

DENS3R: A FOUNDATION MODEL FOR 3D GEOMETRY PREDICTION

SUPPLEMENTARY MATERIALS

Xianze Fang^{1,*} Jingnan Gao^{2,*} Zhe Wang¹ Zhuo Chen² Xingyu Ren² Jiangjing Lv^{1,†}
Qiaomu Ren¹ Zhonglei Yang¹ Xiaokang Yang² Yichao Yan^{2,‡} Chengfei Lv¹

¹Alibaba Group. ²Shanghai Jiao Tong University.

1 ADDITIONAL VISUALIZATION

We provide additional unified geometric prediction results of monocular inputs in Fig. 1 and 2-view inputs in Fig. 2. It can be seen that our model achieves robust and high-quality unified geometric predictions across several scenarios.

2 IMAGE MATCHING VISUALIZATION

We demonstrate our dense and accurate matching results of several challenging cases in Fig. 3. We visualize the matching results from MegaDepth-1500 dataset, ScanNet-1500 dataset and Aachen dataset. It can be seen that our method can handle image-matching for **1)** inputs with different lighting conditions, **2)** inputs taken from different views with large-angle difference, **3)** inputs with small overlapping regions. The accuracy of our method is shown on the upper left of each matching image-pair.

3 MULTI-VIEW RECONSTRUCTION

Apart from multi-resolution image inputs, our model also design a simple yet effective post-processing pipeline and supports multi-view inputs. The results are demonstrate in Fig. 4. It can be seen that our method enables high-quality 3D reconstruction even without known camera poses. In addition, the predictions of Dens3R can be used to initialize robust and accurate 3D reconstructions by integrating the predicted attributes with modern structure-from-motion (SfM) pipelines such as Glomap.

*: Equal Contribution. †: Project Leader. ‡: Corresponding Author.

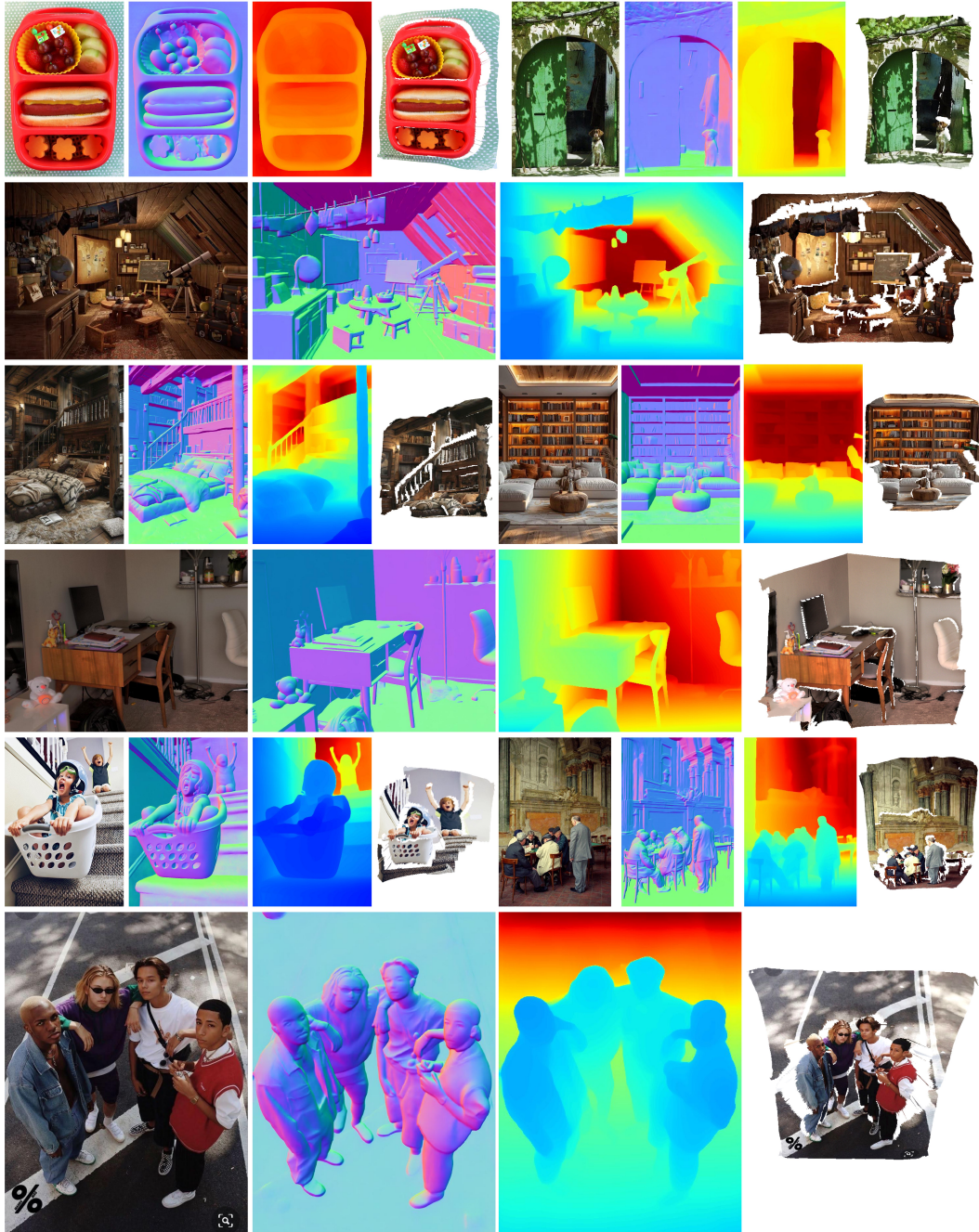


Figure 1: Additional visual results for unified geometric predictions of monocular inputs by our methods

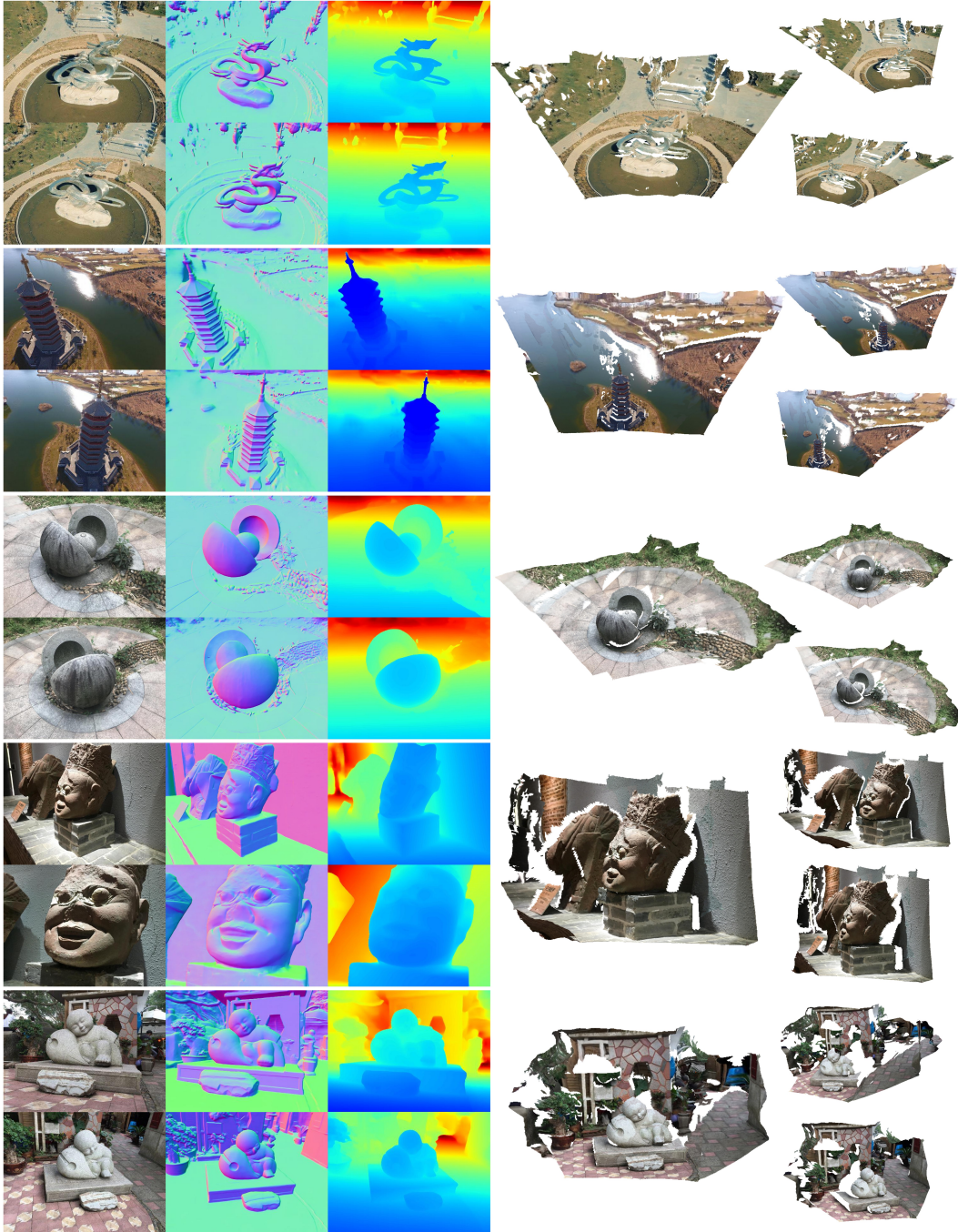
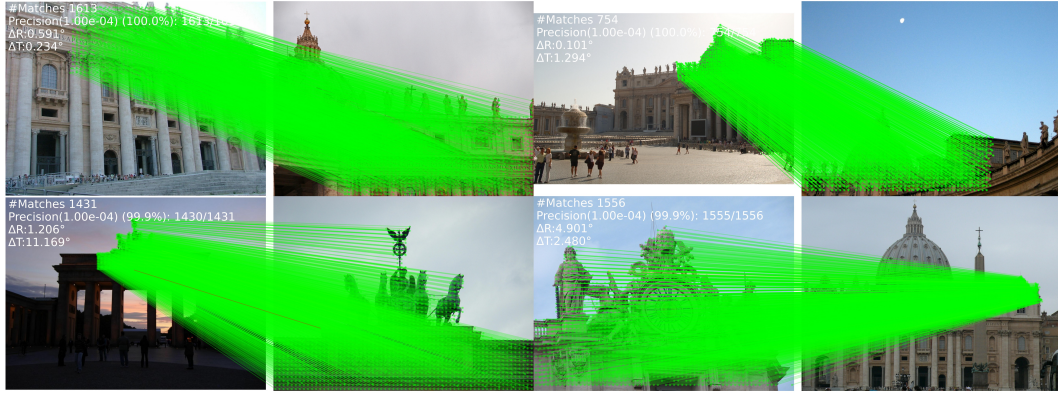
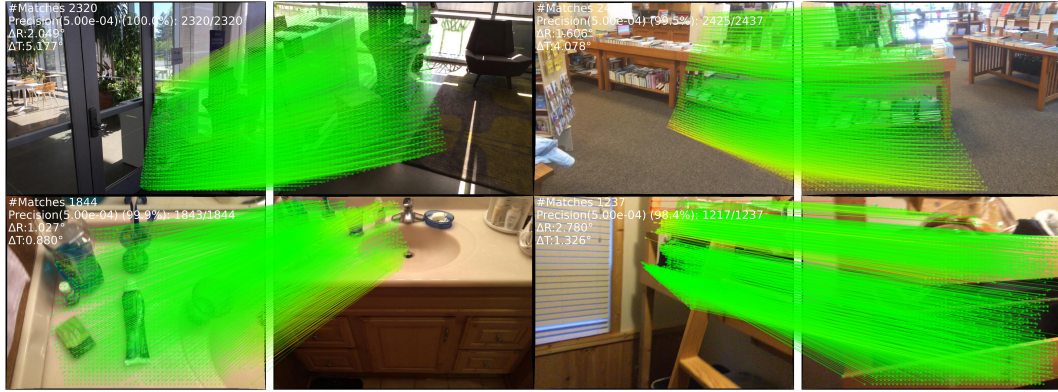


Figure 2: Additional visual results for unified geometric predictions of 2-view images inputs by our method.



a) MegaDepth-1500



b) ScanNet-1500



c) Aachen

Figure 3: Image-matching visualization. We provide the visualization of our dense and accurate image-matching results. The accuracy is also presented on the upper left of each matching image-pair.

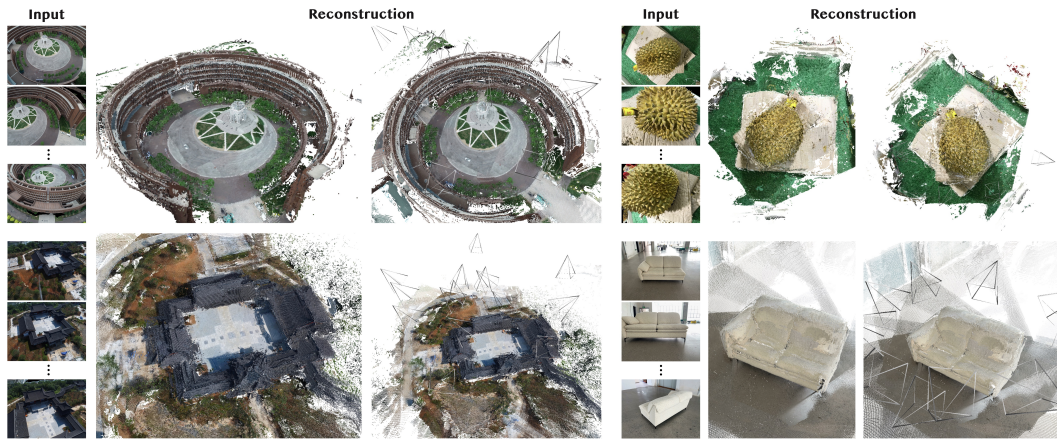


Figure 4: Multi-view reconstruction results. We demonstrate high-quality 3D reconstruction for various scenarios without known camera poses. The predicted camera poses are also shown on the right of the reconstruction results.