Supplementary Material
NDDepth: Normal-Distance Assisted Monocular Depth Estimation

Shuwei Shao[1,2] Zhongcai Pei[2,1] Weihai Chen[2,1*] Xingming Wu[2,1] Zhengguo Li[3]
[1]School of Automation Science and Electrical Engineering, Beihang University, Beijing, China.
[2]Hangzhou Innovation Institute, Beihang University, Hangzhou, Zhejiang, China
[3] SRO department, Institute for Infocomm Research, A*STAR, Singapore
{swshao, peizc, whchen}@buaa.edu.cn ezgli@i2r.a-star.edu.sg

1. Evaluation Metrics

Similar to [3, 2, 1], we adopt the standard evaluation metrics in our experiments, which are described in detail as follows:

- Square Root of the Scale Invariant Logarithmic Error (**SILog**):
  $$\sqrt{\frac{1}{|\mathbf{T}|}\sum_{\mathbf{g}\in\mathbf{T}}(\mathbf{g})^2-\frac{1}{|\mathbf{T}|^2}\left(\sum_{\mathbf{g}\in\mathbf{T}}\mathbf{g}\right)^2};$$
  where $\mathbf{T}$ stands for the set of pixels with valid values and $\mathbf{g}=\log\mathbf{D}-\log\mathbf{D}^{gt}$;
- Relative Absolute Error (**Abs Rel**):
  $\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left|\mathbf{D}-\mathbf{D}^{gt}\right|/\mathbf{D}^{gt}$;
- Relative Squared Error (**Sq Rel, Eigen Split**):
  $\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left(\mathbf{D}-\mathbf{D}^{gt}\right)^2/\mathbf{D}^{gt}$;
- Relative Squared Error (**Sq Rel, Official Split**):
  $\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left(\mathbf{D}-\mathbf{D}^{gt}\right)^2/\mathbf{D}^{gt^2}$;
- Root Mean Squared Error (**RMSE**):
  $\sqrt{\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left(\mathbf{D}-\mathbf{D}^{gt}\right)^2}$;
- Root Mean Squared Logarithmic Error (**RMSE log**):
  $\sqrt{\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left(\log\mathbf{D}-\log\mathbf{D}^{gt}\right)^2}$;
- Inverse Root Mean Squared Error (**iRMSE**):
  $\sqrt{\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left(1/\mathbf{D}-\mathbf{1/D^{gt}}\right)^2}$;
- $\log_{10}$:
  $\frac{1}{|\mathbf{T}|}\sum_{\mathbf{D}\in\mathbf{T}}\left(\log_{10}\mathbf{D}-\log_{10}\mathbf{D}^{gt}\right)^2$;
- Threshold Accuracy ($\delta <$ **thr**):
  $\%$ of $\mathbf{D}$ satisfies $\left(\max\left(\frac{\mathbf{D}}{\mathbf{D}^{gt}},\frac{\mathbf{D}^{gt}}{\mathbf{D}}\right)=\delta<thr\right)$ for $thr=1.25,1.25^2,1.25^3$.

2. More Qualitative Depth and Point Cloud Results

To perform more comparisons against previous state-of-the-art competitors, we display qualitative depth results of NewCRFs [3] and our method on the official split of KITTI dataset, as shown in Figure 1. As can be seen, the proposed method delinates more accurate depth estimates, especially in difficult regions, for example windows in the fourth column. For better evaluation of depth estimates from the 3D shape, we convert depth maps into point clouds and show more qualitative point cloud results on the KITTI dataset and NYU-Depth-v2 in Figure 2 and Figure 3, respectively. It can be seen that the proposed method preserves prominent geometric features, for example, roads in outdoor scenario and floors in indoor scenario, and is capable of recovering the 3D world reasonably.

References

[1] Shubhra Aich, Jean Marie Uwabeza Vianney, Md Amirul Islam, and Mannat Kaur Bingbing Liu. Bidirectional attention network for monocular depth estimation. In *2021 IEEE International Conference on Robotics and Automation*, pages 11746–11752. IEEE, 2021.

[2] Vaishakh Patil, Christos Sakaridis, Alexander Liniger, and Luc Van Gool. P3depth: Monocular depth estimation with a piecewise planarity prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1610–1621, 2022.

[3] Weihao Yuan, Xiaodong Gu, Zuozhuo Dai, Siyu Zhu, and Ping Tan. Neural window fully-connected crfs for monocular depth estimation. In *Proceedings of the IEEE Conference on*

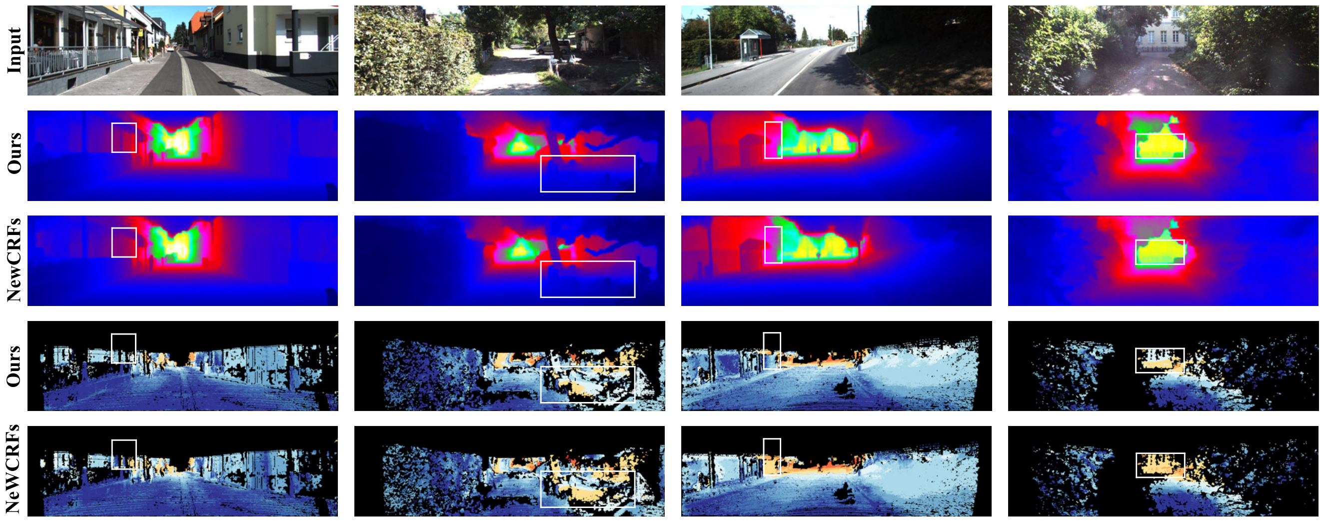

Figure 1. **Qualitative depth results on the official split of KITTI dataset**. The white boxes highlight the regions to emphasize. The second and third rows are depth maps. The fourth and fifth rows are error maps.

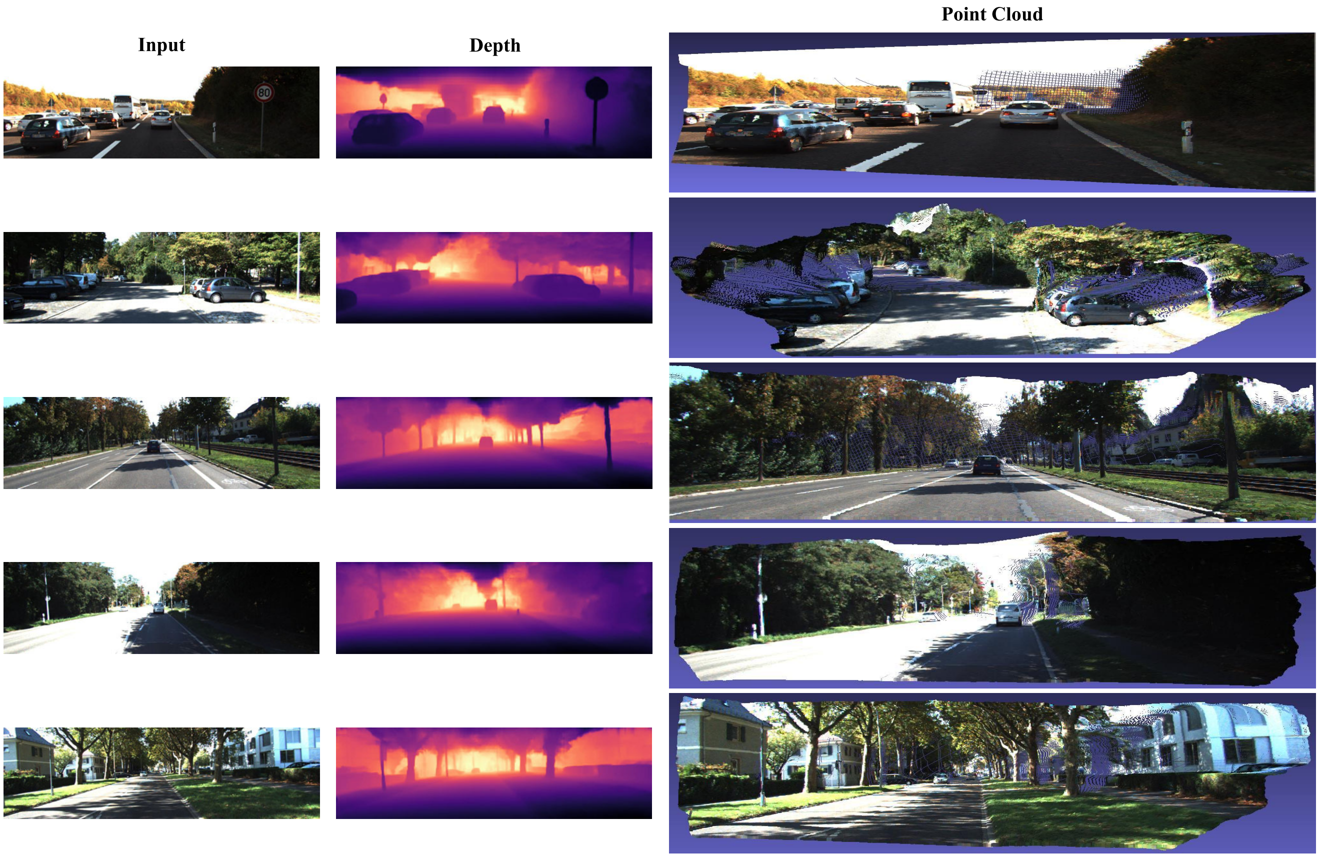

Figure 2. **Qualitative depth and point cloud results on the KITTI dataset**.

Computer Vision and Pattern Recognition, pages 3916–3925, June 2022.

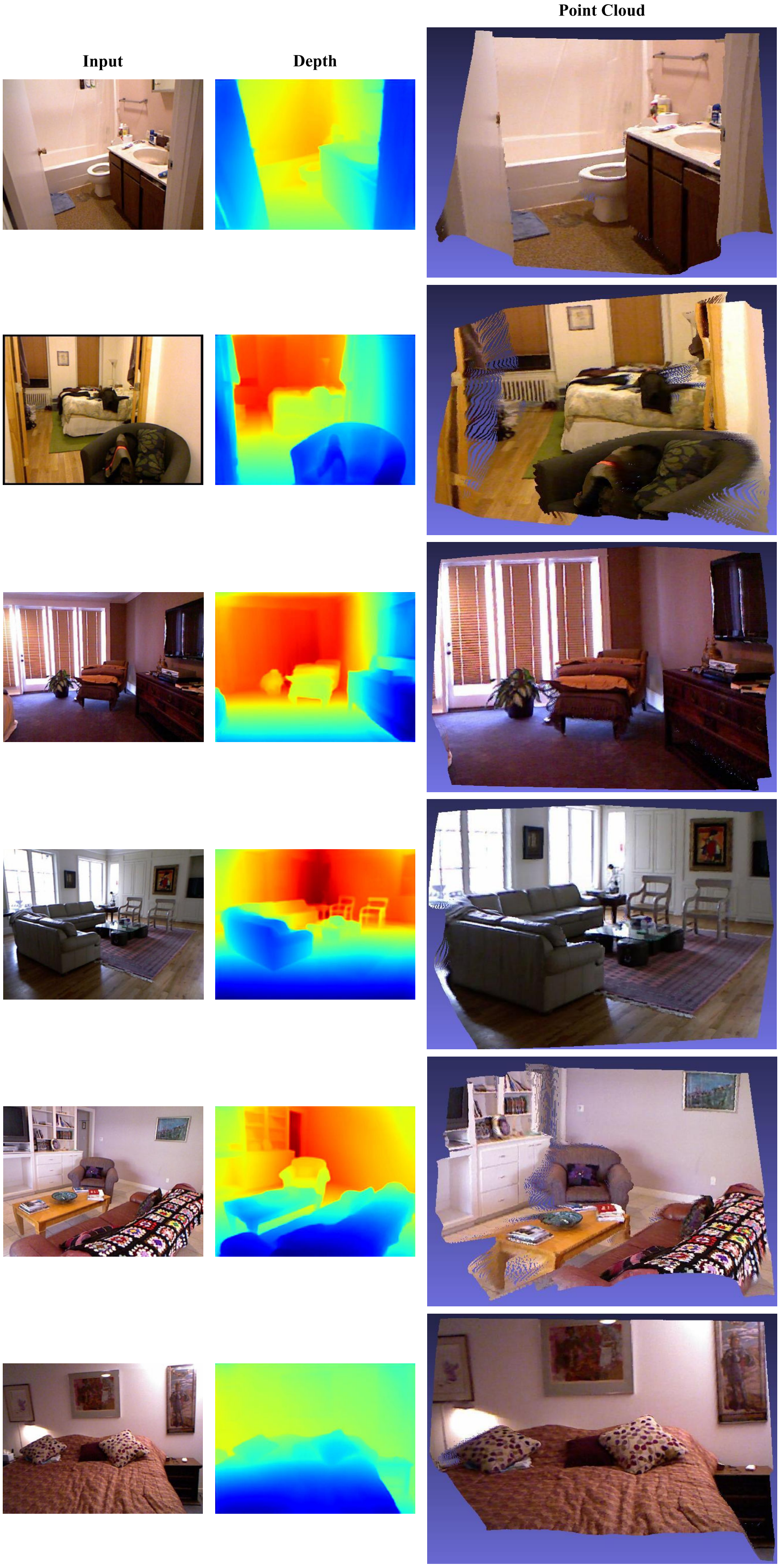

Figure 3. **Qualitative depth and point cloud results on the NYU-Depth-v2 dataset**.