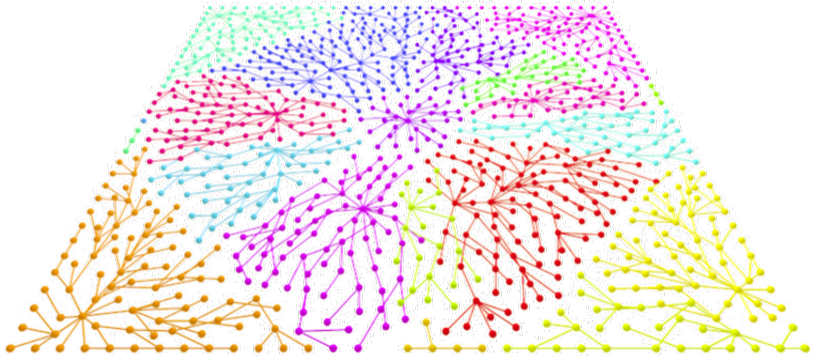
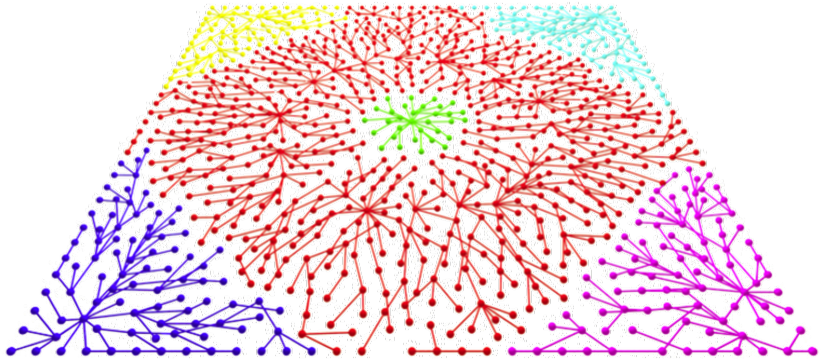
CS233, CME251: Geometric and Topological Data Analysis

Leonidas Guibas
Computer Science Department
Stanford University



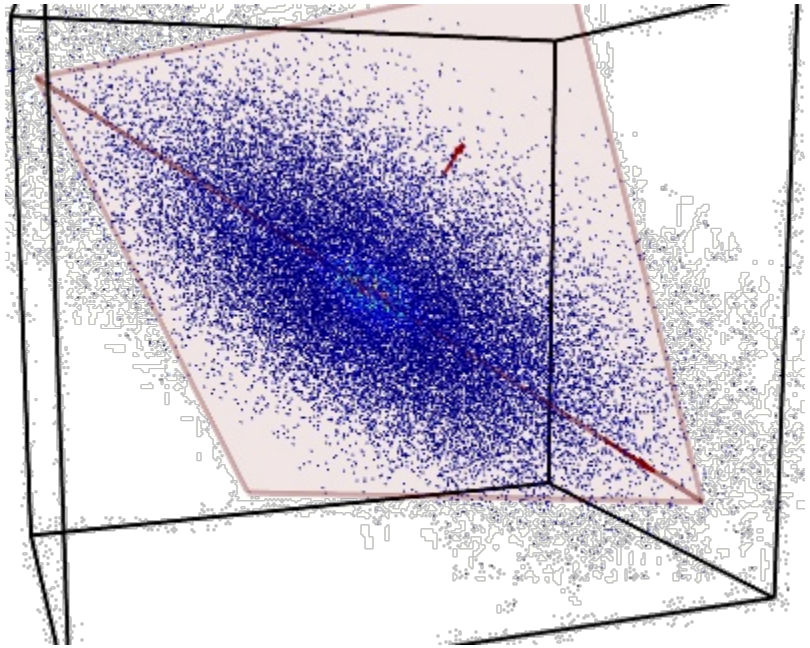


Lecture 3
04 April 2022

Last Time: (PCA)
Principal Components Analysis

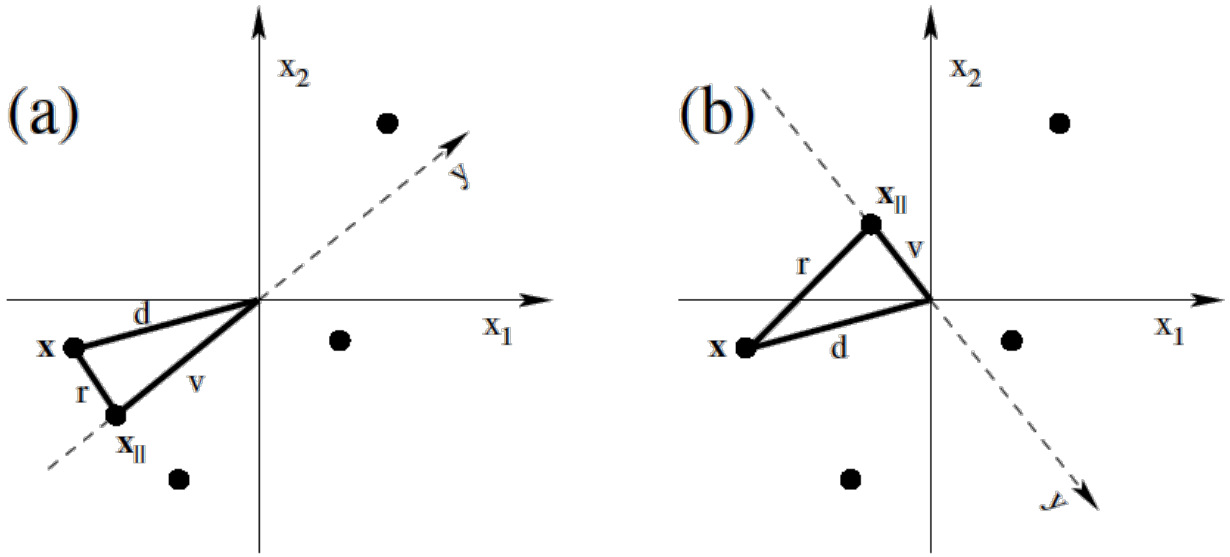
Principal Components Analysis (PCA)

- Introduced by Pearson (1901) and Hotelling (1933) to describe the variation in a set of multivariate data in terms of a set of uncorrelated variables.
- PCA looks for a single lower dimensional subspace that captures most of the variation in the data.
- Specifically, we aim to minimize the error introduced by projecting the data into this linear subspace.



Reconstruction Error and Variance

- Minimizing the reconstruction error is equivalent to maximizing the variance of the projected data.



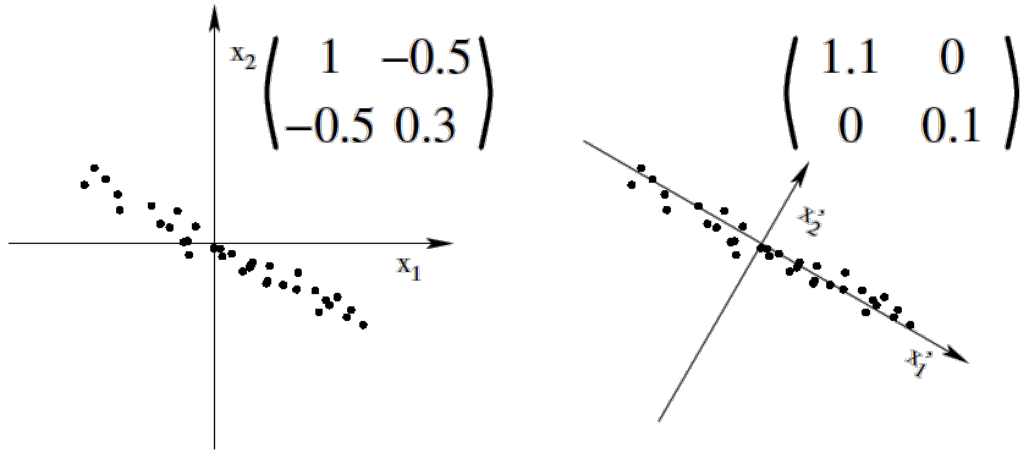
- Remember the mean is 0

$$r^2 + v^2 = d^2$$

NB, projections of centered data are centered

PCA by Diagonalizing the Covariance Matrix

- Finding the direction of maximum variance is easy if the covariance matrix is diagonal.



- So, in general, we want to rotate the coordinate system so as to make the covariance matrix diagonal.
- This is an eigenvalue problem; the eigenvectors of the covariance matrix point in the directions of maximal and minimal variance – so we rotate the axes to align them with the directions of maximal and minimal variance.

The General PCA Problem

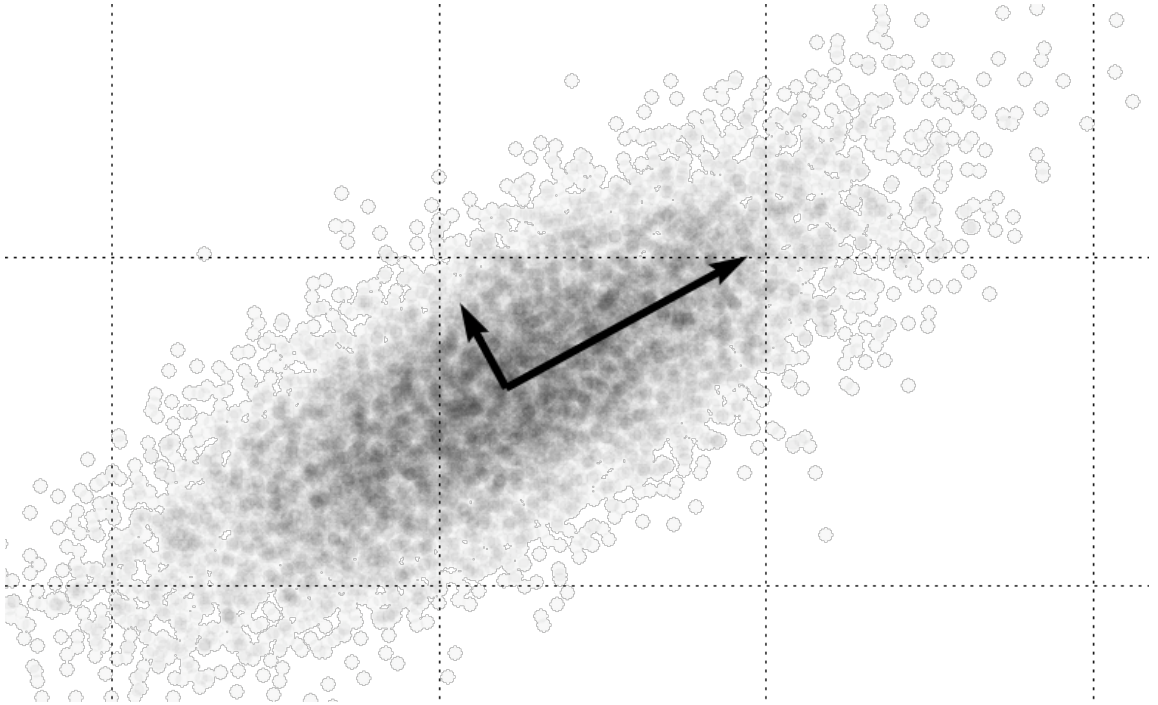
Principal Component Analysis (PCA): Given a set $\{\mathbf{x}^\mu : \mu = 1, ..., M\}$ **of** I**-dimensional data points** $\mathbf{x}^\mu = (x_1^\mu, x_2^\mu, ..., x_I^\mu)^T$ **with zero mean,** $\langle \mathbf{x}^\mu \rangle_\mu = \mathbf{0}_I$**, find an orthogonal matrix** $\mathbf{U}$ **with determinant** $|\mathbf{U}| = +1$ **generating the transformed data points** $\mathbf{x}'^\mu := \mathbf{U}^T\mathbf{x}^\mu$ **such that for any given dimensionality** P **the data projected onto the first** P **axes,** $\mathbf{x}'^\mu_{||} := (x'^\mu_1, x'^\mu_2, ..., x'^\mu_P, 0, ..., 0)^T$**, have the smallest**

$$\text{reconstruction error} \quad E := \langle \|\mathbf{x}'^\mu - \mathbf{x}'^\mu_{||}\|^2 \rangle_\mu \tag{8}$$

among all possible projections onto a P**-dimensional subspace. The row vectors of matrix** $\mathbf{U}$ **define the new axes and are called the *principal components*.**

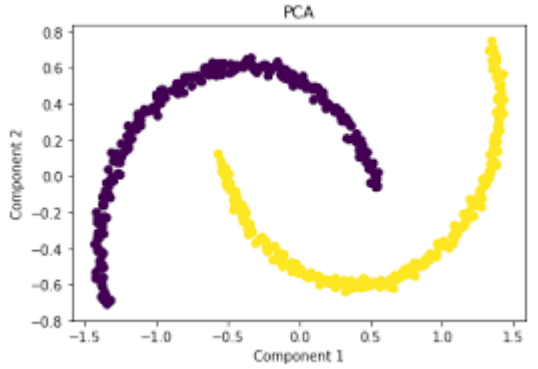
PCA

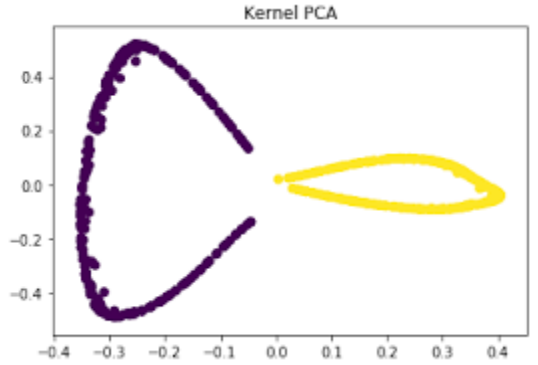
- Key result:
- Exploit spectral analysis of the covariance matrix *C* of the data
- For any integer *p*, the error-minimizing *p*-dimensional subspace is the one spanned by the first *p* eigenvectors of the covariance matrix



Kernel PCA (KPCA)

- Assumption behind PCA is that the data points **x** are well represented by a small-dimensional linear subspace
- Often this assumption does not hold …
- However, it may still be possible that a non-linear transformation $\phi(\mathbf{x})$ "linearizes" the data, but in a higher dimensional space -- then we can perform PCA in the space of $\phi(\mathbf{x})$
- Kernel PCA performs this "lifted" PCA; however, because of "kernel trick," it never computes the mapping $\phi(\mathbf{x})$ explicitly.





Today:
Visual Datasets

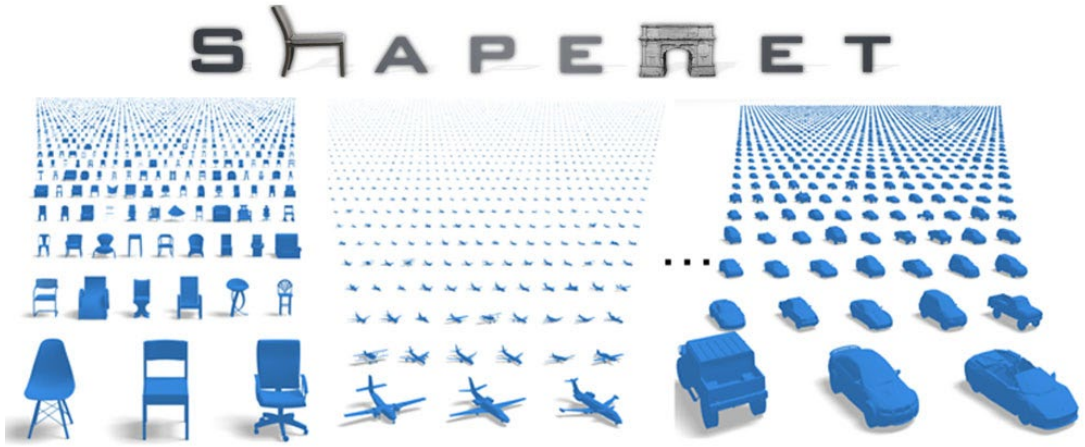
Agenda

Large and high-quality data sets are essential for both training and testing machine learning algorithms

- From semantic networks to visual or geometric data networks
  - WordNet, ImageNet, and ShapeNet
- Approaches for annotation acquisition
  - Difficulty of 3D labels
- From vertical networks to horizontal networks
  - Annotation transportation in ShapeNet

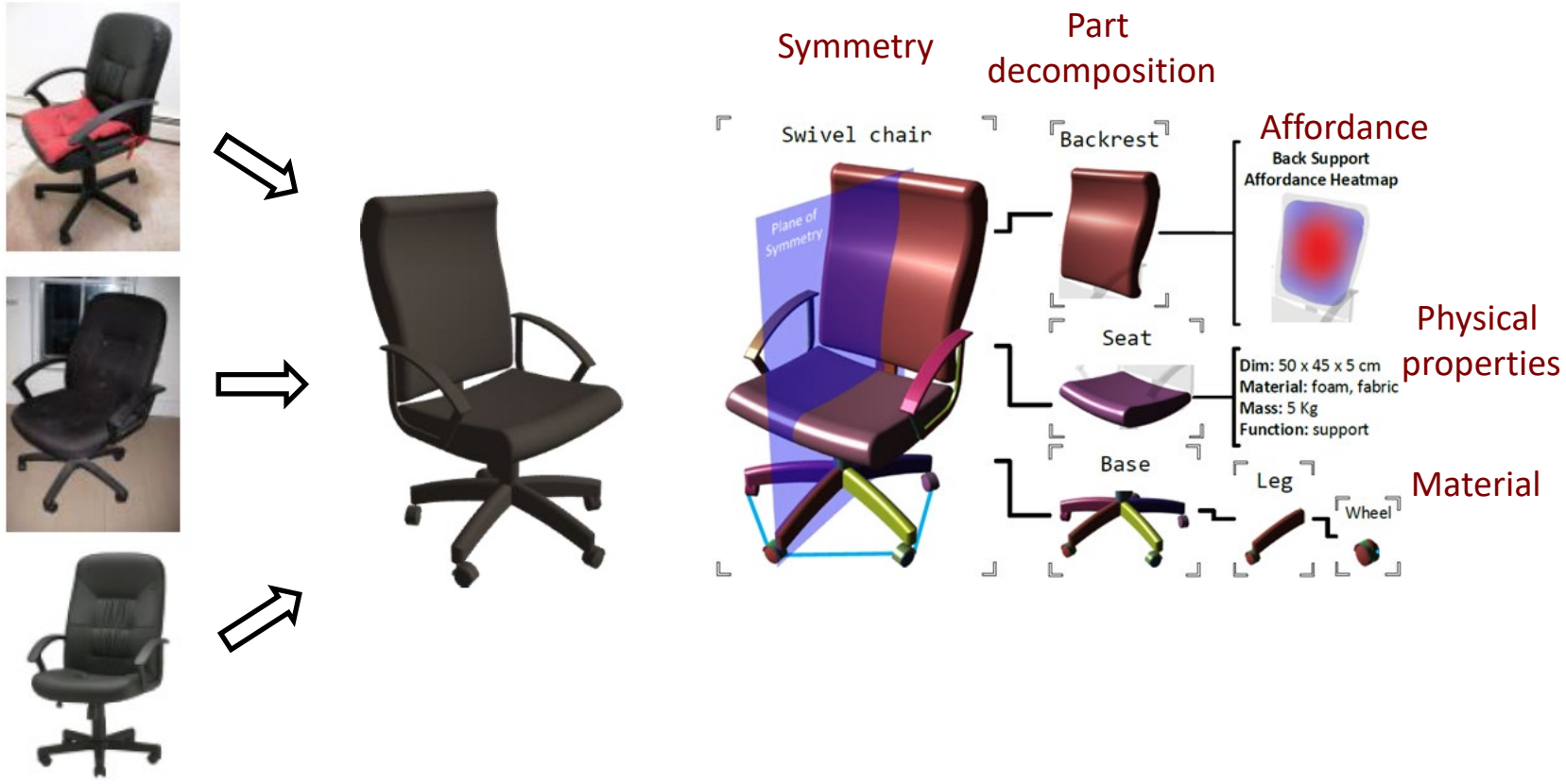
Goal of this Lecture

- Relate geometry and topology to data semantics
- Explain how big visual datasets including ImageNet and ShapeNet are organized



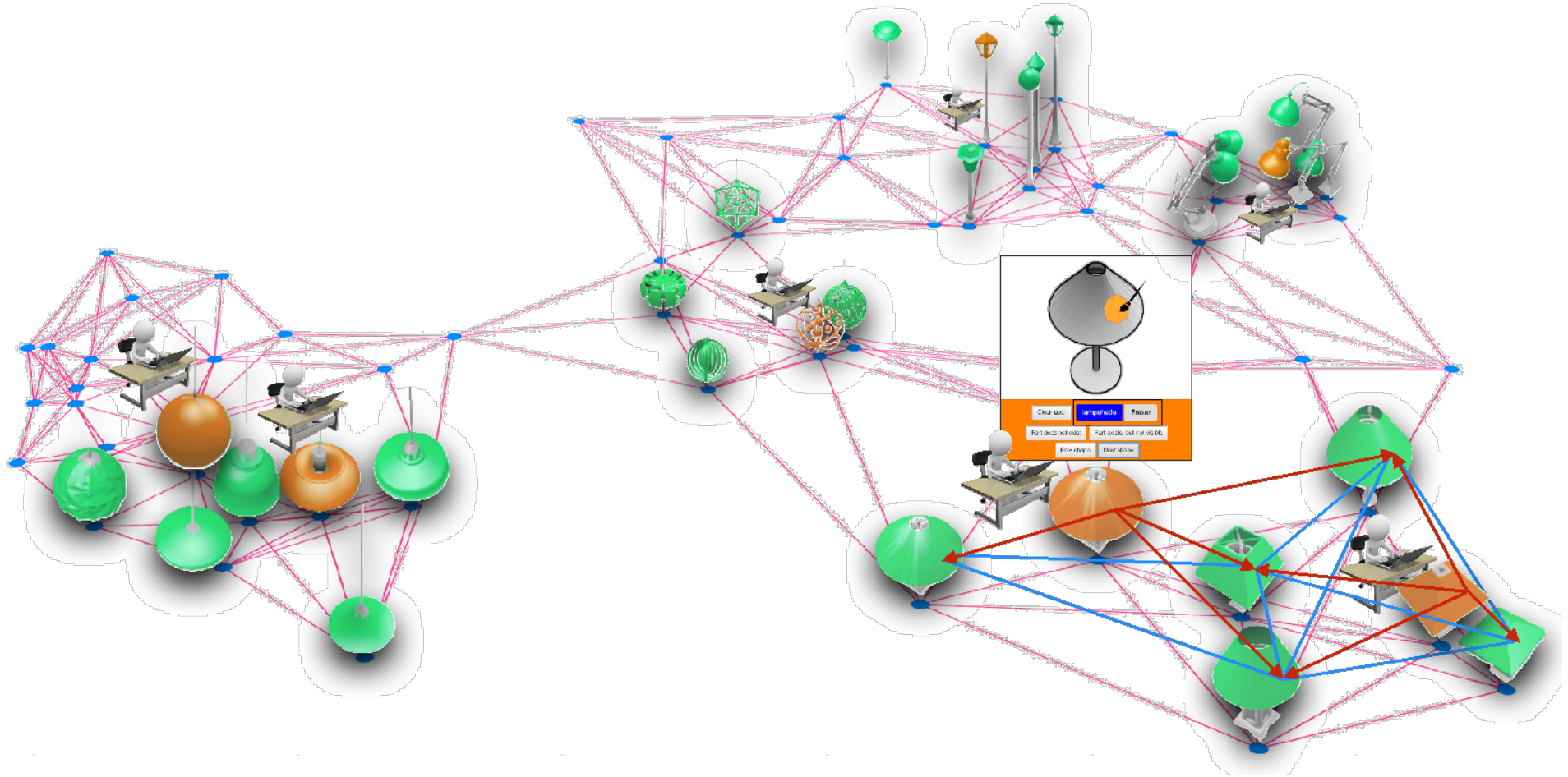
Goal of this Lecture

- Explain how ShapeNet are annotated



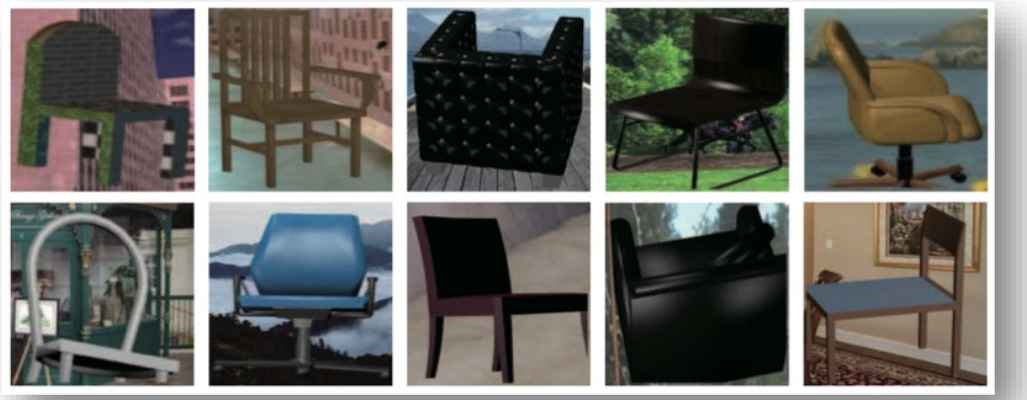
Goal of this Lecture

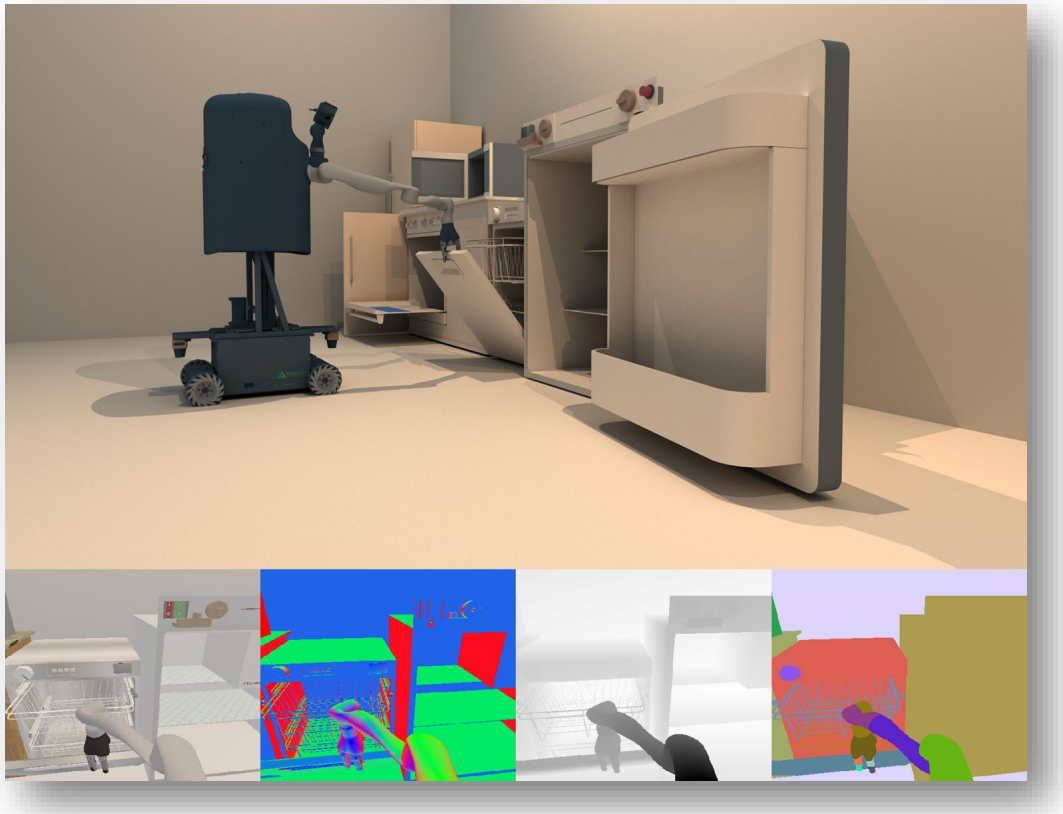
- Show examples of label transportation in a network



Goal of this Lecture

- Using synthetic data and simulation to get geometry + image data





Semantic Networks: Storing Knowledge about the World

Semantic Networks

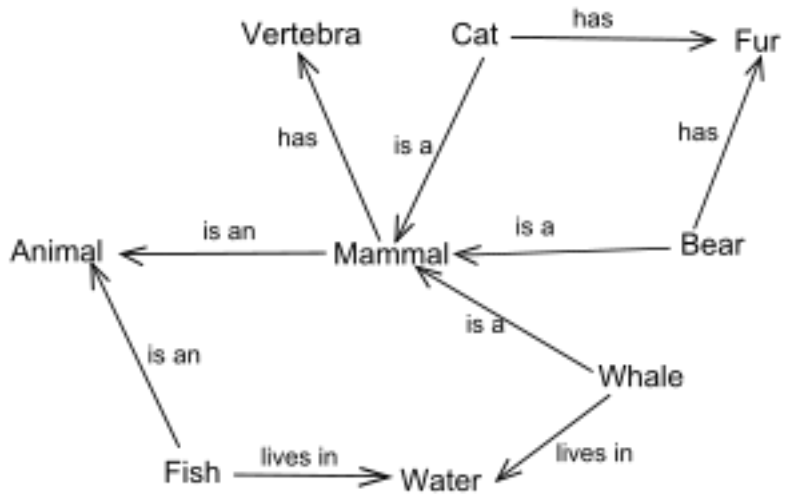
- Also known as **frame networks**
- Encode semantic relations between concepts
- Often used as a form of knowledge representation
- A directed or undirected graph consisting of vertices, which represent concepts, and edges which represent concept relations

Example of a Semantic Net

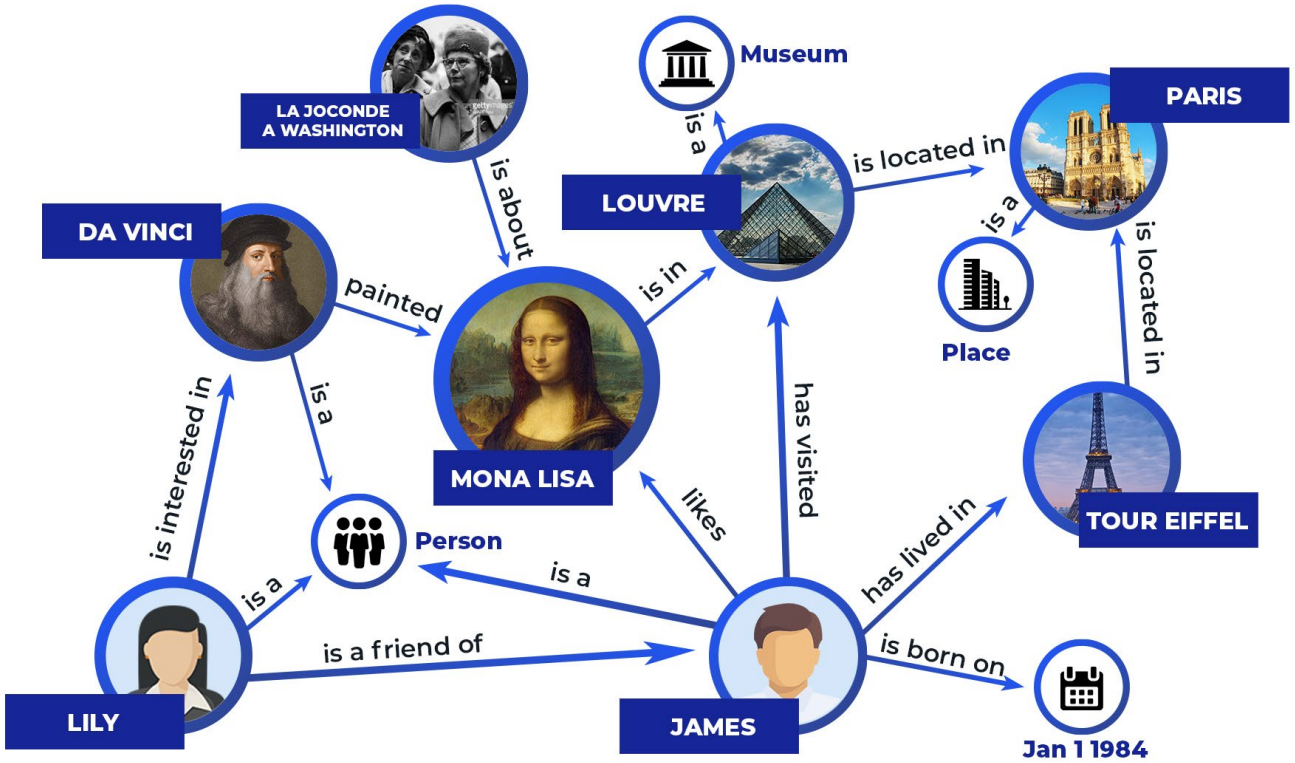
Semantic Net in Lisp

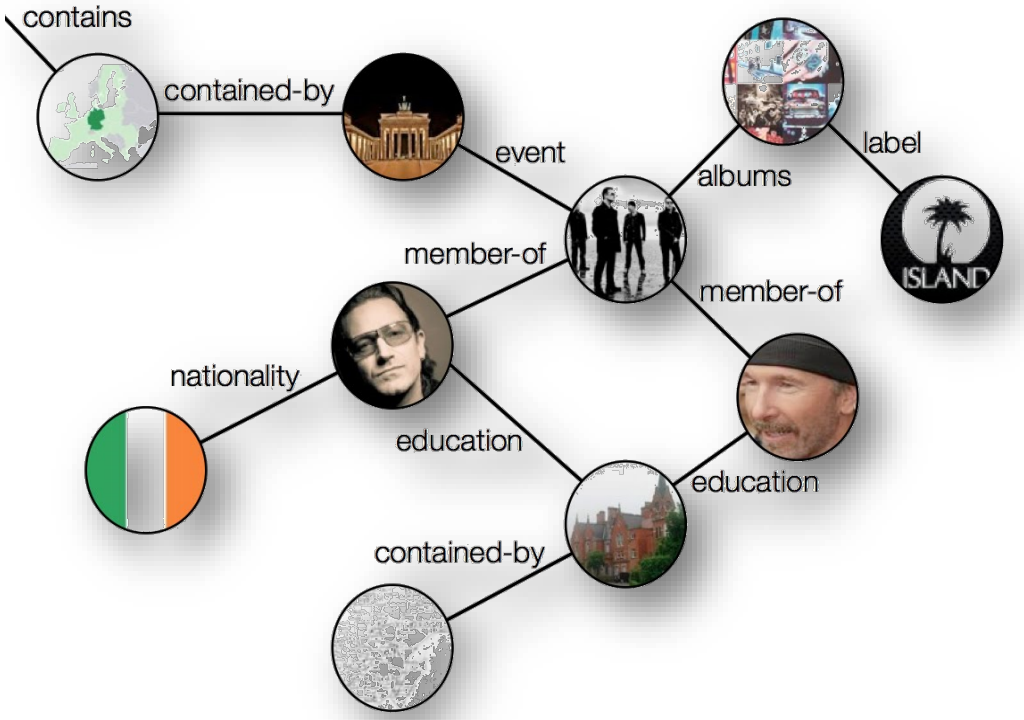
```
(defun *database* ()
'((canary  (is-a bird)
           (color yellow)
           (size small))
  (penguin (is-a bird)
           (movement swim))
  (bird    (is-a vertebrate)
           (has-part wings)
           (reproduction egg-laying))))
```

Graph representation



Google Knowledge Graph





WordNet

What is WordNet?

Original paper by **[George Miller, et al 1990]** cited over 5,000 times

Organizes over 150,000 words into 117,000 categories called *synsets*.

Establishes ontological and lexical relationships in NLP and related tasks.

WordNet

- a lexical database of English
- words -> synonym sets (synsets)

```
dog, domestic dog, Canis familiaris
    => canine, canid
       => carnivore
          => placental, placental mammal, eutherian, eutherian mammal
             => mammal
                => vertebrate, craniate
                   => chordate
                      => animal, animate being, beast, brute, creature, fauna
                         => ...
```

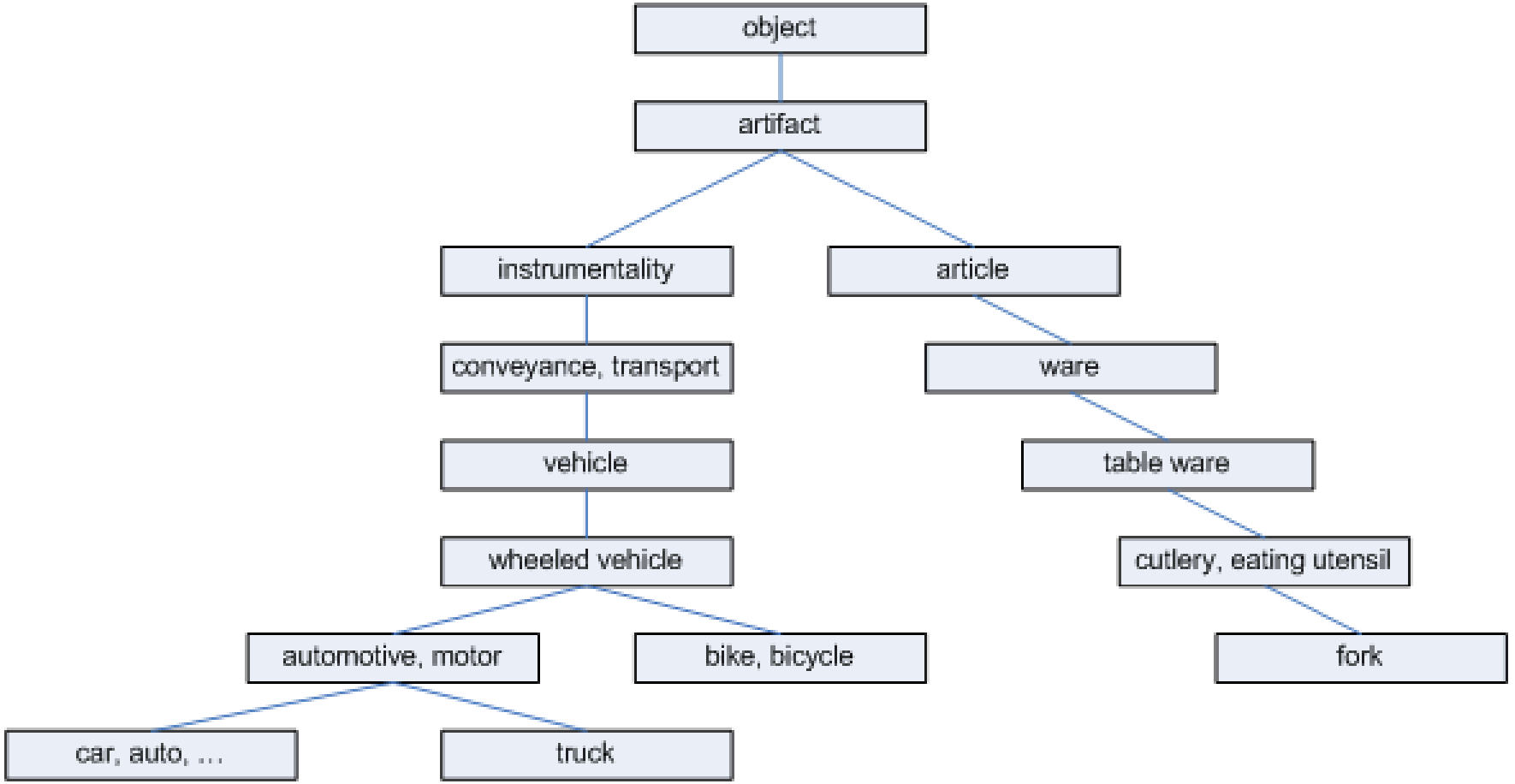
G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. Int. J. Lexicograph.

WordNet

- Important relations between synsets (nouns):

Relation	Definition	Example
Hypernym	From concepts to superordinates	water^1 →liquid
Hyponym	From concepts to subtypes	water^1 →seawater
Has-Part	From groups to their members	water^1 → oxygen
Part-of	From members to their groups	water^1 →ice
Antonym	Opposites	leader→follower

Taxonomy: is-a Relationship



Partonomy: has-a Relationship

- S: (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
  - *direct hyponym* / *full hyponym*
  - ***part meronym***
    - S: (n) accelerator, accelerator pedal, gas pedal, gas, throttle, gun (a pedal that controls the throttle valve) *"he stepped on the gas"*
    - S: (n) air bag (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)
    - S: (n) auto accessory (an accessory for an automobile)
    - S: (n) automobile engine (the engine that propels an automobile)
    - S: (n) automobile horn, car horn, motor horn, horn, hooter (a device on an automobile for making a warning noise)
    - S: (n) buffer, fender (a cushion-like device that reduces shock due to an impact)
    - S: (n) bumper (a mechanical device consisting of bars at either end of a vehicle to absorb shock and prevent serious damage)
    - S: (n) car door (the door of a car)
    - S: (n) car mirror (a mirror that the driver of a car can use)
    - S: (n) car seat (a seat in a car)
    - S: (n) car window (a window in a car)
    - S: (n) fender, wing (a barrier that surrounds the wheels of a vehicle to block splashing water or mud) *"in Britain they call a fender a wing"*
    - S: (n) first gear, first, low gear, low (the lowest forward gear ratio in the gear box of a motor vehicle; used to start a car moving)
    - S: (n) floorboard (the floor of an automobile)
    - S: (n) gasoline engine, petrol engine (an internal-combustion engine that burns gasoline; most automobiles are driven by gasoline engines)
    - S: (n) glove compartment (compartment on the dashboard of a car)
    - S: (n) grille, radiator grille (grating that admits cooling air to car's radiator)
    - S: (n) high gear, high (a forward gear with a gear ratio that gives the greatest vehicle velocity for a given engine speed)
    - S: (n) hood, bonnet, cowl, cowling (protective covering consisting of a metal part that covers the engine) *"there are powerful engines under the hoods of new cowling in order to repair the plane's engine"*
    - S: (n) luggage compartment, automobile trunk, trunk (compartment in an automobile that carries luggage or shopping or tools) *"he put his golf bag in the trunk*
    - S: (n) rear window (car window that allows vision out of the back of the car)
    - S: (n) reverse, reverse gear (the gears by which the motion of a machine can be reversed)
    - S: (n) roof (protective covering on top of a motor vehicle)
    - S: (n) running board (a narrow footboard serving as a step beneath the doors of some old cars)
    - S: (n) stabilizer bar, anti-sway bar (a rigid metal bar between the front suspensions and between the rear suspensions of cars and trucks; serves to stabilize the cl
    - S: (n) sunroof, sunshine-roof (an automobile roof having a sliding or raisable panel) *"'sunshine-roof' is a British term for 'sunroof'"*
    - S: (n) tail fin, tailfin, fin (one of a pair of decorations projecting above the rear fenders of an automobile)
    - S: (n) third gear, third (the third from the lowest forward ratio gear in the gear box of a motor vehicle) *"you shouldn't try to start in third gear"*
    - S: (n) window (a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened)

From Semantic Networks to Visual Data Networks

- Instantiate *concepts* by *exemplars*
- Concepts from WordNet
  - Defined by properties (using language)
- Exemplars from sensor data
  - images (ImageNet)
  - 3D shapes (ShapeNet)
  - videos

Grounding concepts to the real world

- S: (n) chair (a seat for one person, with a support for the back)

Why Go from a Semantic Network to a Visual Data Network ?

- “A picture is worth a thousand words”
- Concepts and their relationships emerge directly from data

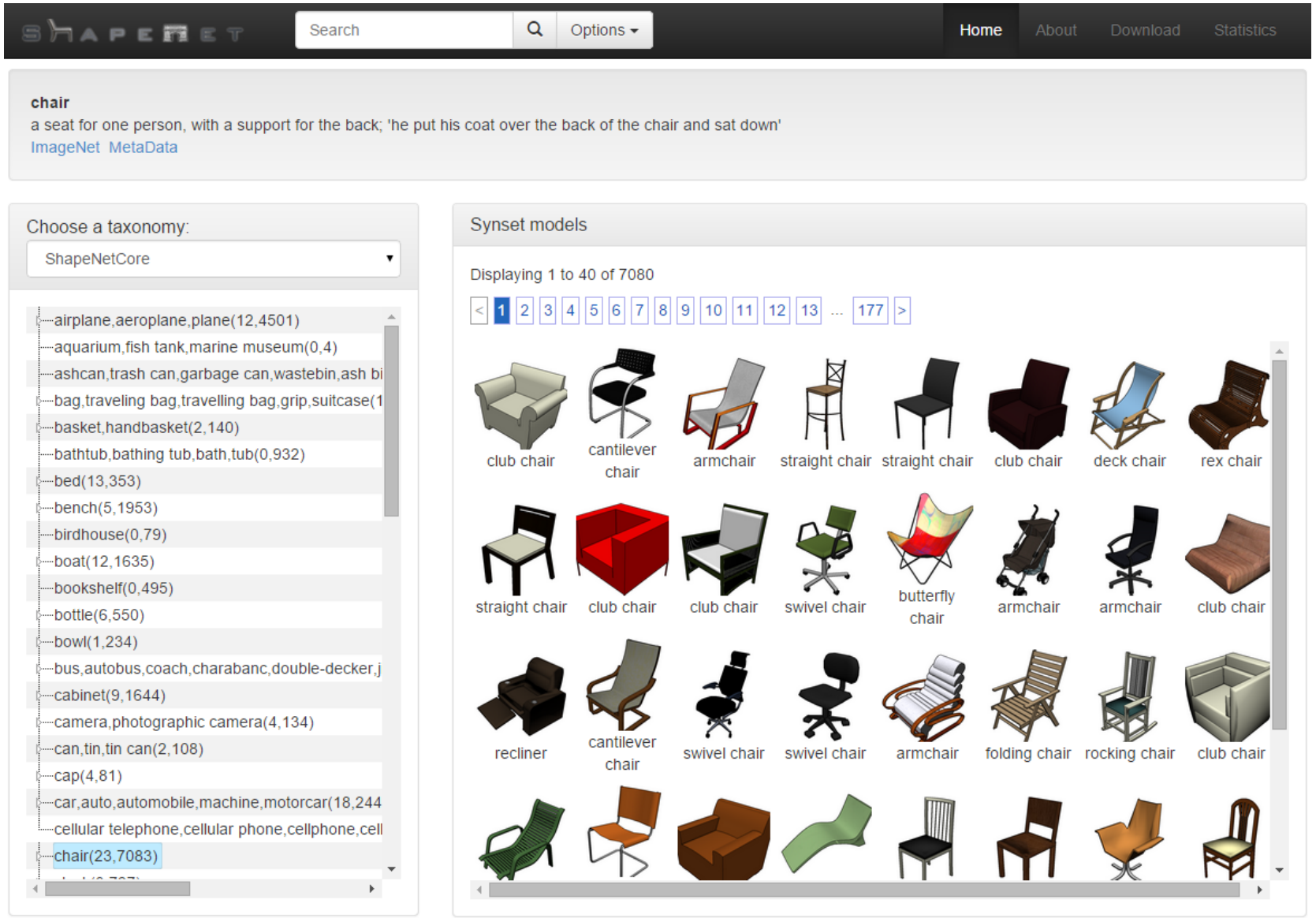
IMAGENET is a knowledge ontology

- Taxonomy (with WordNet backbone)

- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                        - S: (n) living thing, animate thing (a living (or once living) entity)
                          - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                            - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                              - S: (n) physical entity (an entity that has physical existence)
                                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

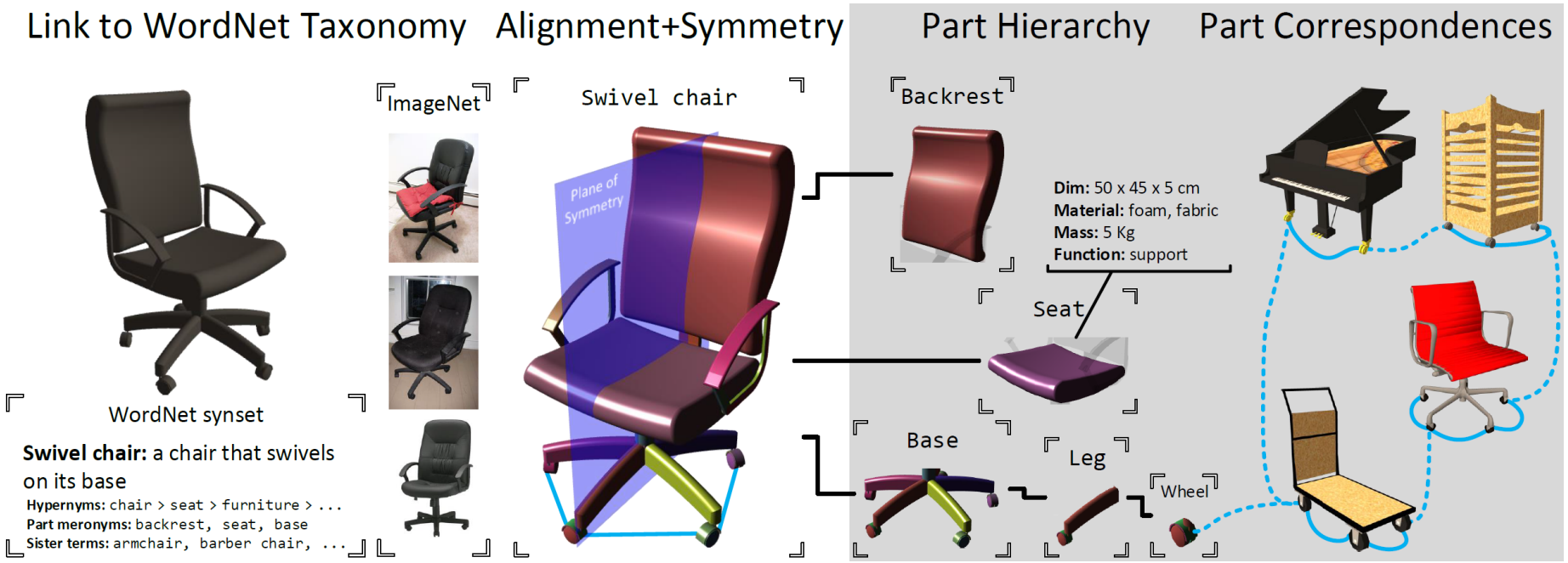
Slide Credit: Fei-Fei Li, Jia Deng

ShapeNet (>3M Models)



Object Knowledge

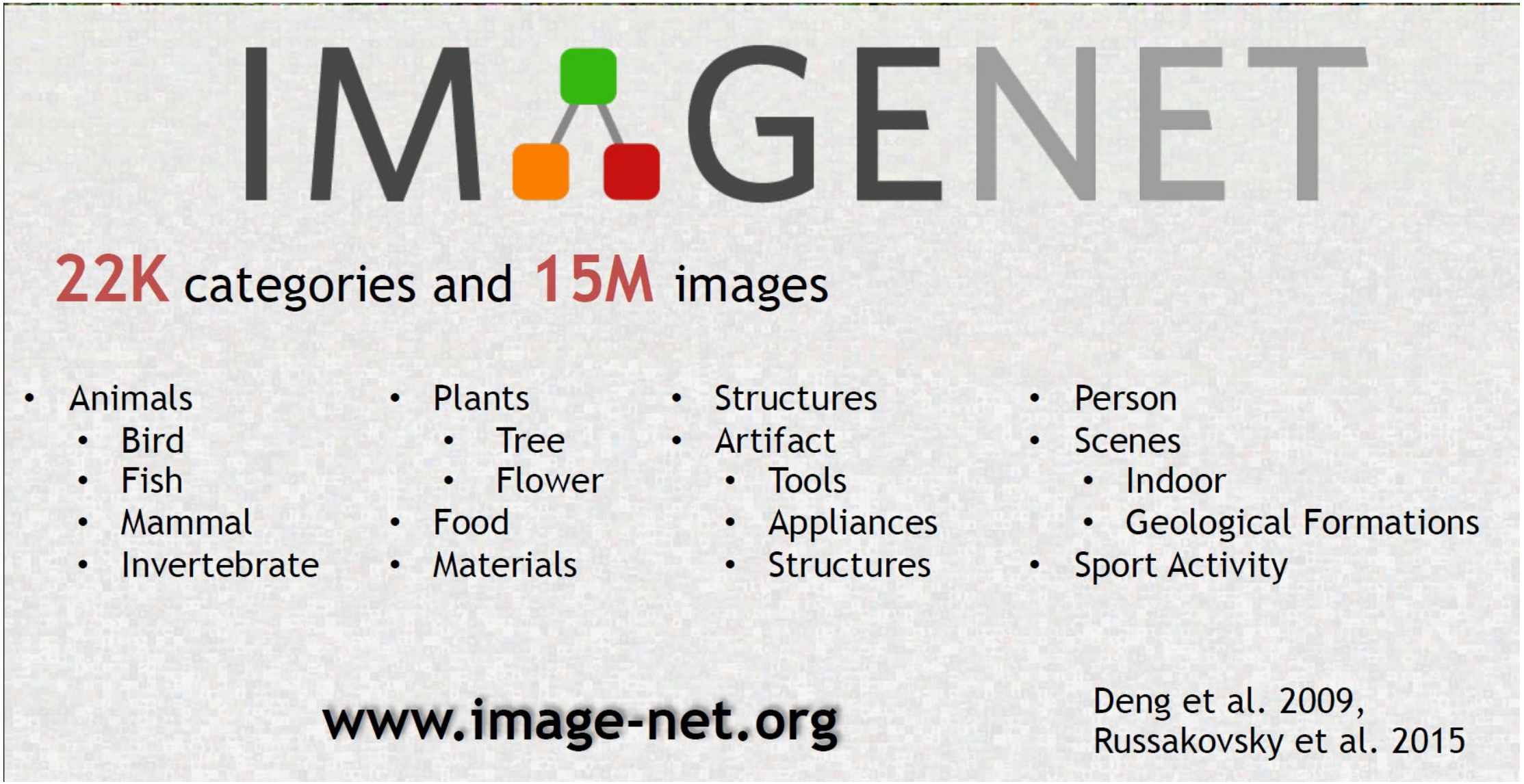
Parts, symmetries, keywords, physical properties, materials, affordances, …



ImageNet

Slide Credit: Fei-Fei Li, Jia Deng

ImageNet


IMAGENET
22K categories and 15M images
Animals
Bird
Fish
Mammal
Invertebrate
Plants
Tree
Flower
Food
Materials
Structures
Artifact
Tools
Appliances
Structures
Person
Scenes
Indoor
Geological Formations
Sport Activity
www.image-net.org
Deng et al. 2009,
Russakovsky et al. 2015

Illustrating WordNet Nodes

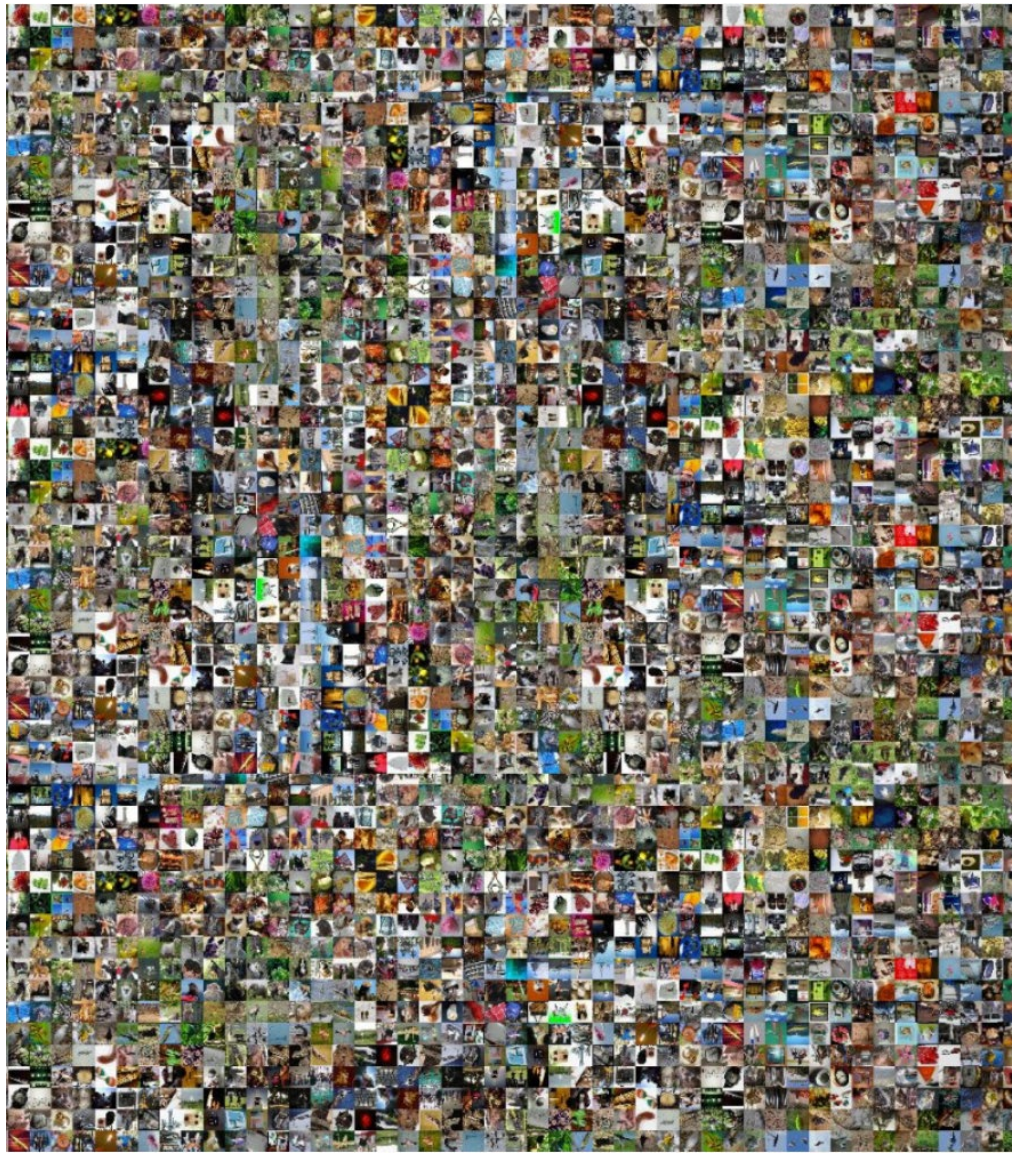
Individually Illustrated WordNet Nodes

jacket: a short coat

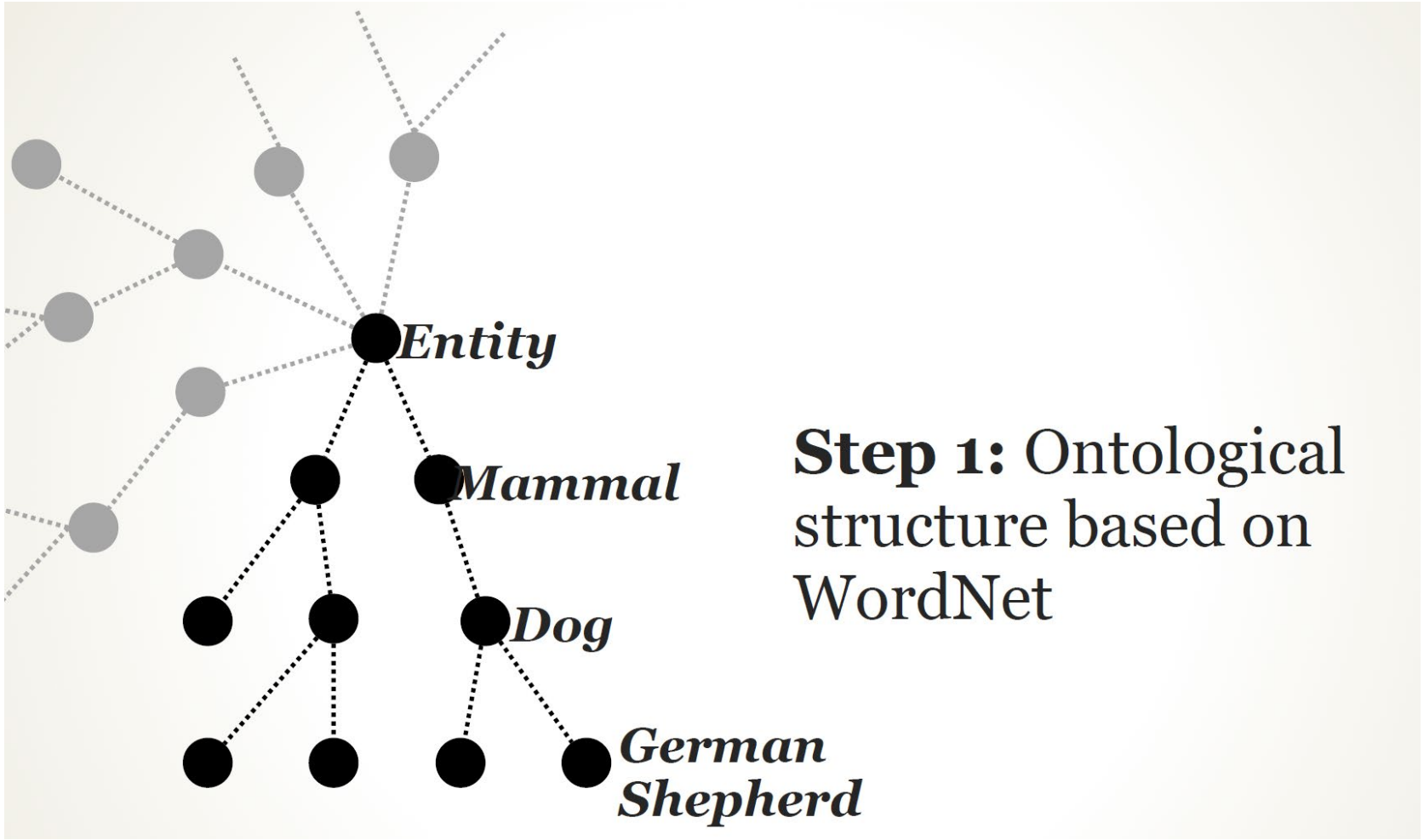
German shepherd: breed of large shepherd dogs used in police work and as a guide for the blind.

microwave: kitchen appliance that cooks food by passing an electromagnetic wave through it.

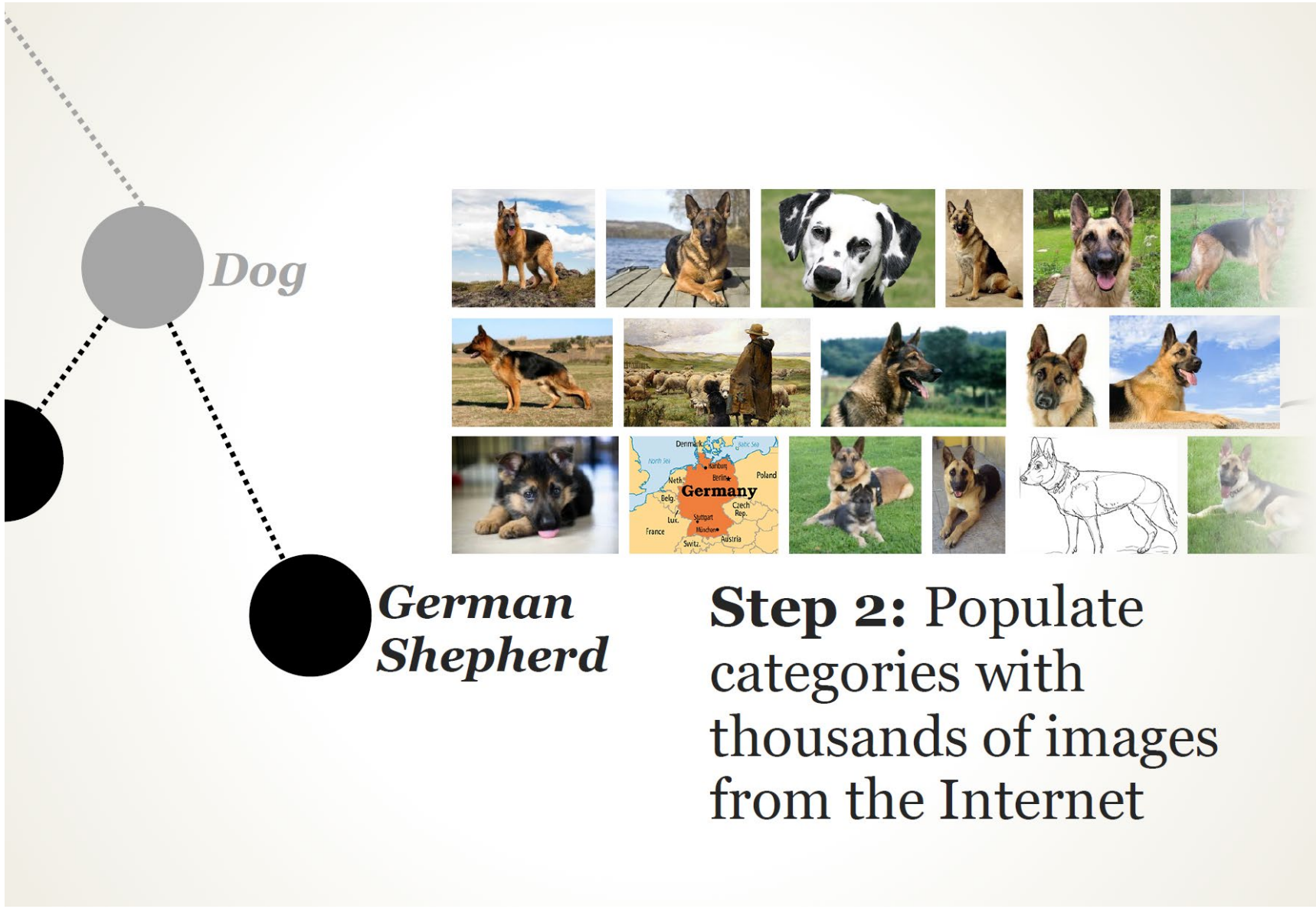
mountain: a land mass that projects well above its surroundings; higher than a hill.



WordNet Ontology



“Illustrating” WordNet



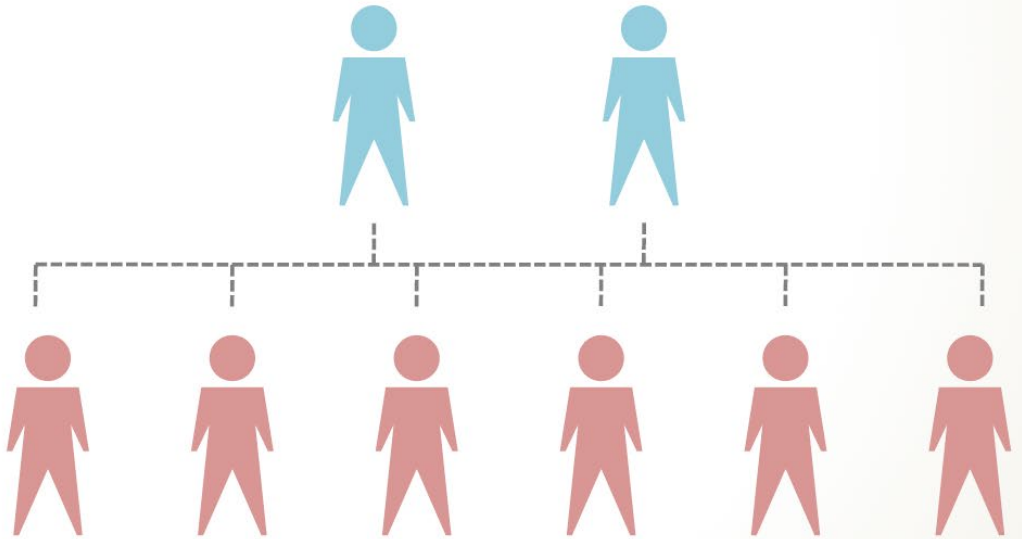
Cleaning Up the Results

Cleaning Up …

1st Attempt: The Psychophysics Experiment

ImageNet PhD Students

Miserable Undergrads



Cleaning Up …

1st Attempt: The Psychophysics Experiment

- # of synsets: **40,000** (subject to: imageability analysis)
- # of candidate images to label per synset: **10,000**
- # of people needed to verify: **2-5**
- Speed of human labeling: **2 images/sec** (one fixation: ~200msec)
- **Massive parallelism (N ~ 10^2-3)**

40,000 × 10,000 × 3 / 2 = 6000,000,000 sec ≈ 19 years

Classify and Collect

2nd Attempt: Human-in-the-Loop Solutions

Towards scalable dataset construction: An active learning approach

Brendan Collins, Jia Deng, Ka
{bmcollin, dengjia, li, feifei

Department of Computer Science, Princeton

Abstract. As computer vision research co
and greater variation within object categor
more exhaustive datasets are necessary. H
ing such datasets is laborious and monoto
in which many images have been automa
category (typically by automatic internet s
relevant images from noise. We present a d
which employs active, online learning to
with minimal user input. The principle ad
vious endeavors is its scalability. We demon
superior to the state-of-the-art, with scala
work.

1 Introduction

Though it is difficult to foresee the future of c
that its trajectory will include examining a g
(such as objects or scenes), that the complexit
categories will increase, and that these catego
variation. It is unlikely that the researcher's
keep pace with the growing need for annotat
work aims to develop a system which can obta
ages with minimal supervision. The particula

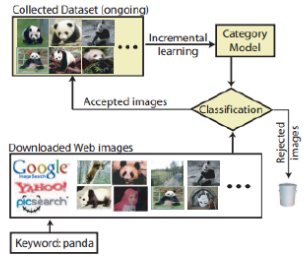
OPTIMOL: automatic Online Picture collecTion via Incremental MOdel Learning

Li-Jia Li[1], Gang Wang[1] and Li Fei-Fei[2]

[1] Dept. of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, USA
[2] Dept. of Computer Science, Princeton University, USA
jialil3@uiuc.edu, gwang6@uiuc.edu, feifeili@cs.princeton.edu

Abstract

A well-built dataset is a necessary starting point for advanced computer vision research. It plays a crucial role in evaluation and provides a continuous challenge to state-of-the-art algorithms. Dataset collection is, however, a tedious and time-consuming task. This paper presents a novel automatic dataset collecting and model learning approach that uses object recognition techniques in an incremental method. The goal of this work is to use the tremendous resources of the web to learn robust object category models in order to detect and search for objects in real-world cluttered scenes. It mimics the human learning process of iteratively accumulating model knowledge and image examples. We



Classify and Collect

2nd Attempt: Human-in-the-Loop Solutions

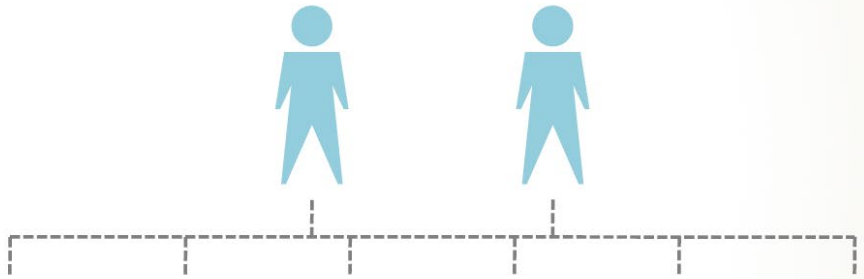
Machine-generated datasets can only match the best algorithms of the time.

Human-generated datasets transcend algorithmic limitations, leading to better machine perception.

Massive Parallelism

3rd Attempt: Crowdsourcing

ImageNet
PhD Students

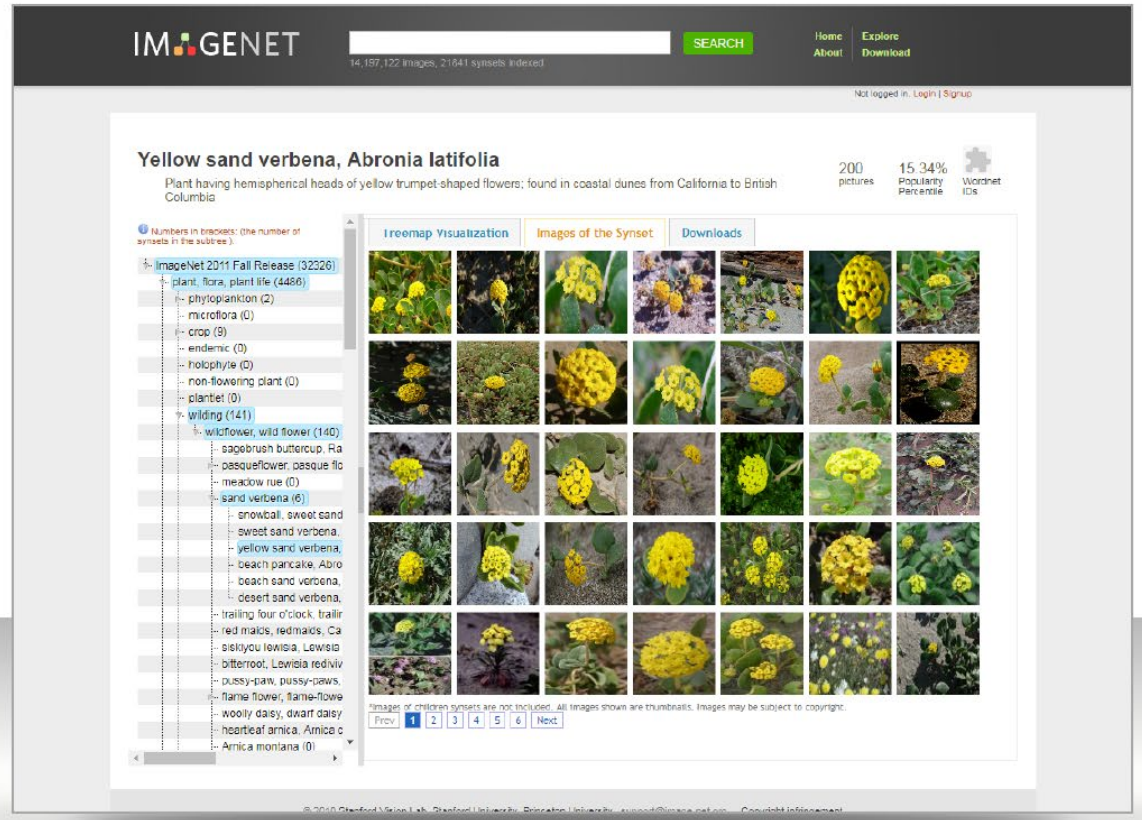


Crowdsourced Labor

amazon mechanical turk™
Artificial Artificial Intelligence

49k Workers *from* 167 Countries
2007-2010

The Result: IMAGENET Goes Live in 2009



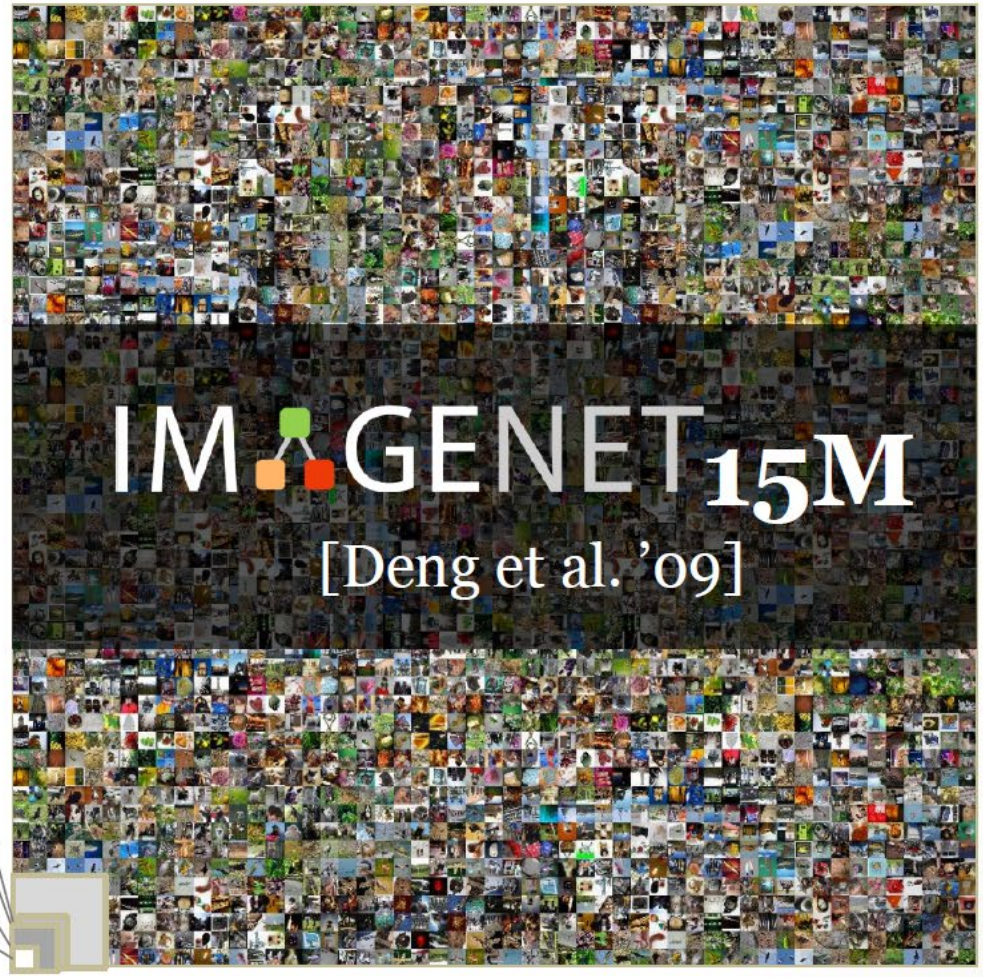
ImageNet Targeted Scale

SUN, 131K
[Xiao et al. '10]

LabelMe, 37K
[Russell et al. '07]

PASCAL VOC, 30K
[Everingham et al. '06-'12]

Caltech101, 9K
[Fei-Fei, Fergus, Perona, '03]



ImageNet Yearly Challenges

1. Training data released: images and annotations
    - For classification, 1000 synsets with ~1k images/synset
2. Test data released: images only (annotations hidden)
    - For classification, ~ 100 images/synset
3. Participants train their models on train data
4. Submit text file with predictions on test images
5. Evaluate and release results, and run a workshop at ECCV/ICCV to discuss result

ImageNet Challenge Tasks

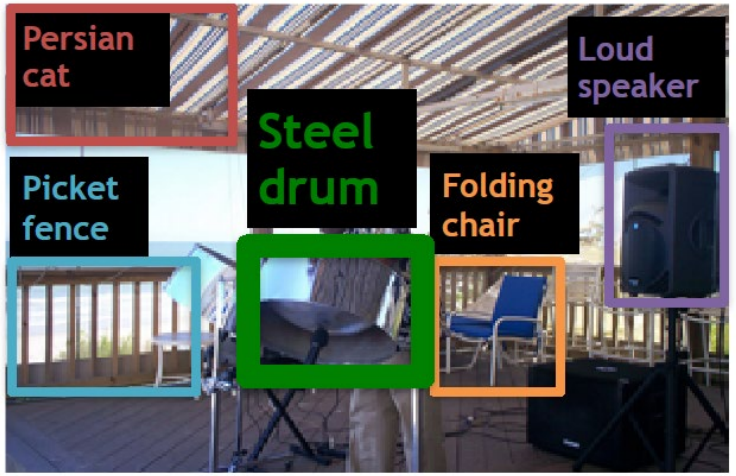
Steel drum

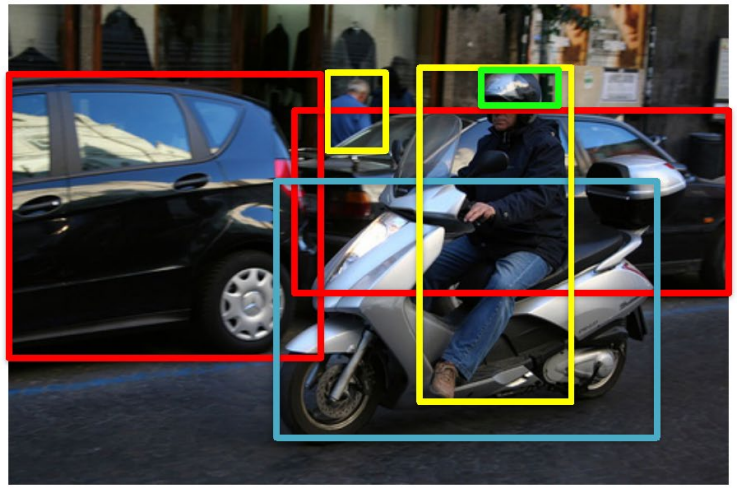
Objects: 1000 classes
Training: 1.2M images
Validation: 50K images
Test: 100K images

Classification

Classification + Localization

Output





Person
Car
Motorcycle
Helmet

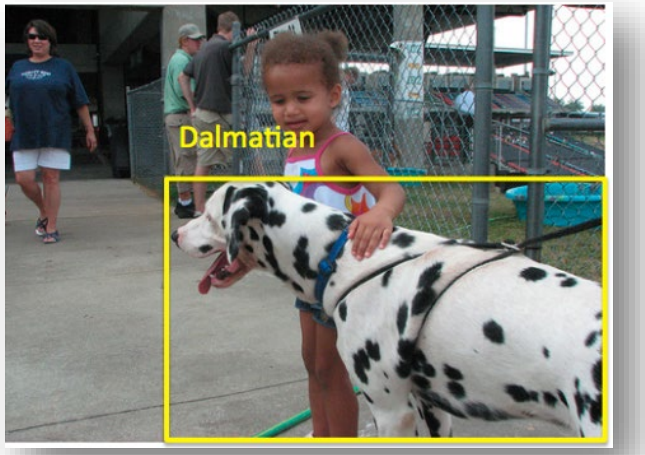
Object Detection

Objects: 200 classes
Training: 450K images, 470K bounding boxes
Validation: 20K images, all bounding boxes
Test: 40K images, all bounding boxes

Limitations of ImageNet

From knowledge representation perspective

- Captures only shallow information in images



Object name, bounding box location

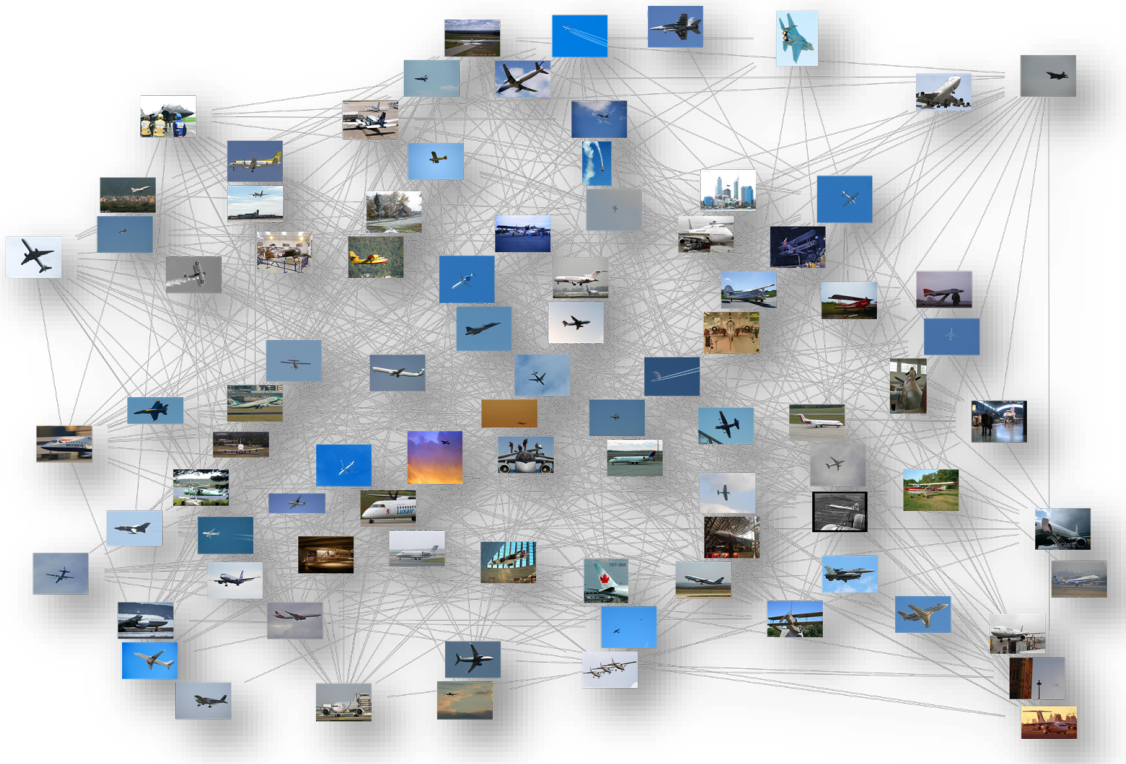
- Geometric and physical knowledge of objects is missing (e.g. ShapeNet)
- Relationships among objects are missing (e.g. VisualGenome)

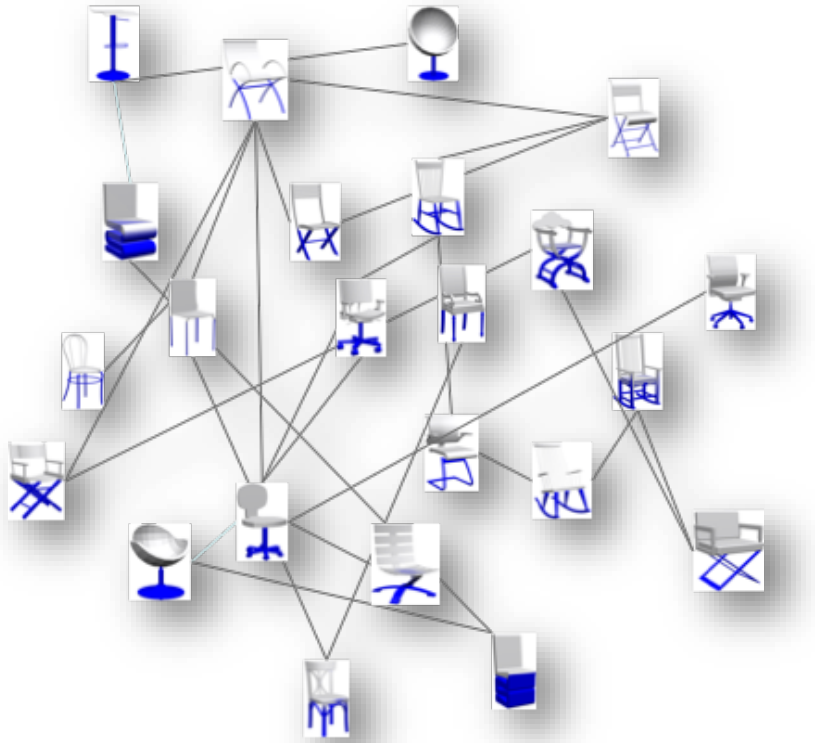
Geometry and Physical Knowledge of Objects

3D Opportunities: Encoding Knowledge

Among all digital representations we have of a real artifact, 3D is the most faithful to the actual physical object

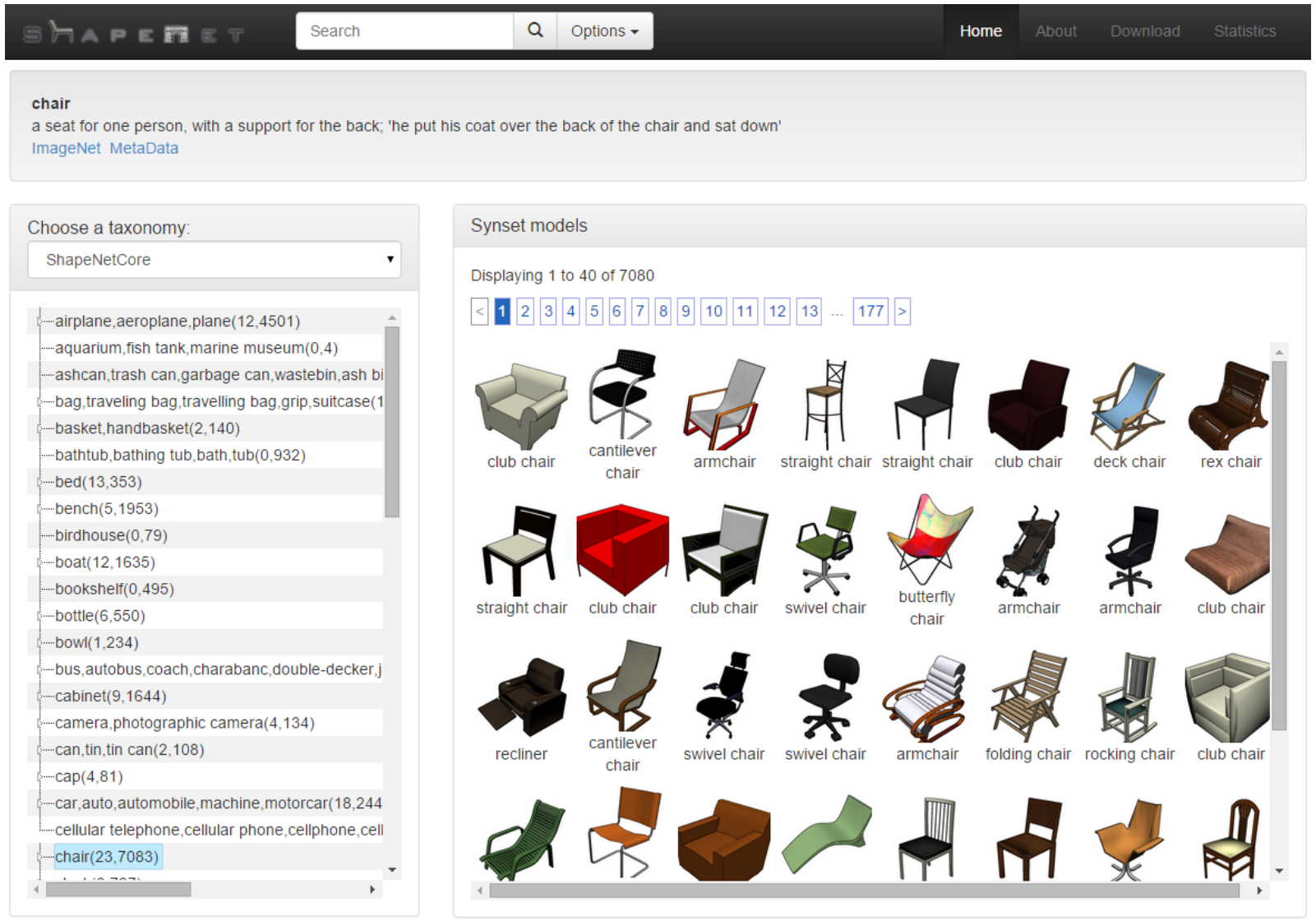
Information Transport





ShapeNet

ShapeNet (>3M Models) https://www.shapenet.org/



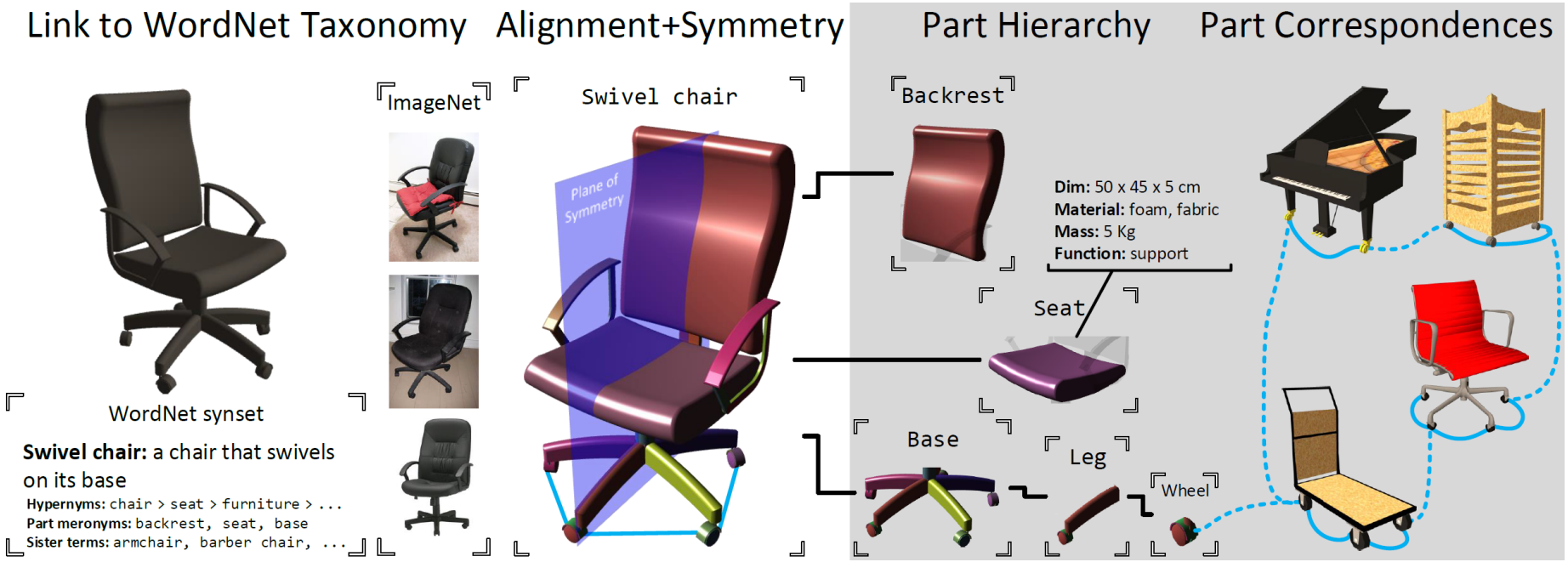
Stanford:
Leonidas Guibas
Pat Hanrahan
Silvio Savarese

Princeton:
Tom Funkhouser
Jianxiong Xiao

UT Austin:
Qixing Huang

Object Knowledge: ShapeNet

Parts, symmetries, keywords, physical properties, materials, affordances, …



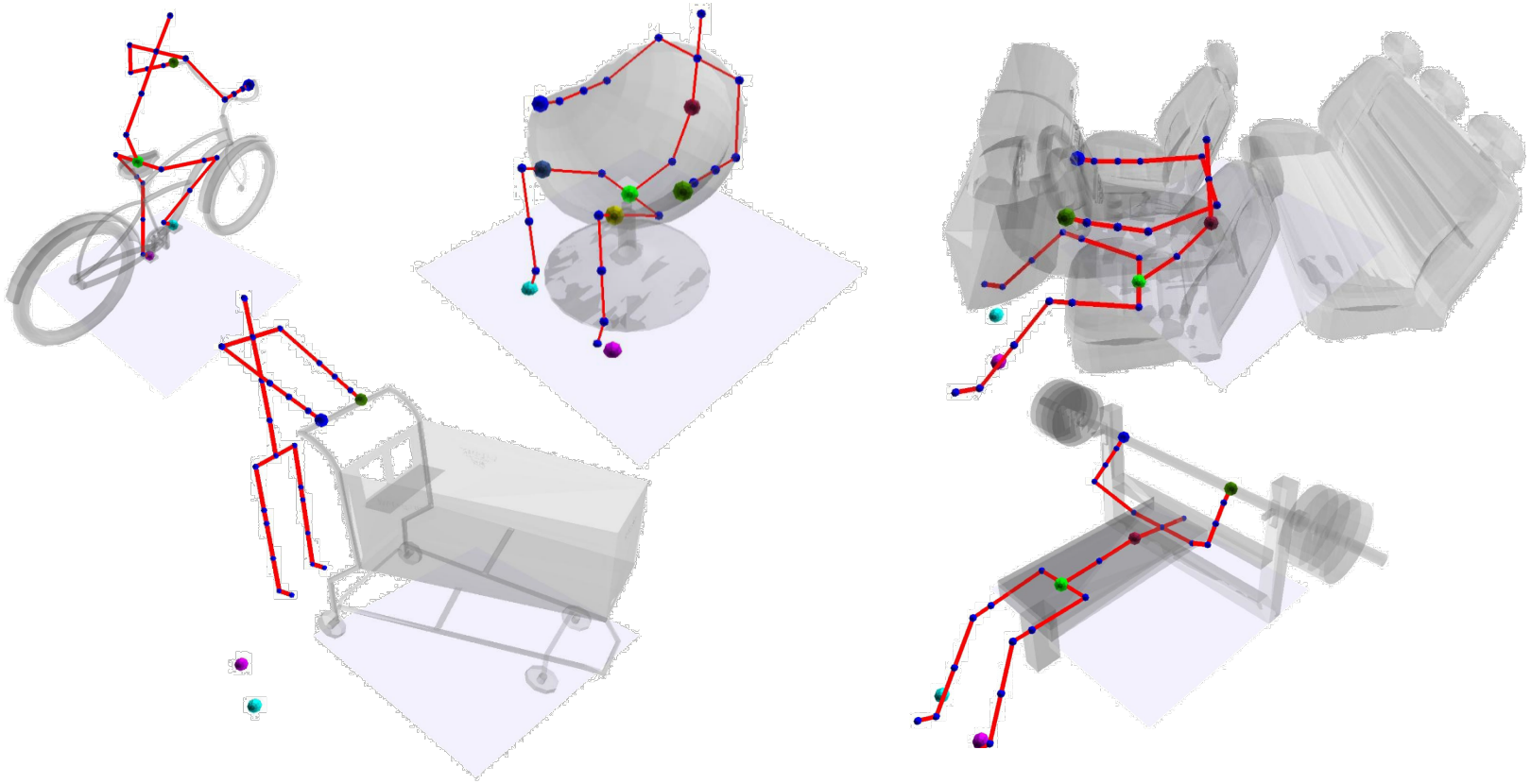
Where is in ShapeNet currently?

- ShapeNetCore
  - 51,300 textured 3D models classified into 55 classes, mostly man-made objects
  - Mesh, point cloud, volumetric representations are provided
  - Consistent orientation within each class
  - Semantic part annotation for a subset

Where is in ShapeNet currently?

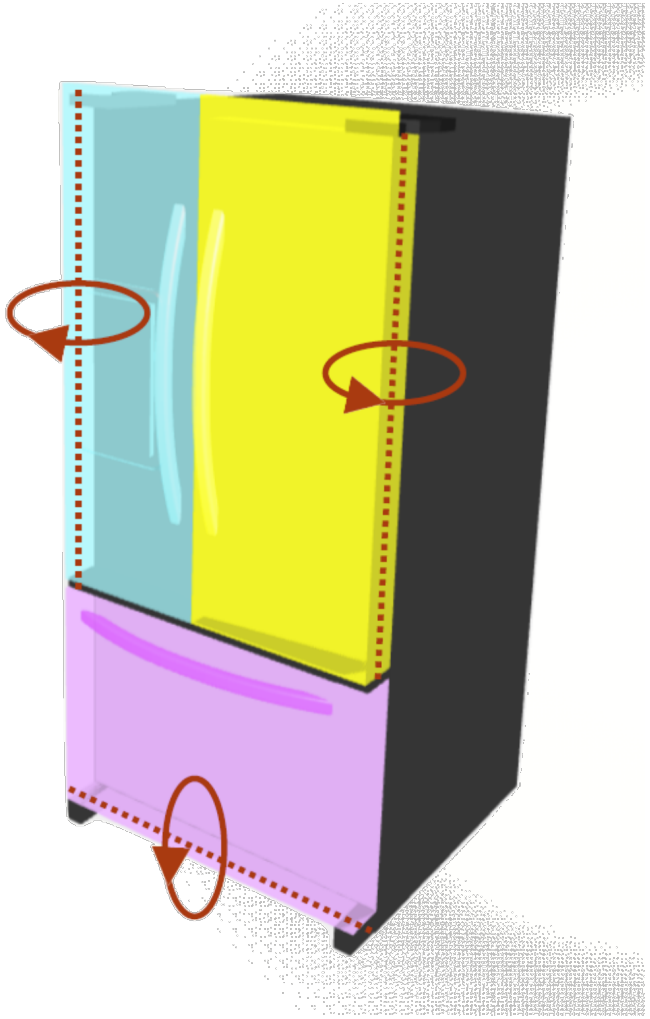
- ShapeNetCore
- ShapeNetSem
  - 12,000 textured models classified into 270 categories, indoor objects
  - Mesh, volumetric representations are provided
  - Consistent orientation within each class
  - Physical dimensions and weights

Object Affordances

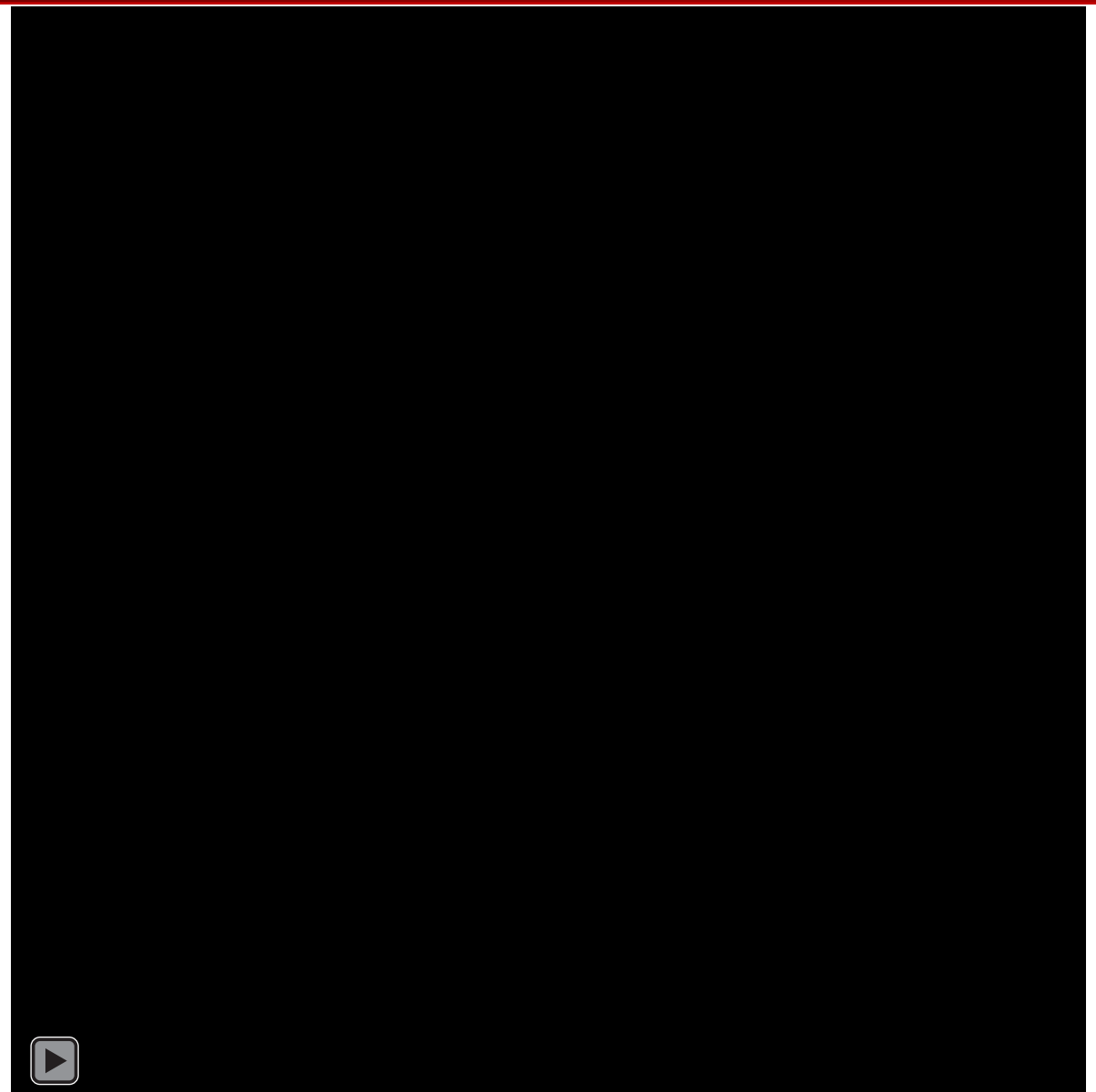


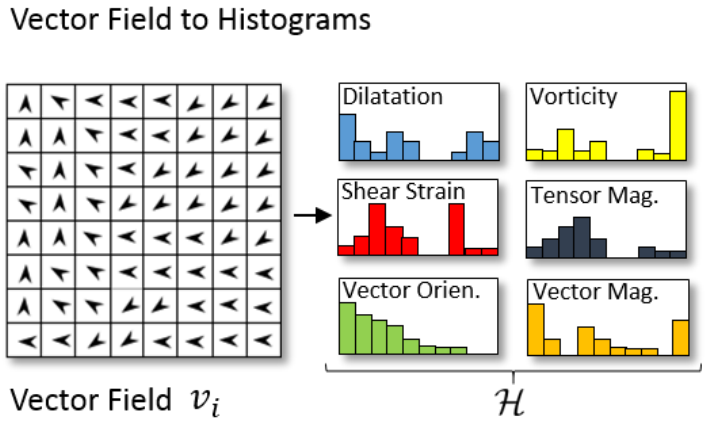
[V. Kim, S. Chaudhuri, L. Guibas, and T. Funkhouser, Siggraph 2014]

Object “Active Sites”



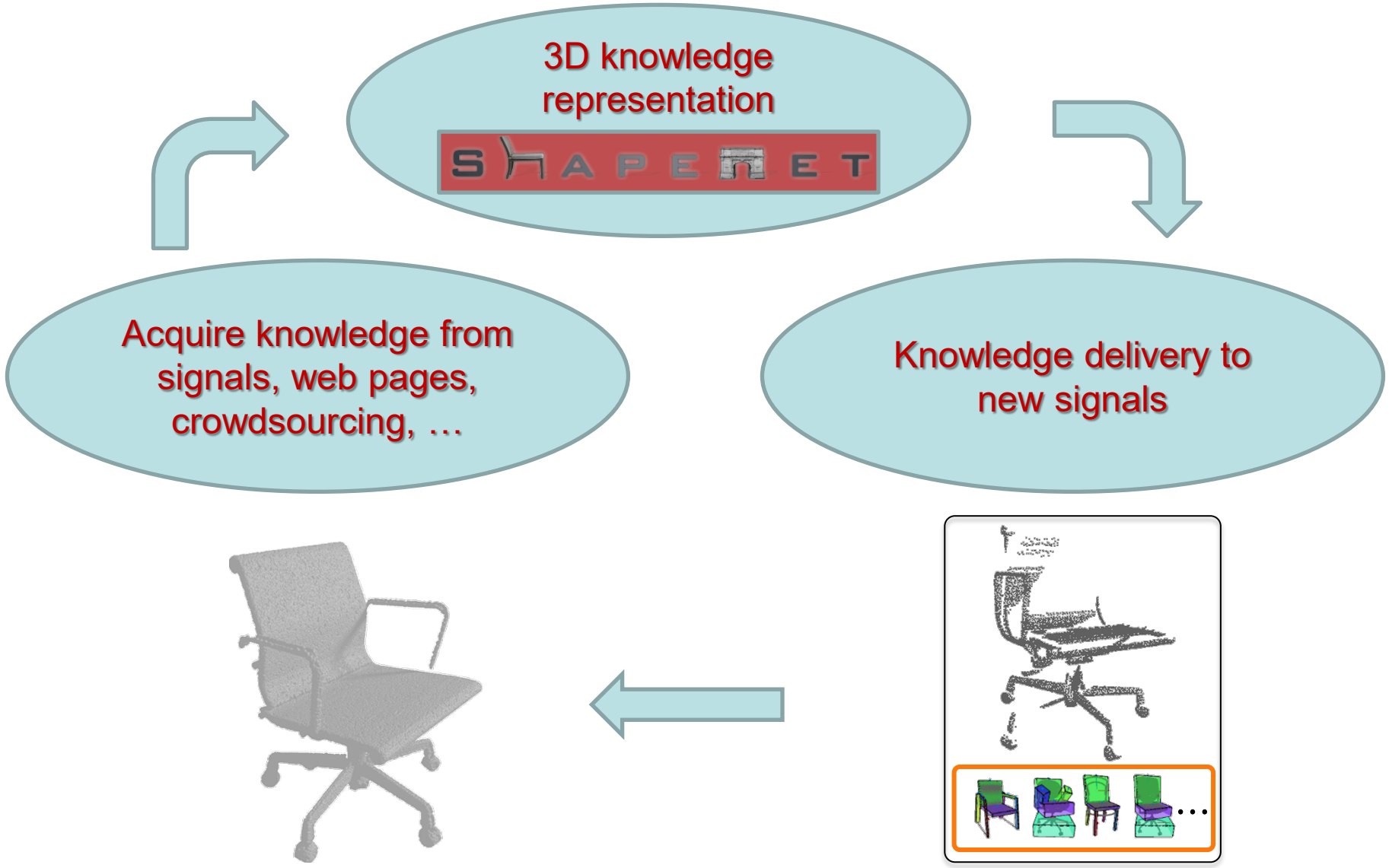
Object Interaction Knowledge





[S. Pirk et al., *Understanding and Exploiting Object Interaction Landscapes*, ACM TOG 2017]

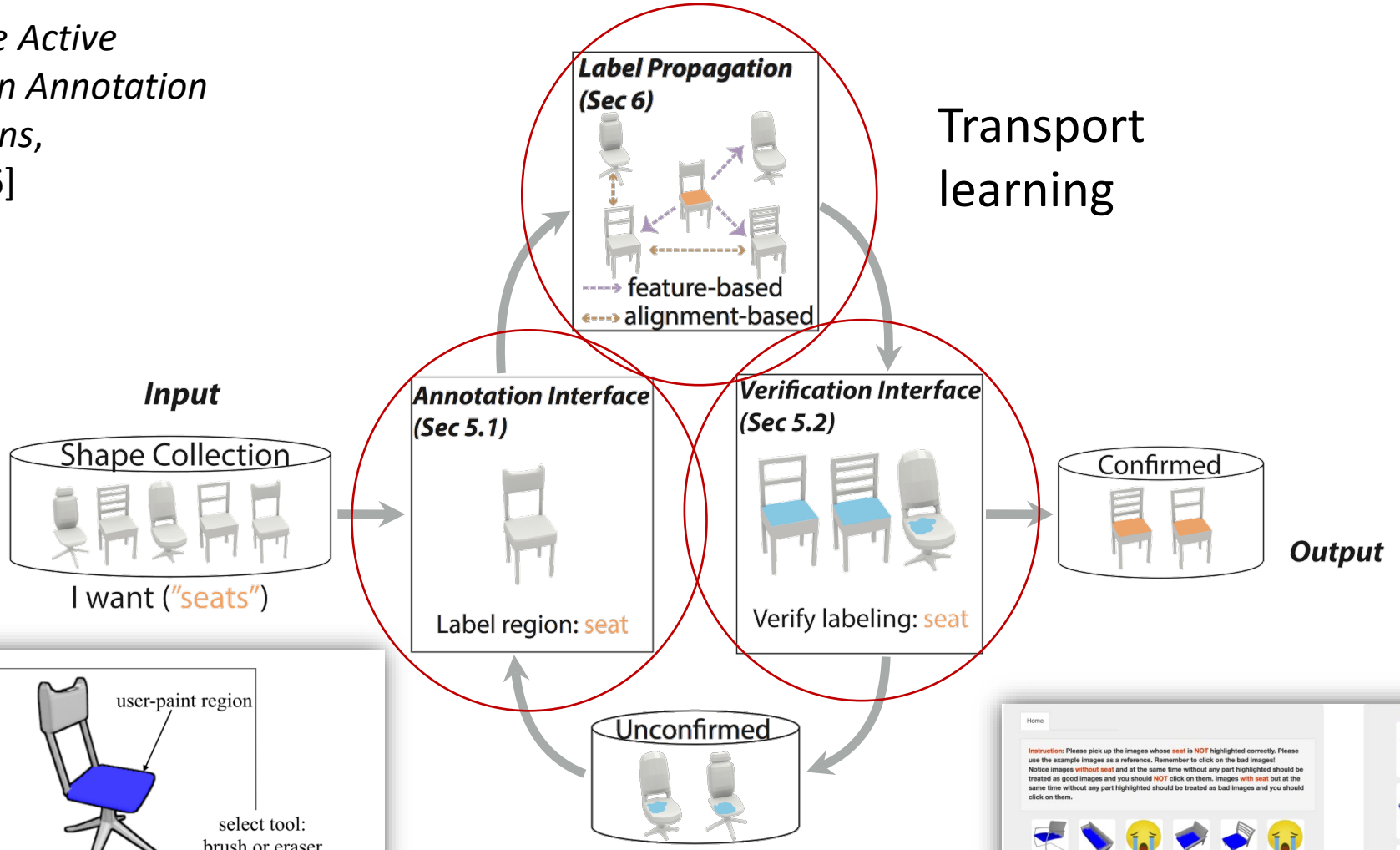
Focus: Knowledge Transport

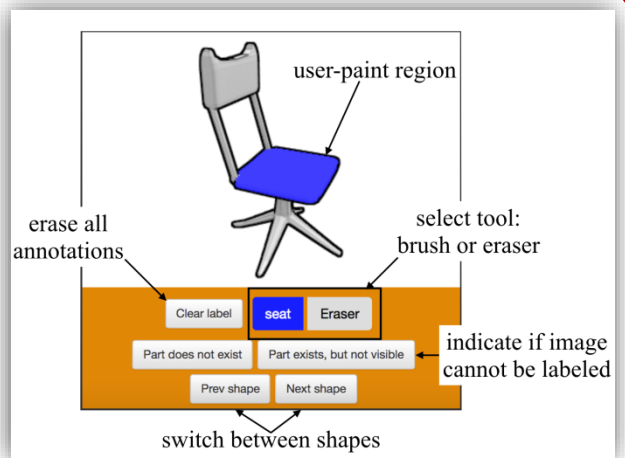


Object Part Annotation

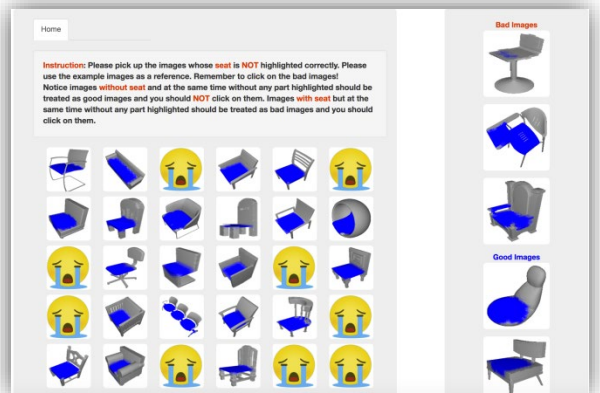
Model Part Annotation

[Li Yi et al., *A Scalable Active Framework for Region Annotation in 3D Shape Collections*, SIGGRAPH Asia, 2016]

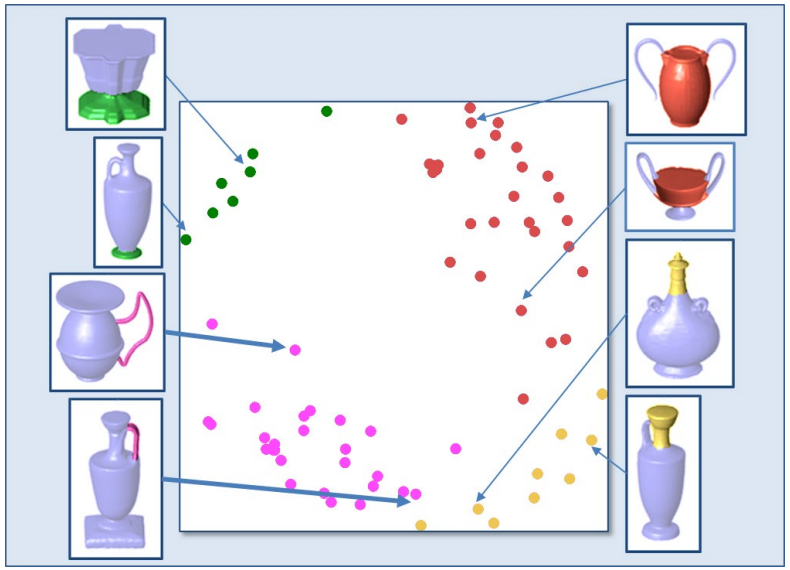




2D annotations

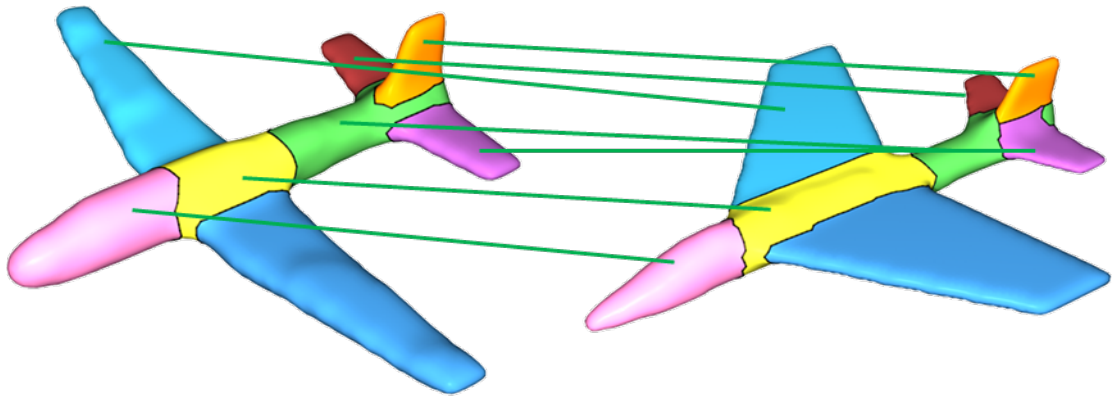


Annotation Propagation



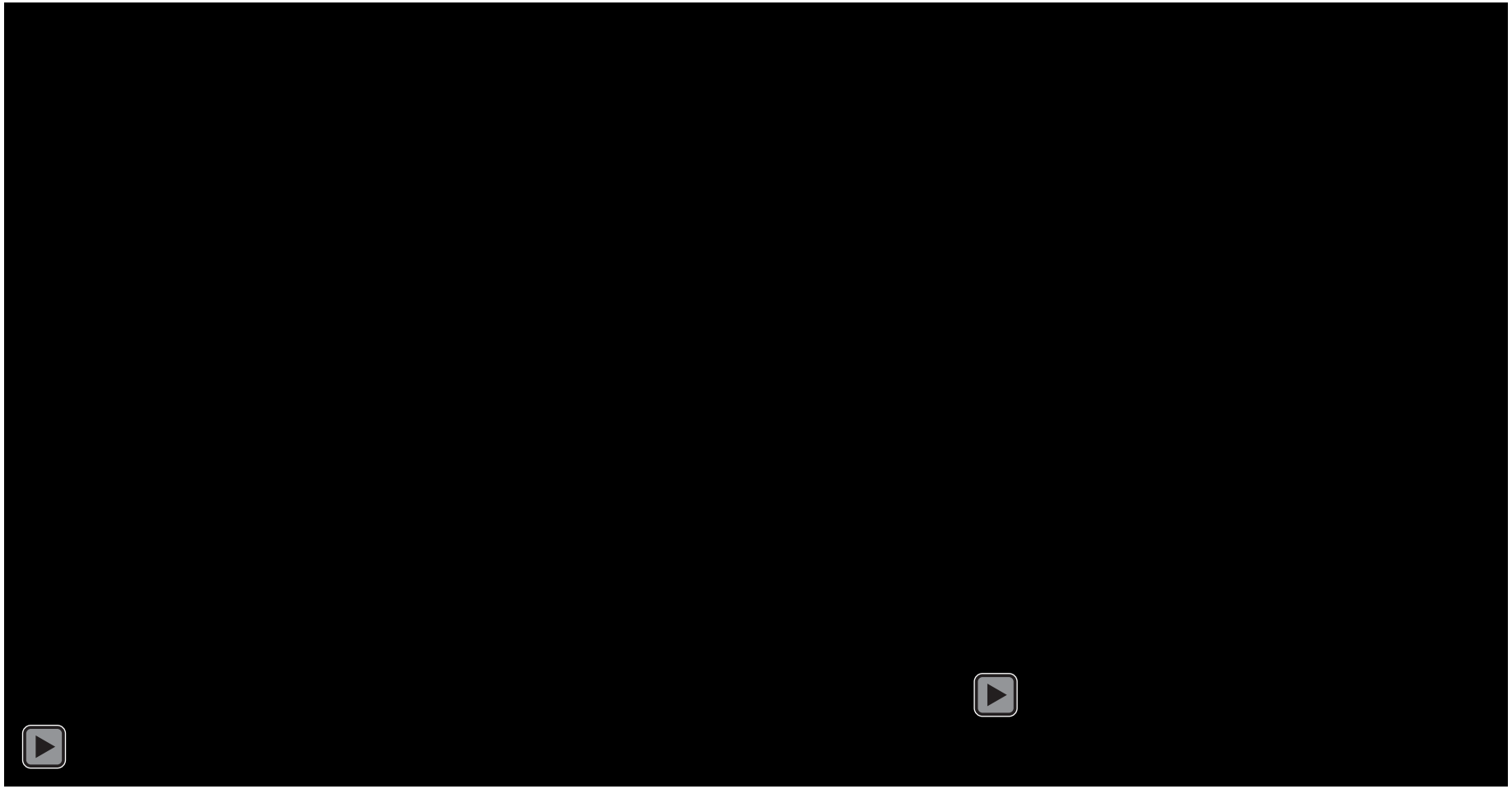
Feature embeddings

Shape correspondences

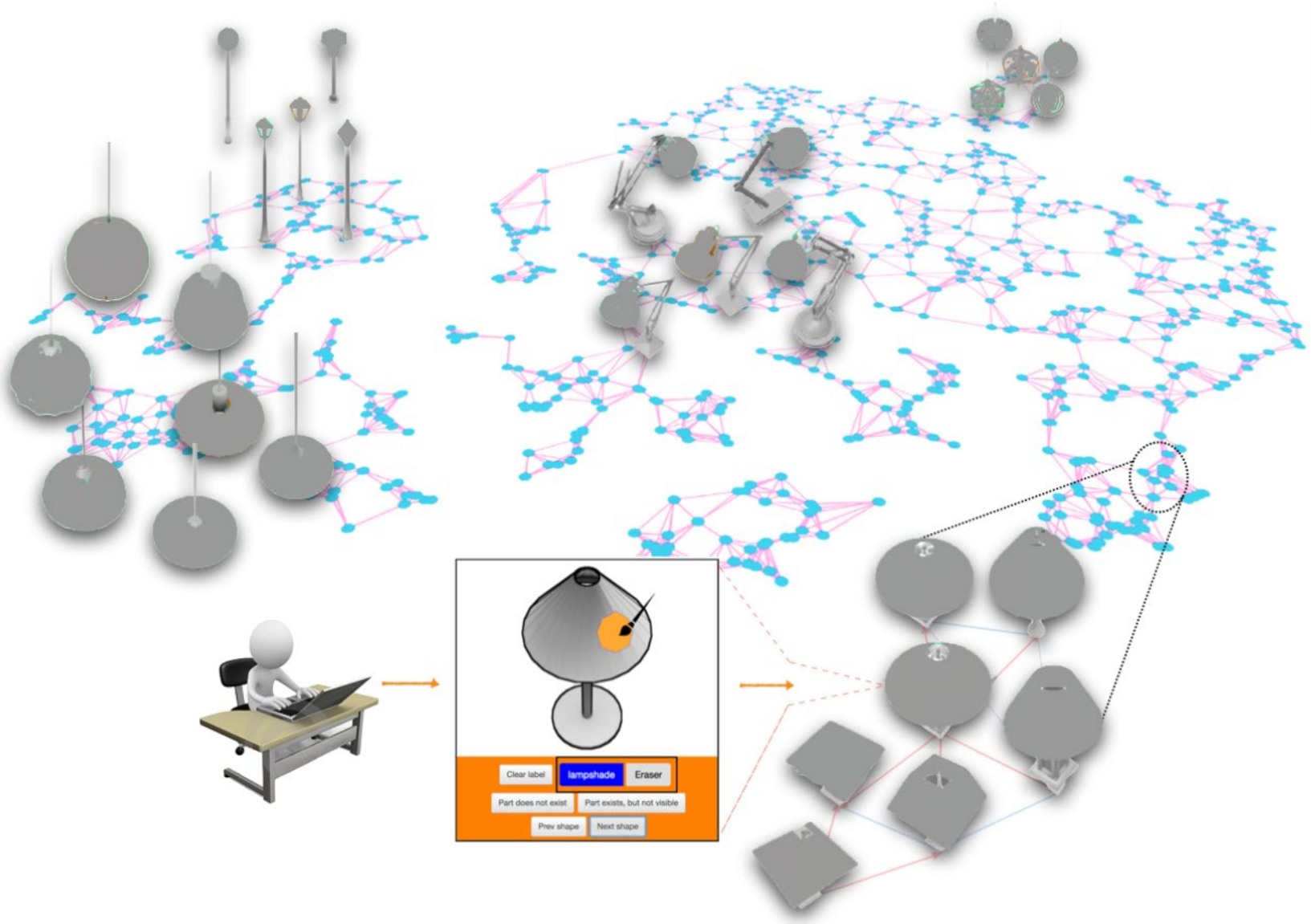


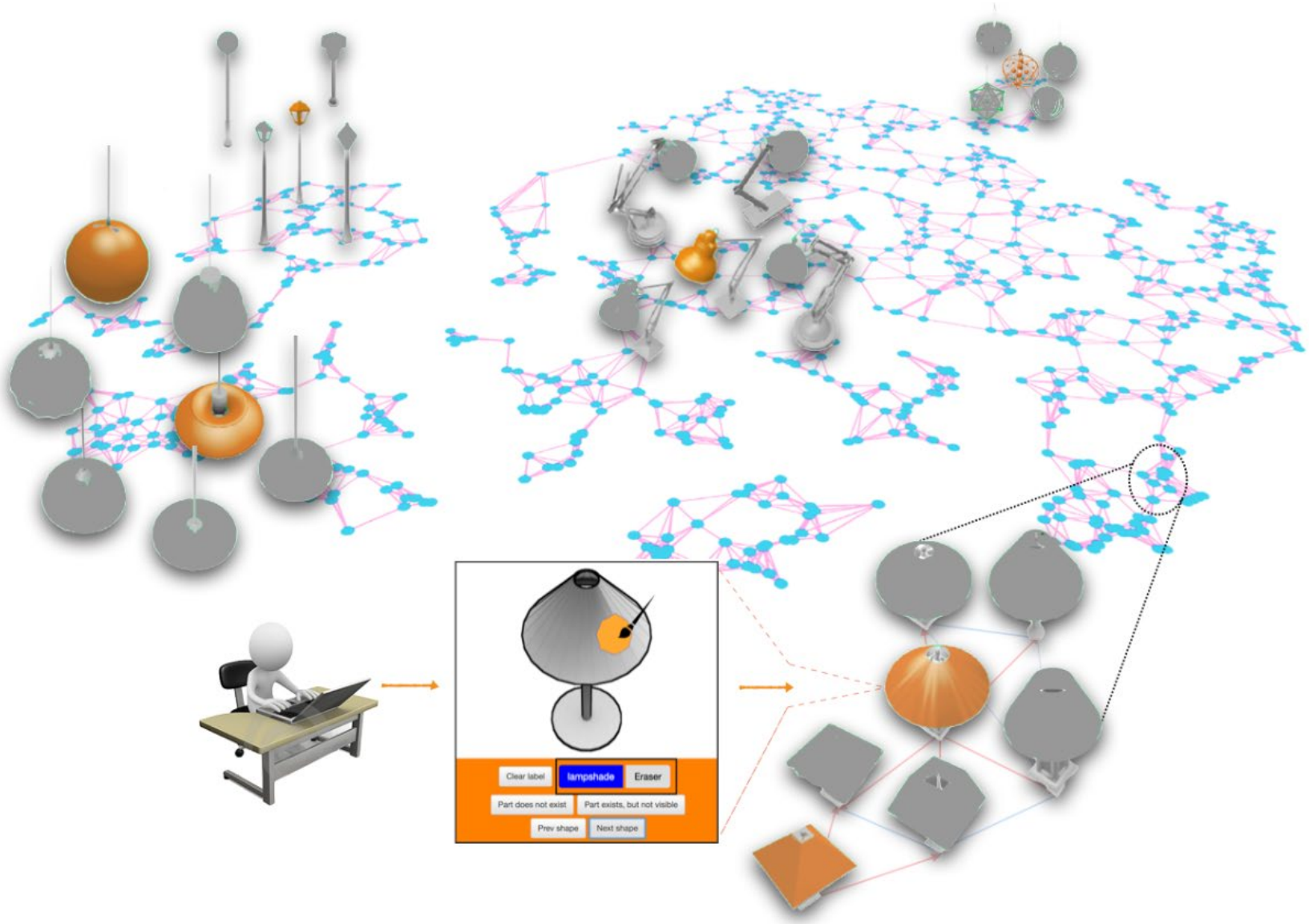
Shape alignments

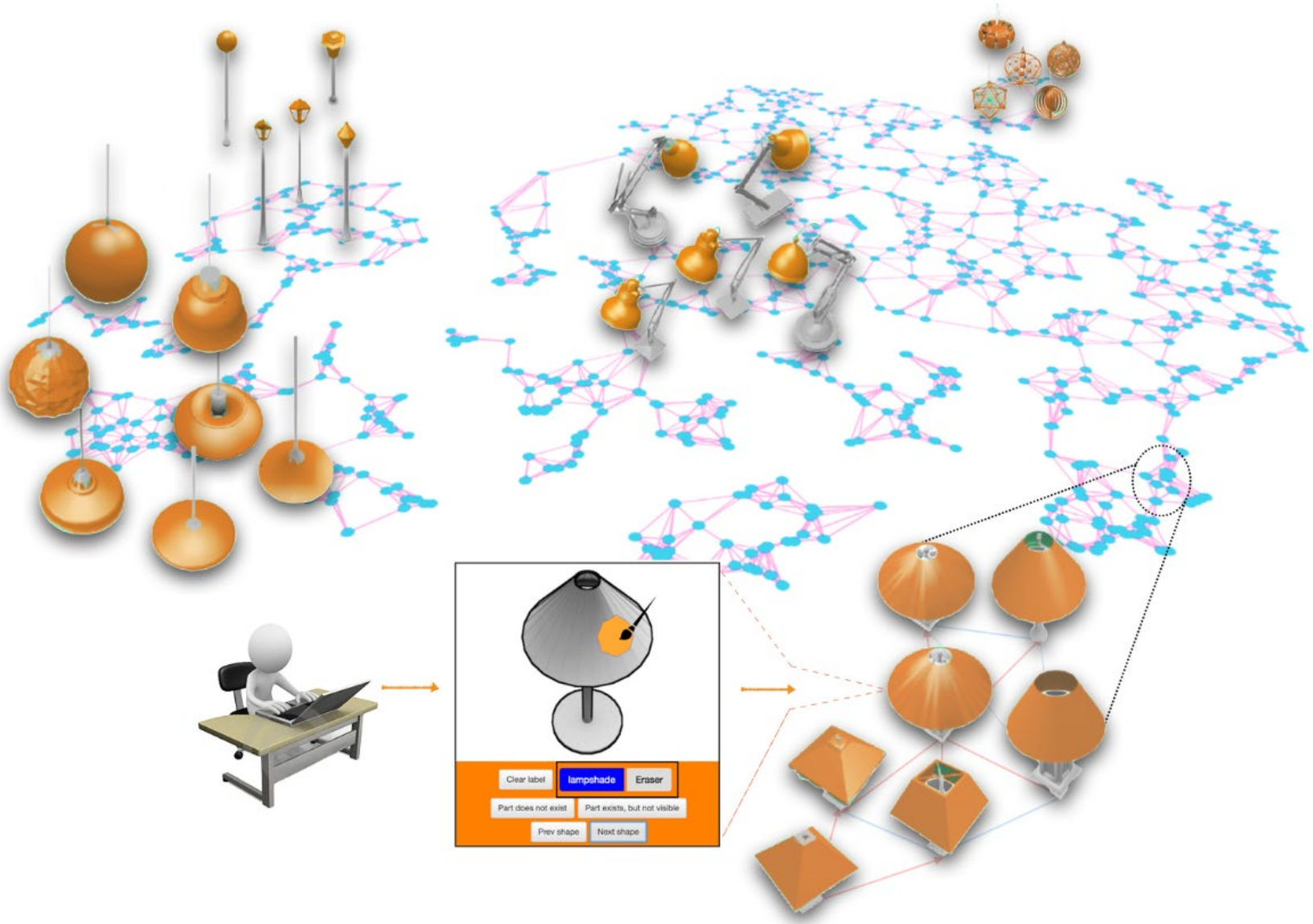
Verification More Efficient Than Annotation

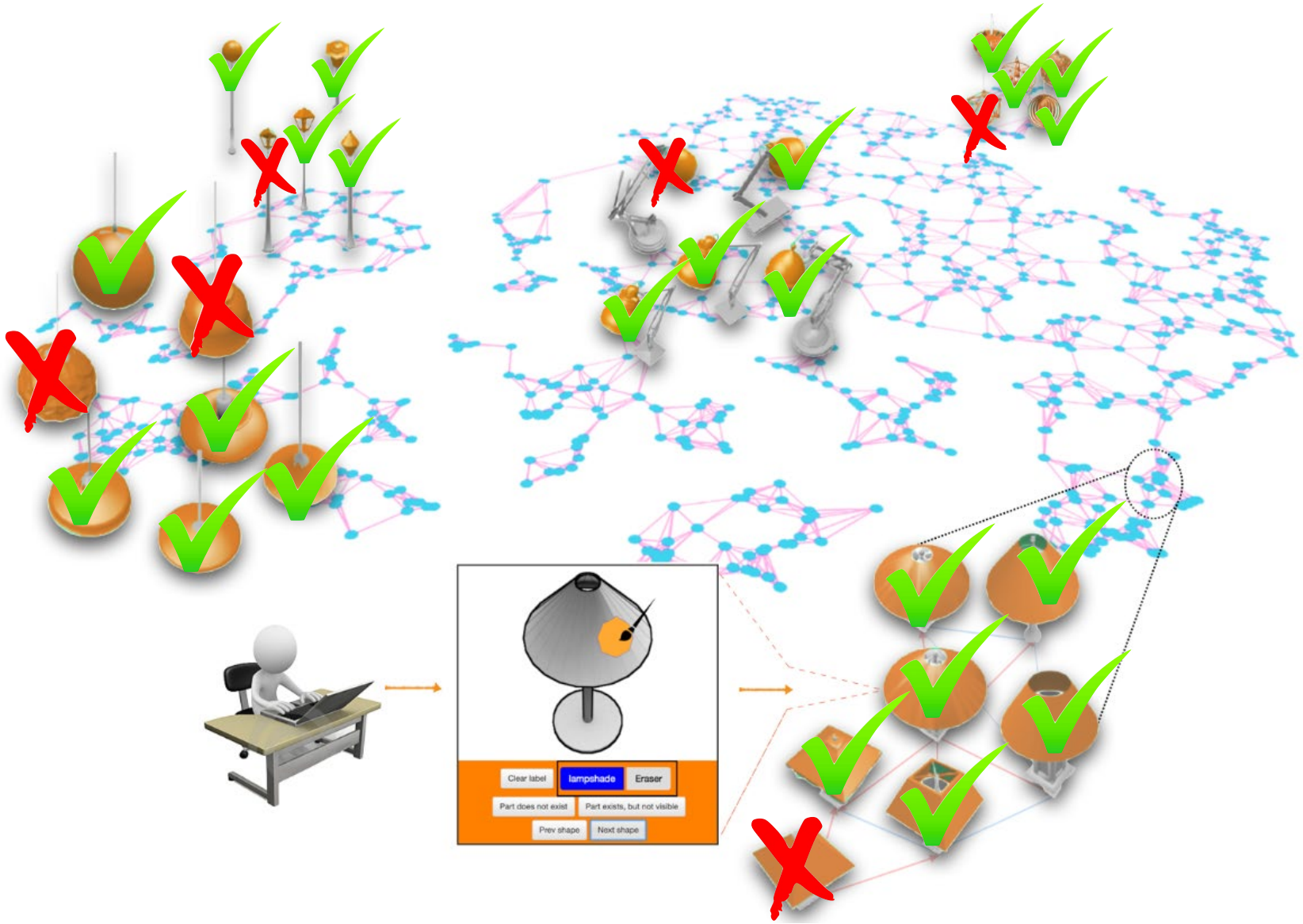


[30,000 shapes in 16 categories, 90,000 parts]

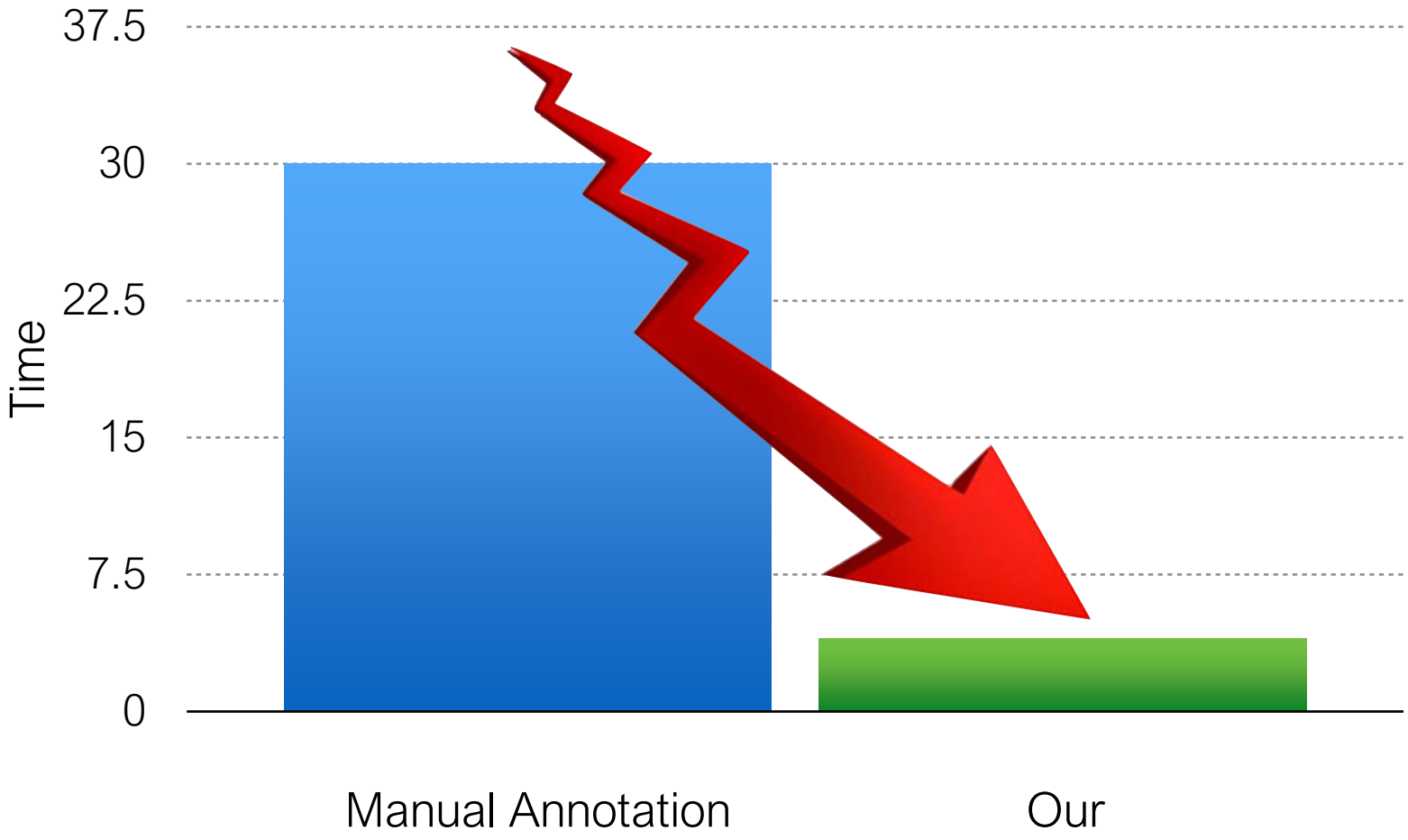

Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape


Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape

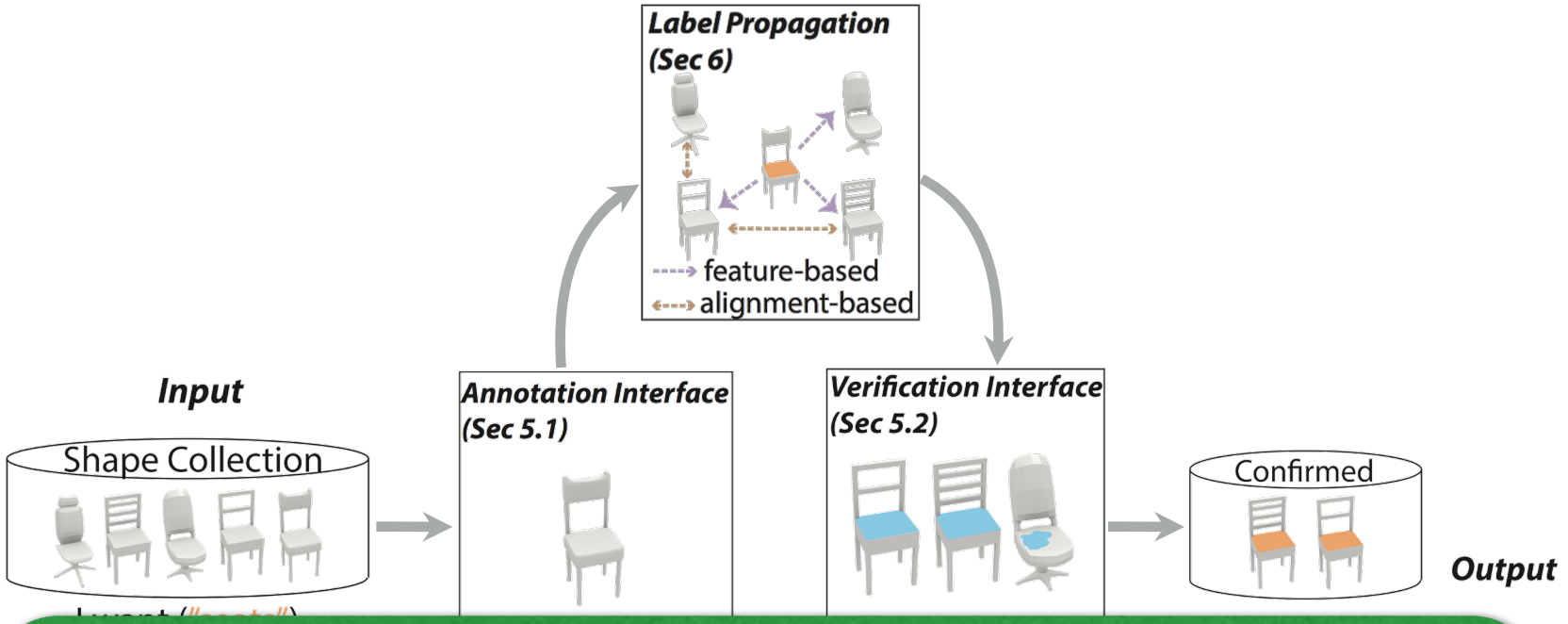

Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape


Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape

Reduced Annotation Cost



Model Part Annotation

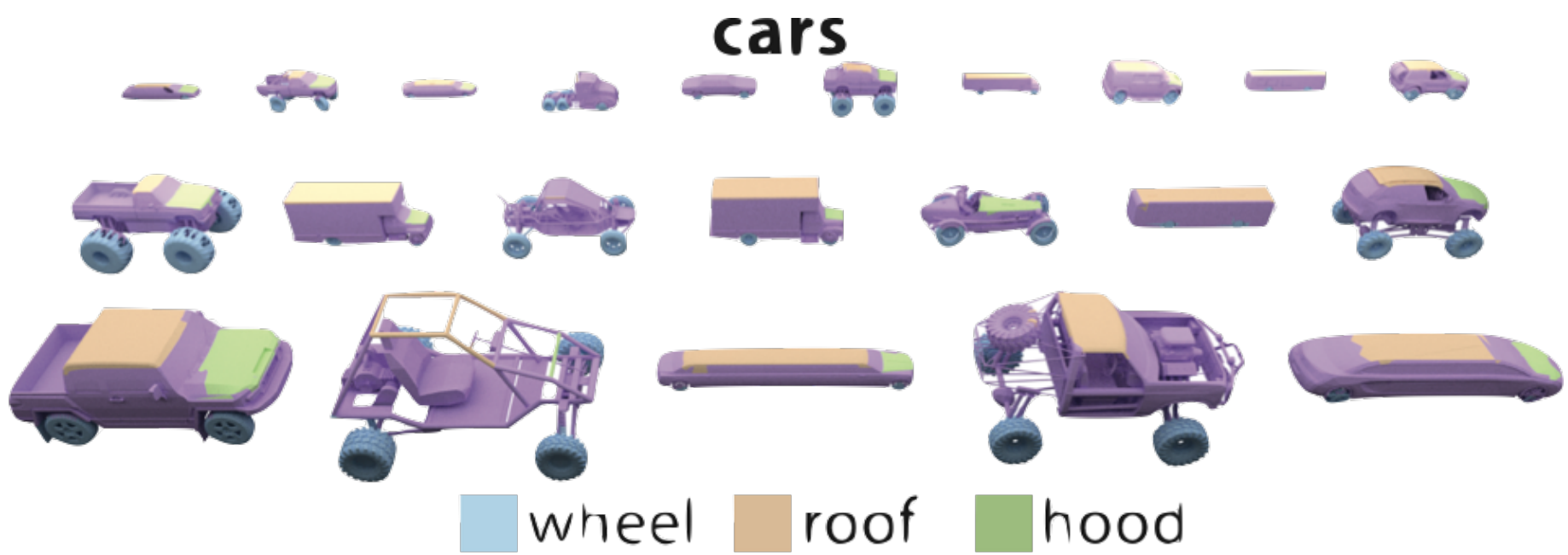


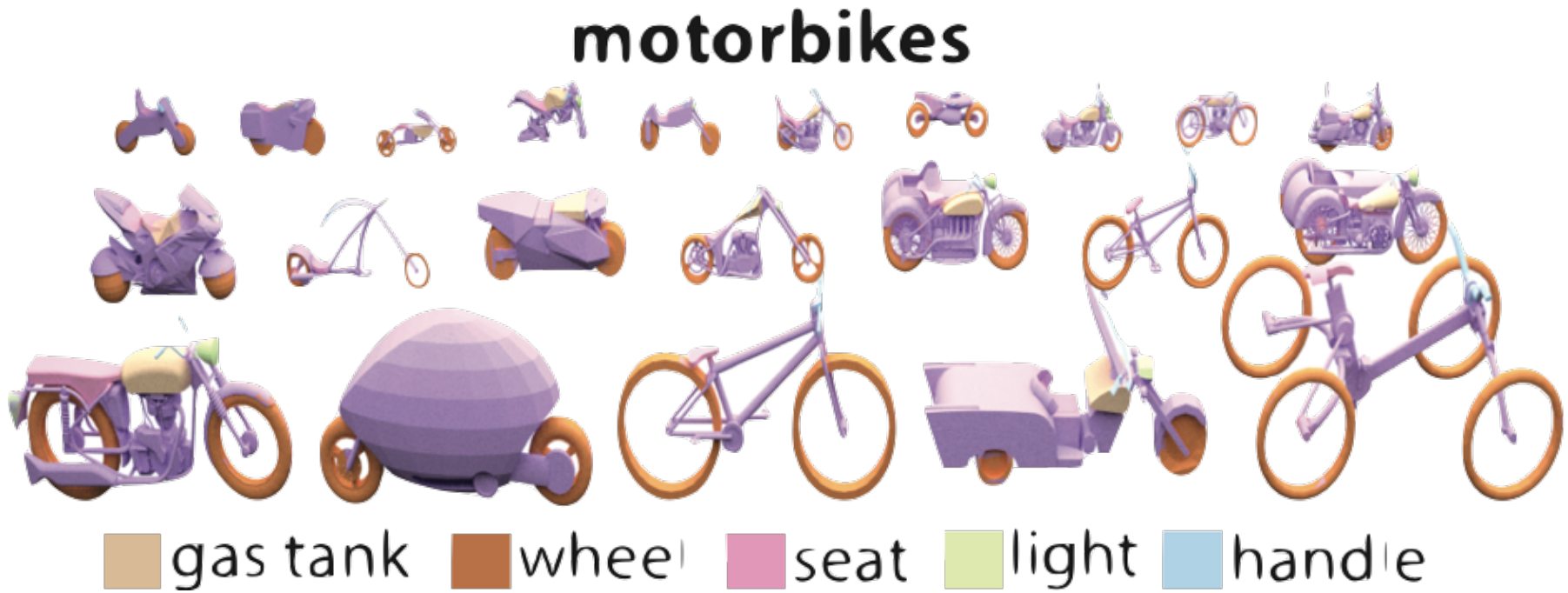
How to issue ***Annotation*** and ***Verification*** tasks

to optimize the ***Utility Function***

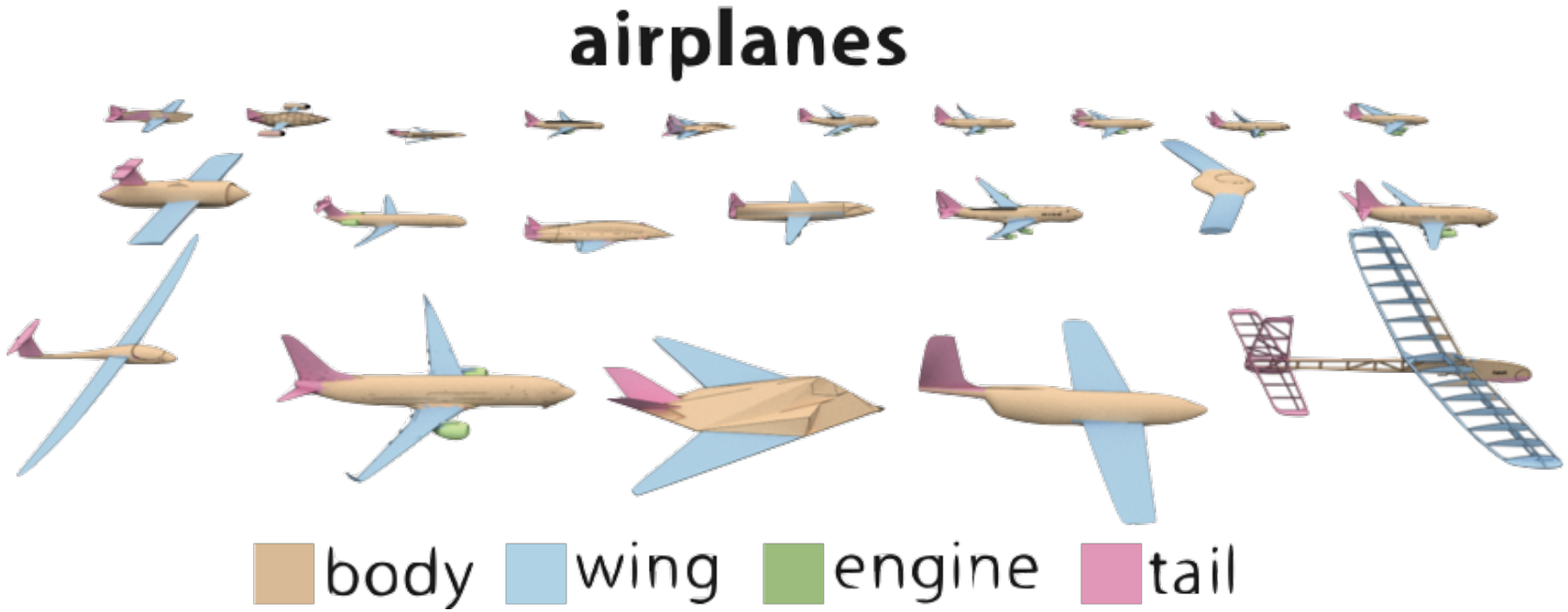
$$\frac{\#(\textit{ACCURATE}\text{ labeling})}{\text{worker time}}$$

[L. Yi, V. Kim, D. Ceylan, I. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, LG, Sigasia 2016]

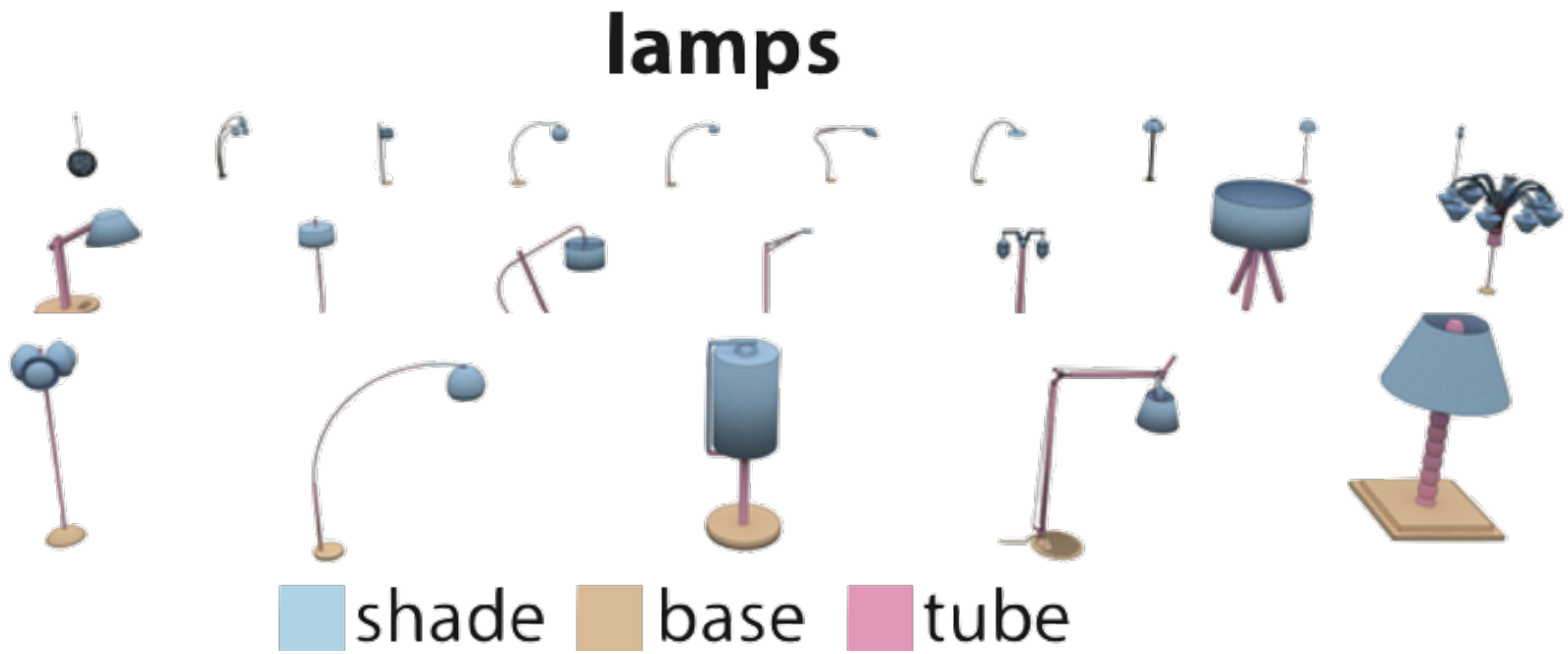

cars
wheel
roof
hood


motorbikes
gas tank
wheel
seat
light
handle

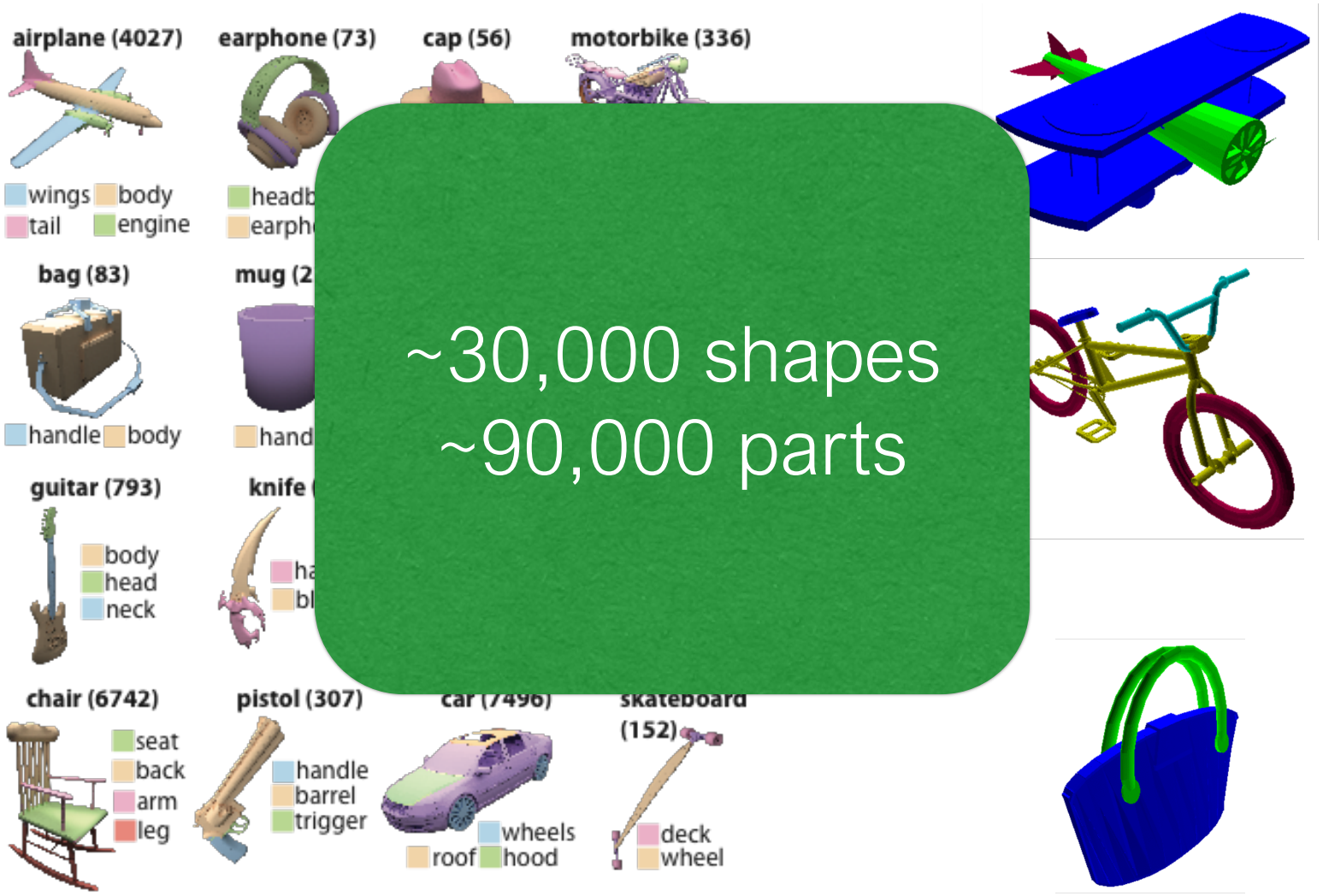

pistols
handle
barrel
trigger


airplanes
body
wing
engine
tail


bags
handle
body

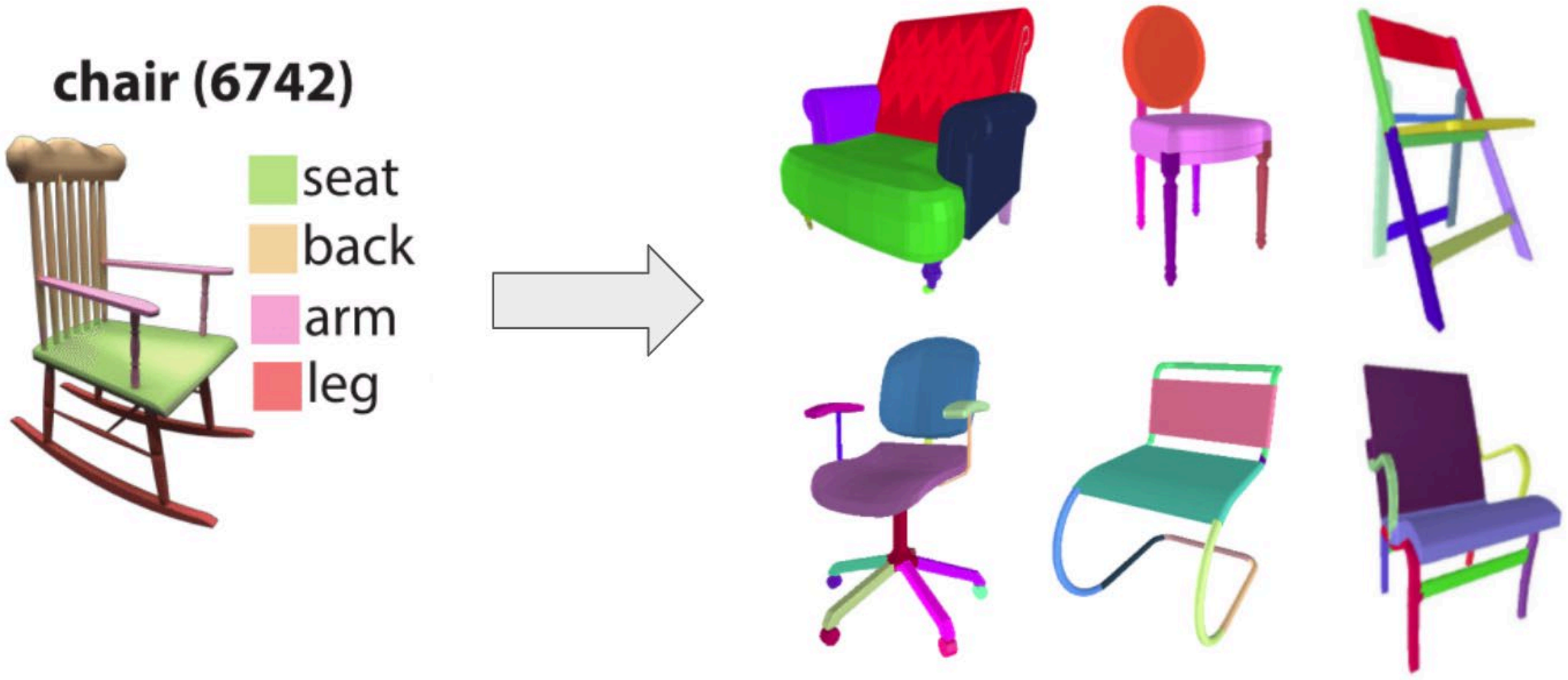

lamps
shade
base
tube

Results


airplane (4027)
wings
body
tail
engine
earphone (73)
cap (56)
motorbike (336)
bag (83)
handle
body
guitar (793)
body
head
neck
chair (6742)
seat
back
arm
leg
pistol (307)
handle
barrel
trigger
car (7496)
wheels
roof
hood
skateboard (152)
deck
wheel
~30,000 shapes
~90,000 parts

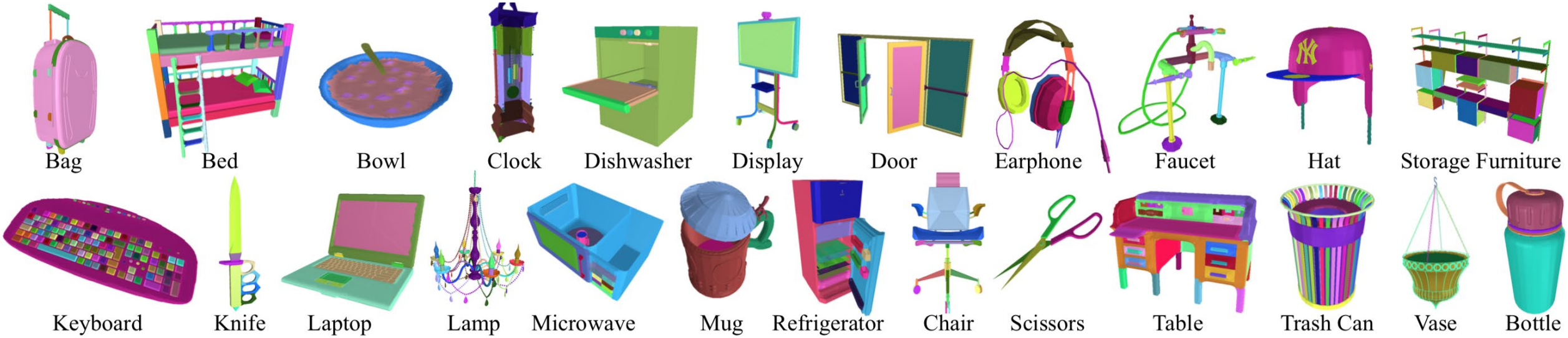
Fine-Grained Part Annotation

PartNet: Fine-Grained and Instance-Level Parts



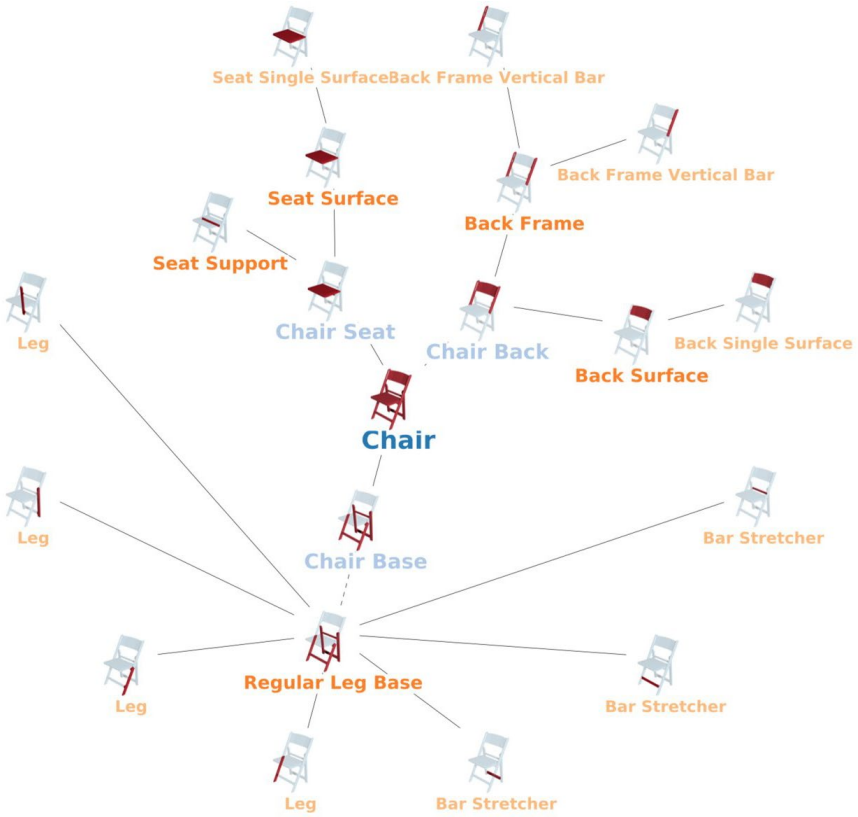
[K. Mo, S. Zhu, A. Chang, L. Yi, S. Tripathi, L. Guibas and H. Su, PartNet: A Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding, CVPR 2019]

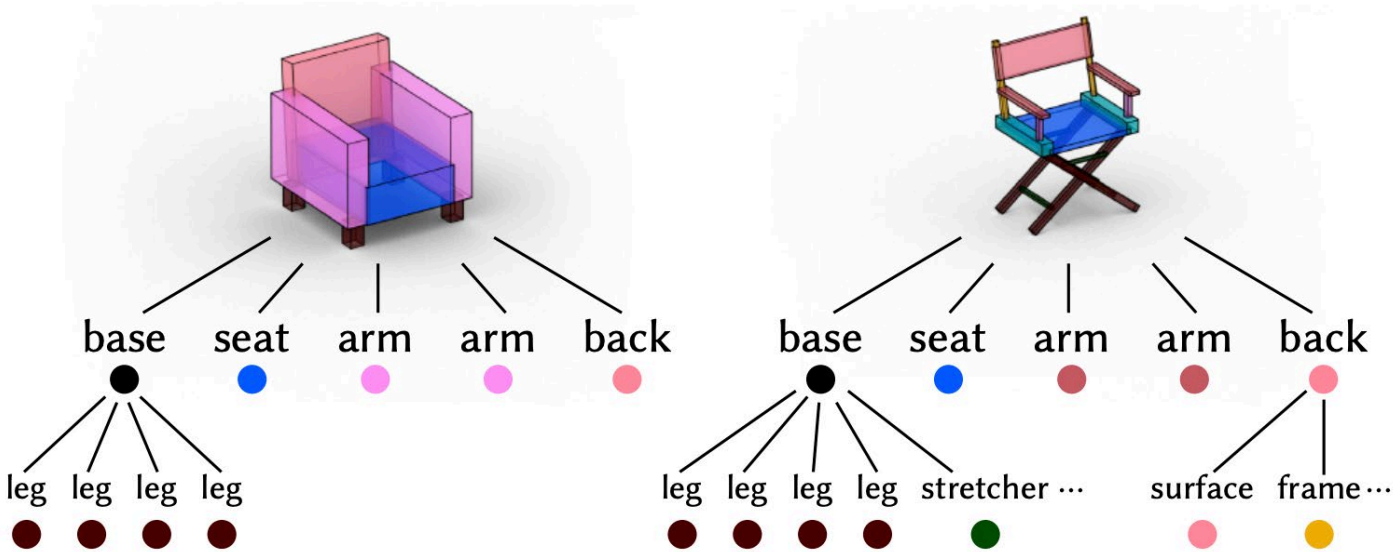
PartNet: Fine-Grained Parts



- Subset of ShapeNetCore
  - 24 common indoor categories, 26,671 shapes, 573,585 parts
  - Avg 18 Part/shape, Max 230
- Human-annotated:
  - more fine-grained parts + instance-level parts

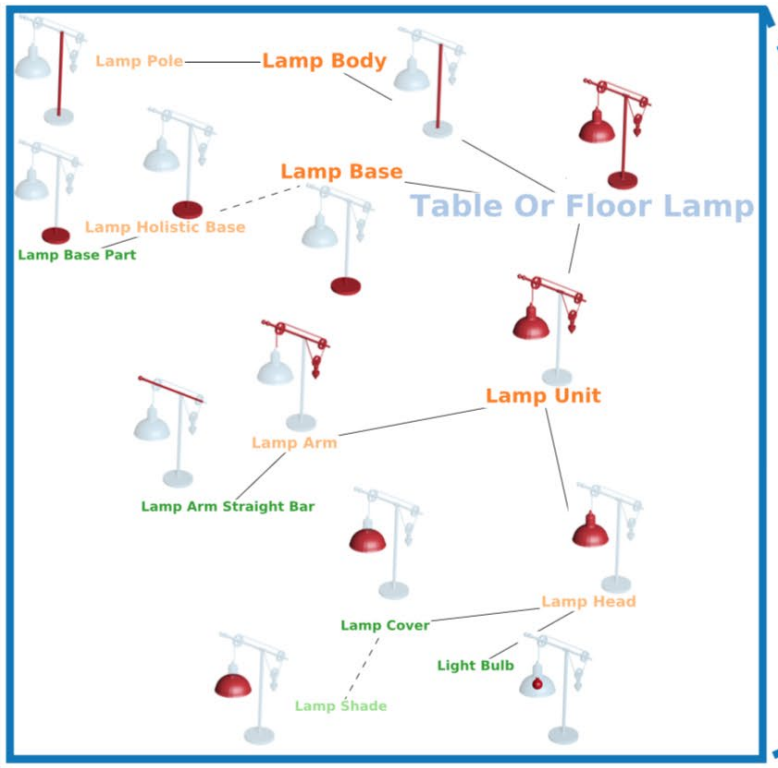
PartNet: Hierarchical and Consistent Parts

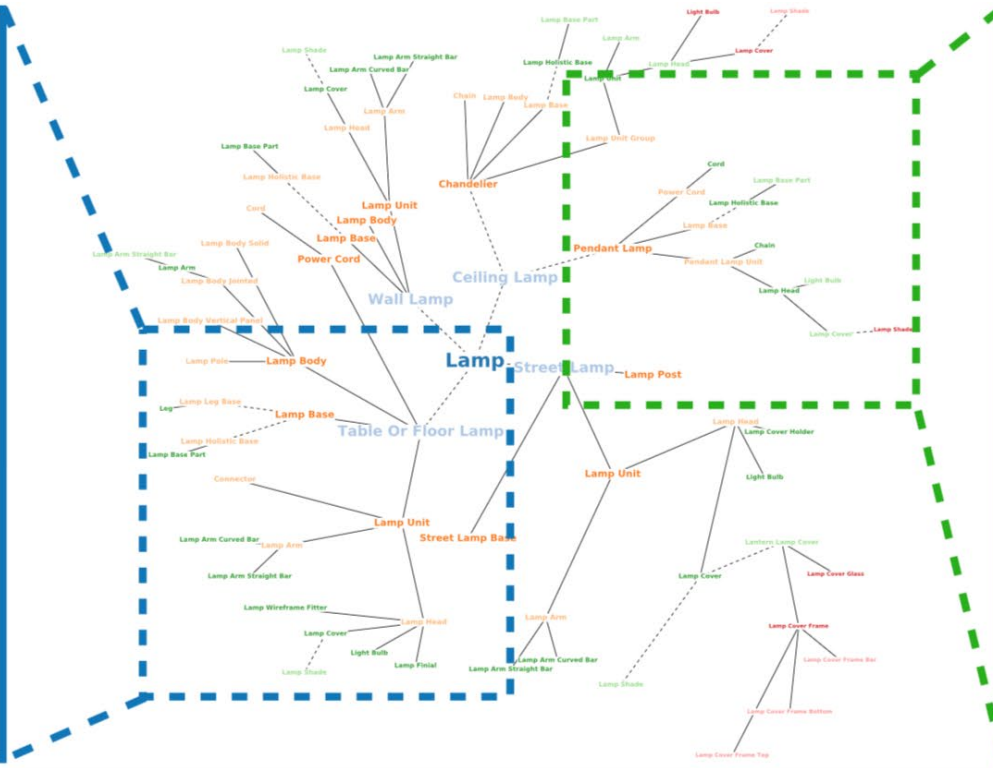


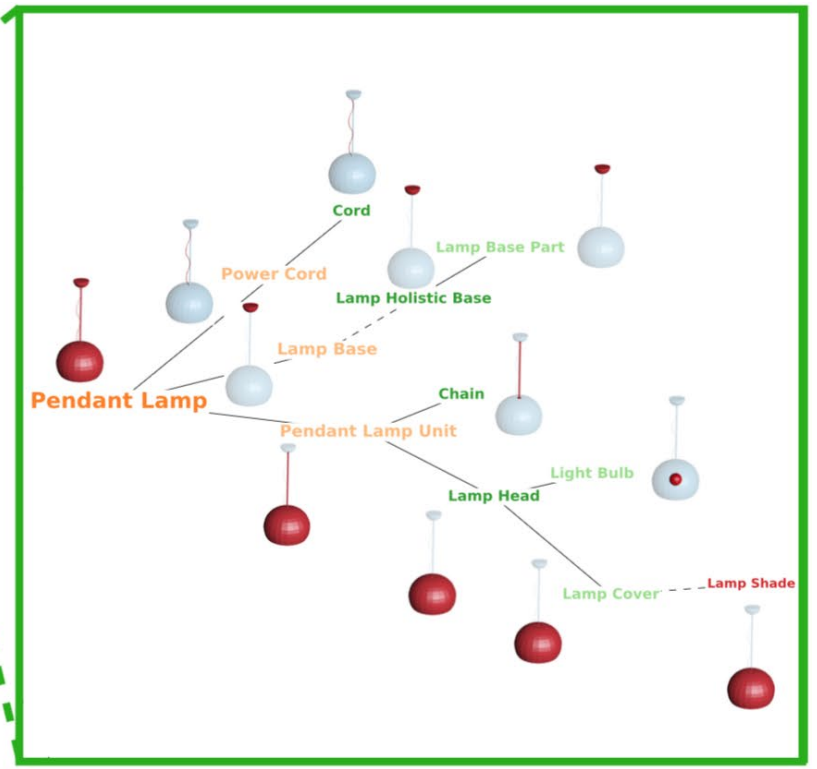


- Provide hierarchical segmentation of shapes: parts at multiple scales
- All shapes from the same category conform to a consistent part template

PartNet: And-Or-Graph Semantic Part-Template





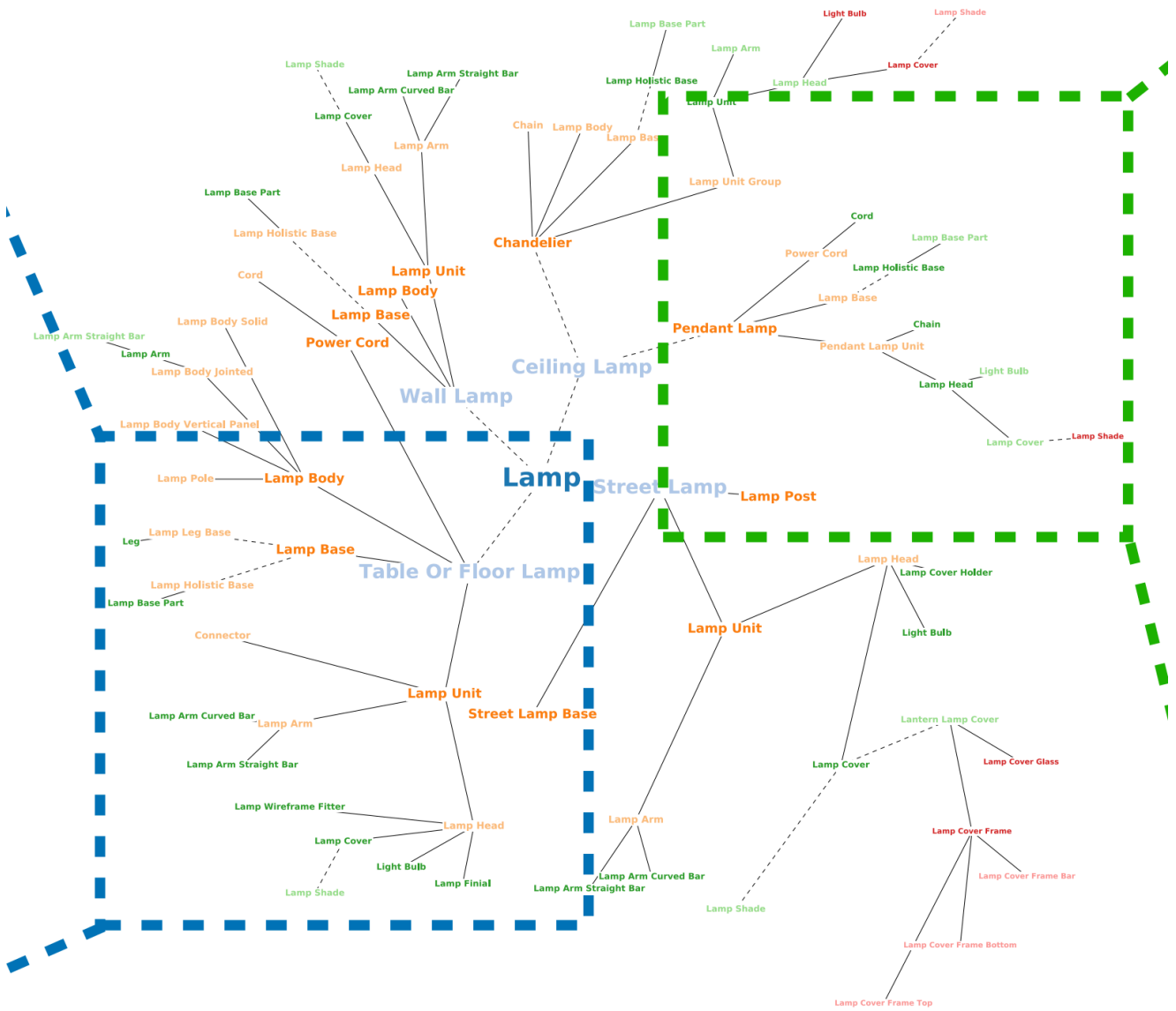


One Table Lamp Example

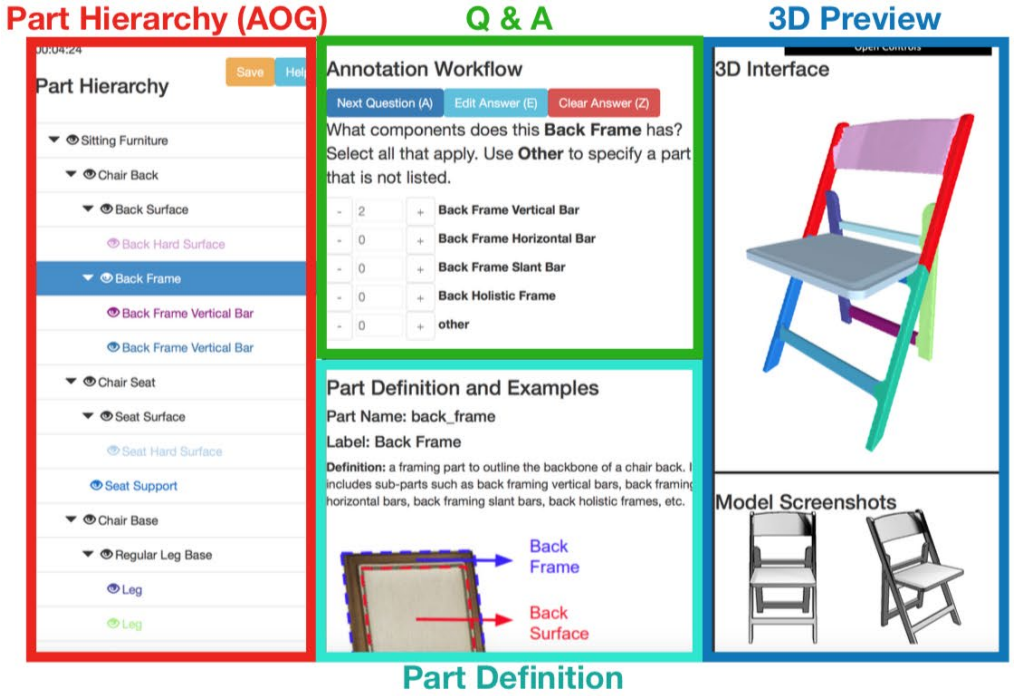
Part Template

One Floor Lamp Example

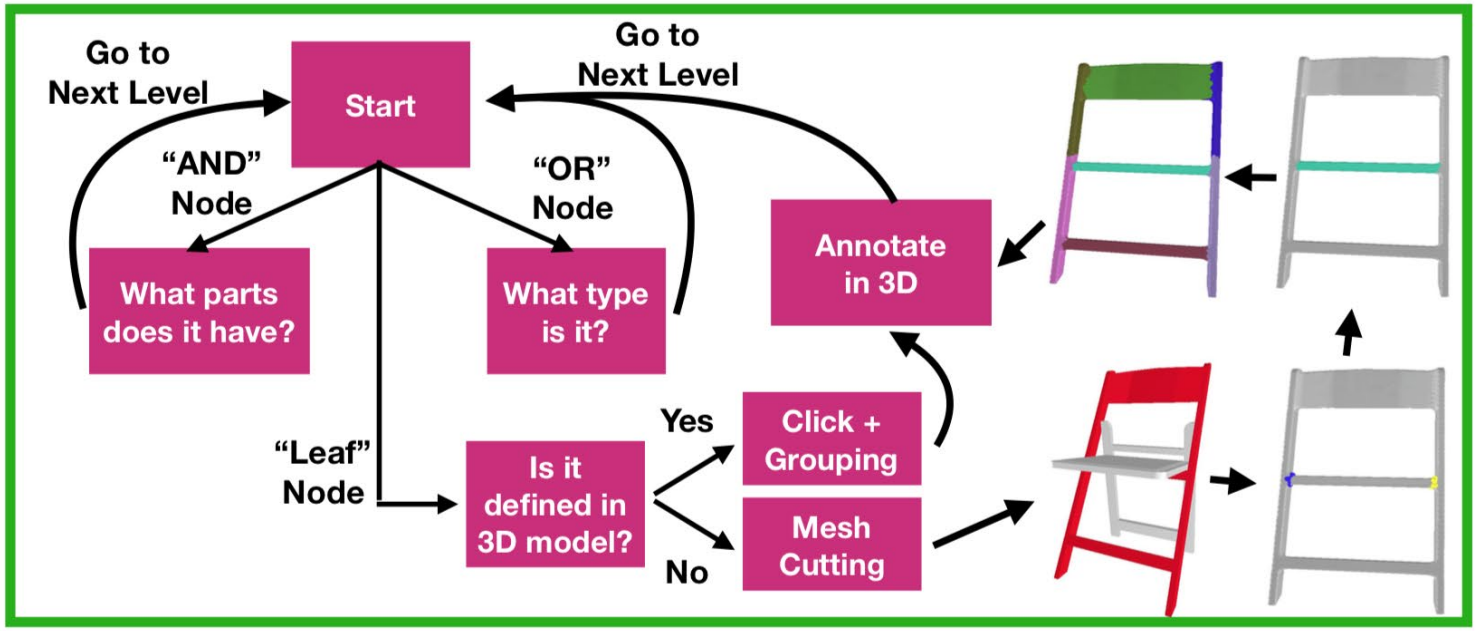
PartNet: And-Or-Graph Semantic Part-Template



PartNet: Annotation Interface



(a) PartNet Main Web Interface

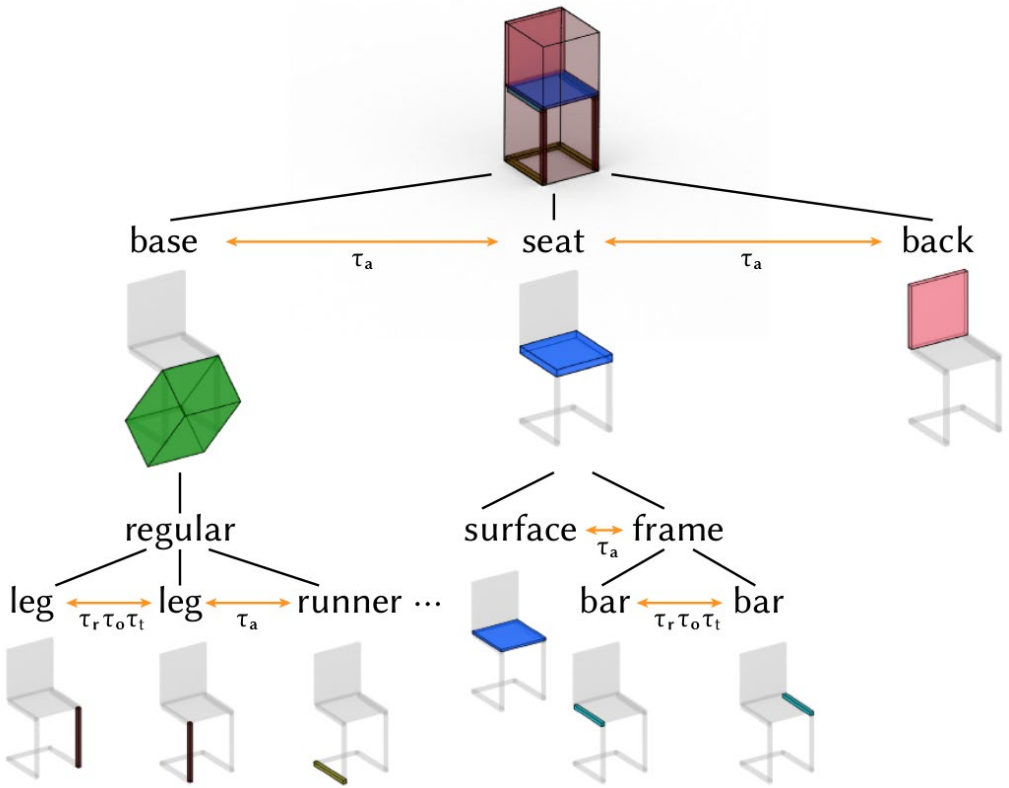


(b) Human Annotation Workflow

(c) Mesh Cutting Interface

- Encourage different users annotate consistently according to the template
- Allow certain freedom to give "other" parts that are not pre-defined

PartNet: Hierarchical Part Graphs



- Part-based structure-aware shape representation
- With intra-part relationships: vertical and horizontal

PartNet-Mobility and SAPIEN: Part Articulation

- Synthetic, 3D Shapes
- Based upon ShapeNet/PartNet
- 14,068 Articulated Parts
- 2,346 Objects, 46 Categories

- Part Articulation
- Physical Simulation

- Support the study of various robotic manipulation tasks

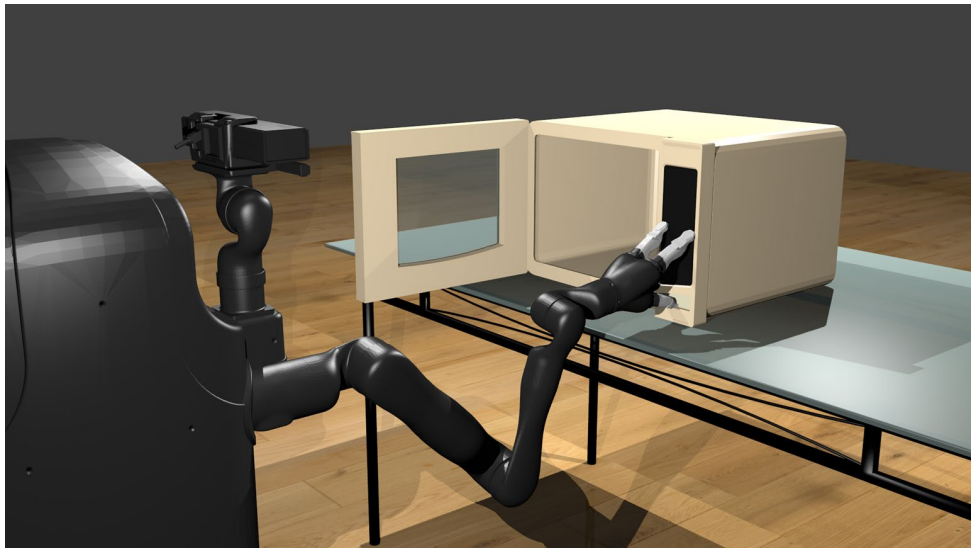
https://sapien.ucsd.edu/

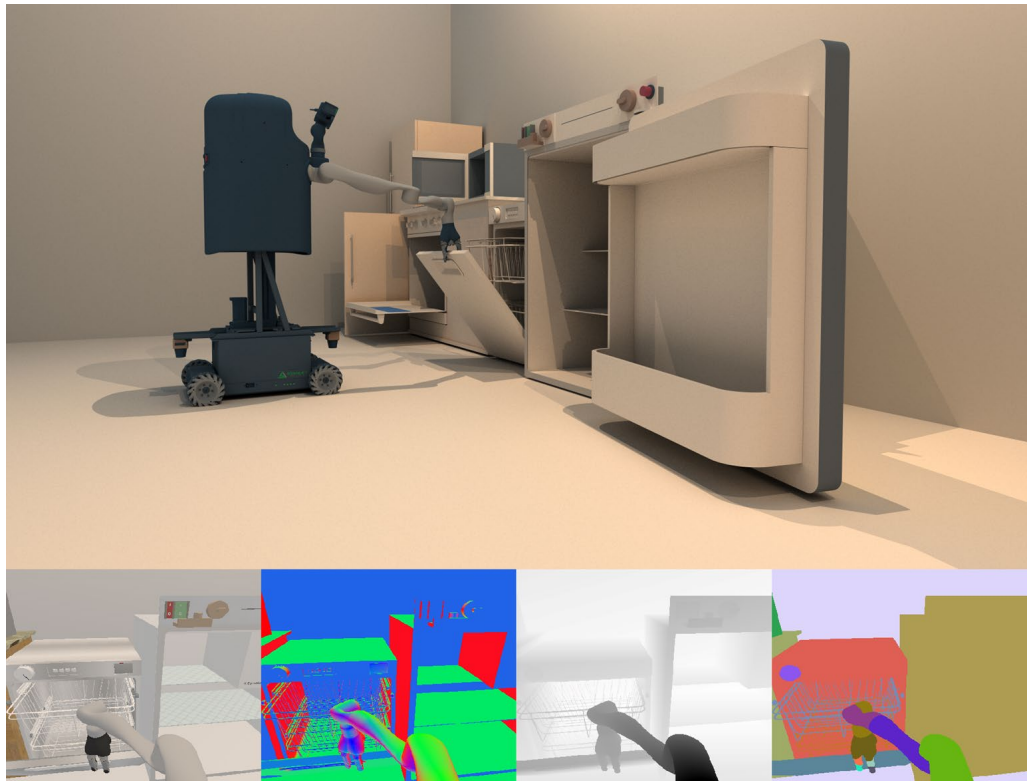
Xiang et al., "**SAPIEN: A SimulAted Part-based Interactive ENvironment**", CVPR 2020

Geometric + Image Data through Simulation

Combing Geometric and Image Data

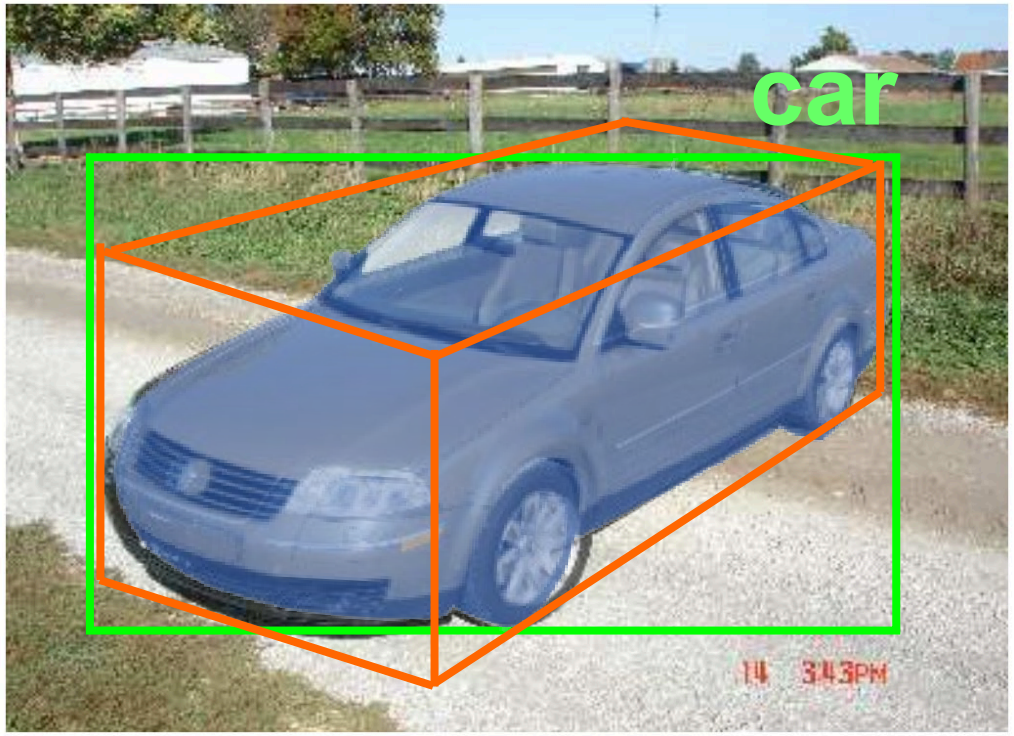
Many real-world tasks (*e.g.* robotics) require image data with 3D annotations.





[F. Xiang, Y. Qin, K. Mo, Y. Xia, H. Zhu, F. Liu, M. Liu, H. Jiang, Y. Yuan, H. Wang, L. Yi, L. Guibas and H. Su, SAPIEN: A SimulAted Part-based Interactive ENvironment, CVPR 2020]

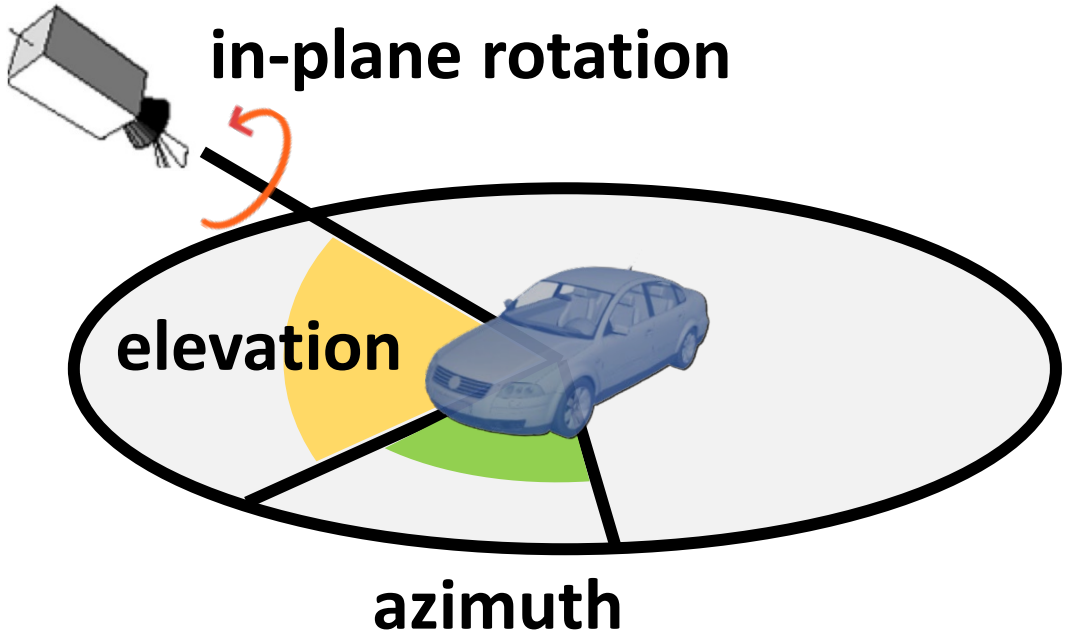
3D Annotations are Difficult



[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

Accurate Label Acquisition is Expensive

Task:
3D Viewpoint Estimation

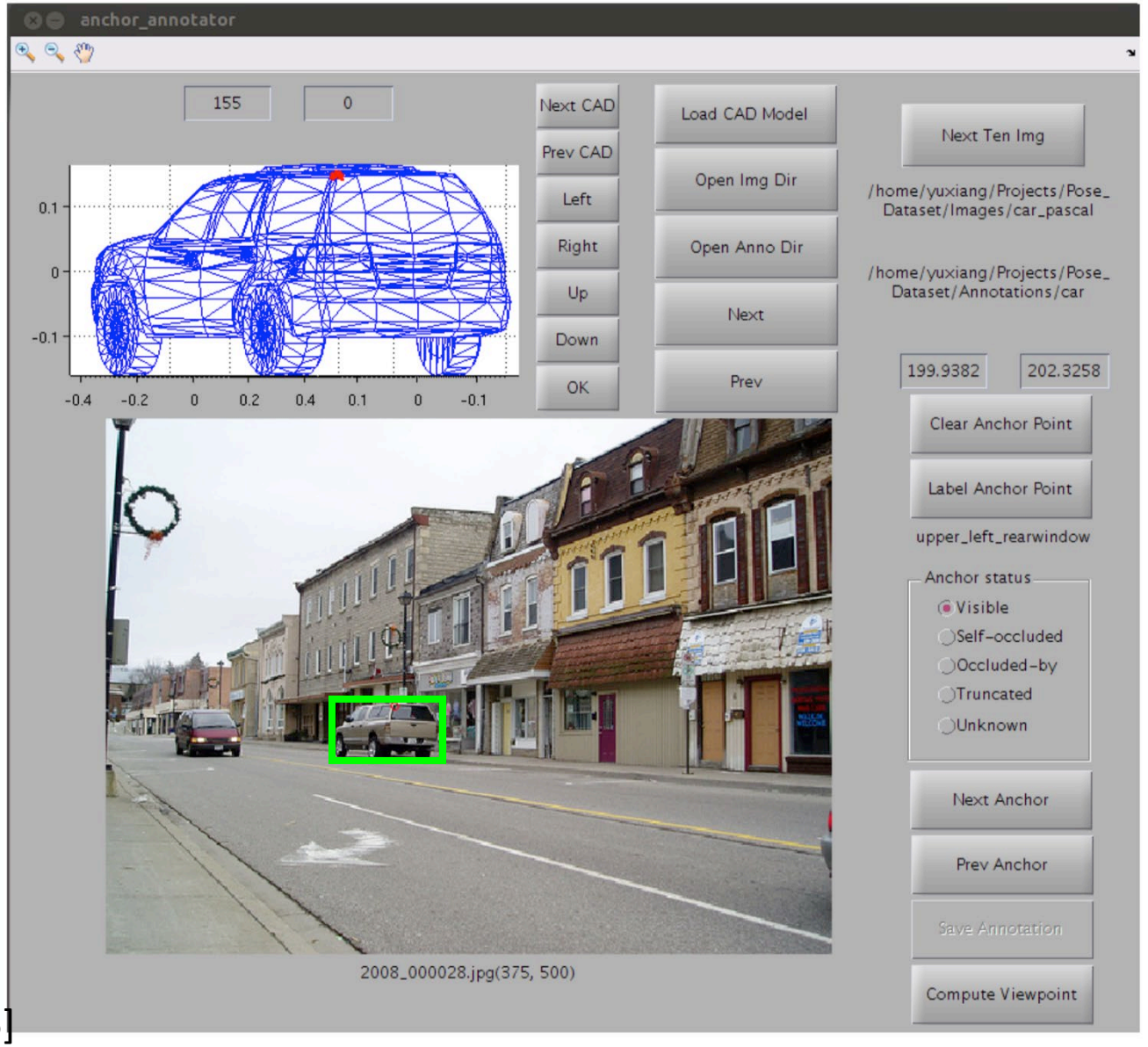


What's the camera viewpoint angles to the SUV in the image?

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

Accurate Label Acquisition is Expensive

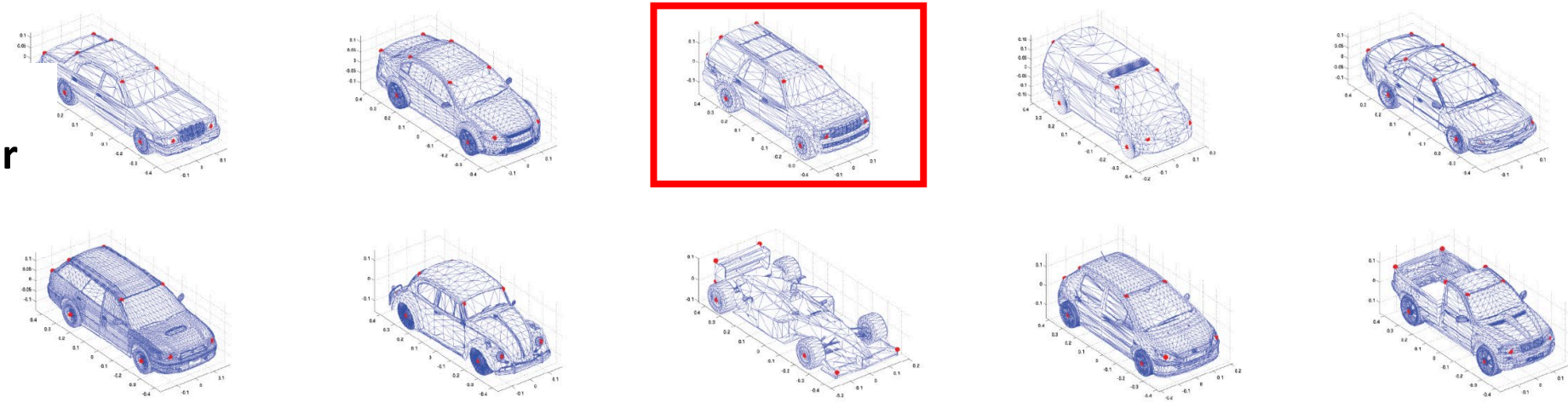
PASCAL3D+ dataset [Xiang et al.]

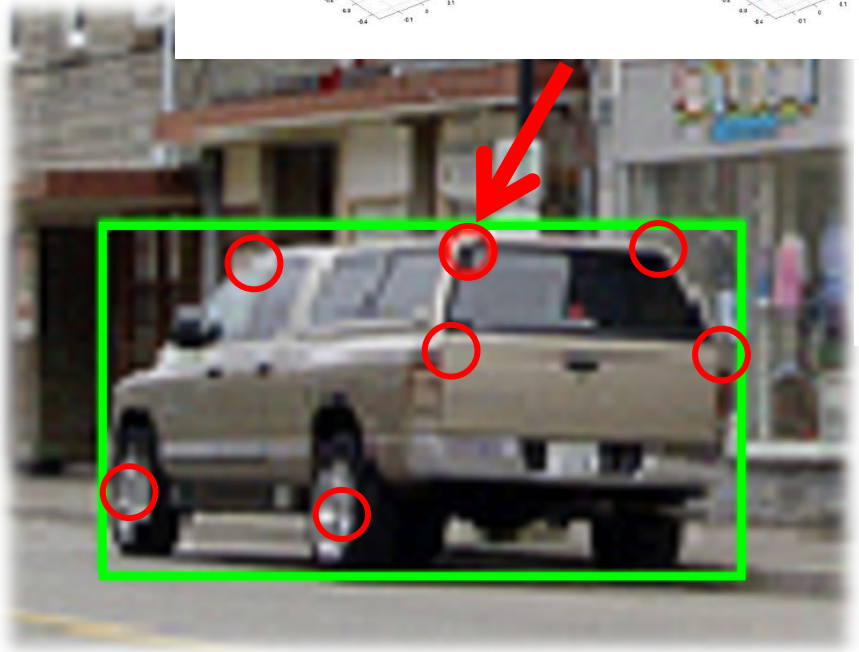


[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

Accurate Label Acquisition is Expensive

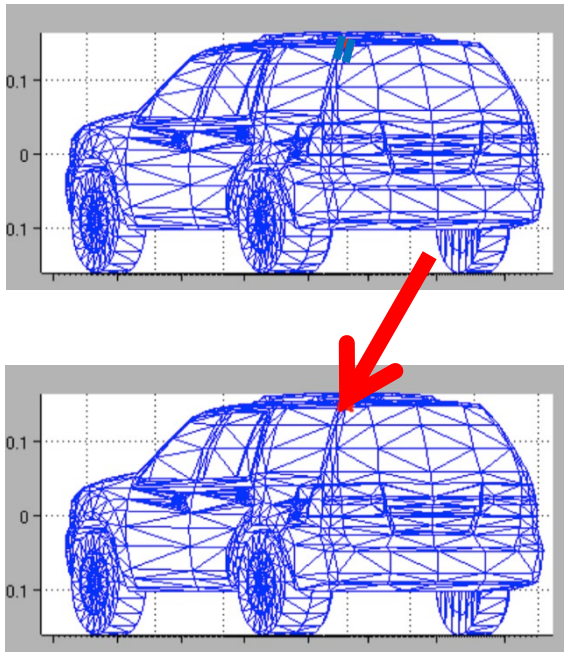
Step1:
Choose similar model





Step2:
Coarse Viewpoint Labeling

Step3:
Label keypoints For alignment

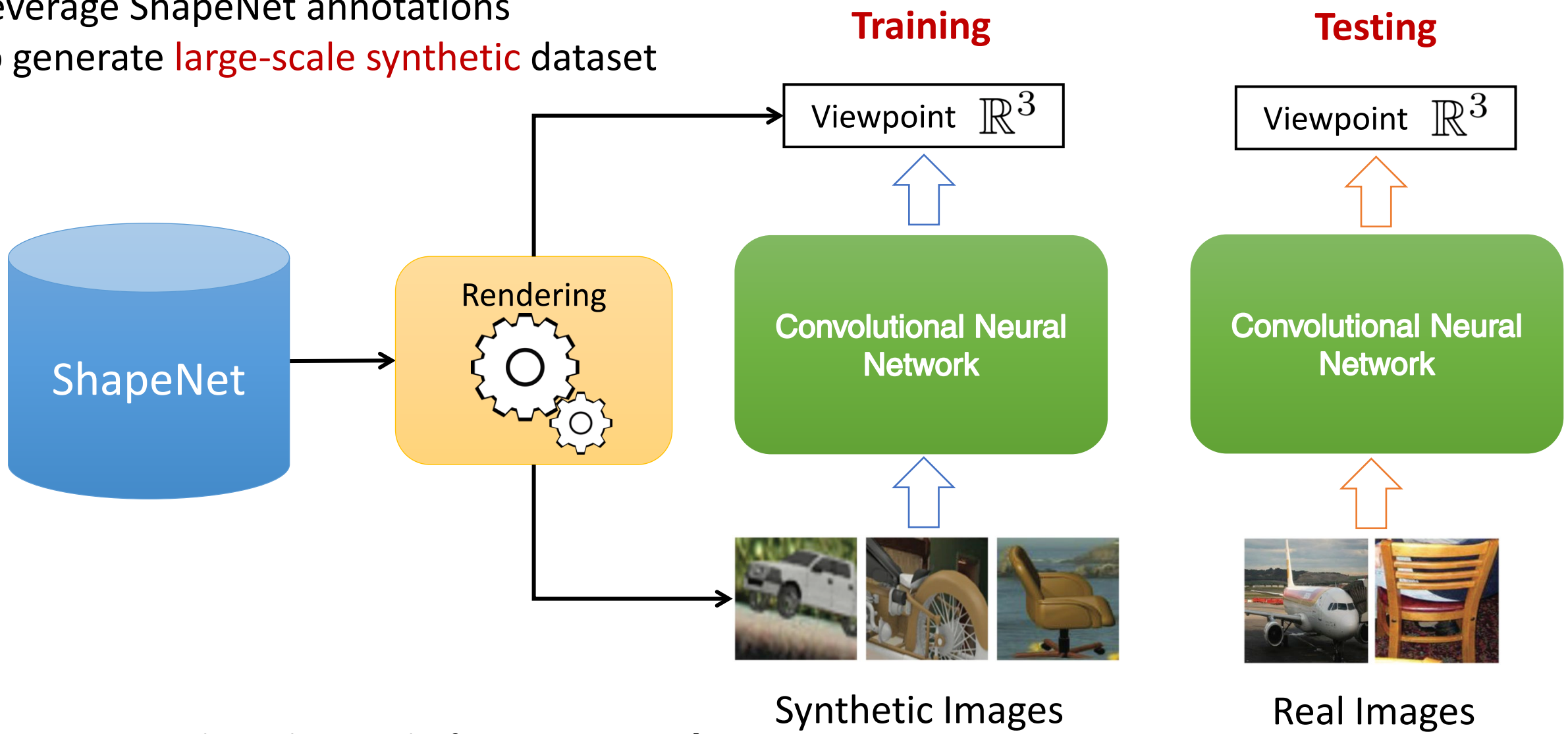


Annotation takes ~1 min per object

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

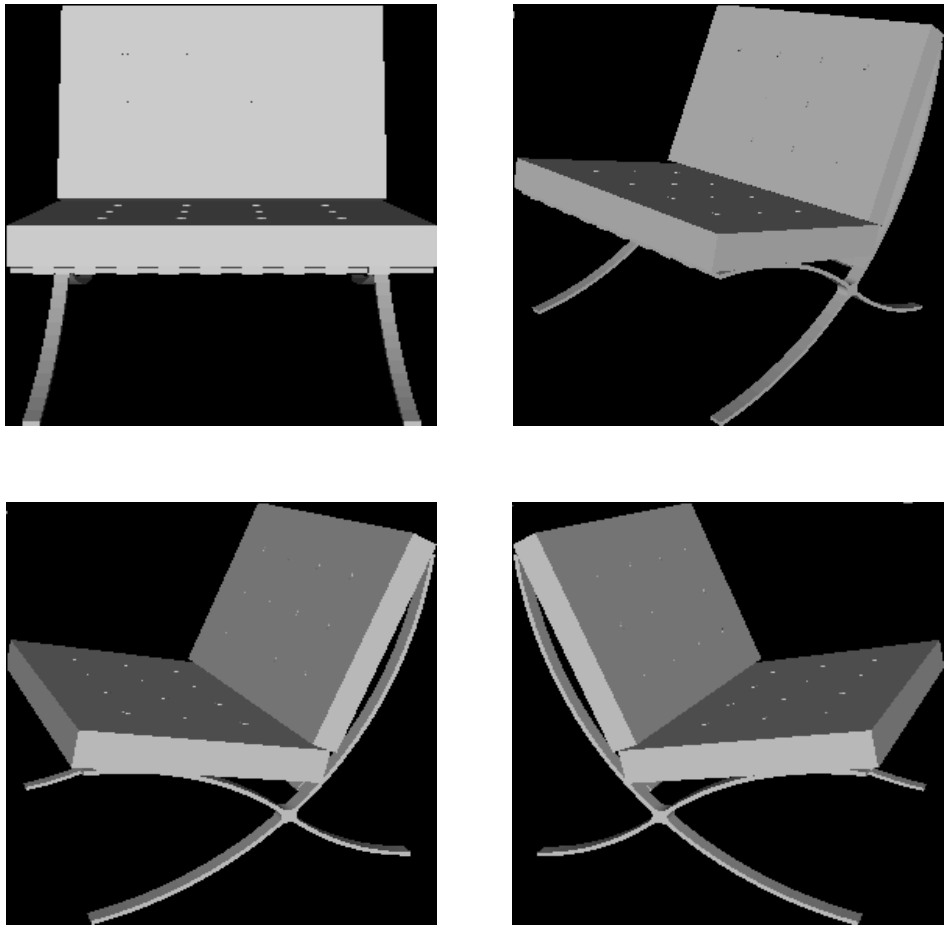
Key Idea: Render for CNN

Leverage ShapeNet annotations
to generate large-scale synthetic dataset



[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

A “Data Engineering” Journey



95% on synthetic val set

47% on real test set ☹

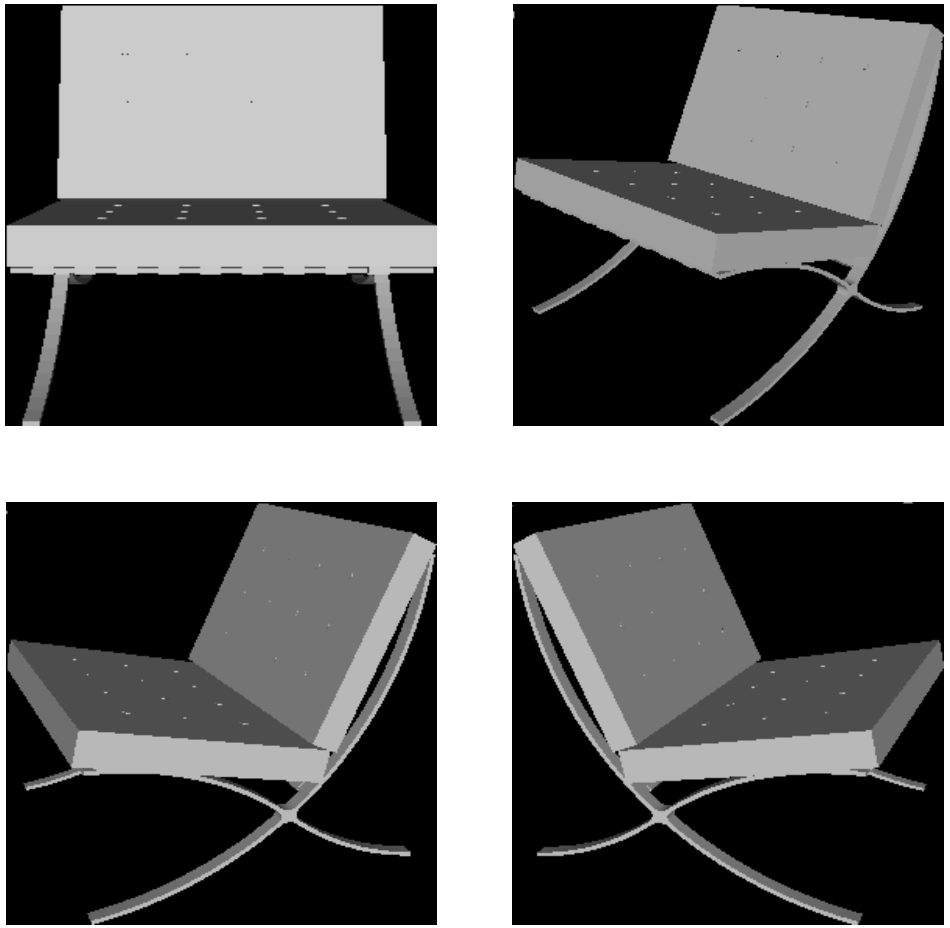
ConvNet:

Ah ha, I know!

Viewpoint is just the brightness pattern!

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

A “Data Engineering” Journey

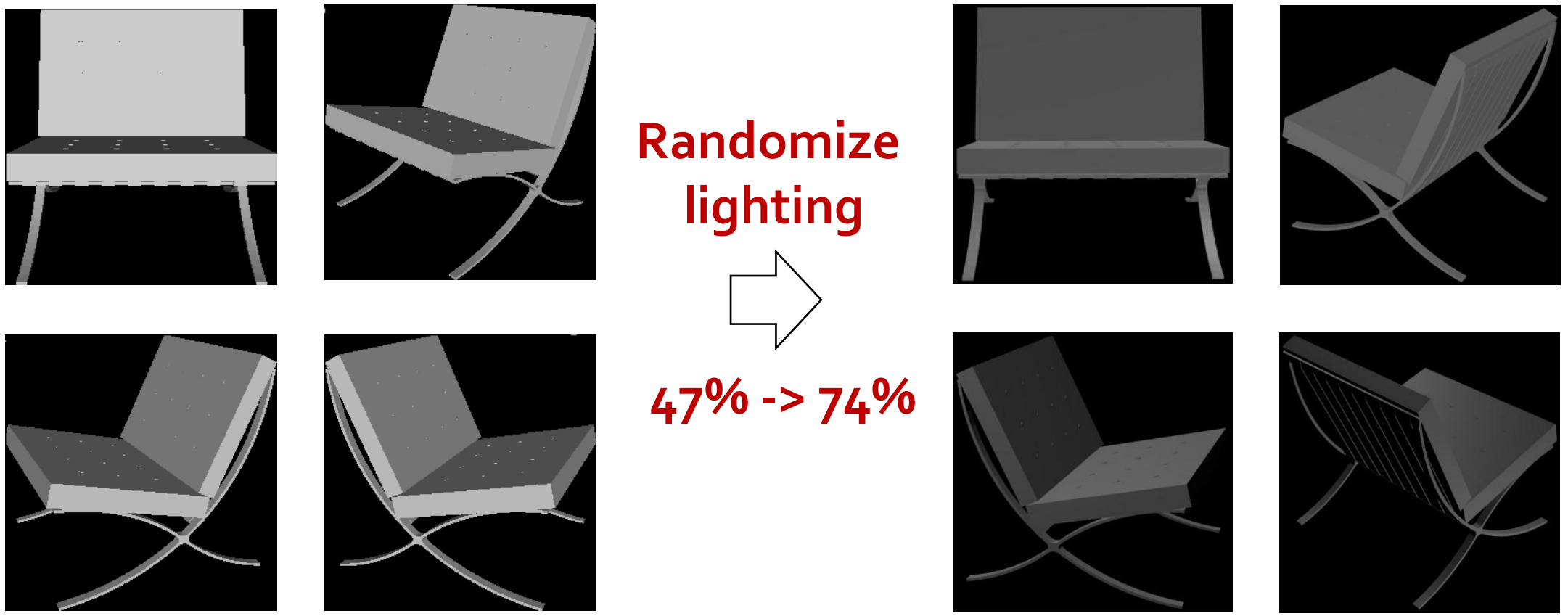


95% on synthetic val set

47% on real test set ☹

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

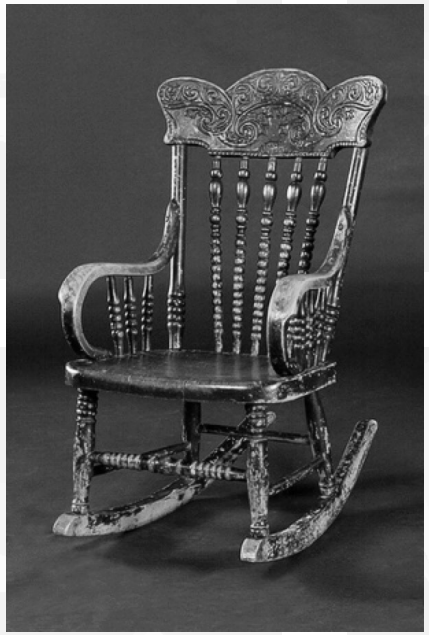
A “Data Engineering” Journey

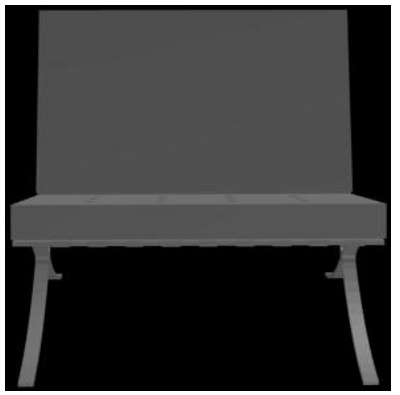


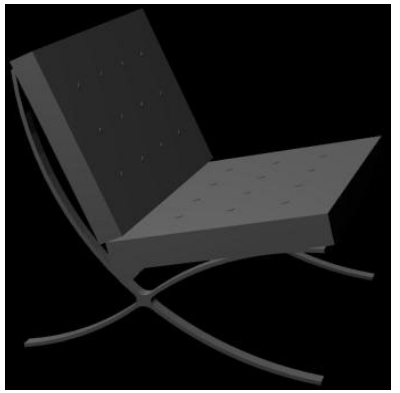
ConvNet: hmm.. viewpoint is not the brightness pattern. Maybe it’s the contour?

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

A “Data Engineering” Journey



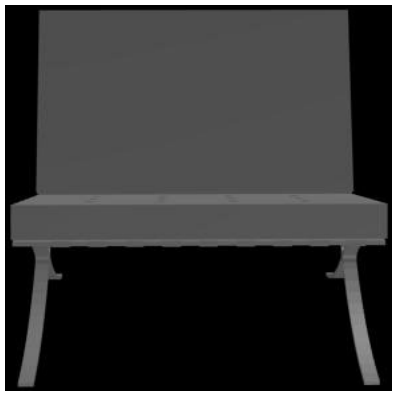


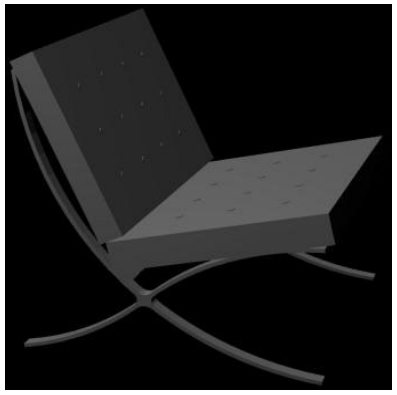


ConvNet: hmm.. viewpoint is not the brightness pattern. Maybe it’s the contour?

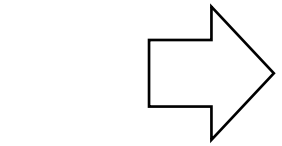
[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

A “Data Engineering” Journey

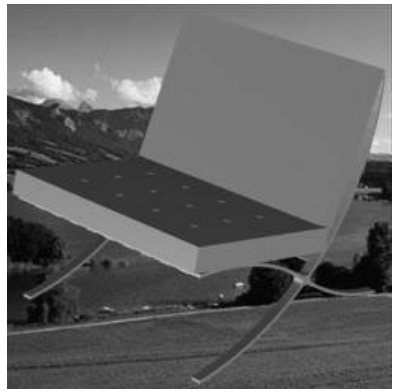
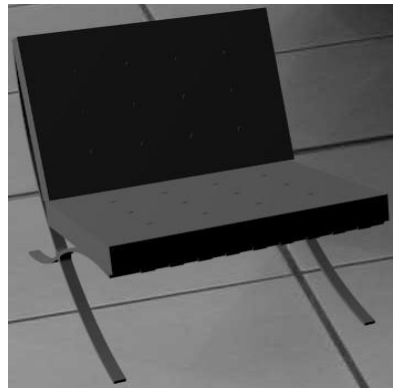





Add backgrounds



74% -> 86%

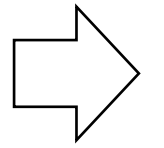



ConvNet: It becomes really hard! Let me look more into the picture.

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

A “Data Engineering” Journey

bbox crop
texture

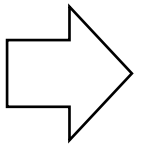


86% -> 93%

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

A “Data Engineering” Journey

bbox crop
texture



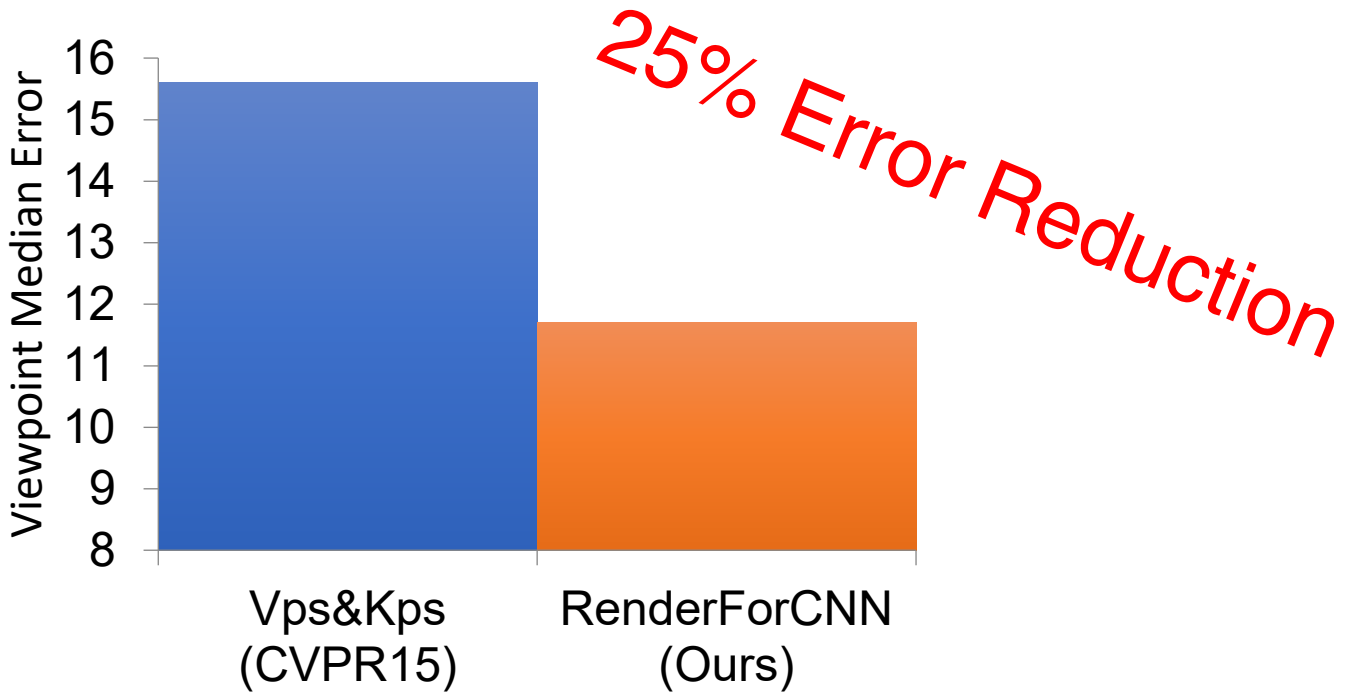
86% -> 93%

Key Lesson: **Don’t give CNN a chance to “cheat” - it’s very good at it. When there is no way to cheat, true learning starts.**

[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

3D Viewpoint Estimation Evaluation

Our model **trained on rendered images** outperforms state-of-the-art model **trained on real images** in PASCAL3D+.



[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

3D Viewpoint Estimation

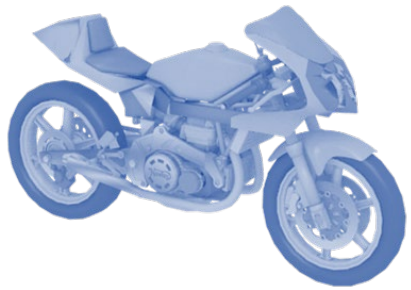








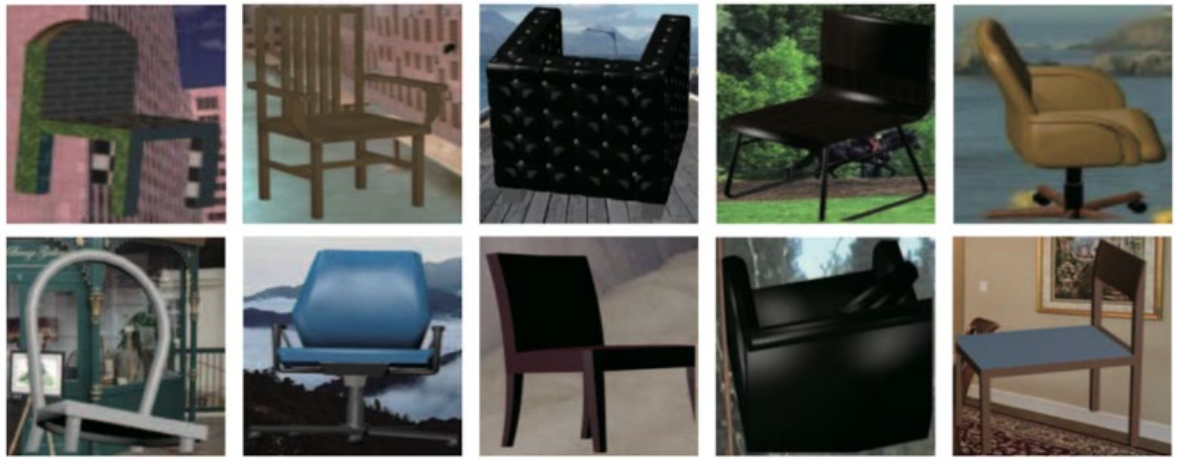


Interactive 3D Simulation

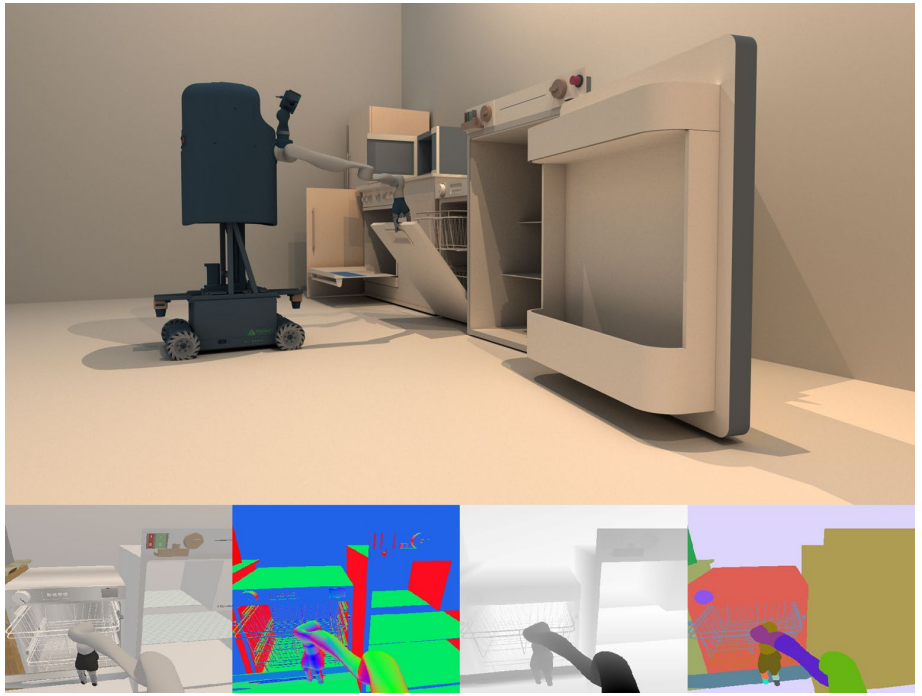
Many tasks require **interaction** with objects and **dense annotations**.

Static, task-specific



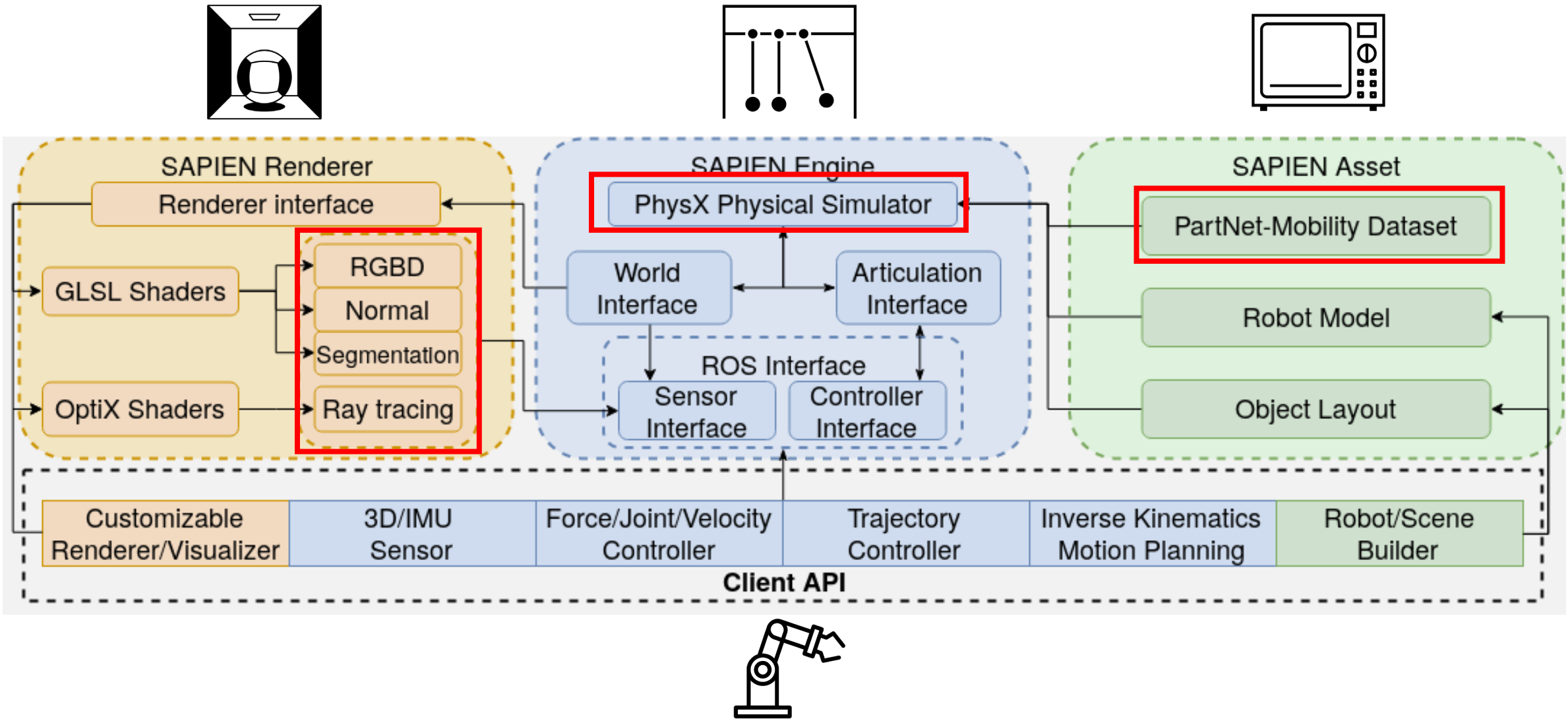
[H. Su, C. Qi, Y. Li, and L. Guibas, *Render for CNN*, ICCV 2015]

Interactive simulation



[F. Xiang et al., SAPIEN: A SimulAted Part-based Interactive ENvironment, CVPR 2020]

Interactive 3D Simulation

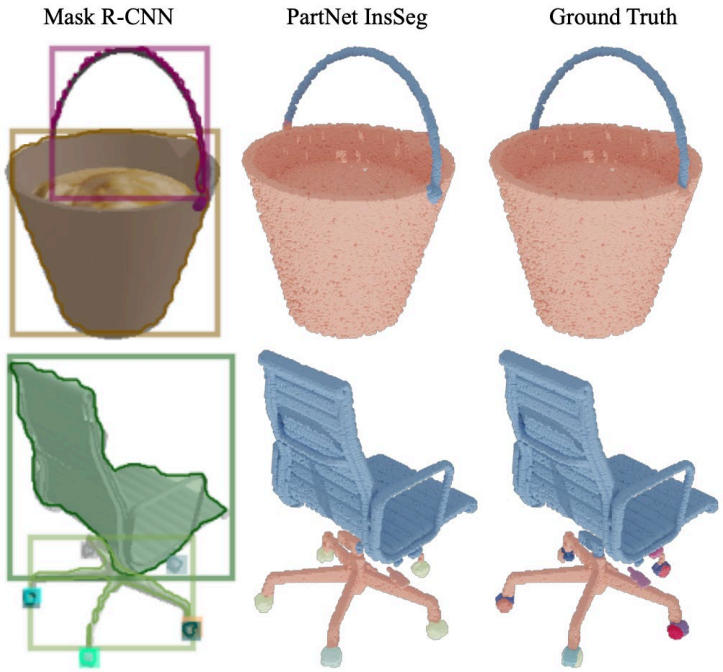


[F. Xiang et al., SAPIEN: A SimulAted Part-based Interactive ENvironment, CVPR 2020]

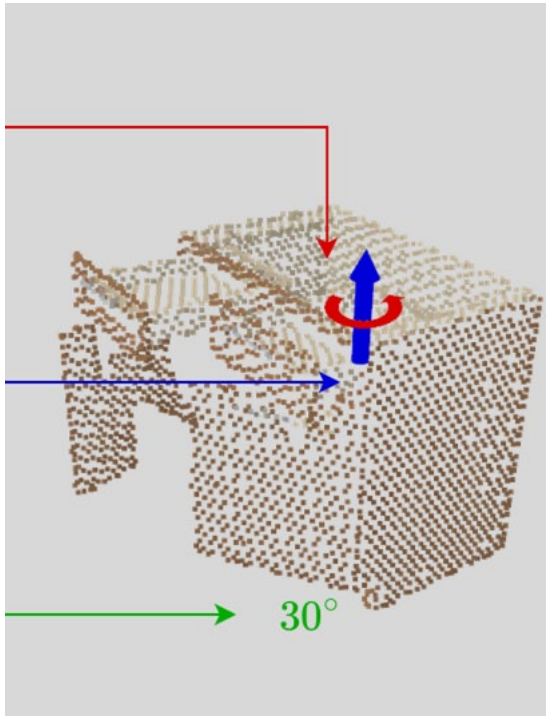
Data for Many Tasks

Simulation provides data for **both vision and interaction** tasks.

Movable Part Segmentation



Motion Attributes Estimation



Robotic Interaction

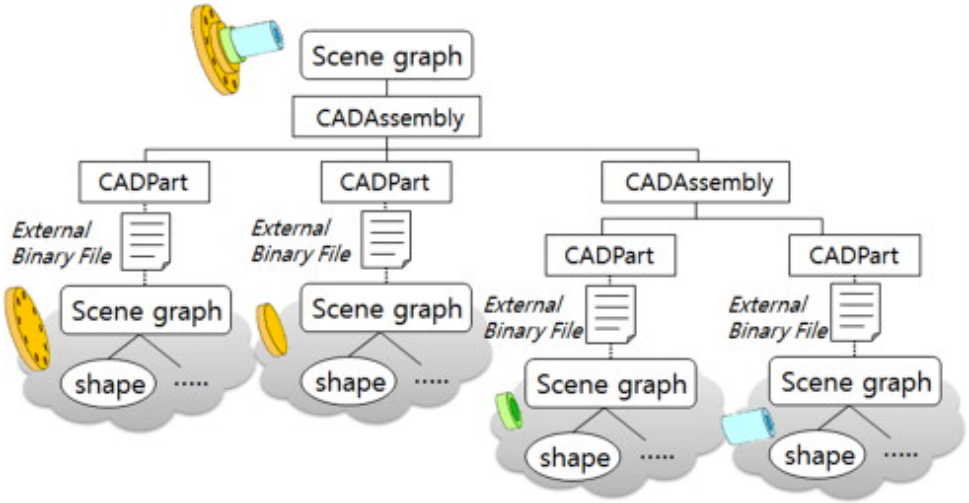
[F. Xiang et al., SAPIEN: A SimulAted Part-based Interactive ENvironment, CVPR 2020]

Learning from Noisy Web Data

Observations

CAD data on the Web often include *scene graphs*:
Part geometry + hierarchical structure

ShapeNet



Kim et al., 2015

Observations — Abundant Shapes

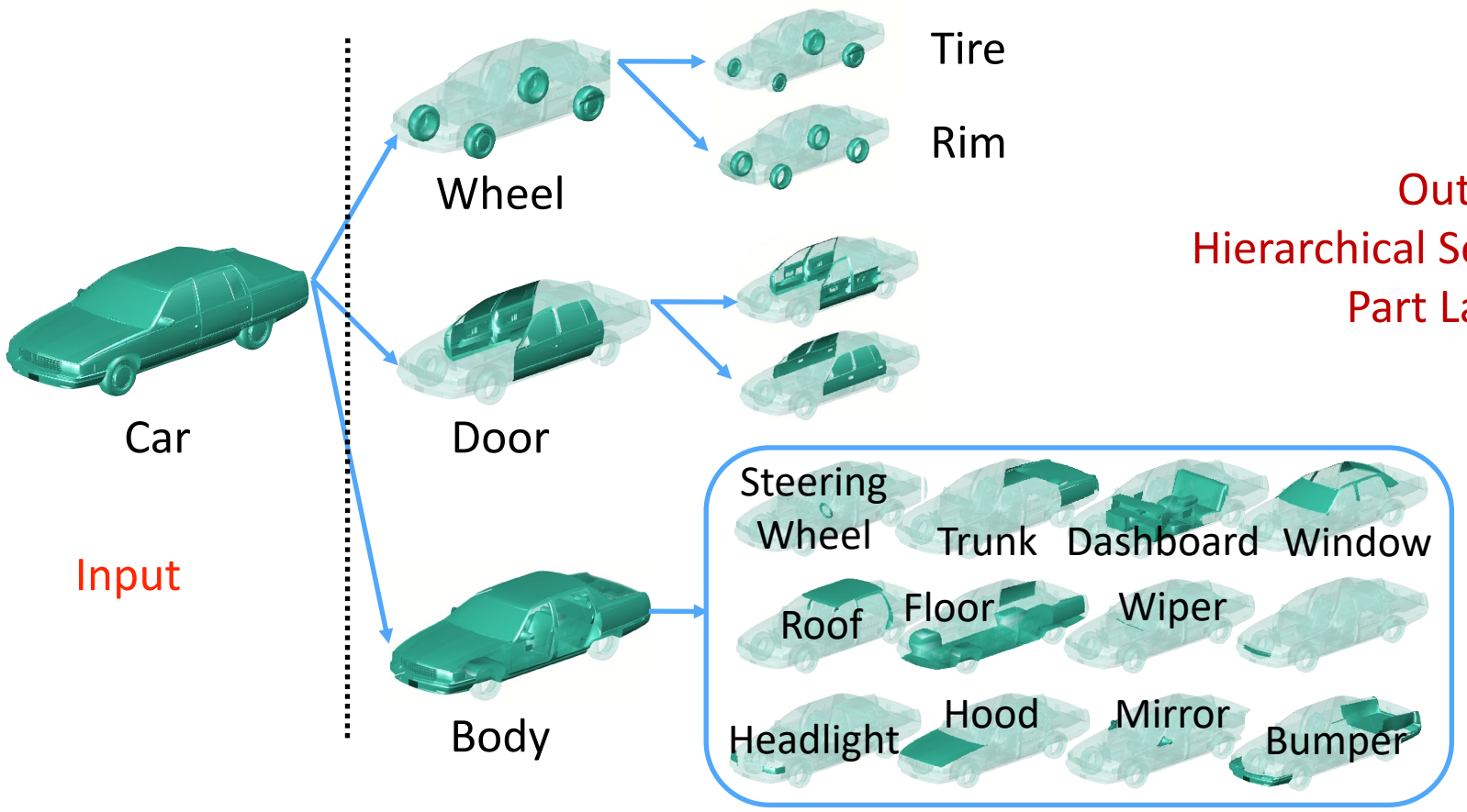
Observations

(+) Provides natural part segmentations.

(–) Inconsistent -- and often unlabeled.

An Application of Horizontal Networks

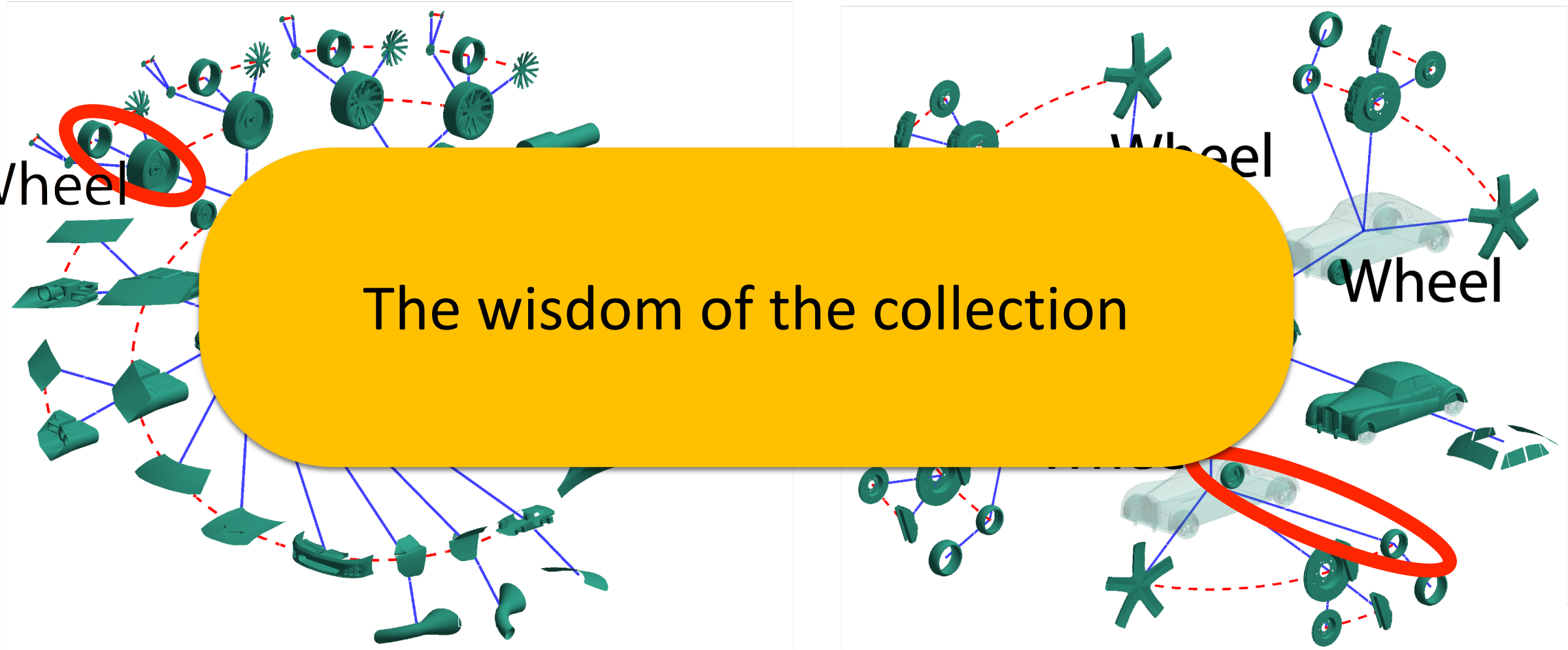
- Learning hierarchical shape segmentation and labeling in a weakly supervised manner from online repositories
- Based on “horizontal” information diffusion



Output:
Hierarchical Segmentation &
Part Labeling

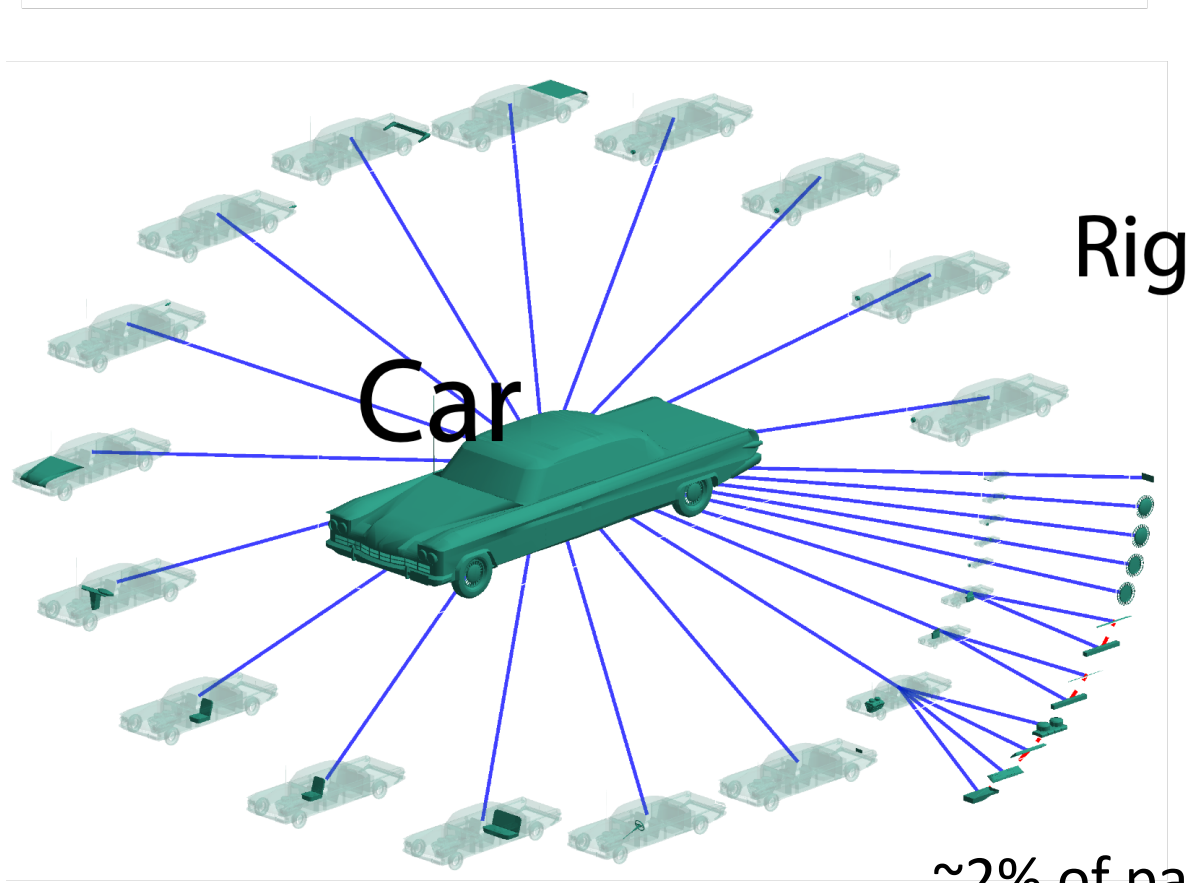
[Yi et al., Learning Hierarchical Shape Segmentation and Labeling from Online Repositories, SIGGRAPH 2017]

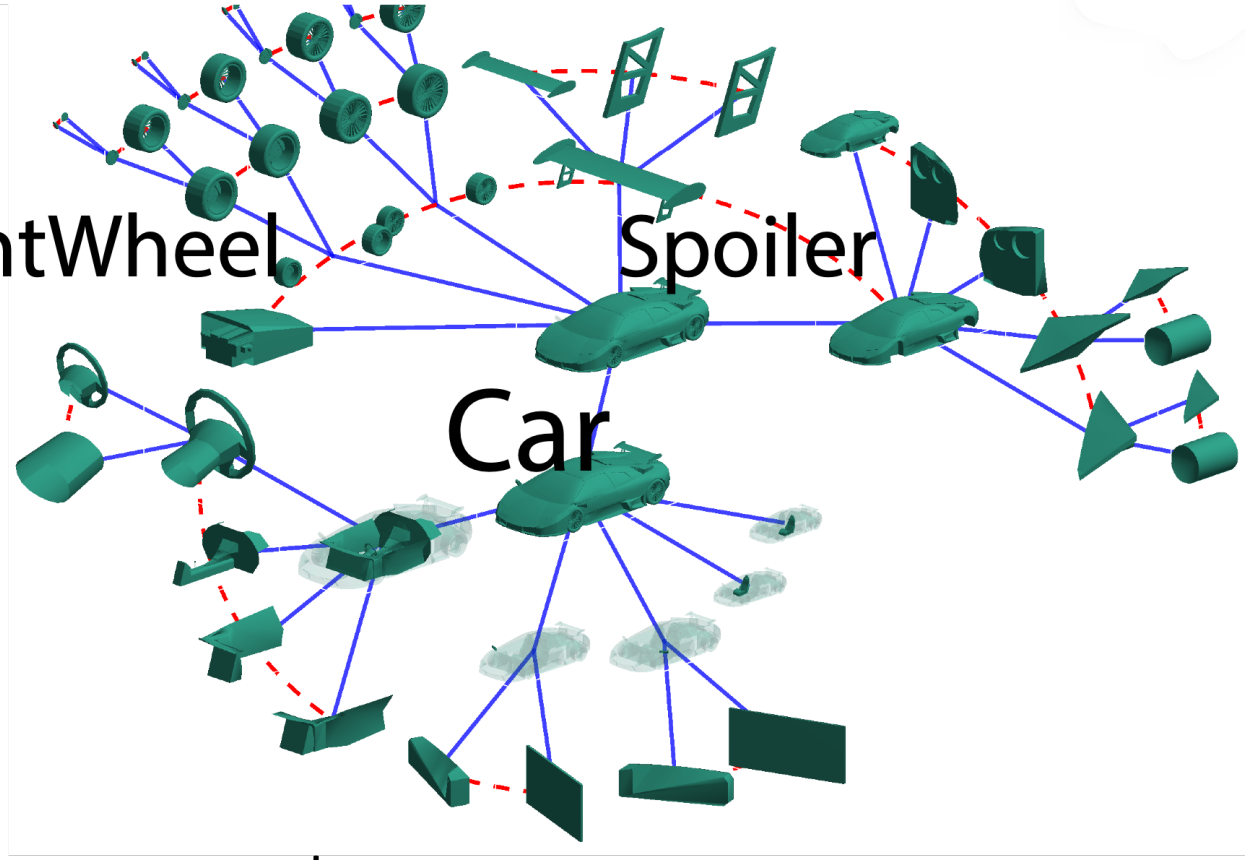
Common Structures in Object Graphs



Challenges - Heterogeneous Data

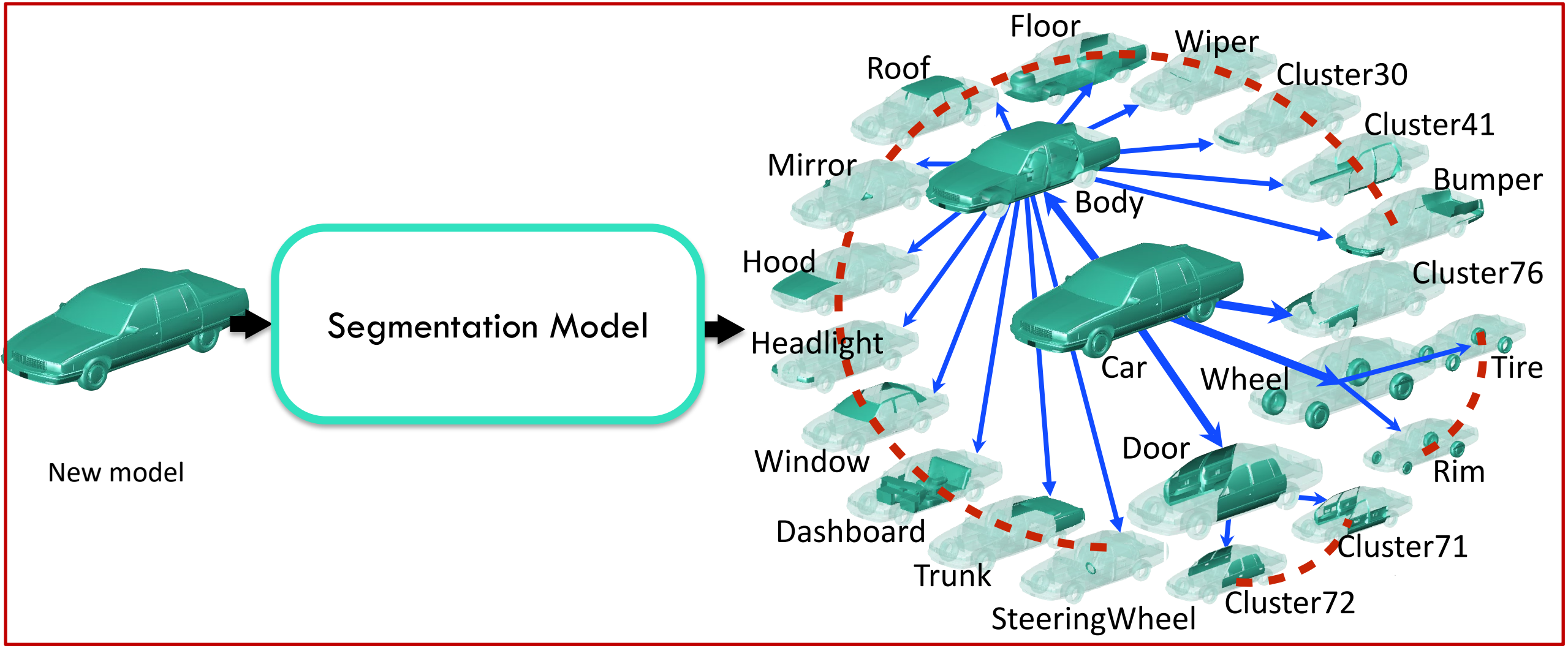
Different Parts, Different Hierarchy, Sparse Tagging



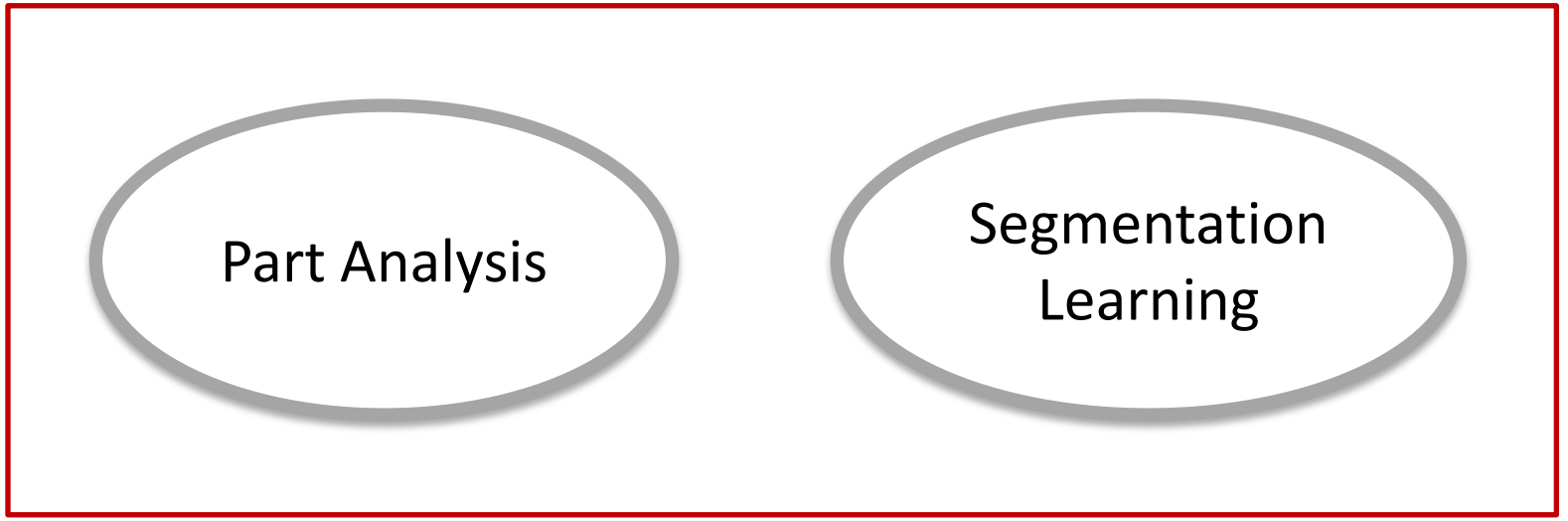


~2% of parts are tagged

Train a Network that Can Segment and Label

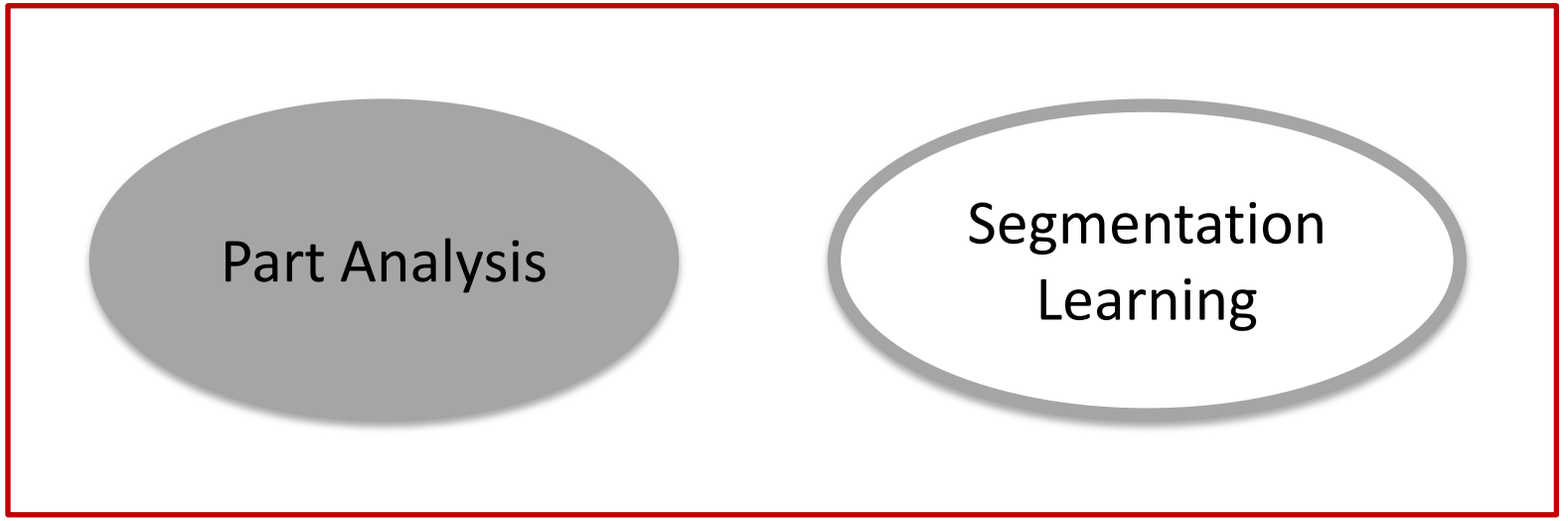


Approach Overview



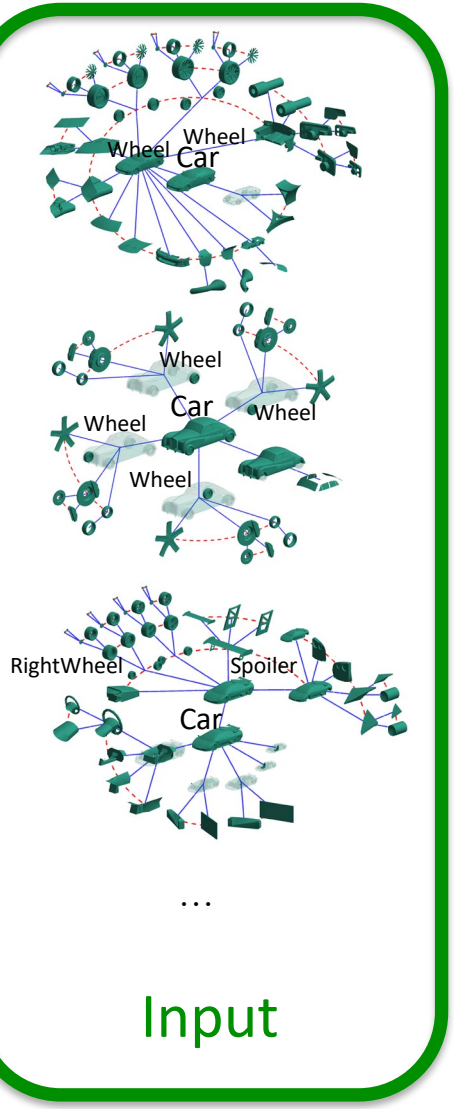
Training Stage

Approach Overview

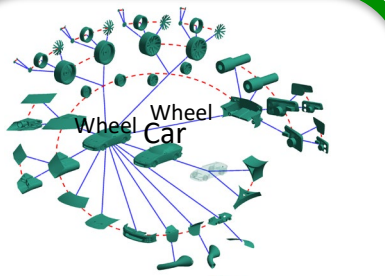


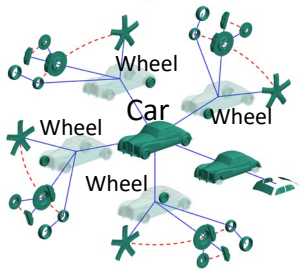
Training Stage

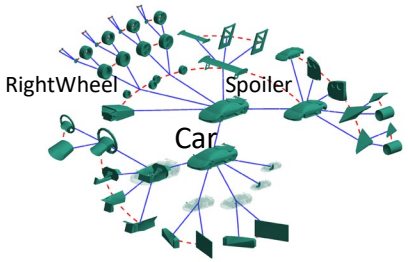
Part Analysis



Part Analysis

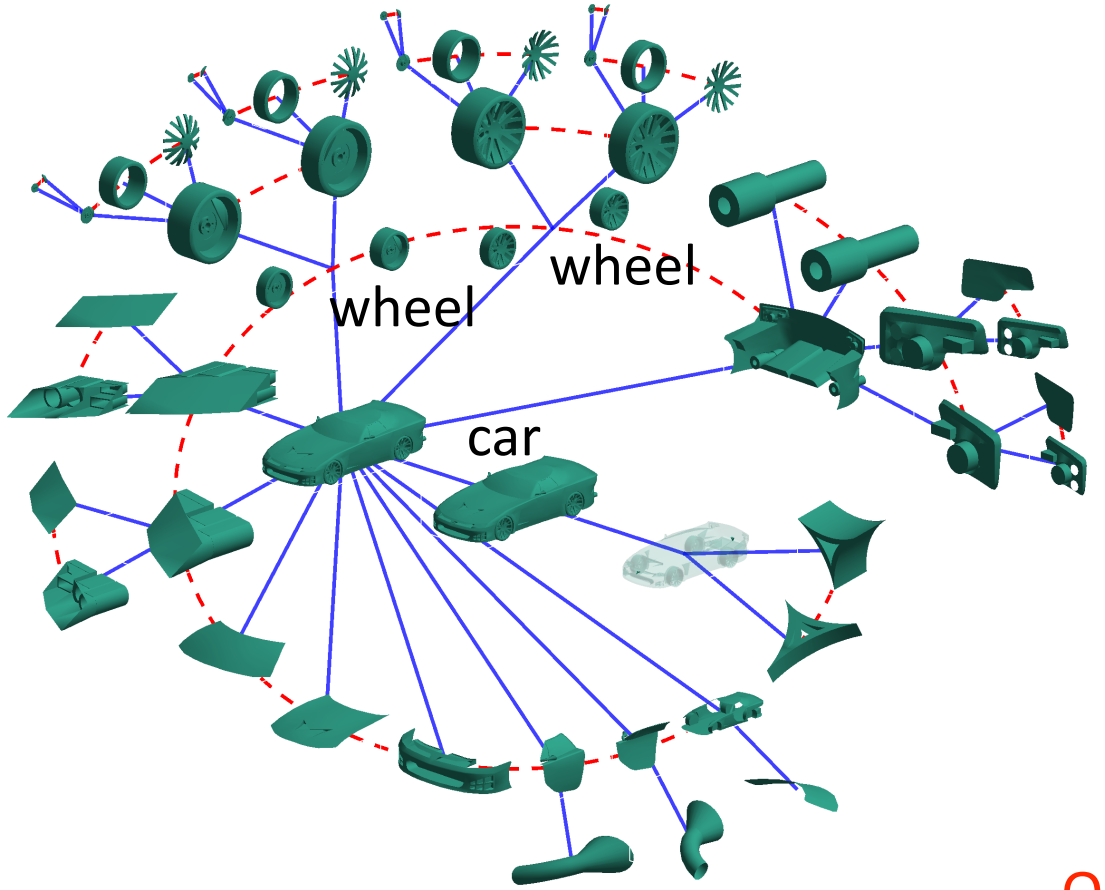






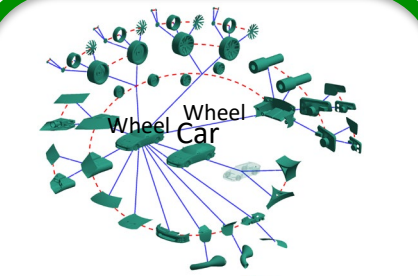
...

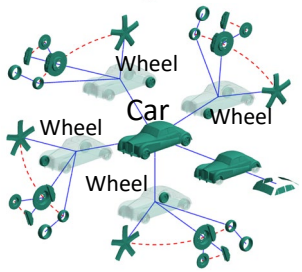
Input

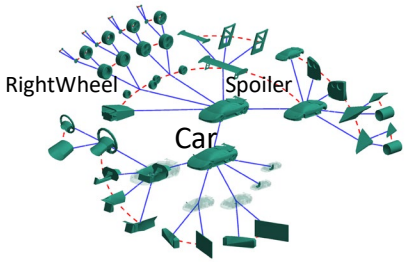


Output

Part Analysis

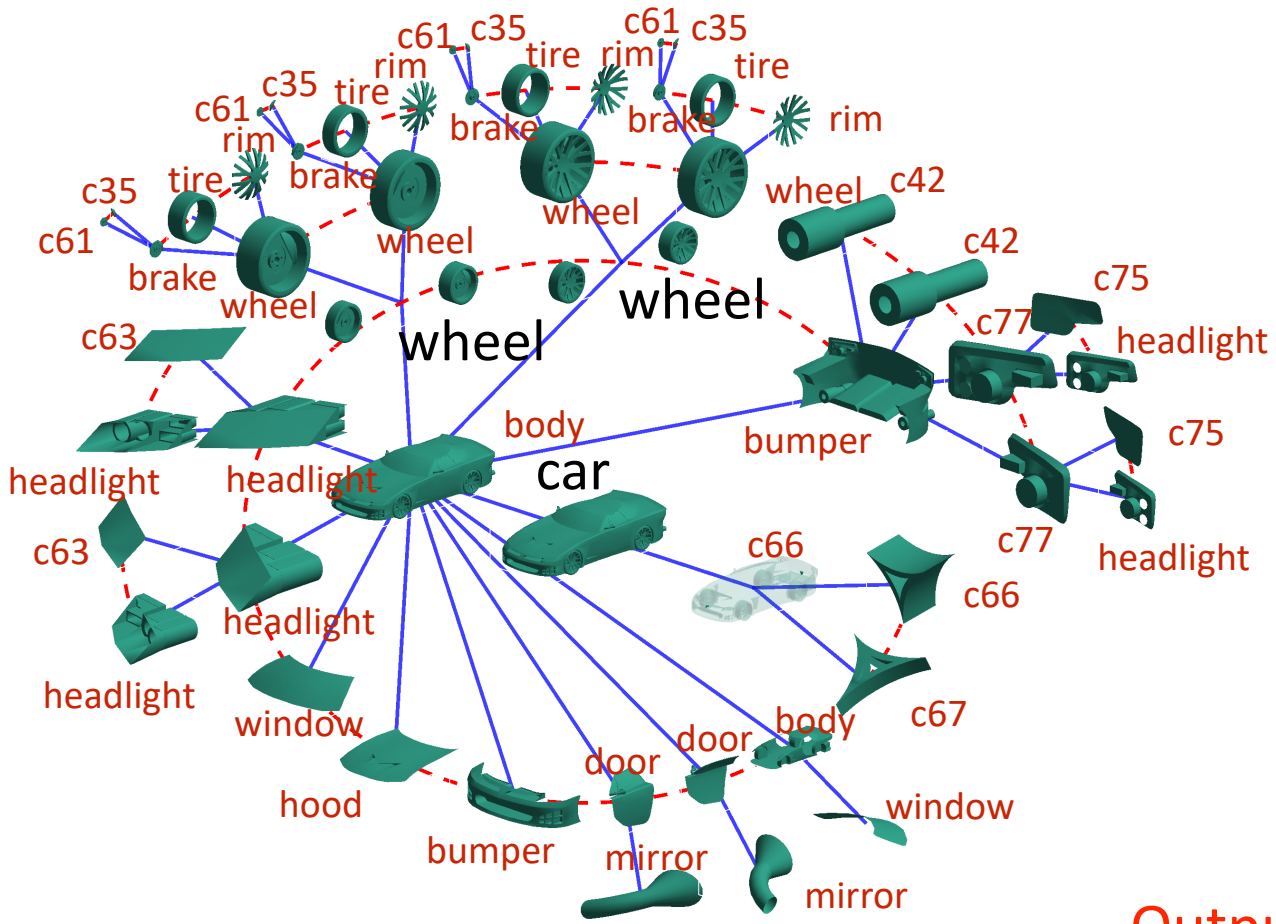






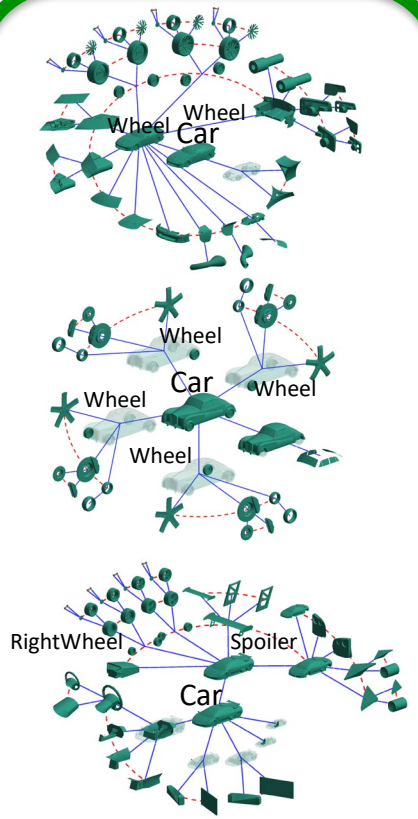
...

Input

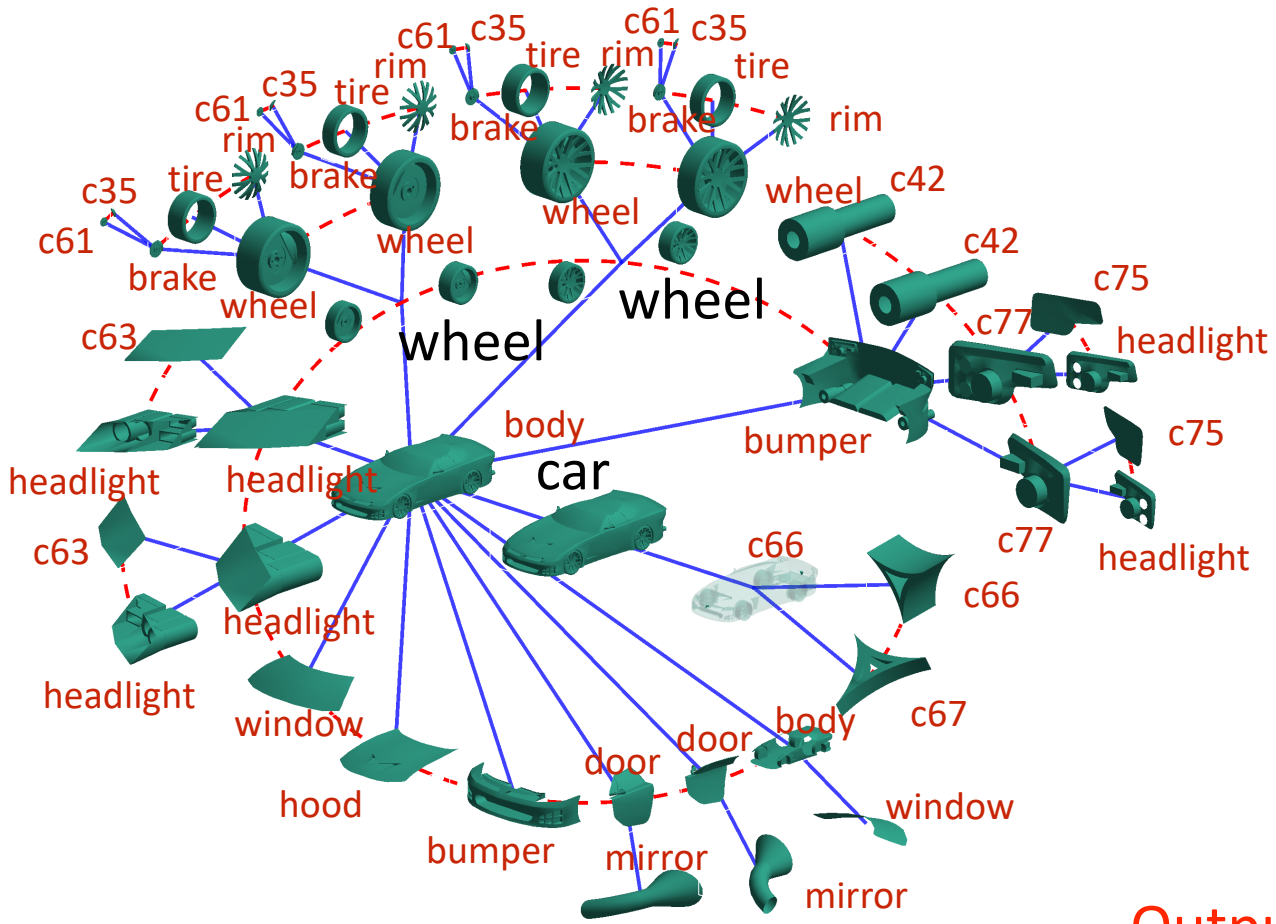


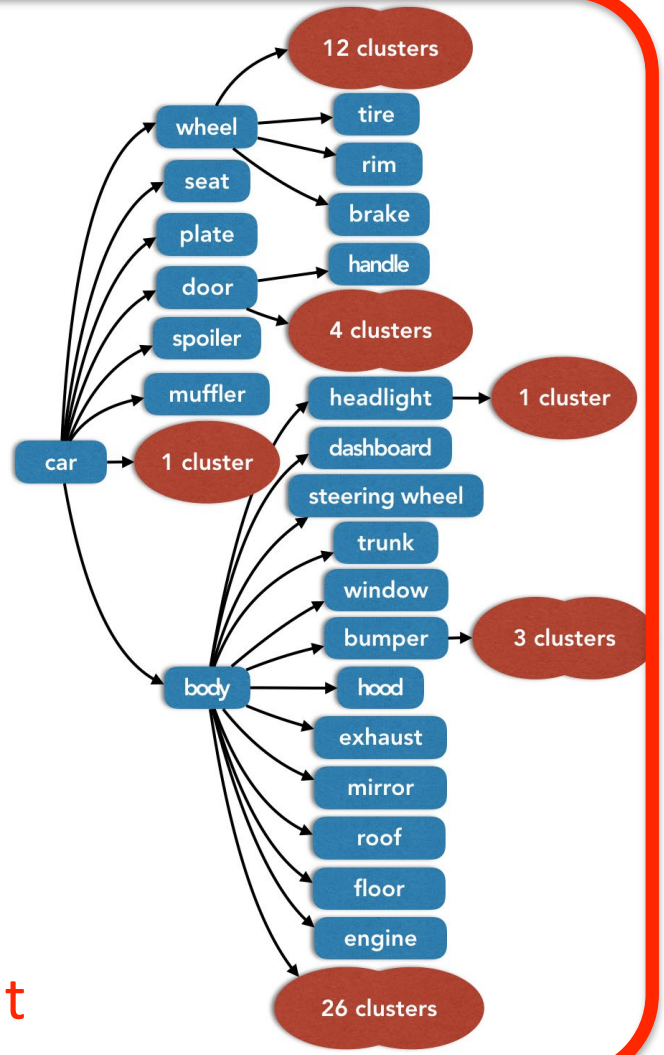
Output

Part Analysis



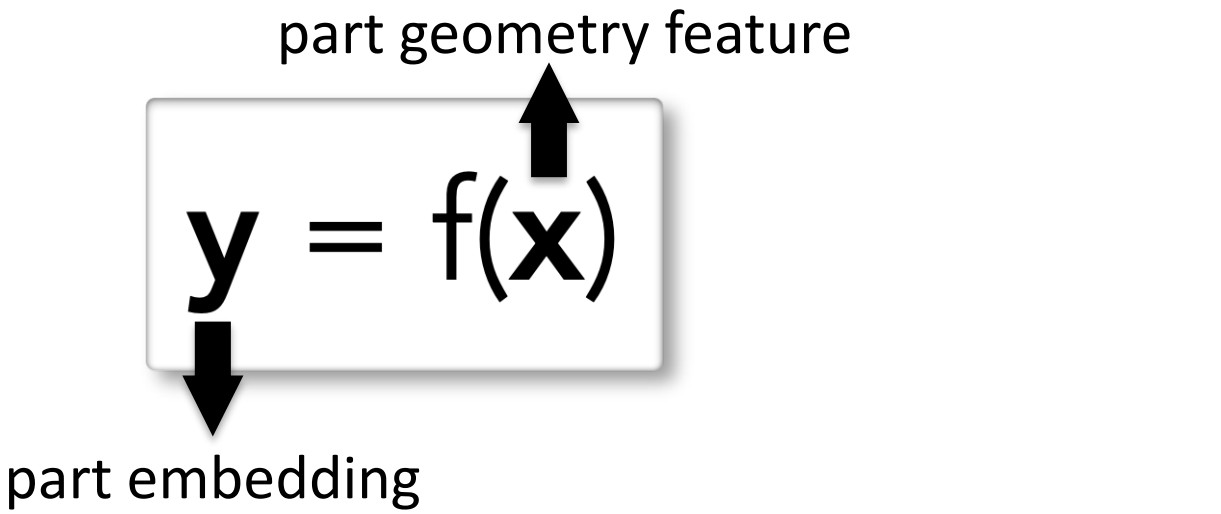
Input





Output

Key Idea — Semi-Supervised Clustering



cluster centroid

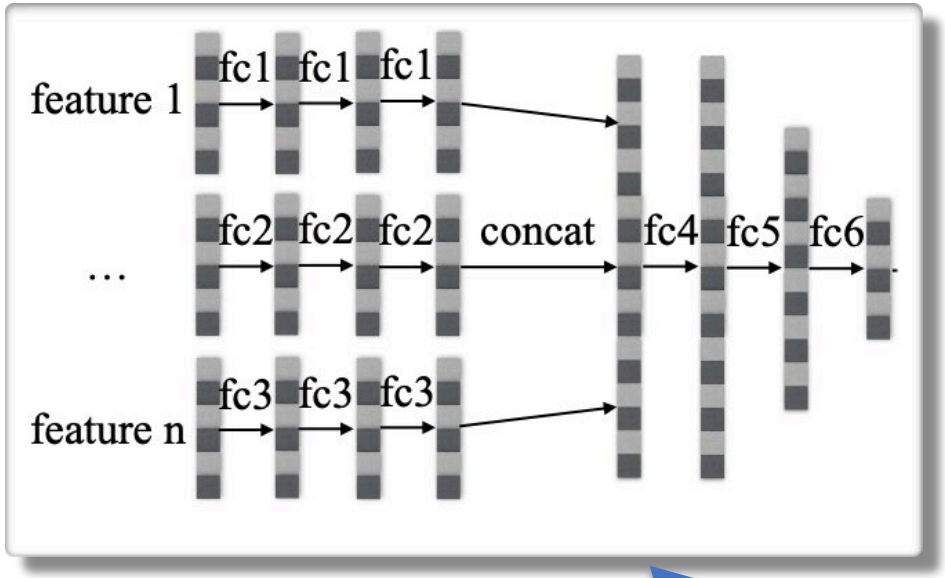
$$\text{clustering:}\quad \min_j \| \mathbf{c}_j - f(\mathbf{x}) \|$$

Key Idea — Semi-Supervised Clustering

part geometry feature

$$y = f(x)$$

part embedding

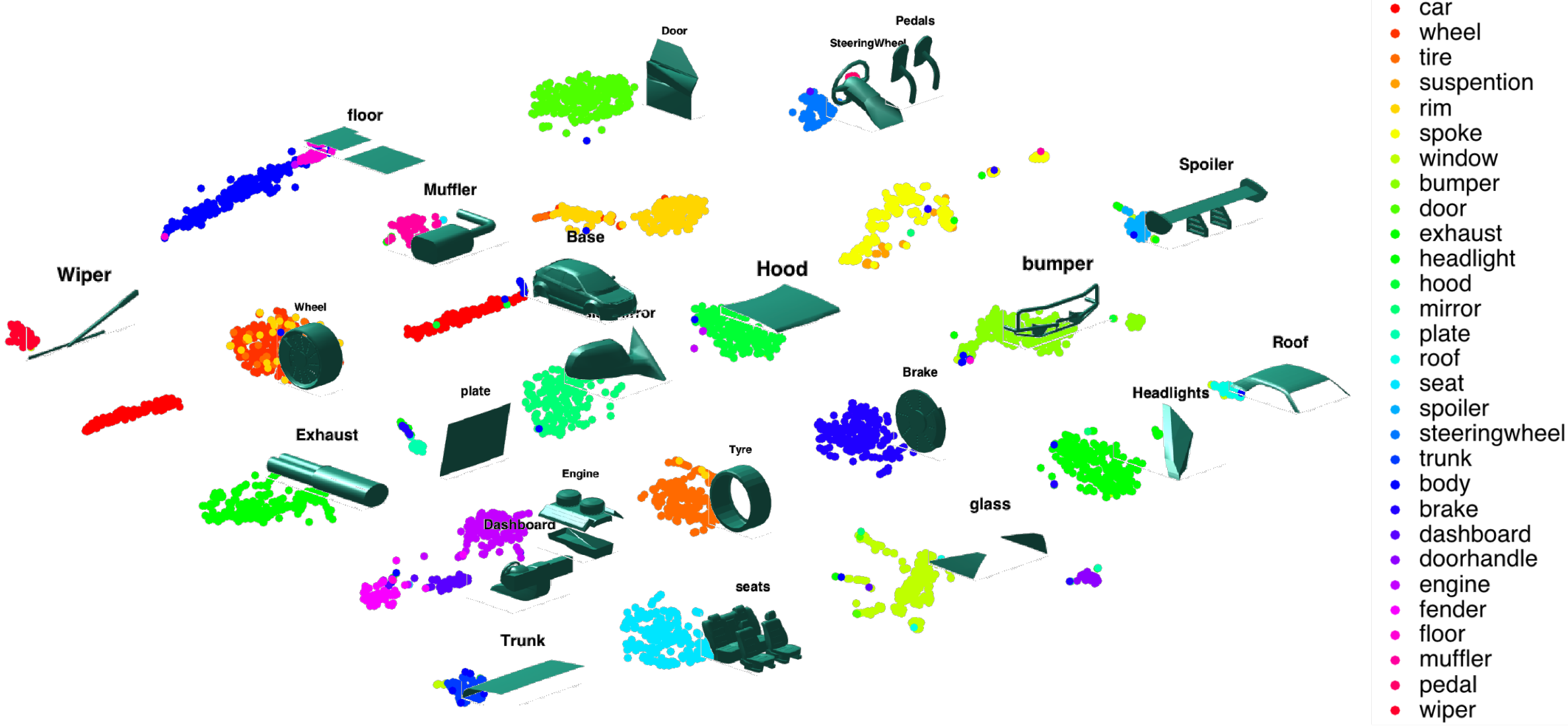


cluster centroid

clustering: $\min_i \| c_i - f(x) \|$

Supervision: sparse tags, inconsistent hierarchies

Key Idea — Semi-Supervised Clustering



EM Algorithm: Objective Function

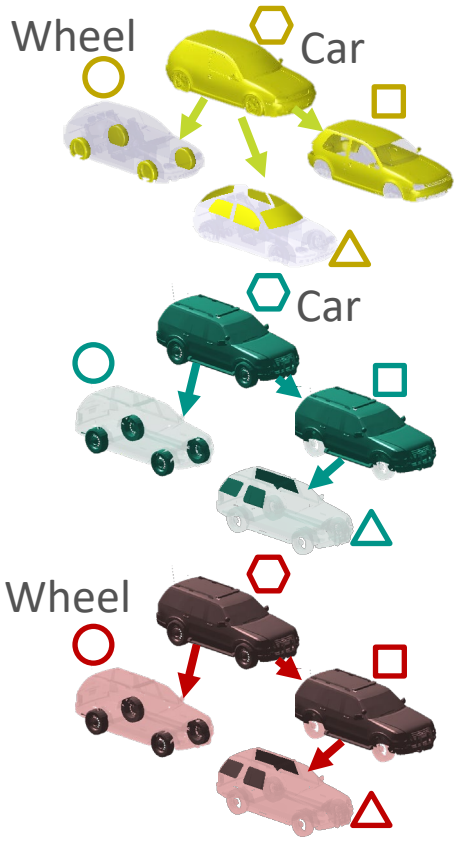
Clustering

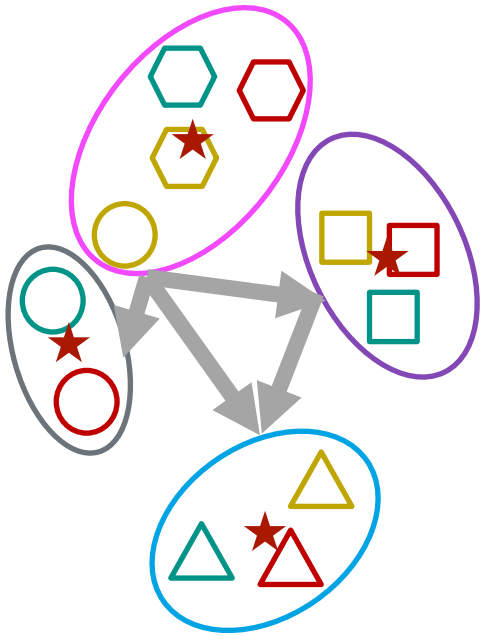
Dissimilarity

$$E(\theta, p, c, M) = \lambda_c E_c + \lambda_s E_s + \lambda_d E_d + \lambda_m E_m - H$$

Similarity

Structure

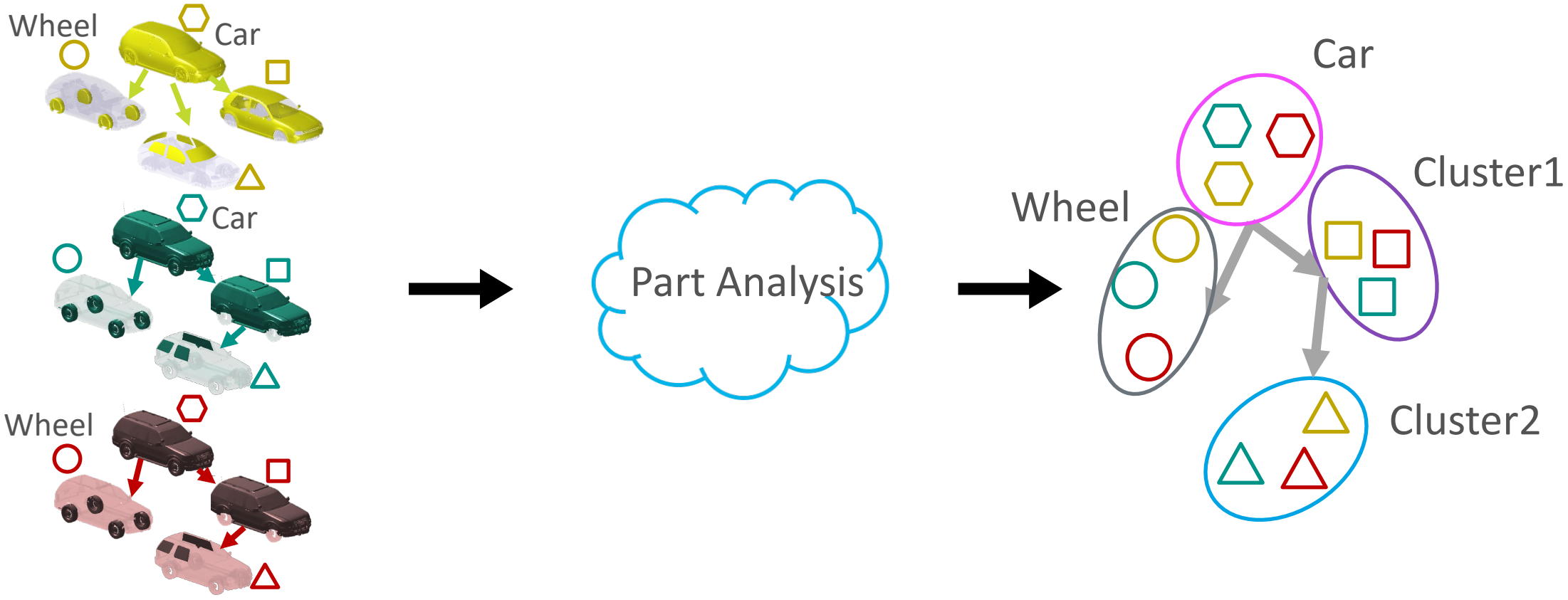




Optimize:

- Embedding network
- Cluster membership and centers
- Hierarchy (soft adjacency matrix)

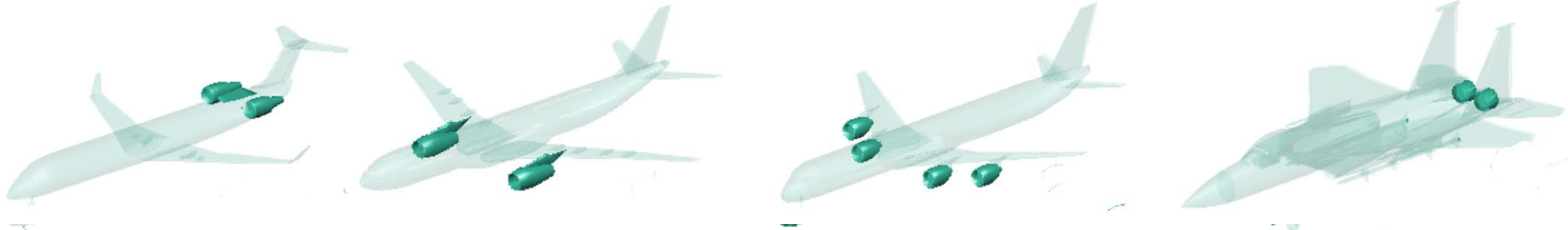
Outputs


Wheel
Car
Car
Wheel
Part Analysis
Car
Cluster1
Wheel
Cluster2

Sample Clusters

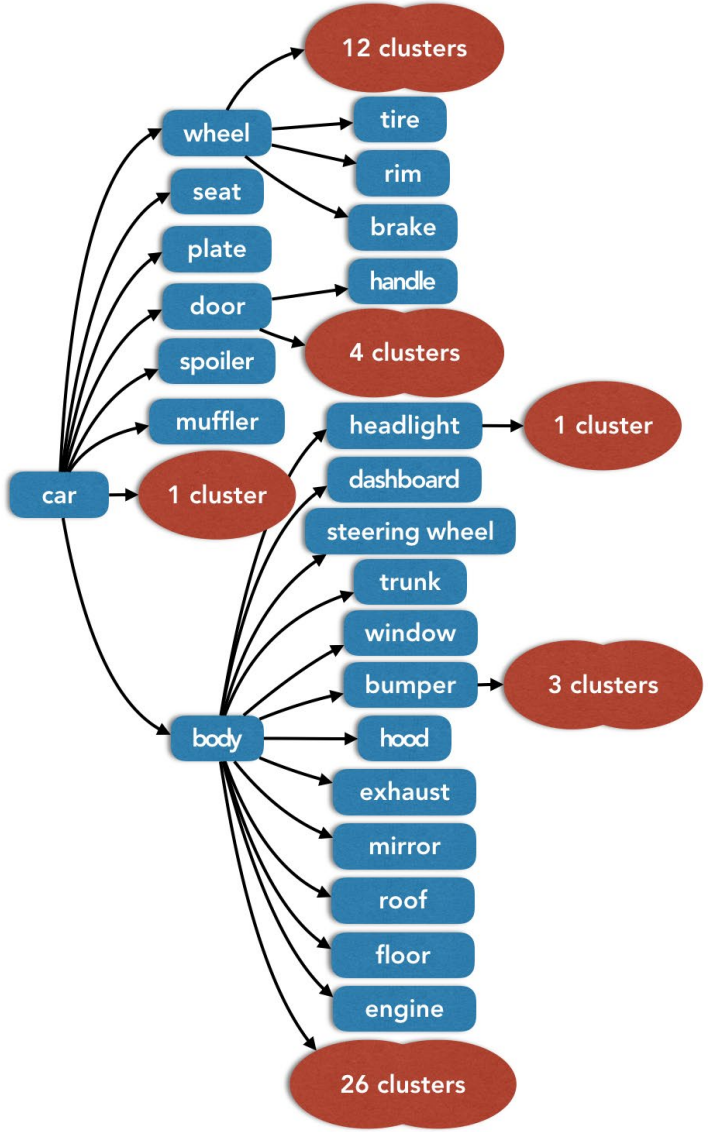
tail

engine

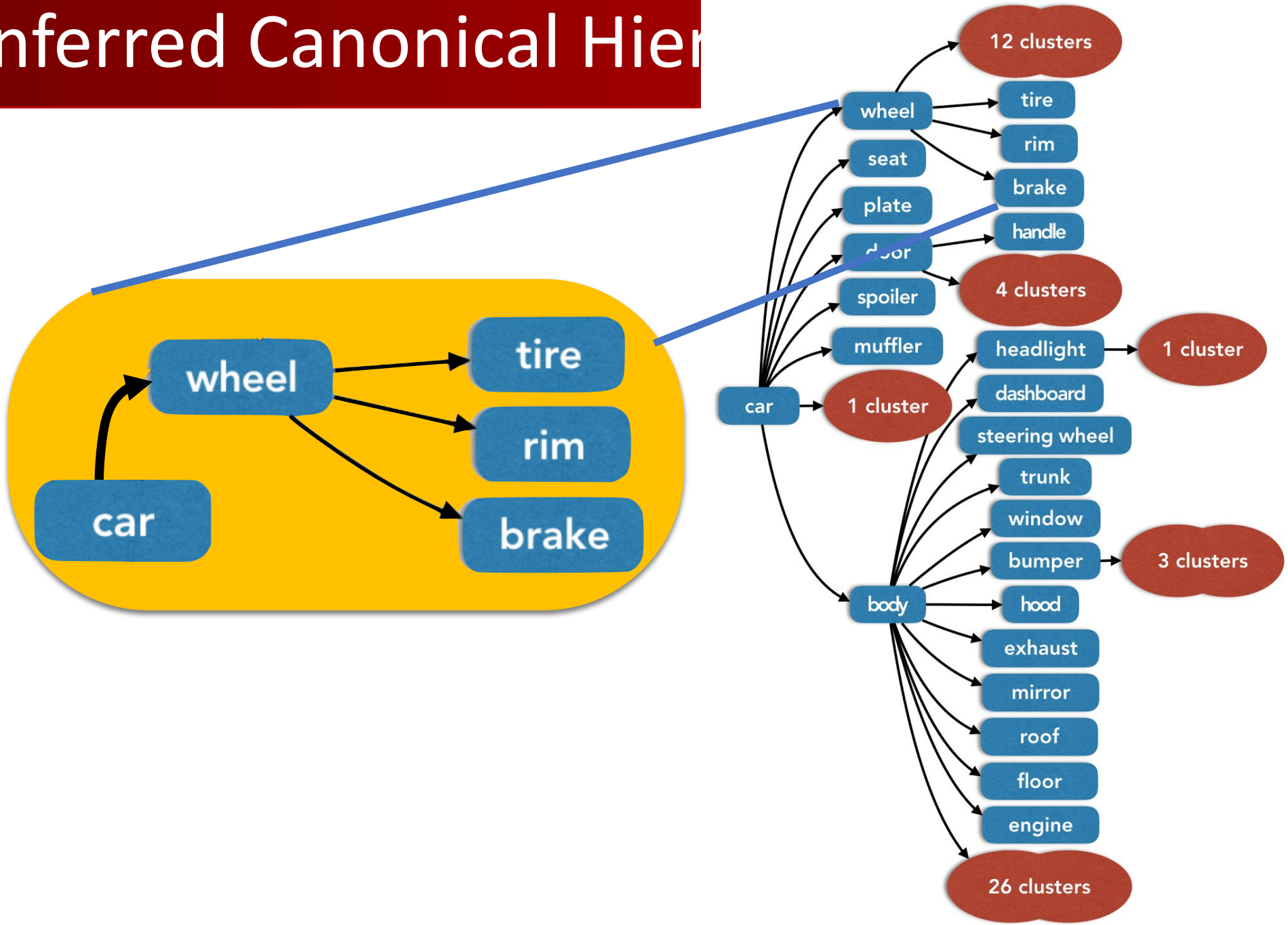


cluster1

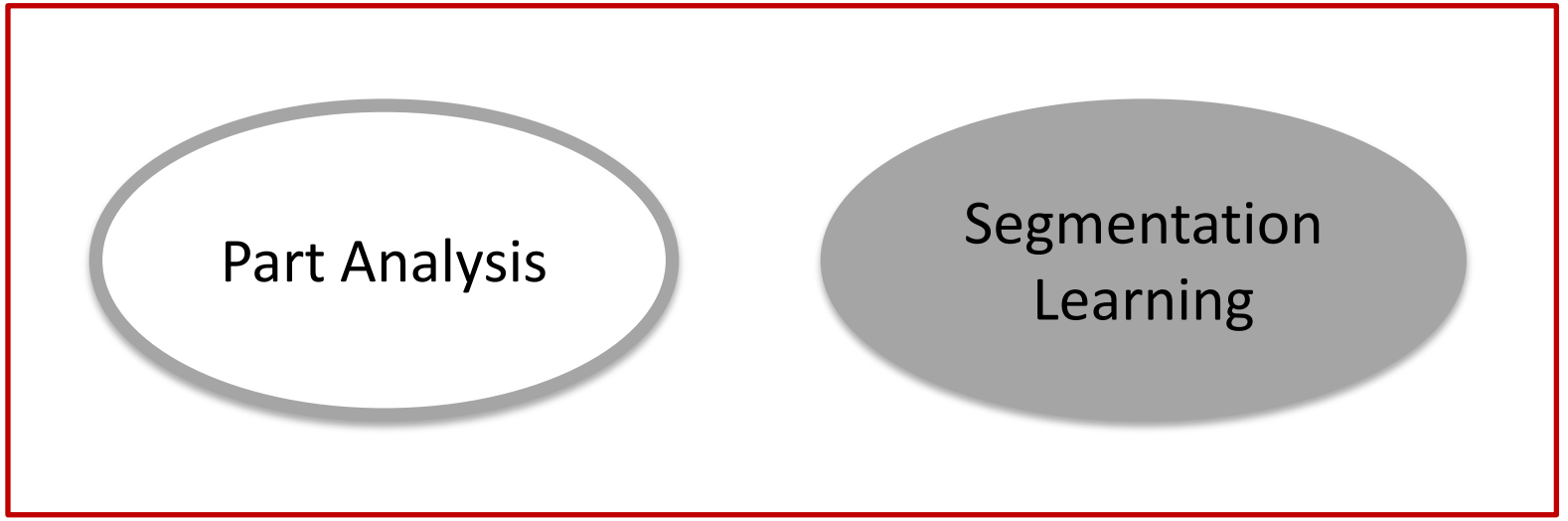
Inferred Canonical Hierarchy



Inferred Canonical Hier

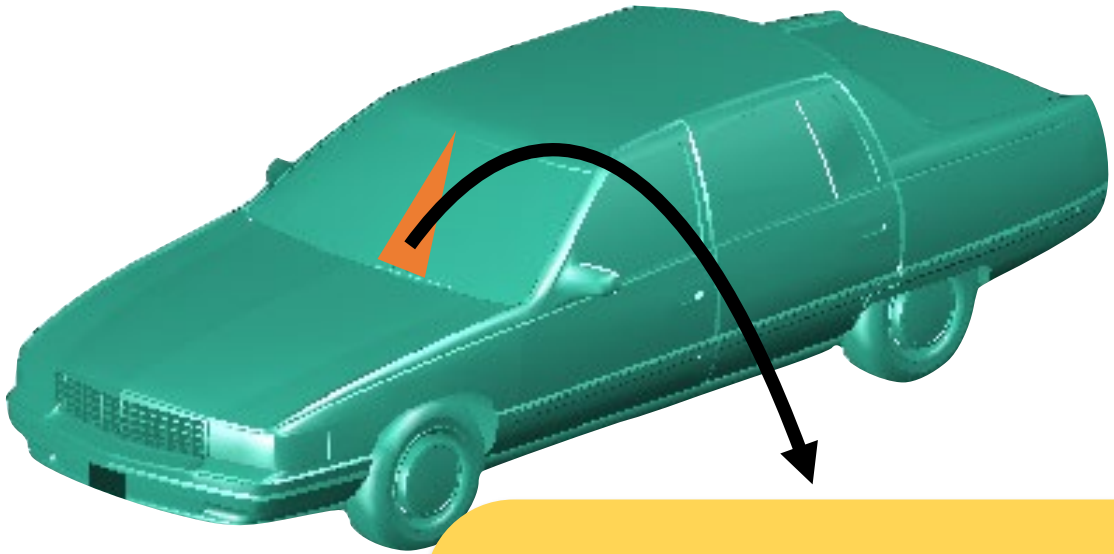


Approach Overview



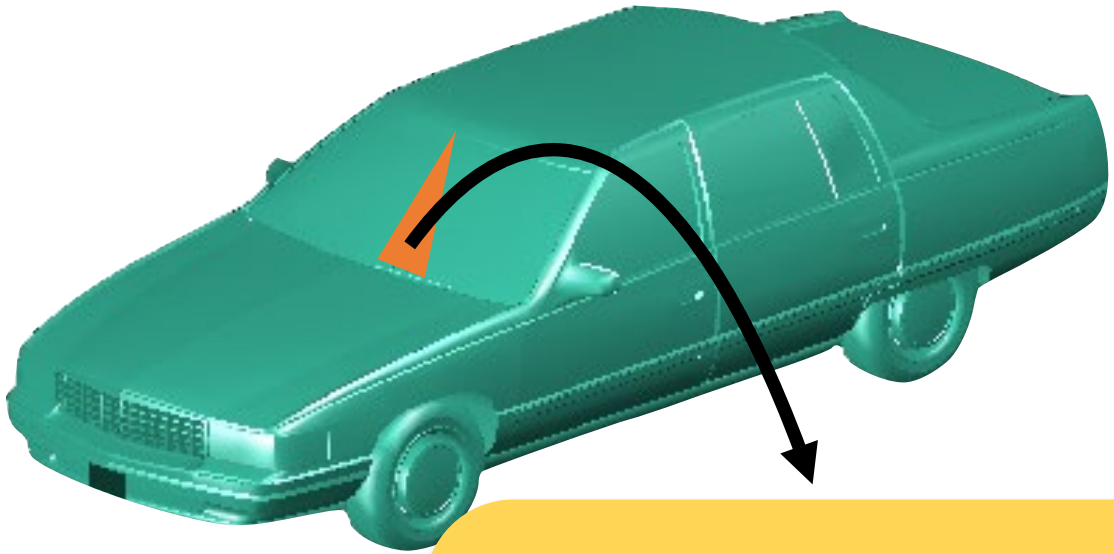
Training Stage

Learning Hierarchical Mesh Segmentation

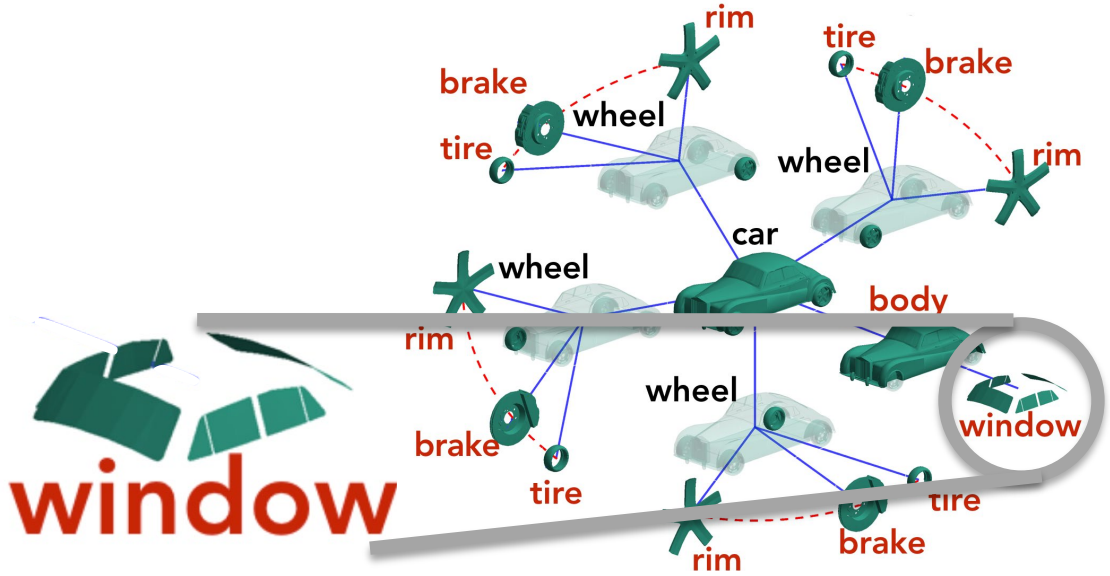


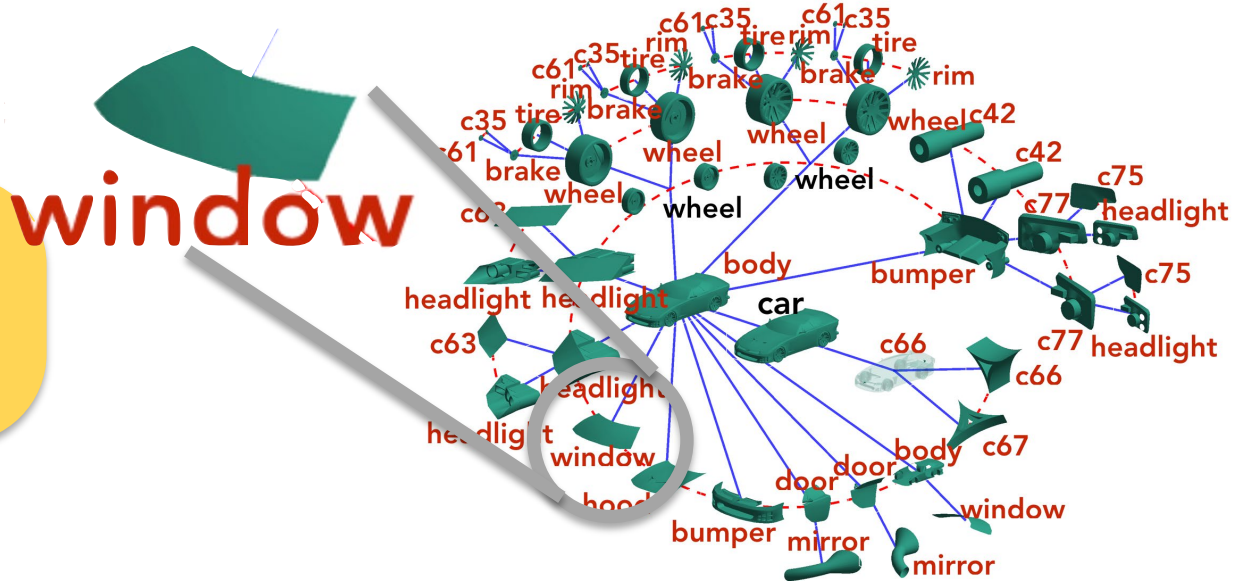
What's the label of this triangle face?

Learning Hierarchical Mesh Segmentation



What's the label of this triangle face?





Learning Hierarchical Mesh Segmentation

A similar MRF formulation as Kalogerakis et al. 2010

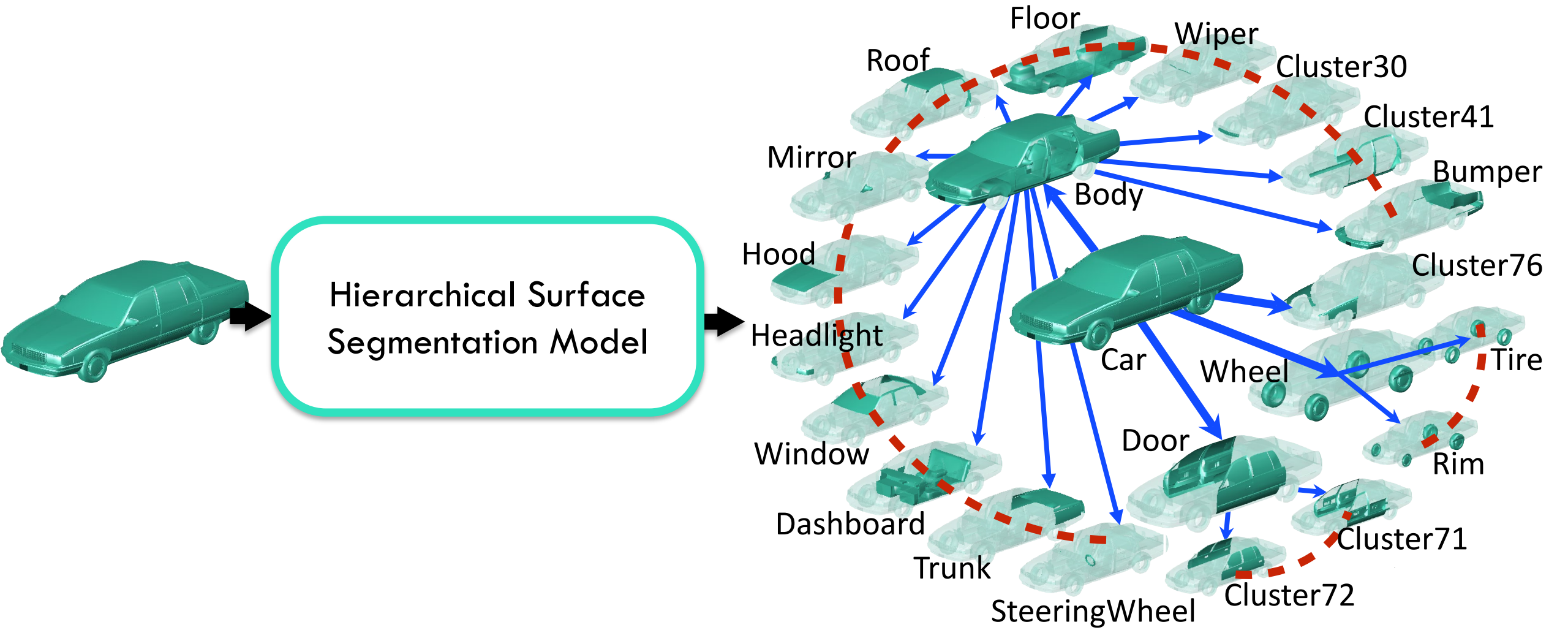
$$E(L) = \sum_{c} \psi_{\text{unary}}(L_c) + \lambda \sum_{u,v \in \mathbb{E}} \psi_{\text{edge}}(L_u, L_v)$$

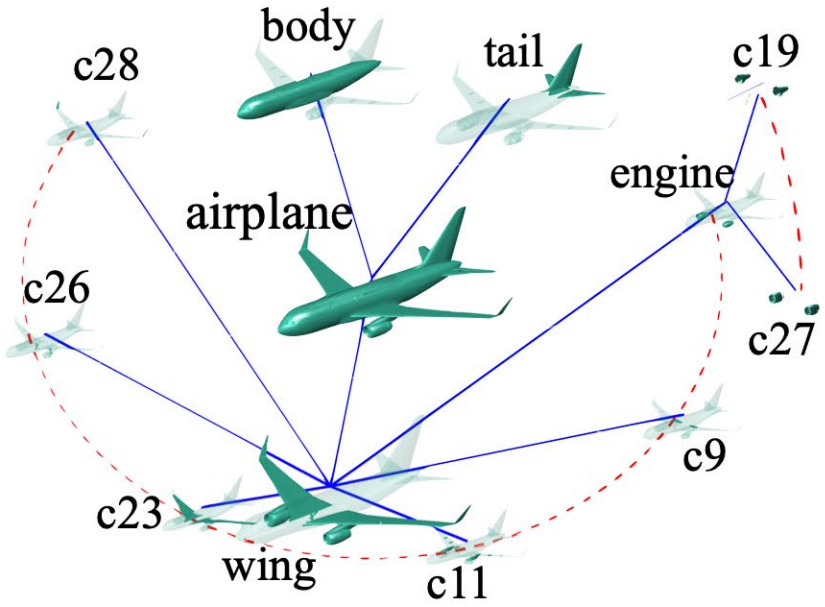
Learned classifier that predicts label confidence from geometry

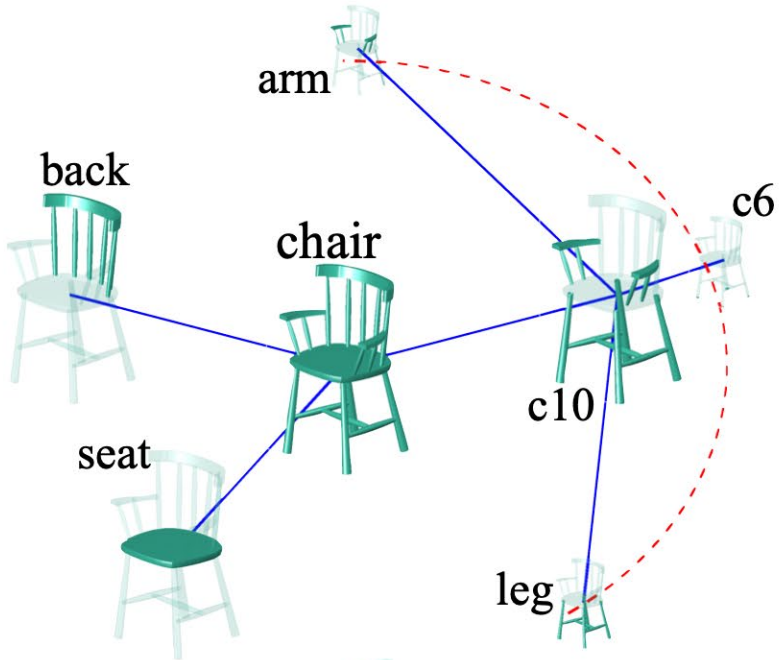
Encodes label structures – adjacent components should have same label

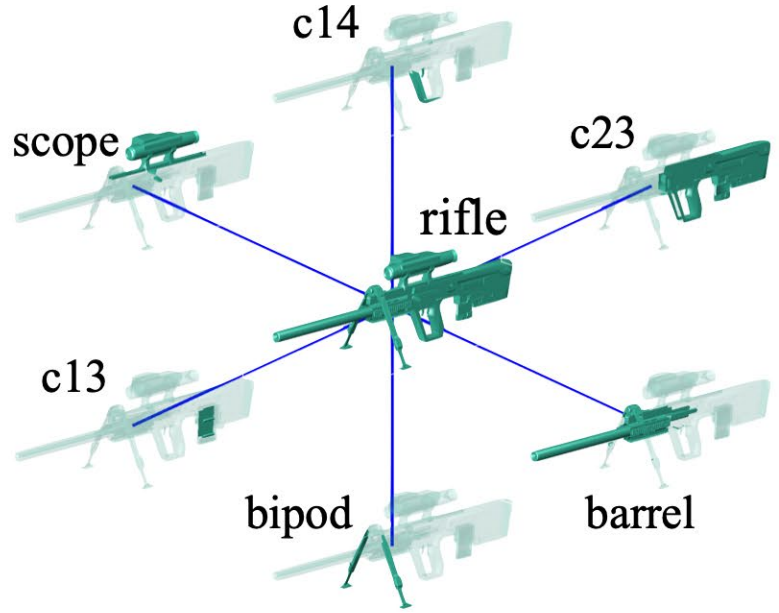
- infer hierarchy,
- handle incomplete training segmentations,
- handle disconnected surfaces

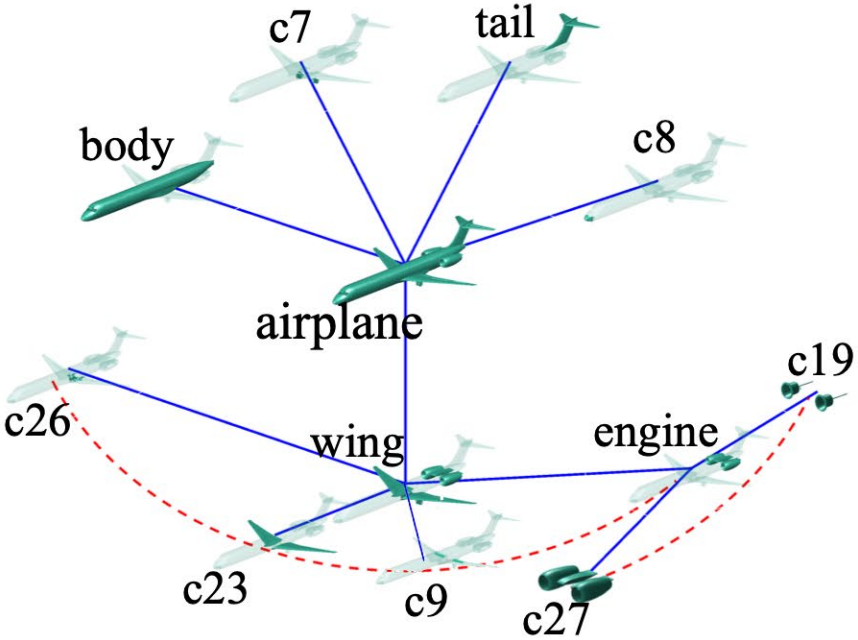
Segmentation Output

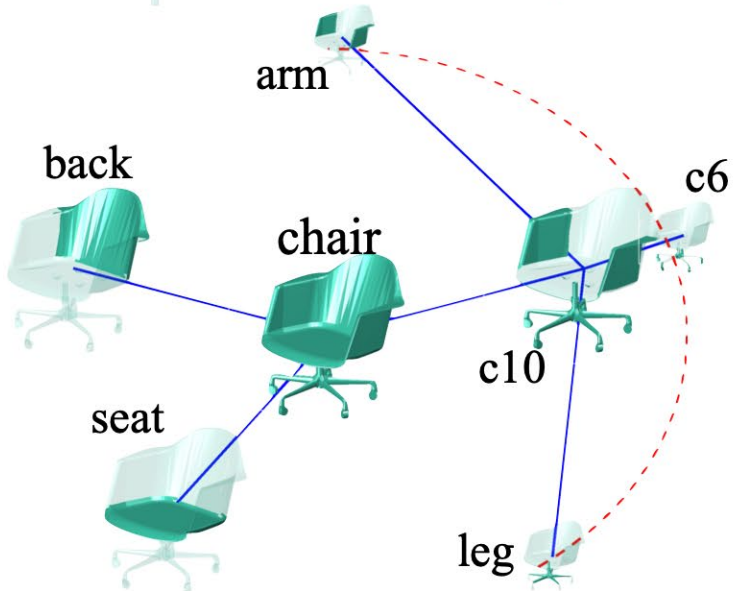


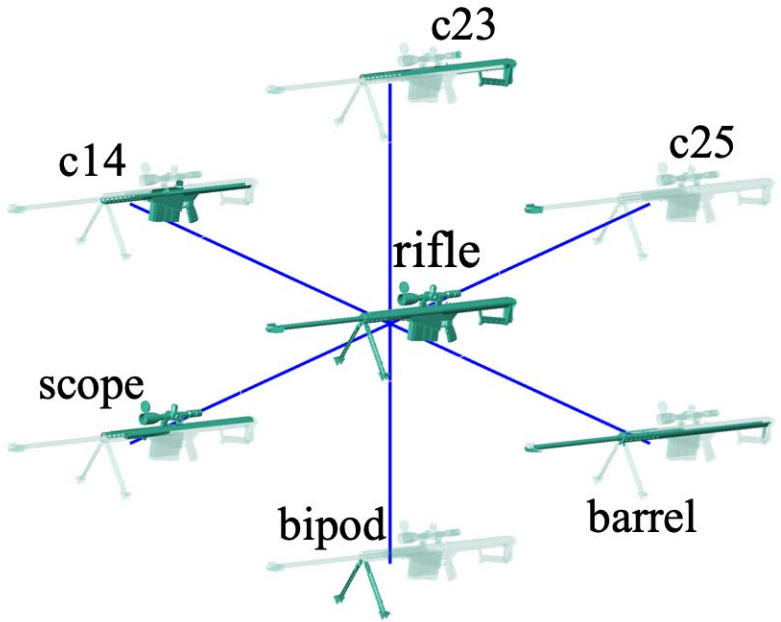

body
c28
tail
c19
engine
airplane
c26
c27
c9
c23
wing
c11


arm
back
c6
chair
c10
seat
leg


c14
scope
c23
rifle
c13
bipod
barrel


c7
tail
body
c8
airplane
c19
c26
wing
engine
c23
c9
c27


arm
back
c6
chair
c10
seat
leg


c23
c14
c25
rifle
scope
bipod
barrel

Take-Home Message

- Distill wisdom from the crowd
- Knowledge emerges while jointly analyzing a collection of shapes
- A novel method for mining massive but sparsely annotated object graphs “in the wild”

Additional 3D Datasets

ABC: with Parametrized Curves and Surfaces

- Synthetic, 3D CAD Models
- > 1M Models
- mostly, Mechanical Parts

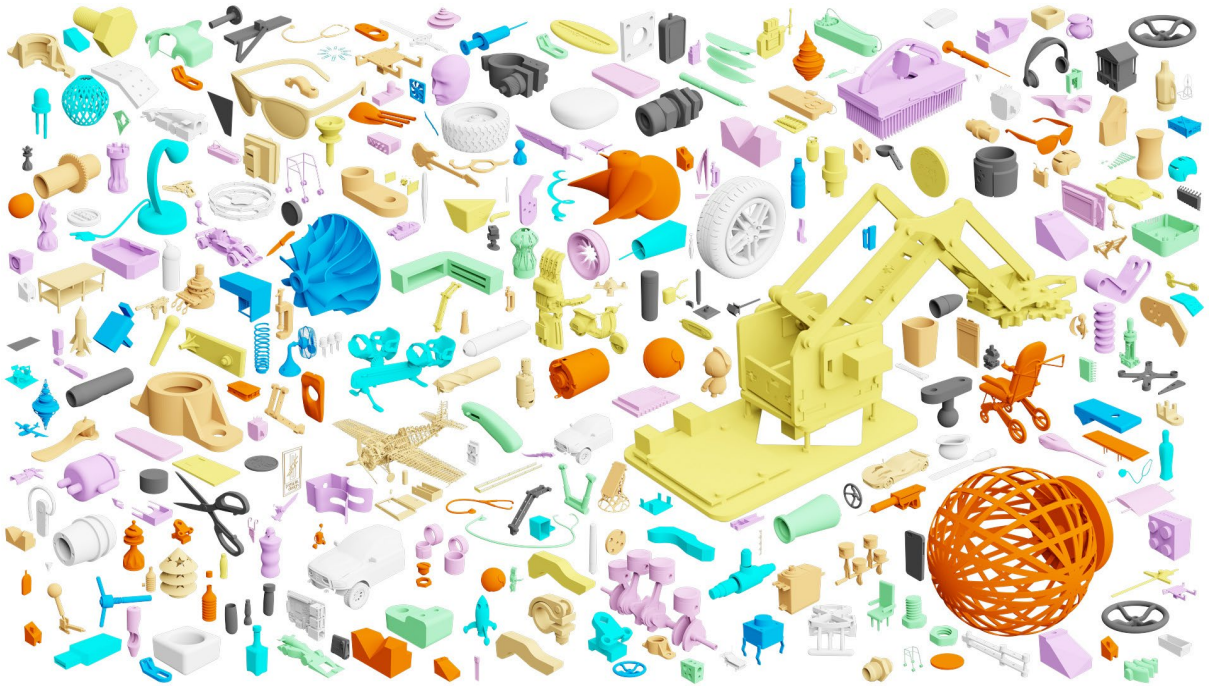
- with Explicitly Parametrized Curves and Surfaces

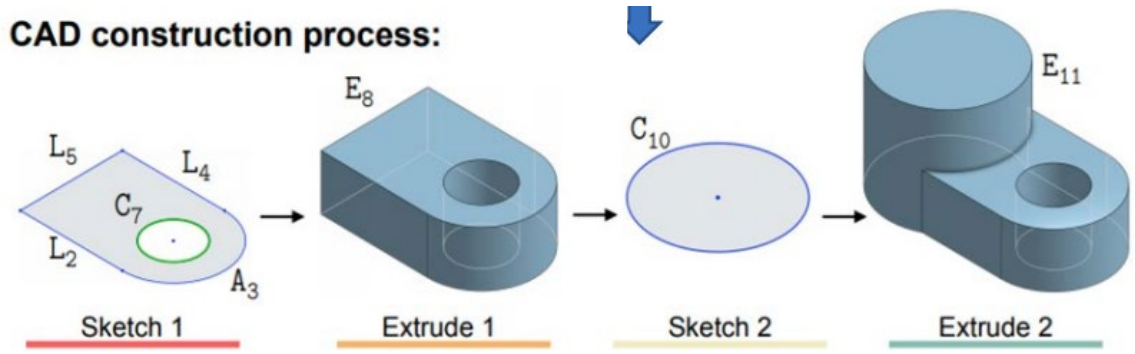
- Normal Estimation
- Surface Parametrization

https://deep-geometry.github.io/abc-dataset/

Sebastian et al., "**ABC: A Big CAD Model Dataset For Geometric Deep Learning**", CVPR 2019

AutoDesk Fusion 360 Gallery Dataset





- Synthetic, 3D CAD Models
- ~ 20K Designs

- Sketch + Extrude
- B-representation

- Reconstruction Dataset
- Segmentation Dataset
- Assembly Dataset

https://github.com/AutodeskAILab/Fusion360GalleryDataset

Willis et al., "**Fusion 360 Gallery: A Dataset and Environment for Programmatic CAD Construction from Human Design Sequences**", ACM Transactions on Graphics (TOG)

CO3D: Common Objects in 3D

CO3D

- Real-world 3D Shape Scans
- 1.5M Multi-view Images
- 19K Objects, 50 Categories

- Novel View Synthesis
- 3D Reconstruction

https://github.com/facebookresearch/co3d

Reizenstein et al., "**Common Objects in 3D: Large-Scale Learning and Evaluation of Real-life 3D Category Reconstruction**", ICCV 2021

3D-Front: Large-scale Synthetic 3D Scenes

- Synthetic, 3D CAD Models
- from Professional Designers

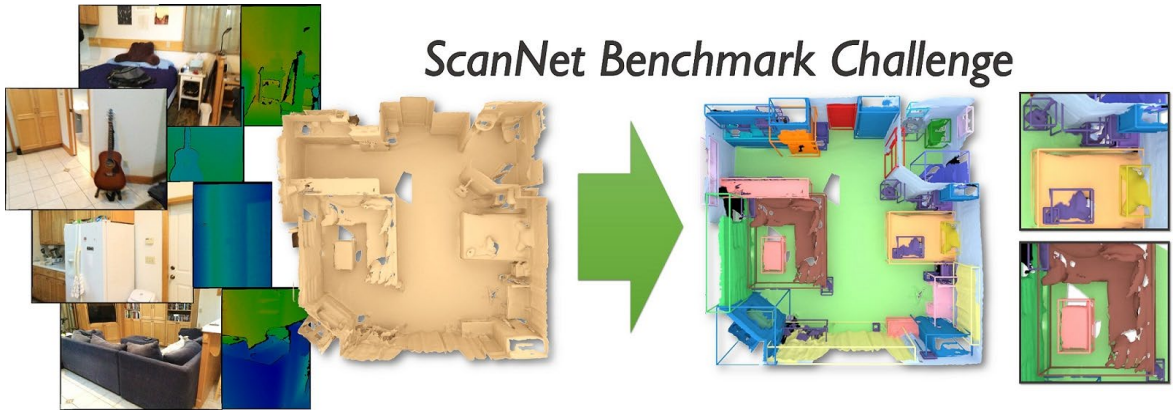
- 18,797 Rooms
- 7,302 Furniture Objects

- Indoor Scene Synthesis
- Texture Synthesis

https://tianchi.aliyun.com/specials/promotion/alibaba-3d-scene-dataset

Fu et al., "**3D-FRONT: 3D Furnished Rooms with layOuts and semaNTics**", ICCV 2021

ScanNet: Large-scale Real-world 3D Scene Scans



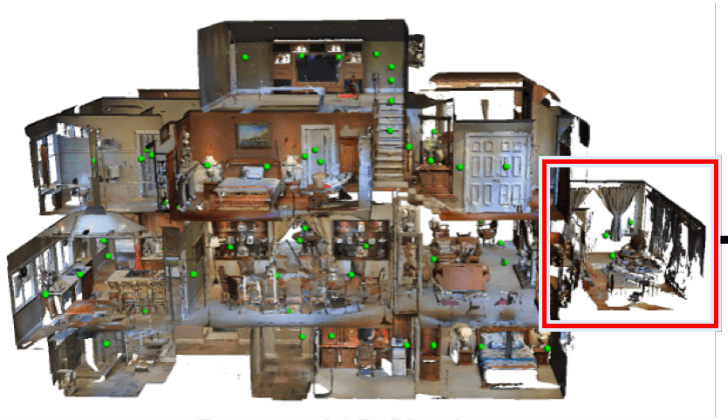
- Real 3D Scene Scans
- RGB-D Videos with 2.5M Views
- 1,500 3D Real Scene Scans

- Semantic/Instance Segmentation
- Object Detection/Classification
- Scene Completion/
  Reconstruction / Generation

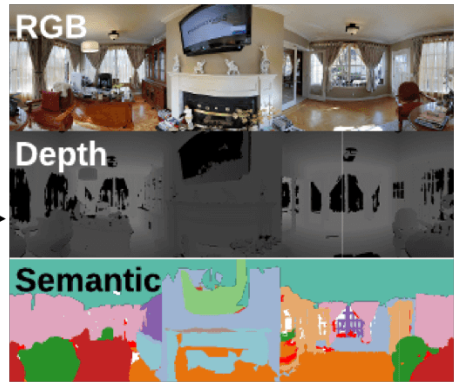
http://www.scan-net.org/

Dai et al., "**ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes**", CVPR 2017

Matterport3D and HM-3D: Larger and Largest



Textured 3D Mesh



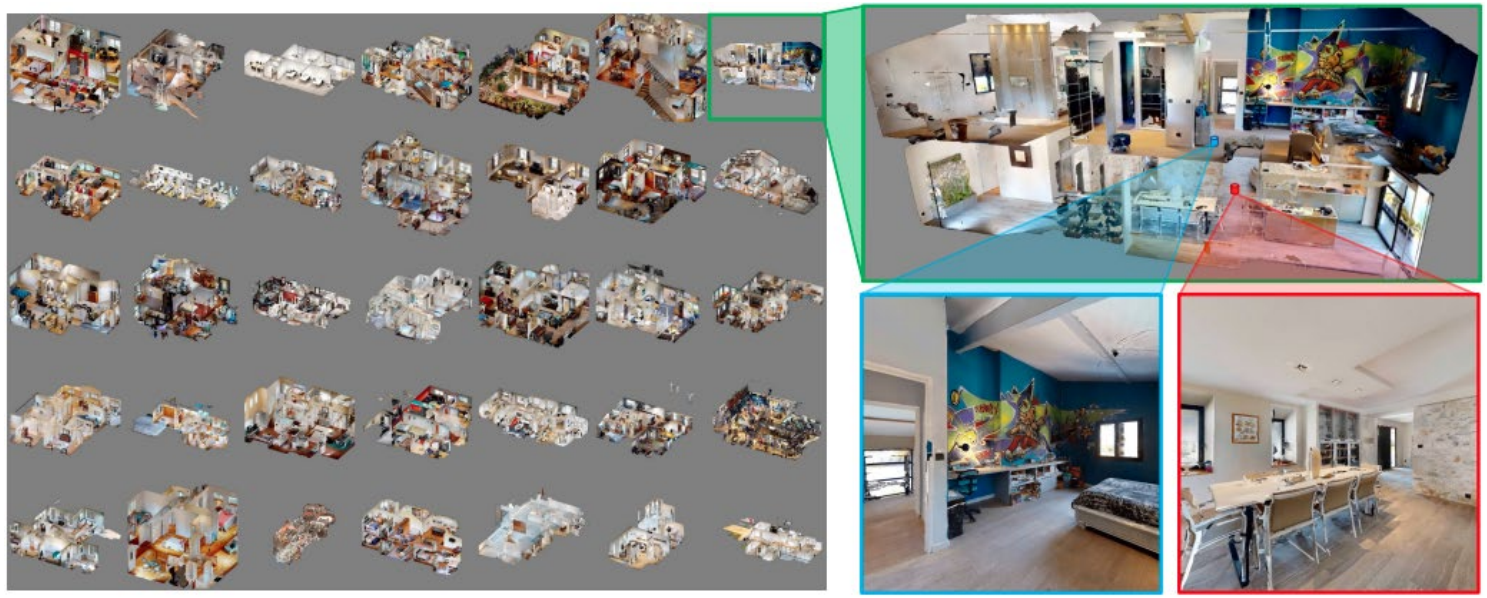
Panoramas

Object Instances

- 194,400 RGB-D Images
- 90 Building-scale Real Scans

https://niessner.github.io/Matterport/

Chang et al., "**Matterport3D: Learning from RGB-D Data in Indoor Environments**", 3DV 2017



- 1,000 Building-scale Real Scans (the Largest until now)

https://aihabitat.org/datasets/hm3d/

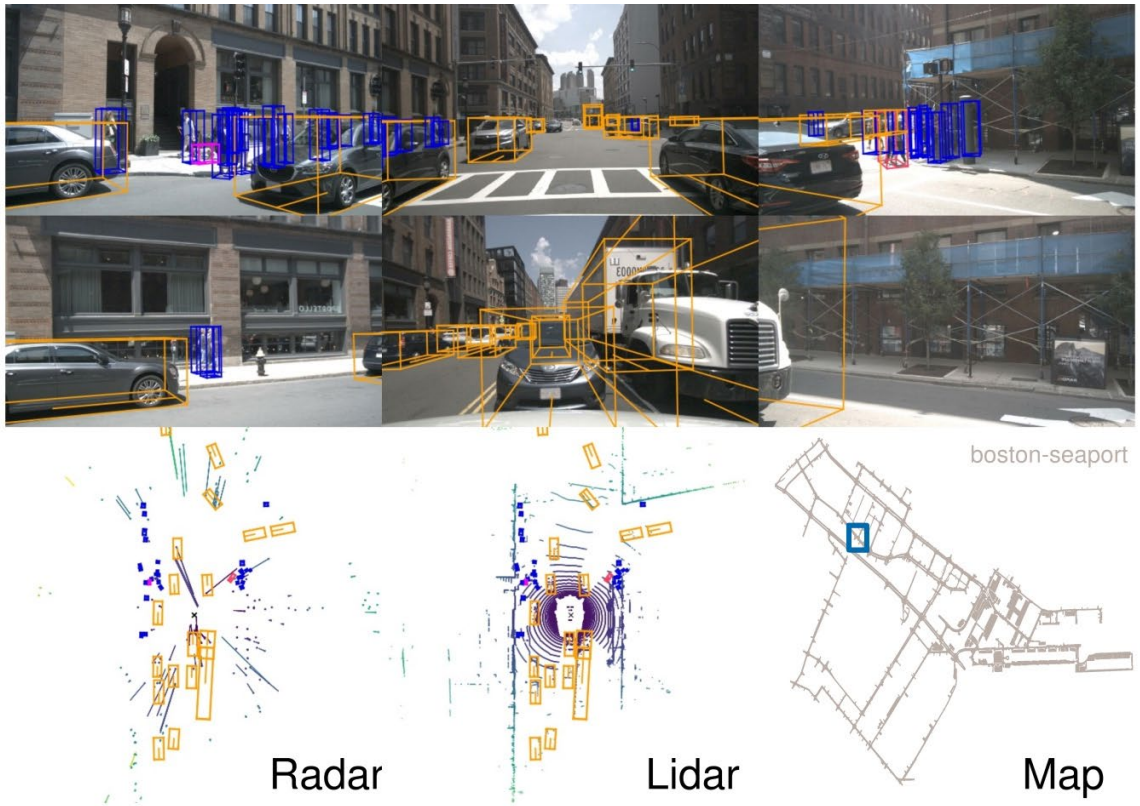
Ramakrishnan et al., "**Habitat-Matterport 3D Dataset HM3D: 1000 Large-scale 3D Environments for Embodied AI**", NeurIPS (dataset) 2021

Kitti and nuScenes: Outdoor Road Scenes for AV

- 389 Images
- 200K Object Annotations

http://www.cvlibs.net/datasets/kitti/

Geiger et al., "**Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite**", CVPR 2012



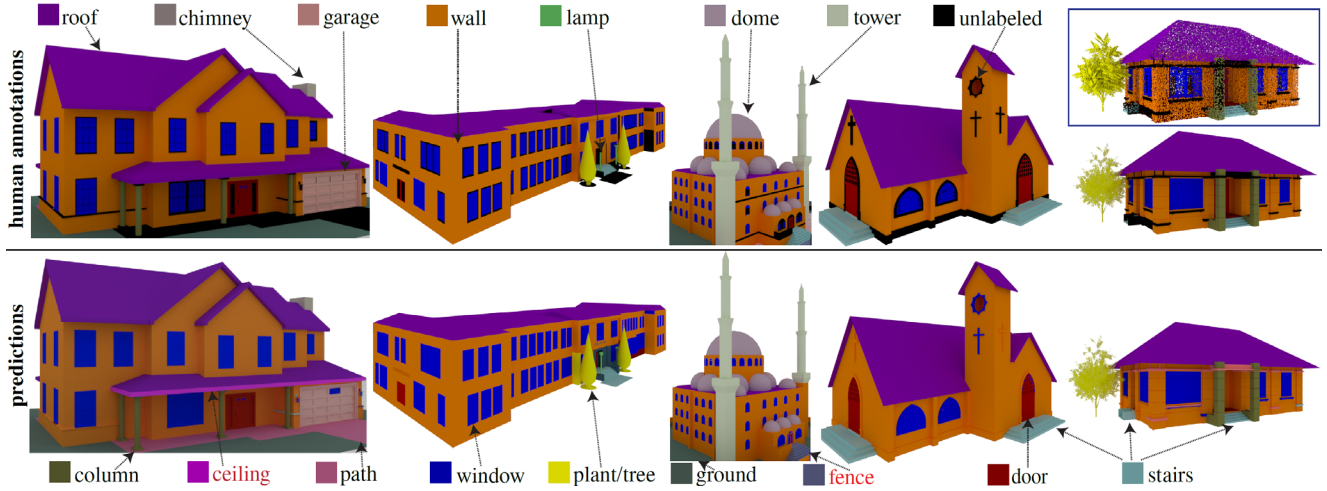
"Ped with pet, bicycle, car makes a u-turn, lane change, peds crossing crosswalk"

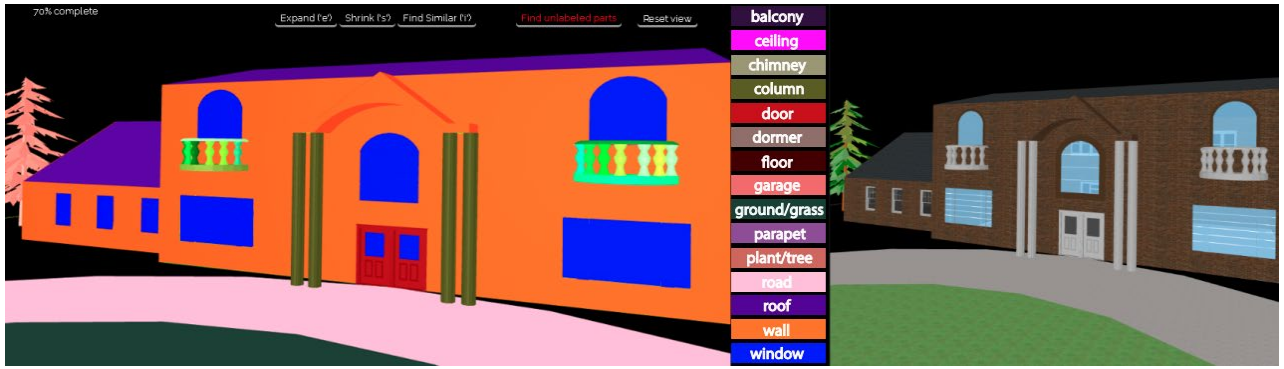
- 1,000 Scenes
- 23 Classes and 8 Attributes

https://www.nuscenes.org/

Caesar et al., "n**uScenes: A multimodal dataset for autonomous driving**", CVPR 2020

BuildingNet: Large-scale 3D Building Models





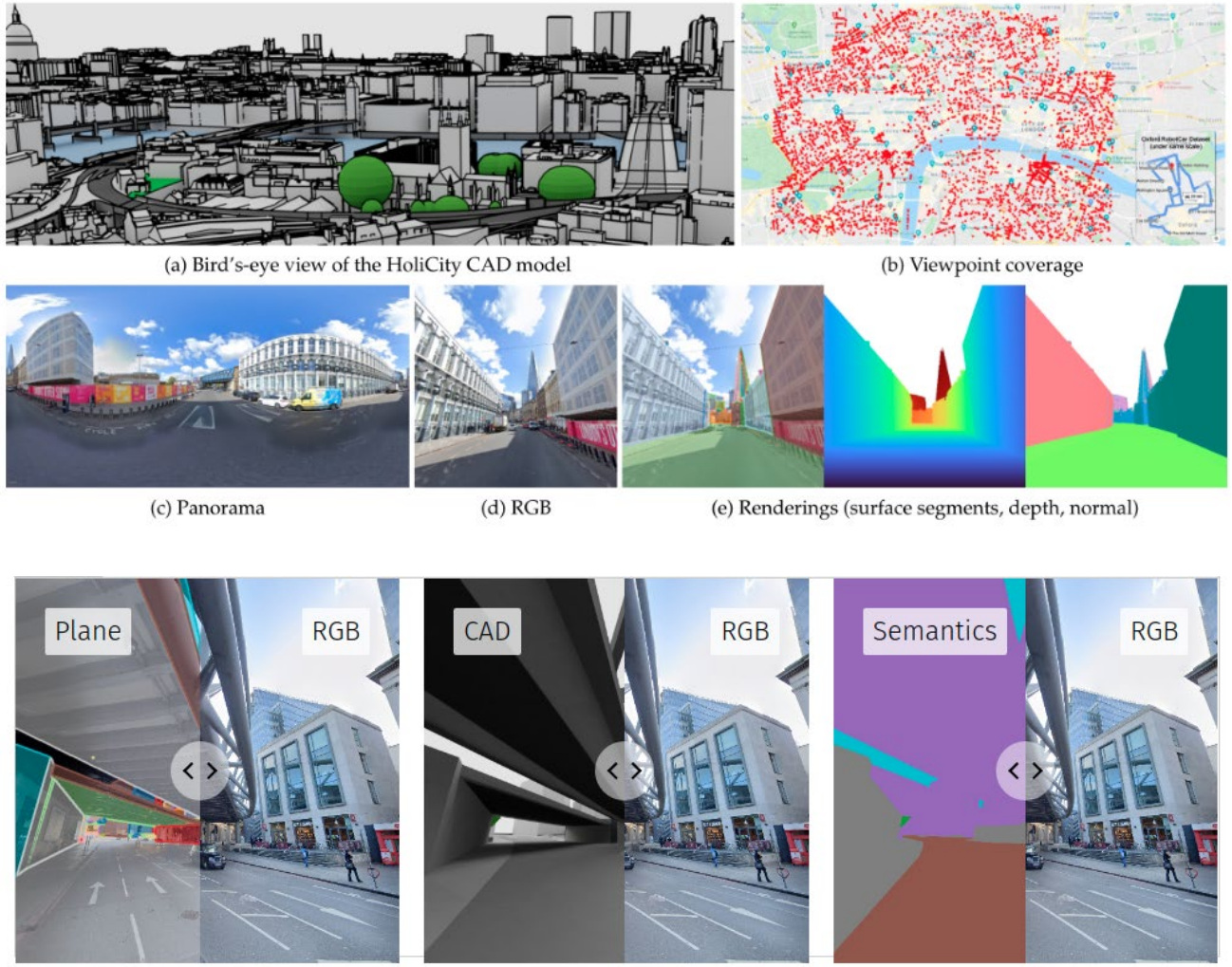
- Synthetic, 3D CAD Models
- 292K Parts, 2K Buildings
- E.g. houses, churches, skyscrapers, town halls, libraries, and castles

- Part Annotations
  - e.g. roof, chimney, wall, lamp
- Edge Annotations
  - e.g. proximity, support, containment

https://buildingnet.org/

Selvaraju et al., "**BuildingNet: Learning to Label 3D Buildings**", ICCV 2021

HoliCity: A City-scale 3D Dataset



- Real-world Scenes in London
- Aligned with 3D CAD Models
- 13312 x 6656 m^2

- 3D Structural Annotations
  - e.g. planes, corners, lines
- Semantic Segmentation

- Support the study of city-scale 3D tasks

https://holicity.io/

Zhou et al., "**HoliCity: A City-Scale Data Platform for Learning Holistic 3D Structures**"

SensatUrban: An Urban-Scale Dataset

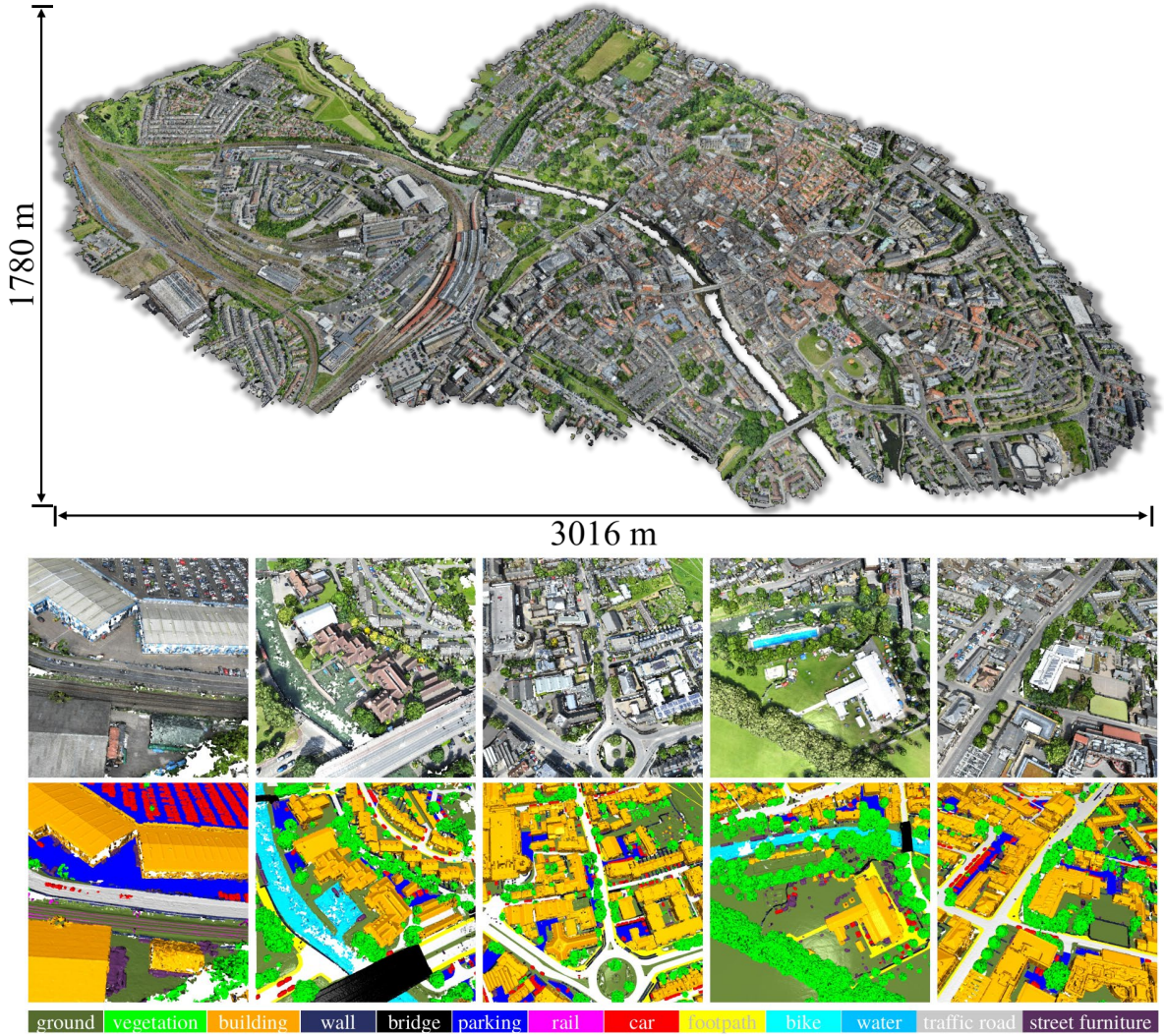


Figure 3: Examples of our SensatUrban dataset. Different semantic classes are labeled by different colors.

https://github.com/QingyongHu/SensatUrban

- 3D Real-world Scans
- 6 km ^ 2 City Landscape
- 13 Semantic Classes
  - e.g. ground, vegetation, car
- Support the study of urban-scale 3D tasks

Hu and Yang et al., "**Towards Semantic Segmentation of Urban-Scale 3D Point Clouds: A Dataset, Benchmarks and Challenges**", CVPR 2021

Summary

- From semantic networks to visual or geometric data networks
  - WordNet, ImageNet, and ShapeNet
- Approaches for annotation acquisition
  - Difficulty of 3D labels
- From vertical networks to horizontal networks
  - Annotation transportation in ShapeNet

Large and high-quality data sets are essential for both training and testing machine learning algorithms

That’s All