Supplemental Material:
Panoptic 3D Scene Reconstruction
From a Single RGB Image

Manuel Dahnert **Ji Hou** **Matthias Nießner** **Angela Dai**

Technical University of Munich

A Additional Quantitative Results

In Table 1, we provide additional ablations on the effect of 3D refinement and completion as well as the 2d features. "Ours w/o 3D" evaluates the performance of the backprojected depth with the 2D instances. "Ours w/o 2D feat." is trained without additional 2D features.

Additionally, in Tables 2, 3, and 4, we show the per-class results of PRQ, SRQ and RRQ on synthetic 3D-Front (3) data. The per-class results for the ablations with ground truth depth information are in Tables 5, 6, 7. This ablation also includes results with Sketch-Aware SSC (2). We also show the per-class results for PRQ, SRQ, and RRQ on real-world Matterport3D data in Tables 8, 9, and 10.

Table 1: Additional quantitative evaluations of Panoptic Reconstruction Quality on 3D-Front (3).

	PRQ	RSQ	RRQ	PRQ	RSQ	RRQ	PRQ	RSQ	RRQ
					Things			*Stuff*	
SSCNet (6) + IC	11.50	32.90	33.00	8.03	32.07	24.69	26.95	36.75	70.25
Mesh R-CNN (4)	-	-	-	20.90	38.00	53.20	-	-	-
Total3D (5)	15.08	36.63	40.15	13.77	34.88	38.89	20.94	44.49	45.85
Ours w/o 3D	-	-	-	8.94	32.58	27.19	-	-	-
Ours w/o IP	20.65	53.87	29.62	8.48	48.30	15.07	**75.40**	78.95	**95.10**
Ours w/o 2D feat.	45.34	55.86	72.64	39.34	50.82	**68.43**	72.30	78.55	91.55
Ours w/o hier.	44.05	55.31	70.54	37.34	50.12	65.33	74.20	78.65	93.95
Ours	**46.77**	**57.35**	**73.13**	**40.52**	**52.52**	**68.43**	74.90	**79.10**	94.25

Table 2: Per-class results of Panoptic Reconstruction Quality (PRQ) on 3D-Front (3).

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor
SSCNet + IC	7.80	16.60	7.90	13.30	12.10	5.50	0.50	0.70	7.90	15.20	38.70
Mesh R-CNN	29.70	13.30	24.10	24.40	28.50	23.50	14.40	1.40	28.70	-	-
Total3D	17.25	4.56	18.76	14.07	19.40	16.79	7.04	8.13	17.97	8.27	33.61
Ours w/o 3D	9.70	2.00	11.10	2.40	13.20	2.10	10.40	**13.70**	15.90	-	-
Ours w/o IP	8.50	25.00	9.30	2.40	11.70	4.20	3.00	0.00	12.20	**64.90**	85.90
Ours w/o hier.	42.70	58.70	32.00	**56.40**	**36.30**	17.00	44.50	0.00	48.50	64.10	84.30
Ours	**47.40**	**58.90**	**36.60**	53.50	35.60	**31.70**	**47.90**	0.00	**53.10**	63.40	**86.40**

B Architecture

We provide detailed versions of the network architecture: Figure 1 shows the 2D feature extraction, depth estimation and mask predictions, as well as the 2D-3D backprojection. Figure 2 shows the sparse, generative 3D U-Net. Each sparse and dense block consists of a 3D ResNet block.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

Table 3: Per-class results of Segmentation Reconstruction Quality (SRQ) on 3D-Front (3).

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor
SSCNet + IC	30.90	31.40	31.90	31.00	34.30	36.70	0.00	0.00	33.40	30.80	41.90
Mesh R-CNN	44.60	30.40	40.90	34.90	42.50	35.20	32.90	30.60	49.40	-	-
Total3D	36.35	29.88	36.93	33.24	35.66	34.15	31.86	37.44	38.43	42.42	46.55
Ours w/o 3D	29.80	27.60	28.40	37.50	32.80	29.80	31.60	**41.10**	34.60	-	-
Ours w/o IP	54.40	**62.40**	47.40	42.00	56.20	44.60	**66.80**	0.00	60.90	**70.80**	87.10
Ours w/o hier.	61.40	58.70	47.90	**59.80**	55.50	49.60	55.20	0.00	63.00	70.60	86.70
Ours	**66.70**	58.90	**51.50**	59.60	**58.50**	**56.10**	56.90	0.00	**64.50**	70.70	**87.50**

Table 4: Per-class results of Recognition Reconstruction Quality (RRQ) on 3D-Front (3).

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor
SSCNet + IC	23.50	45.80	23.50	57.60	34.20	21.10	0.00	0.00	22.40	46.00	92.70
Mesh R-CNN	66.70	43.70	58.80	69.80	**67.10**	**66.70**	43.80	4.70	58.00	-	-
Total3D	47.46	15.25	50.81	42.33	54.41	49.15	22.10	21.70	46.76	19.49	72.20
Ours w/o 3D	32.50	7.30	39.00	6.50	40.20	7.10	32.90	**33.30**	45.90	-	-
Ours w/o IP	15.60	40.00	19.60	5.70	20.80	9.50	4.40	0.00	20.00	**91.60**	98.60
Ours w/o hier.	69.60	**100.00**	66.90	**94.30**	65.40	34.30	80.60	0.00	76.90	90.70	97.20
Ours	**71.10**	**100.00**	**71.10**	89.80	60.80	56.50	**84.20**	0.00	**82.40**	89.70	**98.80**

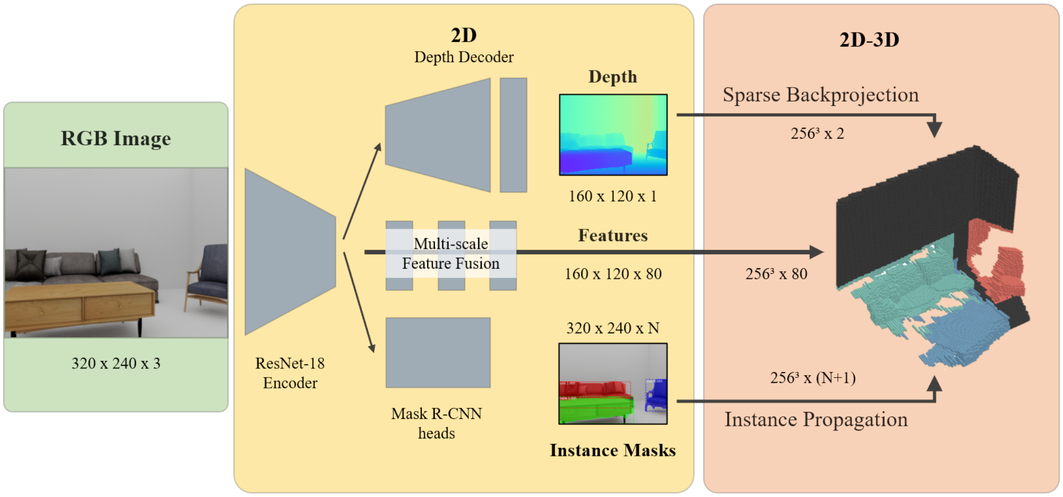

Figure 1: First part of the network architecture.

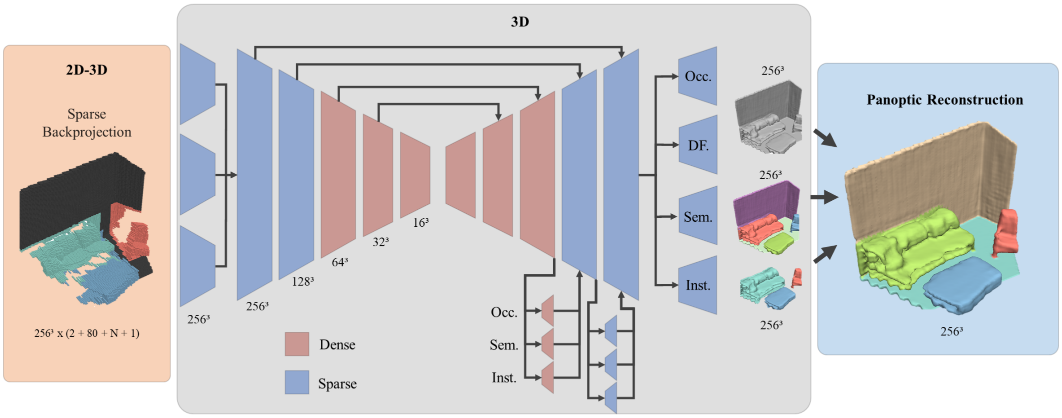

Figure 2: Second part of the network architecture.

Table 5: Per-class results of Panoptic Reconstruction Quality (PRQ) on 3D-Front (3) with ground truth depth information.

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor
SSCNet + IC w/ GT depth	9.10	21.50	14.90	16.20	15.00	9.20	2.40	10.60	12.30	24.80	38.70
Sketch + IC w/ GT depth	21.80	38.20	18.20	32.70	27.10	30.90	25.30	22.40	18.50	29.20	22.00
Mesh R-CNN w/ GT z	38.70	27.00	**45.20**	29.00	**38.90**	30.40	20.50	**44.40**	47.70	-	-
Ours	**42.60**	**58.50**	40.50	**54.50**	36.10	**34.10**	**54.10**	0.00	**54.60**	**75.10**	**78.80**

Table 6: Per-class results of Segmentation Reconstruction Quality (SRQ) on 3D-Front (3) with ground truth depth information.

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor
SSCNet + IC w/ GT depth	32.20	33.60	33.30	32.00	36.90	31.40	29.00	34.00	34.10	31.50	42.90
Sketch + IC w/ GT depth	35.10	39.40	38.30	37.10	36.50	42.10	36.70	36.80	32.70	33.00	30.00
Mesh R-CNN w/ GT z	48.90	33.60	50.40	38.00	46.90	40.30	35.60	**54.40**	56.50	-	-
Ours	**62.30**	**58.50**	**54.50**	**59.40**	**51.90**	**52.20**	**60.70**	0.00	**62.70**	**75.50**	**81.20**

C Data

We use the synthetic data of 3D-Front (3) and real-world 3D scans of Matterport3D (1) for training and evaluation of the panoptic 3D scene reconstruction task. Both datasets are licensed under non-commercial use[1][2]. Collection of the data was obtained by Alibaba and Matterport, respectively, from the designers and owners, and the data anonymized without any offensive content.

---

[1] `https://tianchi.aliyun.com/specials/promotion/alibaba-3d-scene-dataset`

[2] `http://kaldir.vc.in.tum.de/matterport/MP_TOS.pdf`

Table 7: Per-class results of Recognition Reconstruction Quality (RRQ) on 3D-Front (3) with ground truth depth information.

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor
SSCNet + IC w/ GT depth	28.20	63.90	44.60	50.60	40.60	29.20	8.10	31.20	36.20	78.80	90.30
Sketch + IC w/ GT depth	62.10	97.00	47.40	88.00	76.80	73.30	69.00	60.70	56.60	88.50	73.50
Mesh R-CNN w/ GT z	**79.10**	80.20	**89.80**	76.40	**82.80**	**75.30**	57.50	**81.50**	84.40	-	-
Ours	68.30	**100.00**	74.30	**91.80**	69.50	65.30	**89.20**	0.00	**87.10**	**99.40**	**97.00**

Table 8: Per-class results of Panoptic Reconstruction Quality (PRQ) on Matterport3d (1).

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor	Ceiling
SSCNet + IC	0.07	0.11	0.61	0.07	0.53	0.00	0.00	0.00	0.19	0.34	3.96	0.00
Mesh R-CNN	3.10	10.00	**14.80**	12.00	**7.90**	0.00	0.00	**2.80**	**6.00**	-	-	-
Ours	**12.33**	**10.24**	9.75	**14.40**	**8.07**	0.00	0.00	0.00	2.26	**10.92**	**16.54**	**4.8**8

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. *International Conference on 3D Vision (3DV)*, 2017.

[2] Xiaokang Chen, Kwan-Yee Lin, Chen Qian, Gang Zeng, and Hongsheng Li. 3d sketch-aware semantic scene completion via semi-supervised structure prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4193–4202, 2020.

[3] Huan Fu, Bowen Cai, Lin Gao, Lingxiao Zhang, Jiaming Wang Cao Li, Zengqi Xun, Chengyue Sun, Rongfei Jia, Binqiang Zhao, and Hao Zhang. 3d-front: 3d furnished rooms with layouts and semantics. *arXiv preprint arXiv:2011.09127*, 2021.

[4] Georgia Gkioxari, Jitendra Malik, and Justin Johnson. Mesh r-cnn. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9785–9795, 2019.

[5] Yinyu Nie, Xiaoguang Han, Shihui Guo, Yujian Zheng, Jian Chang, and Jian Jun Zhang. Total3dunderstanding: Joint layout, object pose and mesh reconstruction for indoor scenes from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 55–64, 2020.

[6] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image. *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

Table 9: Per-class results of Segmentation Reconstruction Quality (SRQ) on Matterport3d (1).

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor	Ceiling
SSCNet + IC	35.10	27.50	33.70	35.40	35.30	0.00	0.00	0.00	**31.90**	28.60	32.70	0.00
Mesh R-CNN	37.10	**39.10**	**43.80**	38.20	39.40	0.00	0.00	**41.00**	41.50	-	-	-
Ours	**40.30**	35.20	42.20	**38.60**	**47.20**	0.00	0.00	0.00	31.00	**37.40**	**42.40**	**40.30**

Table 10: Per-class results of Recognition Reconstruction Quality (RRQ) on Matterport3d (1).

	Cabinet	Bed	Chair	Sofa	Table	Desk	Dresser	Lamp	Other	Wall	Floor	Ceiling
SSCNet + IC	0.20	0.40	1.80	0.20	1.50	0.00	0.00	0.00	0.60	1.20	12.10	0.00
Mesh R-CNN	8.30	25.60	**33.90**	31.40	**20.10**	0.00	0.00	**6.70**	**14.40**	-	-	-
Ours	**30.60**	**29.10**	23.10	**37.30**	17.10	0.00	0.00	0.00	7.30	**29.20**	**39.00**	**12.10**

D Checklist

1. For all authors...
   - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] The main claims are presented in Section 1, and supported by comparison and ablations in Section 6.
   - (b) Did you describe the limitations of your work? [Yes] See Section 6.5.
   - (c) Did you discuss any potential negative societal impacts of your work? [Yes] See the Broader Impact section.
   - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] Code and data to be release publicly.
   - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Sections 4.1 and 6.
   - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 4.1.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   - (a) If your work uses existing assets, did you cite the creators? [Yes] We use synthetic 3D data from 3D-Front (3) and real-world 3D data from Matterport3D (1).
   - (b) Did you mention the license of the assets? [Yes] See the data section of the supplemental material
   - (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] See the data section of the supplemental material
   - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See the data section of the supplemental material
5. If you used crowdsourcing or conducted research with human subjects...
   - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]