6. Supplementary Materials

6.1. Additional Qualitative Results

We show qualitative comparisons of HoHoNet [6]'s results versua our method using HoHoNet to generate the reference panoramic depth maps in Figure 6 and Figure 7. Same as in other cases (*e.g*. SliceNet and UniFuse), we can see that our method significantly improved the levels of details of the respective panorama-based method.

We show additional qualitative results in Figure 8 and Figure 9. In the first example of Replica360 4K result in Figure 9, we can see that the result of [23] has incorrect relative depths of the white wall w.r.t. the two adjacent walls.

6.2. Quantitative metrics in 360MonoDepth [23]

In [23], they reported different quantitative scores on the Matterport 2K dataset and the Replica360 2K/4K datasets then our experiments. We think the differences came from the different approaches to align an estimated depth map and a ground-truth depth map. In our paper, we used the "median-scaling" approach that is the common practice in all panorama-based methods since OmniDepth [47]. After checking [23]'s paper and code, we think they instead used the least squares-based scaling approach (proposed in the original MiDaS v3 paper [22]), in which both optimal scaling and offset are computed by least-squares to align an an estimated depth map and a ground-truth one. We note that some of their quantitative scores are very different from the ones reported in other papers. For example, they reported RMSE of BiFuse as 0.994 on the Matterport 2K dataset, but it is 0.6259 on the Matterport 1K dataset reported in HoHoNet [31], SliceNet [19], UniFuse [11], and OmniFusion [13]. Our evaluation method reported RMSE of BiFuse as 0.6350 on the Matterport 2K dataset, which is much closer.

6.3. Comparison with OmniFusion [13]

Recently, OmniFusion [13] reported results with new state-of-the-arts quantitative scores among panorama-based methods. At submission, we could not get to run their codes correctly on our machines, so quantitative evaluations on the new Matterport 2K and Replica360 2K/4K datasets are not reported. Comparing to our method, OmniFusion outputs 1K instead of 2K or higher resolution outputs. We also don't find the same levels of details in their results (shown in their paper) as in stitching-based methods ([23] and ours). We expect our method to benefit from the improved accuracy of panoramic depth maps generated by OmniFusion.

6.4. Quantitative evaluation of sharp detail preservation

We also measured several Laplacian-based metrics to quantitatively evaluate sharp details preservation. The results (Table 4) showed that our method significantly outperformed both panoramic methods and 360MonoDepth [23]. These quantitative results are in line with the qualitative comparisons shown in the main paper - both showed that our results outperform other panorama-based methods and 360MonoDepth (another stitching-based method) in terms of sharp detail preservation.

DS	Method	$\|\nabla^2\| \downarrow$	$\|LoG\| \downarrow$
Matterport 2K	Bifuse	0.0450	0.2134
	SliceNet	0.0386	0.1827
	UniFuse	0.0402	0.1843
	360MonoDepth	0.0384	0.1681
	Our (SliceNet)	0.0287	0.1269
	Our (UniFuse)	0.0281	0.1240
Replica360 2K	Bifuse	0.00093	0.0202
	SliceNet	0.00397	0.0811
	UniFuse	0.00134	0.0270
	360MonoDepth	0.00064	0.0122
	Our (SliceNet)	0.00049	0.0107
	Our (UniFuse)	0.00051	0.0110
Replica360 4K	Bifuse	0.00061	0.0132
	SliceNet	0.00446	0.0918
	UniFuse	0.00049	0.0103
	360MonoDepth	0.00038	0.0075
	Our (SliceNet)	0.00024	0.0053
	Our (UniFuse)	0.00027	0.0058

Highlighting: Best, second-best, third-best

Table 4. $\|\nabla^2\|$ and $\|LoG\|$ measure the mean absolute errors of Laplacian and Laplacian of Gaussian (LoG) using the standard 5x5 mask, which are proxies to measure how sharp features of estimated and ground-truth depth maps match.

6.5. Using super-resolution approaches to upsample panoramic depth maps instead of bilinear filtering

We used a recent super-resolution method [16] to upsample 1K outputs of UniFuse and SliceNet to 2K (Matterport3D) and 4K (Replica3604K). The results are very similar to bilinear up-sampling results qualitatively and quantitatively ($< 1\%$ differences by RMSE, MAE, and AbsRel).

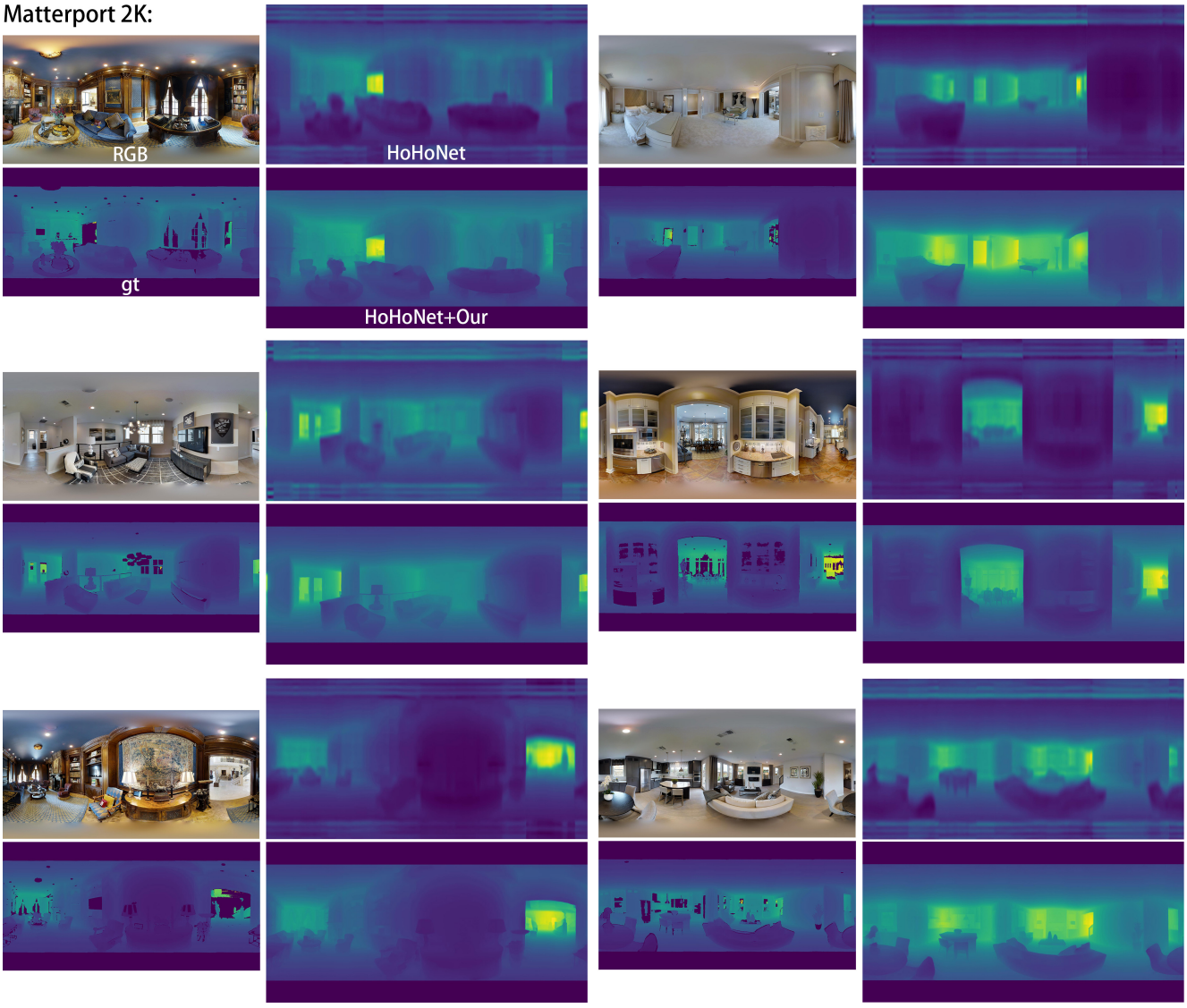

Figure 6. . Qualitative comparisons on the Matterport 2K dataset of HoHoNet and our method using HoHoNet to generate the reference panorama depth maps.

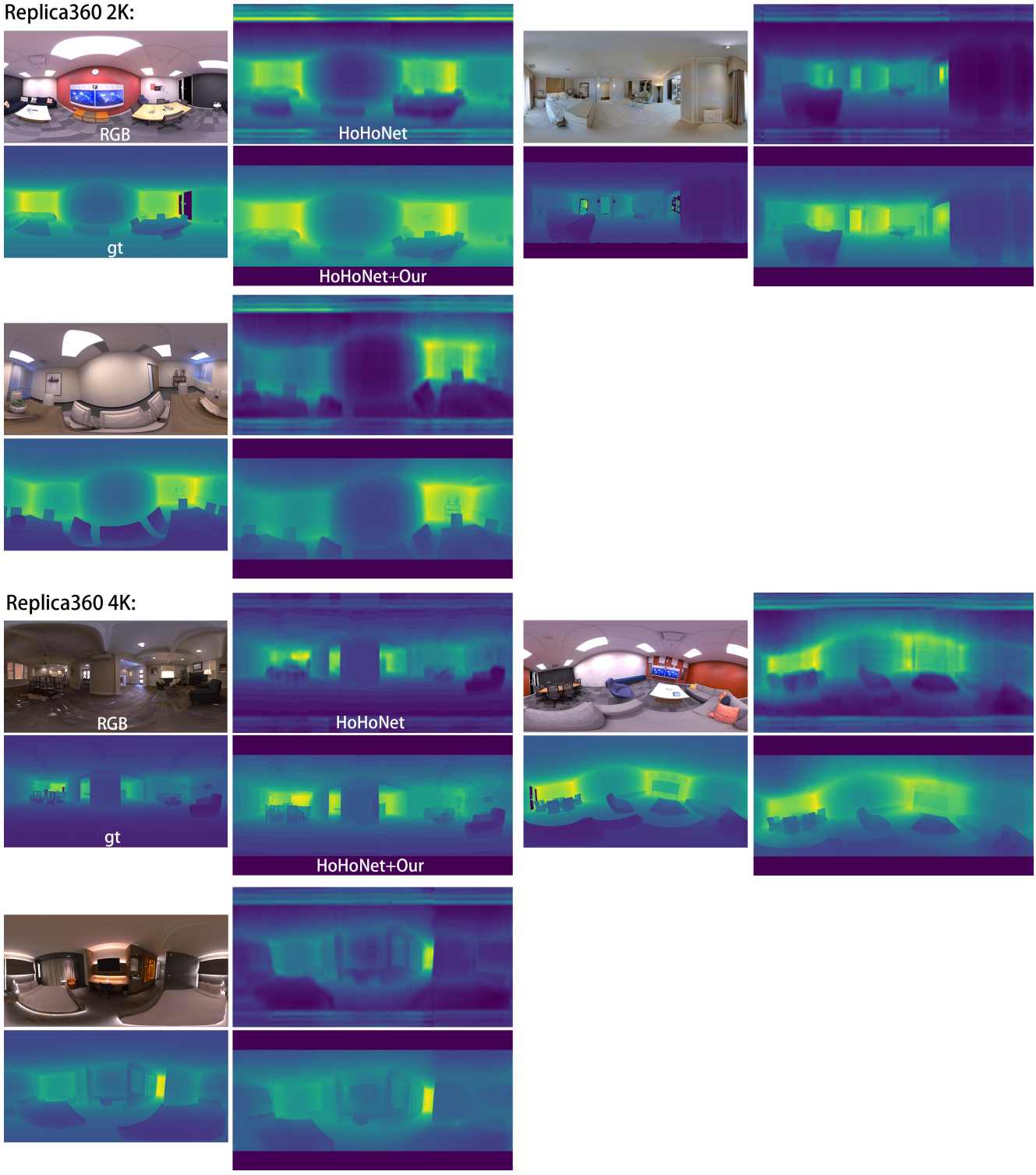

Figure 7. . Qualitative comparisons on the Replica360 2K and 4K dataset of HoHoNet and our method using HoHoNet to generate the reference panorama depth maps.

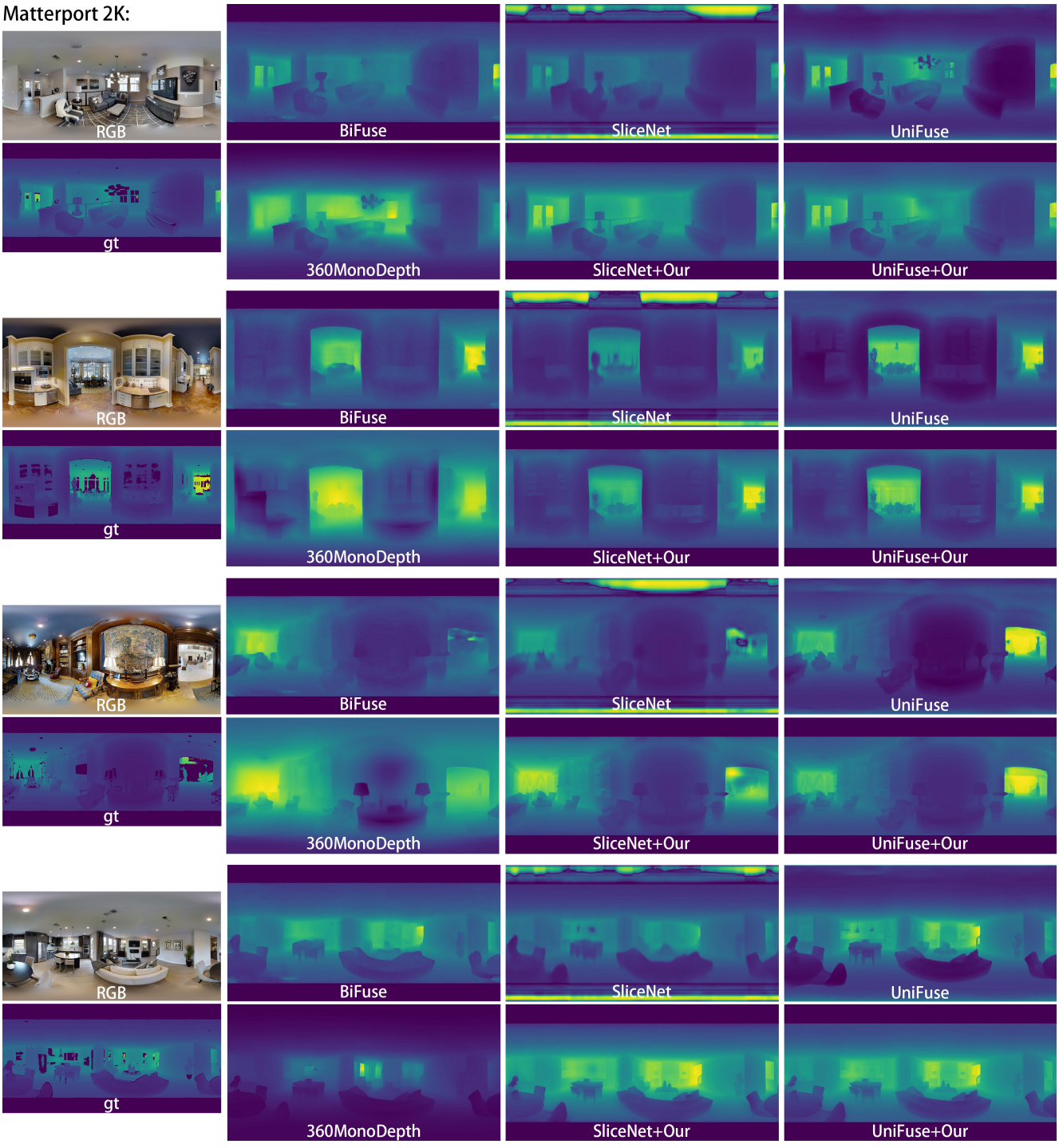

Figure 8. . Additional qualitative comparisons on the Matterport 2K dataset.

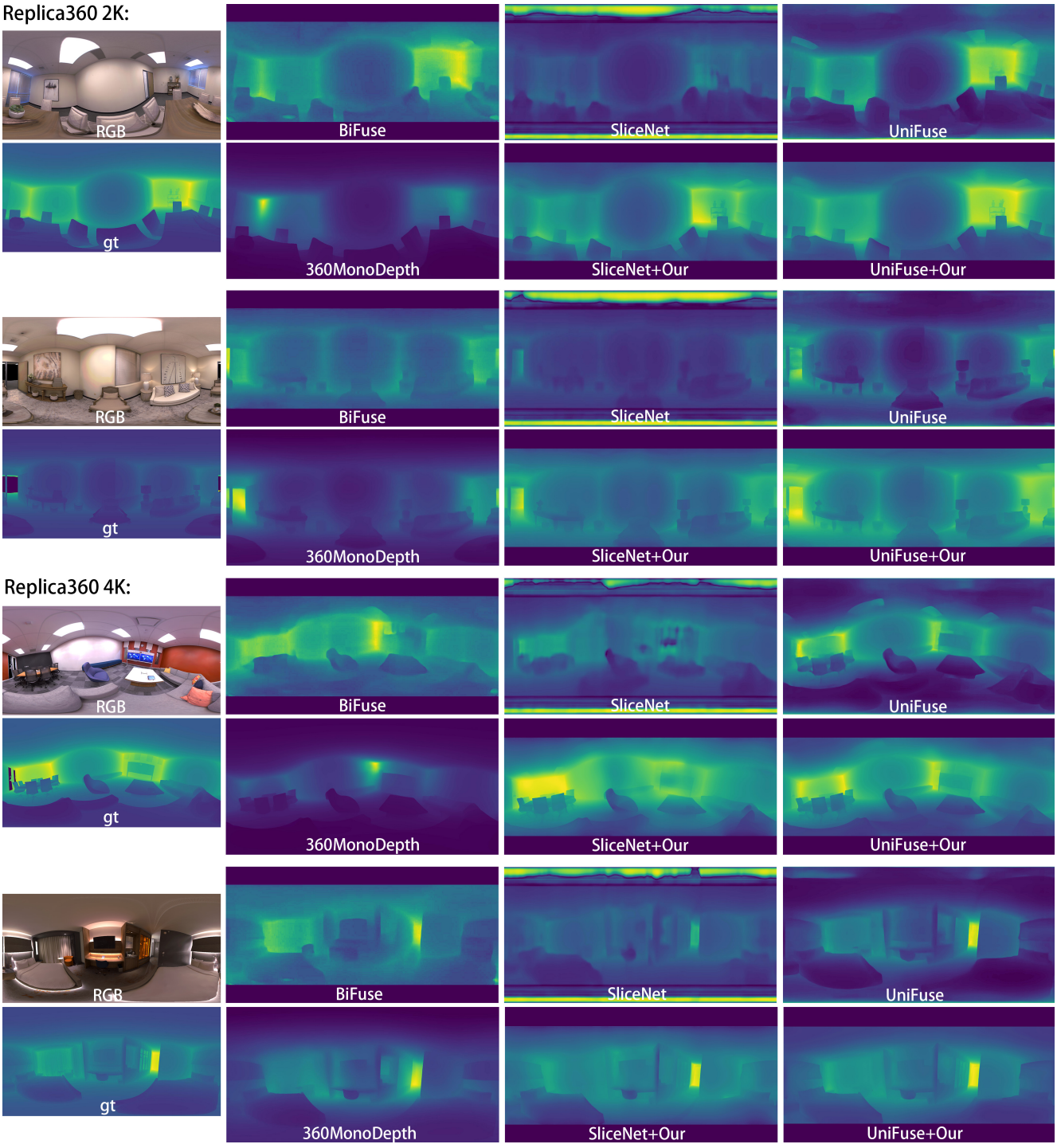

Figure 9. . Additional qualitative comparisons on the Replica 2K and 4K datasets.