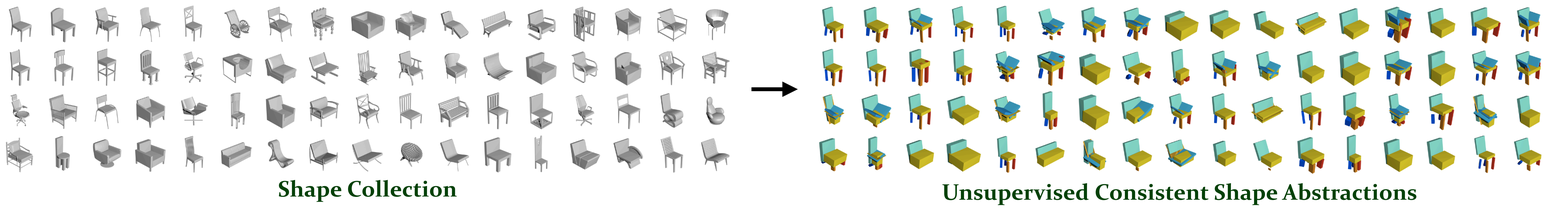
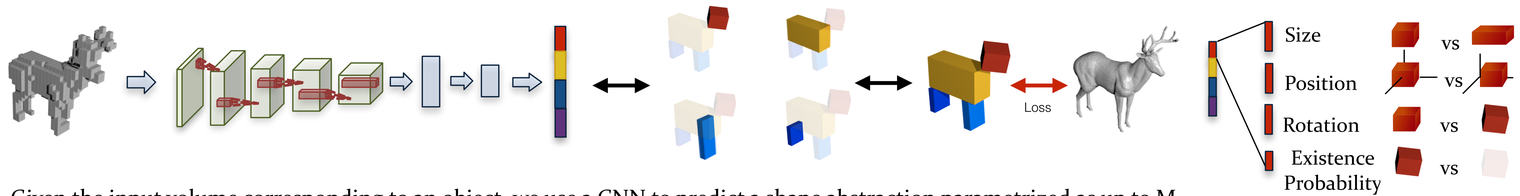
# Learning Shape Abstractions by Assembling Volumetric Primitives

Shubham Tulsiani[1] Hao Su[2] Leonidas J. Guibas[2] Alexei A. Efros[1] Jitendra Malik[1]

[1]University of California, Berkeley [2]Stanford University

## Overview




Given the input volume corresponding to an object, we use a CNN to predict a shape abstraction parametrized as up to M primitives. The obtained abstraction allow an **interpretable** representation for each object as well as provides a **consistent** parsing across shapes e.g. chair seats are captured by the same primitive across the category.

## Loss Function

We want the obtained abstraction to explain the corresponding shape (coverage and consistency loss) as well as be parsimonious (reward for using fewer primitives)

### Coverage Loss

$L_1$( , ) ≡ ⊆

$$L_1(\{(z_m, q_m, t_m)\}, O) = \mathbb{E}_{p \sim S(O)} \|\mathcal{C}(p; \cup_m \bar{P}_m)\|^2$$

Δ( , ) + Δ( , ) + Δ( , ) ... + Δ( , )

$$\mathcal{C}(p; \cup_m \bar{P}_m) = \min_m \mathcal{C}(p; \bar{P}_m)$$

Δ( , ) = min {Δ( , ), Δ( , ), Δ( , ), Δ( , )}

### Consistency Loss

$L_2$( , ) ≡ ⊇

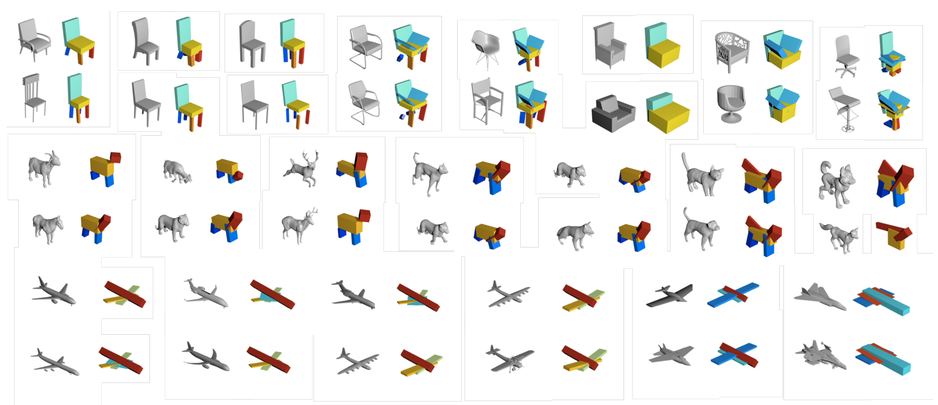
$$L_2(\{(z_m, q_m, t_m)\}, O) = \sum_m \mathbb{E}_{p \sim S(\bar{P}_m)} \|\mathcal{C}(p; O)\|^2$$

⊇ + ⊇ + ⊇ + ⊇

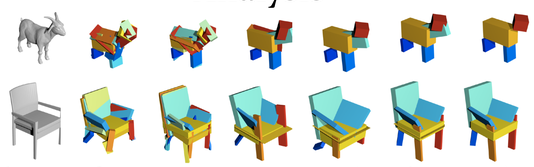
where

⊇ ≡ Δ( , ) + Δ( , ) ...+Δ( , )

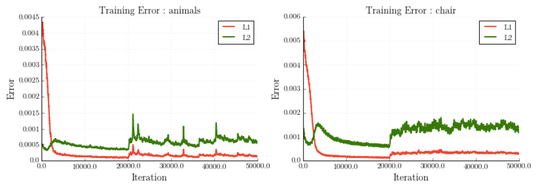
## Results



## Analysis



Predictions after every 10,000 iterations (in columns 2-6). The last column shows the result after removing redundant parts.



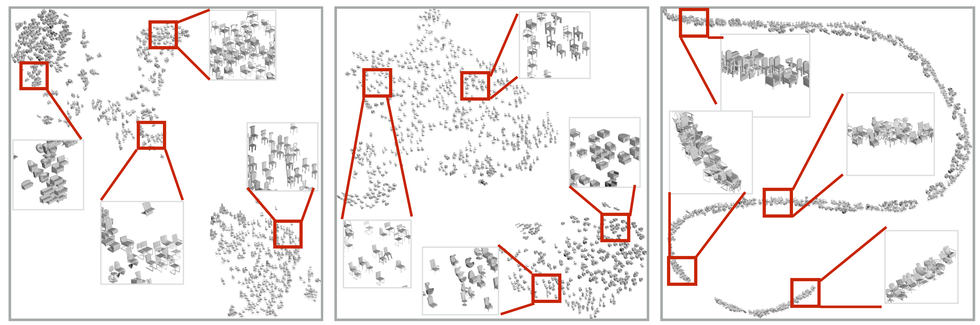

## Applications

### Unupervised Parsing

### Shape Manipulation

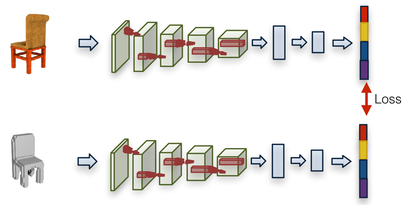
Source mesh (top) deformed(right) to have a shape similar to the target mesh (bottom).

### Interpretable Shape Similarity



a) All primitives b) Chair back, seat primitives c) Chair back orientation.

### Image Based Parsing




Predictions of an image based CNN trained to mimic the output of the learned volume based CNN.

**Acknowledgements**: We thank Saurabh Gupta and David Fouhey for insightful discussions. This work was supported in part by Intel/NSF Visual and Experiential Computing award IIS-1539099, NSF Award IIS-1212798, and the Berkeley Fellowship to ST. We gratefully acknowledge NVIDIA corporation for the donation of Tesla GPUs used for this research.

**Code**

*shubhtuls.github.io/volumetricPrimitives/*