Supplementary Material for MODE: Multi-view Omnidirectional Depth Estimation with 360° Cameras

Ming Li, Xueqian Jin, Xuejiao Hu, Jingzhao Dai, Sidan Du, and Yang Li

Nanjing University, Nanjing, China
{mingli,jcboxq,hxj,dg20230007}@smail.nju.edu.cn,
{coff128,yogo}@nju.edu.cn

1 Implementation Details

We implement the MODE framework with PyTorch. For the omnidirectional stereo matching stage of our framework, we first train the network for 45 epochs with a learning rate of 0.001, and then decay the learning rate to 0.0001 to train the model for 10 epochs. For the depth map fusion stage, we first train the network for 150 epochs with a learning rate of 0.0001, and then fine-tune the fusion network for 20 epochs on the soiled version of Deep360. For stage 1, GPU memory requirement is 54 GB and training time is 130h. For stage 2, GPU memory requirement is 10 GB and training time is 12h.

We use one of the official dataset splits of 3D60[2] that contains 7858 frames for training, 1103 for validation, and 2189 for test in experiments. All the SOTA 360° depth estimation methods are fine-tuned based on this dataset split for comparison.

We use Insta360 ONE X2 cameras to build the multi-view 360° camera system in the real-world environment. The 360° cameras are calibrated with the toolbox Kalibr, in particular [1].

2 More Results of Experiments

We show the qualitative results of omnidirectional stereo matching in Fig. 1. Since the spherical disparity is defined in Cassini projection domain, we present the disparity maps in Cassini projection.

More qualitative comparisons of SOTA 360° depth estimation methods are shown in Fig. 2 and 3. Additional test results of our framework in the real-world environment are presented in the supplementary video at `https://youtu.be/Fw-KR35UWgQ`.

Ming Li and Xueqian Jin contributed equally to this work.
Corresponding authors: Sidan Du, Yang Li

3 Details of Proposed Datasets

The generation details of our proposed 360° dataset Deep360 are presented in the supplementary video as well. More examples of Deep360 are shown in Fig. 4 and 5.

References

1. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE transactions on pattern analysis and machine intelligence **28**(8), 1335–1340 (2006)
2. Zioulis, N., Karakottas, A., Zarpalas, D., Alvarez, F., Daras, P.: Spherical view synthesis for self-supervised 360° depth estimation. In: 2019 International Conference on 3D Vision (3DV). pp. 690–699 (2019). https://doi.org/10.1109/3DV.2019.00081

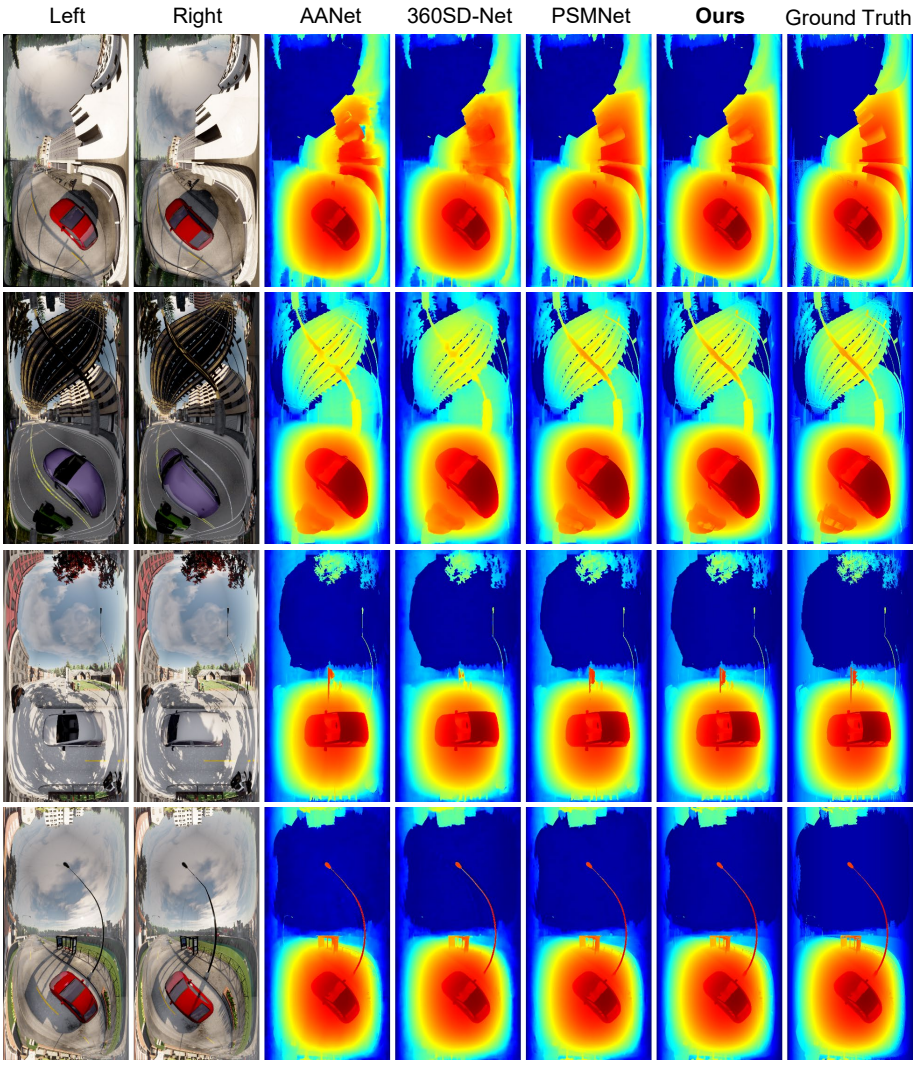

Fig. 1. Qualitative comparisons of omnidirectional stereo matching on Deep360

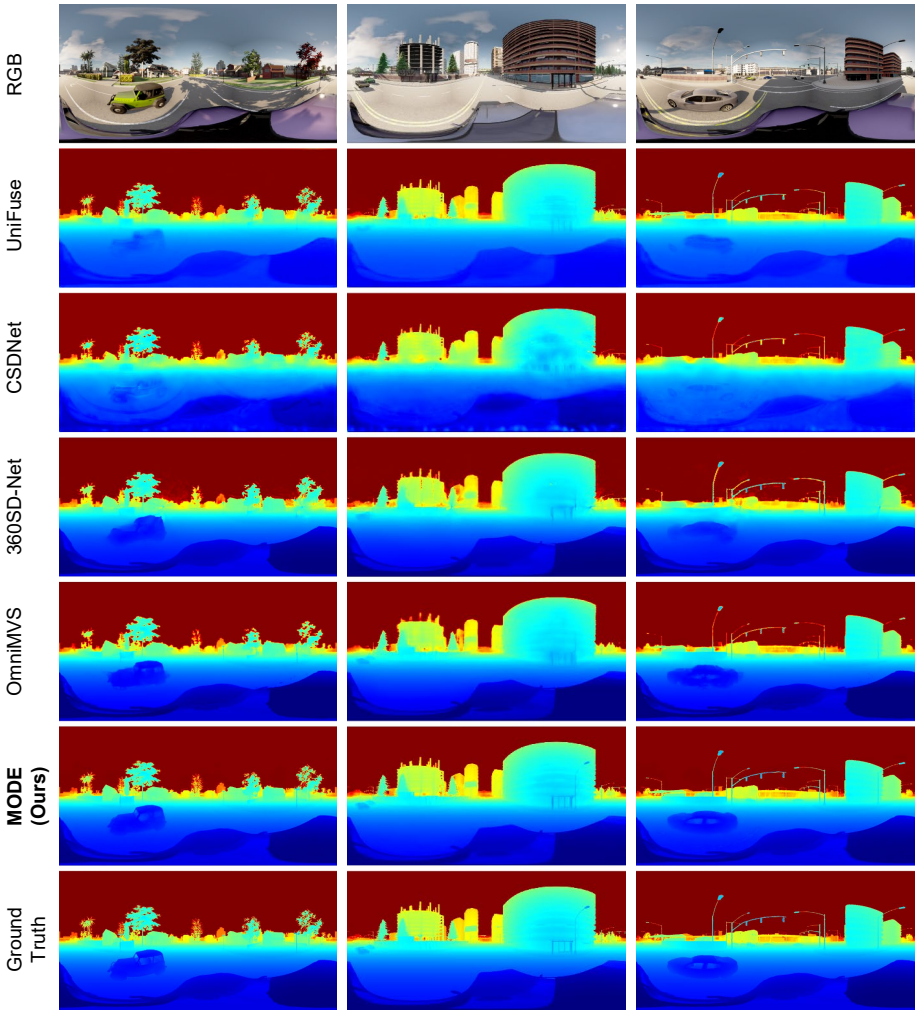

Fig. 2. Qualitative comparisons of omnidirectional depth estimation methods on Deep360

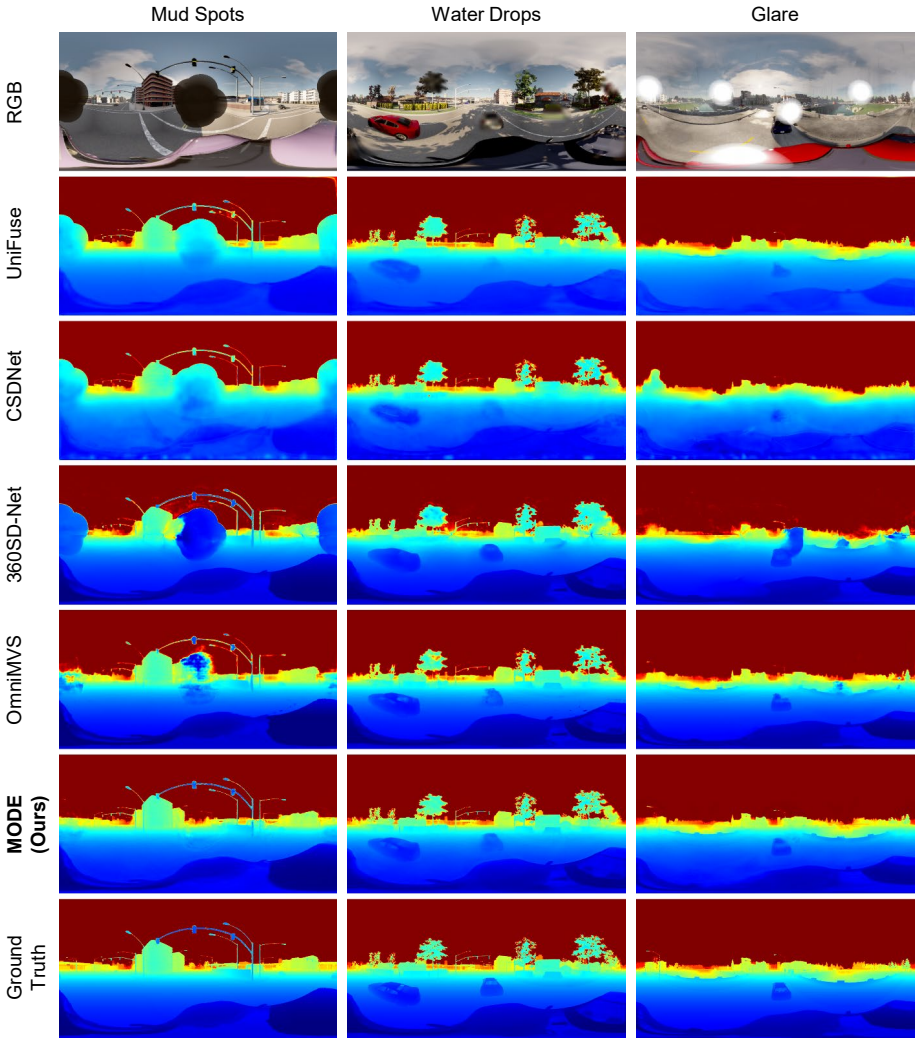

Fig. 3. Qualitative comparisons of omnidirectional depth estimation methods on Deep360 (Soiled)

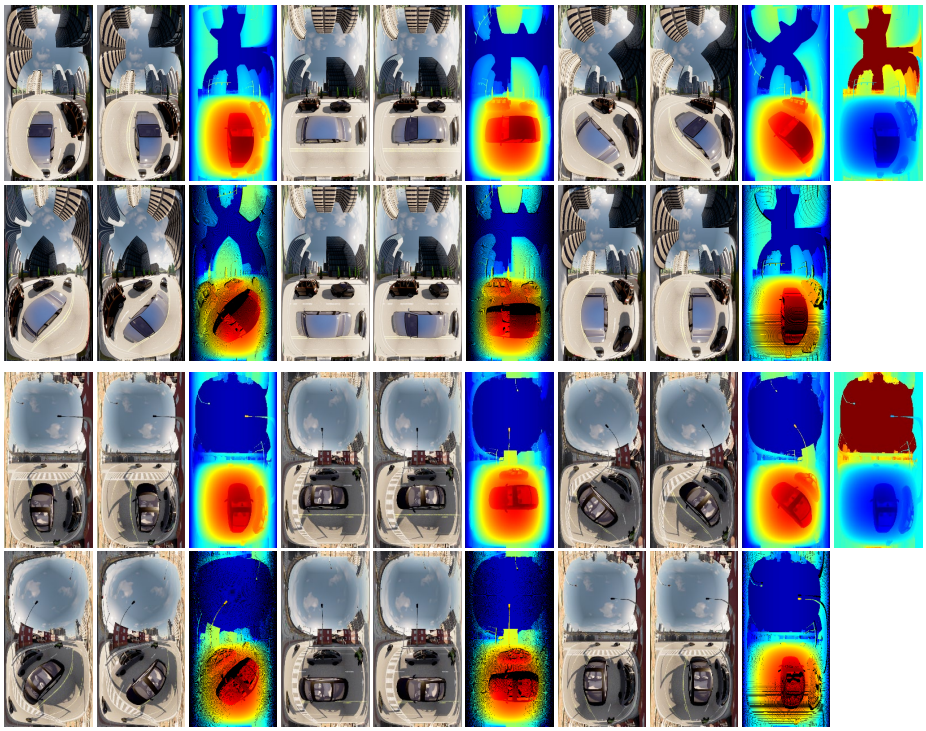

Fig. 4. Deep360 From left: rectified panoramas, disparity maps, and ground truth depth map

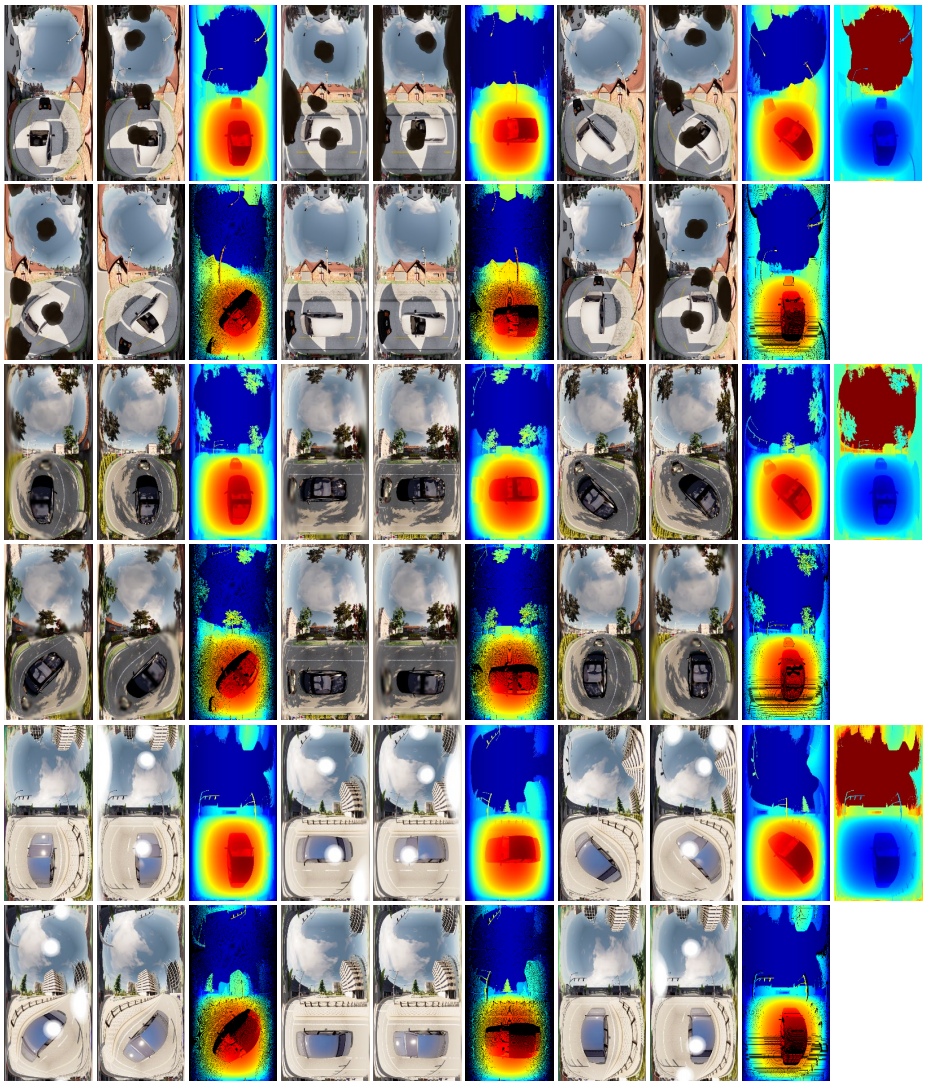

Fig. 5. Deep360 (Soiled) From left: soiled panoramas, disparity maps, and ground truth depth map. From top: mud spots, water drops, and glare