Supplementary Material: Occlusion-shared and Feature-separated Network for Occlusion Relationship Reasoning

Rui Lu[1] Feng Xue[1] Menghan Zhou[1,3] Anlong Ming[1] Yu Zhou[2,*]
[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]Huazhong University of Science and Technology, Wuhan, China
[3]Lenovo Research, Beijing, China
{lurui,xuefeng,mal}@bupt.edu.cn zhoumh3@lenovo.com yuzhou@hust.edu.cn

In this supplementary material, we provide full qualitative analysis for the ablation study. The experiments are conducted on the PIOD dataset.

1. One-branch or Multi-branch Sub-networks

Previous approach DOOBNet [1] adopts a single flow architecture by sharing decoder features that represent high-level features. The shared decoder features reflect the contour cues, which are necessary for both edge and orientation estimations. Besides, edge detection and orientation detection are different in the choice of feature extraction, especially in the case of high semantic layers. We innovatively split the features produced by side-outputs and share decoder features to fit both tasks, respectively. Fig.1 reveals the effectiveness of our design.

2. Necessity for Each Feature

To verify the role of various low-level and high-level features, each feature is removed to construct an independent variant for evaluation. If the low-level features for edge path are removed, the occlusion edge is difficult to be accurately located, leading to decrease in the accuracy of occlusion relationship reasoning (shown in Fig.2(*w/o low-cues*)). If the high-level feature for edge path is removed, the occlusion edge is failed to be detected consistently, which decreases the accuracy at a large margin (shown in Fig.2(*w/o high-cues*)). By capturing spatial and contextual cues from the side-outputs respectively, the network is able to explore specific features for individual predictions.

3. Proportion of Bilateral-Contour Features

Previous works utilize inappropriate feature maps to predict the orientation, which reflects as the characteristic of the edge outline. The features of orientation on both sides of the contour are filtered gradually which are adversely affected by the edge prediction. We take advantage of an **MCL** to perceive the bilateral cues around the contours, and affirm the foreground and background relationship. As shown in Fig.3, fusing bilateral feature and occlusion feature with *64:16* channel ratio in the **BRF** outperforms others.

4. Plain or Stripe Convolution

Plain convolutions perceive information about surrounding small areas. To extract the tendency of edges to extend and bilateral cues around contours, we employ stripe convolutions in orthogonal directions. The convolution kernels possess large receptive field and tend to learn the cues of both directions, respectively. We test stripe convolution kernels with different aspect ratios, which are exhibited in Fig.3. The larger convolution layer takes up too much computation cost, which increases the number of parameters. We evaluate the performance of the model with *11×11 conv* on PIOD and BSDS datasets, the *EPR* (left) and *OPR* (right) are reported in Table.1. Compared with *3×11 conv*, the model with *11×11 conv* achieves limited improvement, while it increases about *50*% gpu memory usage (*10031MB* to *14931MB*).

Table 1: Results of our model with different conv kernel sizes.

Dataset	Scale	ODS	OIS	AP	ODS	OIS	AP
PIOD	*conv = 3×11*	.751	.762	.773	.718	.728	.729
	conv = 11×11	.753	.764	.776	.719	.730	.732
BSDS	*conv = 3×11*	.662	.689	.585	.583	.607	.501
	conv = 11×11	.663	.690	.587	.585	.608	.503

References

[1] Guoxia Wang, Xiaohui Liang, and Frederick Li. Doobnet: Deep object occlusion boundary detection from an image. In *ACCV*, 2018.

*Corresponding Author

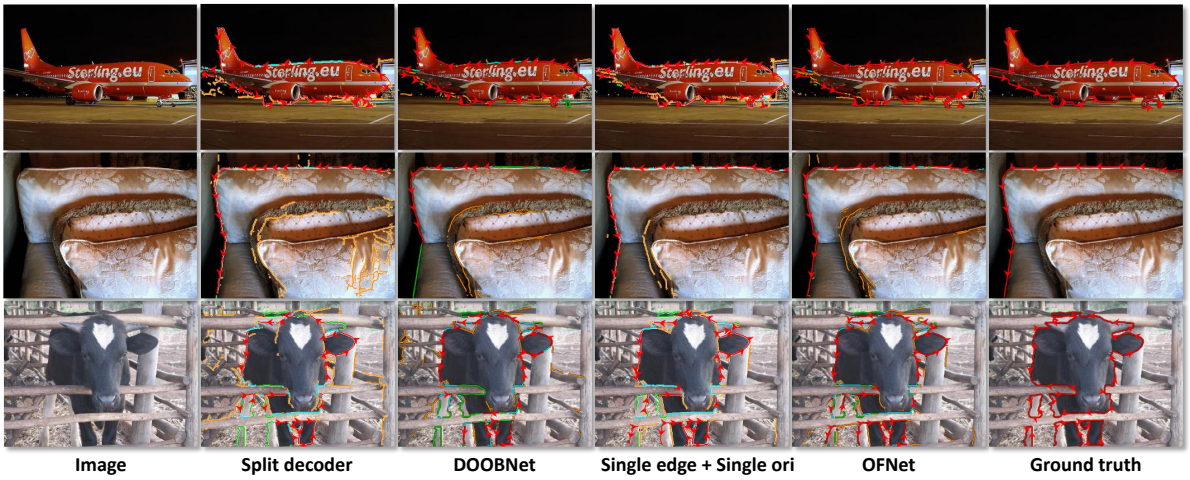

Figure 1: Occlusion relationship of various approaches. The occlusion relationship (the red arrows) is represented by orientation $\theta \in (-\pi, \pi]$ (tangent direction of the edge), using the "left" rule where the left side of the arrow means foreground area. Notably, "red" pixels with arrows: correctly labeled occlusion boundaries; "cyan": correctly labeled boundaries but mislabeled occlusion; "green": false negative boundaries; "orange": false positive boundaries (Best viewed in color). Column 1^{st}: Input image. Column $2^{nd} - 5^{th}$: Output of split decoder, DOOBNet, Single edge + single ori and OFNet. Column 6^{th}: Ground truth.

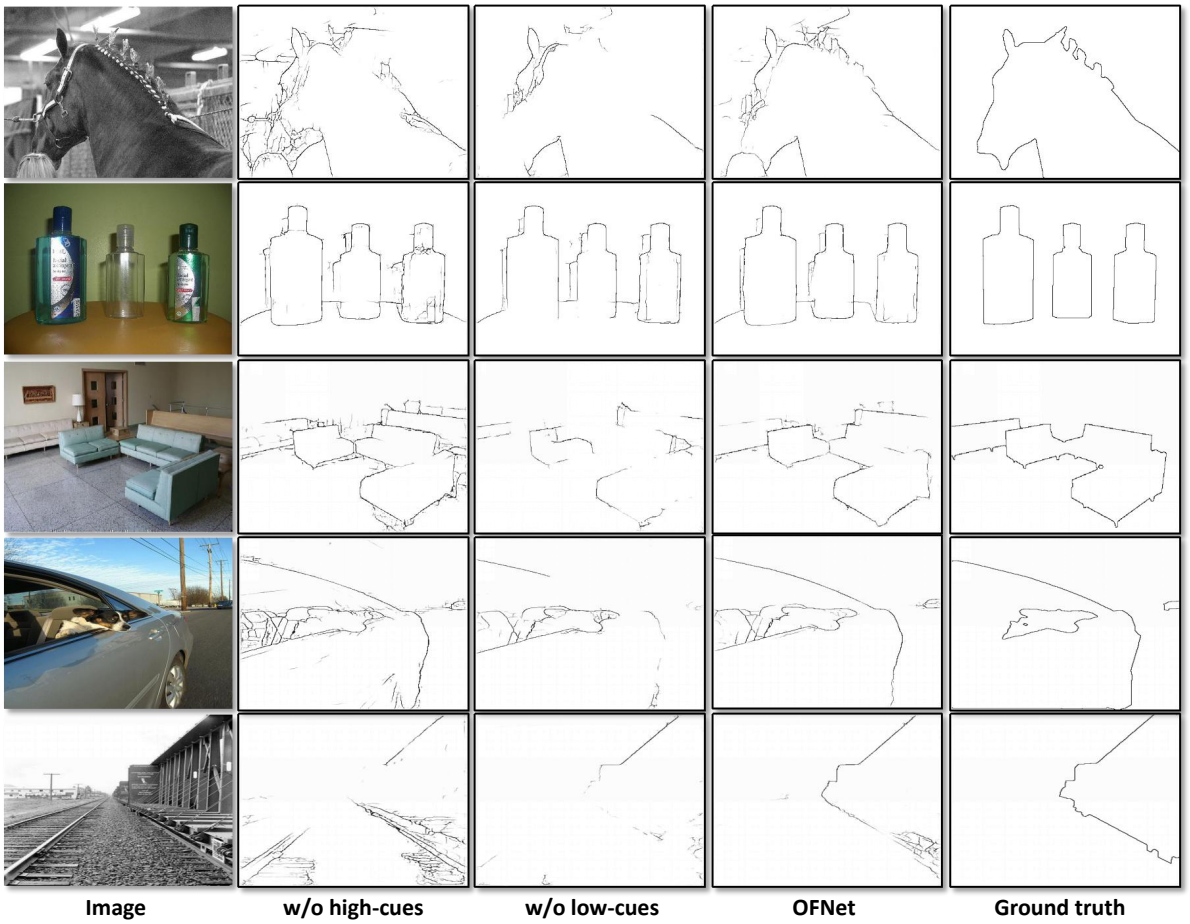

Figure 2: Edge maps of various approaches. Column 1^{st}: Input image. Column $2^{nd} - 4^{th}$: OFNet without high-cues, OFNet without low-cues and OFNet. Column 5^{th}: Ground truth.

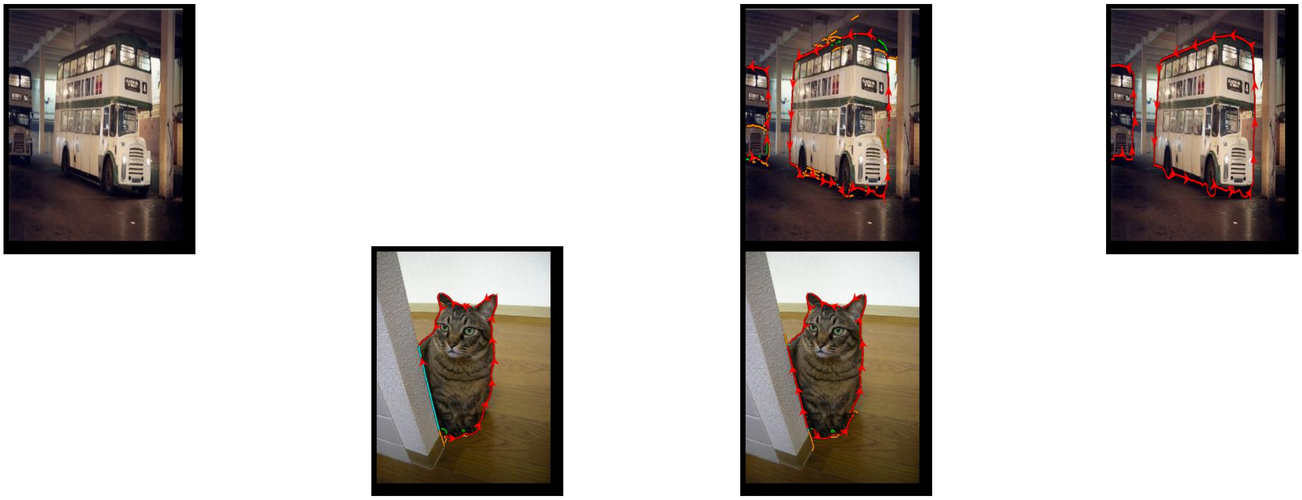

Figure 3: Occlusion relationship of various approaches. Column 1^{st}: Input image. Column $2^{nd} - 6^{th}$: Fusing bilateral feature and occlusion feature with 16:16, 32:16, 48:16, 80:16 and 64:16 channel ratio, respectively. Column 7^{th}: Ground truth.

Figure 4: Edge maps of various approaches. Column 1^{st}: Input image. Column $2^{nd} - 7^{th}$: conv kernel size = 3×3, 3×5, 3×7, 3×9, 3×11 and 11×11. Column 8^{th}: Ground truth.