Supplementary Material
TrainFors: A Large Benchmark Training Dataset for Image Manipulation Detection and Localization

Soumyaroop Nandi, Prem Natarajan, Wael Abd-Almageed
USC Information Sciences Institute, Marina del Rey, CA, USA
{soumyarn,pnataraj,wamageed}@isi.edu

1. More Qualitative Visualization

We reported a few more visualization examples of the image manipulation localization predictions in Fig. 1. We have shown the pixel-map predictions on the five evaluation datasets - Casia [3], Columbia [7], Coverage [11], Nist16 [1], and IMD20 [8] by the four baseline image manipulation detection and localization (IMDL) models - MVSS-Net [2], Cat-Net [5], PSCCNet [6], and ObjectFormer [10] when trained on our TrainFors dataset. We found in our investigation that PSCCNet [6] and MVSS-Net [2] gave comparatively better manipulation localization pixel-maps, PSCC-Net results being more consistent. But again with better prediction, more false alarm predictions are made by PSCCNet, as seen in the Coverage example. All four baseline models failed in predicting a few forged images from the IMD20 dataset as shown in the last row of Fig. 1. The IMD20 consists of real-life manipulated images and the baseline IMDL models were not very promising in predicting the manipulated pixel maps in such cases.

2. Robustness Evaluation: Manipulation Detection

All the previous research [12], [4], [6], [10] reported a robustness evaluation on the pixel level manipulation localization task, but none reported the robustness analysis on the image level-manipulation detection task. It is equally important to report the latter for a fair comparison of the robustness performance of the baseline IMDL models. Moreover, [2] and [5] did not perform any robustness evaluation. Similar to pixel-level robustness evaluation, we added the following distortions to the manipulated images: **Image scaling** with scales=0.78X, 0.25X, **Gaussian Blurring** with a kernel size k=3, 15, added **Gaussian Noise** with a standard deviation σ= 3, 15, and **JPEG Compression** with a quality factor q=50, 100. We also added a mix of these distortions in the mixed column and the No Distortion column. We compared the manipulation detection performance (AUC scores) of the pre-trained models with all the baseline methods on these distorted images and report the results on the Columbia dataset in Tab. 1. We cannot evaluate the robustness of manipulation detection on the Nist16 [1] dataset because there are no pristine (negative) images in the dataset. Objectformer showed promising results on the robustness evaluation of the manipulation detection task when trained on the author-specified backbone network. But as soon as we used the EfficientNetV2 [9] backbone network for pre-training, the image manipulation detection robustness evaluation improved for the other three IMDL models. PSCC-Net [6] achieved the best performance when Gaussian Noise and Gaussian Blurring Distortion were added and MVSS-Net [2] achieved the best performance for resizing the images, JPEG-compressing the raw images, and also on the addition of a mix of all the distortions. Cat-Net's [5] performance on image level manipulation detection task is much better when compared with its performance on the pixel-level manipulation localization task. The reason could be an imbalance between the true positive and false alarm predictions, with false alarm predictions contributing more toward the increase of the manipulation detection rate.

3. Limitations

The major drawback of image manipulation detection and localization (IMDL) tasks is finding a balance between true positive and false alarm predictions. Some of the previous research works showed good model predictions for manipulation localization pixel maps, but they did not report their false alarm rates. The IMDL models fail to detect synthetically generated images from generative methods like - Generative Adversarial Networks (GANs) and Diffusion Models (DMs). A uniform training dataset should be utilized for training and fair evaluation of the IMDL models and we tried to address this by curating the TrainFors dataset in this work.

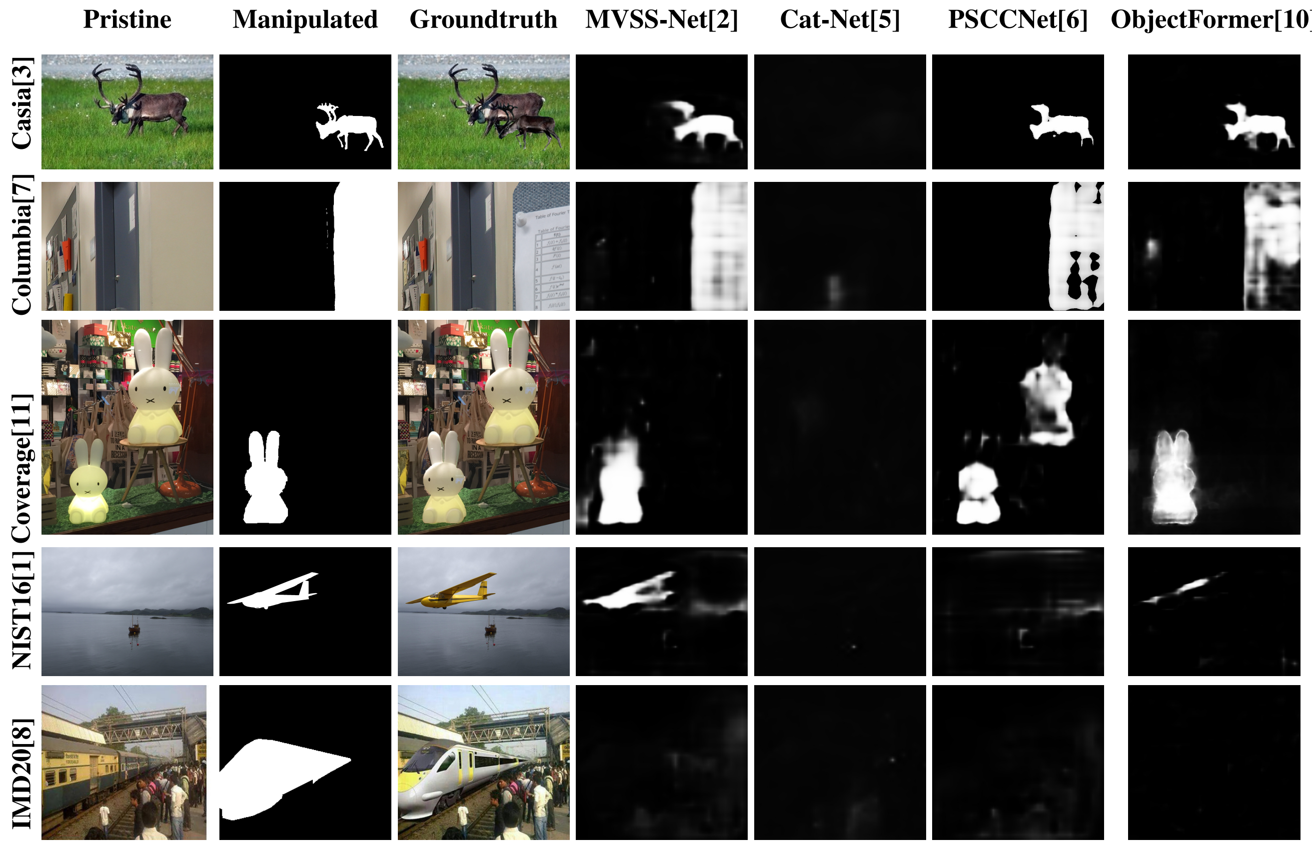

Figure 1: Image Manipulation Localization Prediction Visualization - From top to bottom, we show examples from 5 benchmark evaluation datasets: Casia, Columbia, Coverage Nist16, and IMD20 by 4 baseline IMDL models: MVSS-Net, Cat-Net, PSCC-Net, and ObjectFormer, trained on our proposed TrainFors dataset.

	No Distortion	Resize (0.78X)	Resize (0.25X)	Gau-Blur (k=3)	Gau-Blur (k=15)	Gau-N (σ=3)	Gau-N (σ=15)	JPG-Comp (q=100)	JPG-Comp (q=50)	Mixed
					Columbia					
Author-Specified Backbone										
MVSS-Net[2]	82.1	81.9	81.8	77.9	74.3	68.7	66.5	78.7	78.8	65.3
Cat-Net[5]	81.6	81.5	81.4	77.6	73.8	67.4	65.6	77.9	78.1	64.2
PSCCNet[6]	83.4	82.9	82.8	79.8	77.1	71.4	70.2	80.3	80.5	69.7
ObjectFormer[10]	84.8	84.6	84.5	80.3	77.9	72.7	71.4	82.1	82.8	70.8
EfficientNetV2 [9] Backbone										
MVSS-Net[2]	**85.6**	**85.5**	**85.4**	81.4	78.2	73.7	72.9	**83.6**	**83.8**	**72.4**
Cat-Net[5]	83.5	82.9	82.7	80.6	77.3	72.6	70.8	81.3	81.5	70.1
PSCCNet[6]	85.4	84.8	84.5	**81.6**	**79.1**	**73.8**	**73.4**	83.4	83.5	71.7
ObjectFormer[10]	84.8	84.6	84.5	80.3	77.9	72.7	71.4	82.1	82.8	70.8

Table 1: Robustness Comparison of Image-level Manipulation Detection AUC(%) with various distortions evaluated on Columbia[7] dataset, when pretrained with author-specified backbone networks and EfficientNetV2 [9] backbone network respectively

References

[1] Nimble challenge 2017 evaluation — nist. https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation. (Accessed on 11/14/2020).

[2] Xinru Chen, Chengbo Dong, Jiaqi Ji, Juan Cao, and Xirong Li. Image manipulation detection by multi-view multi-scale supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14185–14193, 2021.

[3] 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. IEEE, 2013.

[4] Xuefeng Hu, Zhihan Zhang, Zhenye Jiang, Syomantak Chaudhuri, Zhenheng Yang, and Ram Nevatia. Span: spatial pyramid attention network for image manipulation localization. In *European Conference on Computer Vision*, pages 312–328. Springer, 2020.

[5] Myung-Joon Kwon, In-Jae Yu, Seung-Hun Nam, and Heung-Kyu Lee. Cat-net: Compression artifact tracing network for detection and localization of image splicing. In

Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 375–384, 2021.

[6] Xiaohong Liu, Yaojie Liu, Jun Chen, and Xiaoming Liu. Pscc-net: Progressive spatio-channel correlation network for image manipulation detection and localization. *arXiv preprint arXiv:2103.10596*, 2021.

[7] Tian-Tsong Ng, Jessie Hsu, and Shih-Fu Chang. Columbia image splicing detection evaluation dataset. *DVMM lab. Columbia Univ CalPhotos Digit Libr*, 2009.

[8] 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.

[9] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[10] Junke Wang, Zuxuan Wu, Jingjing Chen, Xintong Han, Abhinav Shrivastava, Ser-Nam Lim, and Yu-Gang Jiang. Objectformer for image manipulation detection and localization. *arXiv preprint arXiv:2203.14681*, 2022.

[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 *2016 IEEE International Conference on Image Processing (ICIP)*, pages 161–165. IEEE, 2016.

[12] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9543–9552, 2019.