

# Supplementary Material for Biometric Characteristics of Hand Gestures through Joint Decomposition of Cross-Subject and Cross-Session Biases

Aman Verma, Gaurav Jaswal, Seshan Srirangarajan, Sumantra Dutta Roy

The supplementary material has been structured into four sections. Section 1 “Bias decomposition” presents a tabular description of the proposed bias decomposition which has been discussed in the Section 2 “The *GBQA* Framework” of the main manuscript. Next, Section 2 describes the dataset used and corresponding experimentation protocol for the gesture pruning based *GBQA* score estimation. This is followed by Section 3, wherein we provide several evidences to empirically verify the theoretical trade-offs between  $G_{Bias-CU}$  and biometric parameters. For these experiments, we quantify biometric uniqueness and variability using  $d_{UNQ}$  and  $d_{VRB}$  [1]. In particular,  $d_{UNQ}$  computes the average inter-subject distance within a gesture cluster. While  $d_{VRB}$  measures the average intra-subject variation for a gesture. Finally, in the last section, we propose several relative measurement based methods for *GBQA* score estimation. These methods involve comparing the performance in two settings: (i) cross-user, and (ii) cross-user with some technique to intensify and mitigate personality bias.

## 1. Bias Decomposition

In table S1 we outline the biases which causes failure in cross-user and cross-session experiments. It can be noted from the same that cross-user scenario has entanglement between biometric and recognition level generalization biases. With our proposed *GBQA* method, we disentangle generalization parameters ( $G_3$  &  $G_4$ ) from the biometrics one.

## 2. Dataset description and experimentation protocol

In this section, we describe the hand gesture datasets used for validating *GBQA* score computation using gesture pruning method. We report results in two settings: (i) without gesture pruning (the complete dataset) and (ii) with gesture pruning (only the top-performing gestures are utilized for training). The obtained performance is then compared against equal error rates (EER) reported in [1]. We majorly report results with TDSNet [2] as the feature extractor. This choice was motivated by its relevance in both gesture biometrics and recognition.

1. **Soli [3]**: This is a mmWave radar hand gesture recognition dataset has 11 gesture classes, collected from 10 subjects with 25 gesture instances per subject. We specifically chose this dataset because these gesture sequences carry only motion-specific details. We considered gesture from first five subjects in training, while the next five in testing. We select 6 good performing gestures (on the hand gesture recognition task): ‘Pinch Index’, ‘Palm Tilt’, ‘Fast Swipe’, ‘Push’, ‘Finger Rub’ and ‘Circle’.
2. **TinyRadar [4]**: This is a large-scale mmWave radar hand gesture recognition dataset with 30,300 samples. It comprises similar 11 gestures as in Soli. However, the acquisition radar has a different configuration. Also, it has been collected from 26 subjects, which is higher than the former. In order to validate generalizability of the proposed, we have selected the same 6 gestures as in Soli. For training, we have utilized data from first 16 subjects, while testing is performed over the remaining 10 subjects.
3. **SCUT [5]**: This is a large-scale RGB-based hand gesture authentication dataset. It comprises of 6 gestures collected from 193 subjects. The training set consists of 143 subjects while the test set has 50 subjects. After pruning, we considered 3 gestures: ‘Rotate to fist’, ‘Catch and release’, and ‘Bend four finders’.

## 3. Trade-off between biometric parameters and $G_{Bias-CU}$

Our proposed decomposition (refer Sec. 2.3.1, main manuscript) suggests that, with reduction in  $G_{Bias-CU}$  there will be decrement in biometric uniqueness ( $G_1$ ) and variability ( $G_2$ ). To this end, we validate the claim via several experiments. In summary, these experiments involve mitigating  $G_{Bias-CU}$  via different methods. Then, the corresponding impact on biometric uniqueness and variability is analyzed. For mitigating  $G_{Bias-CU}$ , we utilize two strategies: (i) better gesture recognition models and (ii)

Table S1. **Tabular description of the proposed bias decomposition.** We summarize the biases involved in cross-user and cross-session gesture recognition experiments. This summary has been made available in pictorial format in Fig. 2 and Fig. 3 of the main manuscript. Hierarchy of the bias parameters has been tabulated in column titled 'Parent bias'. The table mentions if a bias term is instrumental for biometrics or generalization. Further, it also tabulates the form of bias involved for respective tasks (refer the columns: 'Biometric parameter' and 'Generalization parameter'). The inverse of a bias parameter ( $G^{-1}$ ) characterizes the opposite affect of that bias. 'Disentangled' bias parameter signifies the fact that biometric and generalization cannot be separated at the current stage i.e., it requires further decomposition. It can be noted from the table that the cross-user experiment has uniqueness and stability parameters. On the other hand, the cross-session experiment has only stability parameters. Furthermore, both of these experiments (cross-user and cross-session) have bias parameters, shared between biometrics and generalization.

| Bias parameter                                     | Name                                     | Description                                                                                                                                                                                                                          | Parent bias   | Biometric property | Generalization property | Biometric parameter | Generalization parameter |
|----------------------------------------------------|------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|--------------------|-------------------------|---------------------|--------------------------|
| <b>Cross-user bias (<math>CU_{Test}</math>)</b>    |                                          |                                                                                                                                                                                                                                      |               |                    |                         |                     |                          |
| $P_{Bias}$                                         | Personality bias                         | This originates because of variation in gesture patterns across different subjects in training and testing.                                                                                                                          | $CU_{Test}$   | ✓                  | ✓                       | $P_{Bias}$          | $P_{Bias}^{-1}$          |
| $G_{Bias-CU}$                                      | Cross-user gesture understanding bias    | This is generated if the model has not learnt discriminative enough representation from the training data of different subjects.                                                                                                     | $CU_{Test}$   | ✓                  | ✓                       | Disentangled        | $G_{Bias-CU}^{-1}$       |
| $G_1$                                              | Biometric uniqueness                     | The model is not able to capture generic gesture representations because subjects have highly unique gesture representation. Thus, this bias term encodes the amount of uniqueness between training subjects at intra-gesture level. | $G_{Bias-CU}$ | ✓                  | ✓                       | $G_1$               | $G_1^{-1}$               |
| $G_2$                                              | Biometric stability                      | This term describes if there is high variability in any gesture class at intra-subject level.                                                                                                                                        | $G_{Bias-CU}$ | ✓                  | ✓                       | $G_2^{-1}$          | $G_2^{-1}$               |
| $G_3$                                              | Inter-gesture intra-subject similarity   | This term describes the extent of similarity between different gesture at intra-subject level                                                                                                                                        | $G_{Bias-CU}$ | ✗                  | ✓                       | -                   | $G_3^{-1}$               |
| $G_4$                                              | Inter-gesture inter-subject similarity   | This term describes the extent of similarity between different gesture at inter-subject level                                                                                                                                        | $G_{Bias-CU}$ | ✗                  | ✓                       | -                   | $G_4^{-1}$               |
| <b>Cross-session bias (<math>CS_{Test}</math>)</b> |                                          |                                                                                                                                                                                                                                      |               |                    |                         |                     |                          |
| $S_{Bias}$                                         | Stability bias                           | This accounts for variability between gesture of a subject across different sessions                                                                                                                                                 | $CS_{Test}$   | ✓                  | ✓                       | $S_{Bias}^{-1}$     | $S_{Bias}^{-1}$          |
| $G_{Bias-CS}$                                      | Cross-session gesture understanding bias | This is generated if the model has not learnt discriminative enough representation from the training data of different sessions from a same subject                                                                                  | $CS_{Test}$   | ✓                  | ✓                       | $G_{Bias-CS}^{-1}$  | $G_{Bias-CS}^{-1}$       |
| $G_5$                                              | Biometric stability                      | This term describes if there is high variability in any gesture class at intra-subject and inter-session level.                                                                                                                      | $G_{Bias-CS}$ | ✓                  | ✓                       | $G_5^{-1}$          | $G_5^{-1}$               |
| $G_6$                                              | Inter-gesture intra-subject similarity   | This term describes the extent of similarity between different gesture at intra-subject level                                                                                                                                        | $G_{Bias-CS}$ | ✗                  | ✓                       | -                   | $G_6^{-1}$               |

auxiliary triplet loss objective to improve intra-gesture clustering. In order to quantify uniqueness and variability, we use  $d_{UNQ}$  (inter-subject distance) and  $d_{VRB}$  (intra-subject variation) [1]. These serve as a proxy to  $G_1$  and  $G_2$  respectively. We conduct these experiments on Soli dataset [3]. In the subsequent part, we provide experimental results and analysis:

**Experiment # 1: Better gesture recognition model:** An architecture with better feature modeling capacity can be instrumental in mitigating  $G_{Bias-CU}$ . In this experiment, we check if an improved architecture can achieve lower uniqueness and variability than the original architecture. To this end, we either introduce multi-scale modeling [7] into the architecture or utilize a more domain specific architecture such as TDSNet [2]. In particular, we perform three comparisons: (i) Res3D (3D CNN backbone with residual connections) vs MSRes3D (multi-scale variant of Res3D). (ii) Res3D-ViViT (Res3D with ViViT [8]) vs MSRes3D-ViViT

(Res3D-ViViT with multi-scale 3D CNN backbone). (iii) MSRes3D-ViViT vs TDSNet. A bar graph based presentation of the comparison has been shown in Figure S2. It can be observed from the plot that multi-scale variant attains better performance and lower uniqueness and variability (refer Figure S2, (a) and (b)). Similarly, TDSNet outperforms Res3D-ViViT in terms of gesture recognition performance, and achieves lower biometric parameters (refer Figure S2, (c)). This gives an empirical validation of the proposed bias decomposition. Furthermore, we find some gestures 'Circle' and 'Fast swipe' tend to show limited decrement in uniqueness with the use better models. This aligns with the fact that they are good in biometrics and achieve lower  $EER$ . We observe that the decrease in variability is lesser than the uniqueness. This can be reasoned with the fact that a gesture recognition model clusters samples from the same class irrespective of the identity. Hence, justifying the entanglement of generalization with biometric

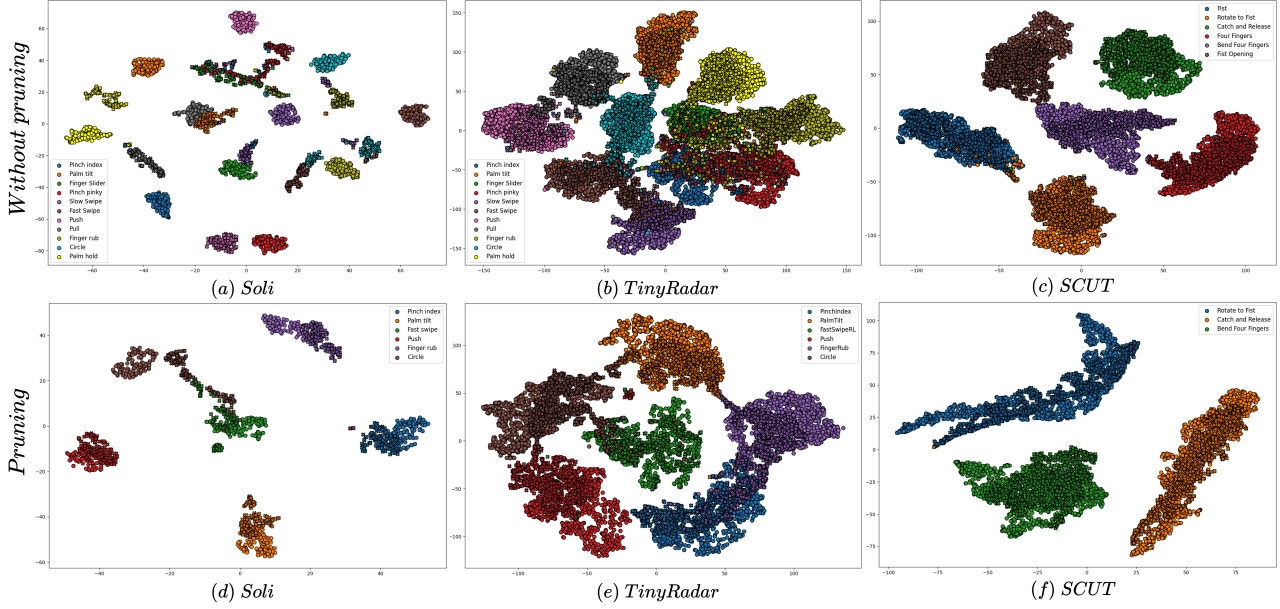


Fig. S1. **tSNE plots with and without gesture pruning.** • and X respectively represent training and testing data. The plot has a two-fold objective: (i) studying impact of gesture pruning, and (ii) analyzing the amount  $P_{Bias}$ . The plot reveals that gesture pruning leads to reduced inter-gesture clustering. This results in a reduction in  $G_{Bias-CU}$ . We also observe that the training and testing data are more disparate in the Soli and TinyRadar datasets, as compared to the SCUT-DHGA dataset. This suggests that the first two have higher  $P_{Bias}$  values as compared to the third. This can be explained by the significantly higher number of subjects in the SCUT-DHGA dataset. As the number of subjects increases, the biometric uniqueness reduces. This results in the test data being less spread apart, compared to the training data.

characteristics. Also, the models that show higher classification accuracy leads to higher reduction in uniqueness (for e.g., TDSNet). This highlights the with better gesture understanding (conversely decrease in  $G_{Bias-CU}$ ), there is decrement in uniqueness and variability.

#### Experiment # 2: Auxiliary gesture clustering objective:

In this experiment, we introduce triplet loss [6] to enforce better intra-gesture clustering and inter-gesture separability. This objective leads towards improved generalization and hence results in mitigating  $G_{Bias-CU}$ . We experiment with different values of the triplet loss margin, and expect decrease in uniqueness and variability with higher margin values. Here, we used TDSNet [2] as the feature extractor. The corresponding trade-off has been illustrated in Figure S3. It can be observed from the same, that with increasing margin ( $m$ ), there is a general downward trend in uniqueness (refer Figure S3 (a)) and variability (refer Figure S3 (b)). This is noted for most of the gestures. Hence, this verifies that there exists an inverse relationship between  $G_{Bias-CU}$  and biometric parameters ( $G1$  and  $G2$ ). It must be noted that decrement in these parameters is non-uniform. This highlights that different gestures have different biometric characteristics. The amount of decrease in uniqueness is more than that variability. This corresponds to the fact that when variability might originate because of outliers, while uniqueness is

easily mitigated with intra-gesture clustering. We also observe some inconsistencies. For example, ‘Pinch index’ and ‘Finger rub’ shows limited depicts some increment in variability with increasing margin. This has several reasons: (i) the margin based loss is dependent upon samples in a batch. Stochasticity if batch formulation leads to suboptimal convergence and thus we might find irregularities in the decrement. (ii) Refer Fig. 5 (a) (main manuscript), it shows that these gestures possess much higher variability in comparison to uniqueness. As a result, the later is more challenging to reduce.

## 4. Relative measurement techniques

We propose following methods for relative measurement based quantification of biometric characteristics.

**Bias Intensification:** As the name suggests, the personality bias is intensified and the entire methodology revolves around on how to increase the bias. To this end, we suggest two strategies: (i) model poisoning and (ii) data poisoning.

(i) *Model Poisoning:* A gesture can be decomposed into gesture content (the gesture class) and gesture style (biometric details of the subject). We aim at making the model forget the content. This can in turn, be captured using neural style transfer [9]. This would culminate in intensifying

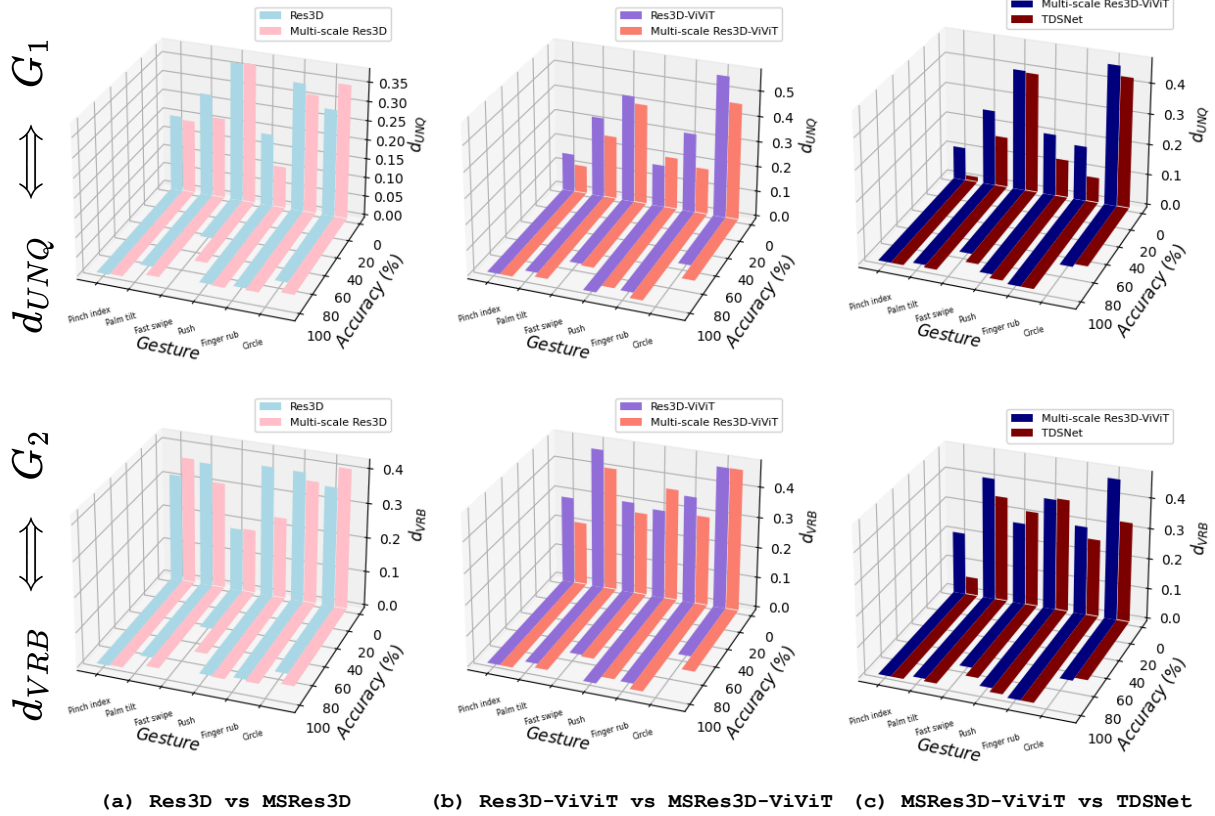


Fig. S2. **The impact of reducing  $G_{Bias-CU}$  on biometric parameters, via better architectures.** The figure studies the reduction in uniqueness and variability with the usage of better models. To this end, we use multi-scale modeling and a domain-specific architecture: TDSNet. We quantify uniqueness and variability with  $d_{UNQ}$  and  $d_{VRB}$  as in our earlier work [1]. These serve as a proxy for  $G_1$  and  $G_2$ , respectively. The figure presents comparison between three pairs of architectures. The first row of the figure presents comparison in terms of  $d_{UNQ}$ , and the second, for  $d_{VRB}$ . For each plot, we also show the classification accuracy of the architectures considered in this comparison. It is evident from the figure that there is a decrement in the uniqueness and variability, with better architectures. The decrease is even higher, when an architecture achieves better results. Furthermore, gestures with better biometric characteristics show a limited reduction in uniqueness. These inferences substantiate the claim that with a decrease in  $G_{Bias-CU}$ , both uniqueness and variability reduce.

personality-related biases.

(ii) *Data Poisoning*: One can apply adversarial perturbations on the input to make the model forget the gesture content and emphasize personality sensitive features. The gesture class which becomes the most susceptible can be referred to as bearing more personality-relevant features.

**Bias Mitigation**: Here, the aim is to curb personality-sensitive features. Hence, for the classes wherein these are most challenging to suppress, can be accredited to bear biometric details. If the user identity information is available, then the model can be poisoned to forget identities by inculcating domain invariance; GRL [10] is an alternative. Using this notion, gestures which tend to be vulnerable are segregated from the ones that have intrinsic biometric character-

istics.

## References

- [1] A. Verma, et al., Quantifying Biometric Characteristics of Hand Gestures Through Feature Space Probing and Identity-Level Cross-Gesture Disentanglement, in: Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FG), 2024, pp. 1–9. 1, 2, 4
- [2] W. Song, et al., TDS-Net: Towards Fast Dynamic Random Hand Gesture Authentication via Temporal Difference Symbiotic Neural Network, in: Proc. IEEE Intl. Joint Conf. on Biometrics (IJCB), 2021, pp. 1–8. 1, 2, 3
- [3] S. Wang, et al., Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-

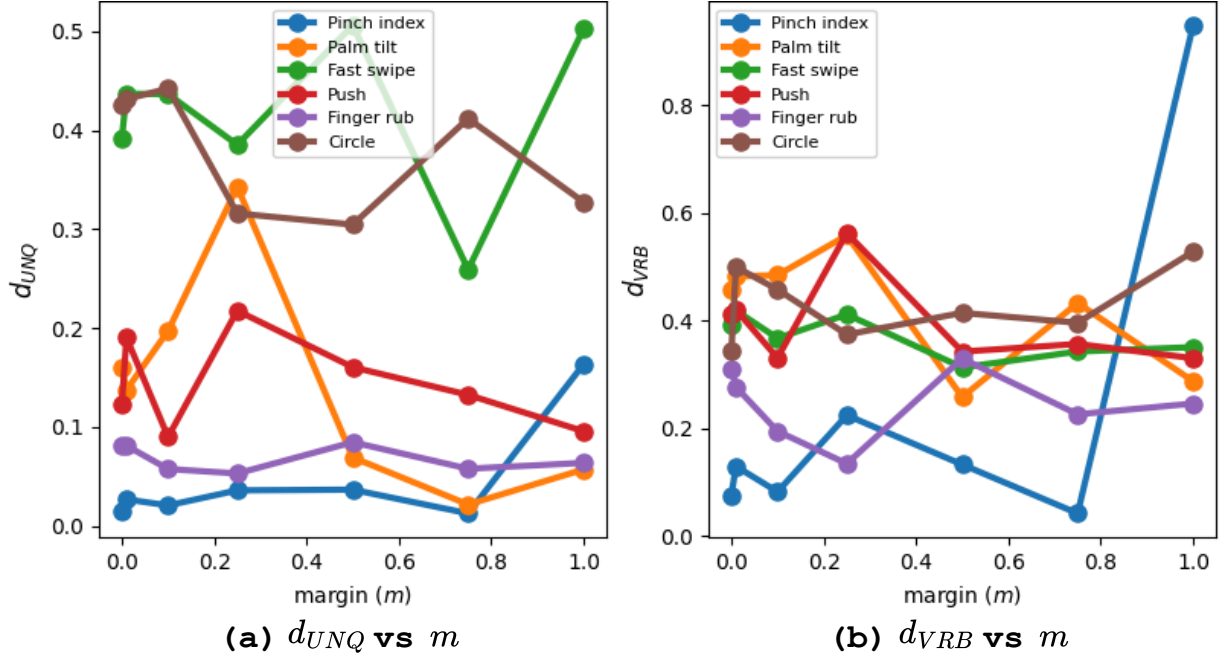


Fig. S3. **The impact of enforcing intra-gesture clustering and inter-gesture separability on biometric bias parameters, with a triplet loss function [6].** We introduce a margin-based triplet loss in cross-user gesture recognition experiments, and plot  $d_{UNQ}$  (biometric uniqueness) and  $d_{VRB}$  (variability) for different margin values. The figure highlights that with an increase in the margin, there is an overall decrease in  $d_{UNQ}$  and  $d_{VRB}$ . This suggests that by improving gesture understanding (conversely reducing  $G_{Bias-CU}$ ), we find a decrease in biometric uniqueness and variability.

- Frequency Spectrum, in: Proc. Symposium on User Interface Software and Technology, 2016, pp. 851–860. 1, 2
- [4] M. Scherer, et al., TinyRadarNN: Combining Spatial and Temporal Convolutional Neural Networks for Embedded Gesture Recognition with Short Range Radars, IEEE Internet Things J. 8 (13) (2021) 10336–10346. 1
- [5] C. Liu, et al., Dynamic-Hand-Gesture Authentication Dataset and Benchmark, IEEE Transactions on Information Forensics and Security 16 (2020) 1550–1562. 1
- [6] F. Schroff, et al., Facenet: A Unified Embedding for Face Recognition and Clustering, in: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2015, pp. 815–823. 3, 5
- [7] S.-H. Gao, et al., Res2Net: A New Multi-Scale Backbone Architecture, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2) (2019) 652–662. 2
- [8] A. Arnab, et al., ViViT: A Video Vision Transformer, in: Proc. Intl. Conf. on Computer Vision (ICCV), 2021, pp. 6836–6846. 2
- [9] Delgado-Santos, et al., GaitPrivacyON: Privacy-Preserving Mobile Gait Biometrics using Unsupervised Learning, Pattern Recognition Letters 161 (2022) 30–37. 3
- [10] I. Joshi, et al., Sensor-Invariant Fingerprint ROI Segmentation using Recurrent Adversarial Learning, in: Proc. Intl. Joint Conf. on Neural Networks (IJCNN), 2021, pp. 1–8. 4