A Closer Look at Geometric Temporal Dynamics for Face Anti-Spoofing - Supplementary Material -

Chih-Jung Chang[12]*, Yaw-Chern Lee[1], Shih-Hsuan Yao[1], Min-Hung Chen[1], Chien-Yi Wang[1]
Shang-Hong Lai[13], Trista Pei-Chun Chen[1]
[1]Microsoft AI R&D Center, Taiwan [2]Stanford University [3]National Tsing Hua University, Taiwan

1. More Implementation Details

GCN Training Augmentation: As described in Sec. 4.2 in the main paper, random rotation and horizontal flip are applied during geometric feature learning. More specifically, given a landmark $v'_{ti} = (x'_{ti}, y'_{ti})$ in an aligned and sub-sampled input sequence, it is first rescaled by $2v'_{ti} - 1$ so that it fits within the range $[-1, 1]$. We later denote the rescaled landmark by v_{ti} for simplicity. Rotating v_{ti} by an angle θ is formulated as:

$$rotate(v_{ti}, \theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_{ti} \\ y_{ti} \end{pmatrix}, \quad (1)$$

and flipping v_{ti} horizontally is done by:

$$flip(v_{ti}) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{x}_{ti} \\ \tilde{y}_{ti} \end{pmatrix}, \quad (2)$$

where $\tilde{v}_{ti} = (\tilde{x}_{ti}, \tilde{y}_{ti})$ is the position of the corresponding landmark of v_{ti} on the other side of the face. During training, we rotate each landmark sequence by a rotation angle θ randomly chosen from $[-30, 30]$, and horizontally flip the landmark sequence with a probability of 0.5.

Facial Movements in CASIA-FASD Dataset: While the spoof types in CASIA-FASD dataset (live, print attack, and video-replay attack) are the same as in the other three datasets used for the cross-dataset evaluation: OULU-NPU, Replay-Attack, MSU-MFSD, the class of print attack in CASIA-FASD can be further divided into two sub-classes: print-warp attack and print-cut attack (Fig. 1). In particular, the facial movements of print-cut attacks in CASIA-FASD, where the motion around the eyes is driven by a live person blinking behind, are very different from the facial dynamics of print attacks in the other three datasets. Therefore, we exclude the print-cut attack in CASIA-FASD from training in all the experiments. This attack is also excluded when evaluated for abnormal movement detection. However, this attack (print-cut attack in CASIA-FASD) is included in all the testing protocols when GAIN is compared to the SOTA methods under the settings in Sec. 4.3 in the main paper.

Figure 1. Print Attack in CASIA-FASD. The print attack in CASIA-FASD can be furthered divided into the sub-classes of print-warp attack (left) and print-cut attack (right).

Standard Temporal-Based Methods (3D-CNN): In Sec. 4.4 in the main paper, we compare GAIN with methods that adopt 3D-CNN [3] to discuss the benefit of extracting geometric information from facial landmarks. In this section, we provide further implementation details of the 3D-CNN-based methods. We utilize MTCNN [9] to align and crop faces to the size 256×256 in each video. The optical flow is then estimated based on the cropped and aligned face images every two frames using the algorithm from [5]. The output is later resized to 64×64 for training efficiency. During training, we randomly sub-sample 64 frames of resized images/flows as the input of the 3D-CNN. At the inference stage, we choose the 64 frames by uniform sub-sampling. The 3D-CNN is trained for 65 epochs with a batch size of 16. The optimizer is SGD with a momentum of 0.9 and a weight decay of 0.0001. The initial learning rate is set to 0.1, and it is decayed with a factor of 0.1 after the 50-th epoch. Same augmentation techniques are applied to the RGB+3D-CNN and Flow+3D-CNN methods: random horizontal flip and rotation by an angle within 30 degrees.

*Work done during the internship at Microsoft AI R&D Center, Taiwan.

Method	O&M&I to C		M&I to C		C&M&I to O		M&I to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MS-LBP [6]	54.28	44.98	51.16	52.09	50.29	49.31	43.63	58.07
IDA [8]	55.17	39.05	45.16	58.80	54.20	44.59	54.52	42.17
CT [1]	30.58	76.89	55.17	46.89	63.59	32.71	53.31	45.16
LBP-TOP [2]	42.60	61.05	45.27	54.88	53.15	44.09	47.26	50.21
MADDG [7]	24.50	84.51	41.02	64.33	27.89	80.02	39.35	65.10
SSDG-M [4]	23.11	85.45	31.89	71.29	25.17	81.83	36.01	66.88
SSDG-R* [4]	9.89	95.28	18.11	88.00	14.03	93.07	18.89	89.87
SSDG-R* + **GAIN**	**8.52**	**96.02**	**18.00**	**89.39**	**12.50**	**95.12**	**15.00**	**92.36**

Table 1. The results of the cross-dataset evaluation with limited (2-to-1) training datasets, the results of 3-to-1 cross-dataset evaluation is listed to the left for comparison. We reproduce SSDG-R as our baseline methods (noted as SSDG-R*).

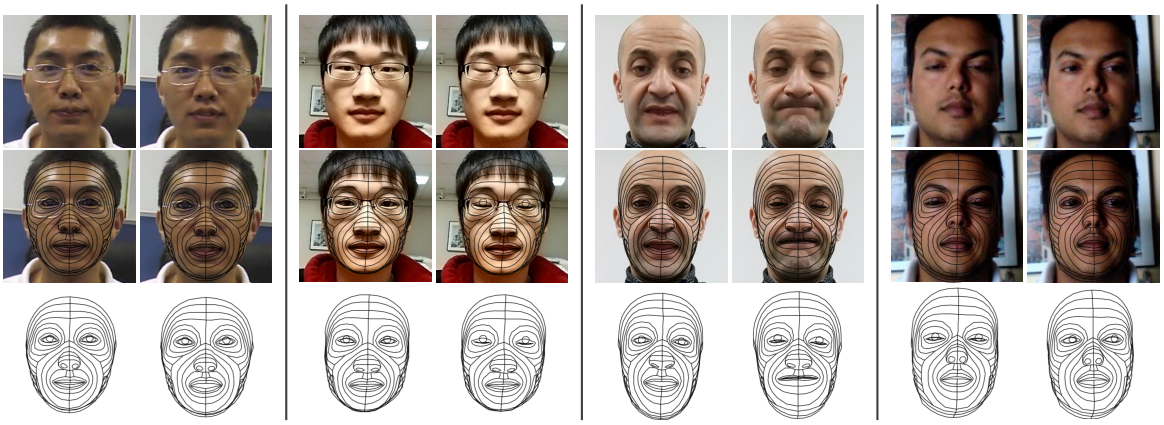

Figure 2. Visualization of dense landmark predictions. The four live subjects (from left to right) are from CASIA-FASD, MSU-MFSD, OULU-NPU, and Replay-Attack.

2. More Results in Cross-Dataset Evaluation

In addition to the results shown in Sec. 4.3 in the main paper, Table 1 provides further cross-dataset evaluations under the setting of limited training datasets (M&I to C, and M&I to O). GAIN once again improves significantly over the baseline photometrics-based method. The result indicates that our proposed GAIN is able to learn discriminative geometric facial dynamics even with limited sources of training data.

3. More Visualization

Dense Landmark Prediction: More dense landmark predictions of live faces are provided in Fig. 2. The live faces are from the four datasets used in the cross-dataset evaluation (CASIA-FASD, MSU-MFSD, OULU-NPU, and Replay-Attack). As shown by the visualization, the fine-grained facial movements are detailedly captured by dense facial landmarks.

t-SNE Visualization: In Fig. 3, we provide the t-SNE visualization of the extracted geometric features by GAIN under the cross-dataset settings of O&M&I to C, O&C&M to I, and I&C&M to O. Close matches between the training and testing datasets can be observed in all three settings.

References

[1] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016. 2

[2] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Lbp- top based countermeasure against face spoofing attacks. In *Computer Vision-ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part I 11*, pages 121–132. Springer, 2013. 2

[3] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6546–6555, 2018. 1

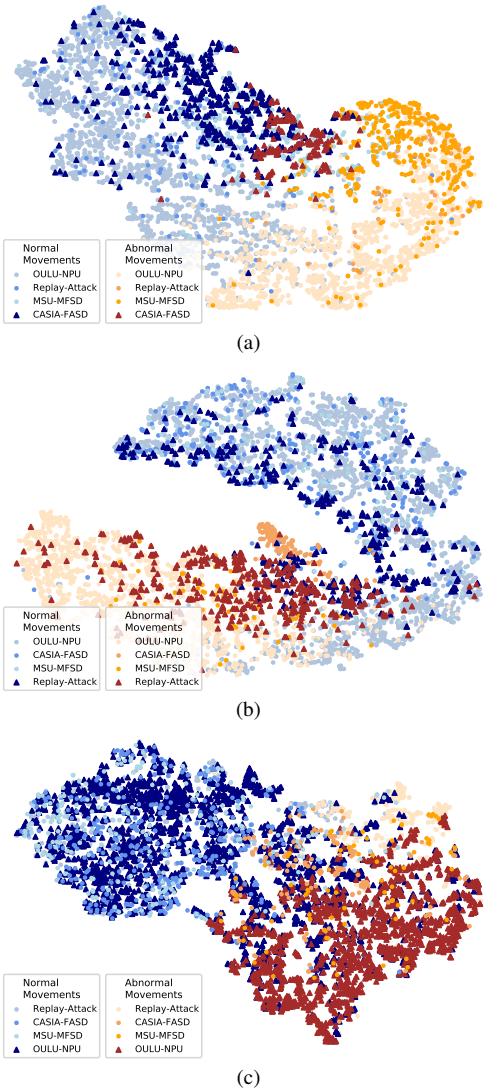

Figure 3. t-SNE visualization of the extracted geometric features by GAIN. The features are extracted under the cross-dataset settings of: (a) O&M&I to C, (b) O&C&M to I, and (c) I&C&M to O.

[4] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020. 2

[5] Ce Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. 1

[6] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *2011 international joint conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011. 2

[7] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10023–10031, 2019. 2

[8] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015. 2

[9] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016. 1