Mirror3D: Depth Refinement for Mirror Surfaces
Supplemental Materials

Jiaqi Tan, Weijie Lin, Angel X. Chang, Manolis Savva
Simon Fraser University
https://3dlg-hcvc.github.io/mirror3d/

A. Mirror3D Dataset Statistics

Here, we provide additional summary statistics for the Mirror3D dataset that we constructed based on RGBD frames from NYUv2 [5], Matterport3D [1], and ScanNet [2]. We also provide several examples of annotated 3D mirror planes to illustrate the diversity of scenarios in which they occur.

Mirror in-frame location distribution. Figure 1 plots the distribution of mirror mask centroids within the image plane per source RGBD dataset and also overall for the entirety of Mirror3D. Overall, mirror centroid points tend to cluster around the upper part of the image. This is expected as mirrors are usually placed to be within human eyesight. We do note clear differences in the distributions for different datasets, which combine to form an overall distribution with fairly dense coverage of the image frame.

Mirror depth distribution. The distance from the camera of the mirror surface is an important characteristic influencing the reliability of the surrounding depth points, and correspondingly the challenge of performing depth estimation or completion on images including mirrors. Figure 2 plots histograms of the maximum mirror depth value for each mirror instance in our dataset, again broken down by source RGBD dataset and overall. We show histograms for both the raw original depth values and the corrected depth values after annotation of the 3D mirror planes. We note that the corrected depth value distributions are tighter, with fewer outliers at larger depth values. This indicates that the annotation mitigates many of the 'floating depth noise' artifacts in the raw data.

Mirror normal distribution. In Figure 3 we plot the distribution of mirror 3D plane normal direction relative to the camera viewing direction. Most mirror normals are clustered around facing the camera center and have relatively small angle deviations from that orientation. There are again interesting differences between the source dataset distributions that we attribute to the use of tripod-based equipment with fixed tilt angles (Matterport3D) vs the use of handheld scanning sensors (NYUv2 and ScanNet).

Mirror pixel ratio distribution. We define the *mirror ratio* as the fraction of image pixels that belong to a mirror instance. In Figure 4, we plot the mirror ratio distribution across source RGBD datasets, and for the overall Mirror3D dataset. We note that there is a relatively broad range of mirror ratios, though few images have ratios higher than 0.5. This is understandable, as that would correspond to scenarios where the mirror takes up most of the image and would thus show the equipment or the operator.

Example annotations. In Figures 5 to 7 we provide several examples of mirror mask and 3D mirror plane annotations from our Mirror3D dataset across the three source RGBD datasets. We show image pairs, with one image showing the mirror masks on the RGB frame, and the other visualizing the 3D mirror plane and depth points 'behind' the mirror plane in each point cloud. Mirrors occur in a variety of scenarios, and the outlier depth points cause many 'floating artifacts' that our mirror plane annotations allow us mitigate.

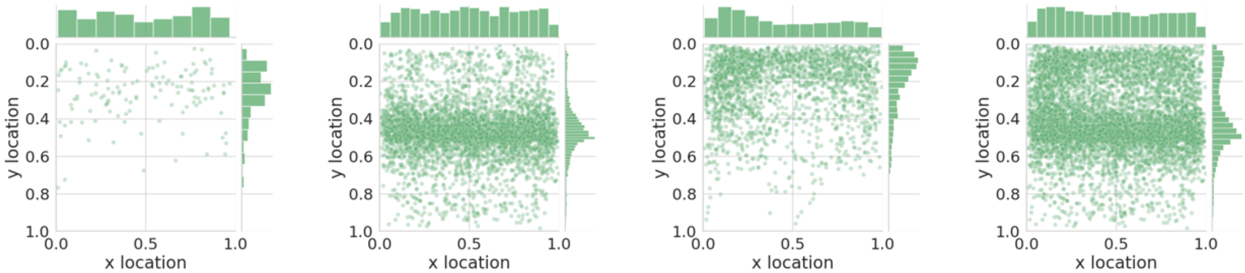
Coarse vs detailed Masks. During our annotation process, we annotate both a 'coarse' and a 'detailed' mask for each mirror plane (see Figure 8). The 'coarse' mask serves as an indication of the entire mirror surface (including occluded areas), while the 'detailed' mask provides a pixel-accurate boundary of the visible mirror surface in each image.

B. Ablations

Here, we present ablations for two aspects of our approach: the choice of mirror anchor normal count and the mirror border threshold.

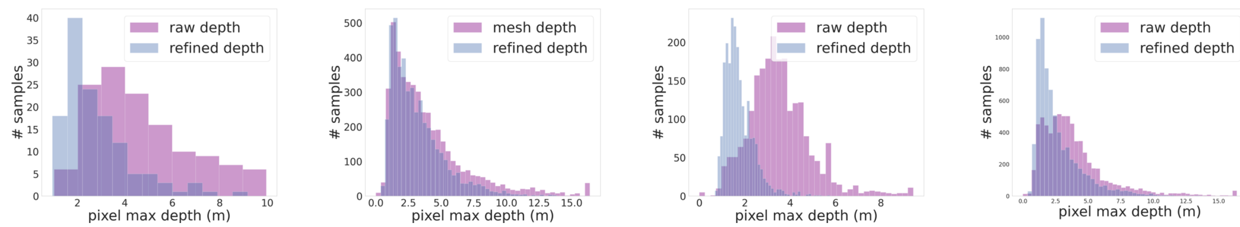
Impact of anchor normal count. We evaluate performance on a spectrum of anchor normal counts (i.e. different values of k for the k-means clustering that we use to select anchor normals in the training set). Table 1 shows that using 10 anchor normals achieves the best performance on both mirror segmentation and mirror normal estimation metrics on the MP3D-mesh (Matterport3D mesh-based depth) dataset.

Impact of mirror border width. We experiment with different mirror border width values for estimating the mirror



(a) NYUv2 (b) Matterport3D (c) ScanNet (d) Overall

Figure 1: Distribution of mirror mask centroids in image plane, plotted by source RGBD dataset (a through c) and overall (d). Each point corresponds to a mirror instance mask. Matterport3D is collected using a static tripod-based sensor and exhibits strong vertical centering of the mirror centroid point. In contrast, ScanNet and NYUv2 were collected with hand-held depth sensors which were frequently held in a slightly downwards facing angle, and avoided capturing mirrors 'head on'. The overall dataset therefore exhibits a somewhat bimodal distribution of mirror centroid along the vertical axis.



(a) NYUv2 (b) Matterport3D (c) ScanNet (d) Overall

Figure 2: Distribution of the maximum depth value per frame by source RGBD dataset (a through c) and overall (d) for both the raw and refined depth in images with mirrors. After correcting the noisy depth values using our 3D mirror plane annotations, the maximum depth is consistently reduced for all datasets showing the reduction of depth outlier points at far distances (usually associated with 'floating points' behind mirror surfaces).

# anchors	AC-AP ↑	AR-L2 ↓	AngErr ↓
3	0.035	0.307	17.779
5	0.044	0.212	12.252
7	**0.046**	0.213	12.286
10	0.041	**0.150**	**8.616**
12	0.002	0.171	9.857

Table 1: Ablation on anchor normal count. We evaluate a spectrum of anchor normal counts in terms of mirror normal estimation. Using 10 anchor normals gives the lowest L2 normal regression error (AR-L2) and normal angle error (AngErr).

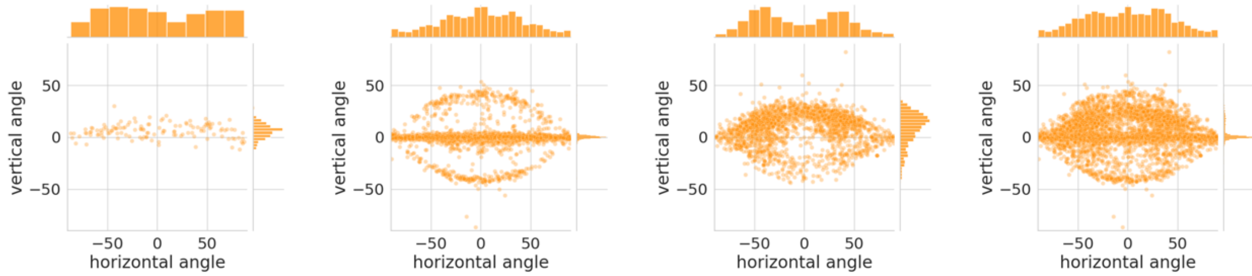
	RMSE ↓			SSIM ↑		
Border width	Mirror	Other	All	Mirror	Other	All
30	0.598	**0.210**	0.293	0.866	0.965	0.951
50	**0.570**	0.211	**0.282**	**0.867**	**0.966**	**0.952**
70	0.602	0.224	0.309	0.863	**0.966**	0.950

Table 2: Ablations on mirror border width. We tested performance on three difference border widths (number of pixels expanded outwards from mirror mask region) on the MP3D-mesh-ref dataset. We find that a border width of 50 produces the lowest RMSE and SSIM metrics.

plane offset on the MP3D-mesh-ref dataset. Table 2 summarizes the results. The mirror border width has a relatively small impact on the accuracy of the estimated depth, with a mirror border width of 50 pixels giving the best results.

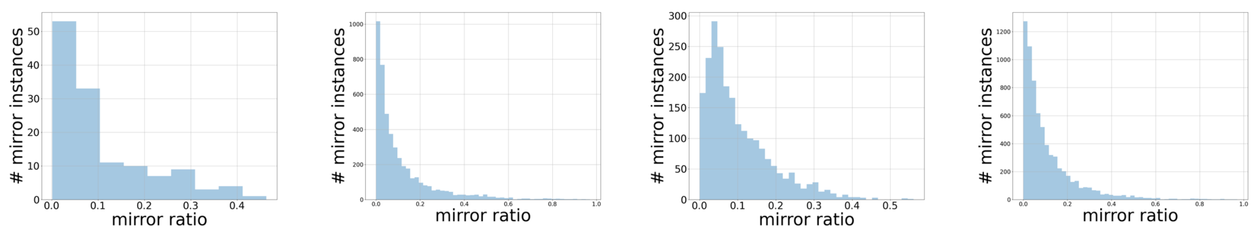
C. Additional qualitative results

We provide more qualitative results presented in a similar layout as the qualitative results figure in the main paper. Figure 9 shows two qualitative comparison results from NYUv2 (top two sets) and two from Matterport3D (bottom



(a) NYUv2 (b) Matterport3D (c) ScanNet (d) Overall

Figure 3: Distribution of mirror normals by source RGBD dataset (a through c) and overall (d). Each point corresponds to the normal of a 3D mirror plane instance. The overall distribution reveals that many normals are at angles that are close to 'facing towards' the camera with small horizontal deviations. We note interesting differences between the distributions for different source datasets. The Matterport3D frames exhibit a tight horizontal vertical angle distribution, which we hypothesize is due to the tripod-based sensor. This is in contrast to both NYUv2 and ScanNet which were both captured with handheld devices and exhibit broader distributions.



(a) NYUv2 (b) Matterport3D (c) ScanNet (d) Overall

Figure 4: Distribution of the ratio of mirror pixels to total pixels, by source RGBD dataset (a through c) and overall (d).

two sets). Note that by using Mirror3DNet to refine depth output from various depth estimation and completion approaches we can significantly reduce the depth error against the corrected ground truth depth (lower error values, particularly in mirror regions shown in the rightmost RMSE visualization column).

D. Additional quantitative results

Here, we compile a complete set of result tables including all the quantitative evaluation metrics we defined in the main paper. In Table 3 we report the full set of metrics for experiments using the NYUv2-ref dataset. Contrast these results with the ones obtained when using the NYUv2-raw dataset, shown in Table 4 to see the impact of using the original raw depth as the ground truth for evaluation. Similarly, in Table 5 we report complete metrics for the experiments using MP3D-mesh-ref, and contrast these metrics against the original dataset depth being used as ground truth in Table 6.

References

[1] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3D: Learning from RGB-D data in indoor environments. In *Proc. of International Conference on 3D Vision (3DV)*, 2017.

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5828–5839, 2017.

[3] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation. *arXiv preprint arXiv:1907.10326*, 2019.

[4] Dmitry Senushkin, Ilia Belikov, and Anton Konushin. Decoder modulation for indoor depth completion. *arXiv preprint arXiv:2005.08607*, 2020.

[5] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from RGBD images. In *Eur. Conf. Comput. Vis.*, pages 746–760. Springer, 2012.

[6] Wei Yin, Yifan Liu, Chunhua Shen, and Youliang Yan. En-

Figure 5: Example 3D mirror plane annotations in our Mirror3D dataset. Source RGBD data from the NYUv2 [5] dataset. In each image pair, the mirror mask is shown as a transparent red on the RGB image, the mirror plane is in light blue on the point cloud, and erroneous depth points that are incorrectly behind the mirror plane in the raw depth are shaded in orange.

forcing geometric constraints of virtual normal for depth prediction. In *Int. Conf. Comput. Vis.*, pages 5684–5693, 2019.

Figure 6: Example 3D mirror plane annotations in our Mirror3D dataset. Source RGBD data from the Matterport3D [1] dataset. In each image pair, the mirror mask is shown as a transparent polygon (with different color signifying each mirror instance) on the RGB image, the mirror plane is in light blue on the point cloud, and erroneous depth points that are incorrectly behind the mirror plane in the raw depth are shaded in orange.

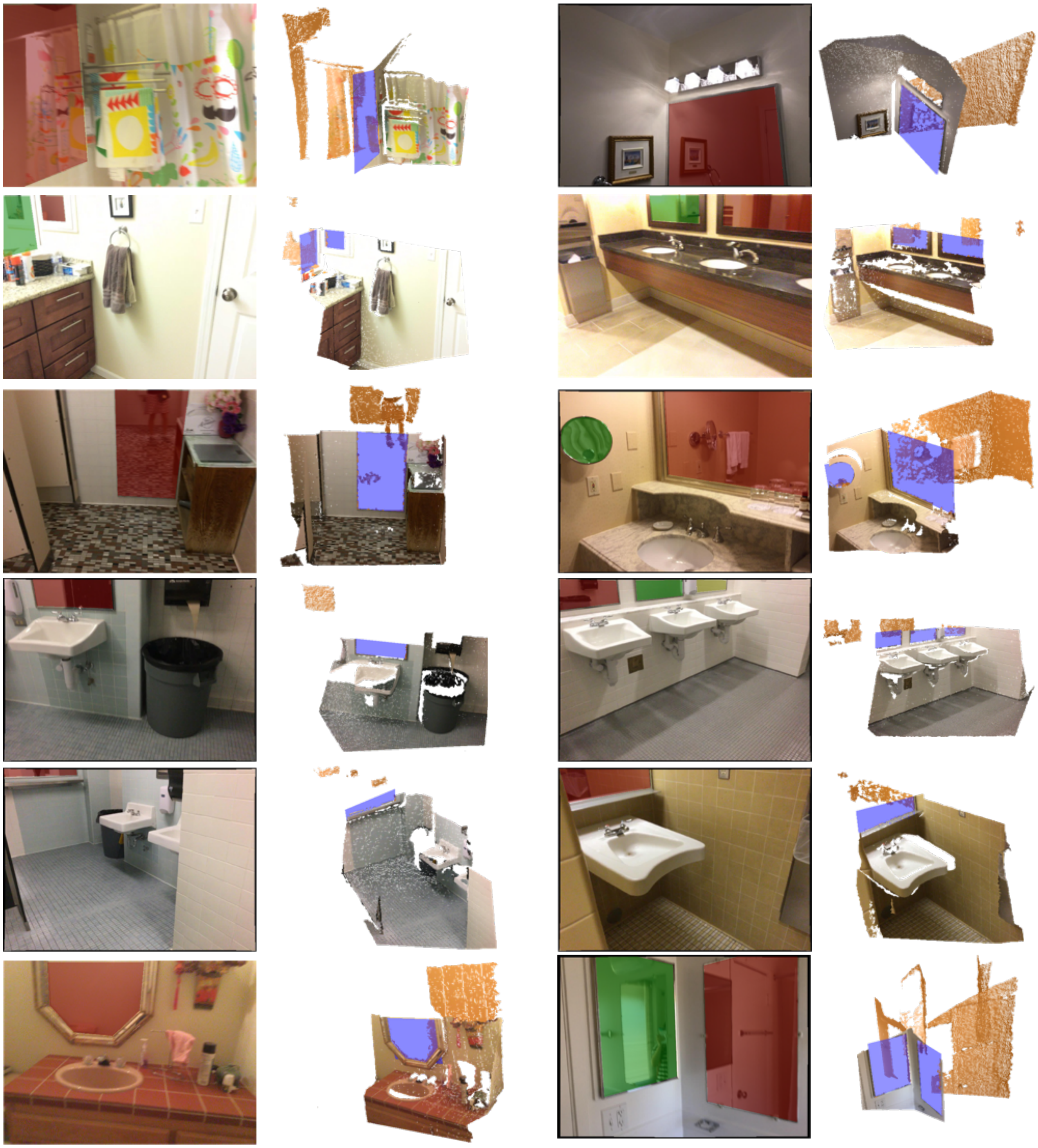


Figure 7: Example 3D mirror plane annotations in our Mirror3D dataset. Source RGBD data from the ScanNet [2] dataset. In each image pair, the mirror mask is shown as a transparent polygon (with different color signifying each mirror instance) on the RGB image, the mirror plane is in light blue on the point cloud, and erroneous depth points that are incorrectly behind the mirror plane in the raw depth are shaded in orange.

Figure 8: Example of 'coarse' and 'detailed' masks in our Mirror3D dataset, images from the Matterport3D [1] dataset. The 'detailed' mask provides a pixel-accurate boundary of the visible mirror surface in each image.

			RMSE ↓			s-RMSE ↓			Rel ↓			SSIM ↑		
Input	Train	Method	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
RGBD	*	Mirror3DNet	0.891	0.077	0.309	0.698	0.155	0.257	0.454	0.008	0.074	0.721	0.984	0.946
RGBD	ref	saic[4]	1.081	**0.074**	0.391	0.782	**0.201**	0.308	0.556	**0.012**	0.099	0.669	0.928	0.884
RGBD	raw	saic[4]	1.170	0.077	0.417	0.849	0.214	0.329	0.601	0.015	0.107	**0.658**	0.926	**0.882**
RGBD	raw	Mirror3DNet + saic [4]	**0.874**	0.095	**0.314**	0.695	0.169	**0.266**	**0.446**	0.018	**0.081**	0.718	**0.922**	0.888
RGB	ref	BTS[3]	**0.472**	0.351	**0.391**	**0.343**	**0.281**	**0.298**	**0.250**	0.158	0.175	0.825	0.832	0.821
RGB	ref	VNL[6]	6.169	5.804	5.882	0.394	0.574	0.563	3.667	3.465	3.503	**0.228**	**0.203**	**0.204**
RGB	raw	BTS[3]	0.971	0.315	0.547	0.702	0.346	0.414	0.501	**0.112**	0.188	0.691	0.856	0.819
RGB	raw	VNL[6]	3.939	2.265	2.725	0.853	0.463	0.526	2.197	1.099	1.296	0.384	0.629	0.583
RGB	raw	Mirror3DNet + BTS[3]	0.801	**0.317**	0.481	0.607	0.319	0.378	0.404	**0.112**	**0.169**	0.753	0.856	0.827
RGB	raw	Mirror3DNet + VNL[6]	3.462	2.262	2.554	0.719	0.410	0.461	1.932	1.099	1.242	0.444	0.628	0.593

			1.05 ↑			1.10 ↑			1.25 ↑			1.25^2 ↑			1.25^3 ↑		
Input	Train	Method	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
RGBD	*	Mirror3DNet	0.194	0.972	0.873	0.299	0.981	0.891	0.465	0.990	0.915	0.695	0.997	0.954	0.837	0.999	0.975
RGBD	ref	saic[4]	0.183	**0.969**	**0.867**	0.264	**0.986**	**0.888**	0.407	**0.995**	0.908	0.592	**0.998**	0.929	0.795	**0.999**	0.961
RGBD	raw	saic[4]	0.186	**0.969**	**0.867**	0.254	**0.986**	0.887	0.380	**0.995**	0.906	0.561	0.997	0.927	0.757	**0.999**	0.958
RGBD	raw	Mirror3DNet + saic [4]	**0.203**	0.958	0.860	**0.304**	0.977	0.887	**0.471**	0.989	**0.915**	**0.697**	0.997	**0.954**	**0.843**	**0.999**	**0.975**
RGB	ref	BTS[3]	**0.255**	0.276	0.268	**0.404**	0.499	0.483	0.700	0.793	0.777	**0.881**	0.961	**0.947**	**0.952**	0.990	0.982
RGB	ref	VNL[6]	0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.004	0.004	0.012	0.023	0.023	0.058	0.063	0.063
RGB	raw	BTS[3]	0.130	**0.360**	0.321	0.284	**0.588**	**0.537**	0.527	**0.885**	0.820	0.681	0.971	0.913	0.827	0.992	0.959
RGB	raw	VNL[6]	0.014	0.034	0.031	0.035	0.067	0.060	0.061	0.165	0.148	0.125	0.400	0.357	0.247	0.607	0.552
RGB	raw	Mirror3DNet + BTS[3]	0.151	**0.360**	**0.324**	0.298	0.584	0.536	**0.577**	0.883	**0.826**	0.770	**0.972**	0.934	0.880	**0.993**	**0.971**
RGB	raw	Mirror3DNet + VNL[6]	0.034	0.034	0.032	0.044	0.067	0.062	0.080	0.164	0.151	0.131	0.396	0.358	0.298	0.605	0.558

Table 3: Additional quantitative metrics for experiments on NYUv2-ref dataset images containing mirrors (NYUv2 [5] with ground truth depth refined using 3D mirror plane annotations).

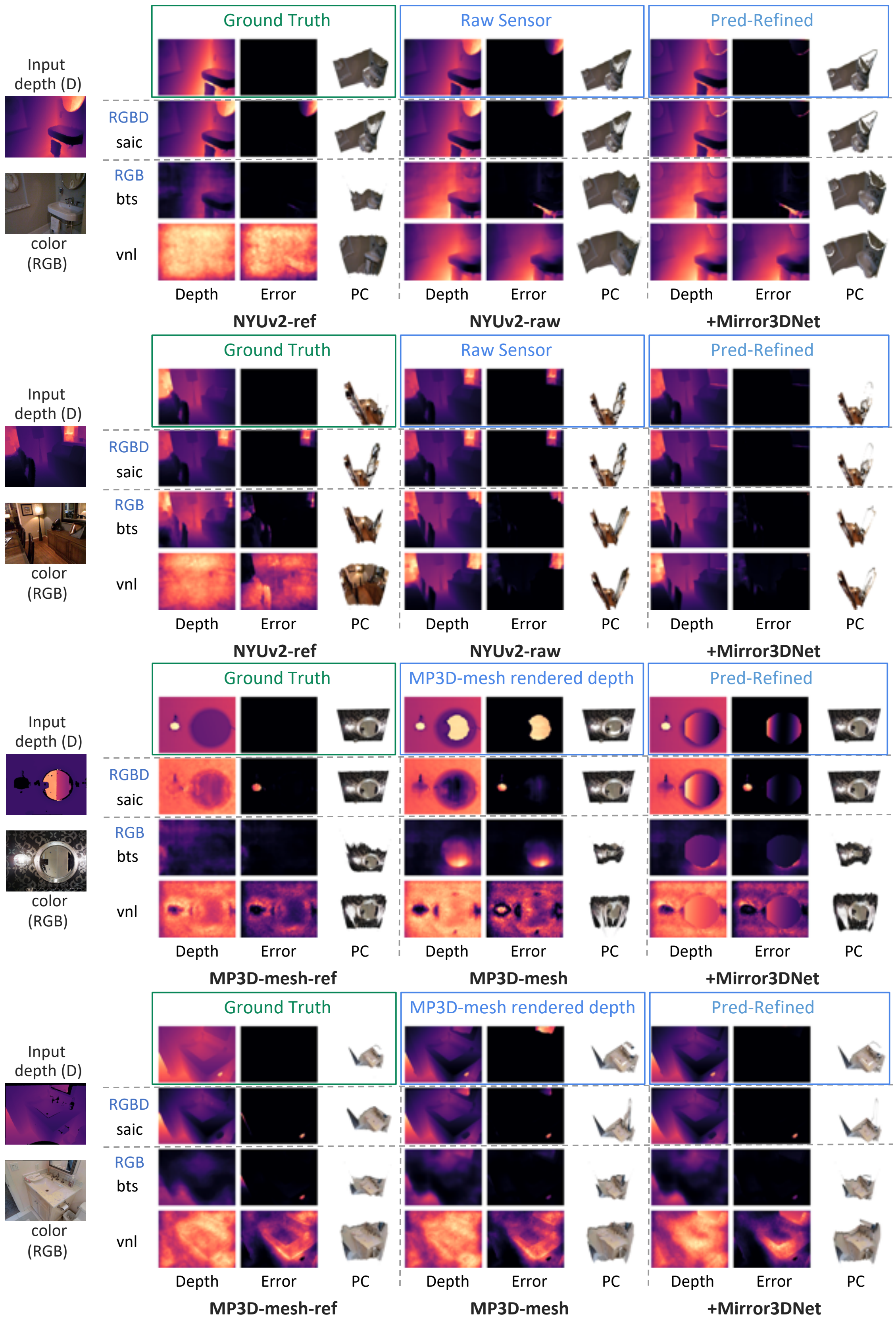


Figure 9: Additional qualitative results. Top two rows results from NYUv2 [5], bottom two from Matterport3D [1].

Input	Train	Method	RMSE ↓			s-RMSE ↓			Rel ↓			SSIM ↑		
			Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
RGBD	*	Mirror3DNet	0.414	0.036	0.179	0.384	0.065	0.171	0.080	0.003	0.017	0.885	0.995	0.974
RGBD	ref	saic[4]	0.201	**0.042**	0.085	0.192	0.046	0.082	0.036	**0.008**	0.011	0.846	0.935	0.919
RGBD	raw	saic[4]	**0.102**	**0.042**	**0.054**	**0.098**	**0.039**	**0.051**	**0.017**	**0.011**	**0.011**	0.862	0.933	0.920
RGBD	raw D	Mirror3DNet + saic [4]	0.484	0.070	0.216	0.450	0.095	0.205	0.093	0.014	0.028	**0.788**	**0.929**	**0.901**
RGB	ref	BTS[3]	**1.024**	0.355	0.538	0.960	0.332	0.470	0.279	0.157	0.174	0.644	0.833	0.796
RGB	ref	VNL[6]	5.238	5.798	5.741	1.070	0.623	0.723	2.181	3.454	3.229	**0.264**	**0.204**	**0.209**
RGB	raw	BTS[3]	0.962	**0.316**	**0.452**	**0.874**	**0.274**	**0.382**	**0.253**	**0.111**	**0.124**	0.621	0.858	0.821
RGB	raw	VNL[6]	3.113	2.254	2.491	0.925	0.371	0.463	1.148	1.092	1.107	0.433	0.633	0.599
RGB	raw D	Mirror3DNet + BTS[3]	1.098	0.318	0.510	0.980	0.288	0.438	0.280	**0.111**	0.131	0.624	0.857	0.817
RGB	raw D	Mirror3DNet + VNL[6]	2.902	2.252	2.416	1.044	0.372	0.509	1.077	1.093	1.090	0.438	0.631	0.598

Input	Train	Method	1.05 ↑			1.10 ↑			1.25 ↑			1.25^2 ↑			1.25^3 ↑		
			Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
RGBD	*	Mirror3DNet	0.783	0.988	0.950	0.805	0.990	0.954	0.832	0.994	0.962	0.889	1.000	0.980	0.954	1.000	0.992
RGBD	ref	saic[4]	0.833	0.984	0.969	0.897	0.994	0.984	0.955	**0.999**	0.994	0.983	**1.000**	0.998	0.995	**1.000**	0.999
RGBD	raw	saic[4]	**0.940**	**0.985**	**0.980**	**0.975**	**0.995**	**0.993**	**0.995**	**0.999**	**0.999**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
RGBD	raw D	Mirror3DNet + saic [4]	0.732	0.973	0.932	0.783	0.985	0.948	0.827	0.993	0.961	0.887	0.999	0.979	0.952	**1.000**	0.991
RGB	ref	BTS[3]	0.145	0.277	0.254	0.250	0.500	0.460	0.432	0.793	0.742	0.686	0.960	0.919	0.874	0.989	0.970
RGB	ref	VNL[6]	0.004	0.001	0.001	0.007	0.001	0.002	0.021	0.004	0.006	0.087	0.023	0.027	0.170	0.064	0.072
RGB	raw	BTS[3]	**0.204**	**0.361**	**0.344**	**0.365**	**0.588**	**0.568**	**0.576**	**0.885**	**0.858**	**0.771**	0.971	**0.954**	**0.898**	0.992	**0.982**
RGB	raw	VNL[6]	0.030	0.034	0.033	0.052	0.067	0.065	0.123	0.166	0.159	0.251	0.401	0.377	0.514	0.610	0.593
RGB	raw D	Mirror3DNet + BTS[3]	0.170	0.361	0.334	0.304	0.585	0.550	0.510	0.883	0.839	0.719	**0.972**	0.943	0.866	**0.994**	0.977
RGB	raw D	Mirror3DNet + VNL[6]	0.043	0.034	0.034	0.073	0.067	0.067	0.141	0.165	0.164	0.291	0.397	0.384	0.549	0.608	0.601

Table 4: Additional quantitative metrics for experiments on NYUv2-raw dataset images containing mirrors (NYUv2 [5] using original raw depth as ground truth). Compare to Table 3 and note the incorrect ranking of methods with this imperfect ground truth.

Input	Train	Method	RMSE ↓			s-RMSE ↓			Rel ↓			SSIM ↑		
			Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
sensor-D	*	*	2.605	1.268	1.048	2.215	1.140	1.020	0.985	0.281	0.203	0.215	0.669	0.730
mesh-D	*	*	0.631	**0.000**	0.177	0.597	**0.035**	0.168	0.162	**0.000**	**0.024**	0.794	**1.000**	0.970
RGBD (sensor-D)	*	Mirror3DNet	1.542	0.897	0.960	1.271	0.969	0.886	0.648	0.130	0.160	0.586	0.798	0.780
RGBD (mesh-D)	*	Mirror3DNet	**0.428**	0.016	**0.150**	**0.393**	0.046	**0.141**	**0.134**	0.001	**0.024**	**0.881**	0.998	**0.978**
RGBD	mesh-ref	saic[4]	**0.308**	**0.316**	**0.358**	**0.281**	**0.322**	0.332	**0.099**	0.055	0.073	**0.861**	0.899	0.889
RGBD	mesh	saic[4]	0.984	0.320	0.595	0.698	0.421	0.486	0.427	**0.054**	0.148	0.692	**0.909**	0.864
RGBD	mesh	Mirror3DNet + saic [4]	0.786	0.553	0.389	0.552	0.456	**0.326**	0.379	0.142	**0.065**	0.786	0.870	**0.909**
RGB	mesh-ref	BTS[3]	**0.572**	**0.634**	**0.658**	**0.391**	**0.489**	**0.489**	**0.262**	**0.297**	**0.305**	**0.788**	**0.776**	**0.769**
RGB	mesh-ref	VNL[6]	1.364	1.410	1.408	0.681	0.861	0.840	0.553	0.594	0.598	0.620	0.630	0.623
RGB	mesh	BTS[3]	1.142	1.033	1.097	0.685	0.680	0.691	0.525	0.413	0.454	0.669	0.757	0.733
RGB	mesh	VNL[6]	1.400	1.432	1.429	0.740	0.917	0.898	0.551	0.591	0.593	0.456	0.440	0.421
RGB	mesh	Mirror3DNet + BTS[3]	1.156	1.034	1.092	0.635	0.677	0.683	0.515	0.413	0.450	0.746	0.757	0.739
RGB	mesh	Mirror3DNet + VNL[6]	1.390	1.424	1.423	0.688	0.877	0.856	0.564	0.600	0.603	0.612	0.475	0.470

Input	Train	Method	1.05 ↑			1.10 ↑			1.25 ↑			1.25^2 ↑			1.25^3 ↑		
			Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
sensor-D	*	*	0.122	0.744	0.803	0.144	0.757	0.822	0.197	0.776	0.837	0.291	0.798	0.849	0.373	0.816	0.855
mesh-D	*	*	**0.661**	**1.000**	**0.943**	0.745	**1.000**	**0.959**	0.822	**1.000**	**0.972**	0.878	**1.000**	0.983	0.902	**1.000**	0.987
RGBD (sensor-D)	*	Mirror3DNet	0.221	0.851	0.824	0.300	0.865	0.837	0.418	0.879	0.849	0.562	0.890	0.860	0.649	0.895	0.866
RGBD (mesh-D)	*	Mirror3DNet	0.651	0.996	0.936	**0.758**	0.998	0.956	**0.847**	0.999	**0.972**	**0.918**	**1.000**	**0.985**	**0.942**	**1.000**	**0.989**
RGBD	mesh-ref	saic[4]	**0.595**	0.875	0.823	**0.787**	0.917	0.887	**0.920**	**0.955**	**0.941**	**0.962**	**0.977**	**0.969**	**0.979**	**0.987**	**0.982**
RGBD	mesh	saic[4]	0.385	**0.893**	0.801	0.518	**0.924**	0.842	0.670	**0.955**	0.889	0.783	0.976	0.927	0.846	**0.985**	0.948
RGBD	mesh	Mirror3DNet + saic [4]	0.453	0.802	**0.869**	0.579	0.844	**0.904**	0.716	0.892	0.940	0.810	0.930	0.961	0.871	0.952	0.970
RGB	mesh-ref	BTS[3]	**0.190**	**0.183**	**0.181**	**0.353**	**0.339**	**0.334**	**0.653**	**0.624**	**0.619**	**0.878**	**0.842**	**0.840**	**0.949**	**0.937**	**0.933**
RGB	mesh-ref	VNL[6]	0.048	0.046	0.046	0.094	0.091	0.090	0.222	0.218	0.217	0.442	0.448	0.446	0.663	0.671	0.671
RGB	mesh	BTS[3]	0.053	0.063	0.060	0.106	0.124	0.118	0.268	0.298	0.285	0.575	0.617	0.598	0.812	0.858	0.838
RGB	mesh	VNL[6]	0.045	0.046	0.046	0.088	0.091	0.091	0.211	0.217	0.216	0.436	0.445	0.446	0.652	0.668	0.669
RGB	mesh	Mirror3DNet + BTS[3]	0.049	0.063	0.059	0.096	0.125	0.117	0.251	0.298	0.286	0.570	0.617	0.599	0.819	0.858	0.840
RGB	mesh	Mirror3DNet + VNL[6]	0.048	0.049	0.048	0.101	0.095	0.094	0.212	0.218	0.215	0.438	0.445	0.444	0.660	0.663	0.663

Table 5: Additional quantitative metrics for experiments on MP3D-mesh-ref dataset images containing mirrors (Matterport3D [1] with ground truth depth refined using 3D mirror plane annotations).

			RMSE ↓			s-RMSE ↓			Rel ↓			SSIM ↑		
Input	Train	Method	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
RGBD (sensor-D)	*	Mirror3DNet	1.610	0.897	1.124	1.363	0.960	1.034	0.643	0.130	0.244	0.550	0.798	0.745
RGBD (mesh-D)	*	Mirror3DNet	**0.194**	**0.016**	**0.054**	**0.184**	**0.021**	**0.053**	**0.050**	**0.001**	**0.004**	**0.936**	**0.998**	**0.993**
RGBD	mesh-ref	saic[4]	**0.438**	**0.316**	**0.374**	**0.397**	**0.319**	**0.346**	**0.140**	0.055	**0.079**	**0.817**	0.899	**0.886**
RGBD	mesh	saic[4]	1.013	0.320	0.584	0.750	0.409	0.483	0.428	**0.054**	0.146	0.672	**0.909**	0.864
RGBD	mesh	Mirror3DNet + saic [4]	0.853	0.321	0.544	0.635	0.395	0.455	0.390	**0.054**	0.141	0.743	0.908	0.869
RGB	mesh-ref	BTS[3]	**0.681**	**0.634**	**0.665**	**0.495**	**0.485**	**0.499**	**0.296**	**0.297**	**0.310**	**0.750**	**0.776**	**0.765**
RGB	mesh-ref	VNL[6]	1.474	1.410	1.418	0.805	0.859	0.855	0.580	0.594	0.603	0.580	0.630	0.620
RGB	mesh	BTS[3]	1.236	1.033	1.101	0.751	0.675	0.696	0.556	0.413	0.460	0.641	0.757	0.731
RGB	mesh	VNL[6]	1.510	1.432	1.439	0.867	0.915	0.913	0.577	0.591	0.598	0.439	0.440	0.418
RGB	mesh	Mirror3DNet + BTS[3]	1.255	1.034	1.097	0.723	0.671	0.689	0.546	0.413	0.456	0.695	0.757	0.735
RGB	mesh	Mirror3DNet + VNL[6]	1.502	1.424	1.433	0.820	0.875	0.871	0.590	0.600	0.609	0.560	0.475	0.465

			1.05 ↑			1.10 ↑			1.25 ↑			1.25^2 ↑			1.25^3 ↑		
Input	Train	Method	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All	Mirror	Other	All
RGBD (sensor-D)	*	Mirror3DNet	0.240	0.851	0.757	0.303	0.865	0.774	0.422	0.879	0.797	0.563	0.890	0.823	0.650	0.895	0.843
RGBD (mesh-D)	*	Mirror3DNet	**0.844**	**0.996**	**0.984**	**0.900**	**0.998**	**0.990**	**0.951**	**0.999**	**0.996**	**0.980**	**1.000**	**0.998**	**0.987**	**1.000**	**0.999**
RGBD	mesh-ref	saic[4]	**0.552**	0.875	**0.820**	**0.719**	0.917	**0.880**	**0.867**	**0.955**	**0.935**	**0.931**	**0.977**	**0.964**	**0.960**	**0.987**	**0.979**
RGBD	mesh	saic[4]	0.410	**0.893**	0.810	0.527	**0.924**	0.848	0.668	**0.955**	0.892	0.785	0.976	0.929	0.851	0.985	0.950
RGBD	mesh	Mirror3DNet + saic [4]	0.441	0.891	0.809	0.560	0.923	0.849	0.702	**0.955**	0.895	0.807	0.976	0.932	0.871	0.985	0.953
RGB	mesh-ref	BTS[3]	**0.184**	**0.183**	**0.179**	**0.335**	**0.339**	**0.331**	**0.609**	**0.624**	**0.612**	**0.847**	**0.842**	**0.834**	**0.931**	**0.937**	**0.930**
RGB	mesh-ref	VNL[6]	0.046	0.046	0.046	0.090	0.091	0.090	0.214	0.218	0.215	0.436	0.448	0.443	0.642	0.671	0.667
RGB	mesh	BTS[3]	0.051	0.063	0.060	0.101	0.124	0.118	0.258	0.298	0.285	0.561	0.617	0.597	0.787	0.858	0.836
RGB	mesh	VNL[6]	0.044	0.046	0.046	0.087	0.091	0.090	0.207	0.217	0.214	0.430	0.445	0.443	0.627	0.668	0.664
RGB	mesh	Mirror3DNet + BTS[3]	0.048	0.063	0.060	0.094	0.125	0.118	0.244	0.298	0.286	0.551	0.617	0.598	0.794	0.858	0.838
RGB	mesh	Mirror3DNet + VNL[6]	0.046	0.049	0.048	0.092	0.095	0.093	0.202	0.218	0.214	0.433	0.445	0.441	0.635	0.663	0.658

Table 6: Additional quantitative metrics for experiments on MP3D-mesh dataset images containing mirrors (Matterport3D [1] using original mesh-rendered depth as ground truth). Compare to Table 5 and note the incorrect ranking of methods with this imperfect ground truth.