Supplementary Materials for "Dynamic Graph Learning with Content-guided Spatial-Frequency Relation Reasoning for Deepfake Detection"

Yuan Wang[1,2,4] Kun Yu[2] Chen Chen[1*] Xiyuan Hu[3] Silong Peng[1,4,5]
[1]Institute of Automation, Chinese Academy of Sciences [2]Alibaba Group
[3]School of Computer Science and Engineering, Nanjing University of Science and Technology
[4]University of Chinese Academy of Sciences [5]Beijing Visystem Co.Ltd
{wangyuan2020, chen.chen}@ia.ac.cn yukun.yk@alibaba-inc.com xiyuan.hu@foxmail.com

1. Experiments

1.1. Experimental Results

Cross-testing Evaluation on FaceForensics++. As shown in Table 1, to further demonstrate the generalization ability of our proposed SFDG method among different manipulated types, we conduct additional experiments on FF++ (LQ) [10] benchmark dataset that contains counterfeit images from four manipulated techniques, i.e., Deepfakes (DF) [15], Face2Face (F2F) [14], FaceSwap (FS) [7], and NeuralTextures (NT) [13]. We train our model on one of them and test it on all four approaches. We refer to the experiment results in DCL [11]. From Table 1, our method consistently surpasses all competitors by a clear margin in most cases. Specifically, when training on DF and testing on F2F, FS and NT, our approach achieves 12.08%, 23.48% and 15.05% gain in terms of AUC respectively. These overwhelming results give explanations that our method sufficiently excavates adaptive frequency features and discover the relation of essential forged clues in spatial and frequency domain via dynamic graph learning, thus improving the cross-manipulation performance.

Robustness Analysis on FF++/WildDeepfake Dataset. In view of the ubiquity of image processing on social media, we investigate the robustness of our proposed model by training on original WildDeepfake [18] dataset and testing on WildDeepfake samples that are subverted by common unseen perturbations suggested by [1,4–6]. As shown in Table 2, a serious of experiments are conducted to demonstrate the robustness of our proposed SFGD against noise and blur perturbations. In detail, we train our SFDG model on the clean data and then insert GaussianNoise, SaltPepperNoise, and GaussianBlur to the test samples. We evaluate the robustness of the forgery detection model utilizing the decay of Acc and AUC respectively. The experiment results verify that our method has outstanding performances in most disturbance cases. Furthermore, to reveal the robustness of our SFGD method, we perform a series of experiments on WildDeepfake datasets under more unseen corruptions, i.e., Compress, Contrast, Saturate and Pixelate as shown in Table 3. We refer the experimental results from RECCE [1]. We can observe that our model still outperforms some current state-of-the-art works significantly except Pixelate disturbance. We principally attribute the performance gain to the proposed multiscale attention maps with rich context information, thus insusceptible to unseen perturbations.

Training	Model	Testing (AUC)			
		DF	F2F	FS	NT
DF	Ef-b4 [12]	99.97	76.32	46.24	72.72
	MADD [17]	99.92	75.23	40.61	71.08
	GFF [8]	99.87	76.89	47.21	72.88
	DCL [11]	**99.98**	77.13	61.01	75.01
	SFDG (Ours)	99.73	**86.45**	**75.34**	**86.64**
F2F	Ef-b4 [12]	84.52	99.20	58.14	63.71
	MADD [17]	86.15	99.13	60.14	64.59
	GFF [8]	89.23	99.10	61.30	64.77
	DCL [11]	91.91	99.21	59.58	66.67
	SFDG (Ours)	**97.38**	**99.36**	**73.54**	**72.61**
FS	Ef-b4 [12]	69.25	67.69	99.89	48.61
	MADD [17]	64.13	66.39	99.67	50.10
	GFF [8]	70.21	68.72	99.85	49.91
	DCL [11]	74.80	69.75	**99.90**	52.60
	SFDG (Ours)	**81.71**	**77.30**	99.53	**60.89**
NT	Ef-b4 [12]	85.99	48.86	73.05	98.25
	MADD [17]	87.23	48.22	75.33	98.66
	GFF [8]	88.49	49.81	74.31	98.77
	DCL [11]	91.23	52.13	79.31	98.97
	SFDG (Ours)	**91.73**	**70.85**	**83.58**	**99.74**

Table 1. Cross database evaluation in terms of AUC (%) on different manipulated types. The gray background indicates the intra-domain performance.

[*] Chen Chen is the corresponding author. This work is supported by the National Key R&D Program of China under Grant 2021YFF0602101, Alibaba Group through Alibaba Research Intern Program.

Method	+GaussianNoise		+SaltPepperNoise		+GaussianBlur	
	ΔAUC(FF)	ΔAUC(Wild)	ΔAUC(FF)	ΔAUC(Wild)	ΔAUC(FF)	ΔAUC(Wild)
Xception [2]	-0.0397	-0.0082	-0.3330	-0.1373	-0.1994	-0.0664
Add-Net [18]	-0.2862	-0.3327	-0.3445	-0.3589	-0.3445	-0.1895
F3Net [9]	-0.0838	-0.0248	-0.3891	-0.3407	-0.2077	-0.1567
MADD [17]	-0.0058	-0.0139	-0.2494	-0.1813	-0.2475	-0.1829
PEL [4]	-0.0031	**-0.0050**	-0.2079	-0.0483	-0.1274	**-0.1216**
SFDG (Ours)	**-0.0018**	-0.0086	**-0.1144**	**-0.0223**	**-0.0867**	-0.4831

Table 2. Robustness evaluation in terms of the decay of AUC under three types of perturbations. Our SFDG performs admirably under several common perturbations.

Methods	Compress	Contrast	Saturate	Pixelate	Average
Xception [2](CVPR′2017)	86.01	81.90	84.96	66.24	79.78
Ef-b4 [12](ICML′2019)	87.63	84.25	86.71	72.93	82.88
RFM [16](CVPR′2021)	83.74	79.77	82.59	71.25	79.35
Add-Net [18](MM′2020)	83.34	89.85	85.13	64.33	80.66
F3-Net [9](ECCV′2020)	86.71	86.53	87.67	73.23	83.54
MADD [17](CVPR′2021)	89.64	89.30	90.37	79.44	87.19
RECCE [1](CVPR′2022)	89.65	91.19	91.74	**83.88**	89.15
SFDG (Ours)	**91.43**	**92.02**	**92.11**	82.71	**89.56**

Table 3. Robustness evaluation in terms of AUC (%) on WildDeepfake dataset. “Average” indicates the mean score across all perturbations. Our SFGD performs more robust than all listed previous methods except pixelate.

Evaluating on DFDC Dataset. DFDC [3] is the most challenging dataset for face forgery detection tasks because of prominent manipulated quality of counterfeit videos in this dataset. Since seldom existing literature report intra-testing performance on it, we train the proposed SFDG model on the whole training set of DFDC and compare the corresponding Acc, AUC and LogLoss evaluation criteria with the re-implement results introduced in RECCE [1]. As shown in Table 4, our method outperforms other state-of-the-art competitors by 7.13% and 3.41% in terms of Acc and AUC, while the LogLoss decreases by 12.7%. The results in Table 4 throw light on the admirably performance of our method under extreme scene variation.

Methods	**Acc**	**AUC**	**LogLoss**
Xception [2]	79.35	89.50	0.492
Ef-b4 [12]	76.45	89.98	0.524
RFM [16]	80.83	89.75	0.581
Add-Net [18]	78.71	89.85	0.507
F3-Net [9]	76.17	88.39	0.520
MADD [17]	76.81	90.32	0.529
RECCE [1]	81.20	91.33	0.434
Ours (SFDG)	**86.99**	**94.44**	**0.379**

Table 4. Experiment results of intra-testing on the DFDC [3] benchmark dataset. Here, smaller logloss represents a better performance. The bold indicates the best in each column.

1.2. Visualizations

Attention Maps Visualization. To further understand the attention maps learning mechanism of our proposed SFGD framework, in an intuitive fashion, we visualize the overall attention maps and multiscale feature maps from different layers (from low level to high level) of the MDAML module. As illustrated in Fig 3, it implicates an apparent trend that the spatial attention maps are well-separated and exhibits the feature response in independent regions with distinct semantic representations, *e.g.*, eyes, mouth and contour. To explain, this intriguing result benefits from the Regional Independence Loss (RIL) which pulls the identical attention maps close to feature center while repelling different attention map centers scattered.

Further, we investigate the effectiveness of our tailored MDAML module and show the visualization results on FF++ and WildDeepfake datasets as illustrated in Fig 1 and Fig 2, respectively. From the figure, we observe that different scales of feature maps concentrate on different activated intensities. To be more specific, the large scale feature maps with high-resolution representations embrace richer and global manipulated traces, while the small scales with low-resolution representations provide focus to more local-

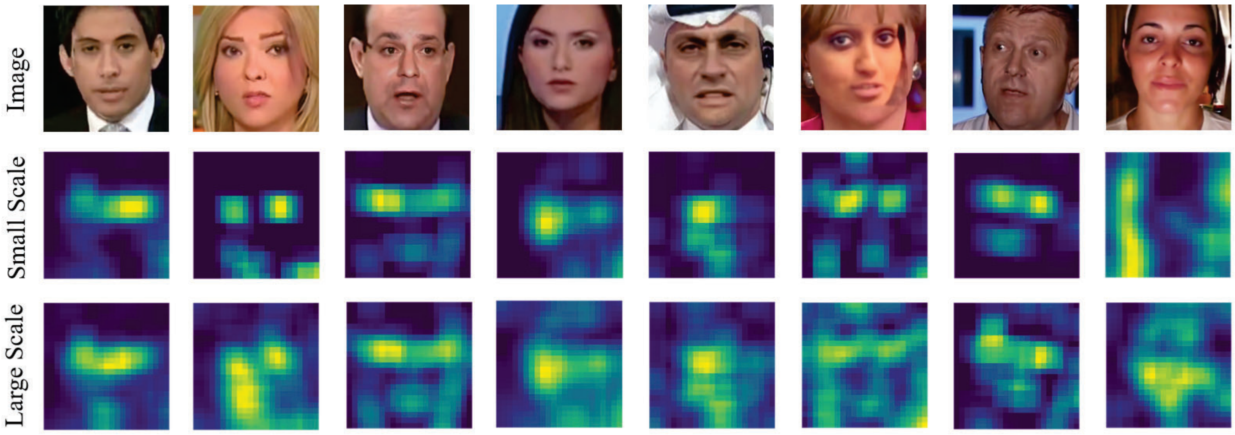

Figure 1. The visualization results of multiscale feature maps in MDAML module on FaceForensics++ [10] dataset. The first row represents the input images with specifical manipulated patterns. The second and third rows show the feature maps at different scales. We merely show one channel of feature maps.

ized and salient features around facial landmarks. The primary object of the MDAML module is to aggregate these multiscale feature maps through hierarchical pyramid network, which can get rid of the disturbance of noise and assist our model to draw attention to the essential discrepancy between real and counterfeit images.

Feature Distribution Visualization. In this section, we further verify the discriminative ability of our proposed SFDG framework. We therefore visualize the learned feature distribution of the baseline model [12], MADD [17] and our approach on FF++ (LQ and HQ) and WildDeepfake dataset utilizing the t-SNE technique. As shown in Fig 4, we randomly sample 5000 images from the FF++ (LQ) and FF++ (HQ) dataset, and 3000 from WildDeepfake. The visualized features of our method are extracted from the layer right before the first fully-connected layer. Observing from the visualization results that compared with the baseline and MADD methods, our approach embeds the same class samples into a relatively compact feature space. This phenomenon tests the validity of the effectiveness of our approach that adequately capture the essential discrepancy between genuine and manipulated faces in spatial-frequency domains through dynamic graph learning, thus improving the generalization ability of our method.

References

[1] Junyi Cao, Chao Ma, Taiping Yao, Shen Chen, Shouhong Ding, and Xiaokang Yang. End-to-end reconstruction-classification learning for face forgery detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4113–4122, 2022. 1, 2

[2] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1251–1258, 2017. 2

[3] Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer. The deepfake detection challenge (dfdc) preview dataset. *arXiv preprint arXiv:1910.08854*, 2019. 2

[4] Qiqi Gu, Shen Chen, Taiping Yao, Yang Chen, Shouhong Ding, and Ran Yi. Exploiting fine-grained face forgery clues via progressive enhancement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 735–743, 2022. 1, 2

[5] Alexandros Haliassos, Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic. Lips don't lie: A generalisable and robust approach to face forgery detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5039–5049, 2021. 1

[6] Liming Jiang, Ren Li, Wayne Wu, Chen Qian, and Chen Change Loy. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2889–2898, 2020. 1

[7] M Kowalski. Faceswap. https://github.com/marekkowalski/faceswap. Accessed: 2020-08-01, 2018. 1

[8] Yuchen Luo, Yong Zhang, Junchi Yan, and Wei Liu. Generalizing face forgery detection with high-frequency features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 16317–16326, 2021. 1

[9] Yuyang Qian, Guojun Yin, Lu Sheng, Zixuan Chen, and Jing Shao. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *Proceedings of the IEEE Conference on European Conference on Computer Vision*, pages 86–103. Springer, 2020. 2

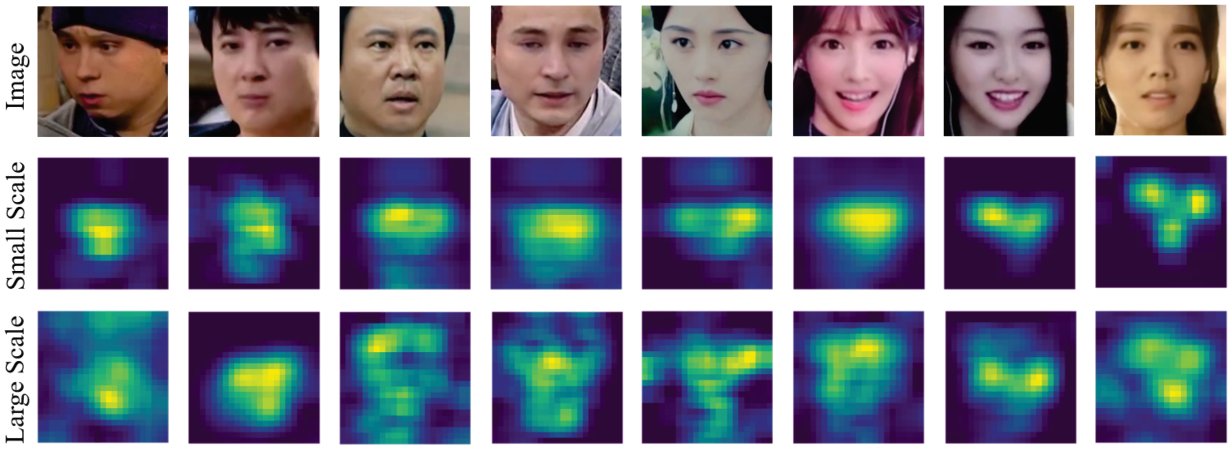

Figure 2. The visualization results of multiscale feature maps in MDAML module on WildDeepfake [18] dataset. The first row represents the input images with specifical manipulated patterns. The second and third rows show the feature maps at different scales. We merely show one channel of feature maps.

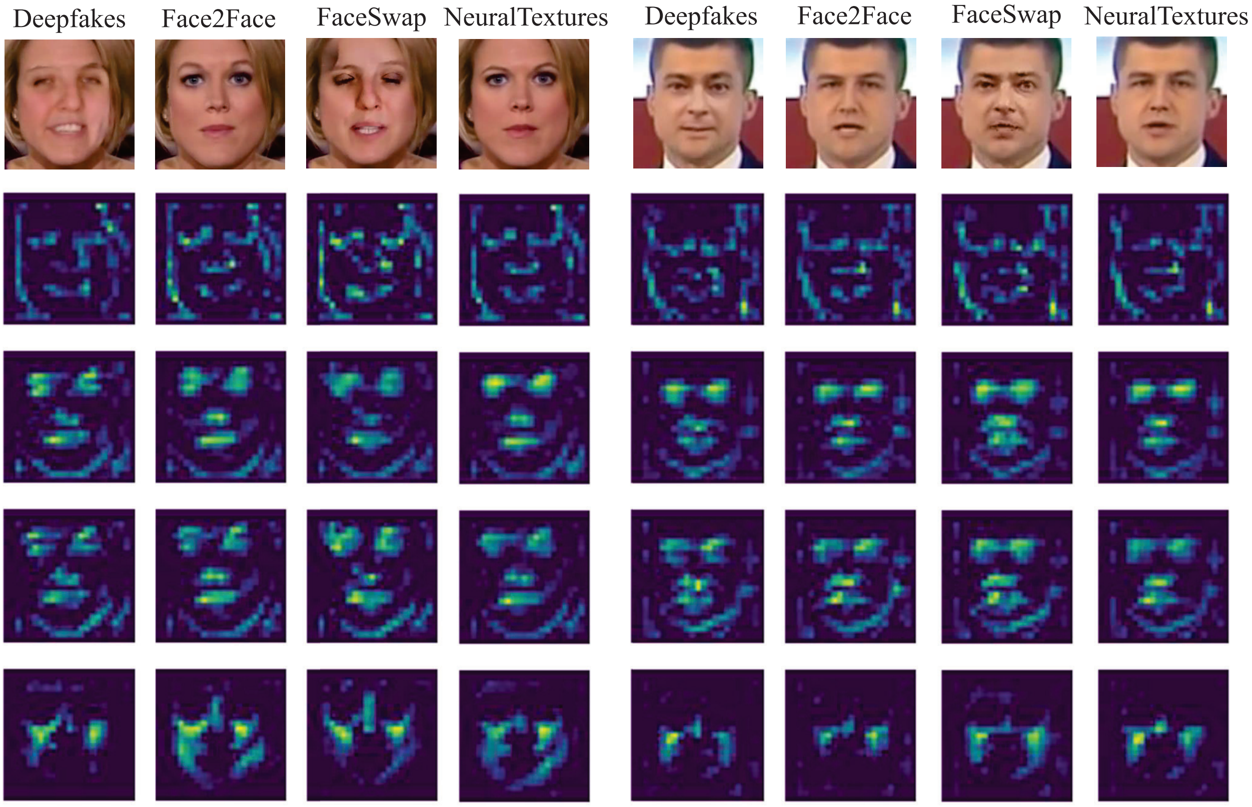

Figure 3. The visualization of overall spatial attention maps trained on FaceForensics++ [10] datasets. The different channels of multiple attention maps are shown in each column.

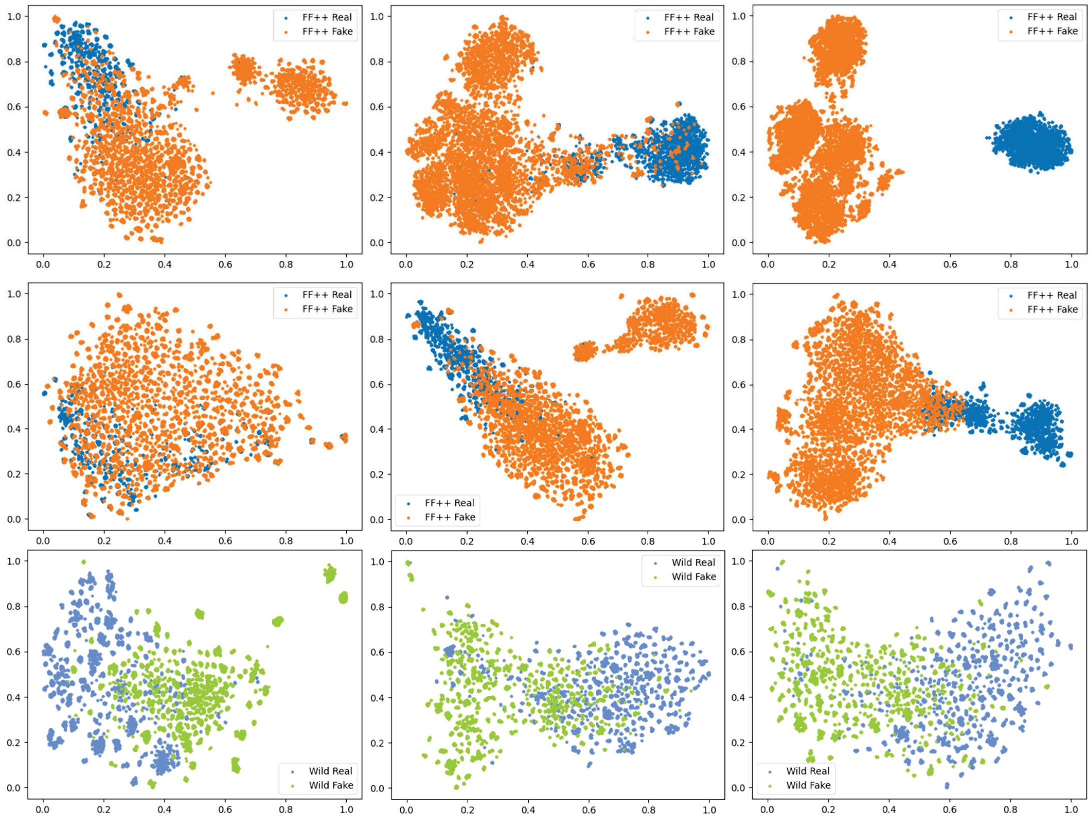

Figure 4. The t-SNE embedding visualization of the feature distribution in the EfficientNet-b4 [12], MADD [17] and SFGD methods. The first two rows display the visualized results on FF++ (HQ) and FF++ (LQ) datasets [10] respectively. The last row shows the visualization on WildDeepfake [18] database. Best viewed in color.

[10] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1–11, 2019. 1, 3, 4, 5

[11] Ke Sun, Taiping Yao, Shen Chen, Shouhong Ding, Jilin Li, and Rongrong Ji. Dual contrastive learning for general face forgery detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2316–2324, 2022. 1

[12] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019. 1, 2, 3, 5

[13] Thies. Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 1

[14] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2387–2395, 2016. 1

[15] M Tora. Deepfakes, 2018. https://github.com/deepfakes/faceswap/tree/v2.0.0. Accessed: 2021-03-29. 1

[16] Chengrui Wang and Weihong Deng. Representative forgery mining for fake face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14923–14932, 2021. 2

[17] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, and Nenghai Yu. Multi-attentional deepfake detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2185–2194, 2021. 1, 2, 3, 5

[18] Bojia Zi, Minghao Chang, Jingjing Chen, Xingjun Ma, and Yu-Gang Jiang. Wilddeepfake: A challenging real-world dataset for deepfake detection. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2382–2390, 2020. 1, 2, 4, 5