AdaStereo: A Simple and Efficient Approach for Adaptive Stereo Matching –Supplementary Material

Xiao Song[1⋆] Guorun Yang[1,3⋆] Xinge Zhu[2] Hui Zhou[1] Zhe Wang[1,4] Jianping Shi[1,5]
[1]SenseTime Research [2]The Chinese University of Hong Kong
[3]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[4]Shanghai AI Laboratory [5]Qing Yuan Research Institute, Shanghai Jiao Tong University
{songxiao,yangguorun,zhouhui,wangzhe,shijianping}@sensetime.com zx018@ie.cuhk.edu.hk

1. Implementation Details

1.1. Specifications of Stereo Models

Here we introduce the network structures of our domain-adaptive stereo models: Ada-ResNetCorr and Ada-PSMNet. Ada-ResNetCorr is extended based on the ResNetCorr [5, 6, 4], which is a baseline model among correlation-based 2-D disparity networks. We extend the ResNetCorr, where "conv1_1" to "conv1_3" from the ImageNet-pretrained ResNet-50 [2] are adopted as the shallow feature extractor. We also replace the single-channel output convolutional layer with the soft-argmin block in 3-D disparity networks [3, 1] for disparity regression. The proposed cost normalization layer is directly applied on the extracted lower-layer features of two views before correlation, which are of $1/2$ spatial size to the input image. The maximum displacement in the correlation layer is set to 128. Ada-PSMNet is extended based on the PSMNet [1], which is a baseline model among cost-volume based 3-D disparity networks. We follow the network structure of the PSMNet except the maximum disparity range is set to 256. The cost normalization layer is directly applied on the extracted lower-layer features of two views before constructing the 4-D cost volume.

1.2. Specifications of Self-Supervised Occlusion-Aware Reconstruction

Here we introduce the network structure of our occlusion prediction network (only adopted during training), which is a small three-layer fully-convolutional network. The first and second convolutional layers both use the 3×3 kernel and the stride of 1 with 32 output channels, followed by a batch normalization layer and a ReLU layer each. The last convolutional layer uses the 1×1 kernel and the stride of 1 with 1 output channel, afterwards the sigmoid layer is adopted for activation. Finally, a full-size occlusion mask is produced, whose element denotes the per-pixel occlusion probability from 0 to 1.

2. Qualitative Results of Color Transfer

As shown in Fig. 1, we provide qualitative results of our color transfer algorithm, from a synthetic dataset (source domain) to varied real-world datasets (target domain). As can be seen, our method enables highly effective image style translations, meanwhile ensuring the semantic invariance without geometrical distortions, which is of vital importance for the low-level stereo matching task.

References

[1] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In *CVPR*, 2018. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[3] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression. In *ICCV*, 2017. 1

[4] Xiao Song, Xu Zhao, Liangji Fang, Hanwen Hu, and Yizhou Yu. Edgestereo: An effective multi-task learning network for stereo matching and edge detection. *IJCV*, 2020. 1

[5] Guorun Yang, Zhidong Deng, Hongchao Lu, and Zeping Li. Src-disp: Synthetic-realistic collaborative disparity learning for stereo matching. In *ACCV*, 2018. 1

[6] Guorun Yang, Hengshuang Zhao, Jianping Shi, Zhidong Deng, and Jiaya Jia. Segstereo: Exploiting semantic information for disparity estimation. In *ECCV*, 2018. 1

⋆ indicates equal contribution.

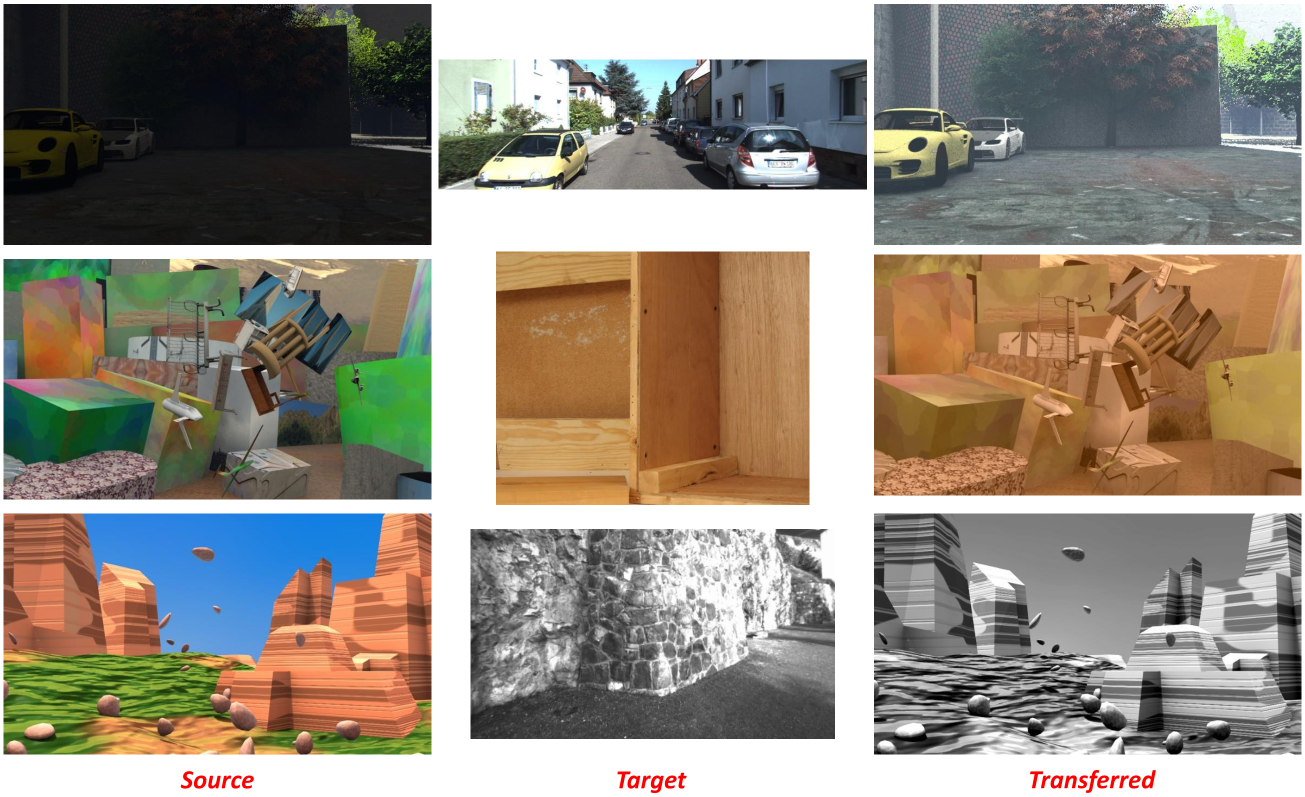

Figure 1. Qualitative results of color transfer from SceneFlow to real-world datasets. Top-down: transfer to KITTI, Middlebury, and ETH3D.