# Supplementary Material for PartAfford: Part-level Affordance Discovery from Cross-category 3D Objects

Chao Xu[1] Yixin Chen[1] He Wang[2,4]
Song-Chun Zhu[2,3,4] Yixin Zhu[2,4] Siyuan Huang[4]

[1] University of California, Los Angeles [2] Peking University
[3] Tsinghua University [4] Beijing Institute for General Artificial Intelligence

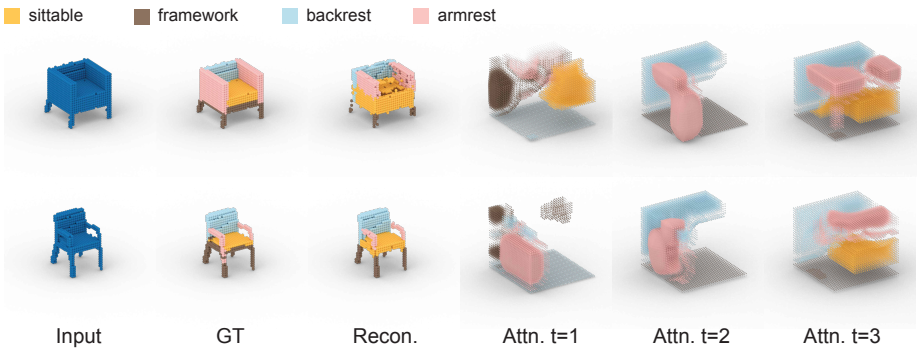


Fig. 1: Example affordance discovery model trained with $T = 3$ attention iterations. Attention is visualized in various colors and point radius. Point is colored according to the predicted affordance label of the slot which attends the most to the point. Point radius positively depends on the maximal attention value across the slots. We use trilinear upsampling to rescale the attention mask to the input resolution $(32 \times 32 \times 32)$.

## 1 Code and Data

Code, data, and instructions needed to reproduce the main experimental results are available in the folder.

## 2 Attention Visualization

Following the settings in [2], we train the models using $T = 3$ attention iterations in the slot attention module. In Fig. 1, we observe that each slot gradually attends to the correct part of the object as the number of attention iterations increases. The attention mask at $t = 3$ depicts the silhouette of the reconstructed shape and parts.

# 3 More Module Ablations

## 3.1 Soft k-means

Slot attention, as a generalized soft k-means algorithm, could be reduced to the soft k-means algorithm according to [2]. It turns out that the reduced model can achieve 23.0% Mean IoU on "sittable" objects, which is slightly better than the Slot MLP baseline but significantly worse than our full model using slot attention (57.3%). It reconstructs the whole shape with a quality (MSE: 0.0096) on par with the full model (MSE: 0.0097). Its affordance set prediction accuracy is the lowest among all models (AP: 0.87). Overall, the soft k-means algorithm cannot effectively segment the shape or discover affordance parts.

## 3.2 Cuboid $L_1$ Norm

As shown in Sec. 4.2 of the main paper, we predict the cuboid scale vector $\boldsymbol{s}^m \in \mathbb{R}^3$ for the $m$-th slot. We regularize the cuboid loss by adding a cuboid scale penalty term, *i.e.*, $L_1$ norm for the scale vector as follows:

$$\mathcal{L}_{\text{scale}} = \sum_m \|\boldsymbol{s}^m\|_1. \tag{1}$$

Without the cuboid $\mathcal{L}_1$ norm, some cuboids may not tightly wrap the reconstructed part if it is too slim.

# 4 Additional Experimental Results

We show additional qualitative results for our full model's affordance discovery results on "sittable" (Fig. 2), "support" (Fig. 3), "openable" (Fig. 4) objects, respectively.

# 5 Part Affordance Dataset

Here we report more details on how we define and annotate the affordance labels in our dataset.

## 5.1 Principles for Affordance Annotation

To keep annotations consistent across object categories, we design a guideline for affordance annotation. Below are some general principles to annotate a leaf part of an object instance with affordances.

*Multiple affordances.* A part can afford multiple kinds of human actions. For example, the seat of chair could afford *sittable* for resting of human body or *support* if one wants to place some books on it. We refer to ConceptNet [4], a giant knowledge database, for annotating common usage of object parts.

*Prioritized fine-grained affordances.* When there are multiple affordances labels for a part, we give the more fine-grained affordance label higher priority. For the chair seat in the example above, *sittable* is prioritized compared with *support* as it is a fine-grained support affordance for body resting.

*Articulation-related affordances.* The PartNet dataset does not contain articulation information, which makes affordances such as *openable* not geometrically distinguishable. Thus, we also generate a set of shapes with *openable* affordance from the PartNet-Mobility dataset by capturing 3D shapes with various opening angles. More geometric variation helps models to learn articulation-related affordances.

### 5.2 Affordance Descriptions

Our description for each affordance contains a brief definition, some supplemental clarification and priority statements if needed, and some example leaf nodes in the part hierarchy of various reasonable objects (full path from root to leaf).

***sittable*:** Indicates whether the object can be used for sitting. Anything sittable of course affords support, and the requirement for a supporting object to be sittable is that it must be both comfortable and safe for human seating. For example, a table is not sittable despite affording support because it is not comfortable. *Sittable* is given priority over potential co-existing affordances like *support*.

*E.g.*, chair/chair_seat/seat_surface.

***support*:** A trait of objects which can safely keep other objects on top of themselves. Common characteristics of support-affording objects are that they are flat and can support multiple objects at once. A key distinction between support-affording objects and non-supporting objects is that support-affording objects will remain stable when other objects are placed on them. For example, a table is a supporting object because it is just as stable with objects on it as it is without. However, a stack of plates is not supporting because they become more unstable as you add more plates.

*E.g.*, table/regular_table/tabletop,
storage_furniture/cabinet/shelf,
bed/bed_unit/bed_sleep_area.

***openable*:** Parts which may be moved with a hinge-like mechanism on an articulated object. Openable objects do not need to afford handles, but usually handles can be found attached to the *openable* part. *Openable* parts are distinct from other moving parts in that they need to swing to some degree to be moved. *Openable* is given priority over potential co-existing affordances like *containment*.

*E.g.*, door/door_body/surface_board,
dish_washer/body/door/door_frame,
microwave/body/door.

***backrest*:** Objects which are reasonably designed for providing support to a person's back. By reasonably designed, we mean either specifically (as in the back of a chair) or can afford back support if a person wanted to sit upright (like a headboard).

*E.g.*, chair/chair_back/back_support.

***armrest*:** Objects which are specifically designed to support an arm. For example, a table can support a variety of things and is thus not an armrest. A chair's arm is the perfect size for a human arm, so it must be an armrest.

*E.g.*, chair/chair_arm.

***handle*:** An object extension which affords the ability to open an attached 'openable' part. Handles are mostly grabbed with hand-wrapping, so it's important to only afford 'handle' to parts which are specifically involved in an opening mechanism, like a door handle.

*E.g.*, storage_furniture/cabinet/cabinet_door/handle,
table/regular_table/table_base/drawer_base/
cabinet_door/handle,
mug/handle,
dish_washer/body/door/handle,
bag/bag_handle.

***framework*:** Any object segment which either: a) helps to define the shape of the object as a whole or b) is an unaffording extension of the object or connector for other segments. For example, a hinge affords framework because it is integral to connects the door to the door frame. Overall, this affordance is the most general of all, so it should only be used when it clearly applies to either case of the definition. For example, most handles do not afford framework because it is both a small part of an object's shape and already affords *handle*. It has the least priority among potential co-existing affordances.

*E.g.*, chair/chair_base.

***containment*:** An affordance of object which can store physical items. The size of the physical items does not matter, so long as they are not too small. For example, anything that can only contain objects smaller than say a marble do not afford containment. Also, items that afford containment must afford security to the items they contain, such that they will not fall out.

*E.g.*, table/regular_table/table_base/drawer_base/drawer,
storage_furniture/cabinet/drawer,
mug/container,
trash_can/container,
refrigerator/body,
bowl/container, bag/bag_body,
bottle/normal_bottle/body.

***liquidcontainment*:** A more specific version of the containment affordance. Objects that afford liquid-containment must be able to safely contain liquid. Examples of these are bottles, bath tubs, *etc.*

*E.g.*, mug/body,
bottle/normal_bottle/body.

***display*:** Something which visualizes information for a useful purpose. Examples of these would be monitor screens or a clock surface.

*E.g.*, display/display_screen/screen,
laptop/screen_side/screen,
clock/table_clock/clock_body/surface.

***cutting*:** The quality of being able to slice through other objects. Certain things that can cut are not considered to have *cutting* affordance if it was used against its intended purpose, like smashing a glass vase. Cutting is only afforded to objects which are specifically designed for cutting, like a blade-edge.

*E.g.*, cutting_instrument/knife/blade_side,
scissors/blade_handle_set/blade.

***pressable*:** A mechanical feature of objects which either have buttons or can interact with a finger. Good examples of these are keyboard keys.

*E.g.*, keyboard/key.

***hanging*:** A part which can be hung on another object. These parts almost always only serve the purpose of hanging the rest of the entire object. An example of this would be a shoulder strap for a handbag.

*E.g.*, bag/shoulder_strap.

***wrapgrasp*:** The 'wrap-grasp' trait is afforded by parts which are explicitly meant to be grabbed in a hand-wrapping motion. Just because a hand can wrap around an object part does not mean it affords wrap-grasp. It must be useful to grip the part in this way. An example of this would be a ladder rung, which a person is meant to wrap their hand around to climb the ladder.

*E.g.*, cup,
bed/ladder/rung.

***illumination*:** The affordance of light emission. This only applies to object parts which are meant to light up a broad area. For example, a monitor screen does not afford illumination despite emitting light because it is not supposed to be used to light up the area around it.

*E.g.*, lamp/table_or_floor_lamp/lamp_unit
/lamp_head/light_bulb.

***lyable*:** Indicates that a human can comfortably rest his/her entire body on the object. These objects are usually flat with a soft surface. *Lyable* is given priority over potential co-existing affordances such as *sittable* and *support*.

*E.g.*, bed/bed_unit/bed_sleep_area/mattress.

***headrest*:** An extension of an object which is oriented so a human head can rest comfortably on it. Examples of these are chair headrests or bedframe headrests.

*E.g.*, bed/bed_unit/bed_sleep_area_pillow,
bed/bed_unit/bed_frame/headboard.

***step*:** A part which affords the human foot climbing or resting functionality. For example, a ladder rung is a step because it affords climbing with both hands and feet. A foot pedestal is also a step because it can be stood on or feet can be rested on it.

*E.g.*, bed/ladder/rung.

***pourable*:** Meant for parts which liquid can flow out of. Things that are pourable may also be dependent on a mechanism for controlling flow, like a bottle cap or a knob.

*E.g.*, bottle/normal_bottle/mouth,
bottle/jug/body.

***twistable*:** These objects can either be detached or provide special functionality by twisting them in a clockwise or counterclockwise motion. Examples include bottle caps and knobs.

*E.g.*, bottle/normal_bottle/lid,
bottle/normal_bottle/mouth.

***rollable*:** A part which can roll to move around. Exceptions to this affordance are objects which roll but stay fixed in place, like a rocking chair.

*E.g.*, wheel.

***lever*:** Any handle which can rotate up to a point. For example, knobs rotate but are not levers because they do not provide handles. Levers must be treated differently from twistable objects or handles because if they are twisted too much they will break.

*E.g.*, lever.

***pinchable*:** An object which is small enough such that it can be manipulated by pinching with two or more fingers. Things that are pinchable must not be heavy, and they usually fit inside the palm of a hand.

*E.g.*, earbud.

***audible*:** Anything which emits sound. This does not include sound emitted indirectly, such as a door creaking when opened, which makes sound as a side-effect.

*E.g.*, headphone/padding.

### 5.3 License

Our dataset is annotated based on PartNet (v0) [3] and PartNet-Mobility (v2.0) [5], both of which are licensed under the terms of the MIT License.

## 6 Supervised Affordance Estimation

Apart from the *PartAfford* task, we also benchmark the state-of-the-art 3D auto-encoder for the supervised affordance segmentation task in a cross-category fashion on our proposed dataset. Evaluation is reported on the Intersection-over-Union (IoU) metric following [3], as shown in Table 1.

Note that the baseline is trained and tested on all object and affordance categories. In contrast, semantic segmentation in PartNet [1,3] is trained on each object category separately. Results show that the algorithm could still achieve high performance, which demonstrates that our affordance annotation is reasonable and consistent.

PartNet normalizes every object shape into a unit bounding sphere. This is not realistic since objects in different categories may have very different dimensions. For example, a mug is much smaller than a bed. If normalized in

| | sittable | support | frame. | contain. | liquid. | openable | display | cutting |
|---|---|---|---|---|---|---|---|---|
| Orig. | 72.44 | 82.92 | 74.74 | 54.01 | 20.34 | 58.62 | 87.26 | **83.29** |
| Scaled | **74.95** | **83.68** | **77.87** | **75.01** | **69.42** | **66.26** | **90.03** | 79.67 |
| | backrest | armrest | press. | handle | illum. | wrapgrasp | lyable | headrest |
| Orig. | 80.51 | 73.19 | 78.70 | 38.09 | 24.57 | 63.07 | **46.24** | **34.73** |
| Scaled | **81.36** | **73.63** | **83.57** | **49.31** | **24.59** | **71.48** | 45.32 | 32.76 |
| | rollable | pourable | twist. | lever | pinch. | audible | | |
| Orig. | 53.33 | 68.15 | 40.24 | 60.34 | 62.07 | **56.67** | | |
| Scaled | **58.43** | **73.89** | **53.97** | **64.24** | **66.84** | 46.11 | | |

Table 1: Supervised affordance segmentation results (category mIoU %). "Orig." refers to the original PartNet dataset with 3D shapes scaled to a unit bounding sphere. "Scaled" refers to rescaling 3D shapes to real-world dimensions. "avg" refers to shape average Intersection-over-Union (IoU).

the same way, cross-category training performance may be hurt to some extent. Thus, we calculate the average real-world 3D dimensions of each object category from metadata of ShapeNet [1] and scale the point cloud in our part affordance dataset according to its object category.


Fig. 2: Additional qualitative results on objects in the "sittable" subset.

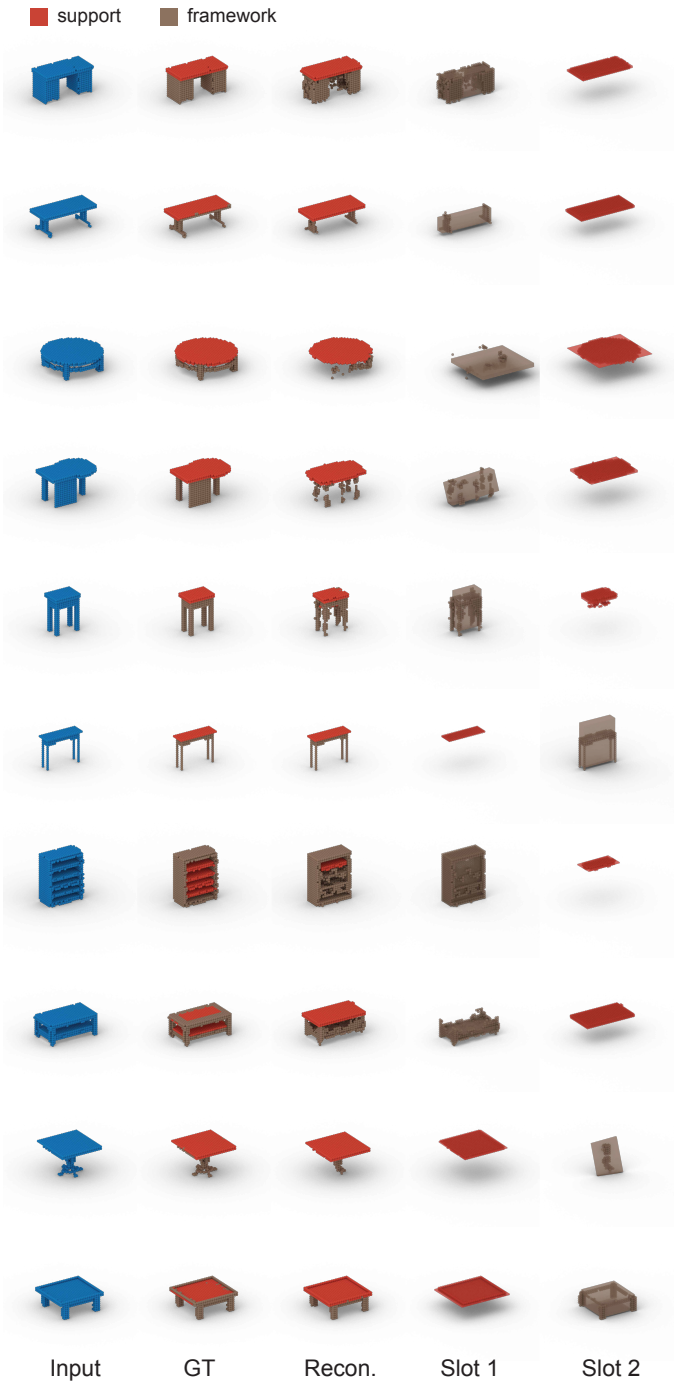


Fig. 3: Additional qualitative results on objects in the "support" subset.

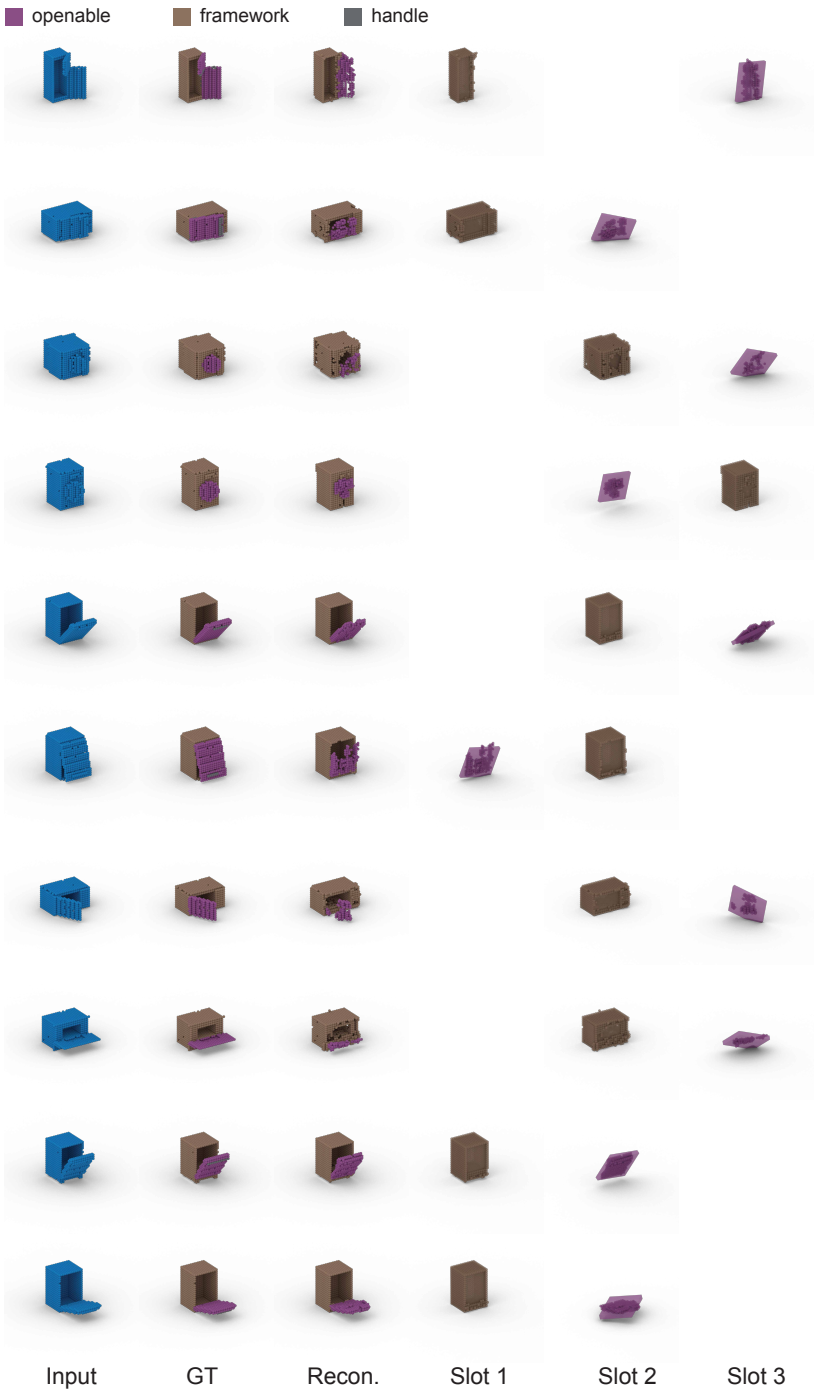


Fig. 4: Additional qualitative results on objects in the "openable" subset.

## References

1. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012 (2015) 6, 7
2. Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., Dosovitskiy, A., Kipf, T.: Object-centric learning with slot attention. Proceedings of Advances in Neural Information Processing Systems (NeurIPS) (2020) 1, 2
3. Mo, K., Zhu, S., Chang, A.X., Yi, L., Tripathi, S., Guibas, L.J., Su, H.: 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 (CVPR) (2019) 6
4. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of AAAI Conference on Artificial Intelligence (AAAI) (2017) 2
5. Xiang, F., Qin, Y., Mo, K., Xia, Y., Zhu, H., Liu, F., Liu, M., Jiang, H., Yuan, Y., Wang, H., et al.: SAPIEN: A simulated part-based interactive environment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 6