Single-Image Depth Perception in the Wild
Supplementary Material

Weifeng Chen **Zhao Fu** **Dawei Yang** **Jia Deng**
University of Michigan, Ann Arbor
{wfchen,zhaofu,ydawei,jiadeng}@umich.edu

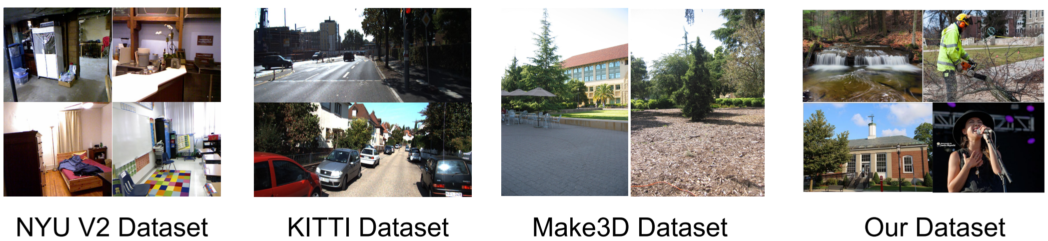

Figure 1: Additional example images from current RGB-D datasets and our Depth in the Wild (DIW) dataset.

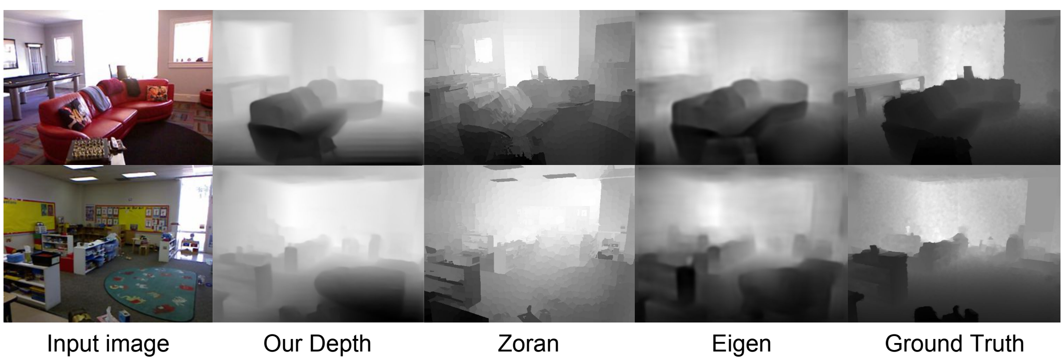

Figure 2: Additional qualitative results on NYU Depth by our method, the method of Eigen et al. [1], and the method of Zoran et al. [2]. All depth maps except ours are directly from [2].

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

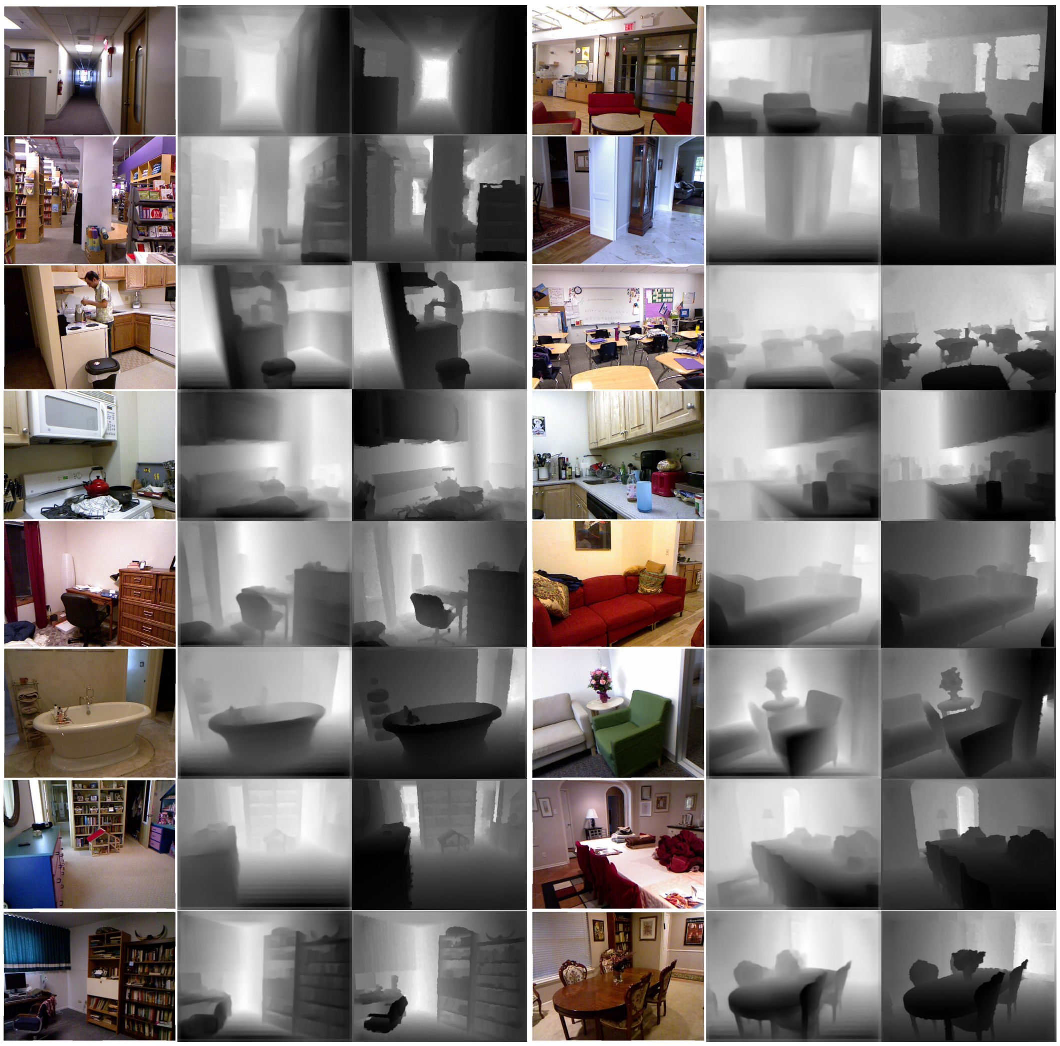

Figure 3: Additional qualitative results on NYU Depth test set by our method. Here we show the original input images and the depth maps by our method, as well as the ground truth.

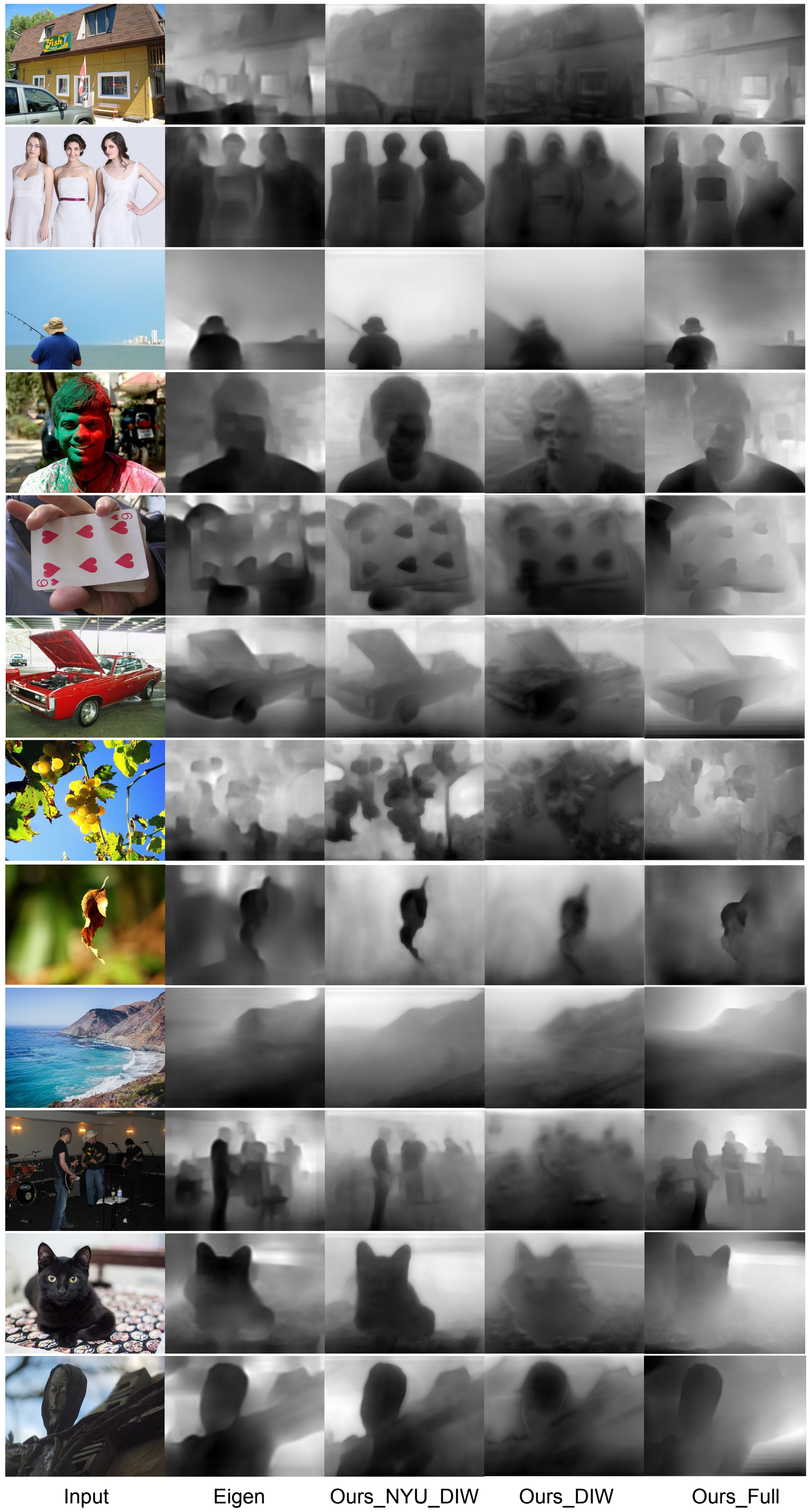

Figure 4: Additional qualitative results on our Depth in the Wild (DIW) dataset by our method and the method of Eigen et al. [1].

Method	RMSE	RMSE (log)	RMSE (s.inv)	absrel	sqrrel
Ours	0.89	0.32	0.25	0.27	0.29
Ours_Full	0.74	0.26	0.21	0.21	0.19
Eigen(V) [1]	**0.64**	**0.21**	**0.17**	**0.16**	**0.12**

Table 1: Table 2 of the main paper reports the metric error of our network trained on relative depth pairs. Here we provide additional results by training our network on the full depth map. The network **Ours** is our network trained on the 795 NYU Depth training subset, and **Ours_Full** is our network trained on the full NYU Depth training set.

Method	WKDR	$WKDR^{=}$	$WKDR^{\neq}$
rand_48K	**34.3%**	**31.7%**	**37.1%**
rand_24K	34.5%	32.6%	36.9%
rand_12K	34.9%	32.4%	37.6%
rand_6K	36.1%	32.2%	39.9%
rand_3K	35.8%	28.7%	41.3%

Table 2: Table 2 of the main paper reports the performance of our network versus the number of randomly sampled non-superpixel point pairs on NYU Depth. Here we report additional results by sampling more pairs. **rand_N** denotes a network trained with N pairs per image.

#Depth Pairs	WKDR	$WKDR^{=}$	$WKDR^{\neq}$
800	**35.6%**	**36.1%**	**36.5%**
500	37.2%	37.7%	38.2%
250	38.0%	37.4%	39.7%
100	42.3%	41.1%	44.0%

Table 3: Table 2 of the main paper reports the performance of our network trained on 800 superpixel point pairs. Here we report additional results by decreasing the number of point pairs.

References

[1] D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” in *ICCV*, 2015.

[2] D. Zoran, P. Isola, D. Krishnan, and W. T. Freeman, “Learning ordinal relationships for mid-level vision,” in *ICCV*, 2015.

[3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.