PartNet: A Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding

Kaichun Mo1, Shilin Zhu2, Angel X. Chang3, Li Yi1, Subarna Tripathi4, Leonidas J. Guibas1,5, Hao Su2
1Stanford University 2UC San Diego 3Simon Fraser University 4Intel AI Lab 5Facebook AI Research

https://cs.stanford.edu/~kaichun/partnet/

Abstract

We present PartNet: a consistent, large-scale dataset of 3D objects annotated with fine-grained, instance-level, and hierarchical 3D part information. Our dataset consists of 573,585 part instances over 26,671 3D models covering 24 object categories. This dataset enables and serves as a catalyst for many tasks such as shape analysis, dynamic 3D scene modeling and simulation, affordance analysis, and others. Using our dataset, we establish three benchmarking tasks for evaluating 3D part recognition: fine-grained semantic segmentation, hierarchical semantic segmentation, and instance segmentation. We benchmark four state-of-the-art 3D deep learning algorithms for fine-grained semantic segmentation and three baseline methods for hierarchical semantic segmentation. We also propose a baseline method for part instance segmentation and demonstrate its superior performance over existing methods.

1. Introduction

Being able to parse objects into parts is critical for humans to understand and interact with the world. People recognize, categorize, and organize objects based on the knowledge of their parts [10]. Many actions that people take in the real world require detection of parts and reasoning over parts. For instance, we open doors using doorknobs and pull out drawers by grasping their handles. Teaching machines to analyze parts is thus essential for many vision, graphics, and robotics applications, such as predicting object functionality [13, 14], human-object interactions [18], shape editing [28, 16], and shape generation [25, 41].

To enable part-level object understanding by learning approaches, 3D data with part annotations are in high demand. Many cutting-edge learning algorithms, especially for 3D understanding [45, 44, 30, 8], intuitive physics [27], and reinforcement learning [48, 29], require such data to train the networks and benchmark the performances. Researchers are also increasingly interested in synthesizing dynamic data through physical simulation engines [20, 43, 29]. Creation of large-scale animatable scenes will require a large amount of 3D data with affordances and mobility information. Object parts serve as a critical stepping stone to access this information. Thus it is necessary to have a large 3D object dataset with part annotation.

With the availability of the existing 3D shape datasets with part annotations [5, 3, 45], we witness increasing research interests and advances in 3D part-level object understanding. Recently, a variety of learning methods have been proposed to push the state-of-the-art for 3D shape segmentation [30, 31, 46, 19, 35, 24, 9, 39, 40, 42, 33, 7, 26, 23]. However, existing datasets only provide part annotations on relatively small numbers of object instances [5], or on coarse yet non-hierarchical part annotations [45], restricting the applications that involves understanding fine-grained and hierarchical shape segmentation.

In this paper, we introduce PartNet: a consistent, large-scale dataset on top of ShapeNet [3] with fine-grained, hierarchical, instance-level 3D part information. Collecting such fine-grained and hierarchical segmentation is challenging. The boundary between fine-grained part concepts are more obscure than defining coarse parts. Thus, we define a common set of part concepts by carefully examining the 3D objects to annotate, balancing over several criteria: well-
defined, consistent, compact, hierarchical, atomic and complete. Shape segmentation involves multiple levels of granularity. Coarse parts describe more global semantics and fine-grained parts convey richer geometric and semantic details. We organize expert-defined part concepts in hierarchical segmentation templates to guide annotation.

PartNet provides a large-scale benchmark for many part-level object understanding tasks. In this paper, we focus on three shape segmentation tasks: fine-grained semantic segmentation, hierarchical semantic segmentation, and instance segmentation. We benchmark four state-of-the-art algorithms on fine-grained semantic segmentation and propose three baseline methods for hierarchical semantic segmentation. We propose the task of part instance segmentation using PartNet. By taking advantages of rich shape structures, we propose a method that outperforms the existing baseline algorithm by a clear margin.

PartNet contains highly structured, fine-grained and heterogeneous parts. Our experiments reveal that existing algorithms developed for coarse and homogeneous part understanding do not work well on PartNet. First, small and fine-grained parts, e.g. door handles and keyboard buttons, are abundant and present new challenges for part recognition. Second, many geometrically similar but semantically different parts require more global shape context to distinguish. Third, understanding the heterogeneous variation of shapes and parts necessitate hierarchical understanding. We expect that PartNet could serve as a better platform for part-level object understanding in the next few years.

In summary, we make the following contributions:

- We introduce PartNet, consisting of 573,585 fine-grained part annotations for 26,671 shapes across 24 object categories. To the best of our knowledge, it is the first large-scale dataset with fine-grained, hierarchical, instance-level part annotations;
- We propose three part-level object understanding tasks to demonstrate the usefulness of this data: fine-grained semantic segmentation, hierarchical semantic segmentation, and instance segmentation;
- We benchmark four state-of-the-art algorithms for semantic segmentation and three baseline methods for hierarchical segmentation using PartNet;
- We propose the task of part instance segmentation on PartNet and describe a baseline method that outperforms the existing baseline method by a large margin.

2. Related Work

Understanding shape parts is a long-standing problem in computer vision and graphics. Lacking large-scale annotated datasets, early research efforts evaluated algorithm results qualitatively and conducted quantitative comparison on small sets of 3D models. Attene et al. [1] compared 5 mesh segmentation algorithms using 11 3D surface meshes and presented side-by-side qualitative comparison. Chen et al. [5] collected 380 surface meshes with instance-level part decomposition and proposed quantitative metrics for evaluation. Concurrently, Benhabiles et al. [2] proposed similar evaluation criteria and methodology. Kalogerakis et al. [17] further assigned semantic labels to the segmented components. Shape co-segmentation benchmarks [38, 11] were proposed to study co-segmentation among similar shapes.

Recent advances in deep learning have demonstrated the power and efficiency of data-driven methods on 3D shape understanding tasks. ShapeNet [3] collected a large-scale 3D CAD models from online open-sourced 3D repositories, including more than 3,000,000 models and 3,135 object categories. Yi et al. [45] took an active learning approach to annotate the ShapeNet models with semantic segmentation for 31,963 shapes covering 16 object categories. In their dataset, each object is usually decomposed into 2~5 coarse semantic parts. PartNet provides more fine-grained part annotations that contains 18 parts per shape on average.

Many recent works studied fine-grained and hierarchical shape segmentation. Yi et al. [44] leveraged the noisy part decomposition inputs in the CAD model designs to learn consistent shape hierarchies. Chang et al. [4] collected 27,477 part instances from 2,278 models covering 90 object categories and studied the part properties related to language. Wang et al. [37] collected 1,016 3D models from 10 object categories and trained neural networks for grouping and labeling fine-grained part components. A concurrent work [47] proposed a recursive binary decomposition network for shape hierarchical segmentation. PartNet provides a large-scale testbed with 573,585 fine-grained and hierarchical shape parts to support this direction of research.

There are also many previous works that attempted to understand parts by their functionality and articulation. Hu et al. [13] constructed a dataset of 608 objects from 15 object categories annotated with the object functionality and introduced a co-analysis method to learn category-wise object functionality. Hu et al. [12] proposed a dataset of 368 mo-

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Shape</th>
<th>#Part</th>
<th>#Category</th>
<th>Granularity</th>
<th>Semantics</th>
<th>Hierarchical</th>
<th>Instance-level</th>
<th>Consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. [5]</td>
<td>380</td>
<td>4,300</td>
<td>19</td>
<td>Fine-grained</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>MCI [37]</td>
<td>1,016</td>
<td>7,537</td>
<td>10</td>
<td>Fine-grained</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Chang et al. [4]</td>
<td>2,278</td>
<td>27,477</td>
<td>90</td>
<td>Fine-grained</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yi et al. [45]</td>
<td>31,963</td>
<td>80,323</td>
<td>16</td>
<td>Coarse</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>PartNet (ours)</td>
<td>26,671</td>
<td>573,585</td>
<td>24</td>
<td>Fine-grained</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1. Comparison to the other shape part datasets.
Figure 2. PartNet dataset. We visualize example shapes with fine-grained part annotations for the 24 object categories in PartNet.

<table>
<thead>
<tr>
<th>All</th>
<th>Bag</th>
<th>Bed</th>
<th>Bowl</th>
<th>Chair</th>
<th>Clock</th>
<th>Dish</th>
<th>Disp</th>
<th>Door</th>
<th>Ear</th>
<th>Faucet</th>
<th>Hat</th>
<th>Key</th>
<th>Knife</th>
<th>Lamp</th>
<th>Lap</th>
<th>Micro</th>
<th>Mug</th>
<th>Frid</th>
<th>Scis</th>
<th>Stora</th>
<th>Table</th>
<th>Trash</th>
<th>Vase</th>
<th>Bottle</th>
</tr>
</thead>
<tbody>
<tr>
<td>#P</td>
<td>32531</td>
<td>186</td>
<td>248</td>
<td>519</td>
<td>247</td>
<td>8176</td>
<td>624</td>
<td>241</td>
<td>1005</td>
<td>285</td>
<td>285</td>
<td>840</td>
<td>287</td>
<td>210</td>
<td>514</td>
<td>3408</td>
<td>485</td>
<td>268</td>
<td>252</td>
<td>247</td>
<td>127</td>
<td>2639</td>
<td>9906</td>
<td>378</td>
</tr>
<tr>
<td>#S</td>
<td>26671</td>
<td>146</td>
<td>212</td>
<td>464</td>
<td>208</td>
<td>6400</td>
<td>579</td>
<td>201</td>
<td>954</td>
<td>245</td>
<td>247</td>
<td>708</td>
<td>250</td>
<td>174</td>
<td>384</td>
<td>2271</td>
<td>453</td>
<td>212</td>
<td>212</td>
<td>207</td>
<td>88</td>
<td>2303</td>
<td>8309</td>
<td>340</td>
</tr>
<tr>
<td>#M</td>
<td>771</td>
<td>20</td>
<td>18</td>
<td>28</td>
<td>20</td>
<td>77</td>
<td>25</td>
<td>20</td>
<td>26</td>
<td>20</td>
<td>19</td>
<td>60</td>
<td>19</td>
<td>18</td>
<td>57</td>
<td>64</td>
<td>20</td>
<td>28</td>
<td>20</td>
<td>20</td>
<td>34</td>
<td>91</td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td>#PS</td>
<td>480</td>
<td>4</td>
<td>24</td>
<td>12</td>
<td>4</td>
<td>57</td>
<td>23</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>15</td>
<td>18</td>
<td>8</td>
<td>3</td>
<td>16</td>
<td>83</td>
<td>8</td>
<td>12</td>
<td>4</td>
<td>13</td>
<td>5</td>
<td>36</td>
<td>82</td>
<td>13</td>
</tr>
<tr>
<td>#PI</td>
<td>573K</td>
<td>664</td>
<td>9K</td>
<td>2K</td>
<td>615</td>
<td>17K</td>
<td>4K</td>
<td>2K</td>
<td>7K</td>
<td>2K</td>
<td>8K</td>
<td>1K</td>
<td>20K</td>
<td>3K</td>
<td>1K</td>
<td>80K</td>
<td>2K</td>
<td>839</td>
<td>2K</td>
<td>981</td>
<td>77K</td>
<td>177K</td>
<td>8K</td>
<td>5K</td>
</tr>
</tbody>
</table>

Table 2. PartNet statistics. Row #A, #S, #M respectively show the number of shape annotations, the number of distinct shape instances and the number of shapes that we collect multiple annotations. Row #PS and #PI show the number of different part semantics and part instances that we finally collect. Row Pmed and Pmax respectively indicate the median and maximum number of part instances per shape. Row Dmed and Dmax respectively indicate the median and maximum hierarchy depth per shape, with root node as depth 0.

bility units with diverse types of articulation and learned to predict part mobility information from a single static segmented 3D mesh. In PartNet, we assign consistent semantic labels that entail such functionality and articulation information for part components within each object category, which makes PartNet useful for such research.

3. Data Annotation

The data annotation is performed in a hierarchical manner. Expert-defined hierarchical part templates are provided to guarantee labeling consistency among multiple annotators. We design a single-thread question-answering 3D GUI to guide the annotation. We hire 66 professional annotators and train them for the annotation. The average annotation time per shape is 8 minutes, and at least one pass of verification is performed for each annotation to ensure accuracy.

3.1. Expert-Defined Part Hierarchy

Shape segmentation naturally involves hierarchical understanding. People understand shapes at different granularities. Coarse parts convey global semantics while fine-grained parts provide more detailed understanding. Moreover, fine-grained part concepts are more obscure to define than coarse parts. Different annotators have different knowledge and background so that they may name parts differently when using free-form annotation [4]. To address these issues, we introduce And-Or-Graph-style hierarchical templates and collect part annotations according to the predefined templates.

Since there are no standard rules of thumb for defining good templates, it is non-trivial to design good hierarchical part templates for a category. Furthermore, the requirement for the designed template to cover all variations of shapes and parts, makes the problem even more challenging. Below we summarize the criteria that we used to guide our template design:

- **Well-defined**: Part concepts are well-delineated such that parts are identifiable by multiple annotators;
- **Consistent**: Part concepts are shared and reused across different parts, shapes and object categories;
- **Compact**: There is no unnecessary part concept and part concepts are reused when it is possible;
- **Hierarchical**: Part concepts are organized in a taxonomy to cover both coarse and fine-grained parts;
- **Atomic**: Leaf nodes in the part taxonomy consist of primitive, non-decomposable shapes;
- **Complete**: The part taxonomy covers a heterogeneous variety of shapes as completely as possible.

Guided by these general principles, we build an And-Or-Graph-style part template for each object category. The templates are defined by experts after examining a broad variety of objects in the category. Each template is designed in a hierarchical manner from the coarse semantic parts to the fine-grained primitive-level components. Figure 3 (middle) shows the lamp template. And-nodes segment a part into small subcomponents. Or-nodes indicate subcategorization for the current part. The combination of And-nodes and Or-nodes allows us to cover structurally different shapes...
using the same template while sharing as much common part labels as possible. As shown in Figure 3 (left) and (right), both table lamps and ceiling lamps are explained by the same template through the first-level Or-node for lamp types.

Despite the depth and comprehensiveness of these templates, it is still impossible to cover all cases. Thus, we allow our annotators to improve upon the structure of the template and to annotate parts that are out of the scope of our definition. We also conduct template refinements to resolve part ambiguity after we obtain the data annotation according to the original templates. To systematically identify ambiguities, we reserve a subset of shapes from each class and collect multiple human annotations for each shape. We compute the confusion matrix among different annotators and address data inconsistencies. For example, we merge two concepts with high confusion scores or remove a part if it is frequently segmented in the wrong way. We provide more details about this in the supplementary material.

### 3.2. Annotation Interface

Figure 4 (a) shows our web-based annotation interface. Based on the template hierarchy, the annotation process is designed to be a single-thread question-answering workflow, traversing the template graph in a depth-first manner, as shown in Figure 4 (b). Starting from the root node, the annotator is asked a sequence of questions. The answers automatically construct the final hierarchical segmentation for the current shape instance. For each question, the annotator is asked to mark the number of subparts (And-node) or pick one among all subtypes (Or-node) for a given part. For each leaf node part, the annotator annotates the part geometry in the 3D interface. To help them understand the part definition and specification, we provide rich textual definitions and visual examples for each part. In addition, our interface supports cross-section and visibility control to annotate the interior structure of a 3D model.

The collected 3D CAD models often include original mesh subgroups and part information. Some of the grouping information is detailed enough to determine the final segmentation we need. Considering this, we provide the annotators with the original groupings at the beginning of the annotation, to speed up annotation. The annotators can simply select multiple predefined pieces to form a part of the final segmentation. We also provide mesh cutting tools to split large pieces into smaller ones following [5], when the original groupings are coarser than the desired segmentation, as shown in Figure 4 (c). The annotators draw boundary lines on the remeshed watertight surface [15] and the mesh cutting algorithm automatically splits the mesh into multiple smaller subcomponents.

In contrast to prior work, our UI is designed for operating directly on 3D models and collecting fine-grained and hierarchical part instances. Compared to Yi et al. [45] where the annotation is performed in 2D, our approach allows the annotators to directly annotate on the 3D shapes and thus be able to pick up more subtle part details that are hidden from 2D renderings. Chang et al. [4] proposes a 3D UI that paints regions on mesh surfaces for part labeling. However, their interface is limited to existing over-segmentations on part components and does not support hierarchical annotations.

### 4. PartNet Dataset

The final PartNet dataset provides fine-grained and hierarchical instance-level part segmentation annotation for 26,671 shapes with 573,585 part instances from 24 object categories. We select categories from ShapeNetCore [3] that 1) are mostly seen in indoor scenes; 2) contain interesting intra-class variation; and 3) provide a huge number of parts. We add 3 new object categories that are commonly present in indoor scenes (i.e. scissors, refrigerators, and doors) and augment 7 of the existing categories with more
5. Tasks and Benchmarks

We benchmark three part-level object understanding tasks: fine-grained semantic segmentation, hierarchical semantic segmentation and instance segmentation.

**Data Preparation.** We only consider parts that can be fully determined by their shape geometry\(^1\). In evaluation, we ignore parts that require additional information to identify, such as glass parts on cabinet doors which requires opacity to identify, and buttons on microwaves which requires texture or color information to distinguish it. We also remove infrequent parts from the evaluation due to the lack of data samples.

We sample 10,000 points from each CAD model with furthest point sampling and use the 3D coordinates as the neural network inputs for all the experiments in the paper. The proposed dataset is split into train, validation and test sets with the ratio 70%: 10%: 20%. The shapes with multiple human annotations are not used in the experiments.

5.1. Fine-grained Semantic Segmentation

Recent advances of 3D semantic segmentation [30, 31, 46, 19, 35, 24, 9, 39, 40, 42, 33, 7, 26, 23] have accomplished promising performance in coarse-level segmentation on the ShapeNet Part dataset [3, 45]. However, few work focus on the fine-grained 3D semantic segmentation, due to the lack of large-scale fine-grained dataset. With the help of the proposed PartNet dataset, researchers can now work on this more challenging task with little overhead.

Fine-grained 3D semantic segmentation requires recognizing and distinguishing small and similar semantic parts. For example, door handles are usually small, 77 out of 10,000 points on average in PartNet, but semantically important on doors. Beds have several geometrically similar parts such as side vertical bars, post bars and base legs. To recognize the subtle part details, segmentation systems need to understand them locally, through discriminative features, and globally, in the context of the whole shape.

**Benchmark Algorithms.** We benchmark four state-of-the-art semantic segmentation algorithms: PointNet [30], PointNet++ [31], SpiderCNN [42] and PointCNN [26]\(^2\). PointNet [30] takes unordered point sets as inputs and extracts features for shape classification and segmentation. To better learn local geometric features, PointNet++ [31] proposes a hierarchical feature extraction scheme. SpiderCNN [42] extends traditional convolution operations on 2D images to 3D point clouds by parameterizing a family of convolutional filters. To organize the unordered points into latent canonical order, PointCNN [26] proposes to learn \(\mathcal{X}\)-transformation, and applies \(\mathcal{X}\)-convolution operations on the canonical points.

We train the four methods on the dataset, using the default network architectures and hyperparameters described in their papers. Instead of training a single network for all object categories as done in most of these papers, we train a network for each category at each segmentation level. We input only the 3D coordinates for fair comparison\(^3\) and train the networks until convergence. More training details are described in the supplementary material.

---

1https://3dwarehouse.sketchup.com
2Although 3D models in ShapeNet [3] come with face normal, textures, material and other information, there is no guarantee for the quality of such information. Thus, we leave this as a future work.
3There are many other algorithm candidates: [46, 19, 35, 24, 9, 39, 40, 33, 7, 23]. We will host an online leadboard to report the performances.
4PointNet++ [31] and SpiderCNN [42] use point normals as additional inputs. For fair comparison, we only input the 3D coordinates.
The method does not perform well on small parts, such as the door handle on the door example, and visually similar parts, such as stair steps and the horizontal bars on the bed frame. How to learn discriminative features that better capture both local geometry and global context for these issues would be an interesting topic for future works.

5.2. Hierarchical Semantic Segmentation

Shape segmentation is hierarchical by its nature. We study hierarchical semantic segmentation that predicts semantic part labels in the entire shape hierarchies that cover both coarse- and fine-grained part concepts. A key problem towards hierarchical segmentation is how to leverage the rich part relationships on the given shape templates in the learning procedure. Recognizing a chair base as a swivel base significantly reduces the solution space for detecting more fine-grained parts such as central supporting bars, starbase legs and wheels. On the other hand, the lack of a chair back increases the possibility that the object is a stool. In contrast to Sec. 5.1 where we consider the problem at each segmentation level separately, hierarchical segmentation requires a holistic understanding on the entire part hierarchy.

Benchmark Algorithms. We propose three baseline methods to tackle hierarchical segmentation: bottom-up, top-down and ensemble. The bottom-up method considers only the leaf-node parts during training and groups the prediction of the children nodes to parent nodes as defined in the hierarchies in bottom-up inference. The top-down method learns a multi-labeling task over all part semantic labels on the tree and conducts a top-down inference by classifying coarser-level nodes first and then finer-level ones. For the ensemble method, we train flat segmentation at multiple levels as defined in Sec. 5.1 and conduct joint inference by calculating the average log-likelihood scores over all the root-to-leaf paths on the tree. We use PointNet++ [31] as the
backbone network in this work. Note that methods listed in Sec. 5.1 can also be used. More architecture and training details are described in the supplementary material.

### Evaluation and Results

Table 4 demonstrates the performances of the three baseline methods. We calculate mIoU for each part category and compute the average over all the tree nodes as the evaluation metric. The experimental results show that the three methods perform similarly with small performance gaps. The ensemble method performs slightly better over the other two, especially for the categories with rich structural and sub-categorization variation, such as chair, table and clock.

The top-down method only considers leaf-node parts in the training. Although the template structure is not directly used, the parent-node semantics of leaf nodes are implicitly encoded in the leaf-node part definitions. For example, the vertical bars for chair backs and chair arms are two different leaf nodes. The top-down method explicitly leverages the tree structures in both the training and the testing phases. However, prediction errors are accumulated through top-down inference. The ensemble method decouples the hierarchical segmentation task into individual tasks at multiple levels and performs joint inference, taking the predictions at all levels into consideration. Though demonstrating better performances, it has more hyper-parameters and requires longer training time for the multiple networks.

### 5.3. Instance Segmentation

The goal of instance segmentation is to detect every individual part instance and segment it out from the context of the shape. Many applications in computer graphics, vision and robotics, including manufacturing, assembly, interaction and manipulation, require the instance-level part recognition. Compared to detecting objects from scenes, parts on objects usually have stronger and more intertwined structural relationships. The existence of many visually similar but semantically-different parts makes the part detection problem challenging. To the best of our knowledge, this work is the first to provide a large-scale shape part instance-level segmentation benchmark.

---

3In our experiments, PointNet++ and PointCNN give the top ranked performance under two different evaluation metrics: part-category mIoU (Table 3) and shape mIoU (Table 2 in supplementary material). We choose PointNet++ because previous works on ShapeNet mostly use shape mIoU as the metric. We reported part-category mIoU in the main paper to make it consistent with mIoU and mAP evaluation metrics used in ScanNet [6].
Table 5. Instance segmentation results (part-category mAP %, IoU threshold 0.5). Algorithm S and O refer to SGPN [36] and our proposed method respectively. The number 1, 2 and 3 refer to the three levels of segmentation: coarse-, middle- and fine-grained.

<table>
<thead>
<tr>
<th></th>
<th>Avg</th>
<th>Bag</th>
<th>Bed</th>
<th>Bott</th>
<th>Bowl</th>
<th>Chair</th>
<th>Clock</th>
<th>Disp</th>
<th>Door</th>
<th>Ear</th>
<th>Fauc</th>
<th>Hat</th>
<th>Key</th>
<th>Knife</th>
<th>Lamp</th>
<th>Lap</th>
<th>Micro</th>
<th>Mug</th>
<th>Frld</th>
<th>Scis</th>
<th>Stora</th>
<th>Table</th>
<th>Trash</th>
<th>Vase</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td></td>
<td>55.7</td>
<td>38.8</td>
<td>29.8</td>
<td>61.9</td>
<td>56.9</td>
<td>72.4</td>
<td>20.3</td>
<td>72.2</td>
<td>89.3</td>
<td>49.0</td>
<td>57.8</td>
<td>63.2</td>
<td>68.7</td>
<td>20.0</td>
<td>63.2</td>
<td>32.7</td>
<td>100</td>
<td>50.6</td>
<td>82.2</td>
<td>50.6</td>
<td>71.7</td>
<td>32.9</td>
<td>49.2</td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td>29.7</td>
<td>15.4</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
<td>58.1</td>
<td>25.4</td>
</tr>
<tr>
<td>S3</td>
<td></td>
<td>29.5</td>
<td>11.8</td>
<td>45.1</td>
<td>19.4</td>
<td>38.2</td>
<td>38.3</td>
<td>78.8</td>
<td>15.4</td>
<td>35.9</td>
<td>37.8</td>
<td>38.3</td>
<td>78.8</td>
<td>15.4</td>
<td>38.3</td>
<td>78.8</td>
<td>15.4</td>
<td>38.3</td>
<td>78.8</td>
<td>15.4</td>
<td>38.3</td>
<td>78.8</td>
<td>15.4</td>
<td>38.3</td>
</tr>
<tr>
<td>Avg</td>
<td></td>
<td>46.8</td>
<td>38.8</td>
<td>19.0</td>
<td>53.5</td>
<td>36.9</td>
<td>39.1</td>
<td>19.3</td>
<td>56.2</td>
<td>84.0</td>
<td>29.9</td>
<td>46.9</td>
<td>50.5</td>
<td>68.7</td>
<td>20.0</td>
<td>50.7</td>
<td>22.9</td>
<td>100</td>
<td>44.2</td>
<td>82.2</td>
<td>30.3</td>
<td>71.7</td>
<td>28.3</td>
<td>27.5</td>
</tr>
<tr>
<td>O1</td>
<td></td>
<td>62.6</td>
<td>64.7</td>
<td>48.4</td>
<td>63.6</td>
<td>59.7</td>
<td>74.4</td>
<td>42.8</td>
<td>76.3</td>
<td>93.3</td>
<td>52.9</td>
<td>57.7</td>
<td>69.6</td>
<td>70.9</td>
<td>43.9</td>
<td>58.4</td>
<td>37.2</td>
<td>100</td>
<td>50.0</td>
<td>86.0</td>
<td>50.0</td>
<td>80.9</td>
<td>45.2</td>
<td>54.2</td>
</tr>
<tr>
<td>O2</td>
<td></td>
<td>37.4</td>
<td>23.0</td>
<td>15.5</td>
<td>62.8</td>
<td>39.7</td>
<td>26.9</td>
<td>47.8</td>
<td>35.2</td>
<td>35.0</td>
<td>31.0</td>
<td>26.9</td>
<td>47.8</td>
<td>35.2</td>
<td>35.0</td>
<td>31.0</td>
<td>26.9</td>
<td>47.8</td>
<td>35.2</td>
<td>35.0</td>
<td>31.0</td>
<td>26.9</td>
<td>47.8</td>
<td>35.2</td>
</tr>
<tr>
<td>O3</td>
<td></td>
<td>36.6</td>
<td>15.0</td>
<td>48.6</td>
<td>29.0</td>
<td>32.3</td>
<td>53.3</td>
<td>80.1</td>
<td>17.2</td>
<td>39.4</td>
<td>44.7</td>
<td>45.8</td>
<td>18.7</td>
<td>34.8</td>
<td>26.5</td>
<td>27.5</td>
<td>23.9</td>
<td>33.7</td>
<td>52.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td></td>
<td>54.4</td>
<td>64.7</td>
<td>28.8</td>
<td>56.1</td>
<td>39.7</td>
<td>46.3</td>
<td>37.5</td>
<td>64.1</td>
<td>86.7</td>
<td>36.6</td>
<td>48.5</td>
<td>57.1</td>
<td>70.9</td>
<td>43.9</td>
<td>52.1</td>
<td>27.6</td>
<td>100</td>
<td>44.2</td>
<td>86.0</td>
<td>37.2</td>
<td>80.9</td>
<td>35.9</td>
<td>36.4</td>
</tr>
</tbody>
</table>

Figure 7. Qualitative results for instance segmentation. Our method produces more robust and cleaner results than SGPN.

Figure 8. Learned instance correspondences. The corresponding parts are marked with the same color.

Finally, we take the average of the mAP scores across different levels of segmentation within each object category and then report the average over all object categories. We compute the IoU between each prediction mask and the closest ground-truth mask and treat a prediction mask as a true positive when the IoU is larger than 0.5.

Figure 7 shows qualitative comparisons for our proposed method and SGPN. Our method produces more robust and cleaner instance predictions. After learning for point features, SGPN has a post-processing stage that merges points with similar features as one component. This process involves many hyper-parameter tuning. Even though most parameters are automatically inferred from the validation data, SGPN still suffers from predicting partial or noisy instances in case of bad thresholding. Our proposed method learns structural priors within each object category that is more instance-aware and robust in predicting complete instances. We observe that training for a set of disjoint masks across multiple shapes gives us consistent part instances. We show the learned part correspondence in Figure 8.

6. Conclusion

We introduce PartNet: a large-scale benchmark for fine-grained, hierarchical, and instance-level 3D shape segmentation. It contains 573,585 part annotations for 26,671 ShapeNet [3] models from 24 object categories. Based on the dataset, we propose three shape segmentation benchmarks: fine-grained semantic segmentation, hierarchical semantic segmentation and instance segmentation. We benchmark four state-of-the-art algorithms for semantic segmentation and propose a baseline method for instance segmentation that outperforms the existing baseline method.

Acknowledgements

This research was supported by NSF grants CRI-1729205, IIS-1763268 and CHS-1528025, a Vannevar Bush Faculty Fellowship, a Google fellowship, and gifts from Autodesk, Google and Intel AI Lab. We especially thank Zhe Hu from Hikvision for the help on data annotation and Lifeng Zhao for the help on preparing hierarchical templates. We appreciate the 66 annotators from Hikvision, Ytuuu and Data++ on data annotation.
References


