HAL3D: Hierarchical Active Learning for Fine-Grained 3D Part Labeling
— Supplemental Document —

Fenggen Yu[1,2*] Yiming Qian[1] Francisca Gil-Ureta[1]
Brian Jackson[1] Eric Bennett[1] Hao Zhang[1,2]
[1]Amazon [2] Simon Fraser University

This supplementary document provides 1) the designs of our label verification and modification interface in Section 1; 2) the evaluation regarding the impact of different hyper-parameter settings in Section 2; 3) the time analysis of the active learning process in Section 3; 4) the effectiveness of active learning on example nodes in Section 4; 5) the detailed node statistics of the active learning process in Section 5; 6) the details about the removed internal nodes in Section 6; 7) more qualitative results in Section 7.

1. User interface

We provide a simple verification user interface for users to verify the labels of each part; see Figure 1 for a screenshot. If the labels of a shape are all correct, the user can mark it with a happy face, otherwise a sad face. The user can also look up the color of each label on the side. In our hierarchical labeling design, a limited number of labels distinguished by different colors are provided to user at each node, effectively eliminating human judgement errors in the process.

We also provide a user interface for user to modify the label of each part; see Figure 2 for a screenshot. The annotated labels are organized in a hierarchical way. Once a label is selected, user can quickly assign it to corresponding parts by clicking them. On the left, we also provide a view of the entire shape for user to check the relative position of the parts at the current node.

We further provide a video ("HAL3D_video.mp4") with captions that gives a detailed illustration of the pipeline of our method.

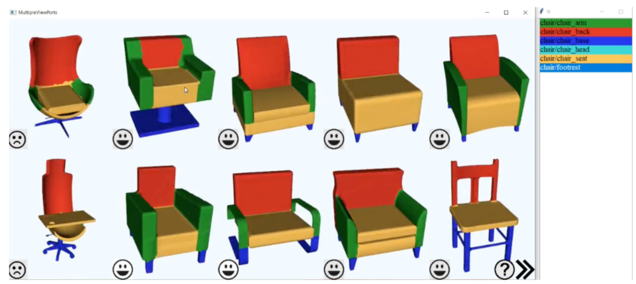

Figure 1: The user interface for part label verification.

Figure 2: The user interface for part label modification at the regular leg base node.

2. Impact of hyper-parameter variations

Our system consists a total of four hyper-parameters:

1. Shape verification batch size (B): The verification shape set is arranged into batches, each batch contains B shapes.
2. Shape modification batch size (Q_1): The number of shapes passed into modification from low-confidence proposals.
3. Verification failure threshold H: If the shape fails verification more than H times, it will go to modification in the next iteration.
4. Verification stopping percentile: The verification step will stop if the number of verified shapes is less than this threshold in the current batch.

To evaluate the impact of these parameter variations on the labeling efficiency, we test our method under different parameter settings. When testing a single parameter, the others are fixed. As shown in Tables 1, 2, 3, 4, changing different settings for all four parameters results in a time variance

*Work done during internship at Amazon

Table 1: Labeling efficiency evaluation under different verification batch size settings

Verification batch size	5	10	20
Time (hours)	4.36	4.34	4.34
# iterations	30	29	29

Table 2: Labeling efficiency evaluation under different modification batch size settings

Modification batch size	10	20	40
Time (hours)	4.25	4.34	4.41
# iterations	31	29	27

Table 3: Labeling efficiency evaluation under different verification failure threshold settings

Verification failure threshold	2	3	4
Time (hours)	4.34	4.39	4.36
# iterations	29	30	28

Table 4: Labeling efficiency evaluation under different verification stopping threshold settings

Verification stop percentile	20%	40%	60%
Time (hours)	4.28	4.34	4.38
# iterations	28	29	25

Table 5: Time analysis (in hours) of HAL3D in PartNet.

Category	Chair	Table	Lamp	Storage
Lab. time	4.34	4.41	3.87	2.16
Machine time	8.53	8.95	3.2	4.72
Clock time	11.41	12.12	6.63	6.29

of less than 10 minutes (about 3.8% of the total time), suggesting that these parameter choices have limited impact on the labeling complexity and subsequently demonstrating the robustness of our method.

3. Time analysis

Table 5 provides the timing results of HAL3D. The lab. time denotes the human labeling time. The machine time represents the running time in fine-tuning and inference during active learning. The clock time represents the total time to finish the active learning process. In HAL3D, the clock time is less than the sum of lab time and machine time since the users can perform active learning at different nodes in parallel. In addition, The clock time can be further reduced when users label different categories in parallel at the same time. Note that the clock time in Table 5 is collected based on the fact that the users only label test data from single category during the process.

Table 6: Effectiveness of active learning on several example nodes of the chair hierarchy tree.

Node name	Root	Arm	Back	Seat
Before AL	66.59	58.30	21.39	58.24
After AL	77.37	74.88	50.73	77.23

4. Effectiveness of active learning

Table 6 shows the effectiveness of active learning on several example nodes of the chair hierarchy tree. We show the prediction IoU change on the hold-out validation set of the chair category from PartNet. It shows that the proposal network is greatly improved after active learning, which iteratively fine-tuning the network with additionally labeled data.

5. Labeling statistics

We present a detailed analysis of our active labeling process for chairs in Table 7. The table includes various statistics such as the number of labels (L), the total number of verified parts during the verification step (PV), the total number of failed shapes in the verification step (FV), the total number of verified parts during the modification step that do not require modification but checking (PV/M), and the total number of modified parts (PM).

Our hierarchical design at each node results in shorter processing time compared to the non-hierarchical design due to reduced task complexity and higher network prediction accuracy at each node. Although the total number of modified parts is slightly larger than the non-hierarchical design, our approach reduces the number of labels required for each part, resulting in shorter overall processing time.

6. Removed internal nodes

We remove internal nodes having less than three children from the original PartNet tree, and the children of those internal nodes become the children of the internal nodes' parents. Firstly, removing those nodes eliminates network training at them, reducing the running time. Secondly, merging the < 3 children of these internal nodes to their parents has marginal impact on the labeling workload at the parent node. As shown in Table 8, there are only around 5% shapes having such internal nodes.

7. More qualitative results

Figures 3, 4, 5, 6 show more qualitative results on the PartNet database. Figures 7, 8, 9, 10, 11 show more qualitative results on the our constructed ABO test set. We also

Table 7: Statistics for active learning at different nodes for our non-hierarchical and hierarchical designs.

Node	L	PV	FV	PV/M	PM	Hours
w/o hier	29	540	101	4296	1301	5.99
Chair	6	3173	139	2648	370	1.59
Arm	6	507	99	291	196	0.52
Back	4	239	124	794	377	0.87
Base	4	346	57	22	28	0.08
Head	2	0	0	18	0	0.01
Seat	3	452	136	560	282	0.68
Footrest	3	0	0	18	0	0.01
Pedestal-B	2	45	1	33	3	0.02
Regular-B	5	792	71	565	177	0.50
Star-B	3	125	20	491	40	0.15
Total	-	5679	647	5440	1473	4.34

show the hierarchy label tree under the visual results, the ”OR” nodes are highlighted with a red bounding box.

Table 8: The removed internal nodes from the original PartNet tree.

Removed internal nodes	Percentile of removed shape
Chair/footrest/chair_seat	9/400=2.3%
Table/picnic_table	4/400=1.0%
Table/game_table/ping_pong_table/table_base/regular_leg_base	9/400=2.3%
Table/game_table/pool_table	6/400=1.5%
Lamp/ceiling_lamp/chandelier/lamp_unit_group/lamp_unit	9/400=2.3%
Lamp/ceiling_lamp/pendant_lamp/pendant_lamp_unit	17/400=4.3%
Lamp/wall_lamp/lamp_unit	19/400=5.0%
Lamp/street_lamp/lamp_unit	17/400=4.3%
Storage_furniture/cabinet/drawer/drawer_box	22/400=5.5%

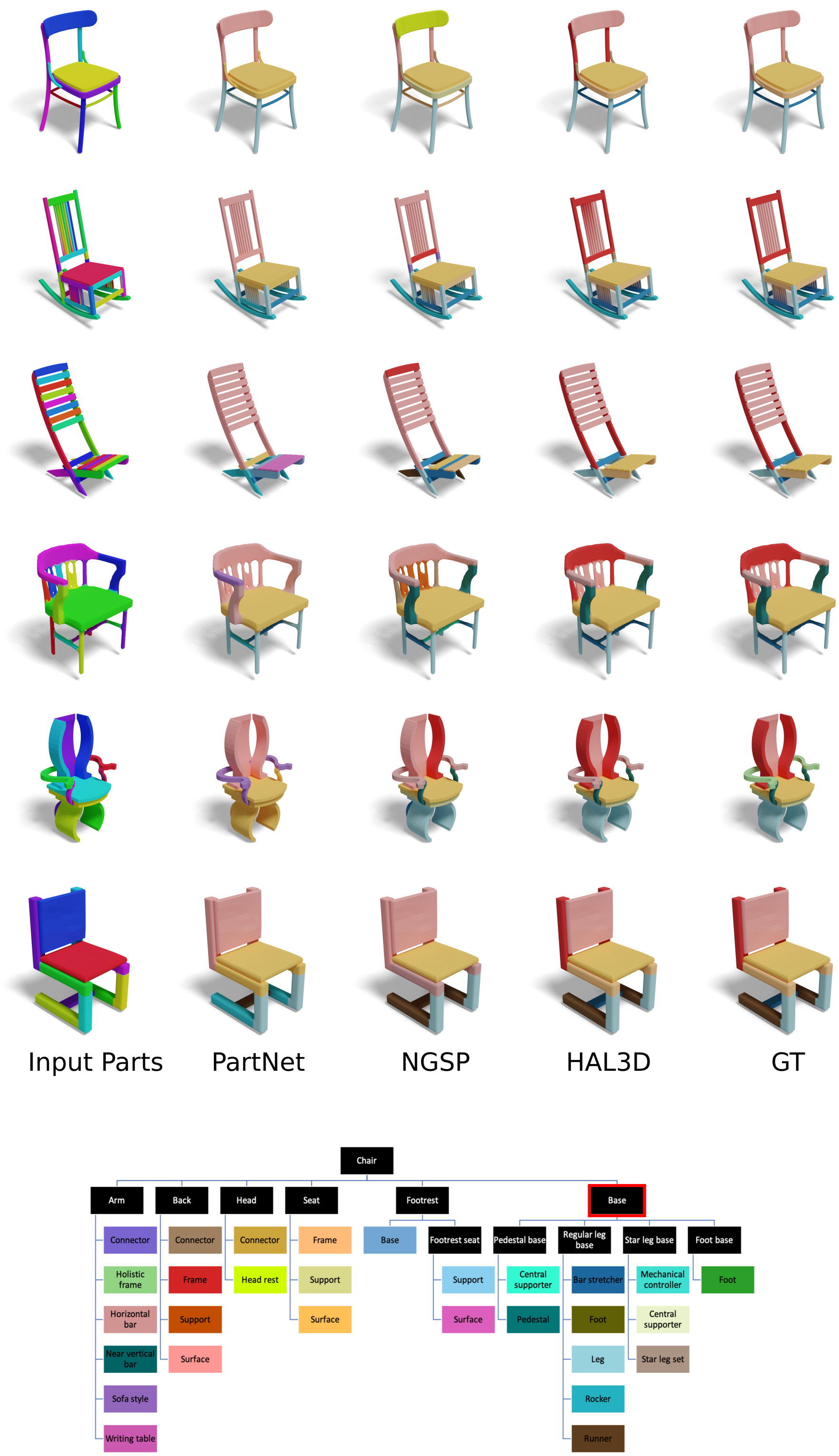

Figure 3: More qualitative evaluations for chairs on PartNet dataset (top) and the corresponding label tree (bottom).

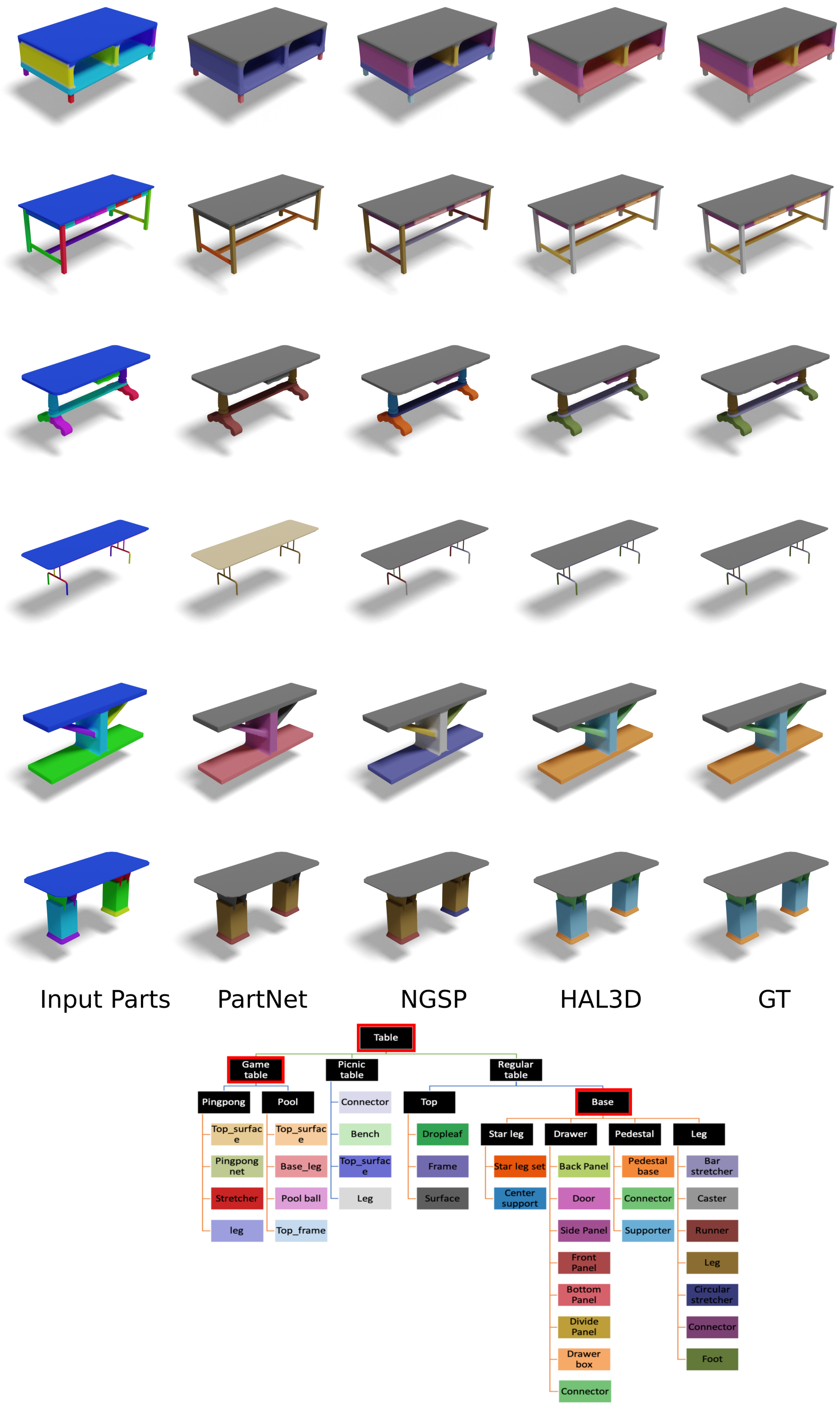

Figure 4: More qualitative evaluations for tables on PartNet dataset (top) and the corresponding label tree (bottom).

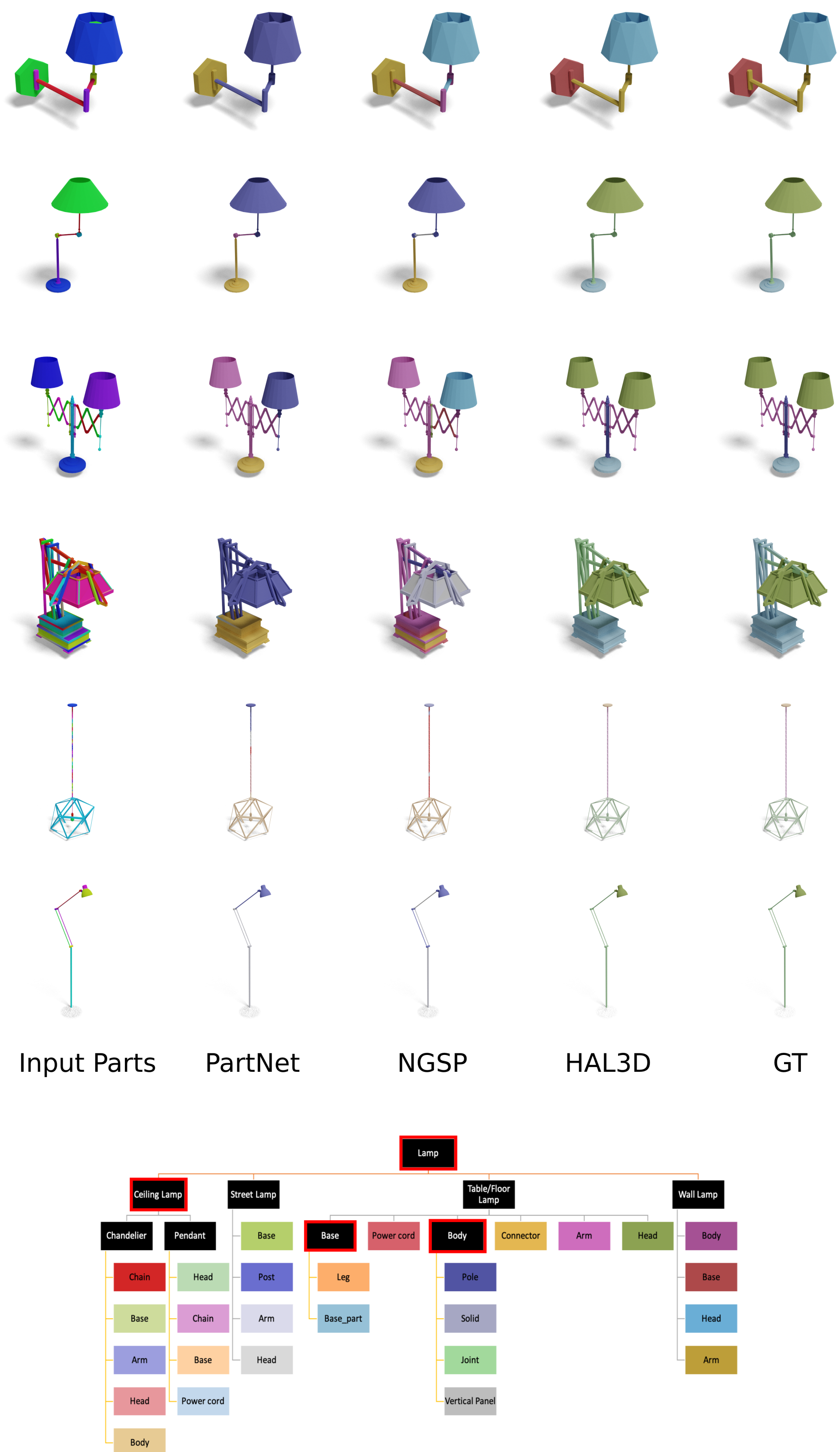

Figure 5: More qualitative evaluations for lamps on PartNet dataset (top) and the corresponding label tree (bottom).

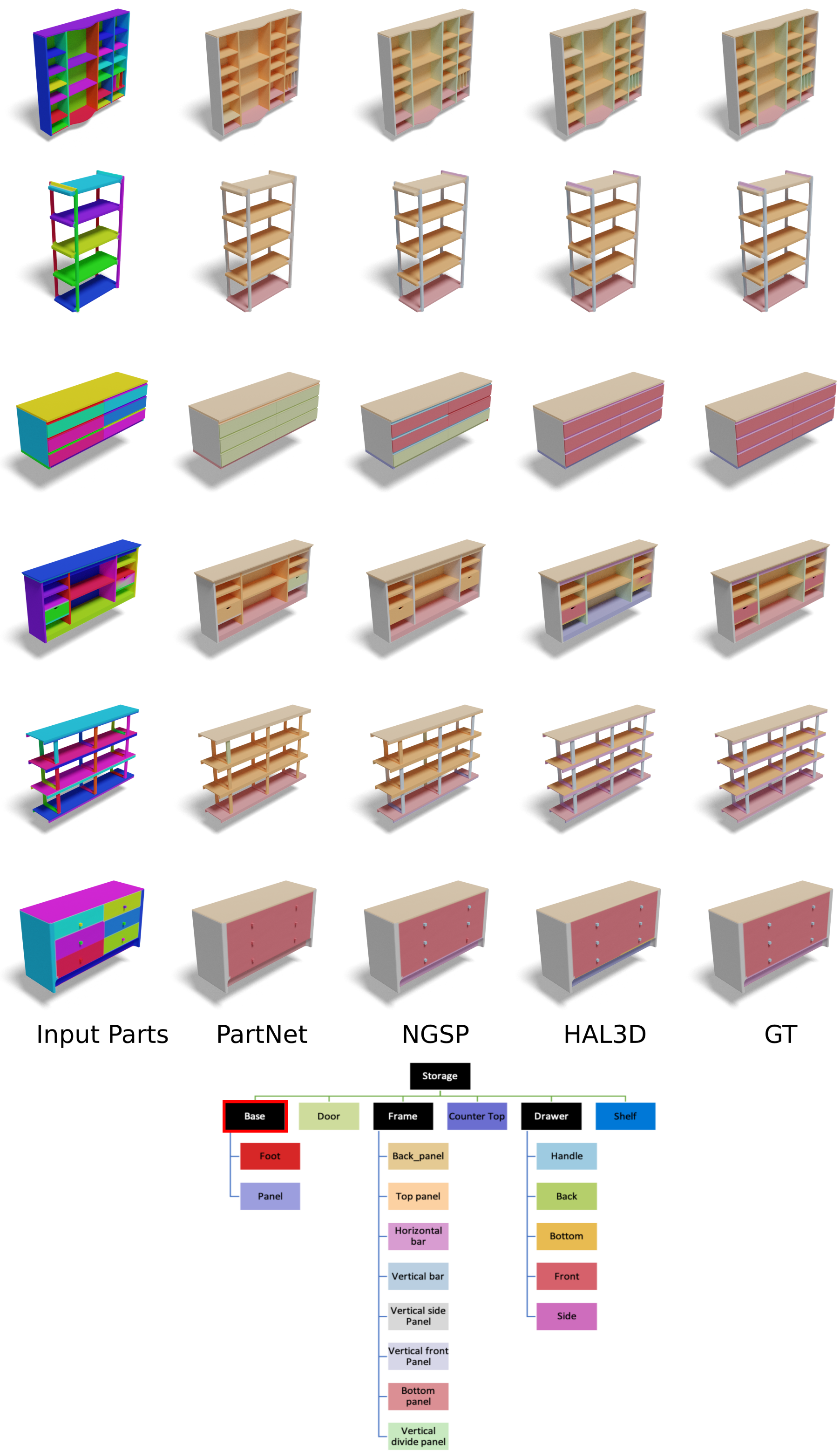

Figure 6: More qualitative evaluations for storage furnitures on PartNet dataset (top) and the corresponding label tree (bottom).

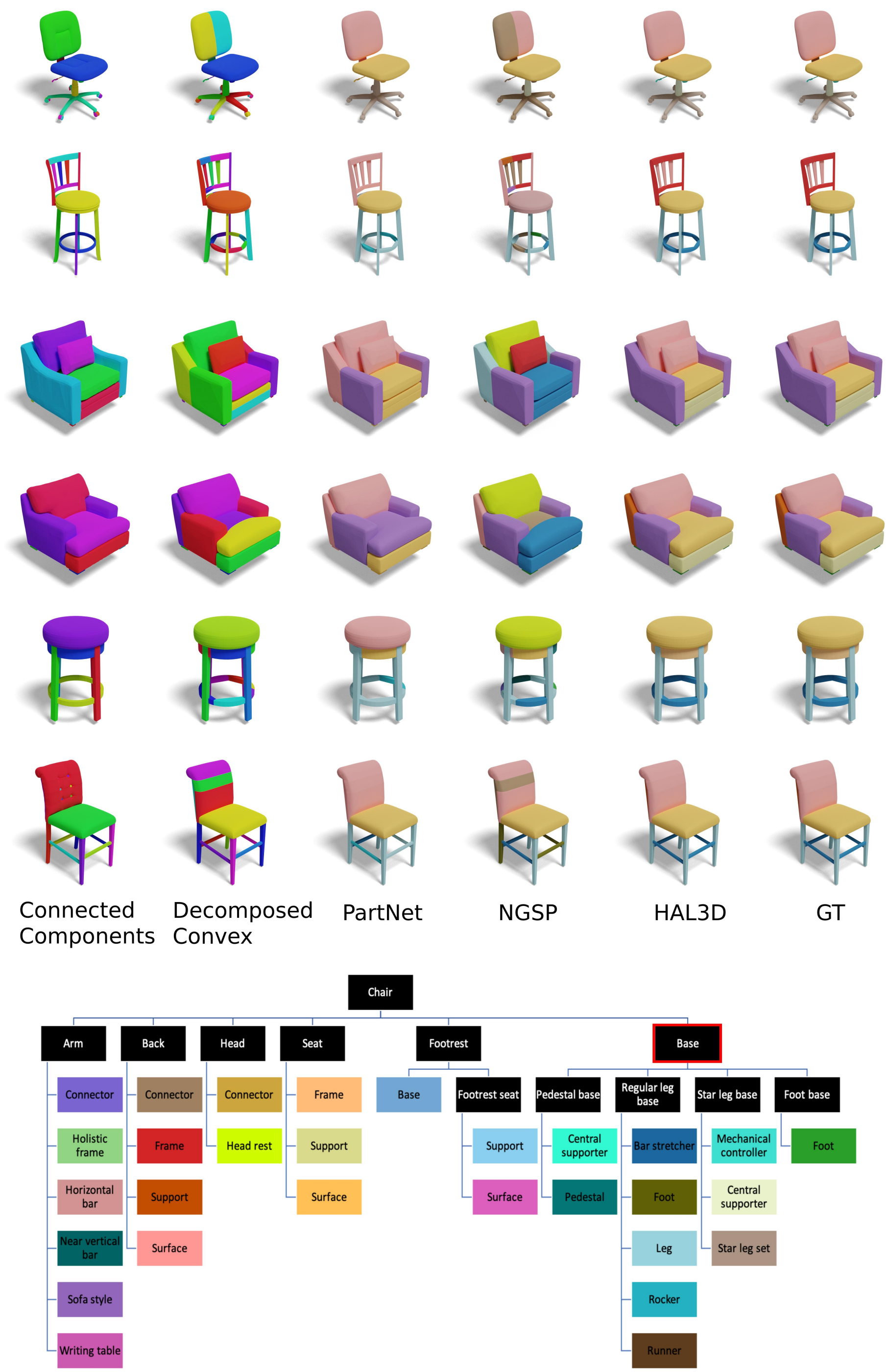

Figure 7: More qualitative evaluations for chairs on ABO dataset (top) and the corresponding label tree (bottom).

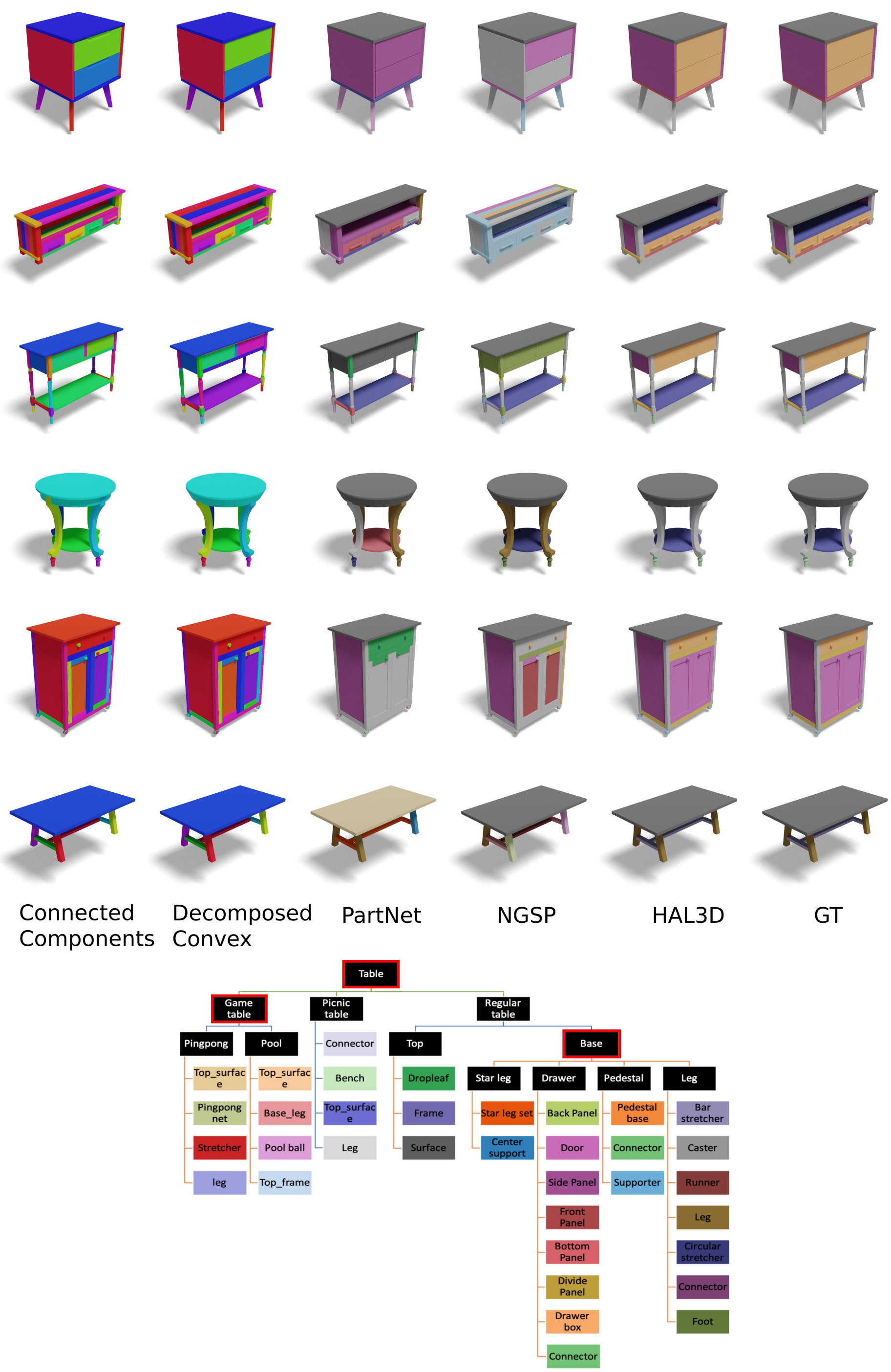

Figure 8: More qualitative evaluations for tables on ABO dataset (top) and the corresponding label tree (bottom).

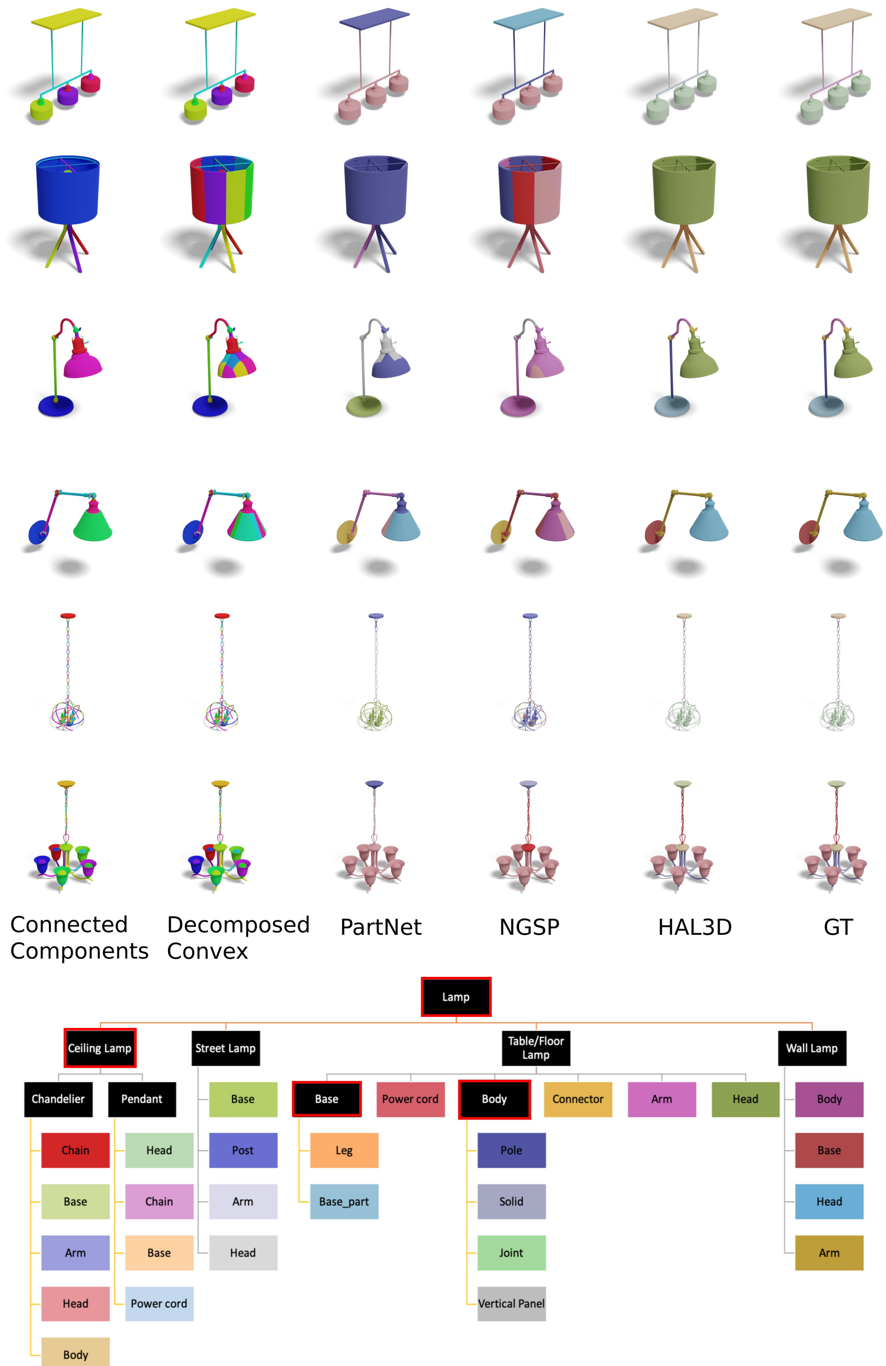

Figure 9: More qualitative evaluations for lamps on ABO dataset (top) and the corresponding label tree (bottom).

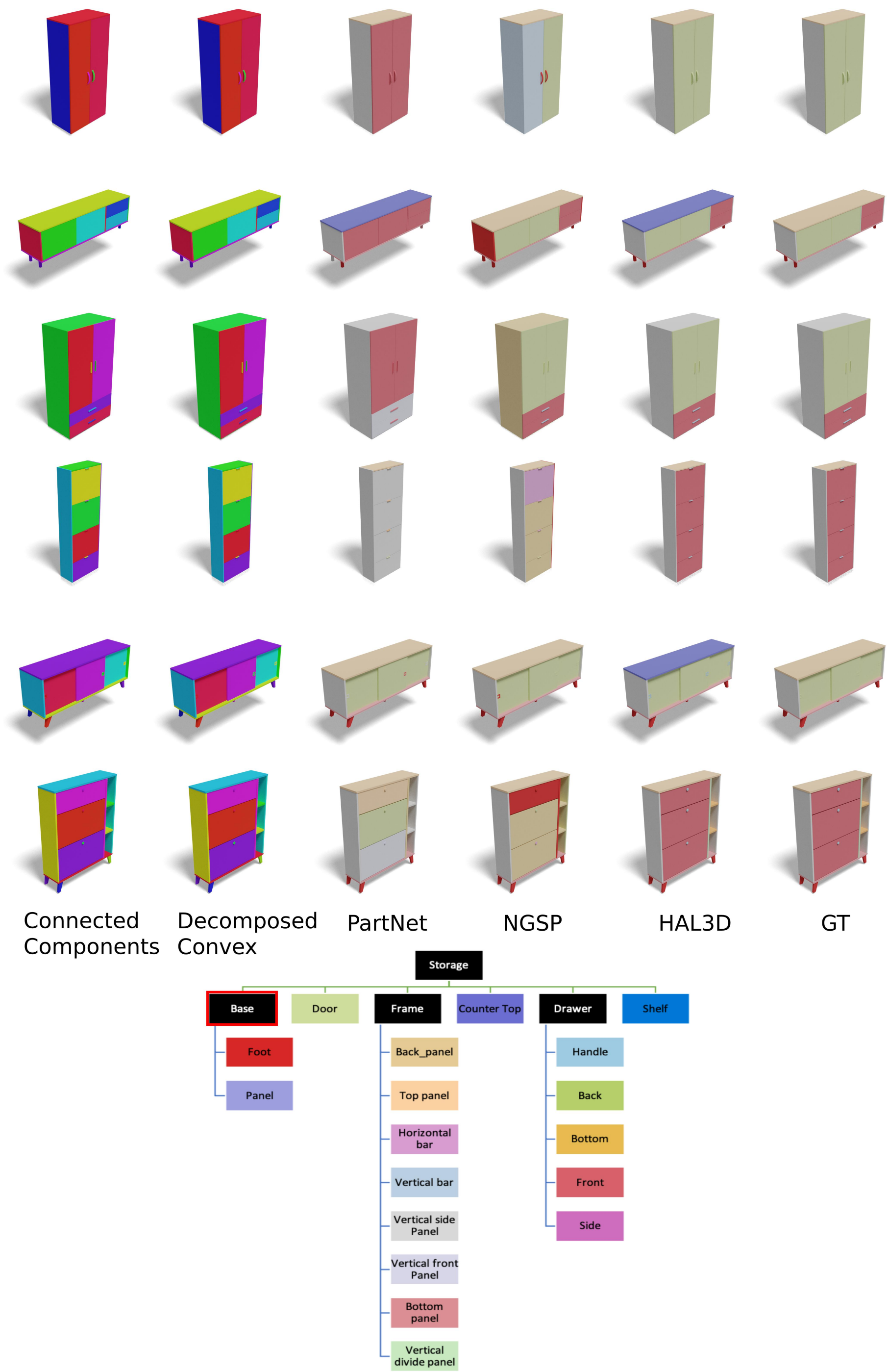

Figure 10: More qualitative evaluations for storage furnitures on ABO dataset (top) and the corresponding label tree (bottom).

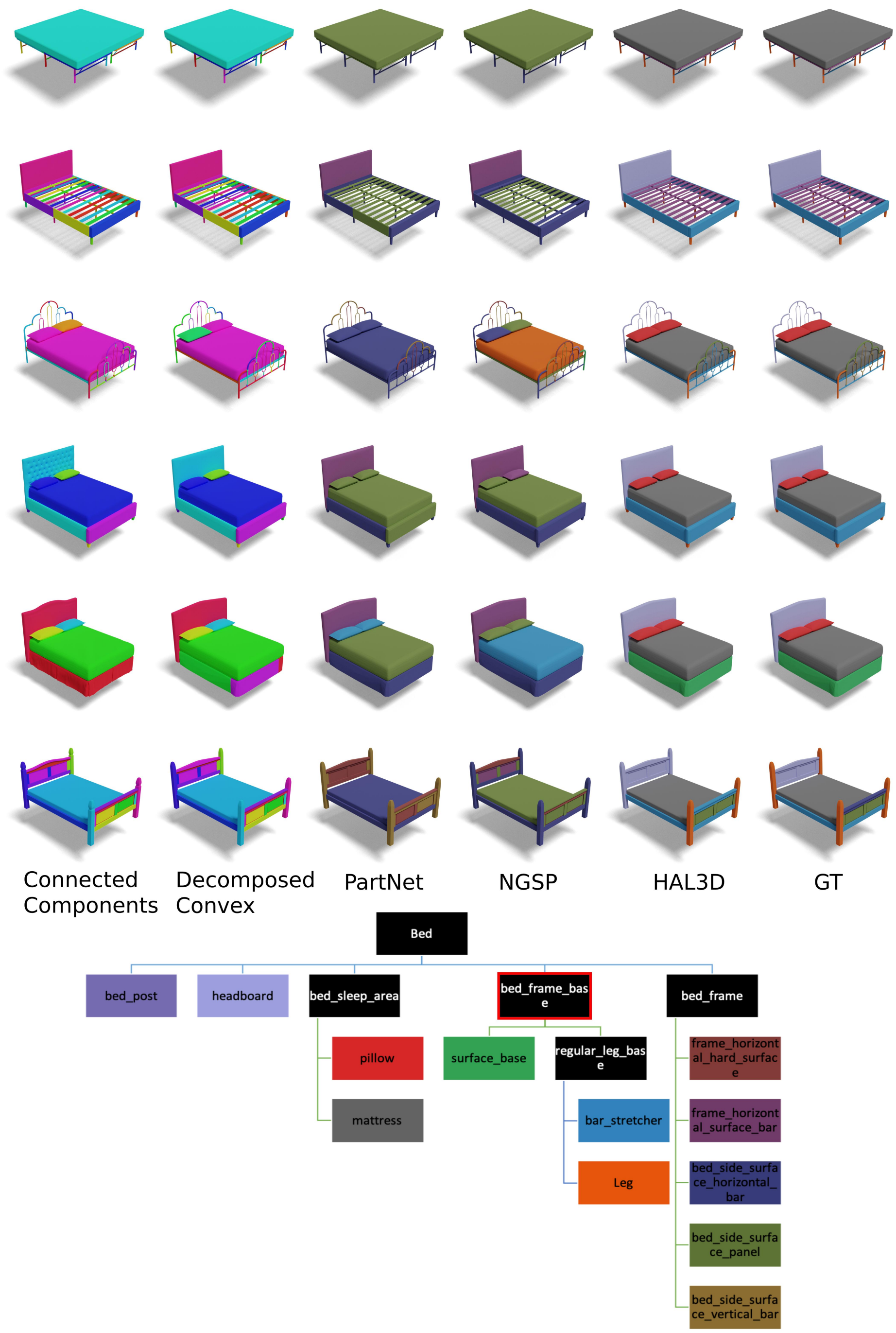

Figure 11: More qualitative evaluations for beds on ABO dataset (top) and the corresponding label tree (bottom).