Data Efficient 3D Learner via Knowledge Transferred from 2D Model

Supplementary Material

In this supplementary material, we show the per-category breakdown comparisons in Section A, and then the qualitative results are showed in Section E. In Section B, we compare with DepthContrast [5] on the efficient ScanNet setup. In Section C, we provide more implement details. Finally, we discuss the effectiveness of image scene parser in Section D.

A Quantitative results of per-category mIoU

We use mean IoU to validate the effectiveness of our approach on ScanNet [2] Data Efficient Benchmark in our main paper. Below, we expand the results presented in the main paper with per-category breakdown.

Limited annotations. Following the official configuration in the *3D Semantic label with Limited Annotations benchmark*, we demonstrate the state-of-the-art results of Table 1 in the main paper. Here we show the detailed performance of each category on 20, 50, 100 and 200 labeled points per scene in Table A, Table B, Table C, and Table D respectively.

Table A. 20-points LA

points	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.550	0.735	0.676	0.601	0.475	0.794	0.288	0.621	0.378	0.430	0.940	0.303	0.089	0.379	0.580	0.531	0.689	0.422	0.852	0.758	0.468
CSC	0.531	0.659	0.638	0.578	0.417	0.775	0.254	0.537	0.396	0.439	0.939	0.284	0.083	0.414	0.599	0.488	0.698	0.444	0.785	0.747	0.440
ViewPointBN	0.548	0.747	0.574	0.631	0.456	0.762	0.355	0.639	0.412	0.404	0.940	0.335	0.107	0.277	**0.645**	0.495	0.666	0.517	0.818	0.740	0.431
OTOC	0.594	0.756	0.722	0.494	0.546	0.795	0.371	0.725	**0.559**	0.488	**0.957**	0.367	**0.261**	**0.547**	0.575	0.225	0.671	0.543	0.904	**0.826**	0.557
Ours	**0.639**	**0.839**	**0.723**	**0.681**	**0.629**	**0.839**	**0.424**	**0.728**	0.538	**0.526**	0.945	**0.427**	0.12	0.511	0.643	**0.547**	**0.781**	**0.566**	**0.905**	0.809	**0.607**

Table B. 50-points LA

points	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.614	0.844	0.731	0.681	0.590	0.791	0.348	0.689	0.503	0.502	0.942	0.361	0.154	0.484	0.624	0.591	0.708	0.524	0.874	0.793	0.536
CSC	0.612	0.747	0.731	0.679	0.603	0.815	0.400	0.648	0.453	0.481	0.944	0.421	0.173	0.504	0.623	0.588	0.690	0.545	0.877	0.778	0.541
ViewPointBN	0.623	0.812	0.743	0.654	0.579	0.800	**0.462**	0.713	0.533	0.516	0.944	0.434	0.215	0.437	0.521	0.601	0.720	0.563	0.884	0.800	0.534
OTOC	0.642	0.725	0.735	0.717	0.635	0.829	0.457	0.639	0.421	0.552	**0.967**	0.460	**0.240**	0.558	**0.788**	0.621	0.720	0.477	**0.915**	**0.842**	0.539
Ours	**0.695**	**0.897**	**0.784**	**0.728**	**0.697**	**0.846**	0.441	**0.77**	**0.615**	**0.585**	0.951	**0.504**	0.232	**0.672**	0.76	**0.655**	**0.772**	**0.599**	0.877	0.834	**0.678**

Table C. 100-points LA

points	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.636	0.694	0.738	0.731	0.653	0.817	0.467	0.651	0.517	0.522	0.946	0.479	0.198	0.575	0.526	0.649	0.747	0.569	0.845	0.803	0.600
CSC	0.644	0.761	0.707	0.703	0.642	0.813	0.436	0.659	0.502	0.516	0.945	**0.487**	0.238	0.538	0.678	0.659	0.739	0.568	0.915	0.811	0.566
ViewPointBN	0.650	**0.778**	0.731	0.688	0.617	0.812	0.446	0.739	0.618	0.540	0.945	0.415	0.204	0.623	0.676	0.594	0.744	0.576	0.868	0.811	0.582
OTOC	0.670	0.734	**0.815**	0.661	0.644	**0.841**	**0.509**	0.741	0.479	0.548	**0.968**	0.461	**0.251**	0.664	0.754	0.656	0.744	0.541	**0.917**	0.844	0.625
Ours	**0.704**	0.774	0.766	**0.764**	**0.687**	0.832	0.413	**0.79**	**0.639**	**0.599**	0.952	0.478	0.222	**0.746**	**0.859**	**0.678**	**0.806**	**0.607**	0.915	**0.847**	**0.703**

Table D. 200-points LA

points	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.653	0.717	0.775	0.754	0.626	0.804	0.391	0.689	0.485	0.572	0.945	0.448	0.232	0.603	0.813	0.591	0.775	0.537	0.885	0.816	0.608
CSC	0.665	0.857	0.756	**0.763**	0.647	**0.852**	0.432	0.684	0.543	0.514	0.948	0.469	0.179	0.599	0.702	0.620	0.789	**0.614**	0.911	0.815	0.607
ViewPointBN	0.669	0.847	0.732	0.724	0.613	0.827	0.443	0.742	0.562	0.551	0.947	0.441	0.218	0.650	0.753	0.621	0.765	0.601	0.905	0.814	0.618
OTOC	0.694	0.760	**0.815**	0.706	0.684	0.840	**0.492**	0.701	0.557	0.596	**0.972**	**0.497**	**0.281**	**0.709**	0.757	0.689	0.789	0.600	0.907	**0.864**	0.6
Ours	**0.709**	**0.877**	0.772	0.744	**0.694**	0.836	0.453	**0.787**	**0.623**	**0.598**	0.953	0.49	0.216	0.682	**0.879**	**0.727**	**0.802**	0.604	**0.922**	0.845	**0.676**

Table E. 1% LR

rate	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.253	0.000	0.412	0.347	0.137	0.564	0.140	**0.361**	0.187	0.249	0.914	0.092	0.055	0.102	0.000	0.048	0.392	0.302	0.000	0.697	0.056
CSC	**0.270**	0.000	0.528	0.331	0.139	0.535	0.118	0.326	0.222	0.292	0.921	0.089	0.163	**0.129**	0.000	**0.131**	**0.463**	0.278	0.000	**0.699**	0.033
ViewPointBN	0.256	0.000	0.479	**0.377**	**0.204**	0.551	**0.205**	0.219	**0.235**	0.224	0.903	**0.092**	0.088	0.122	0.000	0.003	0.354	0.354	0.000	0.676	0.034
Ours	0.263	0	**0.547**	0.235	0.184	**0.566**	0.165	0.249	0.196	**0.309**	**0.938**	0.07	**0.186**	0.069	0	0	0.368	**0.356**	0	0.698	**0.118**

Table F. 5% LR

rate	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.438	0.517	0.659	0.251	0.332	0.783	0.244	0.408	0.411	0.409	0.935	0.206	0.119	0.200	0.048	0.355	0.682	0.414	0.647	0.743	0.391
CSC	0.460	0.472	**0.731**	**0.465**	0.398	**0.817**	0.292	0.442	0.311	0.387	0.939	0.218	0.181	0.302	**0.076**	**0.449**	**0.743**	0.430	0.444	0.737	0.368
ViewPointBN	0.452	0.587	0.569	0.172	0.391	0.769	0.290	**0.512**	0.501	0.373	0.935	0.251	0.173	0.201	0.003	0.352	0.619	0.454	**0.783**	0.719	0.390
Ours	**0.508**	**0.824**	0.53	0.314	**0.479**	0.746	**0.334**	0.49	**0.508**	**0.477**	**0.95**	**0.269**	**0.221**	**0.324**	0.029	0.421	0.626	**0.49**	0.727	**0.782**	**0.62**

Table G. 10% LR

rate	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.555	0.711	0.668	0.622	0.425	0.830	0.433	0.552	0.273	0.440	0.938	0.287	0.096	**0.470**	0.576	**0.612**	0.687	0.438	0.781	0.785	0.474
CSC	0.575	0.671	**0.740**	**0.727**	0.445	**0.847**	0.380	0.602	0.512	0.447	0.942	**0.291**	0.184	0.353	0.468	0.508	0.745	**0.602**	**0.855**	0.765	0.420
ViewPointBN	0.566	0.780	0.659	0.677	**0.484**	0.799	0.419	0.636	0.480	0.432	0.940	0.238	0.124	0.396	**0.609**	0.432	0.735	0.527	0.787	0.752	0.423
Ours	**0.608**	**0.853**	0.689	0.593	0.483	0.83	**0.466**	**0.652**	**0.528**	**0.482**	**0.954**	0.288	**0.25**	0.448	0.595	0.532	**0.748**	0.503	0.822	**0.806**	**0.647**

Table H. 20% LR

rate	avg.	bathtub	bed	book.	cabinet	chair	counter	curtain	desk	door	floor	otherf.	picture	frige	shower	sink	sofa	table	toilet	wall	window
PointContrast	0.603	0.740	0.700	0.700	0.546	0.843	0.419	0.592	0.462	0.513	0.946	0.374	0.104	0.530	0.687	0.571	0.694	0.519	0.850	0.781	0.484
CSC	0.612	0.739	**0.794**	0.687	0.564	**0.850**	0.347	0.590	**0.587**	0.521	0.945	0.358	0.140	0.522	0.496	**0.627**	0.725	**0.598**	0.850	0.792	0.508
ViewPointBN	0.625	**0.873**	0.727	0.709	0.535	0.820	0.402	0.643	0.540	0.501	0.946	0.352	**0.181**	0.535	0.594	0.596	0.685	0.543	**0.927**	0.792	0.592
Ours	**0.663**	0.851	0.77	**0.76**	**0.615**	0.83	**0.439**	**0.67**	0.546	**0.587**	**0.955**	**0.406**	0.177	**0.627**	**0.758**	0.606	**0.74**	0.549	0.888	**0.84**	**0.652**

Limited reconstruction We expand the Table 2 in our main paper and show the per-category results on *3D Semantic label with Limited Reconstructions benchmark* in Table E, Table F, Table G and Table H on 1%, 5%, 10% and 20% percentage of labeled scene. When only 1% of the training data is available, all methods achieve similar performance. When 5%,10%,20% of the training data is given, we outperform the previous state-of-the-art.

B Comparison with DepthContrast [5]

We use the released code and the pre-trained weight by DepthContrast to train on the efficient ScanNet setup. We note that DepthContrast uses Minkowski while we use O-CNN as the base model, so we also run the Minkowski without pre-training for a fair comparison. We present the results of the LR setup in Table I.

Table I. Comparison our pre-training with DepthContrast [5]. The results are presented as mIoU and compared on ScanNet validation set.

base model	pre-training	LR 1%	LR 5%	LR 10%	LR 20%
Minkowski	none	24.8	40.5	53.5	59.8
Minkowski	[5]	27.7	43.0	56.8	61.8
O-CNN	none	18.7	39.7	52.4	59.6
O-CNN	ours	26.3	44.9	56.5	62.7

Without any pre-training, Minkowski outperforms O-CNN. However, O-CNN with our pre-training can overtake Minkowski with DepthContrast pre-training in LR 5% and LR 20%, and achieve similar mIoU in LR 10%. The results show that the proposed pre-training can boost better than the previous work in the LR setup.

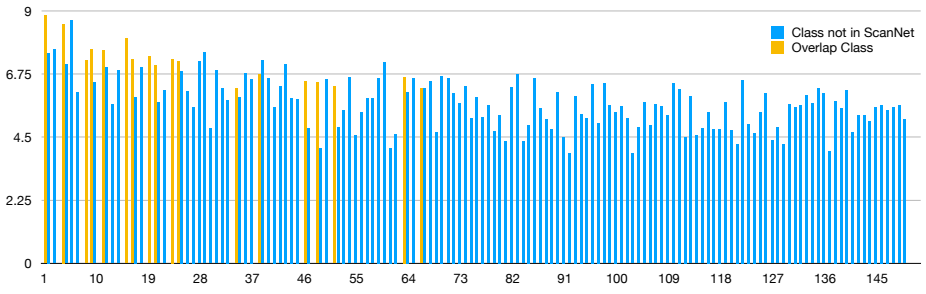
C Additional implement details

In semi-supervised learning, the loss of unlabeled data are weighted sum by different coefficients across downstream tasks. For object classification, the weights of mini-entropy loss and consistency loss referenced in the main paper are 0.01 and 10. For scene semantic segmentation, the weights of mini-entropy loss and consistency loss are 0.25 and 10. Among Mean-Teacher training, the weights of the teacher model were updated each training step by EMA with smoothing hyperparameter $\alpha = 0.999$, and the coefficient of consistency cost will ramp up during the first 30 epochs by a sigmoid-shaped function $e^{-5(1-x)^2}$, where $x \in [0, 1]$.

D Statistic of the pre-training pseudo-labels

We adopt DPT [3], which has been trained on the ADE20k [6] dataset, to predict the categorical probability distribution over 150 classes for all pixels. We sum the predicted probability of pixels in all images in Matterport3D [1] dataset for each class. The histogram of the summed pseudo-label probability are presented in Fig. A. The horizontal axis represents the 150 different classes in ADE20k and the vertical axis shows the log of summed probabilities, then the yellow bars are the overlap classes for ADE20k and ScanNet. As shown in Fig. A, the pseudo-labels for pre-training provide extensive knowledge beyond the target ScanNet dataset and is non-trivial to learn.

Fig. A. The histogram for the summed probability of the pre-training pseudo-label. The x-axis are the 150 different classes of a image scene parse, which is DPT in this work, well-trained on the ADE20k dataset.



E Qualitative results of data efficient on Scannet

Under limited annotations scenario, Figs. C and D are the results of O-CNN [4] models supervised by limited ScanNet annotation. Then Fig. E shows the results of O-CNN models supervised under limited ScanNet reconstruction. We show the visual comparisons between the results of training from scratch and training with our pre-training. We use red frames to highlight the difference, where our pre-training lead to more complete shapes and consistent prediction. The visual results echo our quantitative improvement, which shows that our pre-trained models improve the generalizability for scene understanding by the 2D transferred knowledge.

Fig. B. Corresponding colors to the categories in ScanNet.

Fig. C. Qualitative results for ScanNet LA

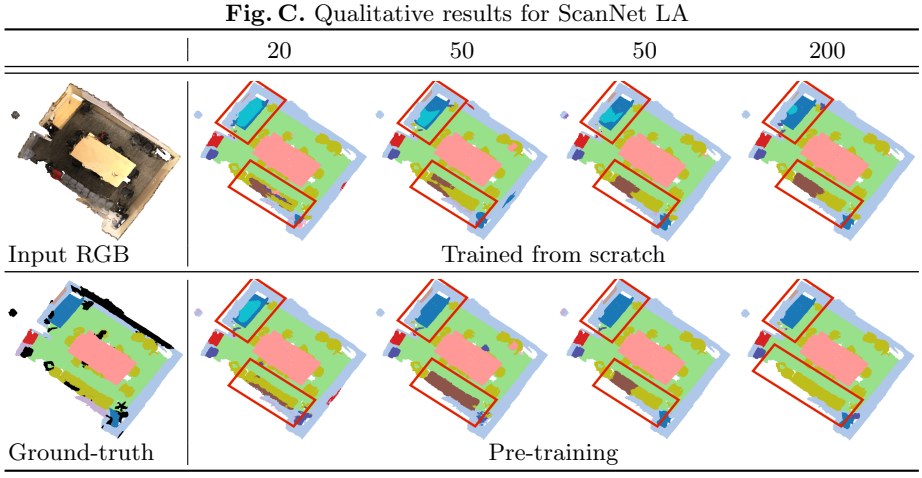


Fig. D. Qualitative results for ScanNet LA

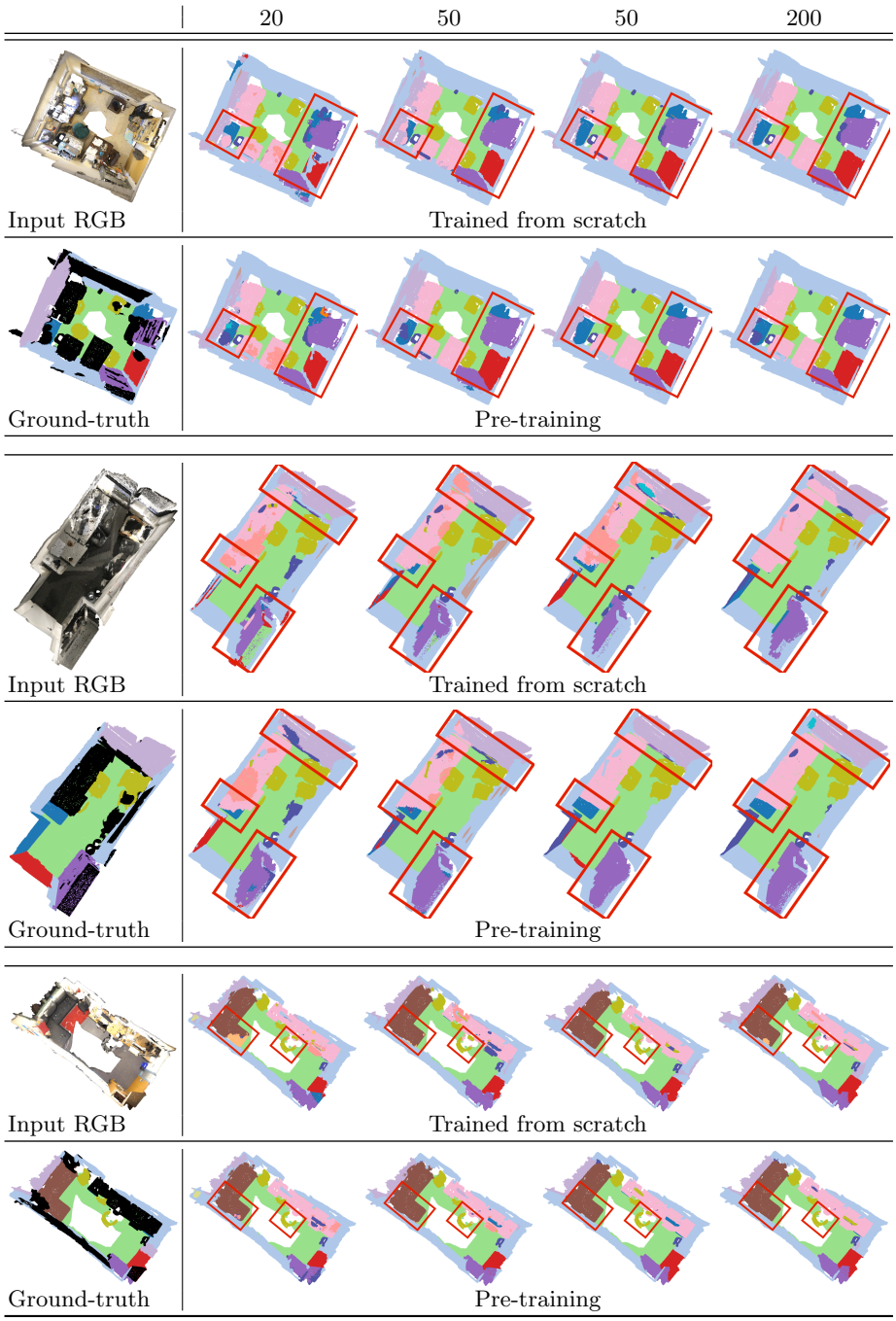
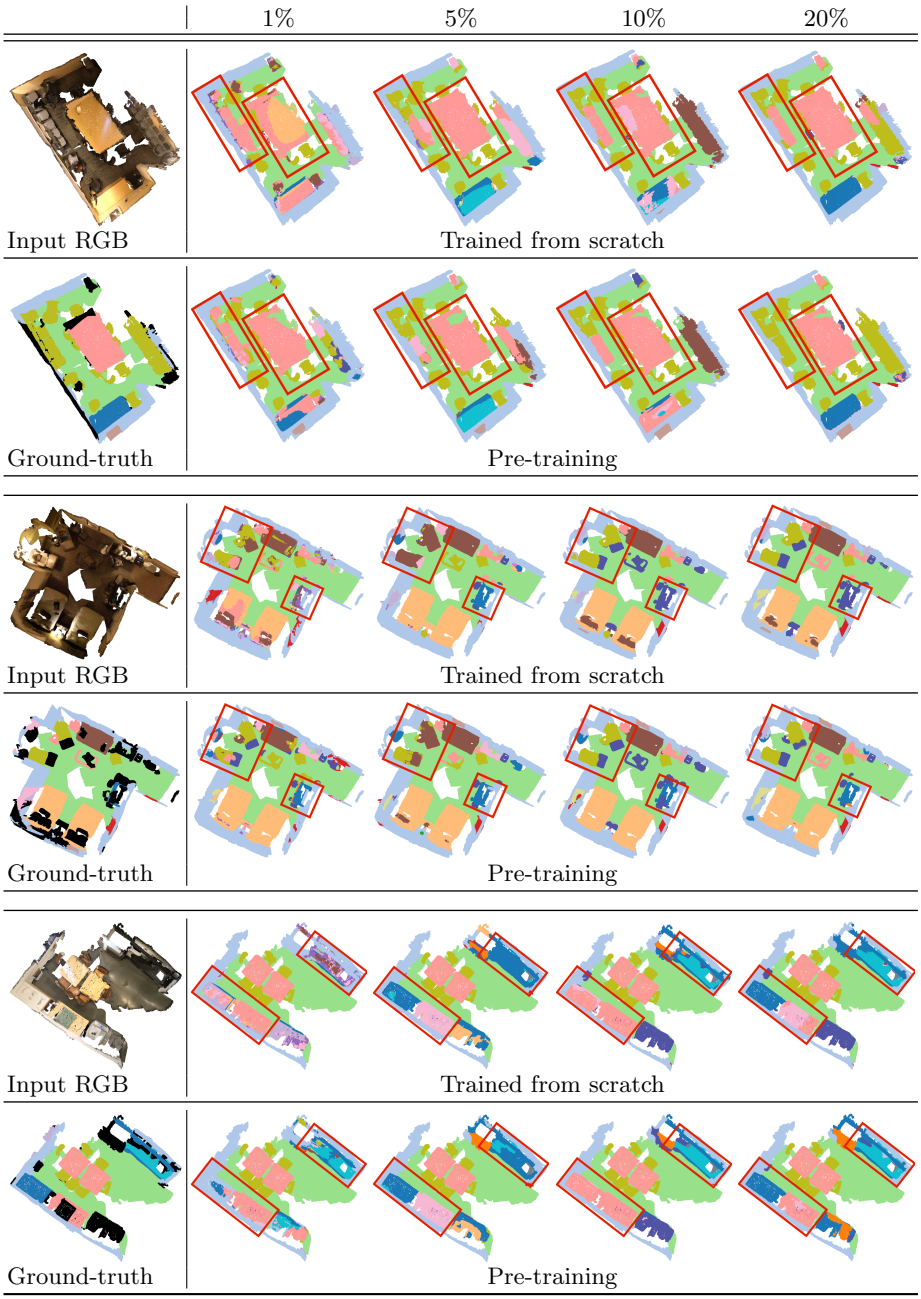


Fig. E. Qualitative results for ScanNet LR



References

1. Chang, A.X., Dai, A., Funkhouser, T.A., Halber, M., Nießner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3d: Learning from RGB-D data in indoor environments. In: 3DV. pp. 667–676 (2017) 3
2. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T.A., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: CVPR. pp. 2432–2443 (2017) 1
3. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. In: ICCV. pp. 12179–12188 (2021) 3
4. Wang, P., Liu, Y., Guo, Y., Sun, C., Tong, X.: O-CNN: octree-based convolutional neural networks for 3d shape analysis. ACM Trans. Graph. pp. 72:1–72:11 (2017) 4
5. Zhang, Z., Girdhar, R., Joulin, A., Misra, I.: Self-supervised pretraining of 3d features on any point-cloud. In: ICCV (2021) 1, 2
6. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ADE20K dataset. In: CVPR. pp. 5122–5130 (2017) 3