Supplementary Material for SPAN: Spatial Pyramid Attention Network for Image Manipulation Localization

Xuefeng Hu[1], Zhihan Zhang[1], Zhenye Jiang[1], Syomantak Chaudhuri[2], Zhenheng Yang[3], and Ram Nevatia[1]

[1] University of Southern California
{xuefeng,zhihanz,zhenyeji,nevatia}@usc.edu
[2] Indian Institute of Technology, Bombay
syomantak@iitb.ac.in
[3] Facebook AI
zhenheny@fb.com

Dataset Characteristics

In Section 4.3 we mentioned that the characteristic difference between Columbia, Coverage and CASIA could be a possible reason of the different scale of performance improvement the SPAN achieved over RGB-N[6]. In Figure 1, three images from each of these three datasets are presented. As shown, in CASIA, the manipulated objects tend to correspond to semantic objects compared to the other two.

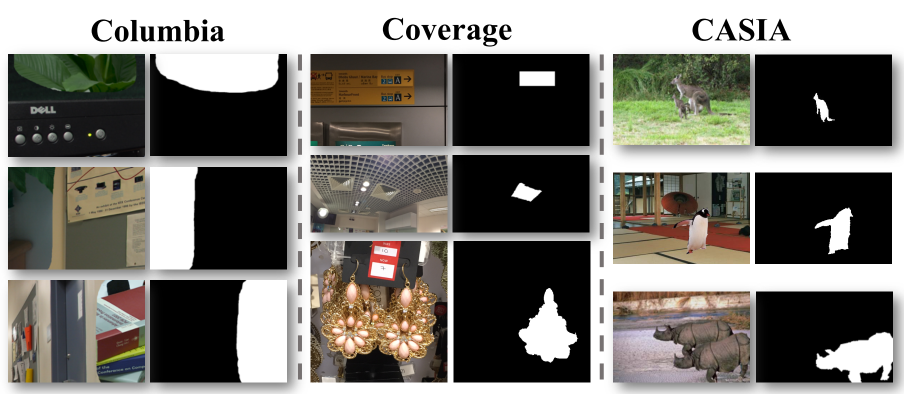

Fig. 1: The characteristic difference between Columbia, Coverage and CASIA.

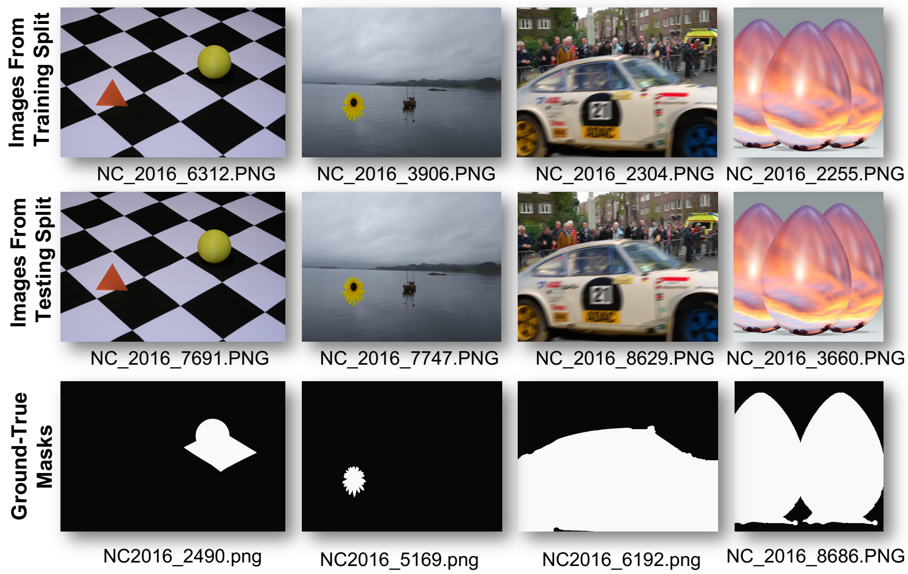

Fig. 2: Demonstration of NIST16[3] information overlapping between its training and testing splits.

NIST16 Image Overlap between Traning and Testing

In Section 4.3 we mentioned that there is an image overlap issue between training and test data on NIST16[3], according to the training and testing split provided by RGB-N[6]. There are 584 images in NIST16 dataset, with 292 pairs of unique image and compressed JPEG version. Following the same practice in [6], the test split contains 160 images. 151 out of the 160 (88%) test images have the other visually similar version in training split and they share the same ground truth mask. Some overlapping images are shown in Figure 2. From top to bottom row, the image in training split, image in testing split and the ground truth mask of these two images are presented.

More Qualitative Results

In Figure 3, we present the qualitative results with comparison with RGB-N[6]. As mentioned in Section 4.3 of the paper, RGB-N (which is built off Faster R-CNN [4]) generates region-level prediction thus only predicts rectangle mask, restricting it from generating more accurate shape of tampered region, as shown in the second row. On the other hand, there are also many test samples where the tampered region is naturally a rectangle, like the one in the first row of Figure 3

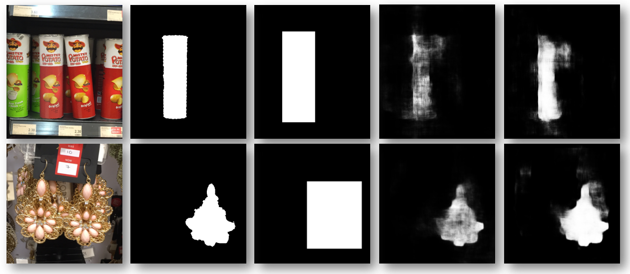

Fig. 3: Qualitative Results of SPAN predictions on Coverage and NIST16 datasets, with comparison to RGB-N predictions. From left to right: Manipulated Image, Ground-truth mask, RGB-N prediction, SPAN (pre-training) prediction and SPAN(fine-tuned) prediction.

We also present more visualization results of SPAN on four different datasets: NIST16 in Figure 4, CASIA in 5, Columbia in Figure 6 and Coverage in Figure 7. We sampled 5 examples (shown in first 5 columns in each image, surrounded by the green box) and 2 additional failure cases (shown in last two columns in each image, surrounded by the red box) from each dataset.

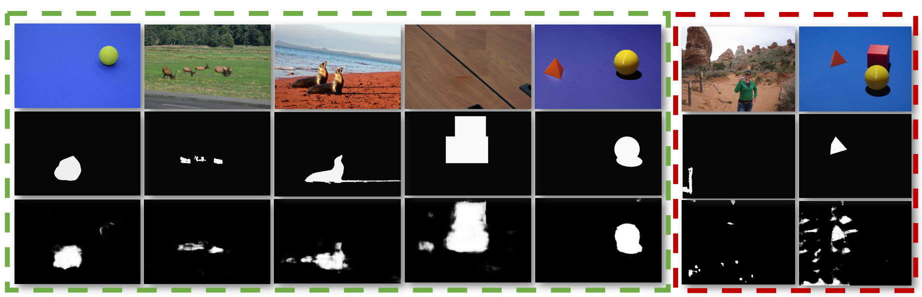

Fig. 4: Qualitative Results on NIST16[3]. Top to bottom: Manipulated Image, Ground-Truth Mask, SPAN(fine-tuned) Prediction Mask. Red box contains failure samples.

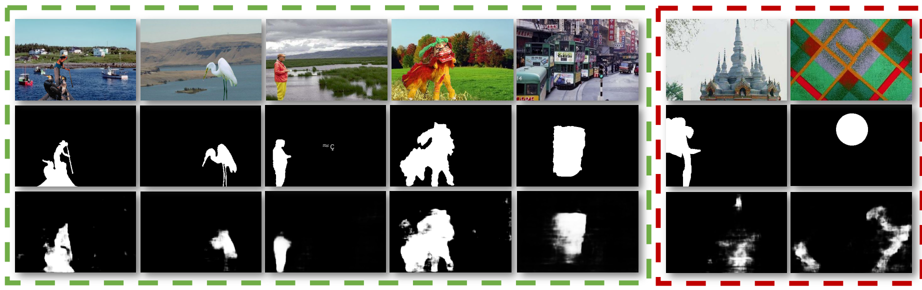

Fig. 5: Qualitative Results on CASIAv1[1]. Top to bottom: Manipulated Image, Ground-Truth Mask, SPAN(fine-tuned) Prediction Mask. Red box contains failure samples.

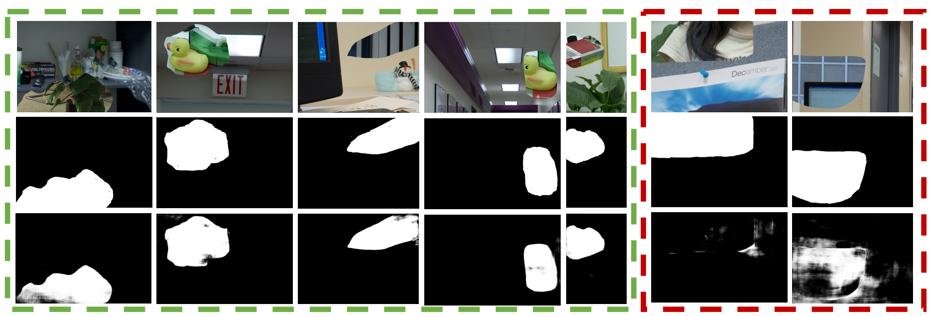

Fig. 6: Qualitative Results on Columbia[2]. Top to bottom: Manipulated Image, Ground-Truth Mask, SPAN Prediction Mask. Red box contains failure samples.

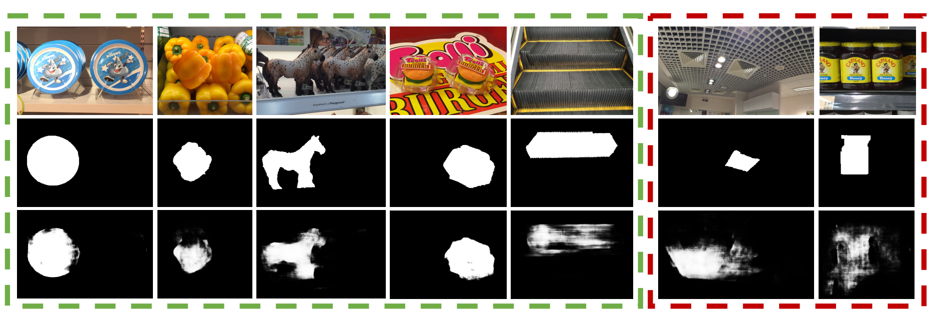

Fig. 7: Qualitative Results on Coverage[5]. Top to bottom: Manipulated Image, Ground-Truth Mask, SPAN(fine-tuned) Prediction Mask. Red box contains failure samples.

References

1. Dong, J., Wang, W., Tan, T.: Casia image tampering detection evaluation database. In: 2013 IEEE China Summit and International Conference on Signal and Information Processing. pp. 422–426. IEEE (2013)
2. Ng, T.T., Hsu, J., Chang, S.F.: Columbia image splicing detection evaluation dataset. DVMM lab. Columbia Univ CalPhotos Digit Libr (2009)
3. NIST: Nist nimble 2016 datasets (2016), https://www.nist.gov/itl/iad/mig/
4. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
5. Wen, B., Zhu, Y., Subramanian, R., Ng, T.T., Shen, X., Winkler, S.: Coverage—a novel database for copy-move forgery detection. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 161–165. IEEE (2016)
6. Zhou, P., Han, X., Morariu, V.I., Davis, L.S.: Learning rich features for image manipulation detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1053–1061 (2018)