Supplementary Materials for On the Over-Smoothing Problem of CNN Based Disparity Estimation

Chuangrong Chen
Sun Yat-sen University
chenchr5@mail2.sysu.edu.cn

Xiaozhi Chen
DJI
cxz.thu@gmail.com

Hui Cheng*
Sun Yat-sen University
chengh9@mail.sysu.edu.cn

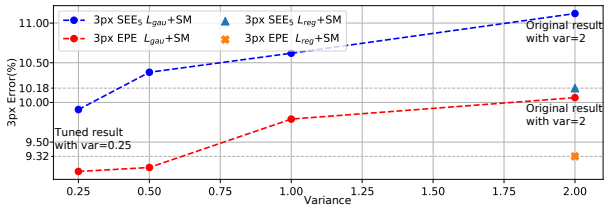

Figure 1. Performance of GCNet [2] on Sintel [1] with different variance.

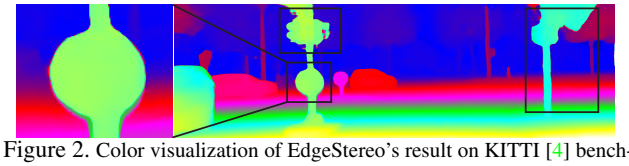

Figure 2. Color visualization of EdgeStereo's result on KITTI [4] benchmark.

1. Discussion

1.1. Small discrepancy of GCNet and PDSNet.

From the experiment results we found a small discrepancy of GCNet [2] and PDSNet [6]. As we didn't tune much the variance for these two networks in experiments, here we conduct more experiments for GCNet on Sintel [1]. Specifically, we train GCNet on Sceneflow [3] dataset with different variance parameter and benchmark then on Sintel. We found that the discrepancy can be eliminated with tuned variance. Fig.1 shows consistent improvements with tuned variance for GCNet on Sintel. Note that GCNet estimates disparity at $\frac{1}{4}$ resolution, with less loss of spatial accuracy. Therefore it has lower 3px EPE with small variance.

1.2. Benefits of edge information on our method.

As our method is based on the multi-modal observation, it doesn't rely on edge information. We leave the integration of edge cues with our method for further research. However, we observe an interesting result from EdgeStereo [5], which is a recent work that integrates edge cues in disparity estimation. As shown in Fig.2, EdgeStereo still introduces over-smoothing estimation on edge regions. Note that the disparity changes smoothly from green to red and blue on the region boundaries. Therefore, we believe addressing the multi-modal estimation is the key to the over-smoothing problem.

*Corresponding author

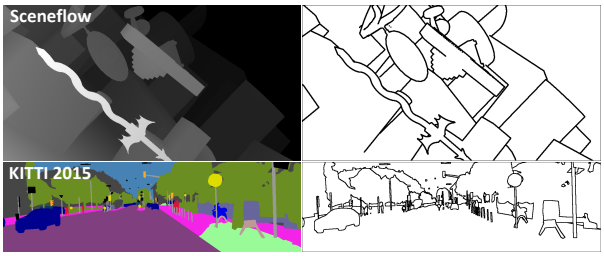

Figure 3. Edge regions extracted. Please zoom in for more details.

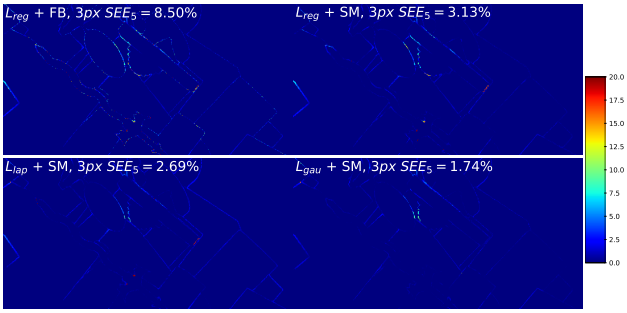

Figure 4. SEE_5 visualization on edge regions for various methods. Please zoom in for more details. The edge map is top-right subfig of Fig.3.

1.3. Robustness of Soft Edge Error.

Fig.3 and 4 present edge regions and loss visualization respectively. For datasets with dense disparity groundtruth, edges are extracted by a simple first-order difference on the disparity map, which is stable and robust. For datasets with sparse disparity groundtruth, computing disparity edges is infeasible. Therefore we use semantic boundaries instead. Although semantic boundaries are a subset of the edges, it's more stable and easier for annotation. Also, for applications like point cloud segmentation, the over-smoothing problem on inter-instance regions are more critical than inner-instance regions. Therefore focusing on semantic edges is preferred, which also benefits the robustness of the *Soft Edge Error* metric.

Region	Max Modal	Wrong Modal	No Modal
All	63.47%	20.08%	16.46%
Edge	72.79%	24.13%	3.08%

Table 1. Statistics of which modal that grountruth lies on.

2. Failure case analysis

2.1. Edge misalignment.

Tab.1 shows relation between groundtruth and modal of probability output. Ideally, groundtruth should lie on modal with max probability but it's not the case for a small portion of pixel, causing edge misalignment. Nevertheless, it still preserves sharp boundary and doesn't cause deterioration on over-smoothing problem. A more powerful backbone should alleviate this limitation and reasoning on occlusion may further help.

2.2. Distant region.

As there is a nonlinear inverse relationship between depth and disparity, minor change of disparity could cause a large change of depth. disparity. Therefore, for depth boundaries in distant region, there is only one modal on disparity and the over-smoothing depth only lead the modal to change slowly instead of the multi-modal phenomenon. We consider this limitation an inherence of stereo geometry.

References

[1] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conf. on Computer Vision (ECCV)*, pages 611–625. Springer-Verlag, 2012. 1

[2] 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 *Proceedings of the IEEE International Conference on Computer Vision*, pages 66–75, 2017. 1

[3] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4040–4048, 2016. 1

[4] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1

[5] Xiao Song, Xu Zhao, Hanwen Hu, and Liangji Fang. Edgestereo: A context integrated residual pyramid network for stereo matching. *Asian Conference on Computer Vision*, 2018. 1

[6] S. Tulyakov, A. Ivanov, and F. Fleuret. Practical Deep Stereo (PDS): Toward applications-friendly deep stereo matching. In *Proceedings of the international conference on Neural Information Processing Systems (NIPS)*, 2018. 1