Learning Semantic Associations for Mirror Detection (Supplementary Material)

Anonymous CVPR 2022 submission

Paper ID 1159

1. Semantic Side-Path

We adopt ResNet50 [7] with multi-grid method [3] as our semantic side-path and train it on the collected semantic annotations as we mentioned in the main paper. Here we provide more details about the collection procedure of the semantic annotations.

To train the semantic side-path, we have collected 4,746 training samples containing pixel-wise semantic annotations for the large-scale mirror detection dataset, PMD [9]. We choose to collect the semantic annotations of PMD dataset [9] for the following considerations:

- **Diversity.** The diversity of PMD dataset [9] is much higher than MSD dataset [15] according to the statistics reported by Lin *et al*. [9]. It indicates that more complex scenes and common objects are captured in the images, which definitely involves stronger semantic associations between objects. We believe that this undoubtedly can enhance model's robustness as more cues are exposed by exploring the semantic associations.
- **Lower the expense.** Since PMD dataset [9] is collected from six public datasets [19, 20, 13, 2, 14, 12, 1], while most of them [19, 20, 13, 2, 14, 12] contain pixel-wise semantic annotations. As a result, we can lower our expense by extracting the semantic annotations from the existing datasets.

However, we notice that different datasets may involved different categories. For example, COCO dataset [10] contains 80 object categories and 91 stuff categories, while ADE20K [19, 20] has up to 3K object categories, which may cause inconsistent problems if use these labels without any pre-process. As a result, we manually unified the categories across different datasets and, at the end, we have manually selected the top 24 classes mainly based on the frequency of occurrence as foreground and regarded the other classes as background. So, there are totally 25 classes in our semantic annotations set. Figure 1 presents one example. It can be seen that the semantic annotations may be inaccurate when encountering mirrors, for example, the reflected person is not considered as a part of mirror. Fortunately, the semantic maps usually expose rich contexts and our method can handle the complex cases successfully by exploiting semantic knowledge and associations. We post more results in Figure 2.

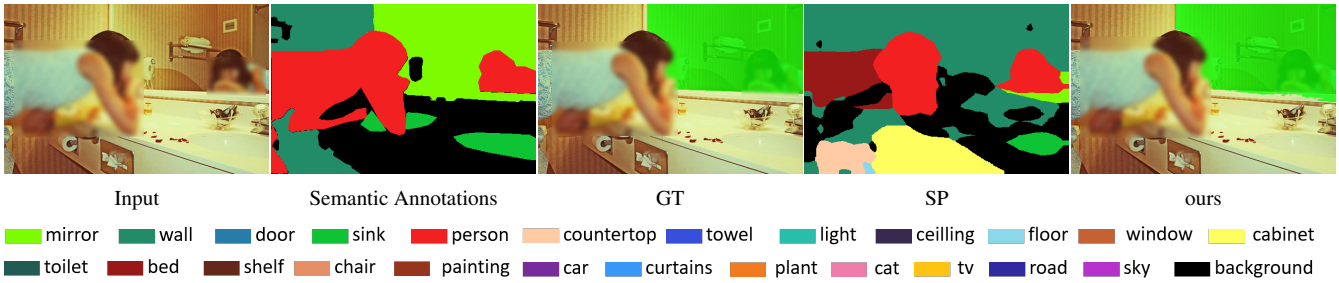

Figure 1. **Semantic Annotations, GT and our model's predictions.**

Figure 2. **Semantic Annotations, GT and our model's predictions.**

2. More Visual Results

2.1. Mirror Detection Results of images from internet

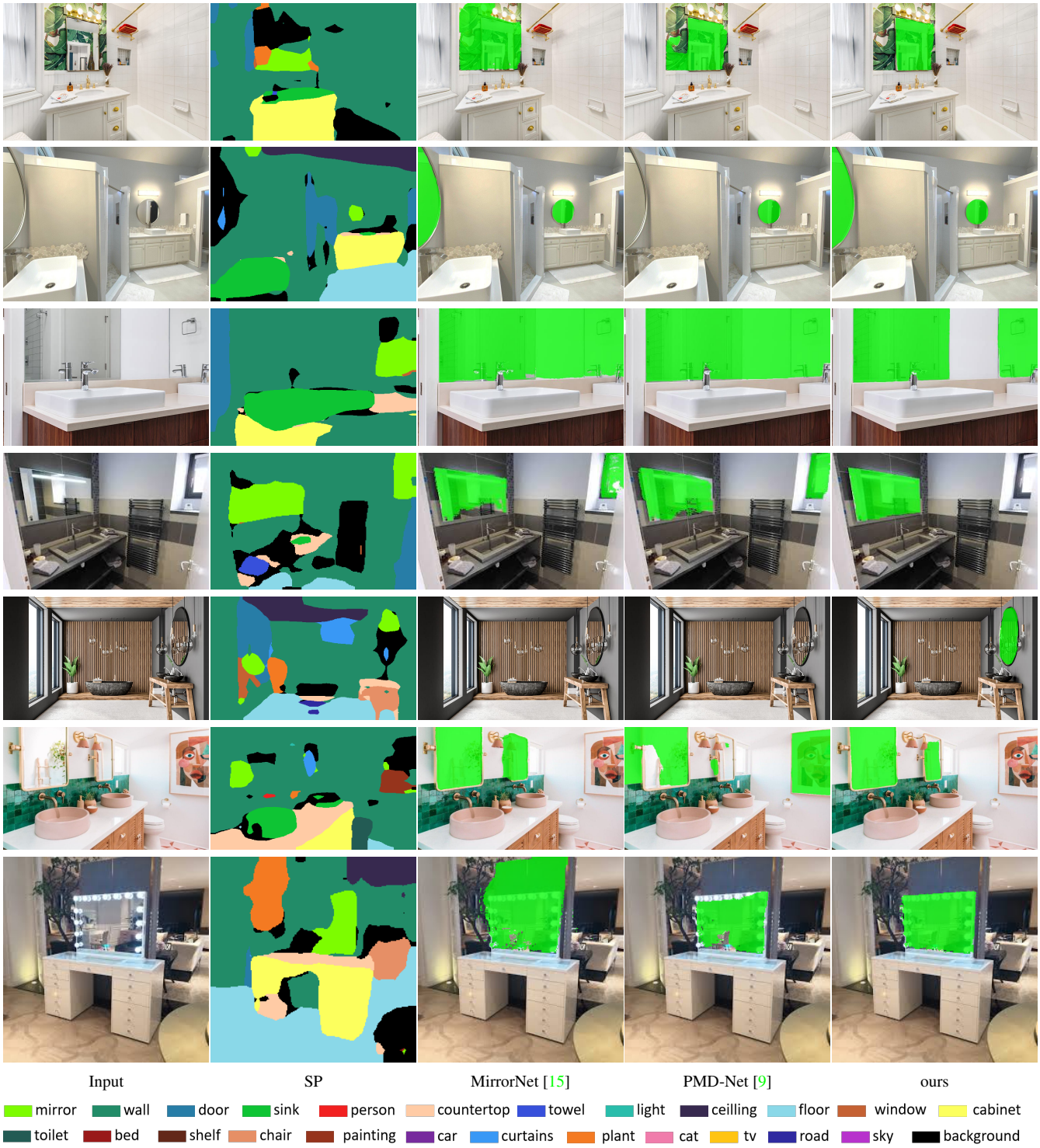

Figure 3. **Visual comparison on images from internet.** In general, our model can be more robust in real-world scenes.

2.2. Mirror Detection Results of images from datasets

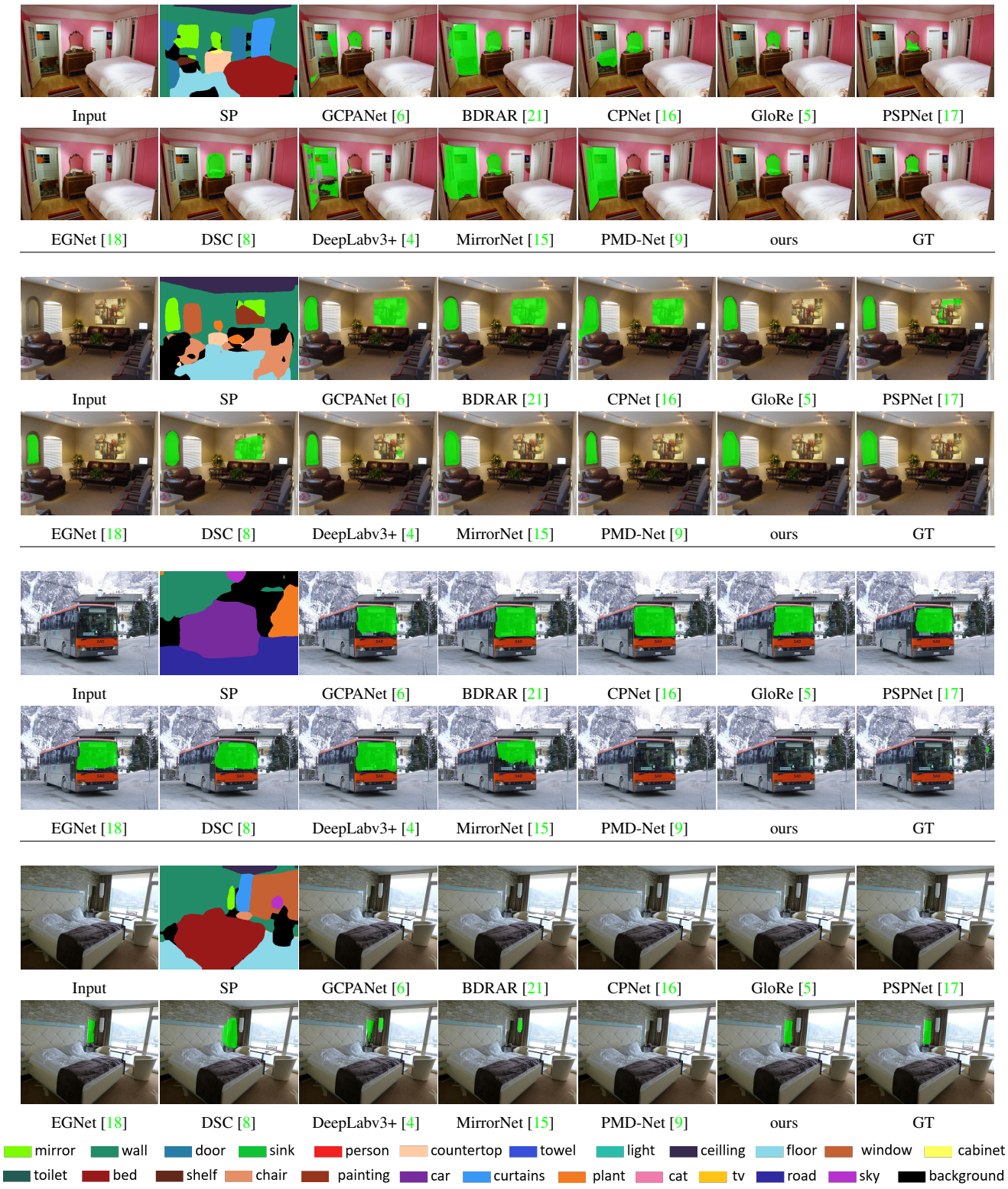

Figure 4. **Visual Comparison on PMD [9].**

Figure 5. **Visual Comparison on MSD [15]**

3. Statement and Acknowledgement

3.1. Licenses of Assets

PMD dataset [9]. Since the images in PMD are collected from six public datasets, all the images inherit their original datasets' licenses. The annotations of PMD are licensed under a BSD 3-clause License and only be used for non-commercial purposes. We access to this dataset via https://jiaying.link/cvpr2020-pgd/.

MSD dataset [15]. You may access to the content of license adopted by this dataset in https://github.com/Mhaiyang/ICCV2019_MirrorNet. The name of license is not clear.

RGBD dataset [11]. We access to this dataset via https://mhaiyang.github.io/CVPR2021_PDNet/index, where you may find the "Terms and Conditions" for this assets.

Images from internet. We collect some examples from internet to evaluate our model. We collect these images manually by searching some keywords on Google. We try our best to avoid involving any personal data.

3.2. Protection on Personal Data/Human Subjects

We highly respect the privacy of personal data/human subjects. This work involves a number of images, which raise a risk of privacy leak. We carefully check our materials and may use blur technique to protect the privacy of personal data/human subjects.

3.3. Others

We will release our code and results once the paper is accepted.

References

[1] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. Material recognition in the wild with the materials in context database. In *2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*. 1

[2] Holger Caesar, Jasper R. R. Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*. 1

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017. 1

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision - ECCV 2018 - 15th European Conference, 2018*. 4, 5

[5] Yunpeng Chen, Marcus Rohrbach, Zhicheng Yan, Shuicheng Yan, Jiashi Feng, and Yannis Kalantidis. Graph-based global reasoning networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019*. 4, 5

[6] Zuyao Chen, Qianqian Xu, Runmin Cong, and Qingming Huang. Global context-aware progressive aggregation network for salient object detection. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*. 4, 5

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*. 1

[8] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*. 4, 5

[9] Jiaying Lin, Guodong Wang, and Rynson W. H. Lau. Progressive mirror detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. 1, 3, 4, 5, 6

[10] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision - ECCV 2014 - 13th European Conference, 2014*. 1

[11] Haiyang Mei, Bo Dong, Wen Dong, Pieter Peers, Xin Yang, Qiang Zhang, and Xiaopeng Wei. Depth-aware mirror segmentation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021*. 6

[12] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan L. Yuille. The role of context for object detection and semantic segmentation in the wild. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*. 1

[13] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from RGBD images. In Andrew W. Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, 2012*. 1

[14] Shuran Song, Samuel P. Lichtenberg, and Jianxiong Xiao. SUN RGB-D: A RGB-D scene understanding benchmark suite. In *2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*. 1

[15] Xin Yang, Haiyang Mei, Ke Xu, Xiaopeng Wei, Baocai Yin, and Rynson W.H. Lau. Where is my mirror? In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*. 1, 3, 4, 5, 6

[16] Changqian Yu, Jingbo Wang, Changxin Gao, Gang Yu, Chunhua Shen, and Nong Sang. Context prior for scene segmentation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. 4, 5

[17] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 4, 5

[18] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet:edge guidance network for salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019. 4, 5

[19] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ADE20K dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 1

[20] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ADE20K dataset. *Int. J. Comput. Vis.*, 127(3):302–321, 2019. 1

[21] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision - ECCV 2018 - 15th European Conference, 2018*. 4, 5