Analysing and Exploiting Complexity Information in On-Line Signature Verification

Miguel Caruana, Ruben Vera-Rodriguez, and Ruben Tolosana

BiDA Lab, School of Engineering, Universidad Autonoma de Madrid
Madrid, Spain
miguel.caruana@estudiante.uam.es, ruben.vera@uam.es,
ruben.tolosana@uam.es

Abstract. This paper proposes an in-depth analysis on how the complexity of signatures affects the performance in on-line signature verification. In signature verification there is a very wide range of signatures from some based on a simple flourish to very complex ones. In this work we consider three different complexity groups: low, medium and high. We carry out an analysis of performance evaluation for each complexity group for both random and skilled forgeries. Two verification systems are used for this analysis, a traditional one based on the popular DTW and a state-of-the-art one based on time aligned recurrent neural networks (TA-RNN) recently proposed. The experiments are carried out over the largest database available to date for on-line signature verification (DeepSignDB). Then, we propose several approaches in order to exploit the information related to the signature complexity with the final aim of improving the signature verification system performance. Our best proposed approach is based on training a system with a balanced number of subjects regarding their type of signature complexity.

Keywords: On-line Signature Verification · Deep Learning · Biometrics.

1 Introduction

On-line handwritten signature verification is one of the most popular behavioral biometrics technologies. In the last 40 years, on-line signature verification has evolved very significantly and has proved to be one of the most reliable and convenient biometric systems in many relevant sectors. As a behavioral biometric trait, there are many factors that affect the performance of on-line signature such as the ergonomics, the quality of the acquisition device, device interoperability [23], usability factors [6], using the finger as the writing tool [16], the effect of aging [24], signature quality [9], signature complexity [22], the limited amount of public databases which has motivated the generation of synthetic signatures, etc. Most of these factors are reviewed in [3].

Complexity is an important factor in some traditional authentication systems such as those based on passwords, where a minimum complexity of the password is needed in order to ensure a minimum level of security. On-line dynamic signature verification systems may suffer from this problem as well. Analyzing the

complexity of the signatures allows us to find out how the complexity of the signatures affects the performance of a verification system. In addition, it would be possible to warn users with vulnerable signatures in terms of their complexity, so they could modify their signature by another more robust one.

In this paper we focus on the effect of the signature complexity in the system performance, which has been shown to be a significant factor. In [2], Alonso-Fernandez *et al.* evaluated the effect of the complexity and legibility of the signatures for off-line signature verification (i.e., signatures with no available dynamic information) pointing out the differences in performance for several matchers.

Different approaches have been proposed to detect the complexity of dynamic handwritten signatures. Houmani *et al.* proposed an entropy measure based on local density estimation by a Hidden Markov Model (HMM) [8]. Miguel-Hurtado *et al.* investigated the creation of a novel mathematical model for the automatic assessment of the signature complexity [12]. A recent approach was carried out by Tolosona *et al.* in [22]. This study proposed a complexity detector based on using the number of strokes applying the well-known writing generation Sigma LogNormal model [5]. Then, once the signatures were classified into three levels of complexity (low, medium and high), optimal time functions associated with each specific complexity level were extracted. Finally, they measured the performance of a signature verification system system based on Dynamic Time Warping (DTW) only using the optimal time functions for each complexity class. In this way, significant improvements in the system's performance were obtained compared to a DTW-based system using the same time functions for all signatures without taking into account their complexity level.

Recently, Vera-Rodriguez *et al.* proposed in [25] a new complexity detection system based on Deep Learning (DL) techniques. This system was developed through a semi-supervised process where an initial model was trained over a medium-size database (BioSecurID [4]), which was then used to classify the complexity of signatures of a much larger database (DeepSignDB [21]). Finally, based on these automatic labels they developed the complexity detection system, achieving results ca. 85% of accuracy compared to the manual labels. This system improved significantly the results of 64% of accuracy achieved previously in [22] in the same experimental conditions.

Deep learning is the state-of-the-art technology used nowadays in many other biometric recognition traits such as the face [14] or the voice [7]. However, most of the state-of-the-art signature verification systems are still based on traditional approaches such as DTW, HMM, and Support Vector Machines (SVM).

Very recently, some works that apply Deep Learning techniques, both based on Recurrent Neural Networks (RNNs) [17, 20, 11, 10, 1], but also based on Convolutional Neural Networks (CNNs) [15, 26] have been published, obtaining promising results. In particular, the strategy proposed in [20] is based on using a large database of on-line signatures (DeepSignDB) with almost 1500 subjects in order to better train a RNN-based system achieving very good results both for skilled and random forgeries.

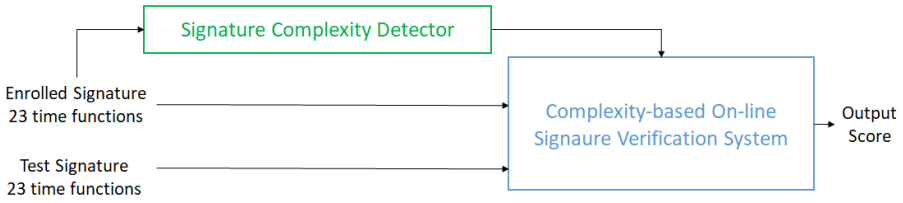


Fig. 1. Architecture of our proposed complexity-based on-line signature verification system.

In this work we first carry out an analysis of the on-line signature verification system performance using the publicly available DeepSignDB database in order to study the diversity of the subjects in terms of their signature complexity. Then, we carry out several experiments in order to exploit the signature complexity with the aim of improving the overall performance of a on-line signature verification system based on recurrent neural networks.

The remainder of the paper is organised as follows. Sec. II describes our proposed on-line verification system specific to signature complexity. Sec. III describes the DeepSignDB database considered in the experiments of this article. Sec. IV describes the experimental protocol, Sec. V reports the results achieved using our proposed approach. Finally, Sec. VI draws the final conclusions and points out some lines for future work.

2 Proposed Approach

This section describes our proposed approach based on detecting the complexity of the enrolled signatures and then applying a complexity-based on-line signature verification system, as shown in Figure 1. The main idea of this work is to analyse how the complexity of the enrolled signatures affects the performance of a verification system and if it is possible to exploit that complexity information in order to develop a complexity-based on-line verification system that can improve its signature verification performance. In order to achieve that, we have defined a simple but effective architecture. In on-line signature verification we will have a test (unknown) signature that is going to be compared to an enrolled signature to verify its identity. First, the enrolled signature goes through a signature complexity detector which classifies the signature into one of three levels of complexity. Then, the two signatures together with the complexity label of the enrolled signature enter the complexity-based on-line verification system. Finally, the verification system provides an output score value (between 0 and 1), which states the similarity of the two signatures under comparison (a value close to 0 would be expected when the two signatures are genuine and a value close to 1 would be expected when the test signature is either a skilled or a

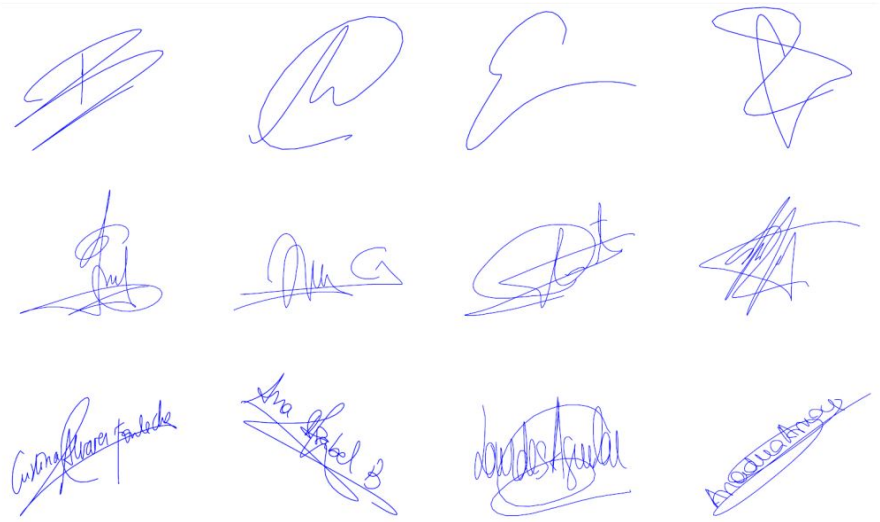


Fig. 2. Examples of signatures of three complexity groups: low complexity (top), medium complexity (middle) and high complexity (bottom).

random forgery). Next, we describe briefly the two technologies used both for signature complexity estimation and on-line signature verification.

2.1 Signature Complexity Detector

The first stage of our proposed system consist of a signature complexity detector. For this we make use of the system proposed in [25]. This system classifies each signature into one of three levels of complexity considered: low, medium and high. Fig. 2 shows examples of signatures in the three levels of complexity considered in this work.

For each signature, signals related to X and Y spatial coordinates and pressure are used to extract a set of 23 time functions [17], which are the input of the system. These time functions are preprocessed following a zero mean and unit standard deviation normalization. The network is composed of two Bidirectional Long Short-Term Memory (BLSTM) layers and a feed-forward layer with a softmax activation, which provides an output score for each of the three complexity levels considered. As described in [25], this system achieves 85% of accuracy of complexity level compared to a manual ground truth for an evaluation dataset from BiosecurID database [4].

2.2 On-line Signature Verification System

We base our proposed complexity-based signature verification system on the TA-RNN system (Time-Aligned Recurrent Neural Networks) proposed in [19],

[20]. For the input of the system, we feed the network with the 23 time functions extracted from the signature, which are the same that are used for the complexity detection system. The TA-RNN architecture is based on two stages: i) a first stage based on time sequence alignment through DTW and ii) a second stage consisting of a neural network, specifically a RNN. The system developed in [20] was comprised of three layers. The first layer is composed of two Bidirectional Gated Recurrent Unit (BGRU) hidden layers with 46 memory blocks each, sharing the weights between them. The outputs of the first two parallel BGRU hidden layers are concatenated and serve as input to the second layer, which corresponds to a BGRU hidden layer with 23 memory blocks. Finally, a feed-forward neural network layer with a sigmoid activation is considered, providing an output score for each pair of signatures.

TA-RNN architecture is used in this paper as one of the baseline systems to compare the results achieved with our complexity-based signature verification system. In our complexity-based signature verification system we carried out different strategies to exploit the signatures complexity. In particular we considered three main approaches: i) training from scratch a specific DL model per complexity level, ii) training a specific DL model per complexity level but applying fine tuning from the general DeepSign model in [20], and iii) training just one model for all types of signatures, but balancing the complexity classes during the training. The different verification systems adapted to each complexity use the same architecture as the one described before. However, for certain experiments some changes in the architecture were made. These details are described in Sect. IV.

3 Database

The DeepSignDB database [21] comprises data from a total of 1526 subjects from four different well-known on-line signature databases: MCYT (330 subjects) [13], BiosecurID (400 subjects) [4], Biosecure DS2 (650 subjects) [9], eBioSign (65 subjects) [16] and a novel signature database comprised of 81 subjects. This database comprises more than 70K signatures acquired using both stylus and finger inputs. Two acquisition scenarios are considered, office and mobile, with a total of 8 different devices. Additionally, different types of impostors and number of acquisition sessions are considered along the database. For the results presented in this work, only signatures performed with pen stylus are considered.

4 Experimental Protocol

The experimental protocol has been designed to analyse two main ideas: i) how the signatures complexity affects on the performance of an on-line signature verification system, and ii) how considering the signatures complexity in the training process can improve the performance of the signature verification system.

Five separate blocks of experiments have been carried out. First, in Exp. 1 we analyse the performance of the DTW and the TA-RNN baseline systems for each

complexity group. In Exp. 2 we propose to train from scratch a specific system per each complexity group using the TA-RNN architecture. Then, in Exp. 3 we analyse a similar approach to Exp. 2 but modifying the system architecture per each complexity group. After this, in Exp. 4 we follow a similar approach to Exp. 2, but fine tuning each complexity group system from the baseline system instead of training from scratch. Finally, in Exp. 5 we followed a different approach, which is based on training just one global system, similar to the baseline, but with data from complexity balanced subjects.

DeepSignDB has been the database used for all the experiments reported in this work. We followed the experimental protocol from [20], and divided it into two different datasets, one for development and one for evaluation. The development set contains 70% of the subjects while the other 30% is the evaluation set. It is important to note that each dataset contains different subjects in order to avoid biased results.

From the original 1084 subjects contained in the development set, we applied the signature complexity detector proposed in [25] to all the genuine signatures. Then, we computed the mean complexity of each subject, taking all their genuine signatures complexity into account. Finally, we classified each subject into the complexity group closer to its mean complexity. At the end we obtained 230 high-complexity subjects, 637 medium-complexity subjects and 217 low-complexity subjects in the development set.

For system training, the development set has a total of 980 subjects. This set has been further divided into two different subsets, one for training (80%) and one for validation (20%).

As the first objective is to analyze the system performance for the three groups of complexity separately, the evaluation dataset was also divided into three subsets, one per each complexity group. This was done based on the complexity level of the enrolment genuine signature of the pairs of signatures under comparison, as Fig. 1 shows. In order to provide a global system performance the scores from the three evaluation sets were put together.

Two impostor scenarios have been considered, skilled and random forgeries. For the skilled forgery case, all available skilled forgery samples are included in the analysis whereas for the random forgery case, one genuine sample of each of the remaining subjects of the same database is considered. This way verification systems are tested with different types of presentation attacks [18]. It is worth noting that the evaluation results reported in this work are only based on one to one comparisons of signatures. Thus, only one enrolment genuine signature is considered per subject, which is an extreme case as normally more than just one signature is considered as enrolment.

The evaluation results are given as the performance of the system in terms of DET curves and Equal Error Rate (EER). The evaluation results are obtained over the three evaluation datasets, one per class, as well as over the complete evaluation dataset. In such a manner we can establish a comparative analysis of the results achieved for each class. Two different systems are used as a baseline

to compare the results achieved in this work. We use a traditional approach, a DTW-based system, and a state-of-art system [20].

5 Experimental work

5.1 Exp. 1 - Analysis of complexity for Baseline Systems

In this experiment we carry out the analysis of the evaluation performance of the two baseline systems (DTW and TA-RNN) for the three signature complexity groups. This way we can obtain the performance of the baseline systems over the same evaluation datasets and carry out a comparative analysis with the following proposed approaches. Tables 1 and 2 show the performance of the system for each complexity group for the case of skilled and random forgeries comparisons respectively.

First of all it is worth noting the much superior performance of the TA-RNN system compared to the traditional DTW approach in particular for skilled forgeries. In both skilled and random forgeries cases the best performance is achieved for the medium complexity subjects and in both baseline systems, with 9.94% and 2.40% EER for the DTW system, and 3.32% EER and 1.19% EER for the TA-RNN system. This can be due to several factors. On the one hand, the number of medium complexity subjects is the highest in the development dataset so it is normal that the performance over this group is better, since it has almost twice as many subjects as the other two classes. On the other hand, the medium complexity signatures might be more stable and robust achieving a lower EER. Low complexity signatures can be easily confused with signatures of different subjects due to the low inter-class variability. High complexity signatures of the same subject can be quite different due to the high intra-class variability. The second best performance is achieved for high complexity signatures leaving the low complexity signatures with the worst performance.

In general terms, the global performance for the baseline systems is 11.13% EER for DTW and 4.20% EER for TA-RNN in skilled forgeries, and 2.62% EER for DTW and 1.51% EER for TA-RNN in random forgeries. Fig. 3 also shows the DET curve for the global evaluation of the TA-RNN baseline system.

Table 1. Performance over skilled forgeries in terms of EER.

	Low	Med	High	Global
Exp. 1 - DTW Baseline	13.27	9.94	11.53	11.13
Exp. 1 - TA-RNN Baseline	5.88	3.32	4.60	4.20
Exp. 2 - 3 Models scratch	8.89	3.66	6.31	5.75
Exp. 3 - Different Architecture	6.80	4.20	5.63	5.11
Exp. 4 - 3 Models Fine Tuning	6.03	3.30	4.60	4.23
Exp. 5 - Prop. 1 Model Balanced	**5.63**	**3.27**	**3.89**	**3.92**

Table 2. Performance over random forgeries in terms of EER.

	Low	Med	High	Global
Exp. 1 - DTW Baseline	2.62	2.40	2.86	2.62
Exp. 1 - TA-RNN Baseline	2.08	1.19	1.84	1.51
Exp. 2 - 3 Models scratch	3.40	1.24	2.22	2.02
Exp. 3 - Different Architecture	1.72	1.45	2.26	1.73
Exp. 4 - 3 Models Fine Tuning	2.00	1.17	1.84	1.54
Exp. 5 - Prop. 1 Model Balanced	**1.50**	**1.09**	**1.62**	**1.32**

5.2 Exp. 2 - Training from Scratch 3 Systems, One System per Complexity

From Exp. 1 it is clear that the system based on TA-RNN clearly outperforms the traditional system based on DTW. Therefore in the following experiments the aim is to exploit the signature complexity to try to improve the already very good performance of the TA-RNN system. As a first approach in order to exploit the signatures complexity in an on-line signature verification system we tried to train three different models from scratch using the TA-RNN architecture, one per complexity group. Each model was trained only using comparisons where the genuine signature belongs to the specific complexity class, but where the impostor signature can belong to any complexity group. It is important to mention that all the subjects available per class were used, so when evaluating the results it is worth taking this into account, as each complexity group has different subjects and therefore different amounts of data to be trained with. Specifically, 181 subjects were used to train the low complexity model, 502 subjects to train the medium complexity model and 184 to train the high complexity model. Tables 1 and 2 show the results achieved by each of the models trained in this experiment following the same evaluation protocol as the baseline systems shown in Exp. 1.

The results obtained in this experiment lead us to a similar conclusion to the previous one regarding the complexity groups as they follow similar trends. However, results achieved with this approach comprised of three systems, one per complexity group, are significantly worse compared to the ones achieved by the baseline TA-RNN system. This can be due to the baseline system having been trained with a much higher number of subjects, and it seems that the baseline system can make use of the information provided by the higher variability of subjects trained with, and provide better results. Only for the case of the medium complexity evaluation the EER of the baseline system and this approach are quite similar as the system with medium complexity has been trained with the highest number of subjects.

In general terms, the global performance for this approach is 5.75% EER for skilled forgeries and 2.02% EER for random forgeries, which is also worse than the performance of the baseline system reported in Exp. 1. Fig. 3 also shows the DET curves with the global evaluation for this approach, showing the worst performance of all experiments carried out.

5.3 Exp. 3 - Training from Scratch and Changing the Baseline Architecture

This experiment was designed to see if any changes in the system's architecture from the original TA-RNN architecture could improve its performance against a particular group of complexity. Two different models were trained, one focused on low complexity signatures and the other focused on improving performance against high complexity signatures. Only these two groups were considered as these are the classes where the worst results were obtained in Exp. 2.

In the model for low complexity the second BGRU layer was removed in order to create a simpler model. The results we achieved using this system were really bad for all types of signatures, not just for low complexity ones. On the other hand, in the model focused on high complexity a third hidden BGRU layer was added before the feed forward layer. This new model trained just with high complexity subjects turned out to give reasonable results in general.

The second proposed architecture just trained with high complexity subjects did get better results. Tables 1 and 2 show the performance achieved by this system. As we can see the results obtained for the high complexity model improve the results obtained by the models of Exp. 2, both in high and low complexity. This can be also seen in general terms in the DET curves shown in Fig. 3.

Despite not being able to improve the results of the baseline system, this new architecture is promising as it achieves results closer to those of the baseline system even though it is trained with a much lower number of subjects.

5.4 Exp. 4 - Fine tuning 3 Systems

In addition to changes in the architecture, in this experiment we also train three systems similar to Exp. 2 but instead of training from scratch we apply fine tuning from the baseline system in Exp. 1 to each complexity group system.

This experiment was designed to find out if the performance of the system could be improved for a particular complexity group by carrying out a fine-tuning of the baseline model using only the subjects of that complexity group. For each complexity, three types of fine-tuning were performed: training only the final layer (fully connected), training the last two layers (second BGRU and fully connected layers) and training all layers (the two BGRU and fully connected layers). Therefore, three different models per complexity group were trained.

Similar results were obtained by the different fine-tuning strategies. The best results were achieved by training only the fully connected layer. Tables 1 and 2 show the achieved results by the three models that apply this strategy.

We can see that there is hardly any difference with the performance of the baseline system. For low complexity signatures there is a small drop in the performance over skilled forgeries and a subtle improvement over random forgeries. Carrying out a fine-tuning with more subjects of low and high complexity might get better results. In addition, having a look at the DET curves in Fig. 3 we can see how the performance achieved with this approach is very similar to the one achieved by the TA-RNN baseline system in general terms.

5.5 Exp. 5 - Training a Global System with Balanced Classes

After having analysed many different approaches training specific systems per complexity group, then we decided to train just one system but using a balanced number of subjects (and samples) per complexity group.

One model was trained using the TA-RNN architecture. To train the system 181 subjects per class were used, in total 543. This way we used all the possible subjects maintaining a balanced number of subjects between the different groups of complexity. Different strategies were followed during the training process, i.e., training from scratch and fine tuning one, two and the three layers from the TA-RNN architecture. Best results were achieved when training the model from scratch.

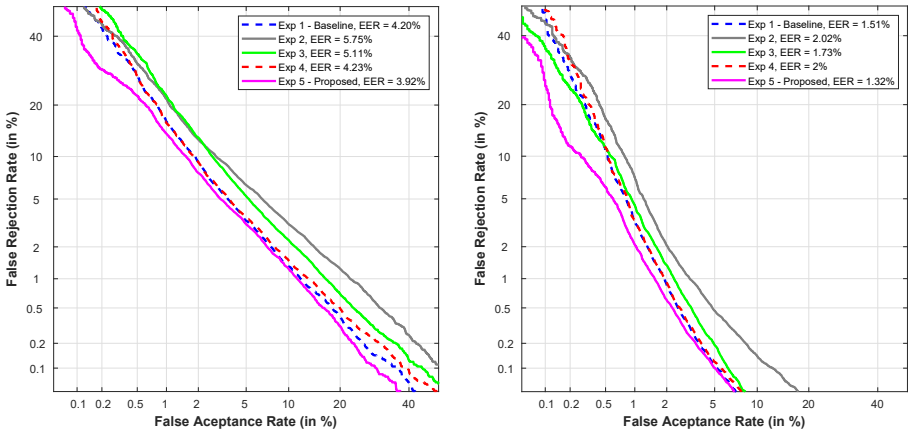
Tables 1 and 2 show the performance of this approach. Despite having been trained with a small number of subjects when comparing with the baseline system (543 Vs. 1084 subjects), we observe that the system achieves better performance for all types of complexity and also in the general evaluation. Analysing the case of skilled forgeries, the improvement of performance for low and medium complexity is small. However, for high complexity there is a relative improvement of 15.44% EER compared to the TA-RNN baseline system. Regarding the global performance the relative improvement achieved with the proposed system is of 6.67% EER skilled forgeries.

Analysing the case of random forgeries, the improvement achieved is more significant compared to the case of skilled forgeries. Here, a very significant relative improvement of 27.89% EER is achieved for the case of low complexity with 1.50% EER. The relative improvements for medium and high complexity are 8.40% and 11.96% EER respectively. Regarding the global performance the relative improvement achieved with the proposed system is of 12.58% EER for random forgeries.

Finally, Fig. 3 shows the DET curves for the proposed system achieving the best system performance at all operating points of false rejection rate and false acceptance rate compared to all the experiments carried out and the baseline systems. This highlights the importance of having balanced classes in the training phase. Analysing the obtained results we can conclude that it is more important to have a balanced number of subjects per class than a high number of unbalanced subjects.

6 Conclusions and Future Work

This paper has proposed an in-depth analysis on how the complexity of signatures affects the performance of an on-line signature verification system. In particular, we have carried out an analysis of the system performance for three groups of subjects regarding their signature complexity: subjects with low, medium and high signature complexity. In general, based on two baseline systems (DTW and TA-RNN), subjects with medium signature complexity achieve the best system performance both for skilled and random comparisons, followed by high



(a) Skilled Forgeries (b) Random Forgeries

Fig. 3. System performance results of our experiments over the DeepSignDB evaluation dataset.

complexity users and finally low complexity users. The reason for this might be that low complexity signatures may have low inter-class variability, while high complexity signatures may have high intra-class variability. Medium complexity signatures are likely to have higher inter-class and lower intra-class variability compared to low and high complexity ones. This information could be taken into account to warn users with low or high complexity signatures, especially low complexity ones, as the system will perform worse over those types of signatures.

Then, different approaches have been proposed in order to exploit the information related to the signature complexity with the final aim of improving the signature verification system performance. In particular we have proposed training specific systems per complexity group following different strategies (training from scratch, trying different system architectures, or applying fine tuning), but none of them improved the performance of the TA-RNN baseline system. Finally, an approach based on training a global system with balanced subjects regarding their complexity has outperformed the TA-RNN baseline system with a relative improvement of 6.67% EER for skilled forgeries and a relative improvement of 12.58% EER for random forgeries. We can conclude that training with a balanced number of subjects regarding their type of signature complexity reduces the number of users needed to achieve a particular performance, thus making training more efficient.

There is still a lot of work to be conducted regarding the complexity of signatures. In this work, our experimental work has been based on using just one enrolled signature. In future works, the case of having access to more than just one enrolled signature will be considered. Also, we will study different types of deep learning architectures, not just based on RNNs, but also on CNNs.

7 Acknowledgments

This work has been supported by projects: PRIMA (MSCA-ITN-2019-860315), TRESPASS (MSCA-ITN-2019-860813), BIBECA (RTI2018-101248-B-I00 MINECO FEDER) and Cecabank. M. Caruana is supported by Ayudas para el fomento de la Investigación en Estudios de Máster-UAM 2019. R. Tolosana is supported by Comunidad de Madrid and Fondo Social Europeo.

References

1. Ahrabian, K., Babaali, B.: On usage of autoencoders and siamese networks for online handwritten signature verification. Neural Computing and Applications (12 2017)
2. Alonso-Fernandez, F., Fairhurst, M., Fierrez, J., Ortega-Garcia, J.: Impact of Signature Legibility and Signature Type in Off-Line Signature Verification. In Proc. IEEE Biometrics Symposium (2007)
3. Diaz, M., Ferrer, M.A., Impedovo, D., Malik, M.I., Pirlo, G. and Plamondon, R.: A Perspective Analysis of Handwritten Signature Technology. ACM Computing Surveys **51**, 1–39 (2019)
4. Fierrez, J., Galbally, J., Ortega-Garcia, J., *et al.*: BiosecurID: A Multimodal Biometric Database. Pattern Analysis and Applications **13**(2), 235–246 (2010)
5. Fischer, A., Plamondon, R.: Signature Verification based on the Kinematic Theory of Rapid Human Movements. IEEE Transactions on Human-Machine Systems **47**(2), 169–180 (2017)
6. Guest, R., Brockly, M., Elliott, S., Scott, J.: An assessment of the usability of biometric signature systems using the human-biometric sensor interaction model. International Journal of Computer Applications in Technology **53**(4), 336–347 (2016)
7. Heigold, G., Moreno, I., Bengio, S., Shazeer, N.: End-to-end text-dependent speaker verification. In: Proc. ICASSP. pp. 5115–5119 (2016)
8. Houmani, N., G.S.S., Dorizzi, B.: A Novel Personal Entropy Measure Confronted to Online Signature Verification Systems Performance. In Proc. International Conference on Biometrics: Theory, Applications and System, BTAS pp. 1–6 (2008)
9. Houmani, N., Mayoue, A., Garcia-Salicetti, S., Dorizzi, B., Khalil, M., Moustafa, M., Abbas, H., Muramatsu, D., Yanikoglu, B., Kholmatov, A., Martinez-Diaz, M., Fierrez, J., Ortega-Garcia, J., Roure Alcobé, J., Fabregas, J., Faundez-Zanuy, M., Pascual-Gaspar, J., Cardeñoso-Payo, V., Vivaracho-Pascual, C.: BioSecure Signature Evaluation Campaign (BSEC'2009): Evaluating On-Line Signature Algorithms Depending on the Quality of Signatures. Pattern Recognition **45**(3), 993 – 1003 (2012)
10. Lai, S., Jin, L.: Recurrent adaptation networks for online signature verification. IEEE Transactions on Information Forensics and Security **14**, 1624–1637 (2019)
11. Li, C., Zhang, X., Lin, F., Wang, Z., Liu, J., Zhang, R., Wang, H.: A stroke-based RNN for writer-independent online signature verification. In: Proc. International Conference on Document Analysis and Recognition (ICDAR). pp. 526–532 (09 2019)
12. Miguel-Hurtado, O., Guest, R., Chatzisterkotis, T.: A new approach to automatic signature complexity assessment. In: Proc. IEEE International Carnahan Conference on Security Technology (ICCST). pp. 1–7 (2016)

13. Ortega-Garcia, J. *et al.*: MCYT Baseline Corpus: A Bimodal Biometric Database. IEE Proceedings Vision, Image and Signal Processing **150**(6), 395–401 (December 2003)
14. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: Proc. British Machine Vision Conference (2015)
15. Sekhar, C. and Mukherjee, P. and Guru, D.S. and Pulabaigari, V.: OSVNet: Convolutional Siamese Network for Writer Independent Online Signature Verification. In: Proc. International Conference on Document Analysis and Recognition (ICDAR) (2019)
16. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J.: Benchmarking Desktop and Mobile Handwriting across COTS Devices: the e-BioSign Biometric Database. PLOS ONE **12** (2017)
17. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: Exploring recurrent neural networks for on-line handwritten signature biometrics. IEEE Access **6**, 5128–5138 (2018)
18. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection, Handbook of Biometric Anti-Spoofing (2nd Edition). Springer (2018)
19. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: BioTouchPass2: Touchscreen password biometrics using time-aligned recurrent neural networks. IEEE Transactions on Information Forensics and Security **5**, 2616–2628 (2020)
20. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: DeepSign: Deep on-line signature verification. arXiv preprint arXiv:2002.10119 (2020)
21. Tolosana, R., Vera-Rodriguez, R., Fierrez, R., Morales, A., Ortega-Garcia, J.: Do you need more data? the DeepSignDB on-line handwritten signature biometric database. In: Proc. 15th IAPR Int. Conference on Document Analysis and Recognition, ICDAR (2019)
22. Tolosana, R., Vera-Rodriguez, R., Guest, R., Fierrez, J., Ortega-Garcia, J.: Exploiting complexity in pen- and touch-based signature biometrics. International Journal on Document Analysis and Recognition (23), 129–141 (2020)
23. Tolosana, R., Vera-Rodriguez, R., Ortega-Garcia, J., Fierrez, J.: Preprocessing and feature selection for improved sensor interoperability in online biometric signature verification. IEEE Access **3**, 478 – 489 (May 2015)
24. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: Reducing the template aging effect in on-line signature biometrics. IET Biometrics **8**(6), 422–430 (June 2019)
25. Vera-Rodriguez, R., Tolosana, R., Caruana, M., Manzano, G., Gonzalez-Garcia, C., Fierrez, J., Ortega-Garcia, J.: DeepSignCX: Signature complexity detection using recurrent neural networks. In: Proc. 15th International Conference on Document Analysis and Recognition, ICDAR (September 2019)
26. Wu, X. and Kimura, A. and Iwana, B.K. and Uchida, S. and Kashino, K.: Deep Dynamic Time Warping: End-to-End Local Representation Learning for Online Signature Verification. In: Proc. International Conference on Document Analysis and Recognition (ICDAR) (2019)