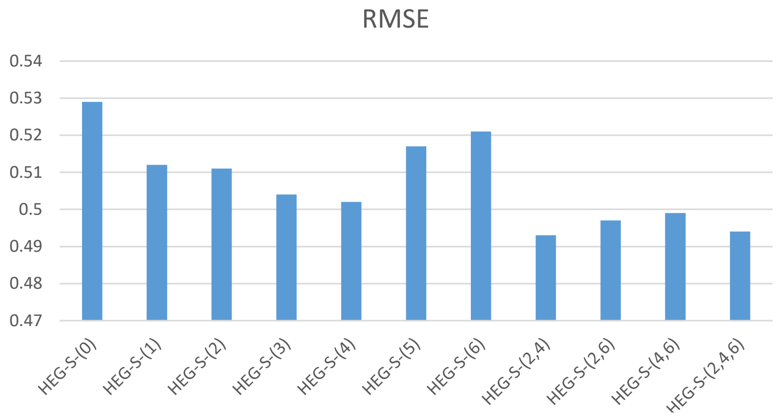
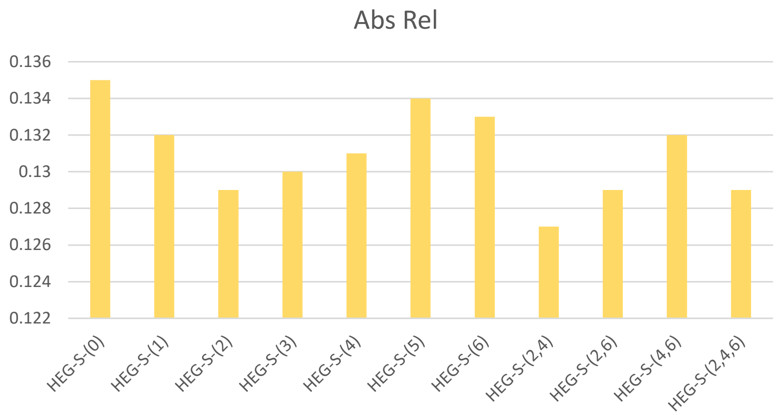
Supplementary Material for "CLIFFNet for Monocular Depth Estimation with Hierarchical Embedding Loss"

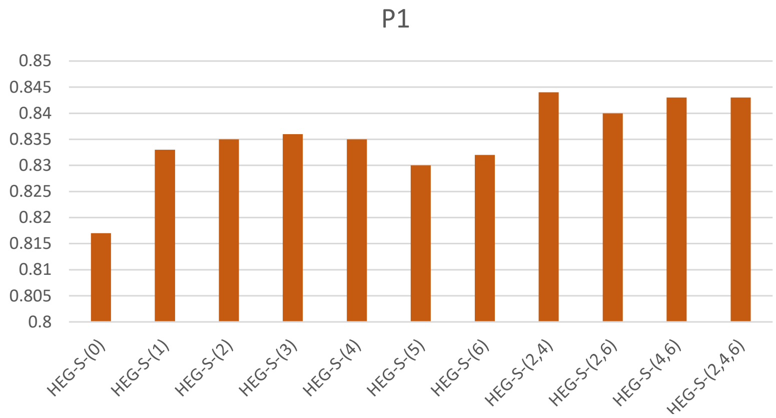
Anonymous ECCV submission

Paper ID 3365

We compare the performance of hierarchical losses computed in embeddings generated by different layers of HEG-S. Denotes HEG-S-(0,i,j,k) as the combination of L1 losses on the 0-th, i-th, j-th, and k-th embedding, where the 0-th embedding is the original depth space. The loss weights for the original depth space and any generated embedding space are firstly set to 1.0 and 10.0, respectively, to find an optimal combination of the embeddings. We then perform grid search to determine the final loss weights for each embedding in the optimal combination. The comparison results on NYU-Depth V2 are as follows.







The comparison results of our methods against Laina *et al.* [3], Zhang *et al.* [4], and Fu *et al.* [2] on Cityscapes [1] are as follows.

Method	Error		Accuracy		
	RMSE	Abs Rel	P1	P2	P3
Laina *et al.* [3]	7.273	0.257	0.765	0.893	0.940
Fu *et al.* [2]	7.103	0.235	0.779	0.905	0.945
Zhang *et al.* [4]	7.104	0.234	0.776	0.903	0.949
CLIFFNet-R	**0.703**	0.231	0.789	**0.907**	0.949
CLIFFNet-S	0.704	**0.229**	**0.790**	**0.907**	**0.950**

We present more qualitative results (depth.mp4) as well as the source code to build the feature pyramid sub-network (fpn.py) and depth prediction sub-network (dpn.py) in the supplementary material.

References

1. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213–3223 (2016)
2. Fu, H., Gong, M., Wang, C., Batmanghelich, K., Tao, D.: Deep ordinal regression network for monocular depth estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2002–2011 (2018)
3. Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: 2016 Fourth international conference on 3D vision (3DV). pp. 239–248. IEEE (2016)
4. Zhang, Z., Cui, Z., Xu, C., Jie, Z., Li, X., Yang, J.: Joint task-recursive learning for semantic segmentation and depth estimation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 235–251 (2018)