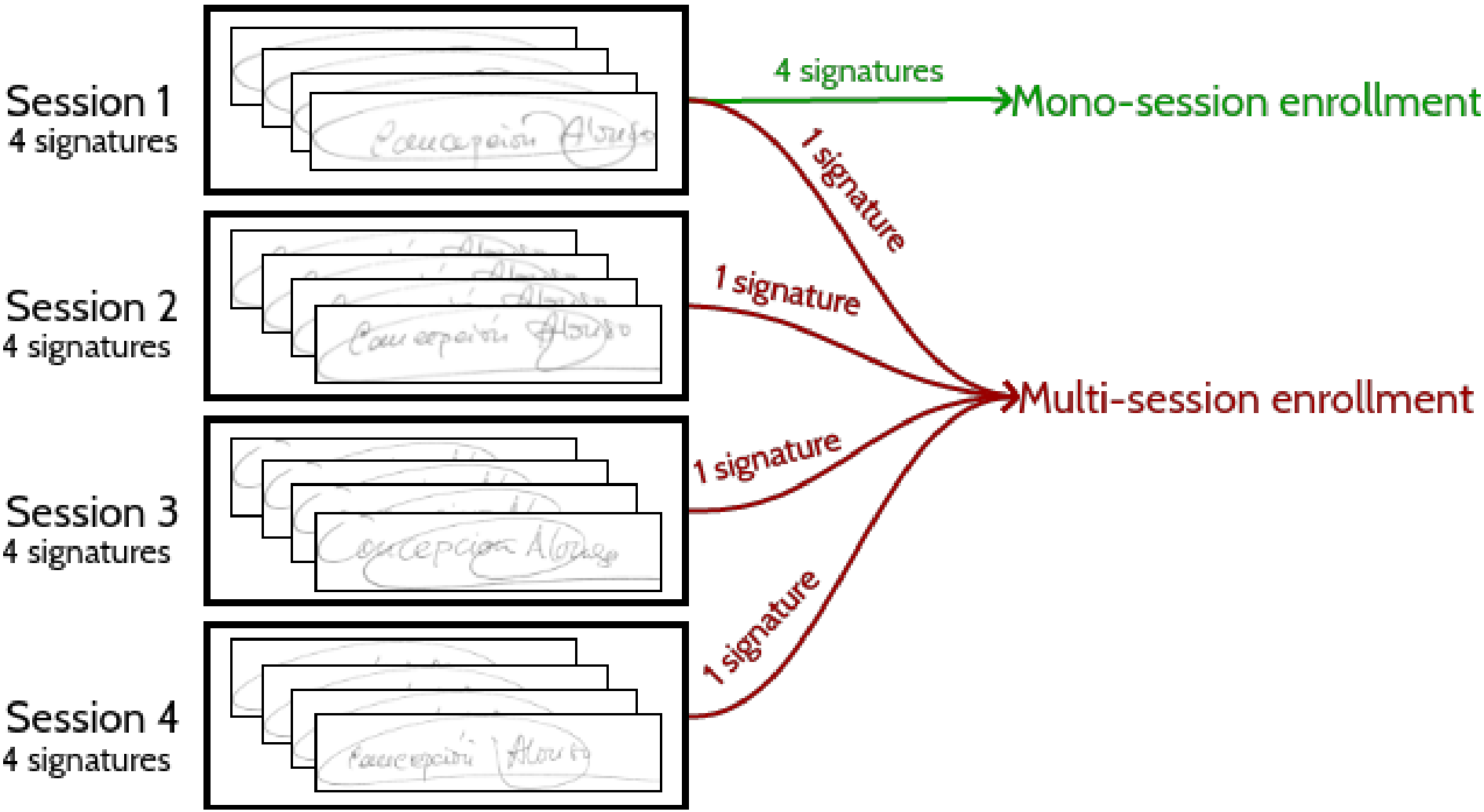


Figure 15: The proposed approach diagram and an example of the synthetic signature generation.

Experiment 2

This experiment is designed to assess whether synthetically increasing the enrollment dataset leads to a better recognition performance. Three different enrollment sets are considered in this experiment:

- 4 real samples belonging to the first acquisition session
- 8 real samples belonging to the first and the second sessions
- 4 real samples belonging to the first session plus 4 synthetic samples belonging to the second session.

Results

The goal of the experiments are:

- Measure the quality of the synthetic images;
- Assess whether using synthetic signatures effects the recognition performance of an offline signature verification system; and
- Analyze the feasibility of using real and synthetic signatures on the enrollment set.

We compare our results with a state-of-the-art method, particularly the approach proposed by Diaz *et al.* Diaz-Cabrera et al., 2014. Specifically, our proposed method is compared with the "Image enhanced" synthetic signatures made available as part of the BiosecurID Fierrez et al., 2010 dataset. The reported EER is achieved for both approaches in the same experimental conditions.

The recognition rates of three types of offline signatures are reported:

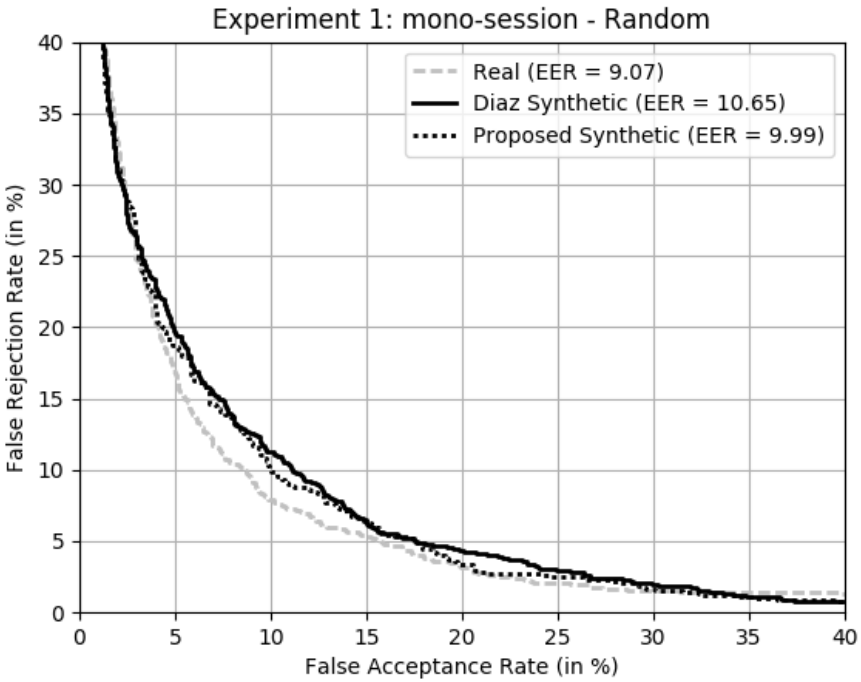
- real offline signatures, which are the corresponding real offline samples of the online signatures used to generate the synthetic samples, they serve us as a ground-truth, i.e., the ideal synthetic signature should be similar to it;
- synthetic signatures generated with the method proposed by Diaz *et*

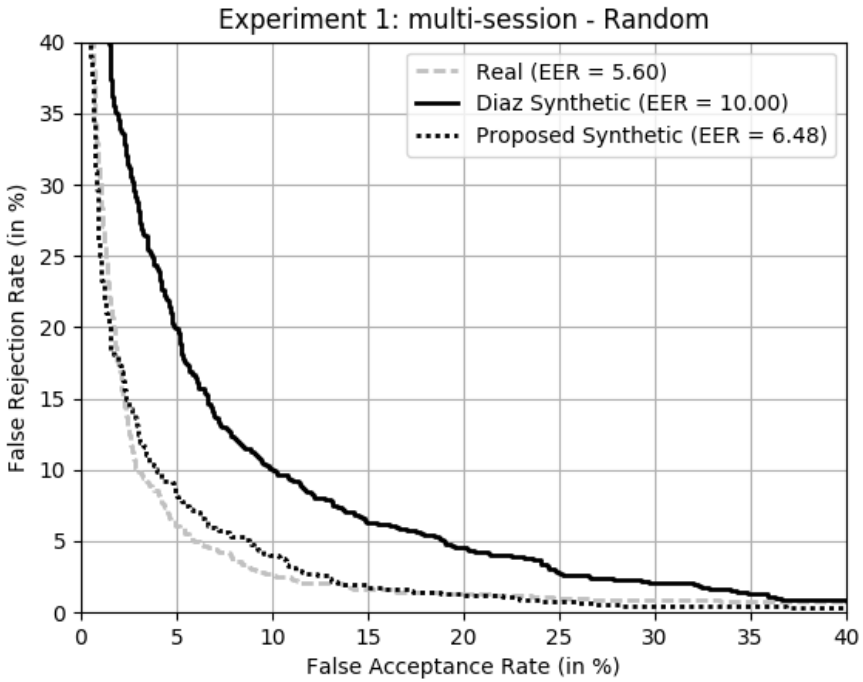
Experiment 1

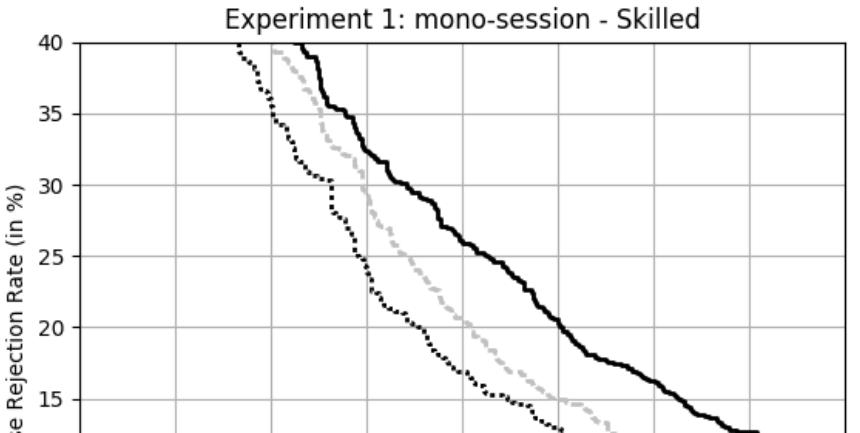
Table 1: EER for real, synthetic samples from Diaz *et al.* Diaz-Cabrera et al., 2014 and our proposed method synthetic offline signatures, for all the approaches considered under the two possible scenarios (i.e., random and skilled forgeries)

Mode	**Skilled Forgeries**		
	Real	**Diaz *et al.***	**Proposed method**
mono-session	20.28%	23.19%	18.38%
multi-session	17.59%	22.27%	16.48%
	Random Forgeries		
	Real	**Diaz *et al.***	**Proposed method**
mono-session	9.07%	10.65%	9.99%
multi-session	5.60%	10.00%	6.48%

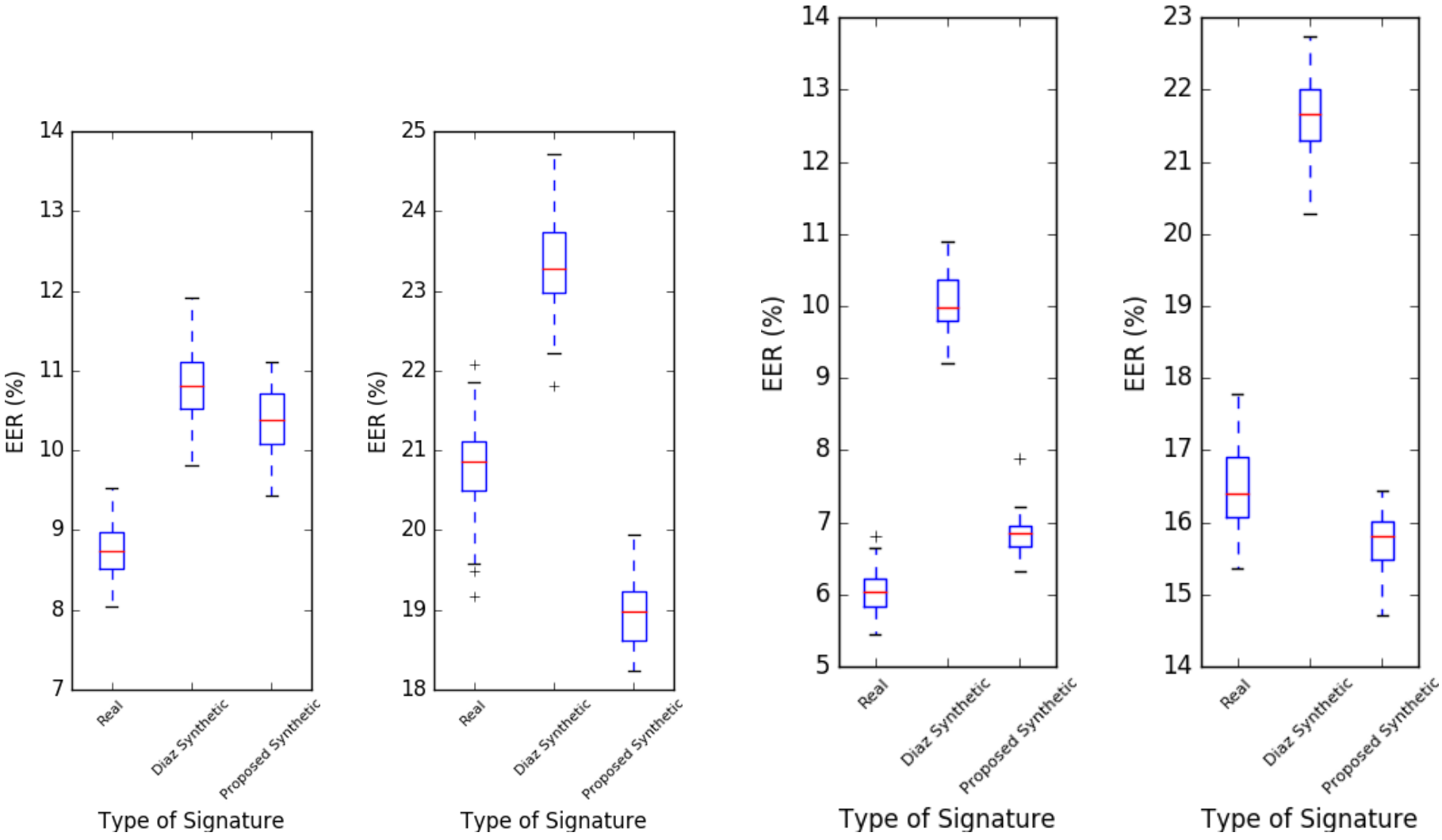
Experiment 1







Experiment 1 - running 30 times



(a) mono-session - random forgeries

(b) mono-session - skilled forgeries

(c) multi-session - random forgeries

(d) multi-session - skilled forgeries

Figure 17: Boxplot comparison for running 30 times the Experiment 1

Experiment 2

Table 2: EER for real, synthetic samples from Diaz *et al.* Diaz-Cabrera et al., 2014 and our proposed method synthetic offline signatures for the Experiment 2 under the two possible scenarios, i.e., random (RF) and skilled forgeries (SF)

Genuine Training	SF	RF
4 real samples	21.55%	10.26%
4 real + 4 real samples	19.72%	7.63%
4 real + 4 synthetic from Diaz *et al.*	24.19%	7.72%
4 real + 4 synthetic from the Proposed Method	19.17%	9.74%

Experiment 2

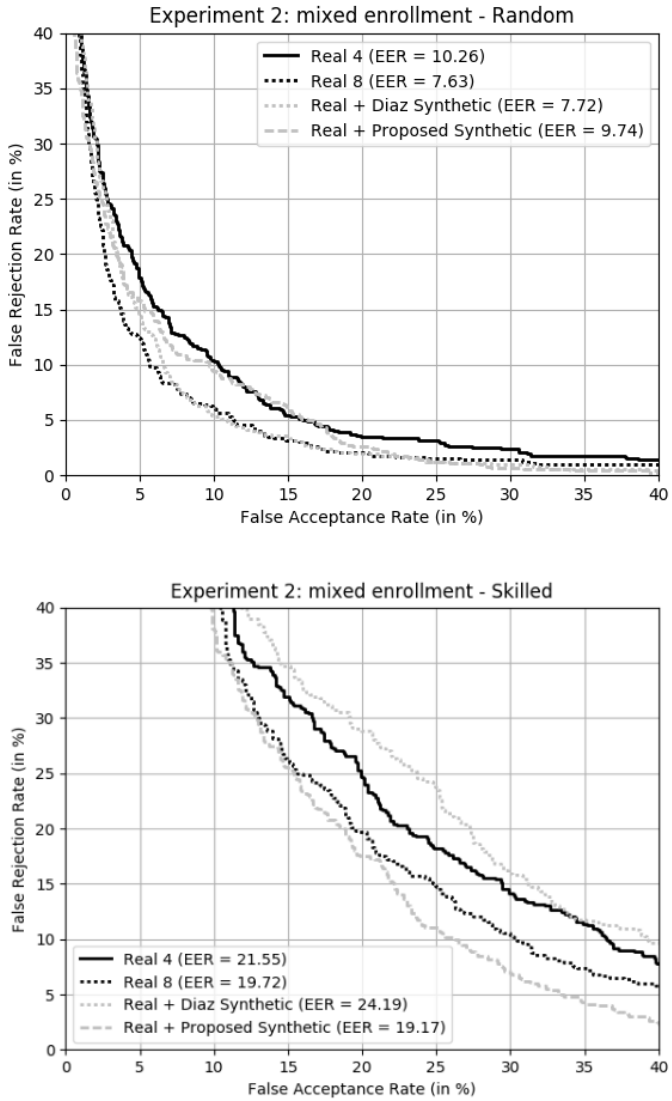
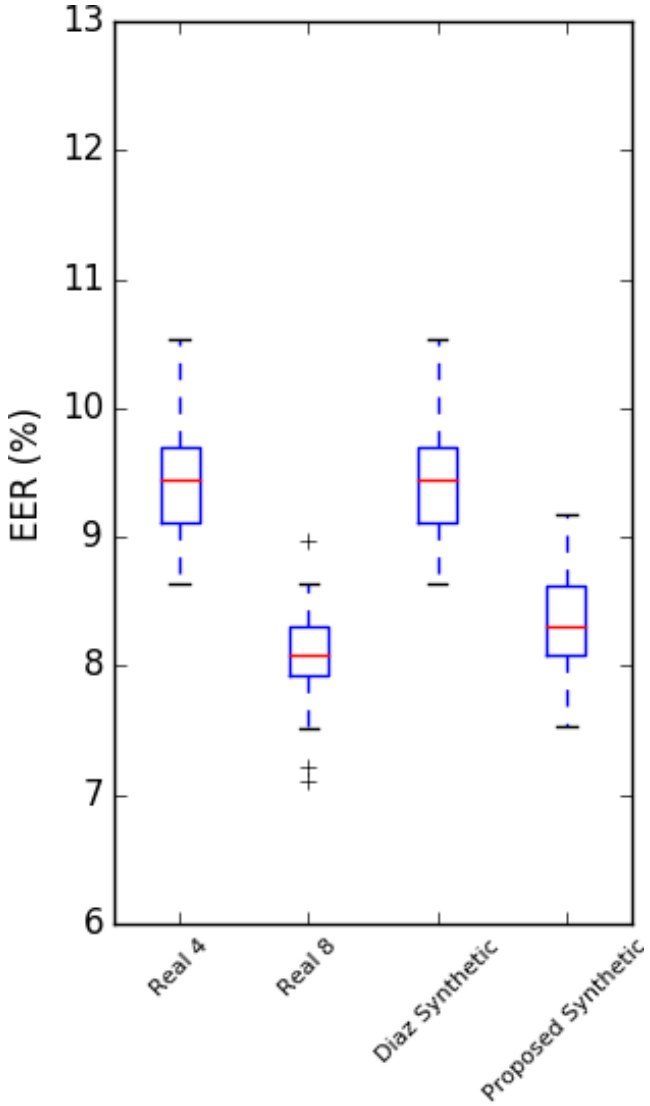
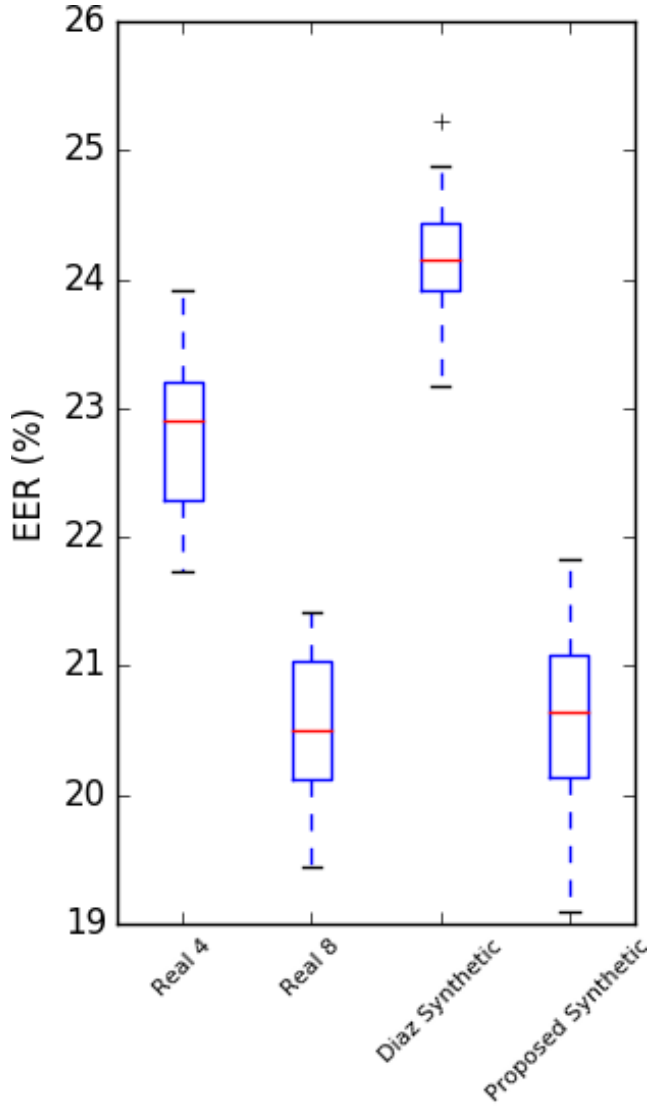


Figure 18: DET curves for real offline signatures and synthetic signatures (from Diaz *et al.* and our proposed method), for the second experiment, for the two

Experiment 2 - running 30 times



(a) mixed enrollment - random



(b) mixed enrollment - skilled

Conclusion

- A Fully Convolutional Neural Network to learn an end-to-end mapping from the online to the offline domain was proposed;
- We show that it is possible to model the "online to offline signature conversion" as a learning from data task.
- We observe that the synthetic offline signatures generated with the proposed method offer a verification performance similar to the one offered by real signatures
- We show that the synthetic signatures present high discriminative power when used to increase the enrollment set under the skilled forgeries scenario.

Future Works

- Explore the optimization of the hyper-parameters of the FCN (such as the number of layers, number of neurons per layer, different architectures)
- Integrate other dynamic features in addition to the pressure on the input of the neural network, such as the velocity.
- Design a model to synthesize offline signatures with bigger resolution.
- Combine real online and synthetically generated offline signatures using the proposed method, when only the online information is available, towards improved recognition results on a dynamic signature verifier.

Bibliography I

Diaz-Cabrera, Moises et al. (2014). "Generation of enhanced synthetic off-line signatures based on real on-line data". In: *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*. IEEE, pp. 482–487.

Ferrer, Miguel A, Moises Diaz-Cabrera, and Aythami Morales (2013). "Synthetic off-line signature image generation". In: *Biometrics (ICB), 2013 International Conference on*. IEEE, pp. 1–7.

Ferrer, Miguel A et al. (2013). "Realistic synthetic off-line signature generation based on synthetic on-line data". In: *Security Technology (ICCST), 2013 47th International Carnahan Conference on*. IEEE, pp. 1–6.

Fierrez, Julian et al. (2010). "BiosecurID: a multimodal biometric database". In: *Pattern Analysis and Applications* 13.2, pp. 235–246.

Bibliography II

Galbally, Javier et al. (2012). "Synthetic on-line signature generation. Part I: Methodology and algorithms". In: *Pattern Recognition* 45.7, pp. 2610–2621.

Galbally, Javier et al. (2015). "On-line signature recognition through the combination of real dynamic data and synthetically generated static data". In: *Pattern Recognition* 48.9, pp. 2921–2934.

Galbally Herrero, Javier et al. (2009). "Synthetic generation of handwritten signatures based on spectral analysis". In: *Proceedings of SPIE-The International Society for Optical Engineering*. Society of Photo-Optical Instrumentation Engineers.

Guest, Richard Matthew, Oscar Miguel Hurtado, and Olaf Henniger (2013). "Assessment of methods for image recreation from signature time-series data". In: *IET biometrics* 3.3, pp. 159–166.

Hafemann, Luiz G, Robert Sabourin, and Luiz S Oliveira (2015). "Offline handwritten signature verification-literature review". In: *arXiv preprint arXiv:1507.07909*.

Bibliography III

Hafemann, Luiz G, Robert Sabourin, and Luiz S Oliveira (2017). "Learning features for offline handwritten signature verification using deep convolutional neural networks". In: *Pattern Recognition* 70, pp. 163–176.

Haykin, Simon S et al. (2009). *Neural networks and learning machines*. Vol. 3. Pearson Upper Saddle River, NJ, USA:

Impedovo, Donato and Giuseppe Pirlo (2008). "Automatic signature verification: the state of the art". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38.5, pp. 609–635.

Jain, Anil K, Karthik Nandakumar, and Arun Ross (2016). "50 years of biometric research: Accomplishments, challenges, and opportunities". In: *Pattern Recognition Letters* 79, pp. 80–105.

Bibliography IV

Melo, Victor Kléber Santos Leite et al. (2017). "Datasets for Handwritten Signature Verification: A Survey and a New Dataset, the RPPDI-SigData". In: *Handwriting: Recognition, Development and Analysis, 2017*. Nova Science Publishers. Chap. 14, pp. 345–361.

Nel, E-M, Johan A Du Preez, and Ben M Herbst (2005). "Estimating the pen trajectories of static signatures using Hidden Markov models". In: *IEEE transactions on pattern analysis and machine intelligence* 27.11, pp. 1733–1746.

Ortega-Garcia, Javier et al. (2003). "MCYT baseline corpus: a bimodal biometric database". In: *IEE Proceedings-Vision, Image and Signal Processing* 150.6, pp. 395–401.

Viard-Gaudin, Christian et al. (1999). "The ireste on/off (ironoff) dual handwriting database". In: *Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the Fifth International Conference on*. IEEE, pp. 455–458.