Exploring Transformers for On-Line Handwritten Signature Verification

Pietro Melzi[1,*], Ruben Tolosana[1], Ruben Vera-Rodriguez[1], Paula Delgado-Santos[1,2], Giuseppe Stragapede[1], Julian Fierrez[1] and Javier Ortega-Garcia[1]

[1]*Universidad Autonoma de Madrid, Spain*
[2]*University of Kent, UK*

Abstract

The application of mobile biometrics as a user-friendly authentication method has increased in the last years. Recent studies have proposed novel behavioral biometric recognition systems based on Transformers, which currently outperform the state of the art in several application scenarios. On-line handwritten signature verification aims to verify the identity of subjects, based on their biometric signatures acquired using electronic devices such as tablets or smartphones. This paper investigates the suitability of architectures based on recent Transformers for on-line signature verification. In particular, four different configurations are studied, two of them rely on the Vanilla Transformer encoder, and the two others have been successfully applied to the tasks of gait and activity recognition. We evaluate the four proposed configurations according to the experimental protocol proposed in the SVC-onGoing competition. The results obtained in our experiments are promising, and promote the use of Transformers for on-line signature verification.

Keywords

biometrics, transformers, signature verification, pattern recognition

1. Introduction

On-line handwritten signature verification is a biometric modality that aims to verify the authenticity of subjects based on their personal signatures. Handwritten signatures have been traditionally used for biometric personal recognition, as they contain unique behavioral patterns that can serve as reliable identifiers [1, 2]. Early approaches focused on extracting dynamic or local features [3, 4], such as pen pressure, stroke sequences, speed, and acceleration, and leveraging machine learning and pattern recognition techniques, such as Dynamic Time Warping (DTW) [4, 5] and Hidden Markov Models (HMM) [6, 7]. With the integration of Deep Learning (DL) [8, 9, 10, 11], on-line signature verification systems have achieved remarkable performance, exhibiting robustness against various forms of forgeries [12] and improving user experience [13]. This technology has wide-ranging applications in areas such as e-commerce, digital banking, and document verification, contributing to the prevention of identity fraud and improving the overall security of on-line transactions [2, 14].

Workshop on Advances of Mobile and Wearable Biometrics, at the ACM International Conference on Mobile Human-Computer Interfaces (MobileHCI), 2023.

*Corresponding author.

Despite the success of DL approches based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [13], recent studies have explored the application of Transformer architectures for other behavioral biometric traits such as gait and keystroke, outperforming the state of the art [15, 16, 17]. Among the multiple advantages of Transformers, we highlight the ability to capture long-term dependencies and interactions, which is especially attractive for time series modeling [18].

In this paper we explore the use of Transformers for on-line signature verification, in which signatures are acquired with pen tablet devices able to capture X and Y spatial coordinates, pen pressure, and timestamps. We investigate four different Transformer configurations: *i)* a Vanilla Transformer encoder [19], *ii)* the THAT Transformer successfully applied to activity recognition [20], *iii)* a Transformer successfully applied to gait recognition [15, 17], and *iv)* a novel configuration based on the Temporal and Channel modules proposed in THAT [20], but with the Transformer encoder proposed in Vanilla [19]. To obtain a fair comparison with the literature, we evaluate the proposed configurations according to the experimental protocol proposed in the SVC-onGoing competition [13]. In particular, we compare the results with our recent Time-Aligned Recurrent Neural Network (TA-RNN) [8]. TA-RNN combines the potential of DTW and RNNs to train more robust systems against forgeries.

2. Methods

We explore a Siamese architecture with four different Transformer configurations for on-line signature verification. From the original time signals acquired by the device (X and Y spatial coordinates and pen pressure), we extract the set of 23 local features proposed in [6], obtaining additional time signals related to velocity, acceleration, geometric aspects of the signature, etc. These time signals are used as input to our Siamese architecture.

2.1. Transformer Configurations

Differently from other behavioral biometrics (*e.g.,* gait and keystroke), on-line signatures usually consist in longer sequences. Hence, the processing of time signals with Transformer-based architectures is computationally expensive. Similarly to some approaches proposed for voice analysis [21, 22], we add convolutional layers prior to Transformers, to aggregate temporal features of the input signals and reduce their dimensionality. A general representation of the proposed architecture is provided in Figure 1.

The CNN layers consist of a combination of 1D convolutional and MaxPooling layers. Convolutional layers extract temporal features by combining consecutive timesteps and augment the feature size, thanks to the 64 channels in the output. MaxPooling layers reduce timesteps, making the signals more suitable for Transformers.

Using the Siamese architecture proposed in Figure 1, we consider four different configurations depending on the Transformer module selected. First, we consider the original Vanilla Transformer encoder [19], with Gaussian Range Encoding instead of Positional Encoding, given its suitability with the time series of interest [15]. The Vanilla encoder processes inputs with size 250×64 and generates a vector of size 64 at each timestep. These vectors are processed by a RNN, whose final state of size 92 is concatenated in the Siamese architecture (see Figure 1).

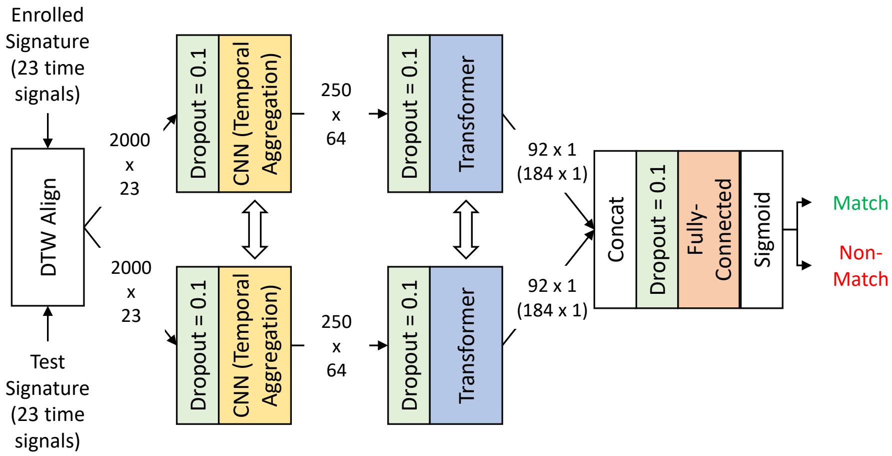

Figure 1: Representation of proposed Siamese architecture based on the combination of CNNs and Transformers.

The second and third approaches are based respectively on the THAT Transformer proposed for activity recognition [20] and the Transformer proposed for gait recognition [15].

Finally, we explore a novel configuration based on the Temporal and Channel modules proposed in THAT [20], but with the Transformer encoder proposed in Vanilla [19]. The Temporal branch is analogous to the one considered in the first configuration. The Channel branch processes inputs with size 64×250 and generates a vector of size 250 for each channel. We average these vectors and apply a Fully-Connected layer to reduce the size of the output to 92 (the same of Temporal branch). The outputs of Temporal and Channel branches are concatenated, being the final vector of size 184 (Figure 2).

3. Experimental Setup

We evaluate our configurations with the experimental protocol proposed in the SVC-onGoing competition [13]. Two publicly available datasets are considered in the competition: DeepSignDB [8], used for development and validation, and SVC2021_EvalDB [13], used for the final evaluation. Different subjects are considered in each dataset. In particular, we focus on the office-like scenario where subjects had to perform signatures using a pen tablet device.

To train our four configurations, we generate random pairs of matching and non-matching signatures from the Development DeepSignDB dataset provided by the SVC-onGoing competition. We consider both random and skilled non-matching pairs for training. Finally, following our previous TA-RNN approach [8], for each signature comparison (i.e., genuine-genuine or genuine-impostor) we align the 23 time signals of each signature pair with DTW, zero-padding them to obtain a fixed length of 2,000 time samples for each signature.

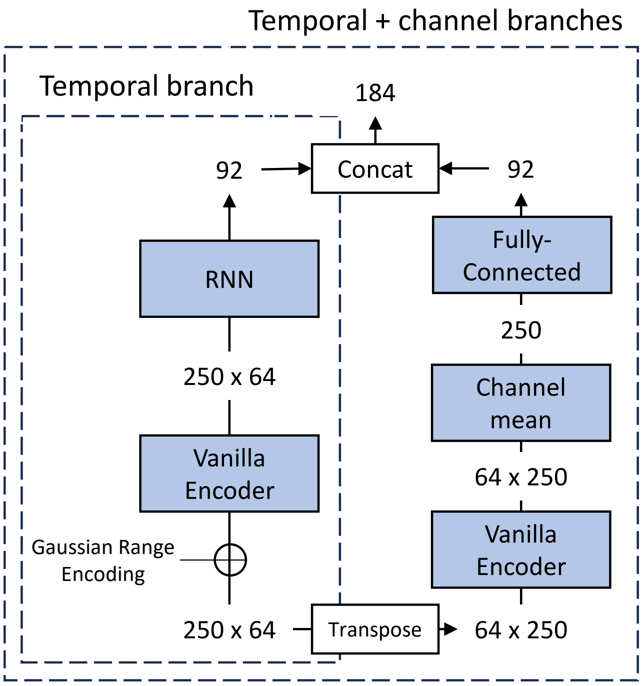

Figure 2: Representation of the proposed configuration based on the original Vanilla encoder, with Temporal and Channel branches. For Vanilla Encoder (with Temporal branch only) refer to the small box.

Table 1
EER results obtained in the SVC-onGoing competition for the proposed configurations. R = Random, S = Skilled, O = Overall.

Configuration	**DeepSignDB**			**SVC2021_EvalDB**		
	R	S	O	R	S	O
TA-RNN [13, 23]	1.87%	5.77%	**4.31%**	1.91%	5.83%	4.08%
Vanilla encoder [19]	3.25%	5.92%	4.64%	3.52%	4.06%	**3.80%**
THAT Transf. [20]	4.57%	8.02%	6.35%	5.13%	6.75%	6.03%
Gait Transformer [15]	3.84%	6.42%	5.13%	3.60%	4.44%	4.10%
Vanilla encoder (Temporal + Channel)	4.13%	6.60%	5.36%	4.58%	4.42%	4.49%

4. Results

The results obtained by evaluating our four Transformer configurations according to the protocol of the SVC-onGoing competition are reported in Table 1. We consider non-match comparisons made of random signatures (R), skilled signatures (S), and an overall combination of the two (O). Random signatures is the type of impostors that always provide the lowest EER, except in the case of the Vanilla encoder with Temporal and Channel branches, evaluated on SVC2021_EvalDB that provides 4.42% EER for skilled non-match comparisons and 4.58% EER for random ones.

The Transformer configurations achieve similar performance compared to the TA-RNN previously presented. From the four Transformer configurations explored, we observe that the Vanilla encoder achieves the best overall results, 4.64% EER and 3.80% EER for the DeepSignDB and SVC2021_EvalDB, respectively. Focusing on the SVC2021_EvalDB dataset considered for the final evaluation of the competition, we can observe how the Vanilla encoder achieves a relative improvement of 7% in comparison to the TA-RNN approach (4.08% EER), showing a better generalisation ability to the new scenarios not considered in training.

The results of Table 1 show how more complex configurations do not improve the results obtained in evaluation. Overall EERs raise from 4.64% to 5.36% for DeepSignDB and from 3.80% to 4.49% for SVC2021_EvalDB when we add the channel branch to the configuration based on Vanilla encoder. Similar results apply to the other two configurations considered, with the Transformer proposed for gait recognition that only get closer to the best performances, with overall EERs of 5.13% and 4.10% in DeepSignDB and SVC2021_EvalDB evaluations.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860813 - TReSPAsS-ETN. With support also from projects INTER-ACTION (PID2021-126521OB-I00 MICINN/FEDER) and HumanCAIC (TED2021-131787B-I00 MICINN).

References

[1] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo, R. Plamondon, A perspective analysis of handwritten signature technology, ACM Computing Surveys 51 (2019) 1–39.

[2] M. Faundez-Zanuy, J. Fierrez, M. A. Ferrer, M. Diaz, R. Tolosana, R. Plamondon, Handwriting biometrics: Applications and future trends in e-security and e-health, Cognitive Computation 12 (2020) 940–953.

[3] M. Martinez-Diaz, J. Fierrez, R. P. Krish, J. Galbally, Mobile signature verification: Feature robustness and performance comparison, IET Biometrics 3 (2014) 267–277.

[4] R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, J. Fierrez, Preprocessing and feature selection for improved sensor interoperability in online biometric signature verification, IEEE Access 3 (2015) 478–489.

[5] R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez, J. Ortega-Garcia, Exploiting complexity in pen-and touch-based signature biometrics, International Journal on Document Analysis and Recognition 23 (2020) 129–141.

[6] J. Fierrez, J. Ortega-Garcia, D. Ramos, J. Gonzalez-Rodriguez, Hmm-based on-line signature verification: Feature extraction and signature modeling, Pattern Recognition Letters 28 (2007) 2325–2334.

[7] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Reducing the template ageing effect in on-line signature biometrics, IET Biometrics 8 (2019) 422–430.

[8] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Deepsign: Deep on-line signature verification, IEEE Transactions on Biometrics, Behavior, and Identity Science 3 (2021) 229–239.

[9] J. Jiang, S. Lai, L. Jin, Y. Zhu, Dsdtw: Local representation learning with deep soft-dtw for dynamic signature verification, IEEE Transactions on Information Forensics and Security 17 (2022) 2198–2212.

[10] R. Tolosana, P. Delgado-Santos, A. Perez-Uribe, R. Vera-Rodriguez, J. Fierrez, A. Morales, DeepWriteSYN: On-line handwriting synthesis via deep short-term representations, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 600–608.

[11] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, BioTouchPass2: Touchscreen password biometrics using time-aligned recurrent neural networks, IEEE Transactions on Information Forensics and Security 15 (2020) 2616–2628.

[12] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Presentation attacks in signature biometrics: types and introduction to attack detection, Handbook of Biometric

Anti-Spoofing: Presentation Attack Detection (2019) Sébastien Marcel, Julian Fierrez, and Nicholas Evans (Eds.). Springer Nature, 439–453.

[13] R. Tolosana, R. Vera-Rodriguez, C. Gonzalez-Garcia, J. Fierrez, A. Morales, J. Ortega-Garcia, J. C. Ruiz-Garcia, S. Romero-Tapiador, S. Rengifo, M. Caruana, et al., Svc-ongoing: Signature verification competition, Pattern Recognition 127 (2022) 108609.

[14] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Benchmarking desktop and mobile handwriting across cots devices: The e-biosign biometric database, PLOS ONE 12 (2017) e0176792.

[15] P. Delgado-Santos, R. Tolosana, R. Guest, F. Deravi, R. Vera-Rodriguez, Exploring transformers for behavioural biometrics: A case study in gait recognition, Pattern Recognition (2023).

[16] G. Stragapede, P. Delgado-Santos, R. Tolosana, R. Vera-Rodriguez, R. Guest, A. Morales, Typeformer: Transformers for mobile keystroke biometrics, arXiv preprint arXiv:2212.13075 (2022).

[17] P. Delgado-Santos, R. Tolosana, R. Guest, R. Vera-Rodriguez, J. Fierrez, M-GaitFormer: Mobile Biometric Gait Verification using Transformers, Engineering Applications of Artificial Intelligence (2023).

[18] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, L. Sun, Transformers in time series: A survey, arXiv preprint arXiv:2202.07125 (2022).

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of Advances in Neural Information Processing Systems, volume 30, 2017.

[20] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, M. Wu, Two-stream convolution augmented transformer for human activity recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 286–293.

[21] K. Chen, Y. Wu, Z. Wang, X. Zhang, F. Nian, S. Li, X. Shao, Audio captioning based on transformer and pre-trained cnn, in: Proceedings of Detection and Classification of Acoustic Scenes and Events, 2020, pp. 21–25.

[22] Q. Kong, Y. Xu, W. Wang, M. D. Plumbley, Sound event detection of weakly labelled data with CNN-Transformer and automatic threshold optimization, IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020) 2450–2460.

[23] C. Gonzalez-Garcia, R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Introduction to Presentation Attacks in Signature Biometrics and Recent Advances, Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment (2023) Sébastien Marcel, Julian Fierrez, and Nicholas Evans (Eds.). Springer Nature, 447–466.