DEEPFAKE DETECTION VIA SEPARABLE SELF-CONSISTENCY LEARNING SUPPLEMENTARY MATERIAL

Lin Lu⋆*, Yunhong Wang*⋆*, Wenqi Zhuo*⋆*, Liang Zhang*†*, Guangshuai Gao*‡*, Yuanfang Guo*⋆

⋆ School of Computer Science and Engineering, Beihang University, China
† CNCERT/CC, Beijing, China
‡ School of Electronic and Information, Zhongyuan University of Technology, China

1. VIDEO-LEVEL COMPREHENSIVE RESULTS

Due to space limitations, we only present the performance of our model on the metric of AUC. Here, we show the results in terms of AUC and AP. Also we show the results within-dataset. Our results only 98.89% on FF++ [1] (within-dataset), indicating a clear impact from NT [2], which significantly lowers the AUC. This is because the artifacts caused by NT are subtle and difficult to capture. However, such decrease is a normal phenomenon since the training process lack specific samples of forgery method. In our future approaches, we will consider how to enhance the detection of subtle artifacts.

Table 1. Comprehensive evaluation of our model in terms of video-level AUC, AP on four datasets.

Method	Test Set	Metrics	
		AUC	AP
SSCL-DFD + SBIs (Ours)	DF[3]	99.88	99.88
	F2F[4]	99.21	99.34
	FS[5]	99.47	99.43
	NT [2]	96.98	97.22
	FF++ [1]	98.89	99.72
	CD2 [6]	96.12	97.92
	DFDC [7]	75.69	76.52
	FFIW [8]	83.27	83.14

2. FRAME-LEVEL RESULTS ON CD2

In the main text, we only present the performance of our model at the video-level. Here, we will present the results at the frame-level. As shown in Table 2. All the methods are trained on FF++ and evaluated on CD2 [6]. Our model outperforms the state-of-the-art method (CGS[9]) and Swin [10] by over 4.97% and 3.36%, respectively. The experimental results demonstrate that our method still maintains a good effectiveness at the frame-level.

Table 2. Cross-dataset evaluation of our model in terms of frame-level AUC on CD2 dataset. The results of comparable methods are directly cited from the original papers for fair comparison.

Method	CD2 (Frame-Level AUC)
Meso4 [11]	54.8
MesoInception4 [11]	53.6
Xception [1]	65.3
UIA-ViT [12]	82.41
CGS [9]	84.97
Two-branch [13]	73.4
Multi-task [14]	54.3
Multi-Attention [15]	67.4
PCL + I2G [16]	81.8
SLADD [17]	79.7
Swin Transformer [10] + SBIs	<u>86.58</u>
SSCL-DFD + SBIs (Ours)	**89.94**

3. REFERENCES

[1] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, “Faceforensics++: Learning to detect manipulated facial images,” in *IEEE CVPR*, 2019, pp. 1–11.

[2] Justus Thies, Michael Zollhöfer, and Matthias Nießner, “Deferred neural rendering: Image synthesis using neural textures,” *ACM TOG*, vol. 38, no. 4, pp. 1–12, 2019.

[3] FaceSwapDevs, “Deepfakes,” https://github.com/deepfakes/faceswap, 2019.

[4] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner, “Face2face: Real-time face capture and reenactment of rgb videos,” in *IEEE CVPR*, 2016, pp. 2387–2395.

[5] Marek Kowalski, “Faceswap,” https://github.com/MarekKowalski/FaceSwap, 2018.

[6] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *IEEE CVPR*, 2020, pp. 3207–3216.

[7] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer, "The deepfake detection challenge (dfdc) dataset," *arXiv preprint arXiv:2006.07397*, 2020.

[8] Tianfei Zhou, Wenguan Wang, Zhiyuan Liang, and Jianbing Shen, "Face forensics in the wild," in *IEEE CVPR*, 2021, pp. 5778–5788.

[9] Ying Guo, Cheng Zhen, and Pengfei Yan, "Controllable guide-space for generalizable face forgery detection," in *IEEE ICCV*, 2023, pp. 20818–20827.

[10] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *IEEE ICCV*, 2021, pp. 10012–10022.

[11] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen, "Mesonet: a compact facial video forgery detection network," in *IEEE WIFS*, 2018, pp. 1–7.

[12] Wanyi Zhuang, Qi Chu, Zhentao Tan, Qiankun Liu, Haojie Yuan, Changtao Miao, Zixiang Luo, and Nenghai Yu, "Uia-vit: Unsupervised inconsistency-aware method based on vision transformer for face forgery detection," in *ECCV*, 2022, pp. 391–407.

[13] Iacopo Masi, Aditya Killekar, Royston Marian Mascarenhas, Shenoy Pratik Gurudatt, and Wael AbdAlmageed, "Two-branch recurrent network for isolating deepfakes in videos," in *ECCV*, 2020, pp. 667–684.

[14] Huy H Nguyen, Fuming Fang, Junichi Yamagishi, and Isao Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," in *IEEE BTAS*, 2019, pp. 1–8.

[15] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, and Nenghai Yu, "Multi-attentional deepfake detection," in *IEEE CVPR*, 2021, pp. 2185–2194.

[16] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, and Wei Xia, "Learning self-consistency for deepfake detection," in *IEEE ICCV*, 2021, pp. 15023–15033.

[17] Liang Chen, Yong Zhang, Yibing Song, Lingqiao Liu, and Jue Wang, "Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection," in *IEEE CVPR*, 2022, pp. 18710–18719.