A Scalable Active Framework for Region Annotation in 3D Shape Collections
Supplemental material

Li Yi[1] Vladimir G. Kim[1,2] Duygu Ceylan[2] I-Chao Shen[3] Mengyan Yan[1]
Hao Su[1] Cewu Lu[1] Qixing Huang[4,5] Alla Sheffer[3] Leonidas Guibas[1]
[1]Stanford University [2]Adobe Research [3]University of British Columbia [4]TTI Chicago [5]UT Austin

This document provides a list of supplemental materials that accompany this paper.

- **Labeling vs segmentation** - We demonstrate performance of our method on segmentation benchmark relative to existing alternatives.
- **Exploring ShapeNetCore** - We demonstrate examples of leveraging our part labels for exploring collections of shapes (see `exploration.mp4` file and Section 2).
- **Optimality of Verification Set** - We prove optimality of our verification set selection procedure in Section 3 of this document.
- **User Interfaces Video** - We submit a video demonstrating key features of annotation and verification user interfaces (see `interface.mp4` file).
- **ShapeNetCore Analysis** - We provide the full set of analysis results for each category of ShapeNetCore showing number of verified models vs human time (Figure 4), and FMF vs fraction of data that has been verified (Figure 5).

1 Labeling vs segmentation

Although the focus of our work is labeling, not segmentation, we found that our simple method for generating segmentation from the labels gives reasonable boundaries. We evaluate our method on PSB benchmark [Chen et al. 2009] using leave-one-out setup proposed in Kalogerakis et al. [2010] and ELM-OPT [Xie et al. 2014]. We tested our method on 7 categories of rigid objects excluding articulated shapes such as human, since our method is not designed for those. One could use intrinsic mapping algorithms to extend our method to non-rigidly deforming shapes. We report the results in Table 1.

	Training	Kalo	ELM-OPT	Ours
Cup	9.8	**9.9**	10.3	10.8
Airplane	7.4	**7.9**	8.9	8.6
Chair	5.2	**5.4**	7.1	5.9
Table	5.9	6.2	**5.9**	7.5
Mech	8.5	**10.0**	15.9	14.8
Bearing	6.8	9.7	15.4	**9.6**
Vase	10.5	16.0	**15.6**	20.9

Table 1: *This table shows the quality of the segmentations our method produces by comparing the rand index scores on some categories of PSB achieved by our method and the prior segmentation algorithms Kalo (Kalogerakis et al.) and ELM-OPT (Xie et al.). Lower rand index scores indicate better performance and are highlighted in bold.*

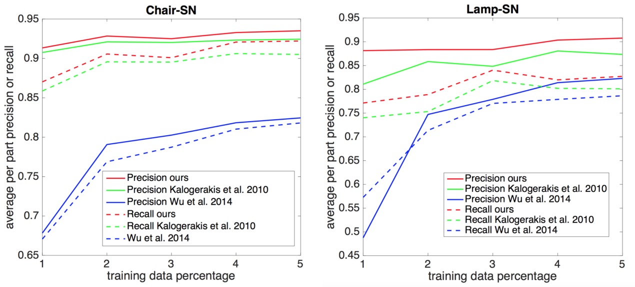

Figure 1: *This comparison corresponds to the one shown in Figure 12 of the paper, but with precision and recall as evaluation metrics.*

In the paper, we compare our method with supervised labeling approaches based on an F1 measurement. Also we evaluate different design choices of our pipeline based on F1 measurement. For a better understanding of the labeling quality of these experiments, we also provide the precision and recall for the final labeling results in Table 2 and 3 correspondingly. In addition, we use precision and recall as evaluation metrics and show the comparison with supervised labeling approaches in Figure 1. Notice unlike traditional retrieval tasks, our experiment does not involve in any parameters influencing the tradeoff between precision and recall, so we show how the average precision and recall change under different training data percentages.

Precision/Recall	Wu et al. [2014]	Kalogerakis et al. [2010]	Ours
Lamp	0.823/0.786	0.873/0.801	**0.907/0.827**
Chair	0.824/0.818	0.924/0.905	**0.935/0.922**

Table 2: *This table gives the average per part precision and recall corresponding to the comparison conducted in Figure 12 of the paper. The numbers reported here are generated when the training data percentage is 5%.*

Precision/Recall	Chair-400	Vase-300
with all component	**0.950/0.953**	**0.909/0.889**
no active selection	0.928/0.946	0.873/0.857
no verification step	0.889/0.896	0.825/0.830
no ensemble learning	0.917/0.933	0.867/0.883
no correspondence term	0.899/0.915	0.821/0.815
no feature-based term	0.913/0.927	0.859/0.878
no learning of weights	0.940/0.948	0.887/0.877
no smoothness term	0.949/0.951	0.902/0.886

Table 3: *This table gives the average per part labeling precision and recall corresponding to the experiment conducted in Figure 13 of the paper. Different variants of our method were tested, each without some feature. The numbers reported here correspond to the end point of each curve in Figure 13 of the paper.*

2 Exploring ShapeNetCore

We can leverage the obtained annotations to explore and gain insights about the data in ShapeNetCore. Instead of considering only

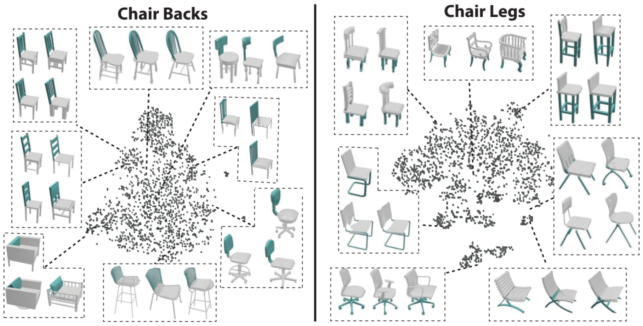

Figure 2: *We demonstrate that our annotations can facilitate exploration of large collections of 3D models. We show embeddings of chairs based on the shape of their backs and legs (see video for an interactive example).*

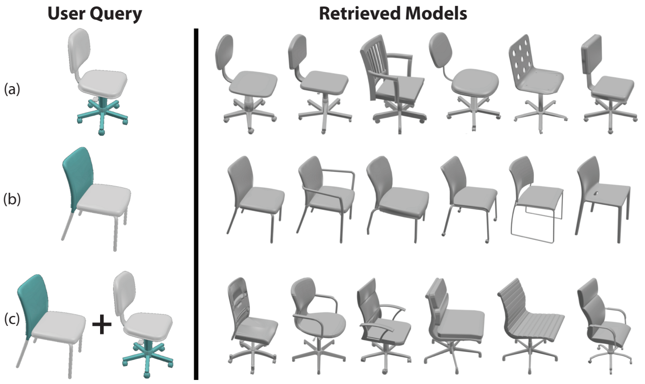

Figure 3: *Part annotations can facilitate shape retrieval. In this example for ShapeNetCore chairs, we show models retrieved based only on the shape of the base (a), the shape of the back (b), and by searching for chairs that combine both of these criteria (c).*

global shape similarities, understanding salient regions enables us to organize data by comparing features in specific regions only. For example, we can embed the models based on their part-to-part similarity as shown in Figure 2 (where part similarity is computed with lightfield descriptors). Note how this quickly allows the user to see different regions due to variations in a particular part. We also enable faceted search in the spirit of Kim et al. [2013] allowing users to select parts of interest and search for shapes that have the desired arrangement of parts (see Figure 3 and video).

3 Optimality of verification set

We demonstrate optimality of our method for selecting the verification set V_k^m (see Section 4 of our paper). Recall that at this stage our method selected annotation set $\mathcal{A}_k^m$, and it will greedily add models to $\mathcal{V}_k^m = \{v_k^m\}$ based on annotation confidences C_{ver}^m until the utility function $E_U^m = \frac{N_{\text{good}}^m}{T^m}$ stops increasing. The input to our algorithm is sorted confidence values: $1 \geq C_{ver}^m[k] \geq C_{ver}^m[k] ... \geq C_{ver}^m[k] \geq 0$ (we assume that sorting does not change shape indexing since it does not affect our derivations), we want to determine $v_k^m \in \{0,1\}$ so that $\mathbb{E}(\mathcal{V}^m) = \frac{N_{\text{good}}^{m-1} + \sum_{k=1}^{N} v_k^m C_{\text{ver}}^m[k]}{T^{m-1} + \sum_{k=1}^{N} v_k^m (\tau_{\text{ident}} + (1 - C_{\text{ver}}^m[k])\tau_{\text{click}})}$ is maximized. Notice that terms that depend on annotation set $\mathcal{A}$ are absorbed into T^{m-1} and can be kept constant at this stage without affecting our derivation.

The optimality of our greedy approach is guaranteed by two observations. First, given a constant number of models to be verified n (i.e., $\sum_k v_k^m = n$), the optimal values are $v_k^m = 1 \ \ \forall k \leq n$ and 0 everywhere else since it minimizes the denominator and maximizes the numerator. Thus, our algorithm is optimal if the number of models to be verified is fixed to n. Let us denote this optimal energy for n models by:

$$\begin{aligned} f(n) &= \max_{V_k^m \text{ s.t. } \sum_k V_k^m = n} E_U^m(V_k^m) \\ &= \frac{N_{\text{good}}^{m-1} + \sum_{k=1}^{n} C_{ver}^m[k]}{T^{m-1} + n\tau_{\text{ident}} + n\tau_{\text{click}} - \sum_{k=1}^{n} C_{ver}^m[k]\tau_{\text{click}}} \end{aligned} \quad (1)$$

We denote numerator in the equation above by f_{num} and denominator f_{denom}. We now prove that our utility function is monotonically decreasing after we greedily pick optimal n: $f(n+1) \leq f(n) \Rightarrow f(n+i+1) \leq f(n+i) \ \ \forall i > 0$.

Suppose $f(n+1) \leq f(n)$ is true, then this is equivalent to:

$$\begin{aligned} & \frac{f_{\text{num}}(n) + C_{\text{ver}}^m[n+1]}{f_{\text{denom}}(n) + \tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+1]\tau_{\text{click}}} \leq \frac{f_{\text{num}}(n)}{f_{\text{denom}}(n)} \\ \Leftrightarrow \quad & \frac{C_{\text{ver}}^m[n+1]}{\tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+1]\tau_{\text{click}}} \leq \frac{f_{\text{num}}(n) + C_{\text{ver}}^m[n+1]}{f_{\text{denom}}(n) + \tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+1]\tau_{\text{click}}} \end{aligned} \quad (2)$$

by utilizing a simple algebraic equality: $\frac{a+c}{b+d} \leq \frac{a}{b} \Leftrightarrow \frac{c}{d} \leq \frac{a+c}{b+d}$ (note that all our terms are positive). Also, since $C_{\text{ver}}^m[n+2] \leq C_{\text{ver}}^m[n+1]$:

$$\frac{C_{\text{ver}}^m[n+2]}{\tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+2]\tau_{\text{click}}} \leq \frac{C_{\text{ver}}^m[n+1]}{\tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+1]\tau_{\text{click}}} \quad (3)$$

Now combining Equations 2 and 3 yields:

$$\frac{C_{\text{ver}}^m[n+2]}{\tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+2]\tau_{\text{click}}} \leq \frac{f_{\text{num}}(n) + C_{\text{ver}}^m[n+1]}{f_{\text{denom}}(n) + \tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+1]\tau_{\text{click}}} \quad (4)$$

By definition:

$$\Leftrightarrow \quad \frac{C_{\text{ver}}^m[n+2]}{\tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+2]\tau_{\text{click}}} \leq \frac{f_{\text{num}}(n+1)}{f_{\text{denom}}(n+1)} \quad (5)$$

and applying $\frac{c}{d} \leq \frac{a}{b} \Leftrightarrow \frac{a+c}{b+d} \leq \frac{a}{b}$:

$$\Leftrightarrow \quad \frac{f_{\text{num}}(n+1) + C_{\text{ver}}^m[n+2]}{f_{\text{denom}}(n+1) + \tau_{\text{ident}} + \tau_{\text{click}} - C_{\text{ver}}^m[n+2]\tau_{\text{click}}} \leq \frac{f_{\text{num}}(n+1)}{f_{\text{denom}}(n+1)} \quad (6)$$

Or equivalently, $f(n+2) \leq f(n+1)$.
Thus, by induction $f(n+i+1) \leq f(n+i)$ for $i > 0$. □

References

CHEN, X., GOLOVINSKIY, A., AND FUNKHOUSER, T. 2009. A benchmark for 3d mesh segmentation. In *ACM SIGGRAPH*, SIGGRAPH '09, 73:1–73:12.

KALOGERAKIS, E., HERTZMANN, A., AND SINGH, K. 2010. Learning 3D mesh segmentation and labeling. In *ACM SIGGRAPH*, 102:1–102:12.

WU, Z., SHOU, R., WANG, Y., AND LIU, X. 2014. Interactive shape co-segmentation via label propagation. *CAD/Graphics 38*, 2, 248–254.

XIE, Z., XU, K., LIU, L., AND XIONG, Y. 2014. 3d shape segmentation and labeling via extreme learning machine. *SGP*.

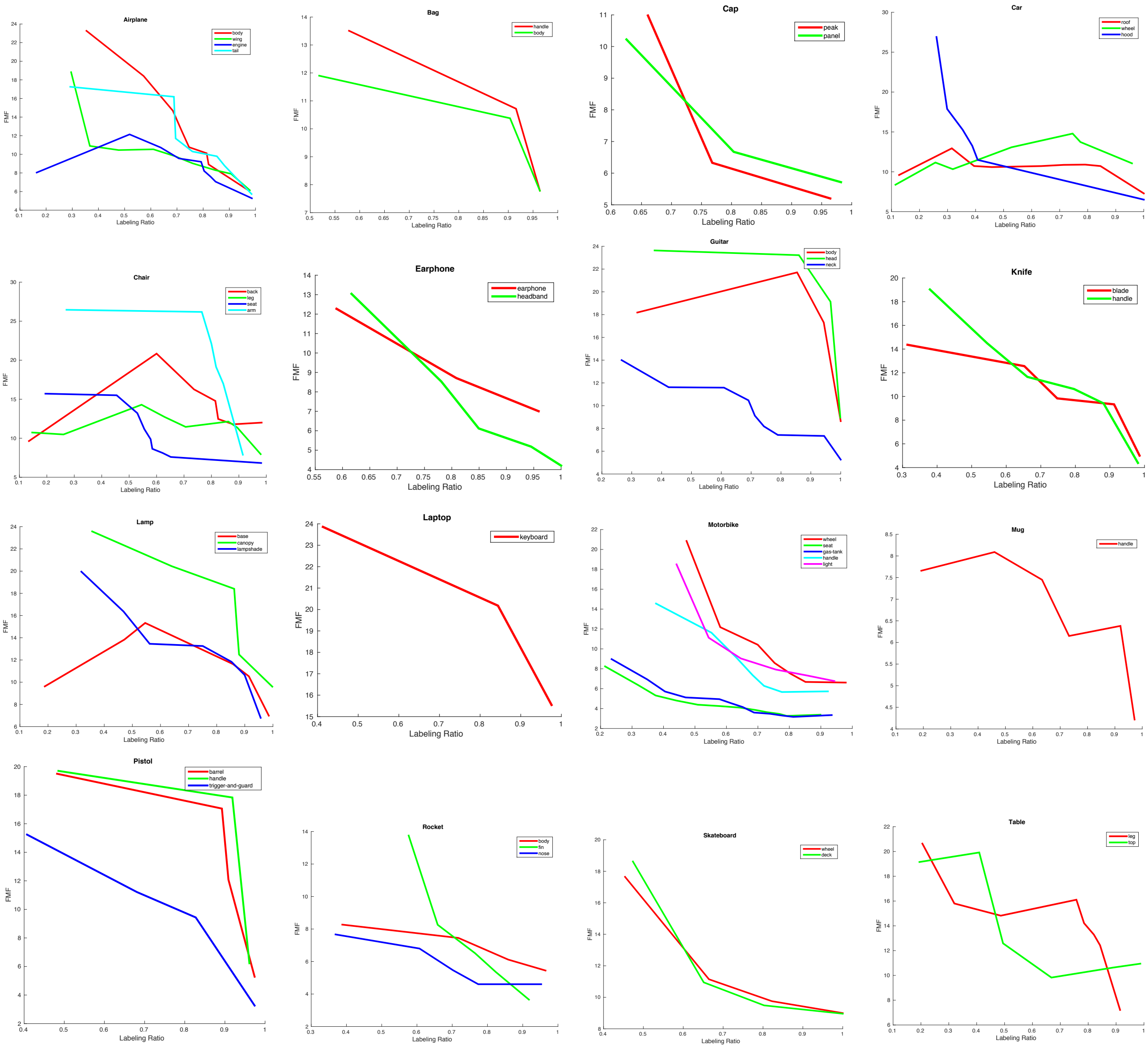

Figure 4: *These plots depict how FMF relates to fraction of labeled data for different labels in different categories of ShapeNetCore. See our paper (Section 8) for more details.*

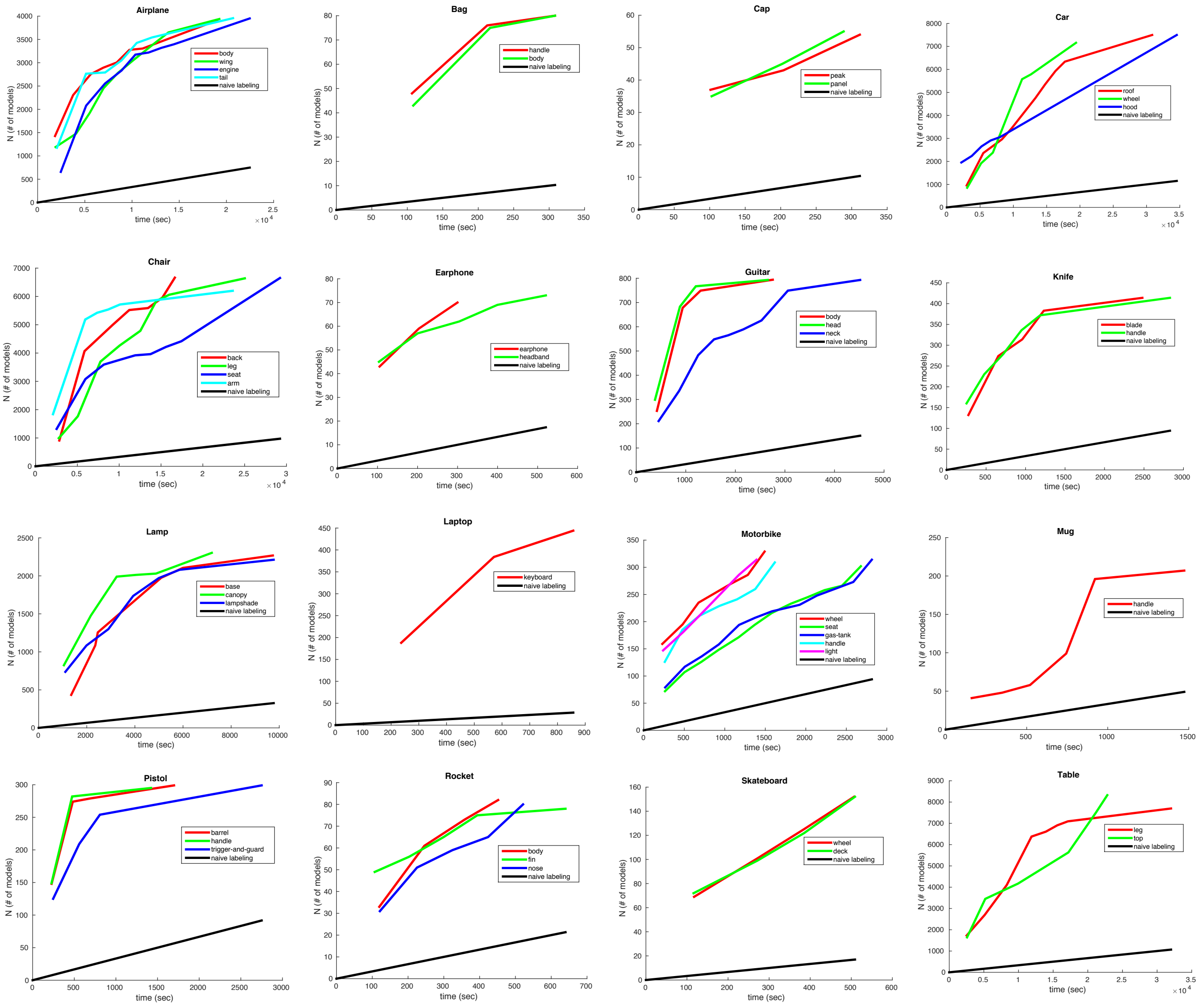

Figure 5: *These plots depict how number of positively-verified models relates to total human work time for different labels in different categories of ShapeNetCore. See our paper (Section 8) for more details.*