Towards Generic Image Manipulation Detection with Weakly-Supervised Self-Consistency Learning — Supplementary Material —

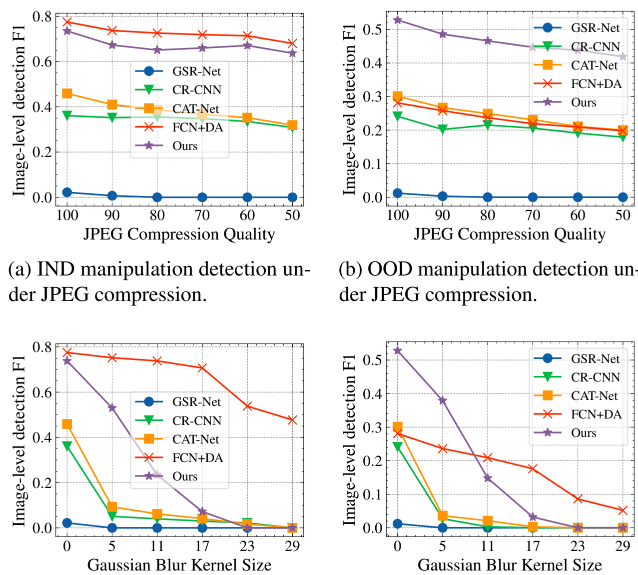
Yuanhao Zhai Tianyu Luan David Doermann Junsong Yuan
University at Buffalo
{yzhai6, tianyulu, doermann, jsyuan}@buffalo.edu

1. Additional Experimental Results

We conduct a set of additional ablation study with MIL-FCN [10] as the baseline.

Robustness comparison with fully-supervised methods. We apply JPEG compression and Gaussian blur separately on CASIAv1 [3] to evaluate the robustness of our method. As shown in Fig. 1a and Fig. 1b, our method effectively defends against JPEG compression, especially under the OOD evaluation, where our method significantly outperforms all competing methods. As for the Gaussian blur, our method resists mild Gaussian blur, but is vulnerable to the blur with large kernel sizes, as shown in Fig. 1c and Fig. 1d. While this limitation highlights the need for further research and optimization to improve our method's resistance to Gaussian blur, it also offers valuable insights into the challenges of robustness in image manipulation detection.

Ablation study on the early fusion architecture. In our method, we use a late fusion architecture to fuse multi-source information. We further evaluate our method on the early fusion architecture, where different sources are concatenated in the channel dimension at input. The results are listed in Tab. 1. Note that multi-source consistency learning and ensemble-supervision inter-patch consistency learning do not apply to the early fusion architecture, and thus are excluded. The results show both adaptive pooling and self-supervision inter-patch consistency learning individually improve the performance of the early fusion architecture, and their combination leads to the best performance, demonstrating the effectiveness of our method under the early fusion architecture. Furthermore, under most settings, the performances in early fusion underperform their counterparts in late fusion. Especially, early fusion performances even underperform several single stream performances in late fusion. Such results show early fusion architecture cannot fully utilize each single source under the weakly-supervised setting, and a late fusion design is needed for the weakly-supervised image manipulation detection and localization.



(a) IND manipulation detection under JPEG compression.

(b) OOD manipulation detection under JPEG compression.

(c) IND manipulation detection under Gaussian blur.

(d) OOD manipulation detection under Gaussian blur.

Figure 1. Robustness evaluation against JPEG compression and Gaussian blur. All methods are trained on CASIAv1 [3]. CASIAv2 [3, 4] is used for IND testing, and the average results on Columbia [6], Coverage [11] and IMD2020 [9] are used for OOD testing. Our method is robust against JPEG compression, and mild Gaussian blur.

AP	IPC	Image-Level				P-F1	C-F1
		AUC	Spe.	Sen.	F1		
-	-	0.597	0.134	0.700	0.225	0.126	0.161
✓	-	0.611	0.158	0.750	0.261	0.142	0.184
-	self	0.641	0.166	0.733	0.271	0.158	0.200
✓	self	**0.682**	0.219	0.771	**0.341**	**0.177**	**0.233**

Table 1. Ablation study on early fusion architecture on IMD2020 [9], where the concatenation of RGB image, Bayar noise map and SRM noise map is fed into a single model. AP is an abbreviation for adaptive pooling.

2. Additional Implementation Details

For unsupervised methods, we use implementations provided by the MKLabl[1], and block size of 2 are used for both CFA1 [5] and NOI1 [8] algorithms.

For the results of fully-supervised methods (HP-FCN [7], Mantra-Net [12], CR-CNN [13], and GSR-Net [14]) on NIST16, Columbia, CASIAv1 and Coverage, we use the reproduced results provided by [1, 2]. And the results of the rest of the data are reproduced by us.

References

[1] Xinru Chen, Chengbo Dong, Jiaqi Ji, Juan Cao, and Xirong Li. Image manipulation detection by multi-view multi-scale supervision. In *ICCV*, 2021. 2

[2] Chengbo Dong, Xinru Chen, Ruohan Hu, Juan Cao, and Xirong Li. Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection. *IEEE TPAMI*, 2022. 2

[3] Jing Dong, Wei Wang, and Tieniu Tan. Casia image tampering detection evaluation database 2010. http://forensics.idealtest.org, 2010. 1

[4] Jing Dong, Wei Wang, and Tieniu Tan. Casia image tampering detection evaluation database. In *2013 IEEE China Summit and International Conference on Signal and Information Processing*, pages 422–426, 2013. 1

[5] Pasquale Ferrara, Tiziano Bianchi, Alessia De Rosa, and Alessandro Piva. Image forgery localization via fine-grained analysis of cfa artifacts. *IEEE Transactions on Information Forensics and Security*, pages 1566–1577, 2012. 2

[6] Yu-Feng Hsu and Shih-Fu Chang. Detecting image splicing using geometry invariants and camera characteristics consistency. In *ICME*, pages 549–552, 2006. 1

[7] Haodong Li and Jiwu Huang. Localization of deep inpainting using high-pass fully convolutional network. In *ICCV*, pages 8301–8310, 2019. 2

[8] Babak Mahdian and Stanislav Saic. Using noise inconsistencies for blind image forensics. *Image and Vision Computing*, pages 1497–1503, 2009. 2

[9] Adam Novozamsky, Babak Mahdian, and Stanislav Saic. Imd2020: A large-scale annotated dataset tailored for detecting manipulated images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 71–80, 2020. 1

[10] Deepak Pathak, Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional multi-class multiple instance learning. In *ICLR*, 2015. 1

[11] Bihan Wen, Ye Zhu, Ramanathan Subramanian, Tian-Tsong Ng, Xuanjing Shen, and Stefan Winkler. Coverage—a novel database for copy-move forgery detection. In *ICIP*, pages 161–165, 2016. 1

[12] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *CVPR*, pages 9543–9552, 2019. 2

[13] Chao Yang, Huizhou Li, Fangting Lin, Bin Jiang, and Hao Zhao. Constrained r-cnn: A general image manipulation detection model. In *ICME*, pages 1–6, 2020. 2

[14] Peng Zhou, Bor-Chun Chen, Xintong Han, Mahyar Najibi, Abhinav Shrivastava, Ser-Nam Lim, and Larry Davis. Generate, segment, and refine: Towards generic manipulation segmentation. In *AAAI*, pages 13058–13065, 2020. 2

[1] https://github.com/MKLab-ITI/image-forensics