Towards Part-Based Understanding of RGB-D Scans

Alexey Bokhovkin[1,2] Vladislav Ishimtsev[2] Emil Bogomolov[2] Denis Zorin[3,2]
Alexey Artemov[2] Evgeny Burnaev[2] Angela Dai[1]

[1]TUM [2]Skoltech [3]NYU

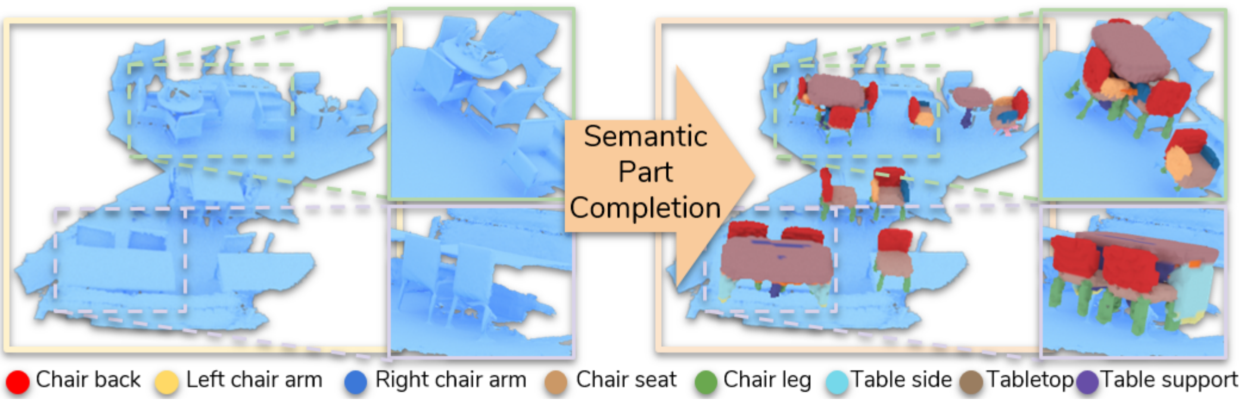

Figure 1: From an input RGB-D scan (left), we propose to detect objects in the scan and predict their complete part decompositions as *semantic part completion*; that is, we predict the part masks for the complete object, inferring the part geometry of any missing or unobserved regions in the scan. To achieve this, we predict the part structure of each detected object to drive a geometric prior-driven prediction of the complete part masks.

Abstract

Recent advances in 3D semantic scene understanding have shown impressive progress in 3D instance segmentation, enabling object-level reasoning about 3D scenes; however, a finer-grained understanding is required to enable interactions with objects and their functional understanding. Thus, we propose the task of part-based scene understanding of real-world 3D environments: from an RGB-D scan of a scene, we detect objects, and for each object predict its decomposition into geometric part masks, which composed together form the complete geometry of the observed object. We leverage an intermediary part graph representation to enable robust completion as well as building of part priors, which we use to construct the final part mask predictions. Our experiments demonstrate that guiding part understanding through part graph to part prior-based predictions significantly outperforms alternative approaches to the task of semantic part completion.

1. Introduction

Recently, we have seen remarkable advances in 3D semantic scene understanding, driven by efforts in large-scale data collection and annotation of 3D reconstructions of RGB-D scanned environments [4, 2], coupled with exploration of 3D deep learning approaches across 3D representations [43, 31, 4, 12, 30, 10, 6]. These have enabled a basis for 3D perception at the level of objects, which is essential for semantic understanding, but lacks finer-grained understanding often critical for enabling interactions with objects and reasoning about functionality.

At the same time, notable progress has been made in part segmentation for shapes [26, 25, 14]. However, these methods have been developed on synthetic datasets such as ShapeNet [3], of objects in isolation; this scenario is much less complex than the objects observed in real-world environments. Thus, we aim to bring these two directions together and propose the task of *semantic part completion*, predicting the part decomposition of objects in real-world 3D environments.

To achieve part-based understanding of a scene, we propose to predict the full part graph for each detected object, and based on the predicted part graph, the geometric masks for each complete part.

To address part-based understanding, we propose an approach to decompose a 3D scan of a scene into its complete object parts, outperforming state-of-the-art alternative approaches for the task.

2. Related Work

3D Object Detection and Instance Segmentation. Following the success of convolutional neural networks for object detection and instance segmentation in 2D images [9, 33, 32, 15], we are now seeing notable advances in 3D object localization and segmentation. Earlier approaches leveraging 3D convolutional neural networks developed methods operating on dense voxel grids using 3D region proposal techniques for detection and segmentation [37, 16]. There are also efficient methods based in sparse voxel backbones [8, 13]. These approaches have now shown impressive performance for instance-level scene understanding; we aim to build upon this and propose to infer finer-grained part decomposition for each object in a 3D scan.

3D Scan Completion. Repairing and completing holes or broken meshes has been well-studied for 3D shapes. There are methods that are focused on repairing small holes [27, 46, 39, 21, 22] as well as structural shape completion [40, 24, 28, 36, 39]. In addition to operating on the limited spatial context of shapes, generative deep learning approaches have also been developed for completion of 3D scenes [38, 7, 5, 17].

Part Segmentation of 3D Shapes. Understanding the structure of a 3D shape by identifying shape parts has been long-studied in shape analysis. Various approaches have been developed for finding a consistent segmentation across a set of shapes without supervision of part labels [11, 19, 35, 18]. To better capture more complex structures in the part layout of shapes, several methods propose to parse object parts as hierarchies [42, 41, 45, 26, 25]. We also adopt a relational inference of parts, but aim to operate on noisy, incomplete real-world scans of scenes with multiple objects, and so propose to combine our hierarchical part decomposition with strong geometric part priors.

3. Method

Overview We address the problem of simultaneous part segmentation and completion of objects of real-world RGB-D scans, which are often noisy and incomplete. An overview of our approach is illustrated on Fig. 2. Given an input 3D scan $\mathbb{S}$, we aim to predict a set of parts for each object in the scan, with each part representing the complete geometry of the part, including any missing or unobserved regions. From $\mathbb{S}$, we first detect a set of object instances $\mathbb{O} = \{o_i\}$ in the scene, as 3D bounding box locations and class category predictions. For each detected object in $\mathbb{O}$, we then convert it into a 32^3 occupancy grid representation, to inform our part segmentation and completion.

We then predict the part segmentation and completion for each detected object $o_i \in \mathbb{O}$, resulting in a set of volumetric binary part masks. First, for a detected object o_i, we predict its semantic part structure T_i, with elements representing part class types, and the adjacency relations between the parts. This enables encoding the high-level, semantic part structure of the shape, which both facilitates completion of the shape structure, as missing parts are easy to identify in their semantic part structure, as well as guides the prediction of the geometry of each part.

Object Detection From an input 3D scan, we first detect objects in the scene. We leverage a state-of-the-art 3D object detection approach, MLCVNet [44], as our object detection backbone. The input scan sampled to a point cloud, and object proposals are produced by voting [29], leveraging global contextual information at various scales.

Since our object detection backbone predicts axis-aligned bounding boxes for each object, we additionally predict the orientation r_i of each object o_i by classifying the angle in $n_\alpha = 8$ bins of discretized angles ($\{0°, 45°, \ldots, 315°\}$).

Semantic Part Decomposition For a detected object o_i from the scan, represented as a 32^3 occupancy grid of the scan geometry within its predicted bounding box, we aim to capture its high-level part structure from its cluttered and partial observation. We predict the semantic part structure T_i of the object; this facilitates completion of the object by predicting its high-level structure, as well as enables our prior-guided part geometry prediction.

We first encode the occupancy grid of o_i with four 3D convolutional blocks, and extract a feature encoding z_i of dimension 128. We then decode z_i into a semantic part prediction, constructing a part set T_i with each element represented by its predicted part category and a 128-dimensional feature encoding. Inspired by StructureNet [25], we leverage a message-passing graph neural network for our semantic part prediction which enables relational inference between semantic parts. From z_i, we predict part elements using an MLP to predict $n_{parts} = 10$ latent vectors $\{z'_k\}$ that correspond to potential parts of o each of which has a tuple $t_k = (e_k, s_k)$, where e_k is the probability of part existence, s_k is the one-hot representation of the part category label. We leverage this part semantic information to guide our final part decomposition as geometric part masks.

Prior-guided Geometric Part Decomposition We then predict the final part decomposition by generating complete part masks for each element in the predicted semantic part arrangement T_i. Rather than directly reconstructing the part

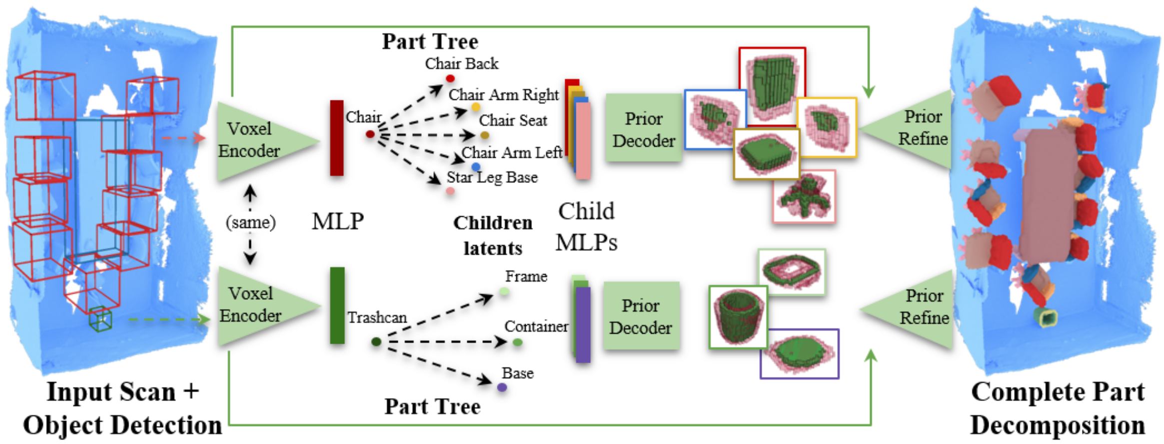

Figure 2: Overview of our approach. From an input scan, we detect objects as their 3D bounding boxes, and for each object (a chair and a trash can visualized top and bottom, respectively), we predict their semantic part structure, which is then used to guide a geometric prior-based part mask prediction. This results in a part decomposition of the scene where each object is decomposed into its complete part geometry, including any missing or unobserved regions.

geometry of each predicted part, we observe that object parts often maintain very similar geometry structures. That is, we construct geometric part priors to aid in generating our complete part mask predictions.

We construct our geometric part priors by k-means clustering of the binary part masks in the train set, inspired by the ShapeMask [23]. For each part type, we find $K = 10$ centroids of the part masks, and perform the clustering on the part masks in 32^3 grids of the canonical object space. This produces a set of part priors $\{P_1, \ldots, P_M\}$ with $M = n_{\text{classes}}K$. Examples of priors are visualized in Figure 2.

To predict the part geometry associated for an element in the predicted semantic part set T_i with feature encoding z'_k and predicted part type t, we use a one-layer MLP which takes as input z'_k and predicts a set of weights w_m used to construct an initial part reconstruction as $P_k^{\text{coarse}} = \sum_{m=1}^{M_t} w_m P_m^r$, where $w = \text{softmax}(\phi(z'_k))$, ϕ is a linear layer and r denotes rotated prior according to predicted orientation. We employ a proxy loss on this initial part reconstruction, using a mean squared error with a target part mask.

We then refine the predicted P_k^{coarse} using four 3D convolutional blocks taking as input the concatenation of the geometry of o_i and P_k^{coarse} to produce P_k^{refine}; we then obtain the final part mask prediction, $P_k = P_k^{\text{coarse}} + P_k^{\text{refine}}$.

Training Details In order to train our approach, we leverage the Scan2CAD dataset [1] in combination with PartNet [26]. Scan2CAD contains annotations of CAD models from ShapeNet [3] aligned to the 3D scans of ScanNet [4], and we use the part annotations of PartNet for these ShapeNet CAD models to obtain our ground truth part decompositions of the 3D scans. We train our part decomposition with an Adam optimizer, using a batch size of 24, learning rate of 0.001, and weight decay of 0.01.

4. Results

We evaluate our proposed approach in comparison to alternative approaches for semantic part completion on real-world RGB-D scans. We use scans from the ScanNet dataset [4], as well as the Scan2CAD [1] annotations of CAD model alignments from ShapeNet [3] to the ScanNet scans, coupled with the PartNet [26] annotations for the part decomposition.

Comparison to alternative approaches. In Table 1, we compare to several state-of-the-art approaches for part segmentation and scan completion, coupled together to provide a complete part decomposition of the objects in a scan. As an alternative approach for this task, we consider scan completion followed by object detection and part instance segmentation. We employ the state-of-the-art scan completion approach SG-NN [5] to generate a prediction for the complete geometry of a partial scan observation, and then apply the object detections of with MLCVNet [44], obtain a final complete part decomposition by the state-of-the-art instance segmentation of PointGroup [20]. We also compare to StructureNet [25] on MLCVNet detections. We additionally consider a UNet [34] composed of 3D volumetric convolutions as a baseline for the final part segmentation, which helps to indicate the performance of a similar approach without the use of geometric priors or semantic part relations before predicting the final part masks. These approaches do not consider explicit part structure reasoning, whereas our prediction of semantic parts and their relations helps to guide or prior-based decomposition for a more effective complete part decomposition.

	Chamfer Distance (↓)								IoU (↑)							
Method	chair	table	cab.	bkshlf	bed	bin	class avg	inst avg	chair	table	cab.	bkshlf	bed	bin	class avg	inst avg
SG-NN + MLCVNet + PointNet++	0.078	0.111	0.111	0.062	0.084	0.197	0.107	0.097	2.3	3.7	0.5	2.7	4.8	0.5	2.5	2.2
SG-NN + MLCVNet + UNet	0.050	0.118	0.080	0.053	0.083	0.108	0.082	0.073	17.5	6.4	7.6	12.4	13.3	13.9	11.9	13.3
SG-NN + MLCVNet + PointGroup	0.074	0.102	0.100	0.063	0.091	0.140	0.095	0.093	5.1	1.5	1.0	4.5	4.5	0.9	2.9	2.9
MLCVNet + StructureNet	**0.029**	0.095	**0.065**	0.037	0.076	0.106	0.068	0.057	13.8	0.5	3.8	9.0	3.9	9.3	6.8	8.9
Ours	0.033	**0.089**	0.069	**0.033**	**0.054**	**0.096**	**0.062**	**0.053**	**22.1**	**7.7**	**13.0**	**18.1**	**17.3**	**22.0**	**16.7**	**18.3**

Table 1: Evaluation on semantic part completion on Scan2CAD [1]. We compare with state-of-the-art approaches for scan completion [5], followed by object detection [44], and then part segmentation [20, 25, 30]. By leveraging part structures to guide our prior-based approach, we obtain more accurate part decompositions.

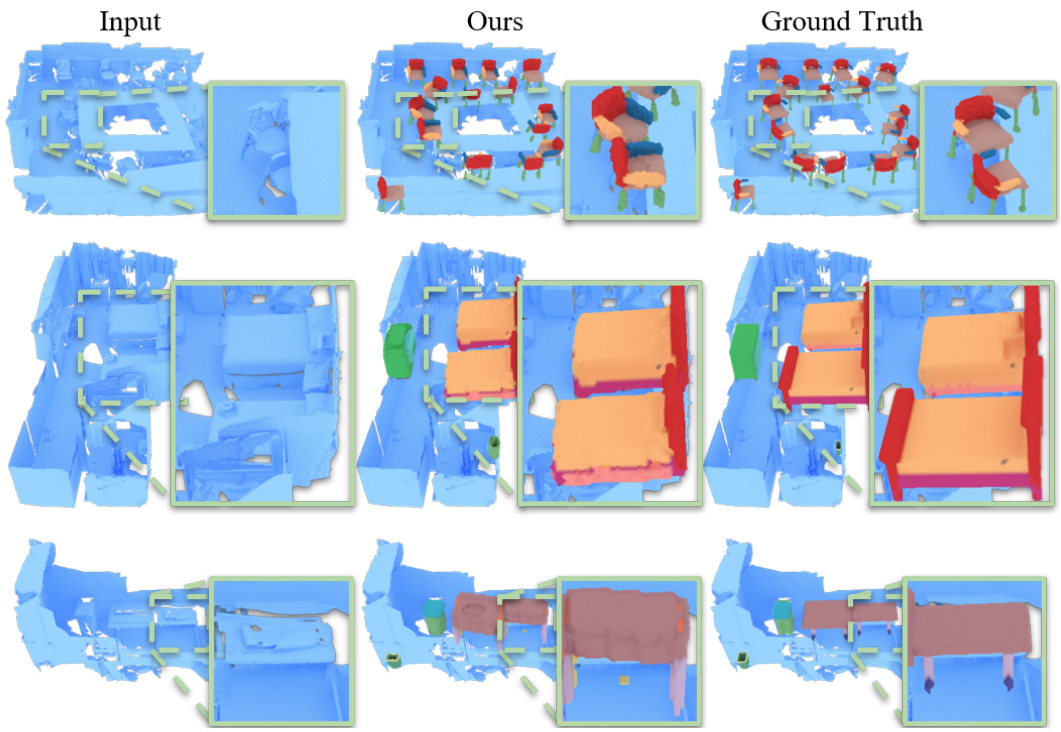

Figure 3: Qualitative results on real-world ScanNet [4] scenes using Scan2CAD [1] and PartNet [26] targets. Our approach effectively predicts each object's complete geometry as a decomposition into semantic parts.

In Figure 3, we show qualitative results: with part structure reasoning, our approach can predicted clean and non-intersecting parts, disregarding the balance across parts in terms of quantitative distribution in a dataset and in a single instance. Our part structure guided priors enable more effective and accurate part decompositions of the objects in the scenes in comparison to baselines.

Limitations. While our approach for semantic part completion shows promise towards a finer-grained, semantically part-based understanding of 3D environments, we believe there are many avenues for further development. For instance, a dense volumetric representation of parts may suffice for functionality analysis of furniture-type objects, but can struggle to generate very high resolution parts for small objects. Furthermore, objects are currently considered independently for each part decomposition, where relational inference between objects in a scene would help to explain noisy or unobserved part regions.

5. Conclusion

In this paper, we have presented a new approach for the semantic part completion task of predicting a geometrically complete part decomposition for each object in a 3D scan. We show that our structural and prior-guided reasoning about object parts notably outperforms alternative approaches on this task. We believe that our approach makes an important step towards part-based understanding of 3D environments, and opens up new possibilities for part-level functionality analysis, autonomous agent interactions with an environment, and more.

References

[1] Armen Avetisyan, Manuel Dahnert, Angela Dai, Manolis Savva, Angel X. Chang, and Matthias Nießner. Scan2cad: Learning CAD model alignment in RGB-D scans. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 2614–2623, 2019. 3, 4

[2] Angel X. Chang, Angela Dai, Thomas A. Funkhouser, Maciej Halber, Matthias Nießner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from RGB-D data in indoor environments. In *2017 International Conference on 3D Vision, 3DV 2017, Qingdao, China, October 10-12, 2017*, pages 667–676, 2017. 1

[3] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 1, 3

[4] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Niessner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 3, 4

[5] Angela Dai, Christian Diller, and Matthias Nießner. Sg-nn: Sparse generative neural networks for self-supervised scene completion of rgb-d scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2020. 2, 3, 4

[6] Angela Dai and Matthias Nießner. 3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 452–468, 2018. 1

[7] Angela Dai, Daniel Ritchie, Martin Bokeloh, Scott Reed, Jürgen Sturm, and Matthias Nießner. Scancomplete: Large-scale scene completion and semantic segmentation for 3d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2018. 2

[8] Francis Engelmann, Martin Bokeloh, Alireza Fathi, Bastian Leibe, and Matthias Nießner. 3d-mpa: Multi-proposal aggregation for 3d semantic instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9031–9040, 2020. 2

[9] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 2

[10] Georgia Gkioxari, Jitendra Malik, and Justin Johnson. Mesh r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9785–9795, 2019. 1

[11] Aleksey Golovinskiy and Thomas Funkhouser. Consistent segmentation of 3d models. *Computers & Graphics*, 33(3):262–269, 2009. 2

[12] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 9224–9232, 2018. 1

[13] Lei Han, Tian Zheng, Lan Xu, and Lu Fang. Occuseg: Occupancy-aware 3d instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2940–2949, 2020. 2

[14] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 1

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2

[16] Ji Hou, Angela Dai, and Matthias Nießner. 3d-sis: 3d semantic instance segmentation of rgb-d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4421–4430, 2019. 2

[17] Ji Hou, Angela Dai, and Matthias Nießner. Revealnet: Seeing behind objects in rgb-d scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2098–2107, 2020. 2

[18] Ruizhen Hu, Lubin Fan, and Ligang Liu. Co-segmentation of 3d shapes via subspace clustering. In *Computer graphics forum*, volume 31, pages 1703–1713. Wiley Online Library, 2012. 2

[19] Qixing Huang, Vladlen Koltun, and Leonidas Guibas. Joint shape segmentation with linear programming. In *Proceedings of the 2011 SIGGRAPH Asia Conference*, pages 1–12, 2011. 2

[20] Li Jiang, Hengshuang Zhao, Shaoshuai Shi, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Pointgroup: Dual-set point grouping for 3d instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4867–4876, 2020. 3, 4

[21] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006. 2

[22] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 2

[23] Weicheng Kuo, Anelia Angelova, Jitendra Malik, and Tsung-Yi Lin. Shapemask: Learning to segment novel objects by refining shape priors. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9207–9216, 2019. 3

[24] Niloy J Mitra, Leonidas J Guibas, and Mark Pauly. Partial and approximate symmetry detection for 3d geometry. *ACM Transactions on Graphics (TOG)*, 25(3):560–568, 2006. 2

[25] Kaichun Mo, Paul Guerrero, Li Yi, Hao Su, Peter Wonka, Niloy Mitra, and Leonidas J Guibas. Structurenet: Hierarchical graph networks for 3d shape generation. *arXiv preprint arXiv:1908.00575*, 2019. 1, 2, 3, 4

[26] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 909–918, 2019. 1, 2, 3, 4

[27] Andrew Nealen, Takeo Igarashi, Olga Sorkine, and Marc Alexa. Laplacian mesh optimization. In *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pages 381–389, 2006. 2

[28] Mark Pauly, Niloy J Mitra, Johannes Wallner, Helmut Pottmann, and Leonidas J Guibas. Discovering structural regularity in 3d geometry. In *ACM SIGGRAPH 2008 papers*, pages 1–11. 2008. 2

[29] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9277–9286, 2019. 2

[30] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 77–85, 2017. 1, 4

[31] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5648–5656, 2016. 1

[32] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 2

[33] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 2

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[35] Oana Sidi, Oliver van Kaick, Yanir Kleiman, Hao Zhang, and Daniel Cohen-Or. Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. In *Proceedings of the 2011 SIGGRAPH Asia Conference*, pages 1–10, 2011. 2

[36] Ivan Sipiran, Robert Gregor, and Tobias Schreck. Approximate symmetry detection in partial 3d meshes. In *Computer Graphics Forum*, volume 33, pages 131–140. Wiley Online Library, 2014. 2

[37] Shuran Song and Jianxiong Xiao. Deep sliding shapes for amodal 3d object detection in rgb-d images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 808–816, 2016. 2

[38] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1746–1754, 2017. 2

[39] Pablo Speciale, Martin R Oswald, Andrea Cohen, and Marc Pollefeys. A symmetry prior for convex variational 3d reconstruction. In *European Conference on Computer Vision*, pages 313–328. Springer, 2016. 2

[40] Sebastian Thrun and Ben Wegbreit. Shape from symmetry. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1824–1831. IEEE, 2005. 2

[41] Oliver Van Kaick, Kai Xu, Hao Zhang, Yanzhen Wang, Shuyang Sun, Ariel Shamir, and Daniel Cohen-Or. Co-hierarchical analysis of shape structures. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013. 2

[42] Yanzhen Wang, Kai Xu, Jun Li, Hao Zhang, Ariel Shamir, Ligang Liu, Zhiquan Cheng, and Yueshan Xiong. Symmetry hierarchy of man-made objects. In *Computer graphics forum*, volume 30, pages 287–296. Wiley Online Library, 2011. 2

[43] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1912–1920, 2015. 1

[44] Qian Xie, Yu-Kun Lai, Jing Wu, Zhoutao Wang, Yiming Zhang, Kai Xu, and Jun Wang. Mlcvnet: Multi-level context votenet for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10447–10456, 2020. 2, 3, 4

[45] Li Yi, Leonidas Guibas, Aaron Hertzmann, Vladimir G Kim, Hao Su, and Ersin Yumer. Learning hierarchical shape segmentation and labeling from online repositories. *arXiv preprint arXiv:1705.01661*, 2017. 2

[46] Wei Zhao, Shuming Gao, and Hongwei Lin. A robust hole-filling algorithm for triangular mesh. *The Visual Computer*, 23(12):987–997, 2007. 2