Secure Triplet Loss for End-to-End Deep Biometrics

João Ribeiro Pinto
INESC TEC & Universidade do Porto
Porto, Portugal
joao.t.pinto@inesctec.pt

Jaime S. Cardoso
INESC TEC & Universidade do Porto
Porto, Portugal
jaime.cardoso@fe.up.pt

Miguel V. Correia
INESC TEC & Universidade do Porto
Porto, Portugal
mcorreia@fe.up.pt

Abstract—**Although deep learning is being widely adopted for every topic in pattern recognition, its use for secure and cancelable biometrics is currently reserved for feature extraction and biometric data preprocessing, limiting achievable performance. In this paper, we propose a novel formulation of the triplet loss methodology, designated as secure triplet loss, that enables biometric template cancelability with end-to-end convolutional neural networks, using easily changeable keys. Trained and evaluated for electrocardiogram-based biometrics, the network revealed easy to optimize using the modified triplet loss and achieved superior performance when compared with the state-of-the-art (10.63% equal error rate with data from 918 subjects of the UofTDB database). Additionally, it ensured biometric template security and effective template cancelability. Although further efforts are needed to avoid template linkability, the proposed secure triplet loss shows promise in template cancelability and non-invertibility for biometric recognition while taking advantage of the full power of convolutional neural networks.**

I. Introduction

Biometric recognition systems are quickly replacing traditional authentication and identity control systems in almost all contexts and applications. While traditional systems are at risk when the user forgets, loses, or shares the credentials, biometric systems do not require the user to know or carry any external credentials, as they perform recognition using intrinsic characteristics such as facial features, fingerprints, iris, voice characteristics, or physiological signals (such as the electrocardiogram) [1]–[3].

However, when a traditional access control system is attacked and the contents of its database are accessed by an intruder, the only things that become compromised are the access credentials, which can easily be changed to avoid greater losses. On an attacked biometric system, what becomes compromised is a part of the individual, difficult to change, as the database stores intrinsic personal data from each of the users. Hence, the development of biometric systems requires redoubled efforts to ensure the security of the stored biometric data [2], [4].

Cryptography and information theory concepts can be used for data encryption and protection in biometric systems. However, biometric systems require special methods, adapted to adequately deal with the variability of the biometric traits and measurements [4], [5]. These methods need to ensure no positive matches occur between signals of different subjects while avoiding negative matches between samples of the same subject, despite the intrasubject trait variability.

Besides accounting for trait variability, data protection methods in biometric systems need to verify three other properties. First, the stored templates need to be easily and effectively cancelable if these become compromised (this property is called cancelability or revokability) [6], [7]. Additionally, the transformation from trait measurements to templates should be as close to irreversible as possible, as it should be impossible or infeasible for attackers to retrieve an approximation of the original trait using a compromised template (non-invertibility property) [4], [5]. Also, it should be hard for an attacker to know whether two samples from different systems belong to the same individual (non-linkability property) [8].

Several methods have been proposed for securing biometric templates [8]–[14]. Most approaches are based on salting, biohashing, or cryptographic protection methods [5]. Although some methods have been proposed for secure biometrics using deep learning [15], [16], none are end-to-end as they require separate decision processes able to deal with hashed or key-transformed templates. Furthermore, the additional separate processes of template protection often influence negatively the performance of the biometric algorithms [4].

In this work, we propose an adaptation of the triplet loss technique that enables training of end-to-end convolutional neural networks for secure biometric authentication. The proposed method is able to safeguard the biometric templates using binary keys while dismissing any additional process beyond the deep network. Hence, it allows one to take full advantage of the capabilities of end-to-end deep neural networks while still ensuring the security of the stored biometric data.

The proposed methodology was implemented and evaluated for electrocardiogram-based biometric authentication, on the University of Toronto ECG Database. End-to-end convolutional neural networks have shown improved performance

This work was financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalization - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project "POCI-01-0145-FEDER-030707", and within the PhD grant "SFRH/BD/137720/2018".

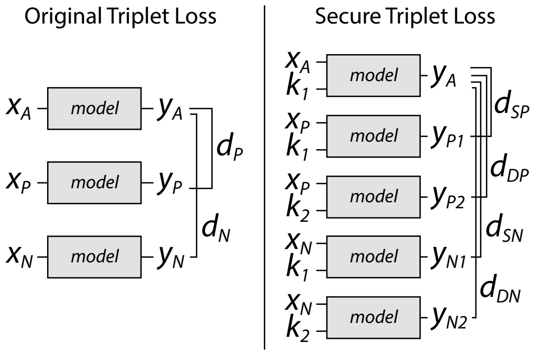

Fig. 1. Comparison between the model training schemas of the proposed secure triplet loss method (right) and the original triplet loss (left).

in off-the-person ECG-based biometrics [17], [18]. Although the literature includes some work in ECG template protection [19]–[21], no approaches have been formulated for end-to-end networks. Hence, this work evaluated the effect of the inclusion of template security measures in such models. Besides the authentication performance, the experiments also assess the proposed method's capabilities in template cancelability, non-invertibility, and non-linkability.

II. Secure Triplet Loss

The triplet loss [22] has been widely used in deep learning to train networks to accurately determine whether or not two samples belong to the same class [18], [23], [24]. During training, such networks receive three inputs (a triplet), in parallel: one is the anchor (x_A, the reference with identity i_A), the second is the positive sample (x_P, with identity $i_P = i_A$), and the third is the negative sample (x_N, with identity $i_N \neq i_A$). In biometrics, triplets are groups of three biometric trait measurements (images or signals): the anchor and positive inputs correspond to the same individual, unlike the negative input.

For each input, the network will output a representation: *e. g.*, for the anchor, $y_A = f(x_A)$. The three representations are then compared using a measure of distance or dissimilarity $d(y_1, y_2)$, and the network is optimized through the minimization of the triplet loss function:

$$l = \max\left[0, \alpha + d(y_A, y_P) - d(y_A, y_N)\right], \tag{1}$$

which lead representations of the same class to be more similar than those of different classes, maximizing $d(y_A, y_N)$ and minimizing $d(y_A, y_P)$. The loss also aims to enforce a minimum margin α between the two distances.

This is a generally successful strategy when training neural networks for biometric authentication (verifying if the identities of a stored template and a current biometric measurement match). However, it does not address the important issue of security in biometrics, especially the topic of cancelability.

Hence, the proposed training method modifies the triplet loss to make the final sample representations cancelable (as illustrated in Fig. 1). Besides the triplet inputs (x_A, x_P, and x_N), the network will also receive two different keys (k_1, k_2) that are bound with the inputs by the network itself.

Unlike the original triplet loss, x_P and x_N are processed by the network twice. First they will be combined with k_1 and then with k_2. The anchor x_A is only bound with k_1. Thus, five representations will be obtained: $y_A = f(x_A, k_1)$, $y_{P1} = f(x_P, k_1)$, $y_{P2} = f(x_P, k_2)$, $y_{N1} = f(x_N, k_1)$, $y_{N2} = f(x_N, k_2)$. From these, four distances are computed: $d_{SP} = d(y_A, y_{P1})$ (with matching identities and keys), $d_{DP} = d(y_A, y_{P2})$ (with matching identities but different keys), $d_{SN} = d(y_A, y_{N1})$ (with different identities but matching keys), and $d_{DN} = d(y_A, y_{N2})$ (with non-matching identities and keys).

The objective is to minimize d_{SP}, when both the identities and the keys match, and maximize the remaining three distances. Hence, the loss is computed through:

$$l = \max\left(0, \alpha + d_{SP} - d_n\right), \tag{2}$$

where d_n will result of the combination of all three distances to be maximized. One option is to set $d_n = \min(\{d_{SN}, d_{DP}, d_{DN}\})$, with the three distances to be maximized being considered equally relevant. This results in:

$$l = \max\left[0, \alpha + d_{SP} - \min(\{d_{SN}, d_{DP}, d_{DN}\})\right]. \tag{3}$$

Alternatively, one can opt for a loss formulation where d_n is randomly chosen randomly among $\{d_{SN}, d_{DP}, d_{DN}\}$:

$$l = \max\left(0, \alpha + d_{SP} - d_n\right), \; d_n \in \{d_{SN}, d_{DP}, d_{DN}\}. \tag{4}$$

Choosing only one of these three distances means inference is only needed for three input combinations (x_A with k_1, x_P with k_1, and the sample-key pair corresponding to the chosen d_n), hastening the training process. The random choice of d_n for each triplet ensures the balanced optimization of the model according to all three distances.

As with triplet loss, α will enforce a margin between positive and negative distances. In this case, the loss involves four distances, since it also takes into account whether or not the keys match. By minimizing the loss in either Eq. (3) or Eq. (4), the network learns to deal with the intrasubject and intersubject variability of the biometric trait. More importantly, it learns to recognize when the keys do not match, even if the identity is the same. Hence, if the stored templates become compromised, they can easily be invalidated through a key change.

III. Experimental Settings

The proposed training methodology was applied to off-the-person electrocardiogram-based biometric authentication, using signals from the University of Toronto ECG Database (UofTDB) [25]. This database includes electrocardiogram (ECG) recordings from 1019 individuals, using dry electrodes on the fingers at 200 Hz sampling frequency, on up to six sessions over six months, and up to five subject positions.

Five-second segments (1000 samples at 200 Hz sampling frequency) from the last 100 subjects in the database were used for training, while the data from the remaining 918 subjects

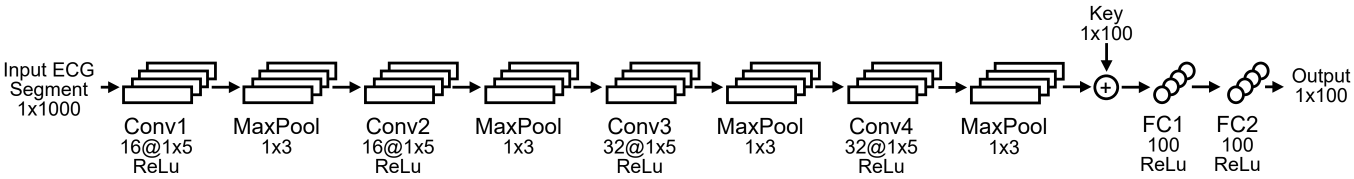

Fig. 2. Architecture of the model trained for ECG-based authentication.

were reserved for testing. The anchors were taken from the first 30 seconds of the recording of each subject, while the positive and negative samples were drawn randomly among the remaining data from the same individual or a different subject, respectively.

For training, 90 000 triplets were generated, as well as 10 000 for validation, and another 10 000 for testing. The networks trained with the original and modified triplet loss functions used the same sets of triplets. Pairs of different arrays, each with 100 binary values, were randomly generated to serve as keys.

The binding of an input and a key happened before the first dense layer, after the input underwent processing in all convolutional layers. Before being received by the first dense layer, the vector of flattened feature maps ($s(x)$) was concatenated with the key k (after its normalization to unit $l2$ norm). The vector resulting from the concatenation of $s(x)$ and k is used by the dense layers to obtain the respective representation $y = f(s(x), k)$.

The network was implemented on Python using Keras with Tensorflow as backend. Its architecture is depicted in Fig. 2. Euclidean and normalized Euclidean distance [26] were used as distance measures, respectively, for training and testing in authentication. The model was trained using the Adam optimizer with an initial learning rate of 0.0001, for a maximum of 500 epochs, with early stopping based on validation loss (patience of 20 epochs).

The security of the method was evaluated, using the privacy leakage rate, the secrecy leakage, and the secret key rate. The privacy leakage rate is used to measure non-invertibility, and can be computed through the expression:

$$\frac{H(X|Y)}{H(X)} = 1 - \frac{I(X;Y)}{H(X)}, \tag{5}$$

where X is the input biometric measurement, Y is the output of the model, $H(X)$ denotes the entropy of X, $H(X|Y)$ denotes the conditional entropy of X given Y, and $I(X;Y)$ denotes the mutual information between X and Y. The privacy leakage rate should be as high as possible: even when one has all knowledge of Y, obtaining information on X should be impossible.

The secrecy leakage measures the mutual information between the stored template Y and the key K, through the expression $I(Y;K)$. The keys are public, unlike the templates, so they should reveal as little information as possible on the templates. Hence, the secrecy leakage should be close to zero. Finally, the secret key rate measures the uncertainty in the model output Y, and is computed through $H(Y)$. The higher it is, the more variability the output presents, and the harder it is to successfully attack the biometric system.

These require the computation of some information theoretical measures, such as entropy and mutual information. This is very difficult in biometrics, due to the high dimensionality of the inputs and the feature sets and their variability. Hence, the viable option is to estimate those measures. In this work, entropy and mutual information were estimated using the methods proposed in [27] and in [28], respectively, through their Python implementation described in [29]. These methods, based on nearest neighbor statistics, were shown to be more accurate than the alternatives [30]. Although the original papers propose setting $k \in [2, 4]$, this parameter was adjusted to $k = 10$ to avoid errors regarding negative mutual information estimation results.

The template linkability analysis followed the method described by Gomez-Barrero *et al.* [8]. The aforementioned test samples were paired into mated (different biometric samples from the same identity with different keys) and non-mated instances (different identities and keys). These have been used to compute $p(s|H_m)$ and $p(s|H_{nm})$: the probability density functions of the dissimilarity score s given the instances are, respectively, mated (hypothesis H_m) or non-mated (hypothesis H_{nm}). From the likelihood ratio $LR(s) = p(s|H_m)/p(s|H_{nm})$, $D_{\leftrightarrow}(s)$ and the $D_{\leftrightarrow}^{sys}$ linkability metric were computed (Eq. (6) and Eq. (7), respecively).

$$D_{\leftrightarrow}(s) = \begin{cases} 0, & \text{if } LR(s) \leq 1 \\ 2\left(\left(1 + \mathrm{e}^{-(LR(s)-1)}\right)^{-1} - \frac{1}{2}\right), & \text{if } LR(s) > 1 \end{cases} \tag{6}$$

$$D_{\leftrightarrow}^{sys} = \int_{s_{min}}^{s_{max}} D_{\leftrightarrow}(s) \cdot p(s|Hm) \, \mathrm{d}s \tag{7}$$

IV. Results and Discussion

After training the network using the original and the modified version of the triplet loss (see Fig. 3 for the evolution of loss over training time), the authentication performance was evaluated through the analysis of false acceptance (FAR) and rejection rates (FRR), receiver-operating characteristic curves (ROC, presented in Fig. 4), and equal error rates (EER).

The proposed method achieved 10.63% EER, versus 12.55% EER for the original triplet loss. These results stand in stark contrast to those reported in the literature, which point to a five-fold average increase in authentication error when security measures are included [4]. The advantage of the proposed method likely results from using an end-to-end deep

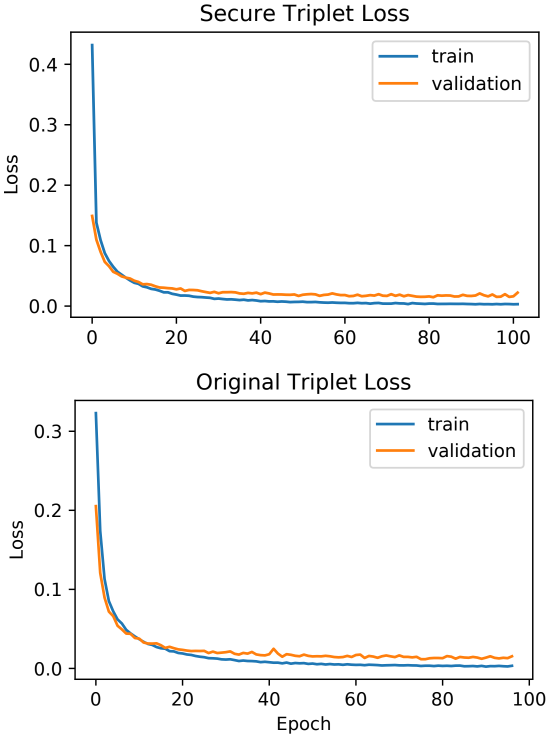

Fig. 3. Train and validation loss evolution over training epochs for the network trained with the modified (top) and original triplet loss (bottom).

neural network, thus taking full advantage of their enhanced robustness to biometric variability.

With the loss formulation in Eq. (4), the trained model offered 11.96% EER. As expected, there has been a small performance decline relative to the formulation in Eq. (3), but the error is still lower than that of the model trained with the original triplet loss. Hence, the more efficient formulation can be used in circumstances where time is limited and small performance losses are admissible. For brevity, all results presented below refer to the formulation in Eq. (3).

These results can also be compared with the performance offered by state-of-the-art methods, implemented and tested in the same conditions in [18]. Despite the inclusion of security measures, the proposed methodology improves over the results of all implemented methods, including the end-to-end models trained with the original triplet loss (13.93% EER) or using transfer learning from identification (13.70% EER).

Regarding the information-theoretical security measures, the network offered the same, approximately perfect privacy rate result when trained with either the original or the modified triplet loss. This may be linked with claims in the literature that, for appropriately optimized convolutional neural networks, the mutual information between inputs and outputs is minimized. When optimized, internal representations compress the input and maintain only the information needed for the task at hand [31]. This makes CNN models very useful to ensure non-invertibility for biometric templates.

The analysis of secrecy leakage for the proposed method rendered similar results to the privacy leakage rate, with a perfect score of zero. Once again, this may be related to the nature of deep neural networks but is, nevertheless, highly beneficial for biometric security. As for secret key rates, the proposed method proved superior to the original triplet loss, with 103.73 bits of output entropy versus 14.20 bits. This means the secure method proposed in this paper should be harder to successfully attack than a network trained with the original triplet loss formulation.

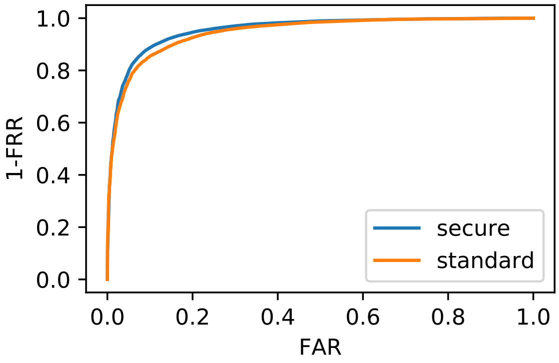

Fig. 4. Comparison between the ROC curves of the model trained with the proposed methodology (in blue) and with the original triplet loss (in orange).

Using Singular Value Decomposition, some plots of samples in 2D projections were analyzed to study variability and separability in the network outputs for different subjects and different keys (see Fig. 5). In these plots, it is easy to see that triplet loss improves class separability, as intended, making classes more compact and clustered. Using the secure modification of triplet loss proposed in this paper, however, the classes become more scrambled as, even though the identities match, the keys are different. This is a useful property for the sake of template cancelability.

One final two-dimensional projection, in Fig. 6, shows the effects on the separability of different keys on data of the same subjects and the same matching keys on data from different subjects. It is visible that data from a given subject shift when the key is changed. In the case illustrated by the figure, the clusters of each subject data suffer a shift towards the bottom-right, which would be sufficient to invalidate templates corresponding to canceled keys.

On the other hand, one might argue that separability would be very weak if every enrolled individual had the same key, as the network might rely on keys for its identity decision. However, the loss formulation guides the network training to avoid such behavior, and the equal error rate result, which addresses this specific aspect, is lower than when using the original triplet loss. In fact, considering the data used, the proposed method exceeds the state-of-the-art in terms of EER (see [18]). Moreover, as keys are randomly generated binary arrays, one can force keys to be different for each subject, or increase their length to make such issue even more unlikely to happen. It would even be possible to force new subject keys to be sufficiently different from past keys, to ensure a sufficiently

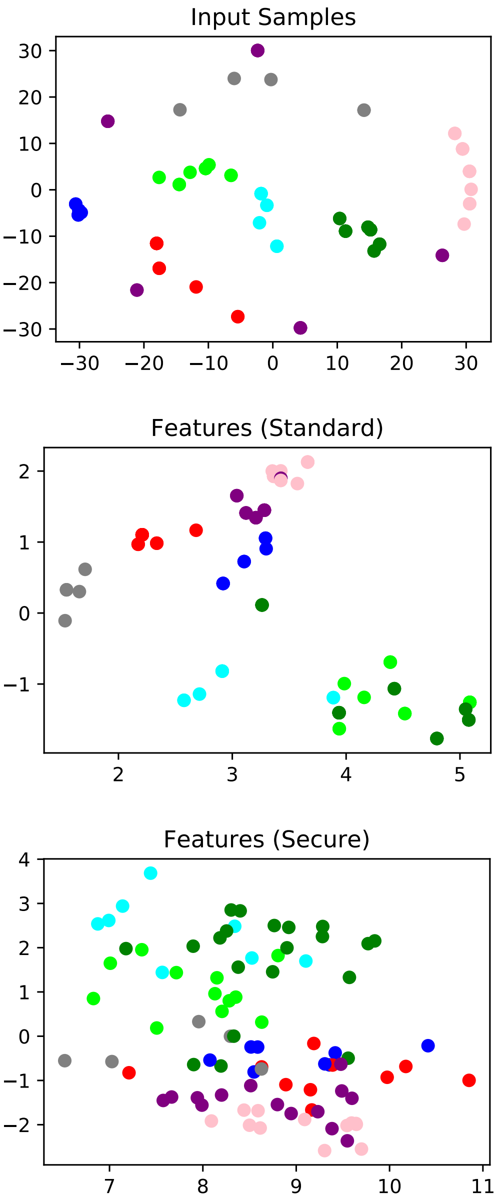

Fig. 5. Illustration of separability between classes in the original data (top), the network's output with triplet loss (center), and the network's output with the proposed method (bottom). Each dot represents a biometric sample, and different colors denote different subjects. The dispersion observed on the secure network's outputs is caused by the binding of the inputs with different keys.

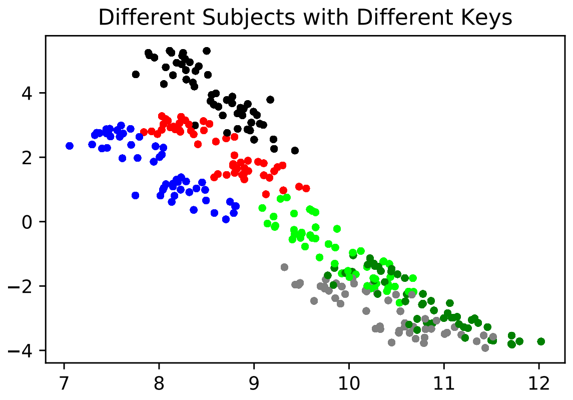

Fig. 6. Scatter plot of the network outputs from samples of different subjects (in different colours) binded with two different keys.

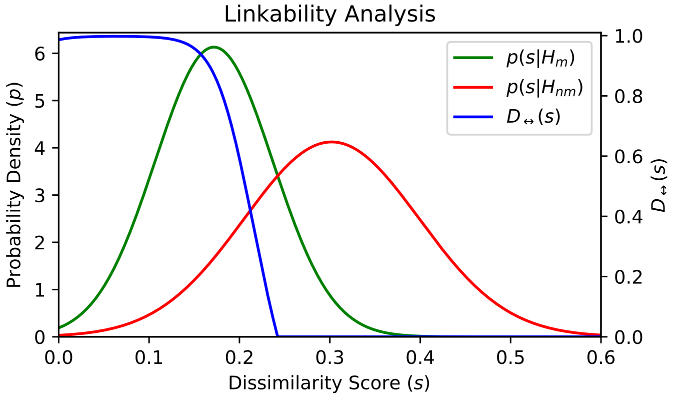

Fig. 7. Linkability analysis of the proposed model.

large shift and an effective cancellation of past templates.

At last, regarding the linkability of templates across different systems, the results of the experiments are presented in Fig. 7. The proposed secure triplet loss model offered $D_{\leftrightarrow}^{sys} = 0.67$. Comparing this value and the $D_{\leftrightarrow}(s)$ curve with the guidelines and examples offered in [8], the method is between semi-linkable and fully linkable. This means it would be relatively easy for an attacker to discover whether two samples with different keys belong to the same subject. This is an important shortcoming of the method that stands in stark contrast with the promising recognition performance, cancelability, and non-invertibility results presented above.

The end-to-end model apparently learned a balance between the weights of identities and keys in the output dissimilarity scores. However, non-linkability would require it to disregard identity when keys don't match, in order to offer the similar dissimilarity scores in these cases regardless of whether the identities match. Identity should only influence the dissimilarity score when the keys of both templates correspond. Future research endeavors should focus on adapting the network to promote this desired behavior and avoid template linkability.

V. Conclusion

This work proposes an adaptation of the widely used triplet loss training scheme to ensure biometric template security in convolutional neural networks. The proposed method is applicable to end-to-end models, which dismiss any separate processes and take full advantage of the capabilities of deep learning. Template cancelability is achieved through the combination of the templates with easily changeable keys.

When tested for ECG-based biometric authentication, the proposed method revealed several advantages over the original triplet loss. Template cancelability is ensured and easy to achieve by changing the respective key. Contrarily to the great majority of existing template security methods, the proposed methodology does not degrade biometric performance. In fact, performance results are improved over the use of the original triplet loss.

The privacy leakage rate is approximately perfect, which means it is almost impossible to retrieve the original signal given the stored template. Finally, the secret key rate is high, even with low dimensionality of the stored templates (100 features), which makes successful attacks more difficult.

Nevertheless, some limitations remain, especially regarding non-linkability. The results of the linkability analysis show the greatest shortcoming of the proposed method. It should be adapted to make it harder for attackers to know whether two templates with different keys belong to the same person. Furthermore, although the proposed method exceeds the state-of-the-art in ECG biometrics, the equal error rate results obtained with ECG are still far above those offered by other biometric traits such as face or fingerprints.

In the future, the proposed method should first be adapted to give appropriate weight to the keys and identities on the dissimilarity scores, according to whether or not the keys match. The training loss could be further adapted so that the network is specifically optimized to avoid template linkability. Then, the proposed method could be applied and evaluated for other biometric traits, to assess its wider applicability, security, and performance advantages.

Acknowledgments

The authors wish to acknowledge the creators and administrators of the UofTDB database (University of Toronto, Canada), which was essential for this work.

References

[1] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Springer Publishing Company, Incorporated, 2011.

[2] T. Ignatenko and F. M. Willems, "Biometric security from an information-theoretical perspective," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 2–3, pp. 135–316, 2012.

[3] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, Current Challenges, and Future Possibilities in ECG Biometrics," *IEEE Access*, vol. 6, pp. 34 746–34 776, 2018.

[4] K. Nandakumar and A. K. Jain, "Biometric Template Protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 88–100, 2015.

[5] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Adv. Signal Process*, vol. 2008, pp. 113:1–113:17, 2008.

[6] P. Punithavathi and G. Subbiah, "Can cancellable biometrics preserve privacy?" *Biometric Technology Today*, vol. 2017, no. 7, pp. 8–11, 2017.

[7] M. Tarek, O. Ouda, and T. Hamza, "Robust cancellable biometrics scheme based on neural networks," *IET Biometrics*, vol. 5, no. 3, pp. 220–228, September 2016.

[8] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, and J. Fierrez, "Unlinkable and irreversible biometric template protection based on bloom filters," *Information Sciences*, vol. 370-371, pp. 18–32, 2016.

[9] A. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme," in *Proceedings of the 6th ACM Conference on Computer and Communications Security*, New York, NY, USA, 1999, pp. 28–36.

[10] A. Teoh, D. Ngo, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognition*, vol. 37, no. 11, pp. 2245 – 2255, 2004.

[11] Y. Sutcu, H. T. Sencar, and N. Memon, "A secure biometric authentication scheme based on robust hashing," in *Proceedings of the 7th Workshop on Multimedia and Security*, August 2005, pp. 111–116.

[12] M. Gomez-Barrero, C. Rathgeb, G. Li, R. Ramachandra, J. Galbally, and C. Busch, "Multi-biometric template protection based on bloom filters," *Information Fusion*, vol. 42, pp. 37–50, 2018.

[13] P. Drozdowski, S. Garg, C. Rathgeb, M. Gomez-Barrero, D. Chang, and C. Busch, "Privacy-preserving indexing of iris-codes with cancelable bloom filter-based search structures," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 2360–2364.

[14] K. B. Raja, R. Raghavendra, M. Stokkenes, and C. Busch, "Biometric template protection on smartphones using the manifold-structure preserving feature representation," in *Selfie Biometrics: Advances and Challenges*, A. Rattani, R. Derakhshani, and A. Ross, Eds. Cham: Springer International Publishing, 2019, pp. 299–312.

[15] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, "Multibiometric secure system based on deep learning," *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 298–302, 2017.

[16] R. K. Pandey, Y. Zhou, B. U. Kota, and V. Govindaraju, "Deep secure encoding for face template protection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2016, pp. 77–83.

[17] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Deep Neural Networks For Biometric Identification Based On Non-Intrusive ECG Acquisitions," in *The Biometric Computing: Recognition and Registration*, K. V. Arya and R. S. Bhadoria, Eds. Boca Raton FL, United States: CRC Press, 2019, ch. 11, pp. 217–234.

[18] J. R. Pinto and J. S. Cardoso, "A end-to-end convolutional neural network for ECG based biometric authentication," in *10th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2019)*, 2019.

[19] S. Wu, P. Chen, A. L. Swindlehurst, and P. Hung, "Cancelable Biometric Recognition With ECGs: Subspace-Based Approaches," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1323–1336, May 2019.

[20] H. Kim and S. Y. Chun, "Cancelable ECG Biometrics Using Compressive Sensing-Generalized Likelihood Ratio Test," *IEEE Access*, vol. 7, pp. 9232–9242, 2019.

[21] H. Kim, M. P. Nguyen, and S. Y. Chun, "Cancelable ECG biometrics using GLRT and performance improvement using guided filter with irreversible guide signal," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017, pp. 454–457.

[22] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," *Journal of Machine Learning Research*, vol. 11, pp. 1109–1135, 2010.

[23] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[24] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person Re-Identification by Multi-Channel Parts-Based CNN With Improved Triplet Loss Function," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[25] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, "On Evaluating ECG Biometric Systems: Session-Dependence and Body Posture," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, Nov. 2014.

[26] Wolfram Language and System Documentation Center, "Normalized square euclidean distance," 2010, (last accessed on 22-11-2019). [Online]. Available: http://reference.wolfram.com/language/ref/NormalizedSquaredEuclideanDistance.html

[27] L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," *Problemy Peredachi Informatsii*, vol. 23, pp. 9–16, 1987.

[28] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, p. 066138, Jun 2004.

[29] P. Brodersen, "Entropy estimators," 2017, (last accessed on 22-11-2019). [Online]. Available: https://github.com/paulbrodersen/entropy_estimators

[30] G. Doquire and M. Verleysen, "A comparison of multivariate mutual information estimators for feature selection," in *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM,*, 2012, pp. 176–185.

[31] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," in *2015 IEEE Information Theory Workshop (ITW)*, April 2015, pp. 1–5.