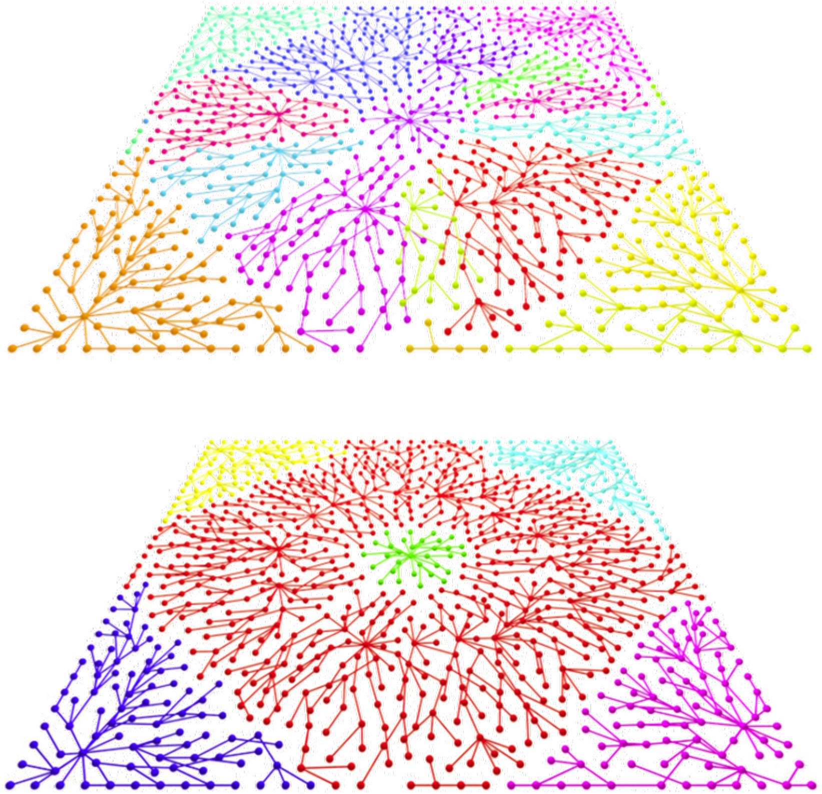
CS233, CME251: Geometric and Topological Data Analysis

Leonidas Guibas
Computer Science Department
Stanford University

Lecture 2
8 April 2020



Leonidas Guibas Laboratory
Geometric Computing

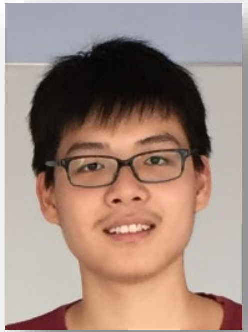
The Principals

- Leonidas (Leo) Guibas (CS & EE)
  - Instructor
- Panos Achlioptas (CS)
  - Course Assistant
- Yueqi Duan (CS)
  - Course Assistant
- Kaichun Mo (CS)
  - Course Assistant
- Samir Chowdhury (Neuropsychiatry)
  - Guest lecturer
- Carrie Petersen (CS)
  - Admin


Panos


Leo


Yueqi


Kaichun

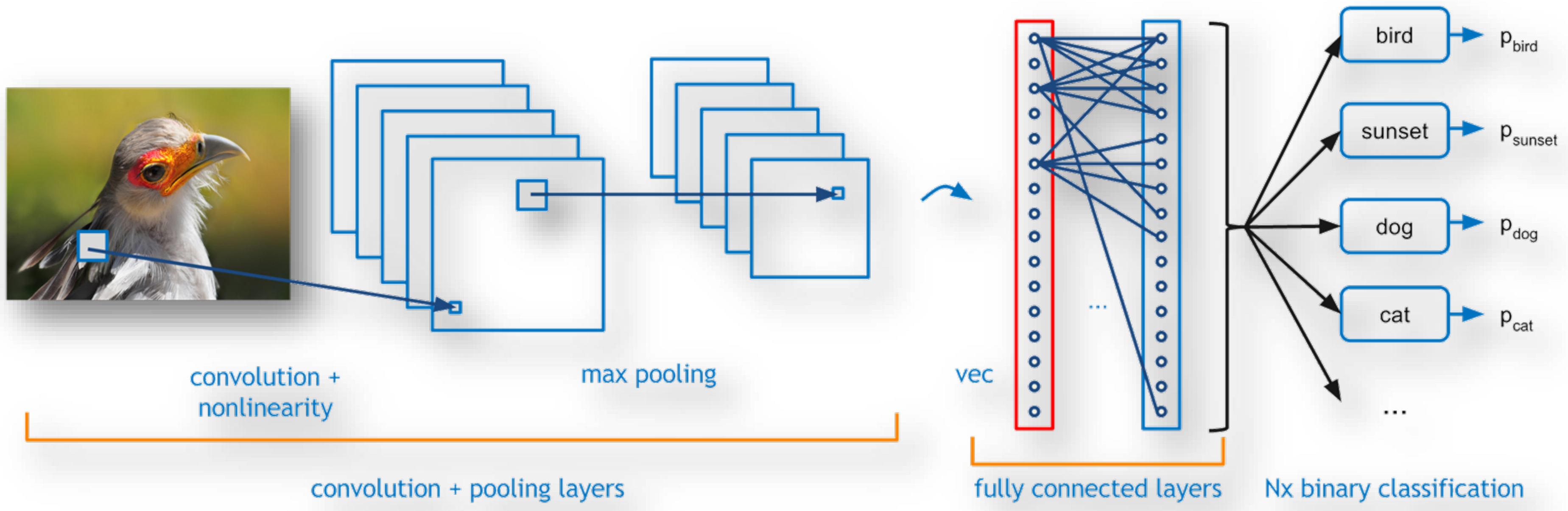

Samir


Carrie

From Last Time:
Deep Learning
Supervised Methods

Deep Learning Networks

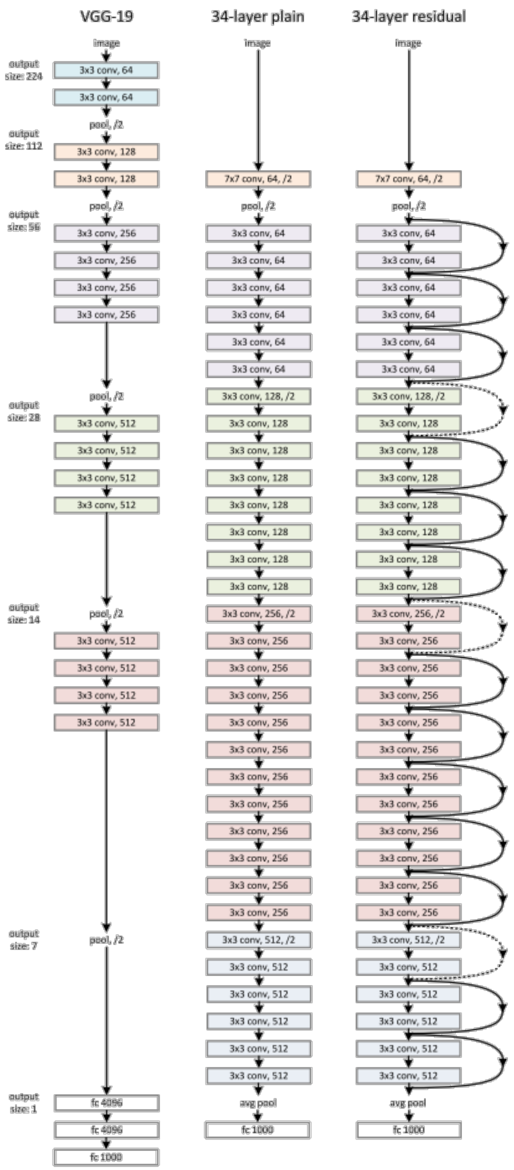
- Feed forward networks



Classification

Success Made Possible By ….

[He et al., 2015]
Novel deep architectures



Plenty of annotated data

Lots of computing power

Semantic Annotations for Visual/Geometric DataSets

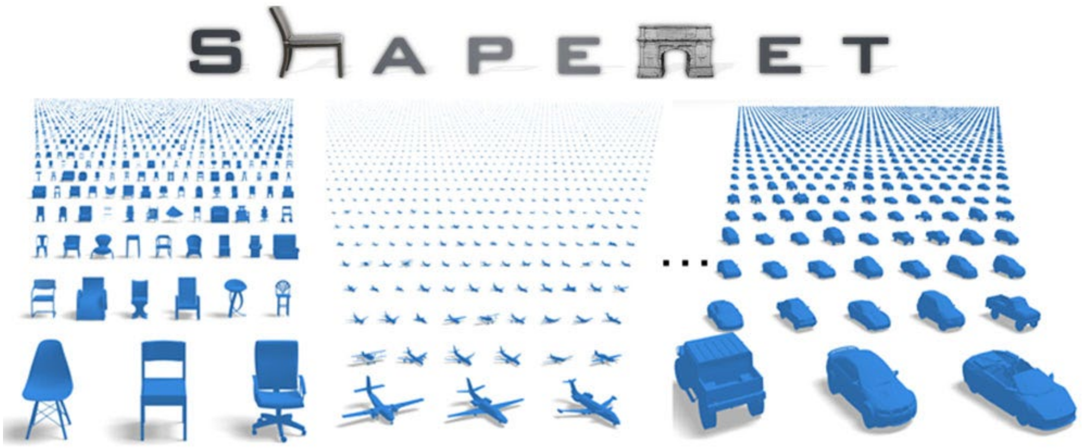
Agenda

- From semantic networks to visual or geometric data networks
  - WordNet, ImageNet and ShapeNet
- Approaches for annotation acquisition
- 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

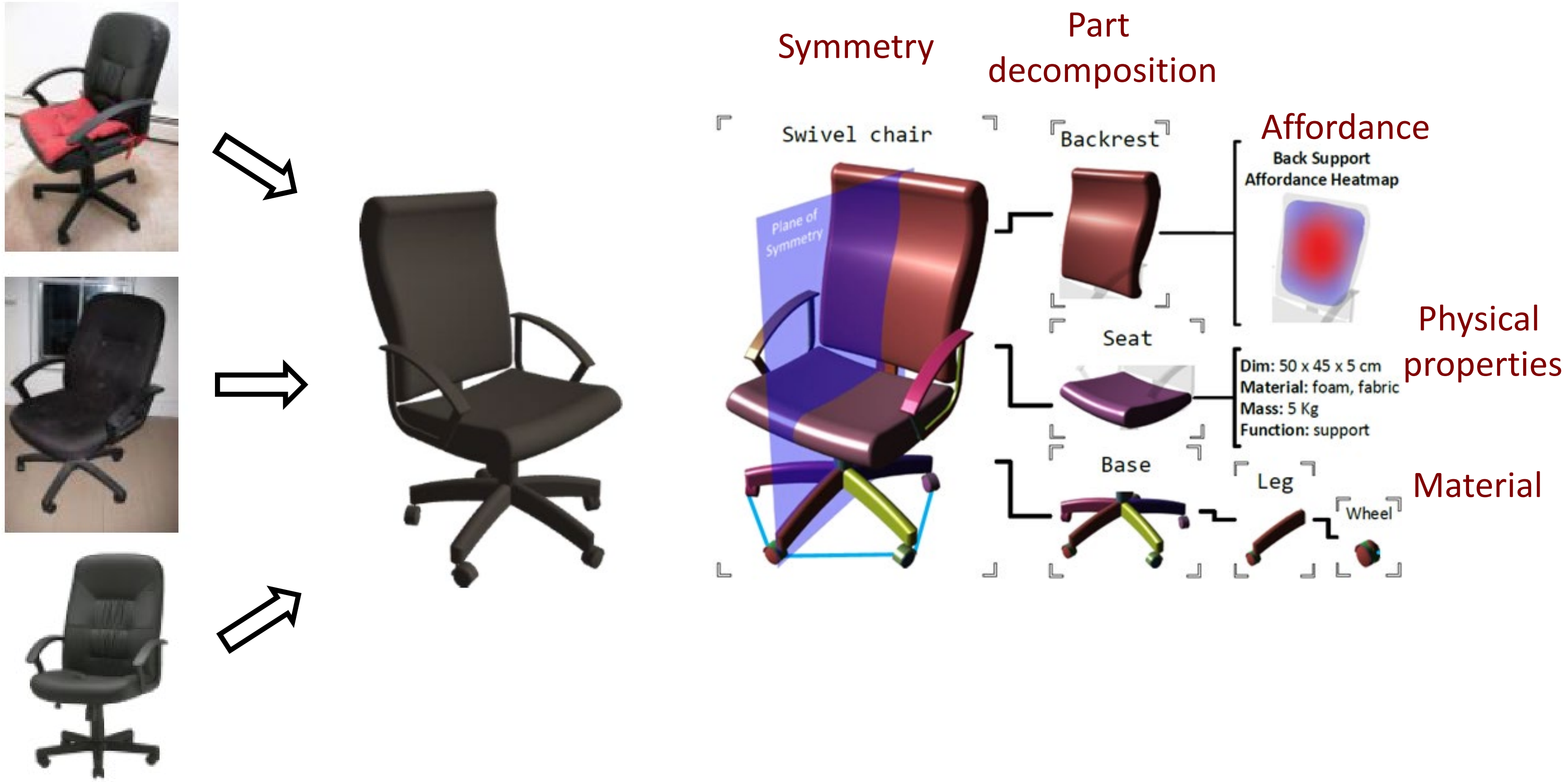
Goal of this Lecture

- Explain how big visual datasets including ImageNet and ShapeNet are organized



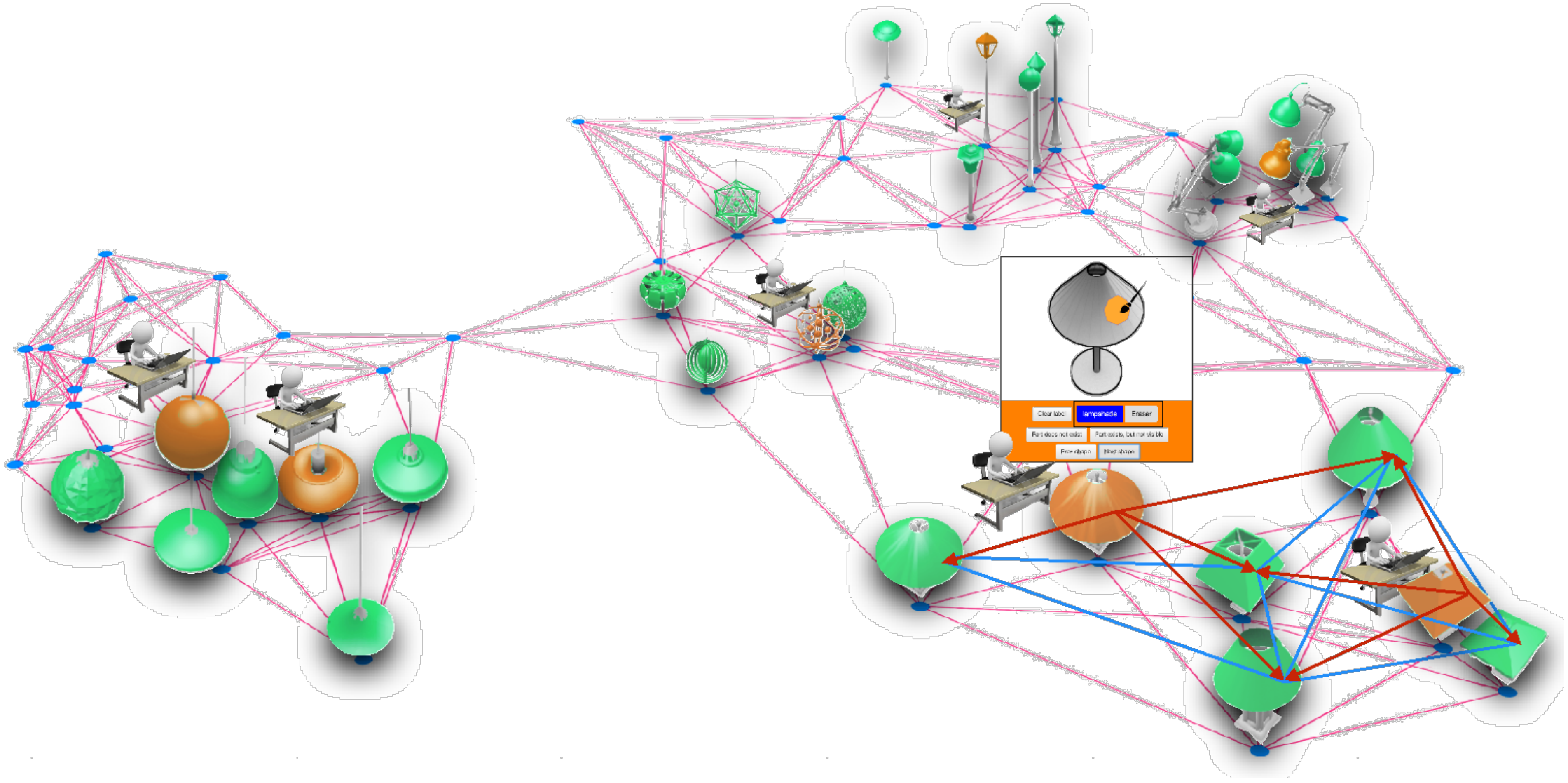
Goal of this Lecture

- Explain how ShapeNet are annotated



Goal of this Lecture

- Show examples of label transportation in a network



Semantic Networks: Storing Knowledge about the World

Semantic Networks

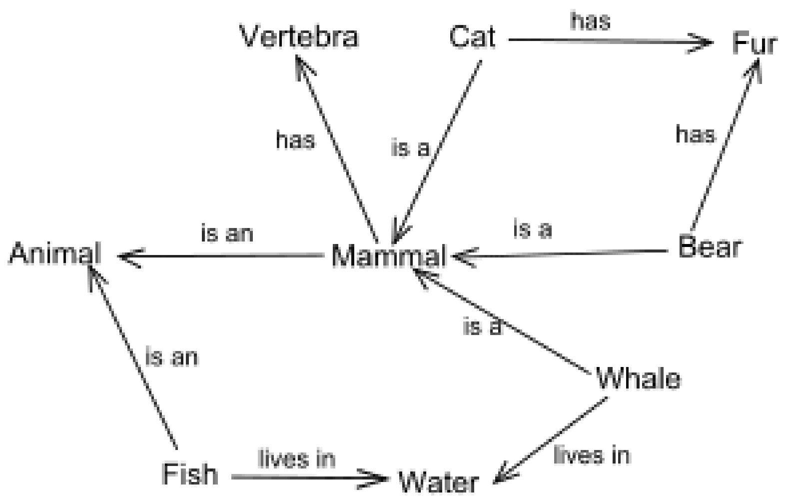
- Also known as **frame networks**
- Encode semantic relations between concepts
- Often used as a form of knowledge representation
- A directed or undirected graph consisting of vertices, which represent concepts, and edges which represent concept relations

Example of a Semantic Net

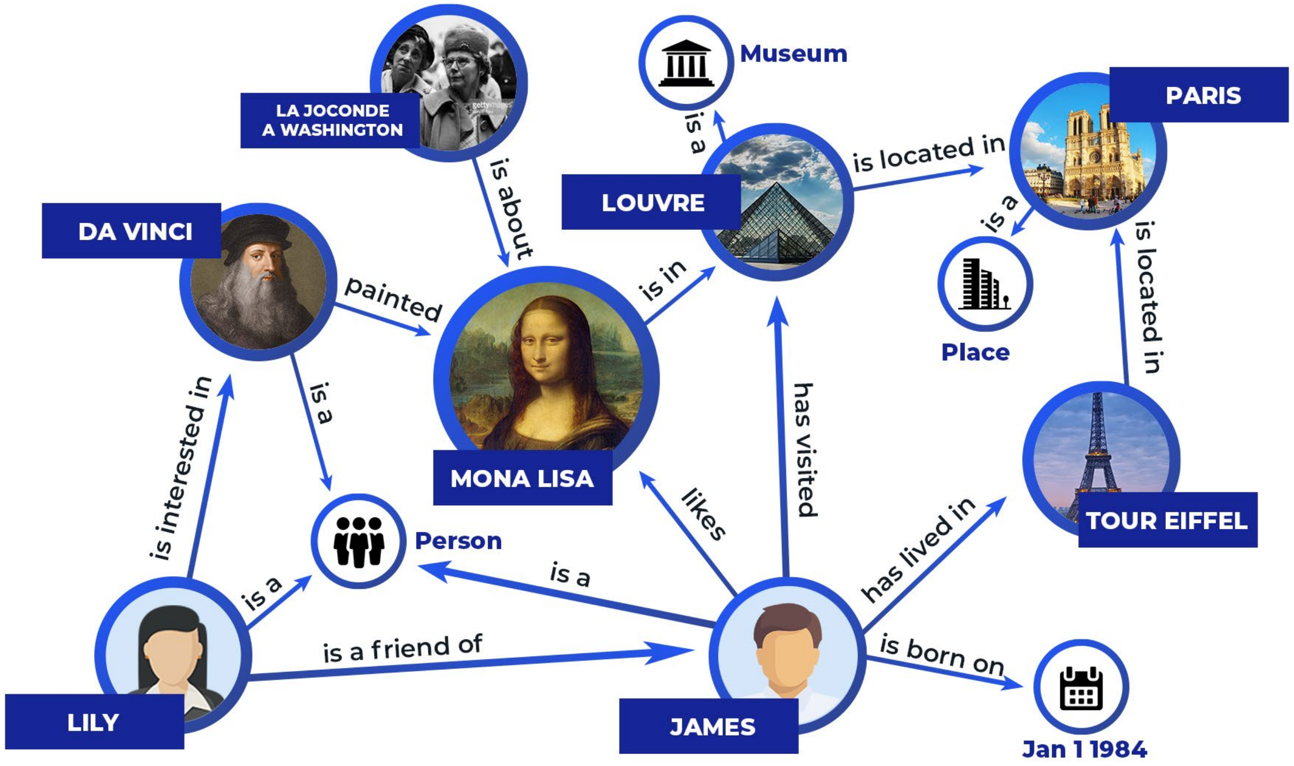
Semantic Net in Lisp

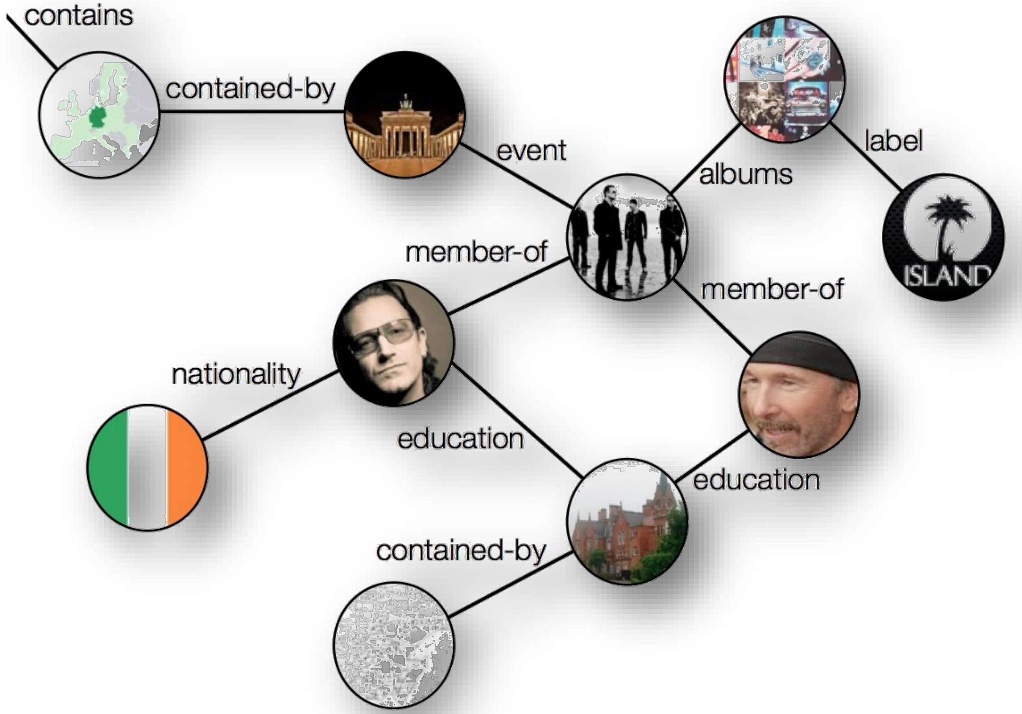
```
(defun *database* ()
'((canary  (is-a bird)
               (color yellow)
               (size small))
   (penguin (is-a bird)
                  (movement swim))
   (bird    (is-a vertebrate)
               (has-part wings)
               (reproduction egg-laying))))
```

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
                   => ...
```

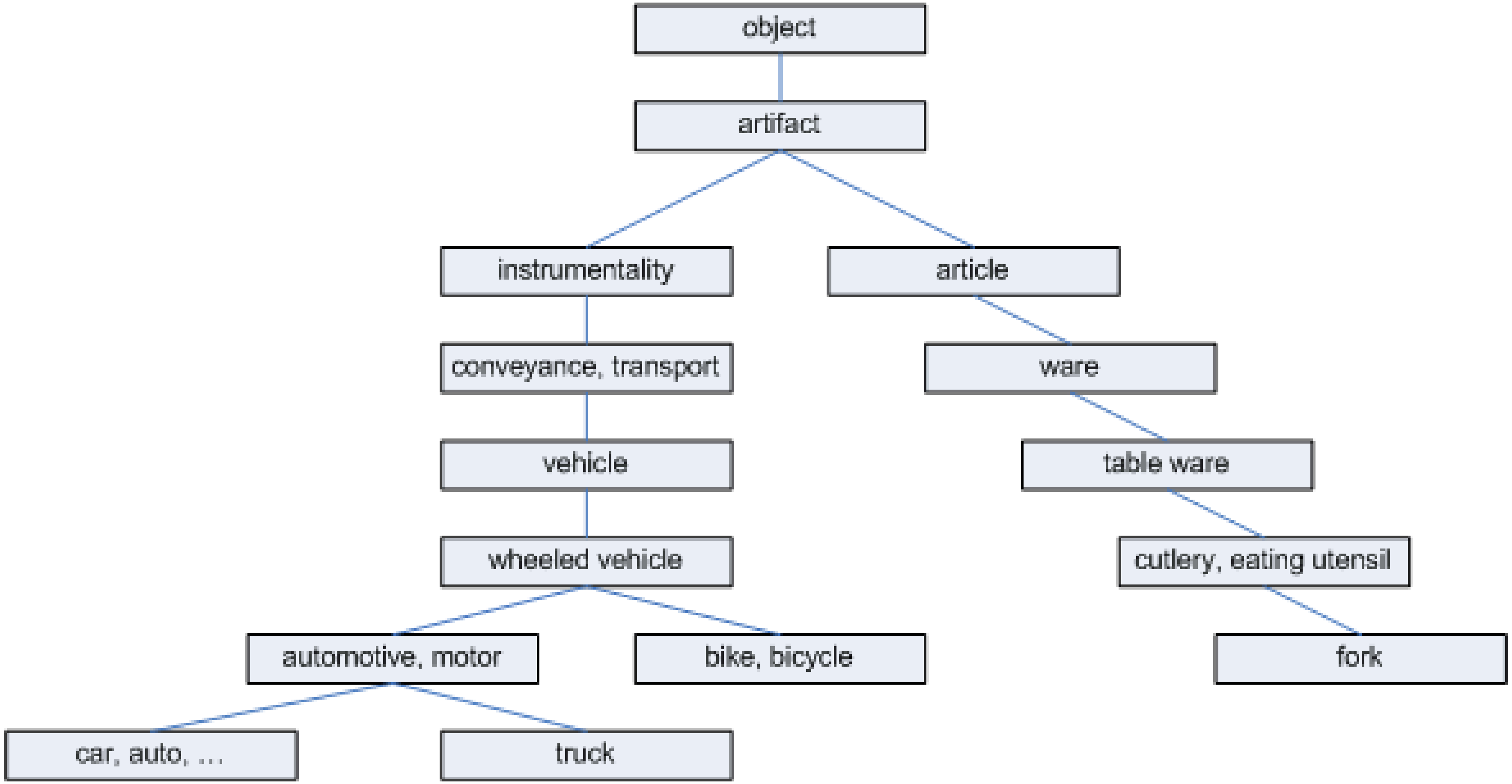
G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990.
WordNet: An online lexical database. Int. J. Lexicograph.

WordNet

- Important relations between synsets (nouns):

Relation	Definition	Example
Hypernym	From concepts to superordinates	$\text{water}^1 \rightarrow$ liquid
Hyponym	From concepts to subtypes	$\text{water}^1 \rightarrow$ seawater
Has-Part	From groups to their members	$\text{water}^1 \rightarrow$ oxygen
Part-of	From members to their groups	$\text{water}^1 \rightarrow$ ice
Antonym	Opposites	leader $\rightarrow$ 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 ch
    - 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

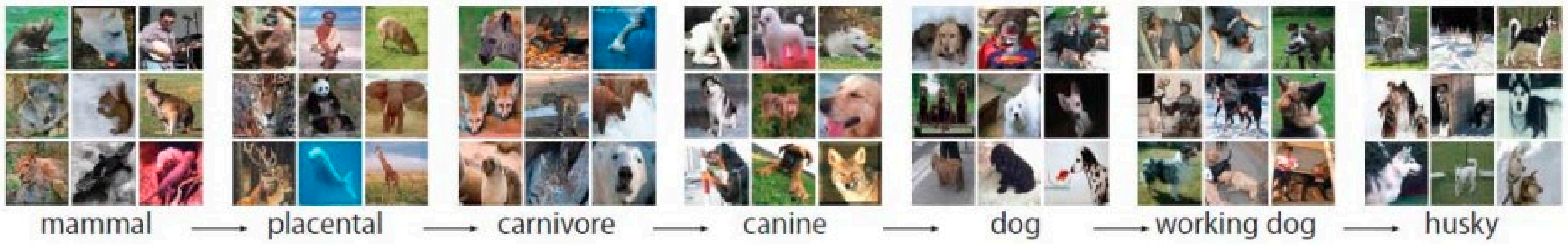
- S: (n) chair (a seat for one person, with a support for the back)

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

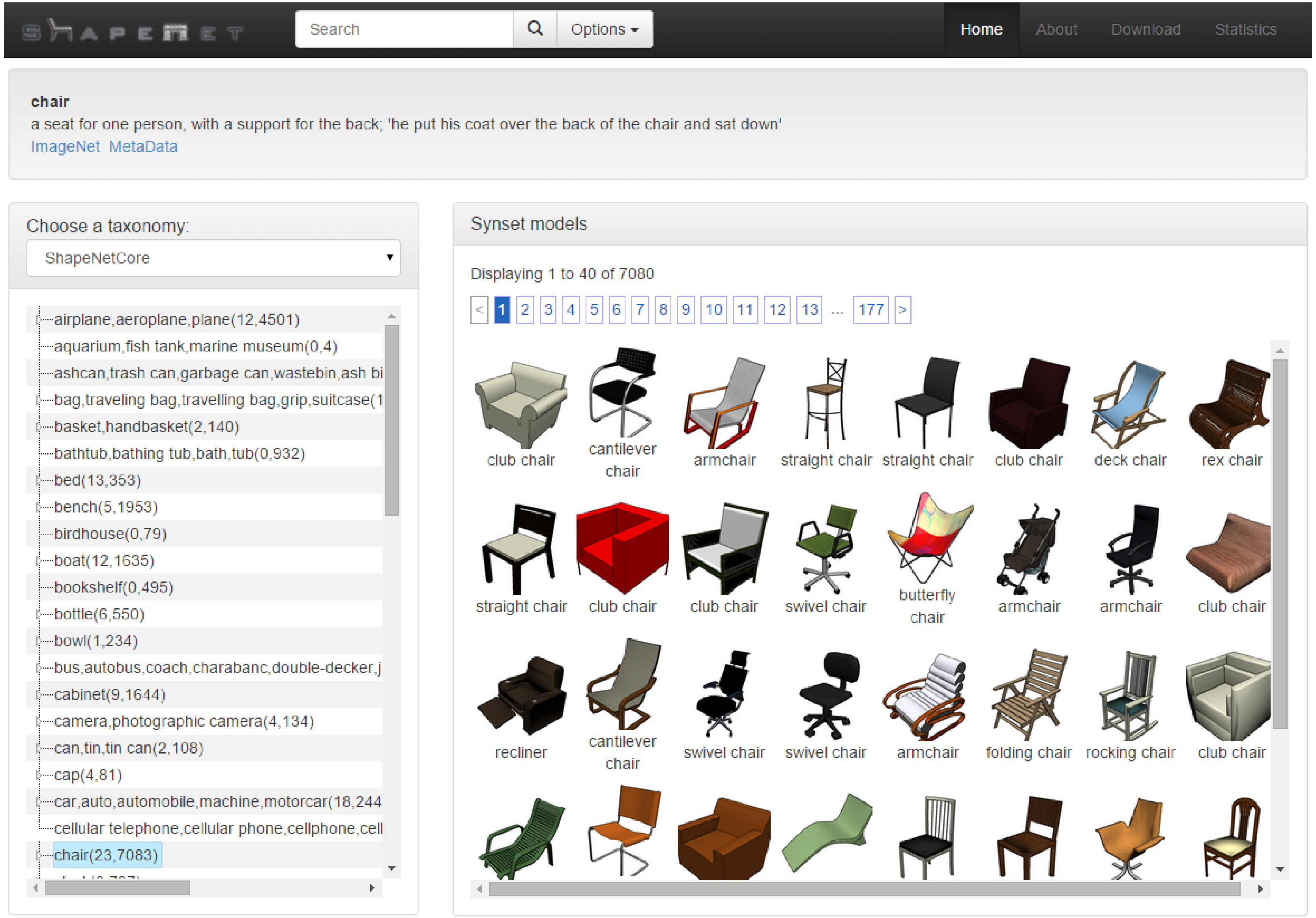
- Taxonomy (with WordNet backbone)



- 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))

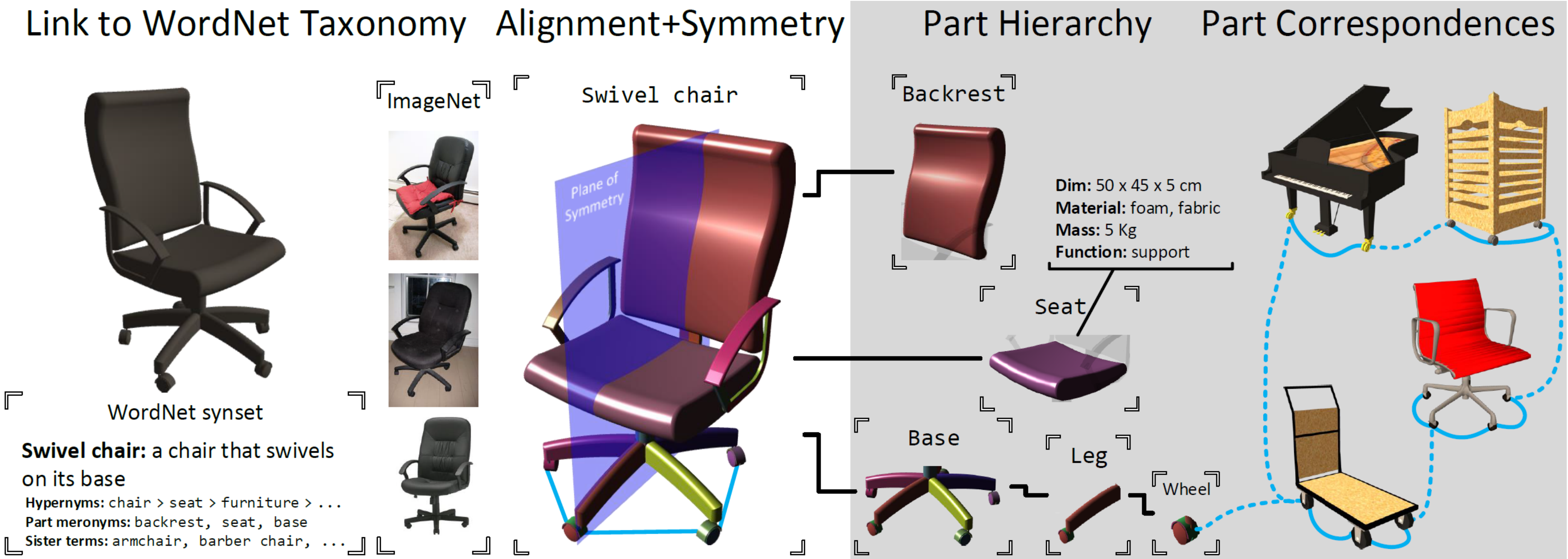
Slide Credit: Fei-Fei Li, Jia Deng

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

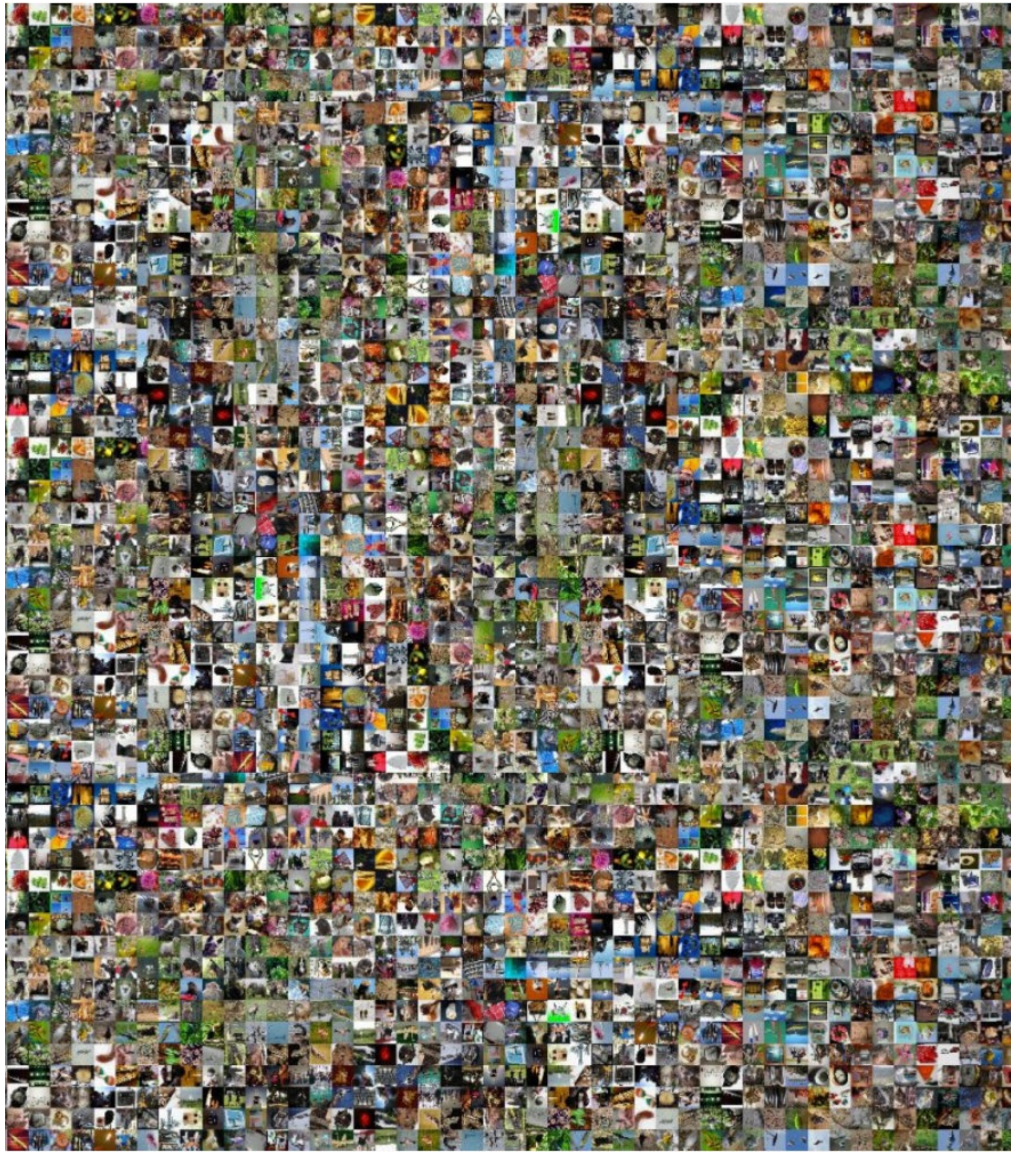
Individually Illustrated WordNet Nodes

jacket: a short coat

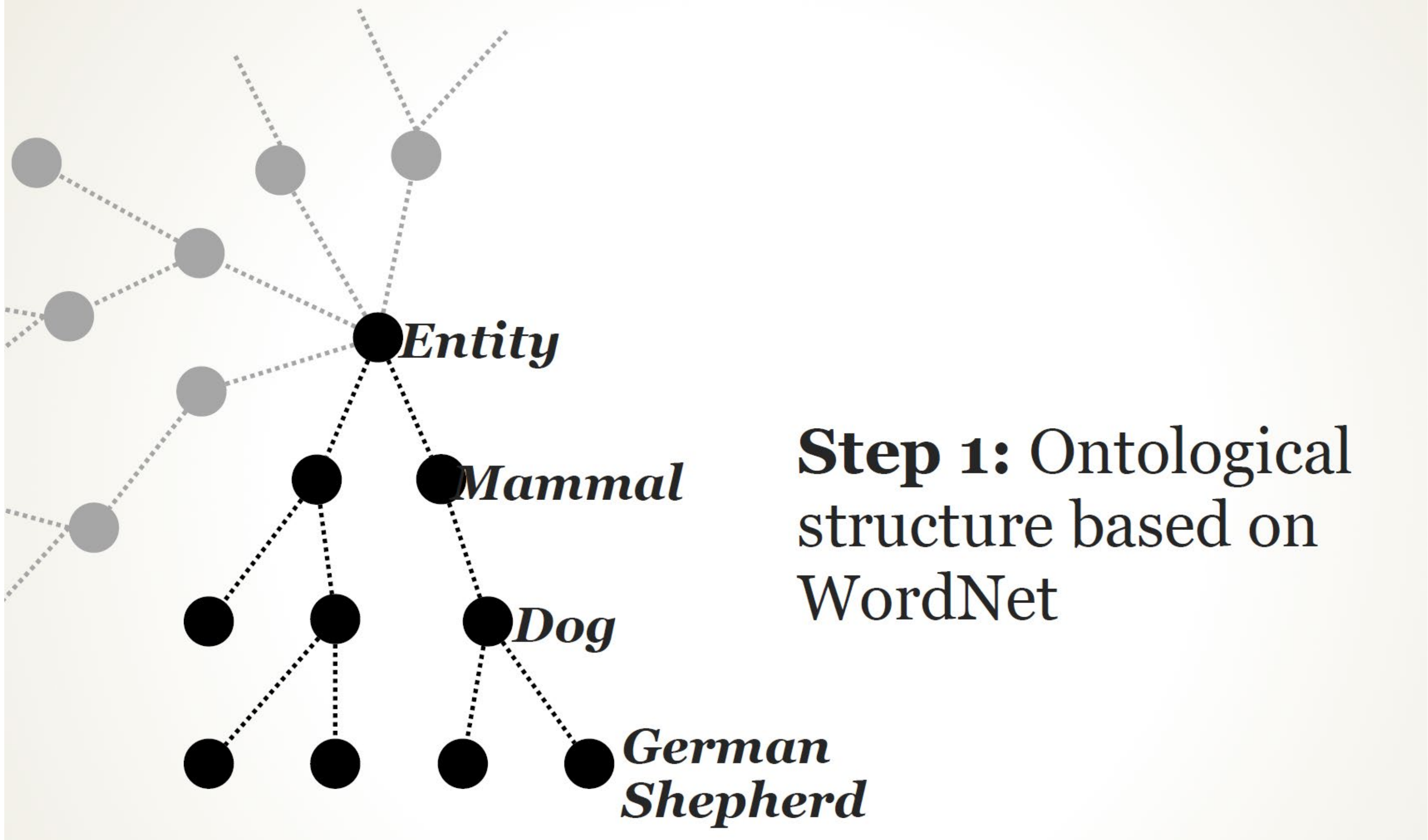
German shepherd: breed of large shepherd dogs used in police work and as a guide for the blind.

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

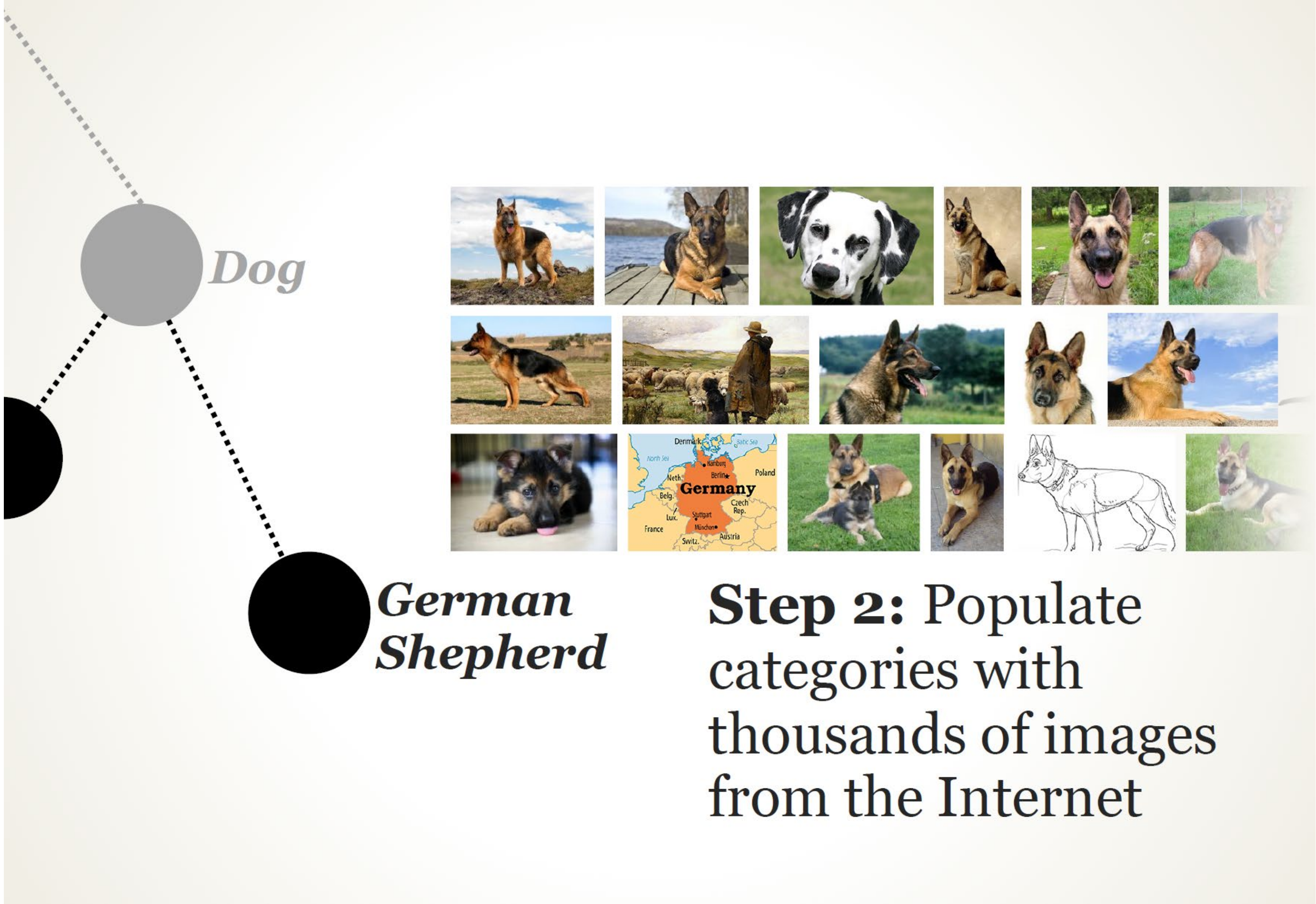
mountain: a land mass that projects well above its surroundings; higher than a hill.



WordNet Ontology

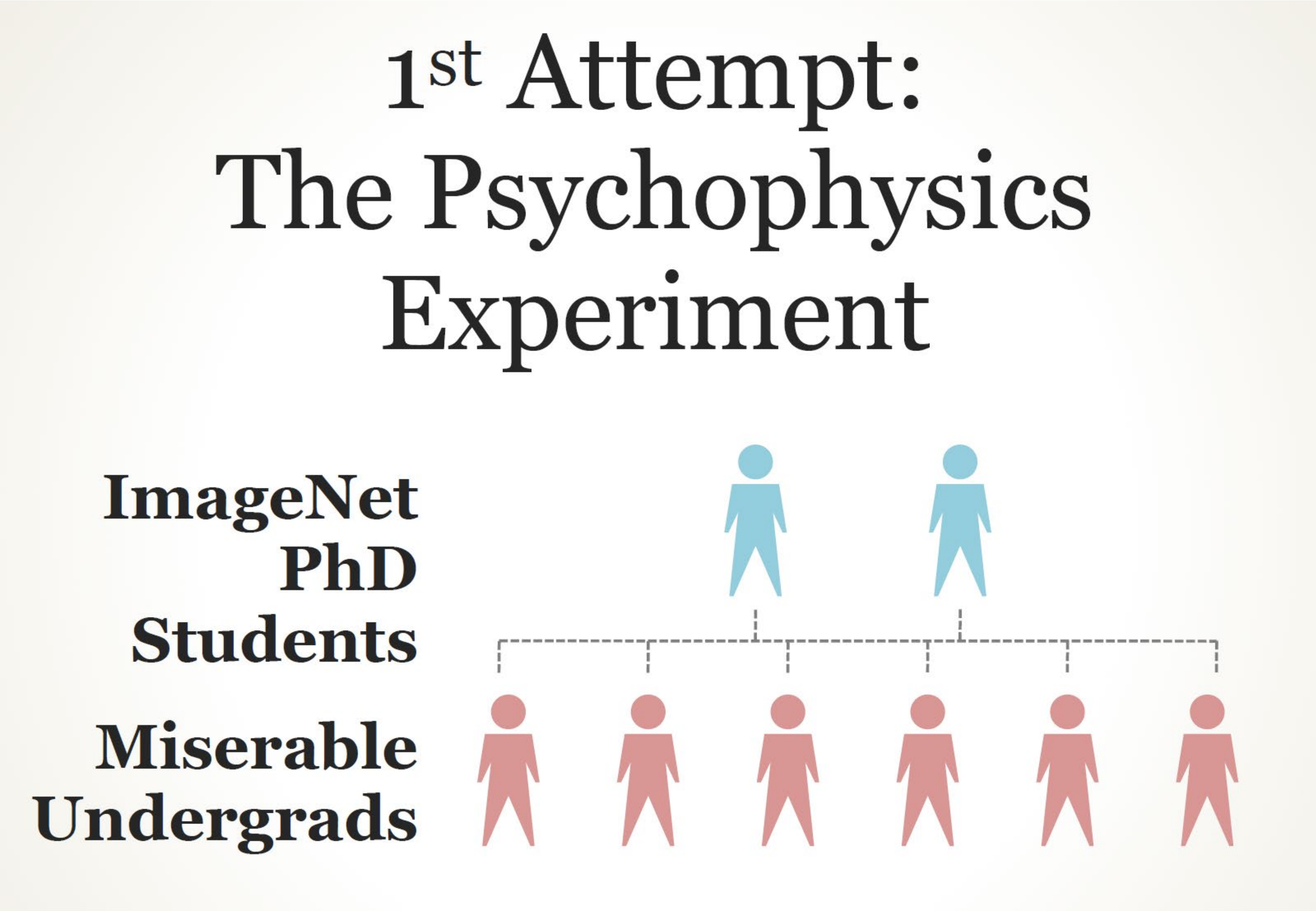


“Illustrating” WordNet



Cleaning Up the Results

Cleaning Up …


1st Attempt:
The Psychophysics
Experiment
ImageNet
PhD
Students
Miserable
Undergrads

Cleaning Up …

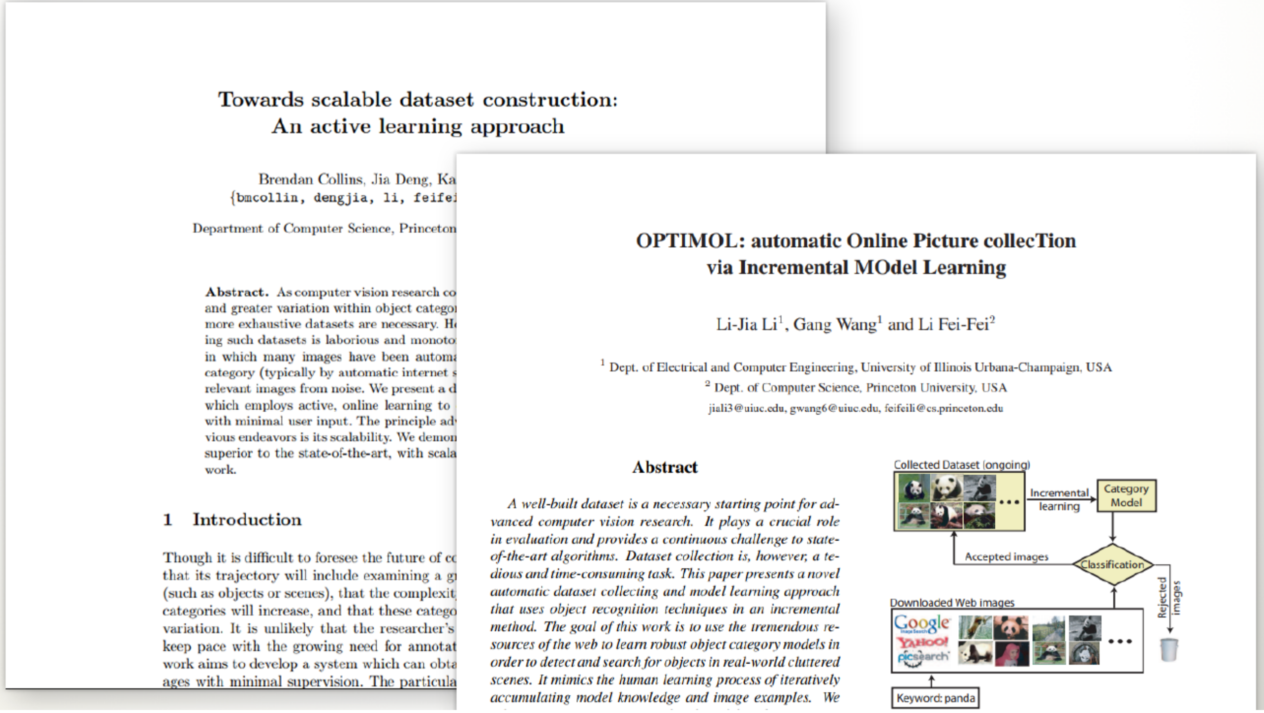
1st Attempt: The Psychophysics Experiment

- # 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, feife

Department of Computer Science, Princeton

Abstract. As computer vision research co
and greater variation within object catego
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

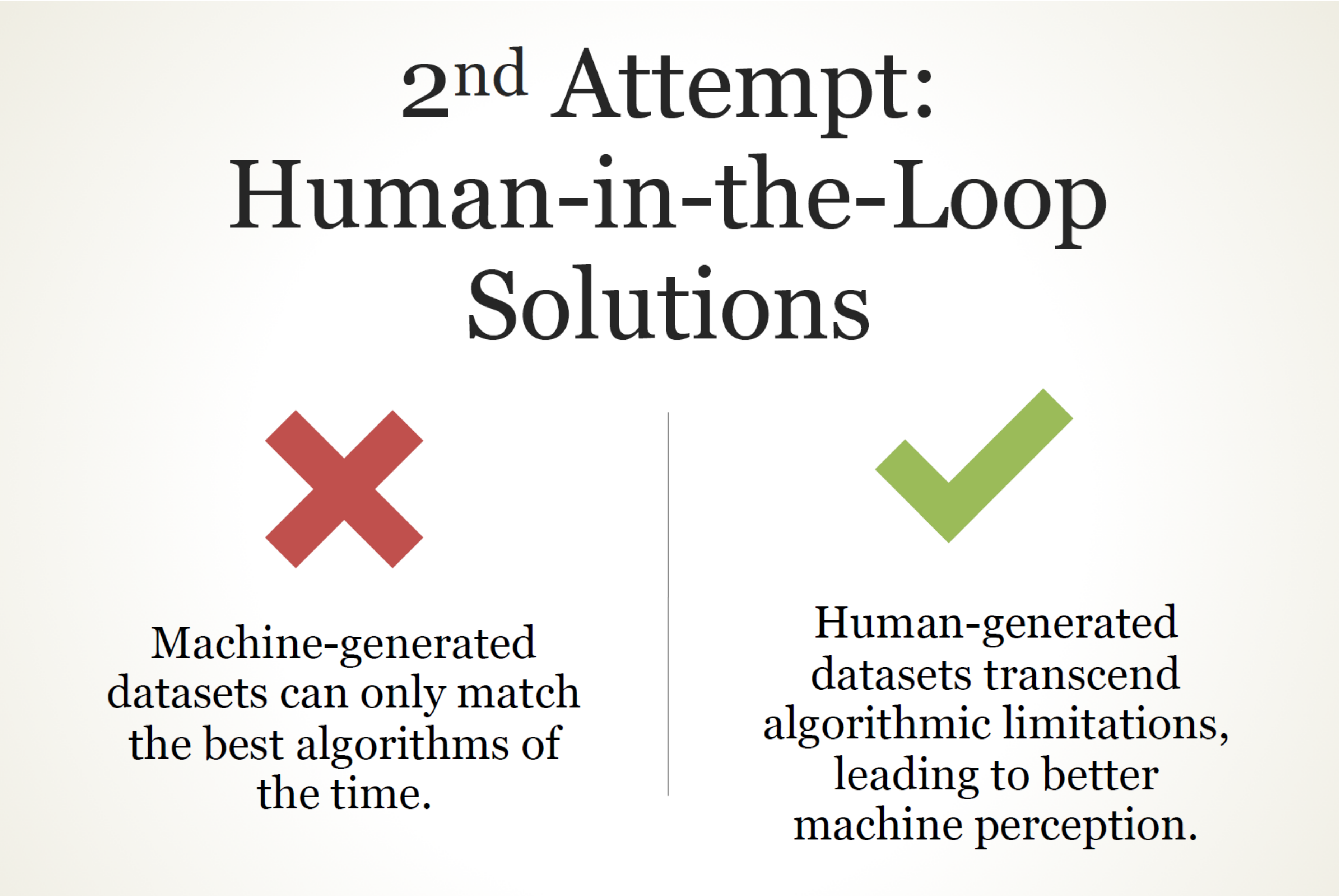
Li-Jia Li[1], Gang Wang[1] and Li Fei-Fei[2]

[1] Dept. of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, USA
[2] Dept. of Computer Science, Princeton University, USA
jiali3@uiuc.edu, gwang6@uiuc.edu, feifeili@cs.princeton.edu

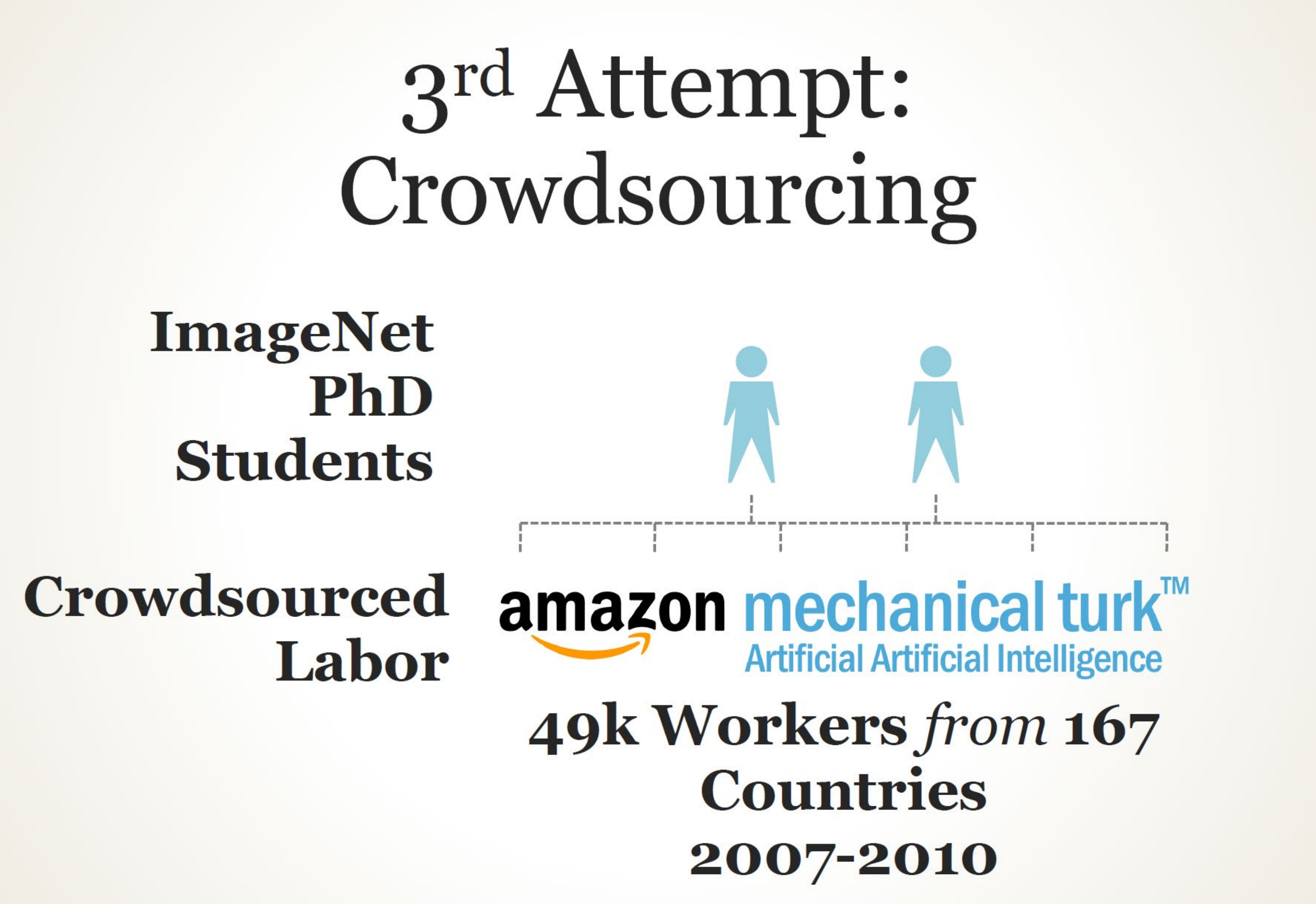
Abstract

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

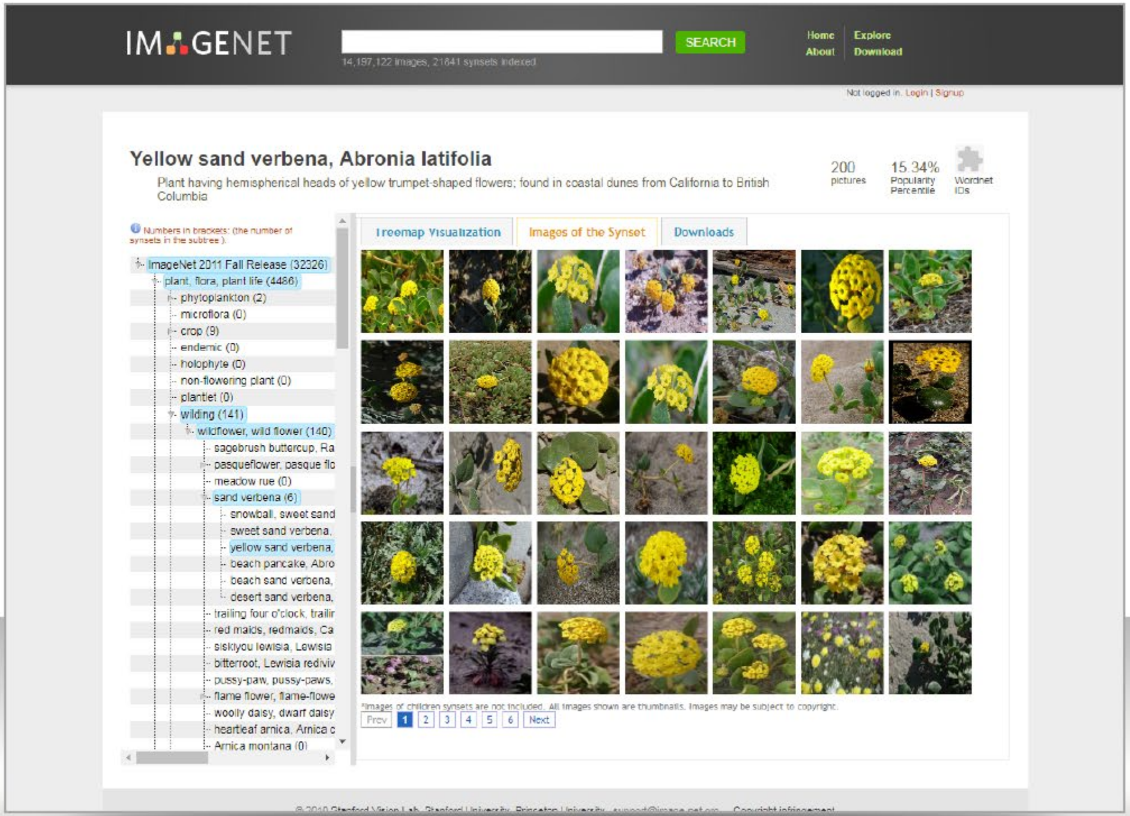
Classify and Collect



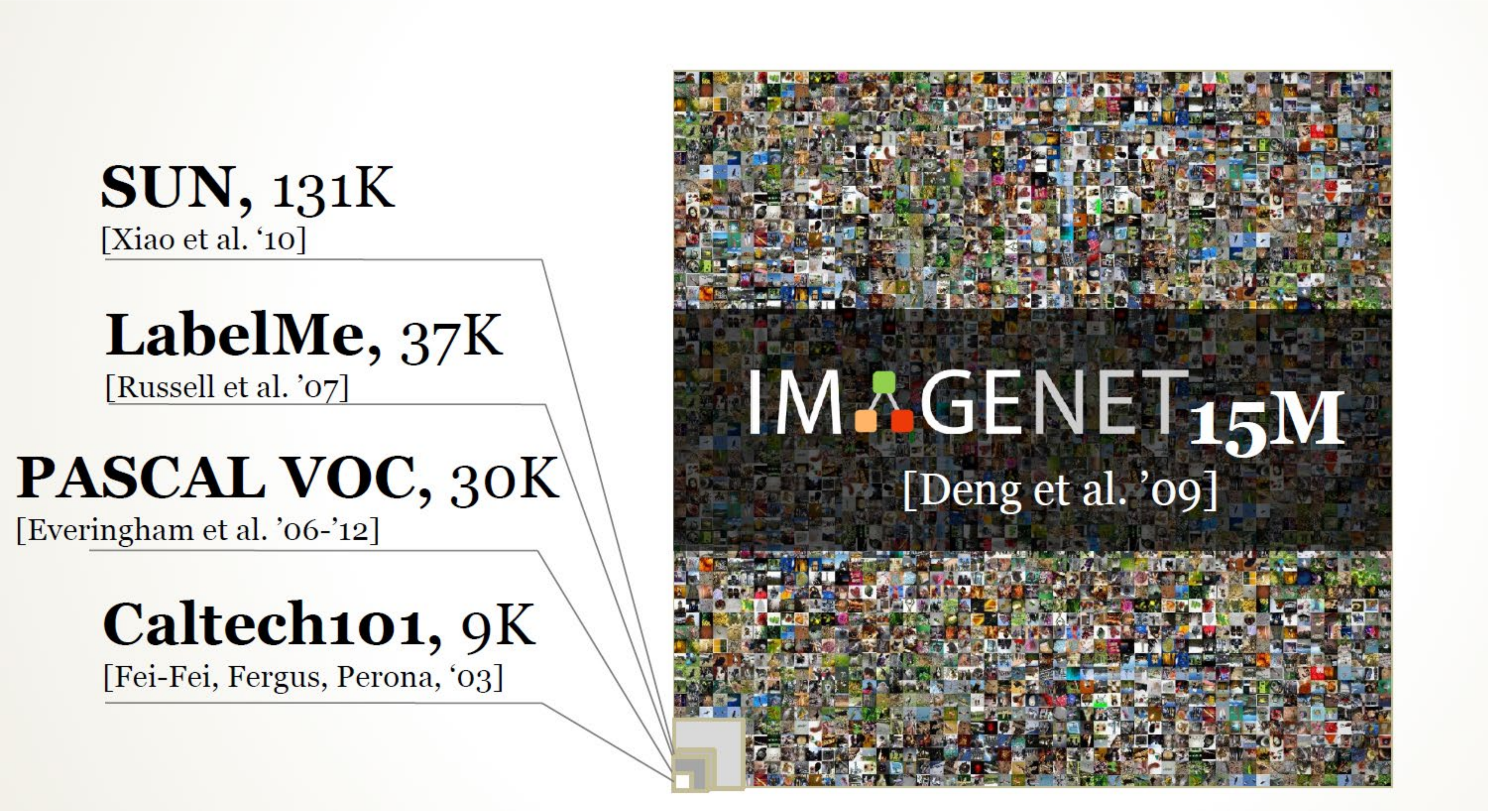
Massive Parallelism



The Result: IMAGENET Goes Live in 2009



ImageNet Targeted Scale



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

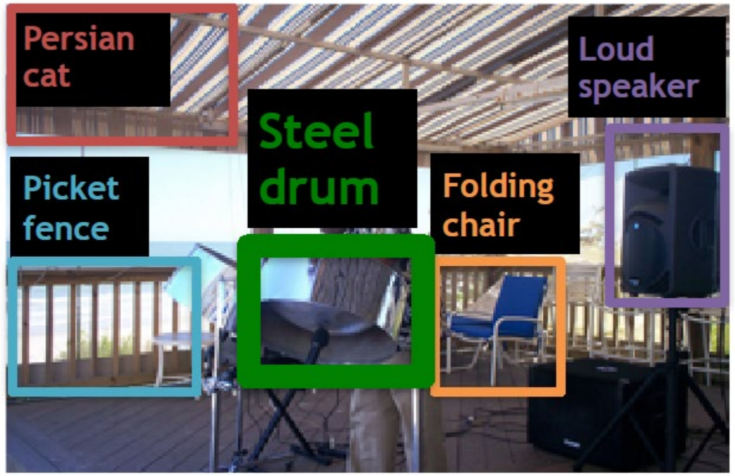
Steel drum

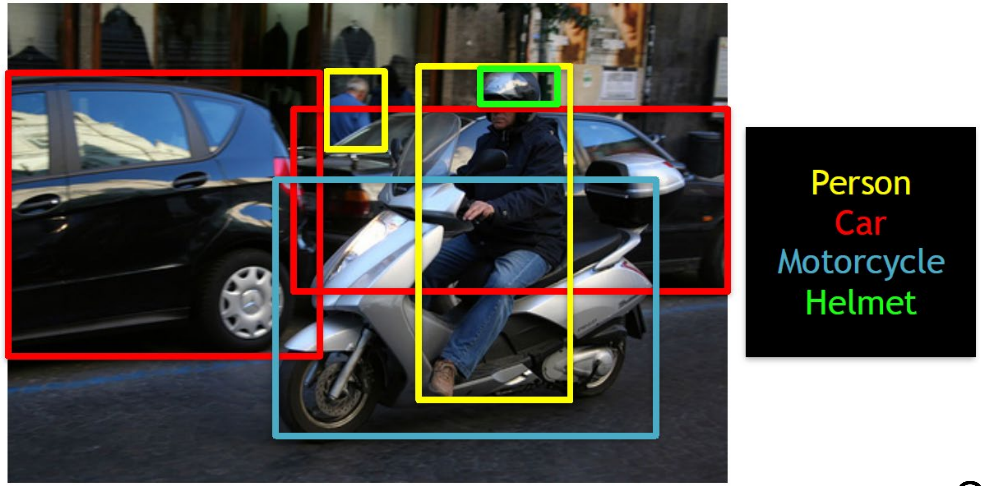
Objects: 1000 classes
Training: 1.2M images
Validation: 50K images
Test: 100K images

Classification

Classification + Localization

Output





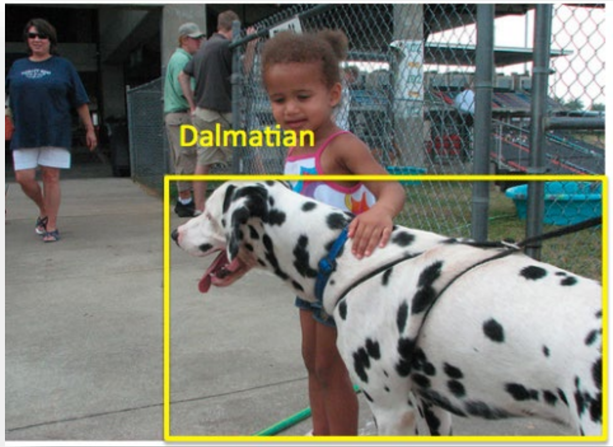
Object Detection

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

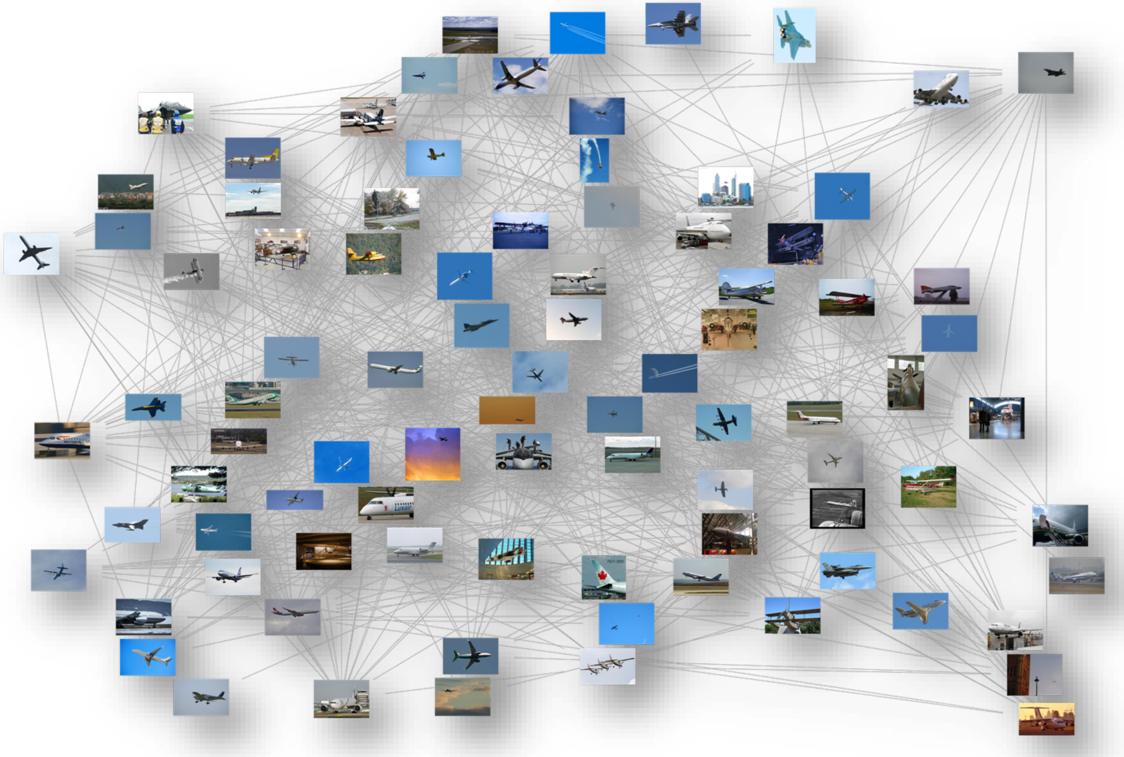
- Geometric and physical knowledge of objects is missing (e.g. ShapeNet)
- Relationships among objects are missing (e.g. VisualGenome)

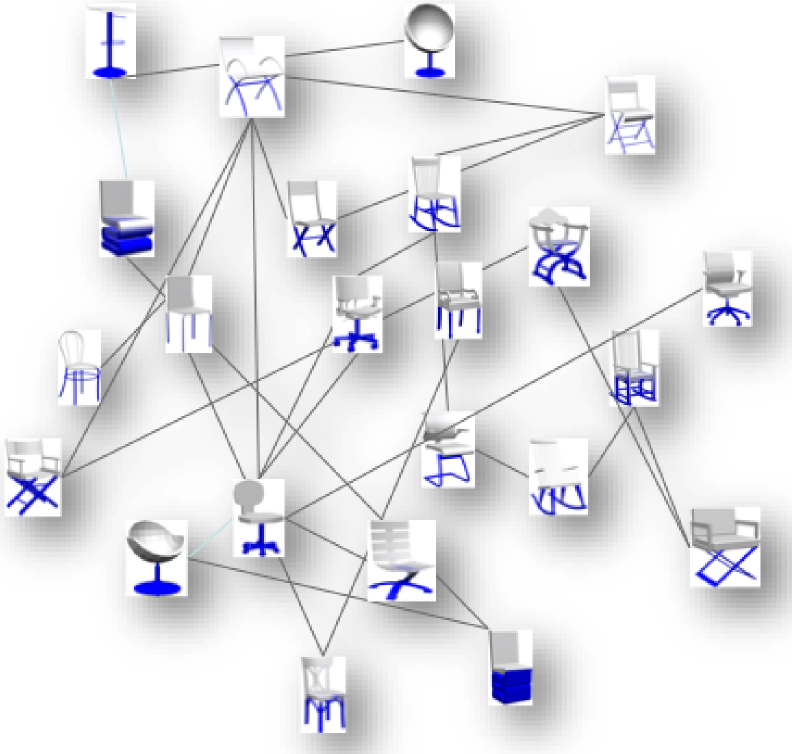
Geometry and Physical Knowledge of Objects

3D Opportunities: Encoding Knowledge

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

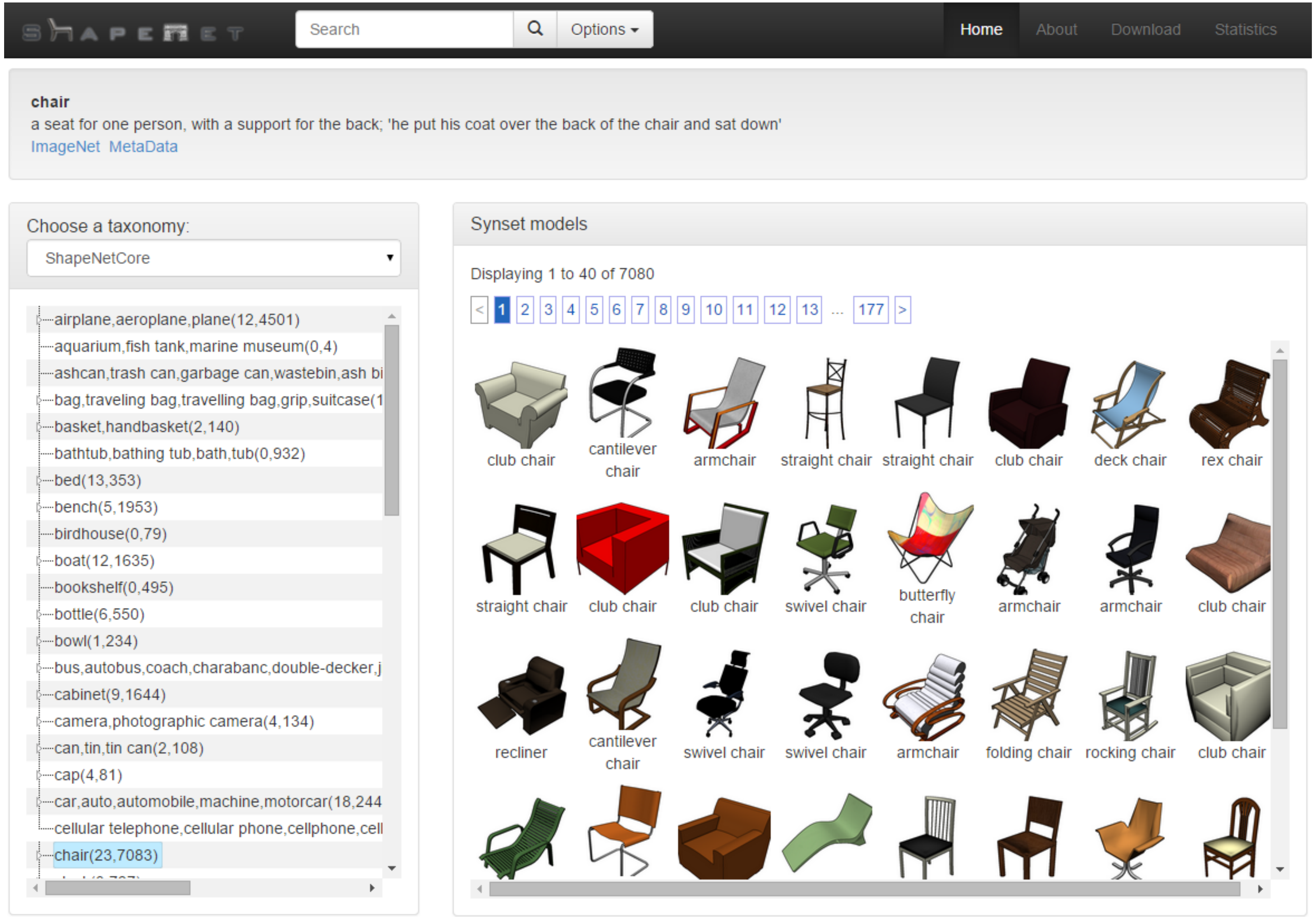
Information Transport





ShapeNet

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



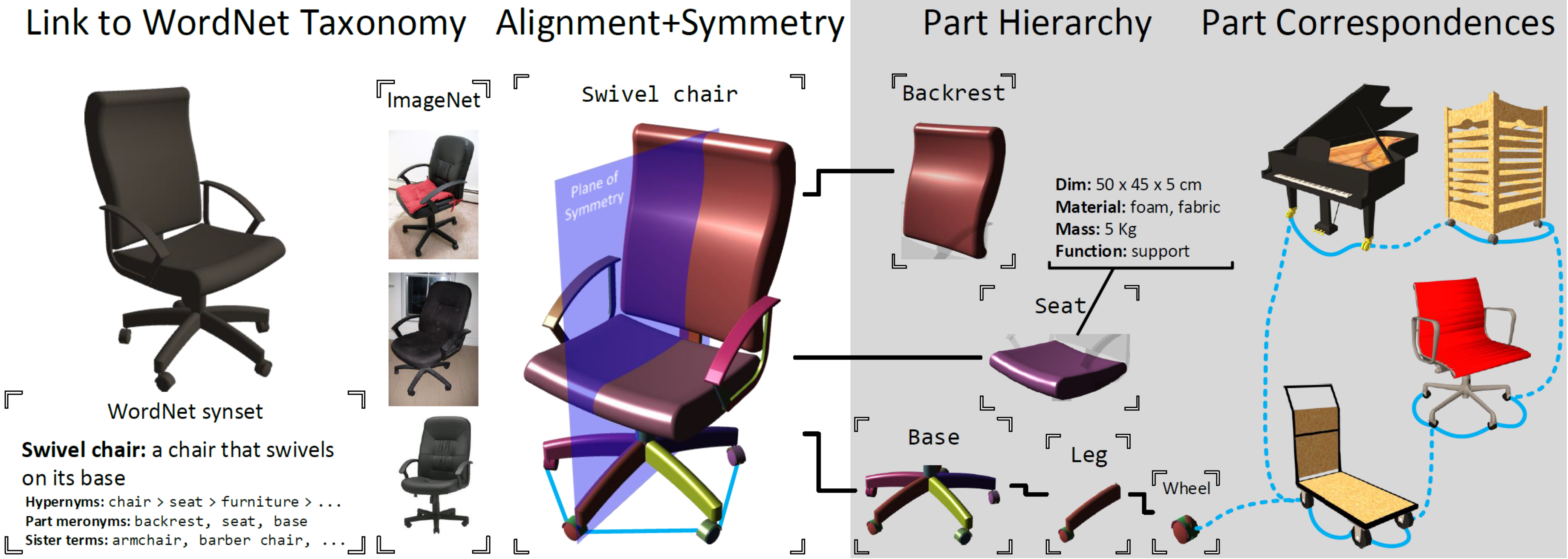
Stanford:
Leonidas Guibas
Pat Hanrahan
Silvio Savarese

Princeton:
Tom Funkhouser
Jianxiong Xiao

UT Austin:
Qixing Huang

Object Knowledge: ShapeNet

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



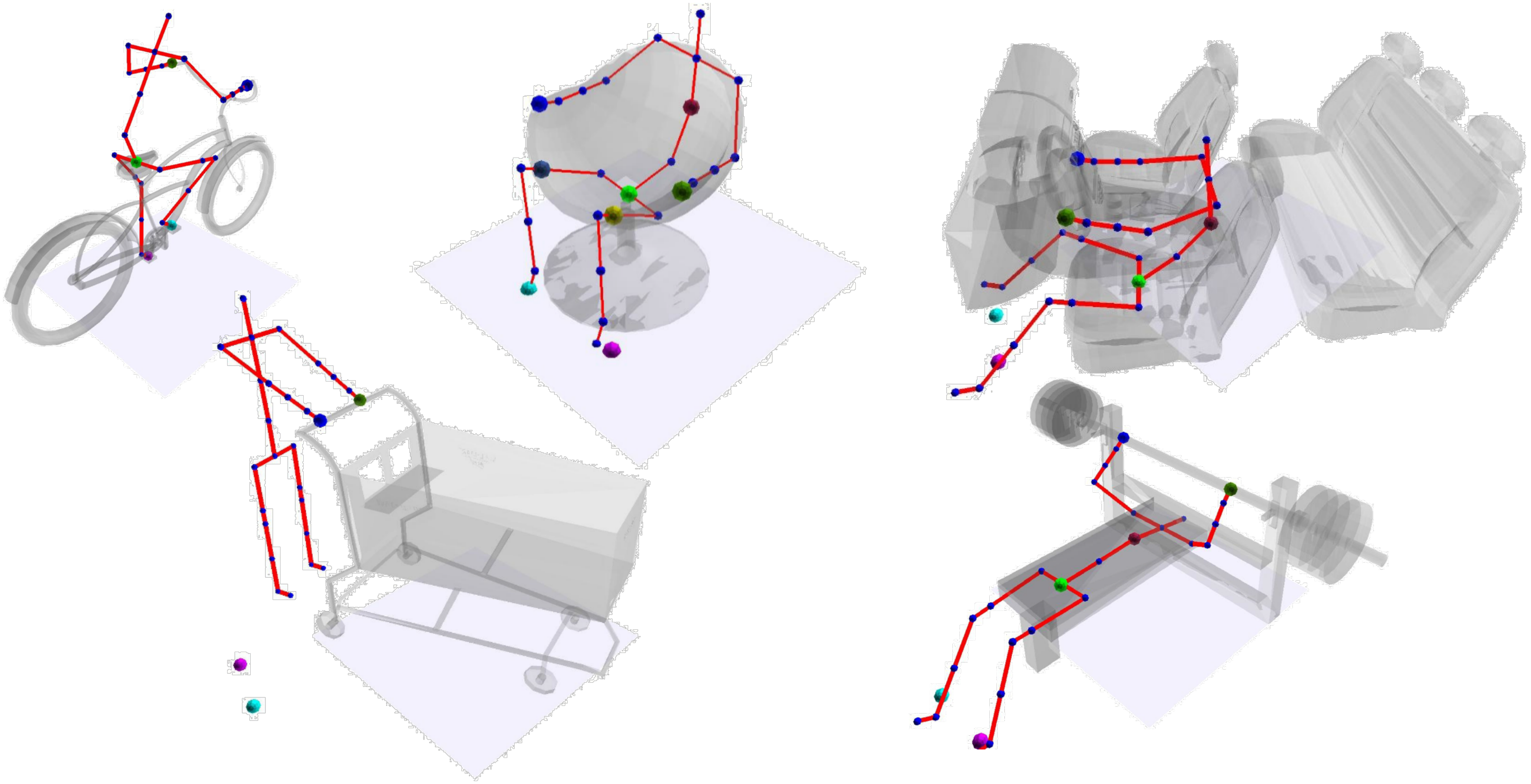
Where is in ShapeNet currently?

- 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?

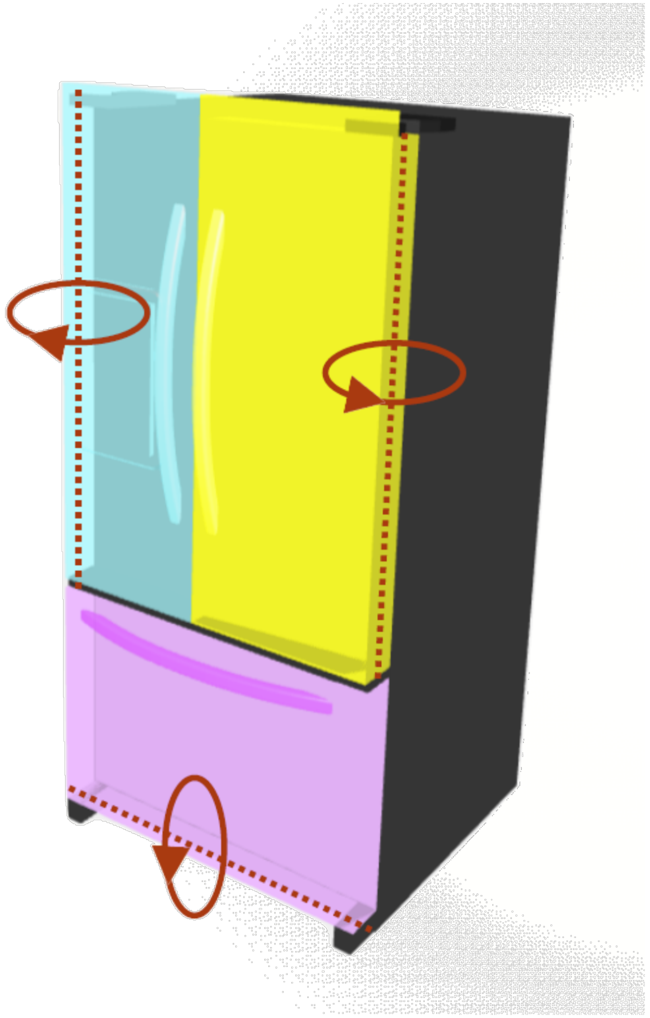
- ShapeNetCore
- ShapeNetSem
  - 12,000 textured models classified into 270 categories, indoor objects
  - Mesh, volumetric representations are provided
  - Consistent orientation within each class
  - Physical dimensions and weights

Object Affordances



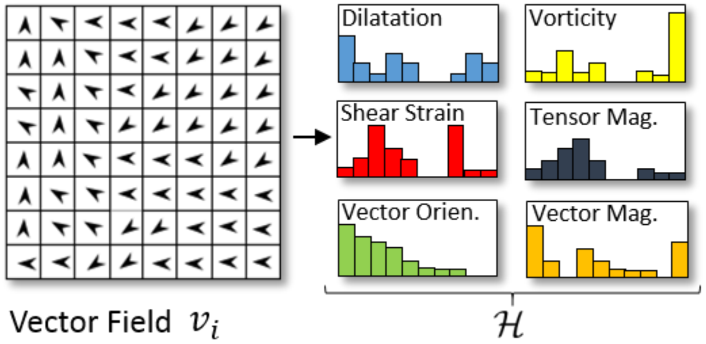
[V. Kim, S. Chaudhuri, L. Guibas, and T. Funkhouser, Siggraph 2014]

Object "Active Sites"

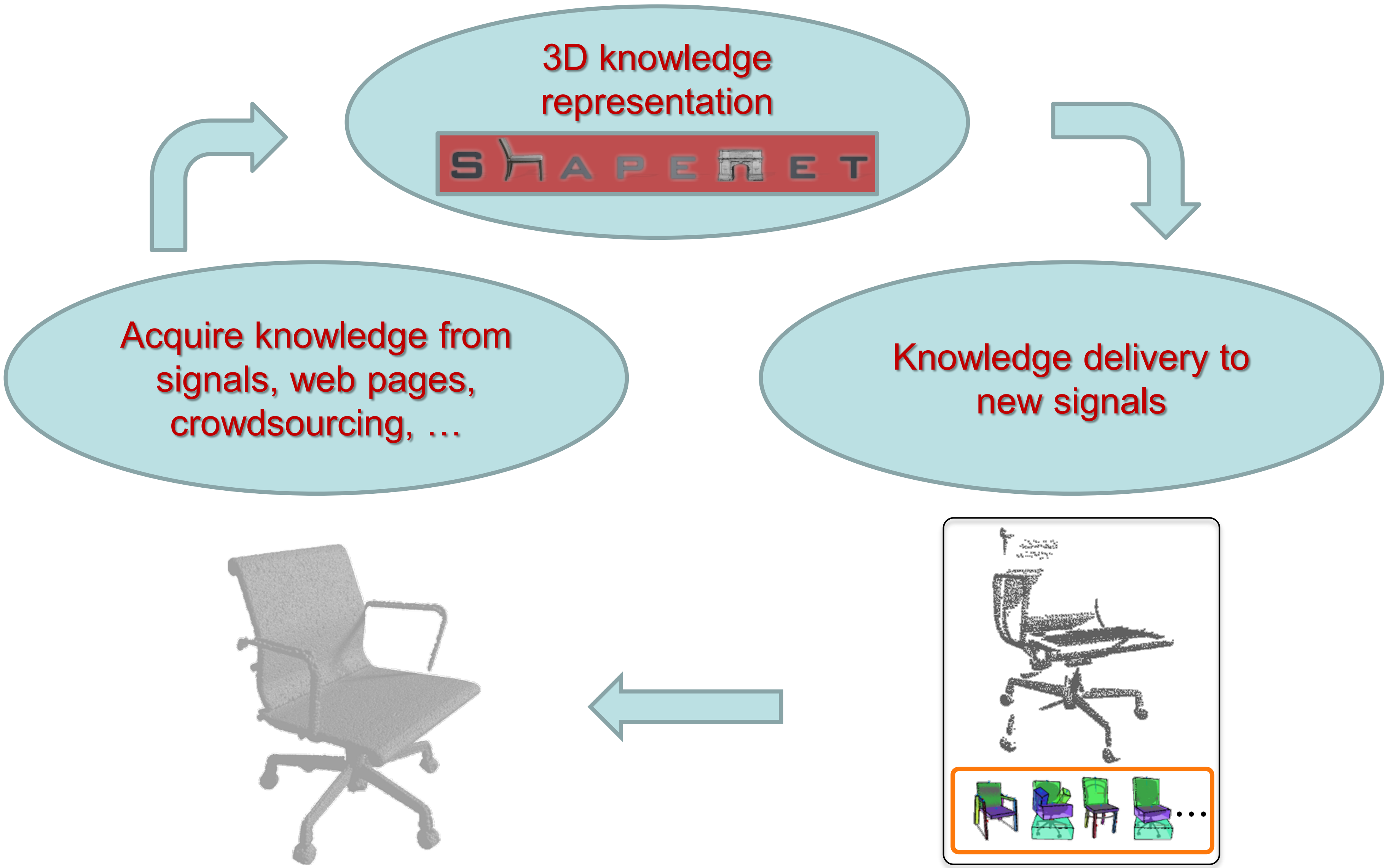


Object Interaction Knowledge

Vector Field to Histograms

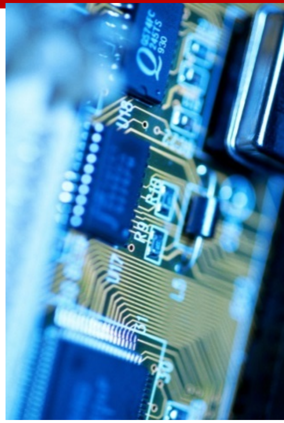


Focus: Knowledge Transport

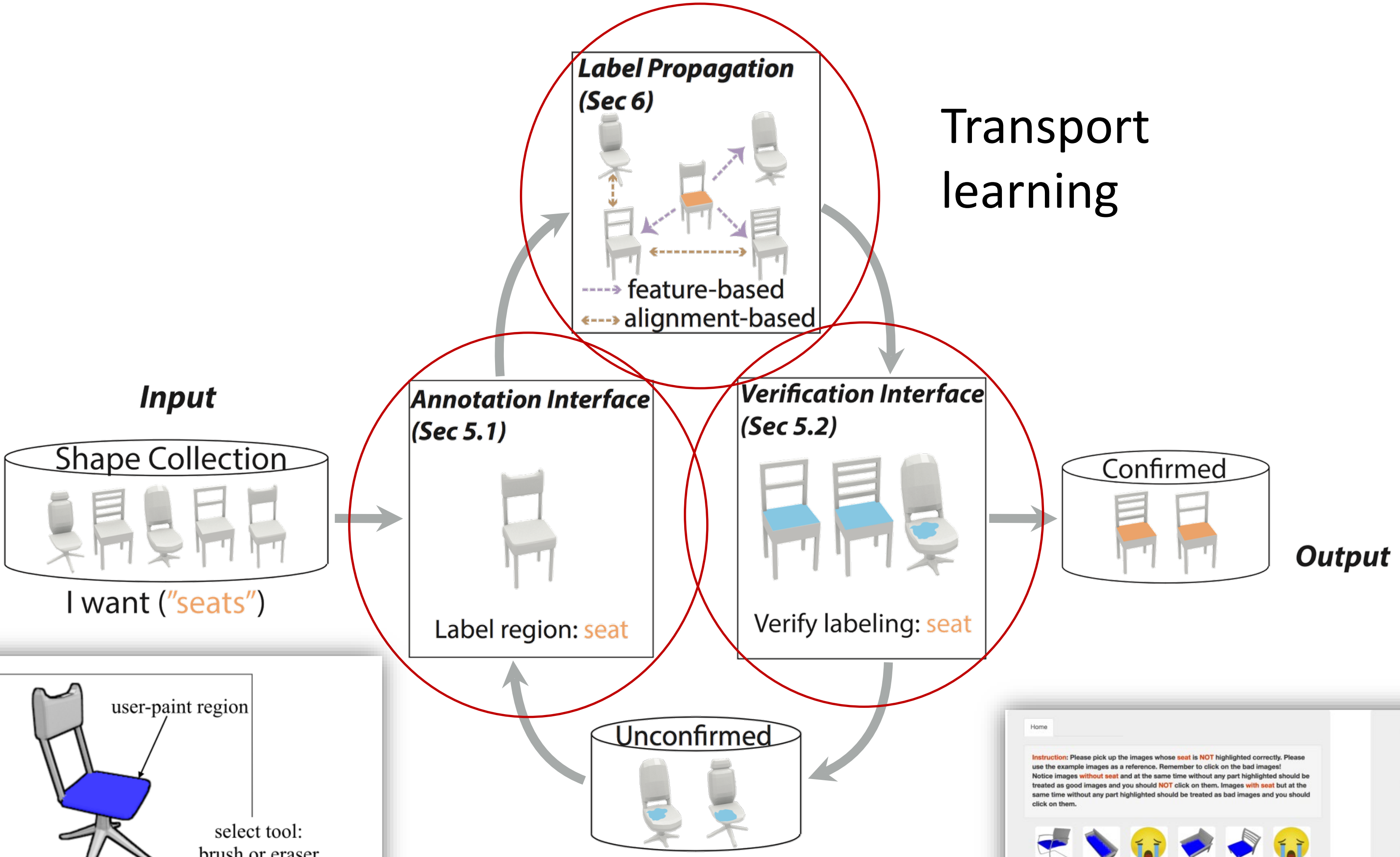


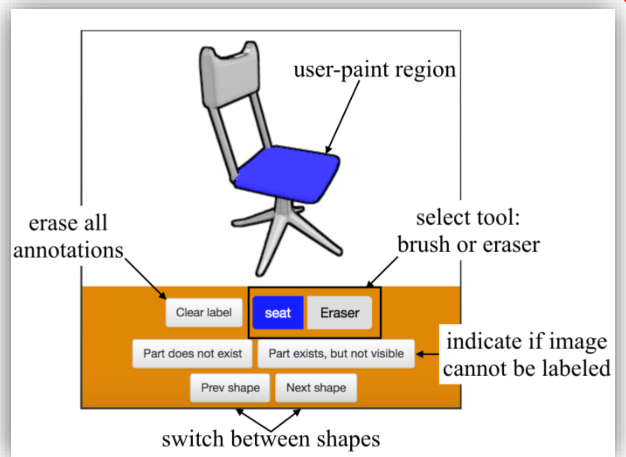
Object Part Annotation

Model Part Annotation

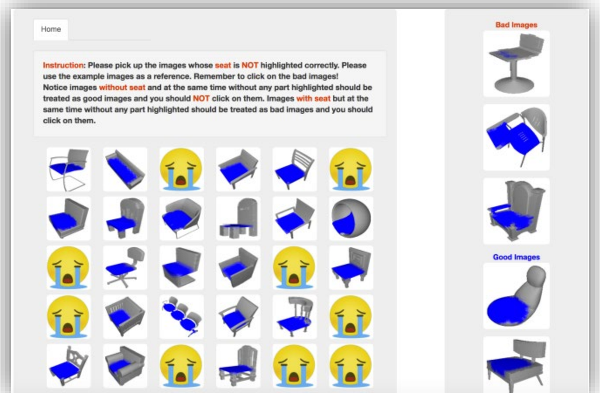


Transport learning

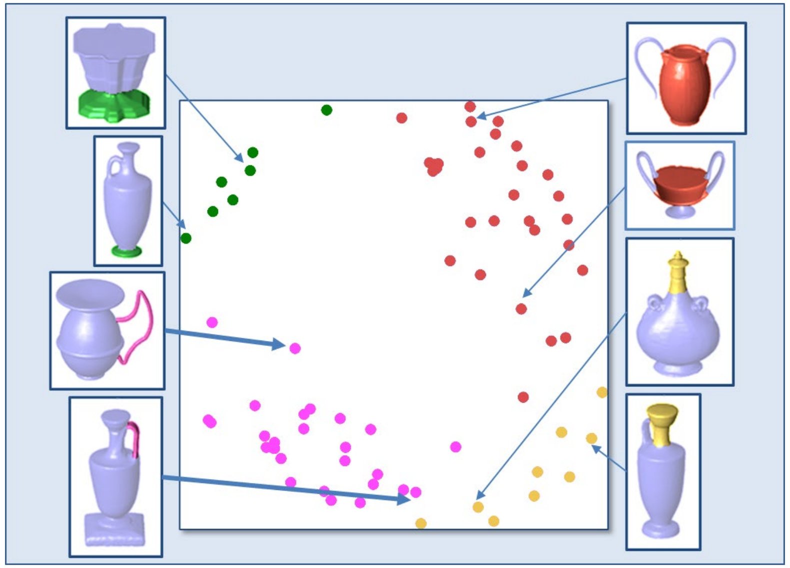




2D annotations

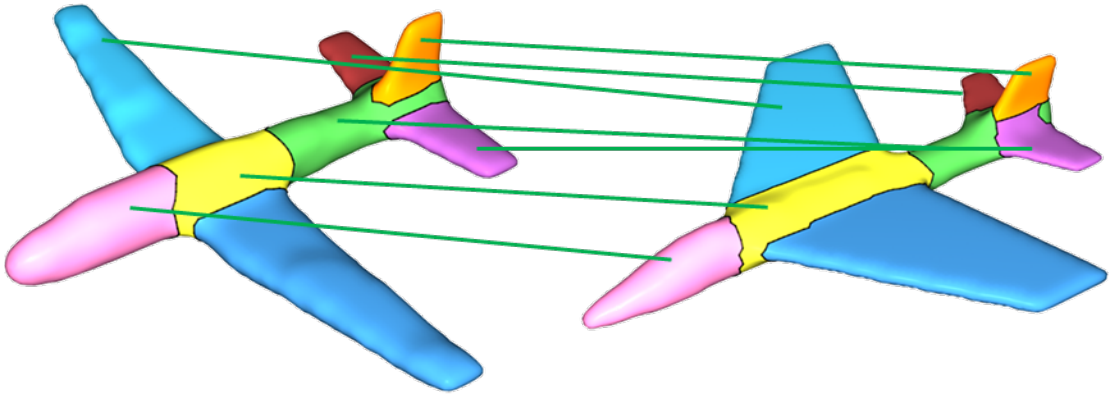


Annotation Propagation



Feature embeddings

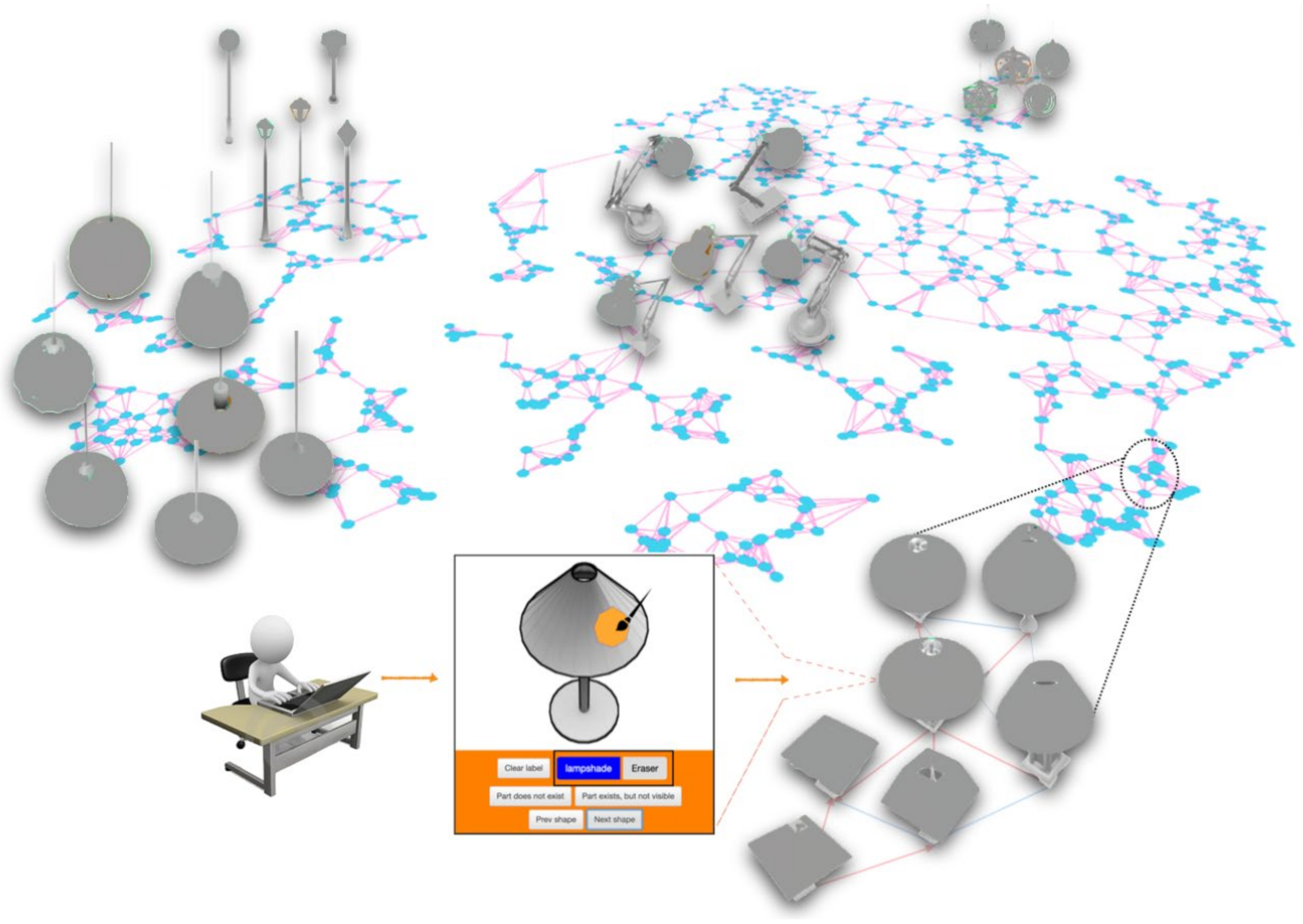
Shape correspondences

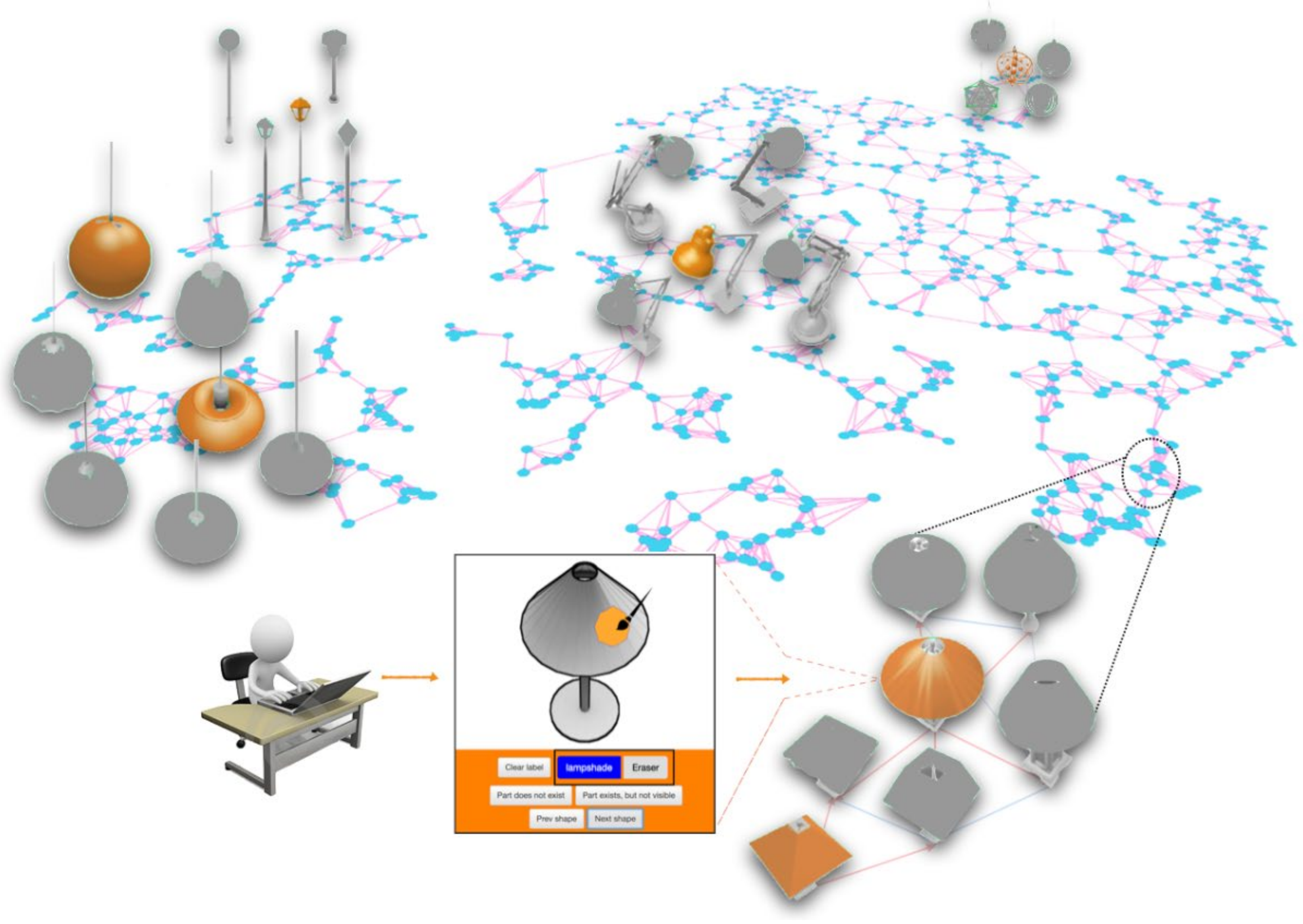


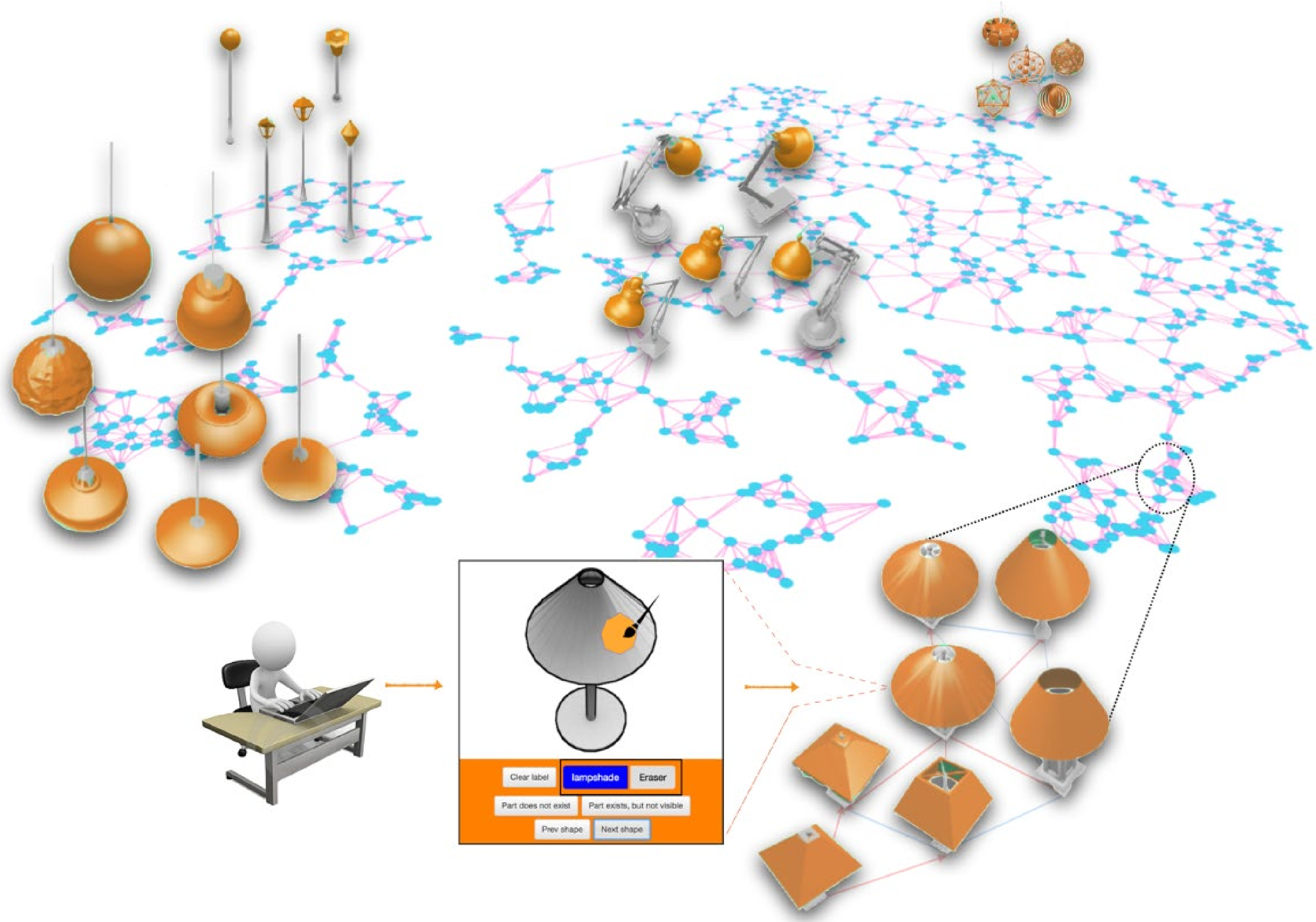
Shape alignments

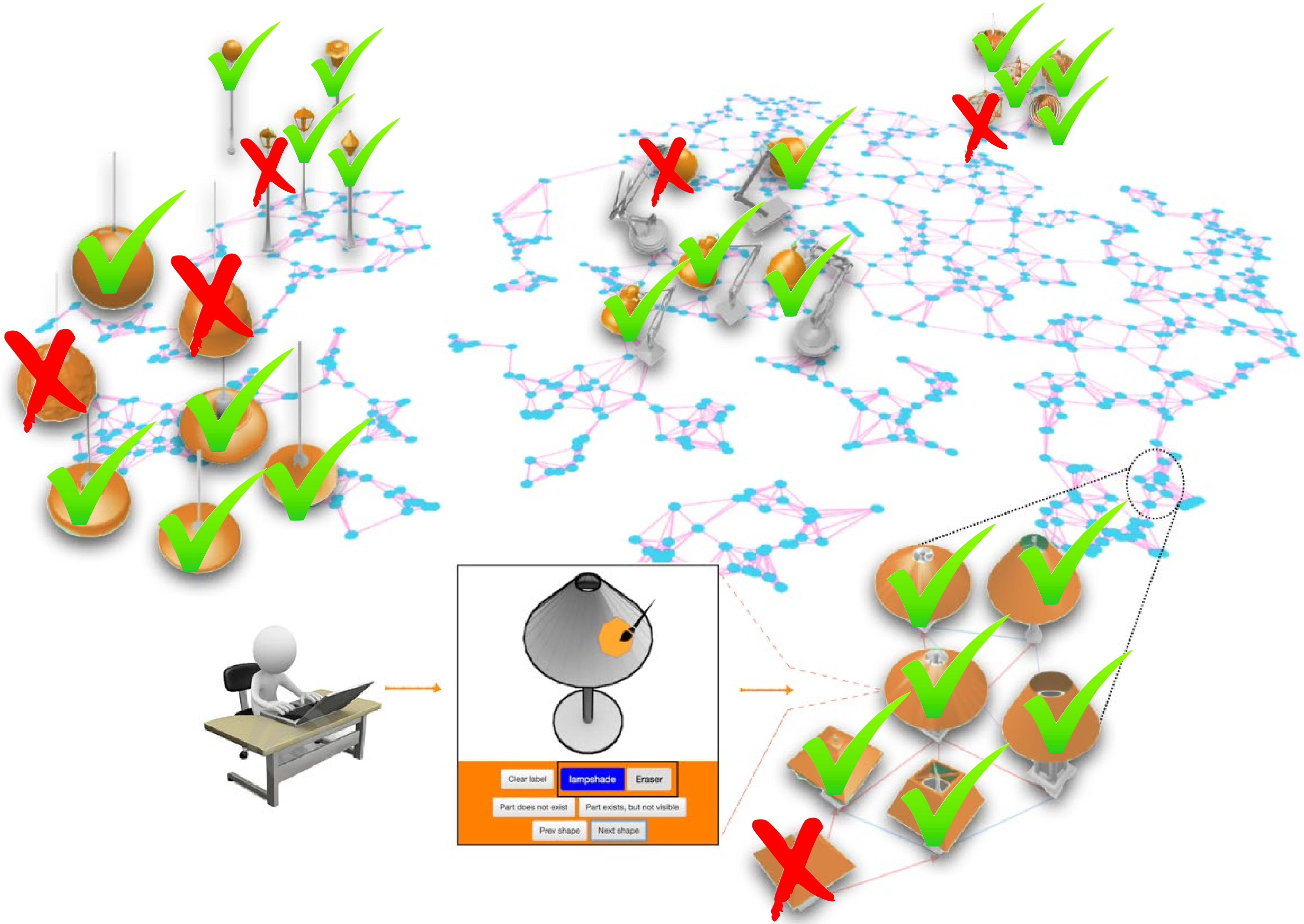
Verification More Efficient Than Annotation

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

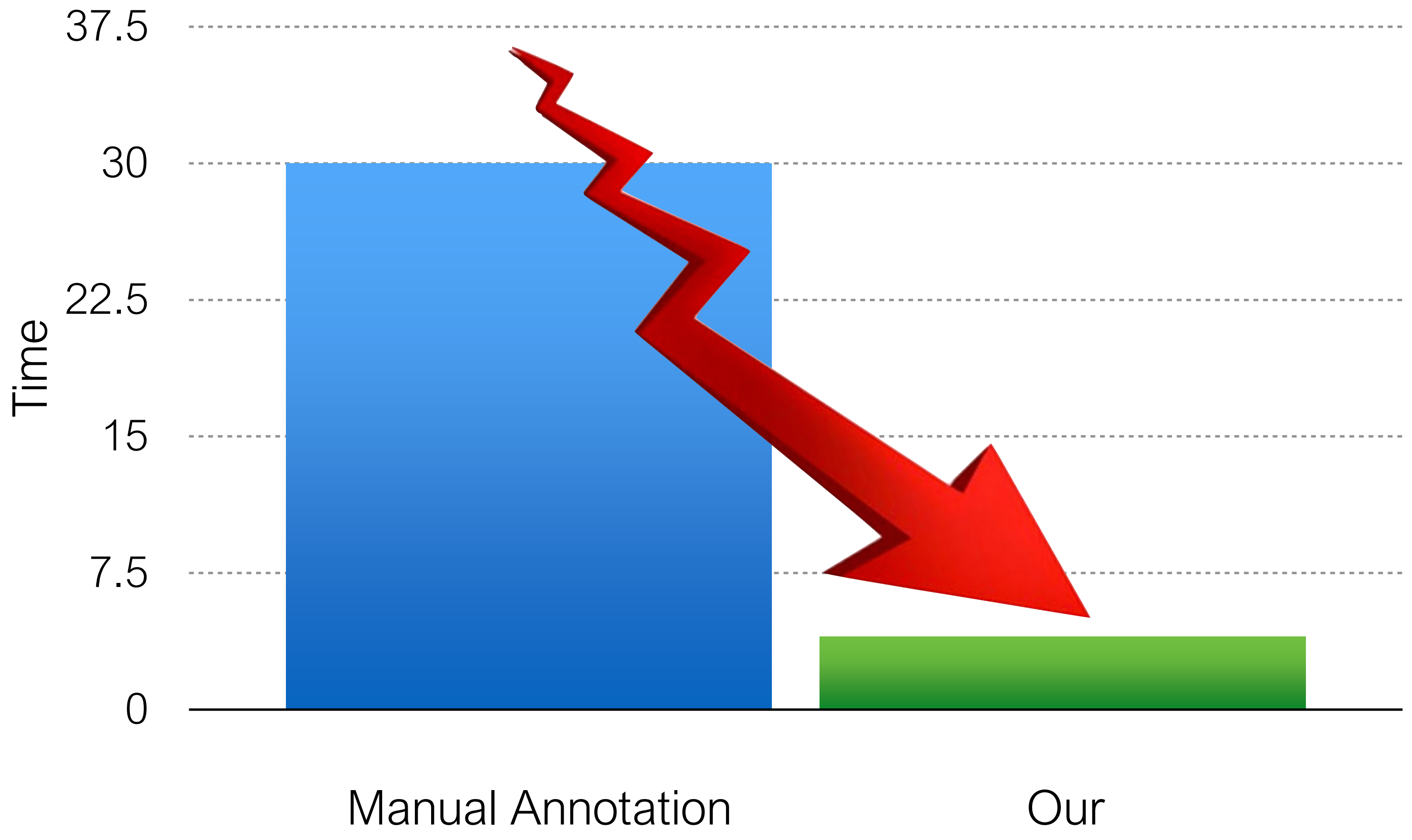

Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape


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

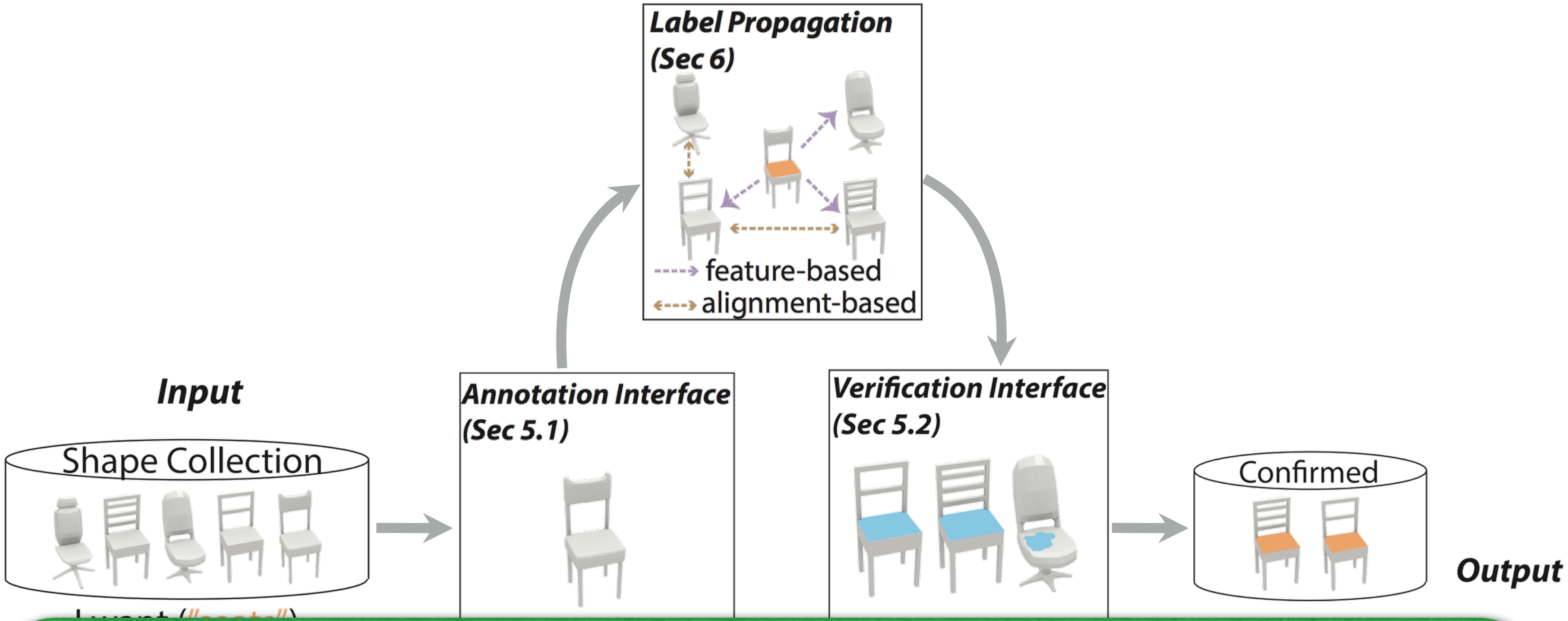

Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape

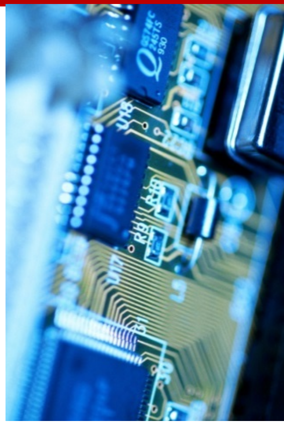

Clear label
lampshade
Eraser
Part does not exist
Part exists, but not visible
Prev shape
Next shape

Reduced Annotation Cost



Model Part Annotation





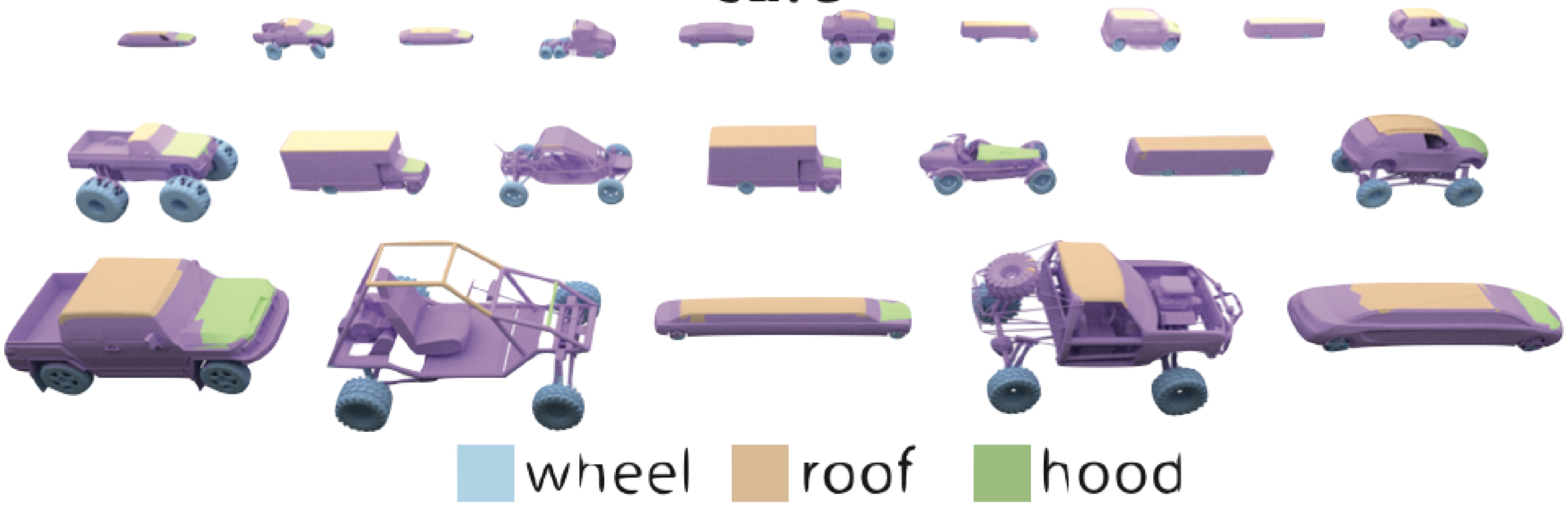
How to issue ***Annotation*** and ***Verification*** tasks

to optimize the ***Utility Function***

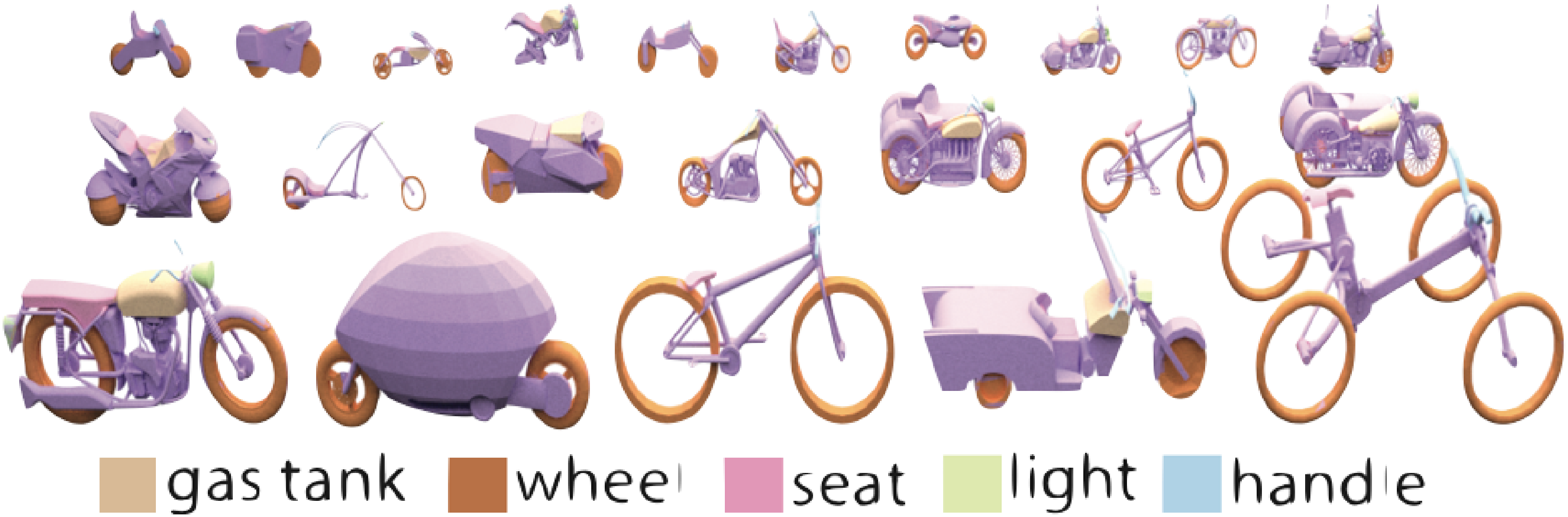
$$\frac{\#(\textit{ACCURATE}\ \text{labeling})}{\text{worker time}}$$

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

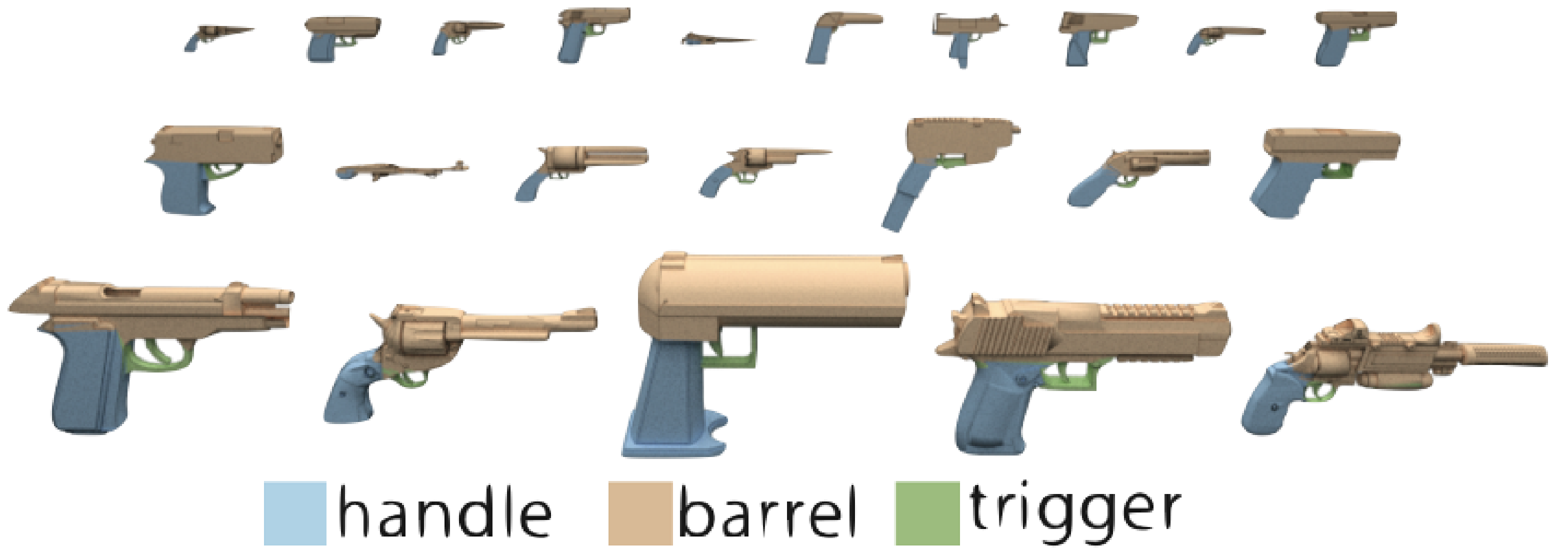
cars



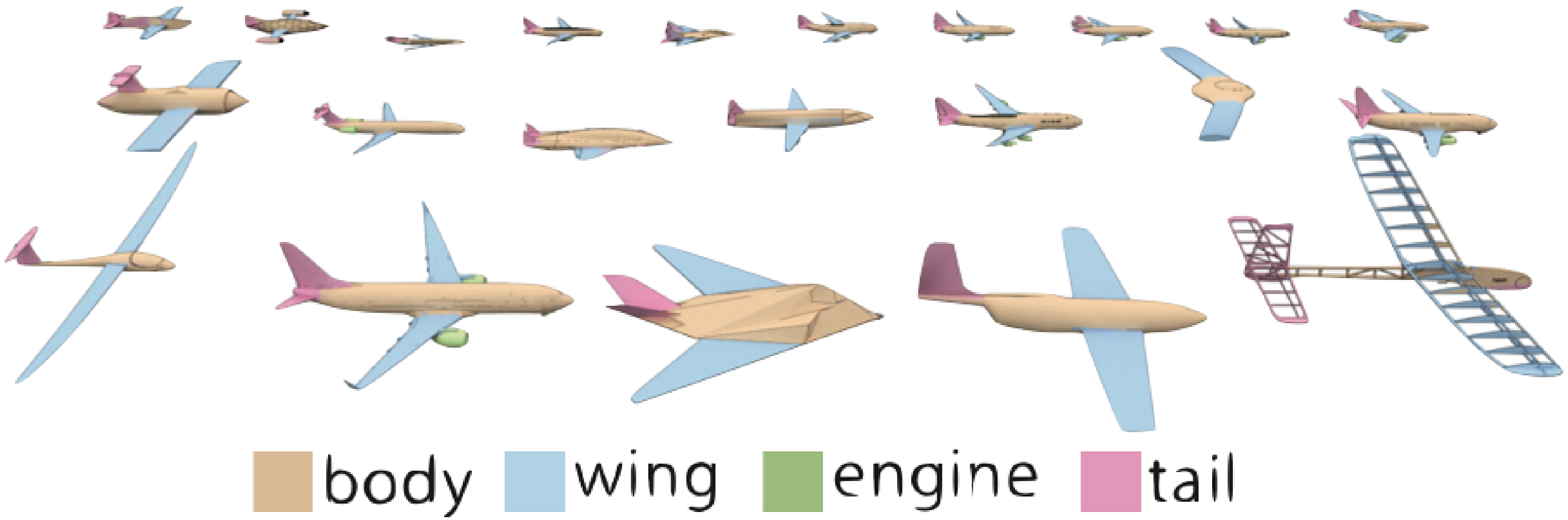
motorbikes



pistols

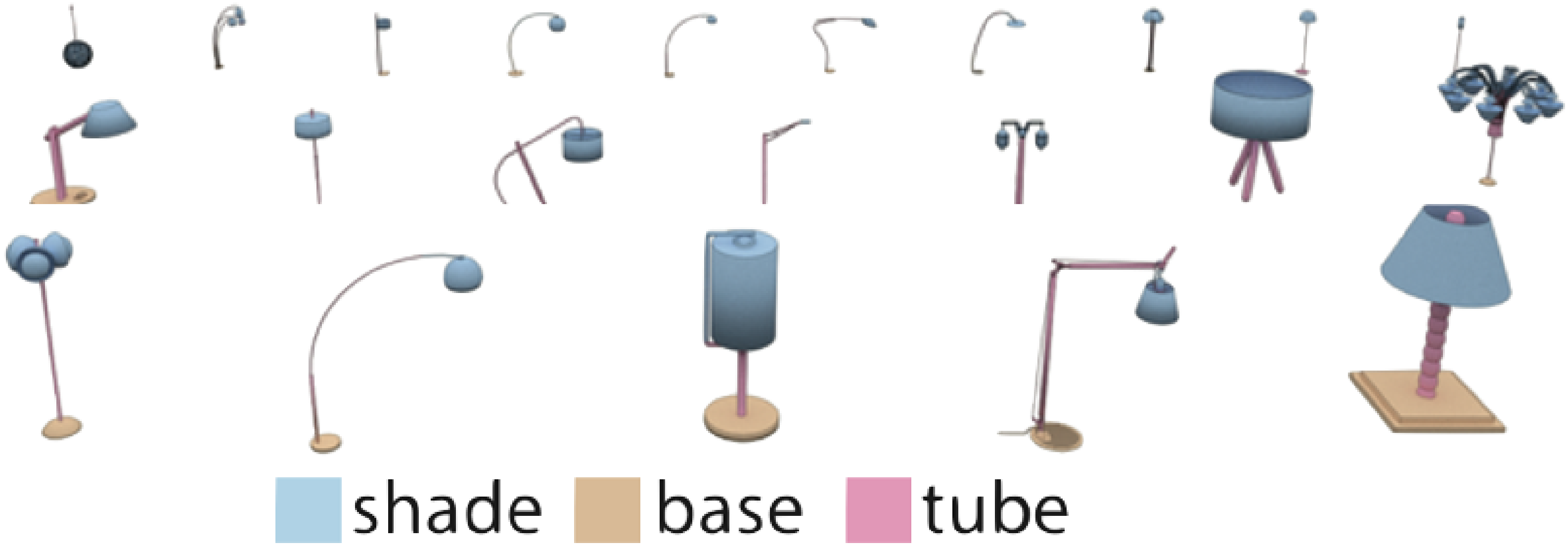


airplanes

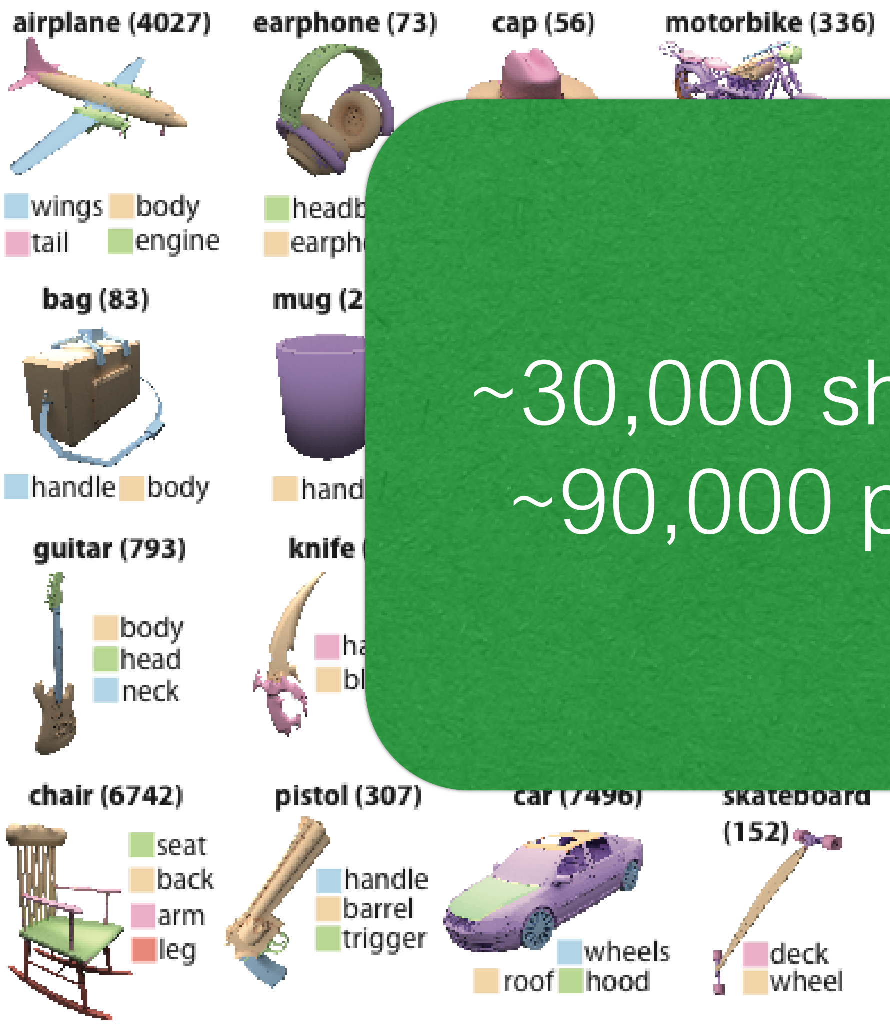


bags

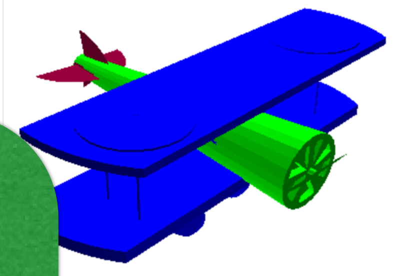
lamps

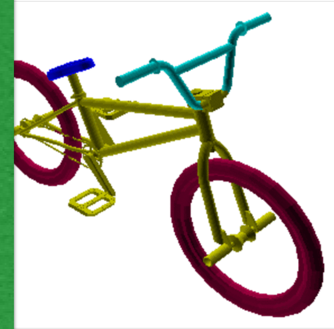


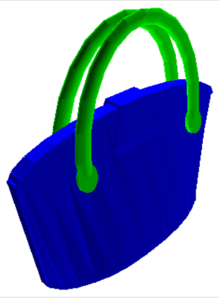
Results



~30,000 shapes
~90,000 parts

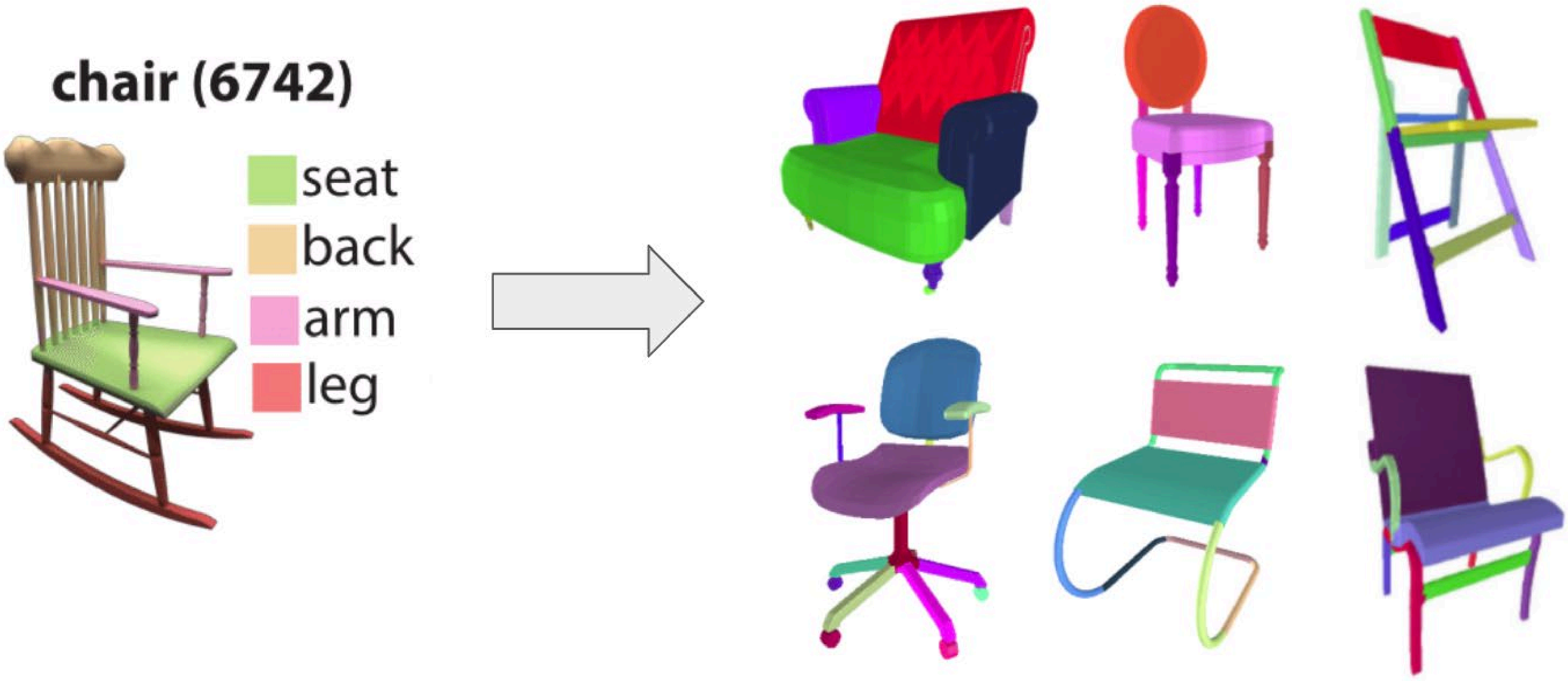






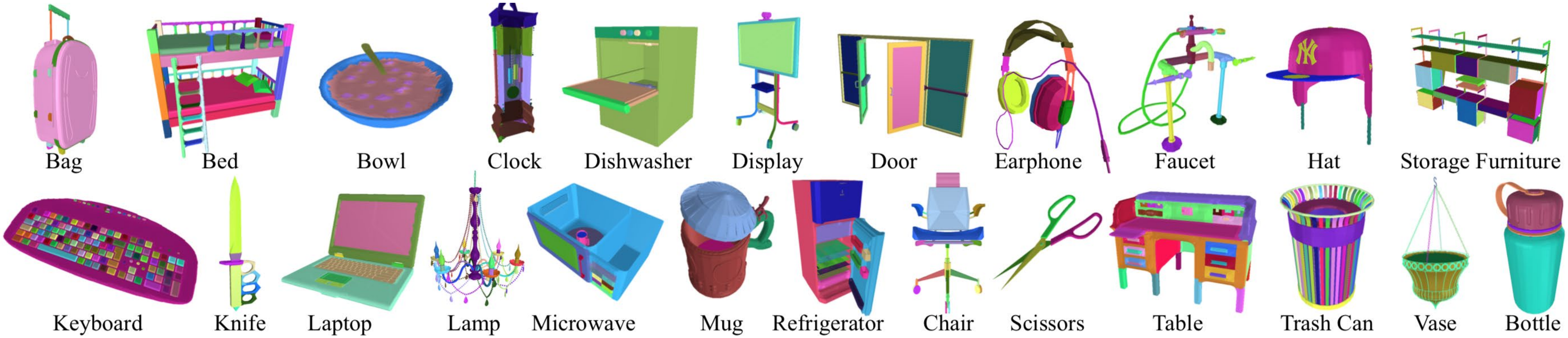
Fine-Grained Part Annotation

PartNet: Fine-Grained and Instance-Level Parts



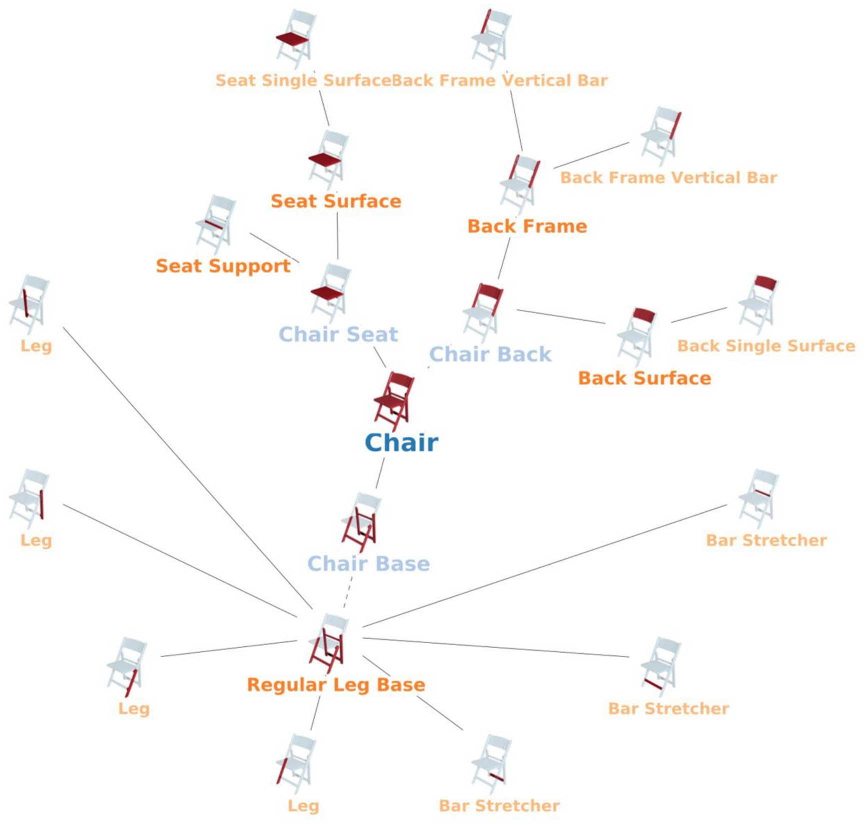
[K. Mo, S. Zhu, A. Chang, L. Yi, S. Tripathi, L. Guibas and H. Su, PartNet: A Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding, CVPR 2019]

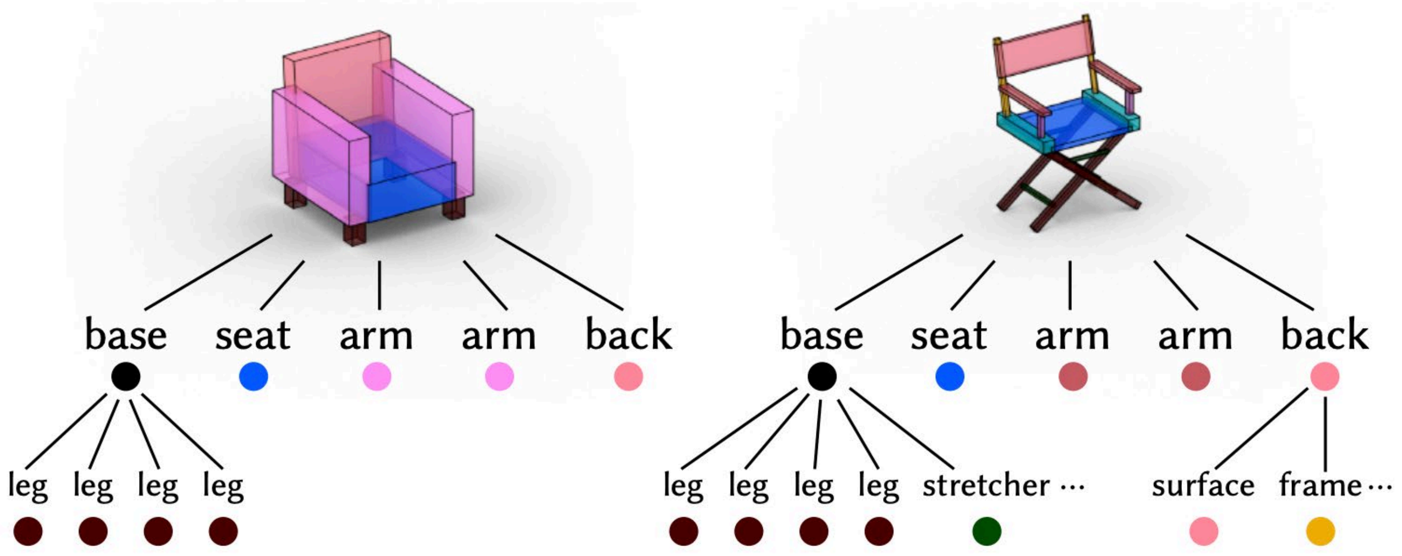
PartNet: Fine-Grained Parts



- 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

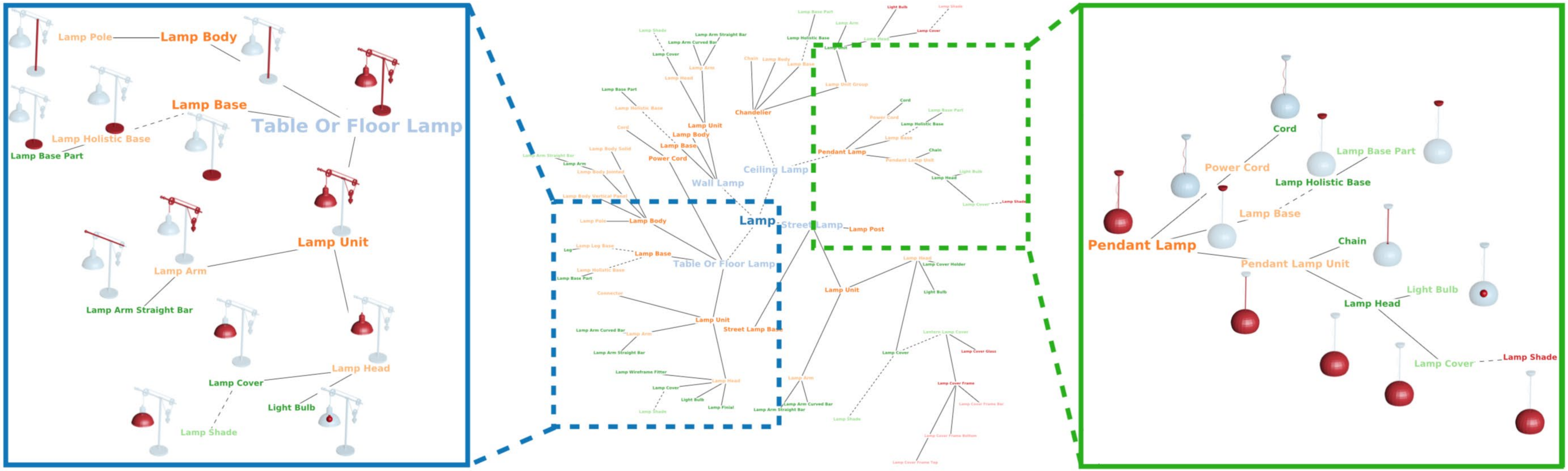
PartNet: Hierarchical and Consistent Parts





- 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

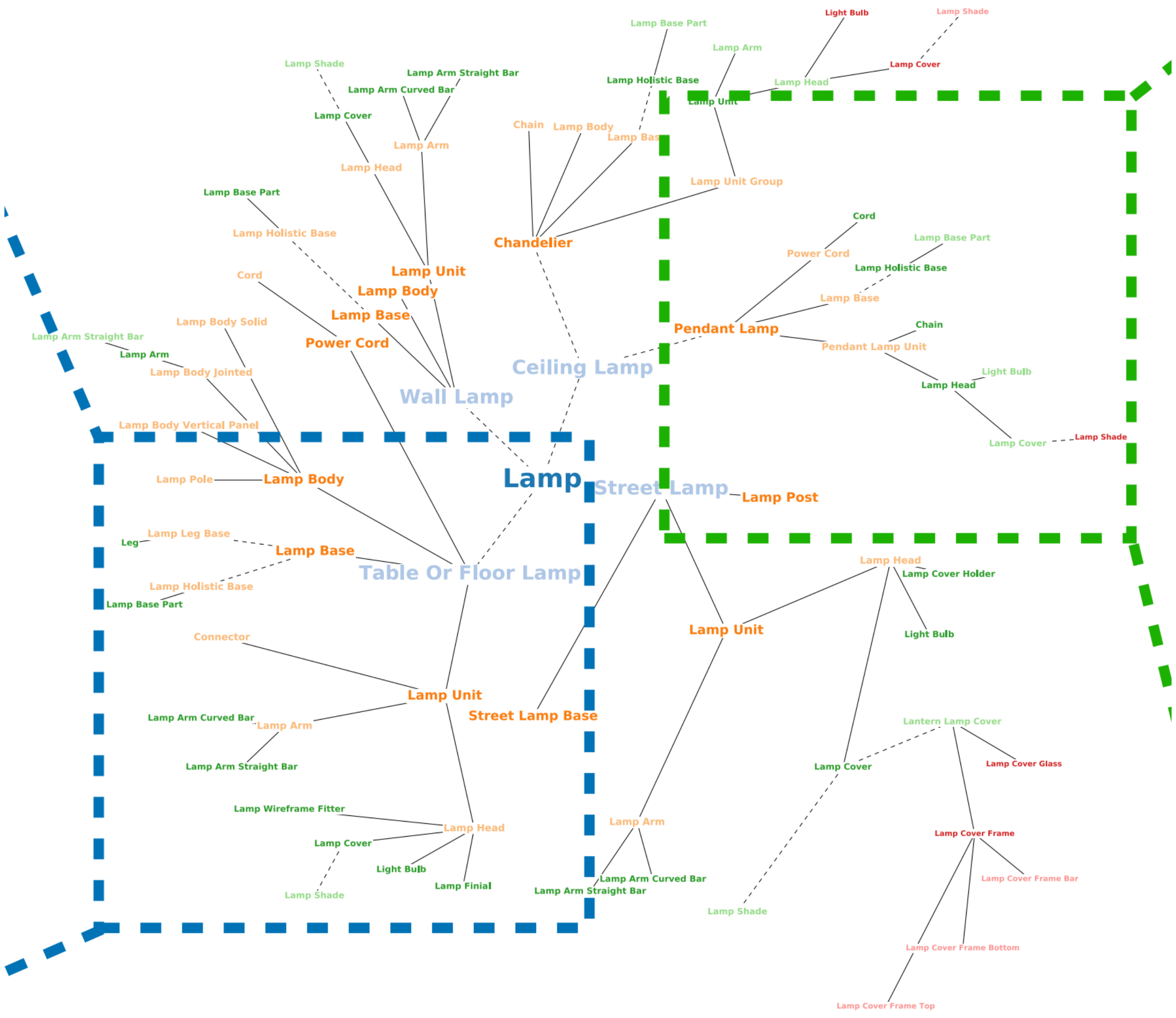


One Table Lamp Example

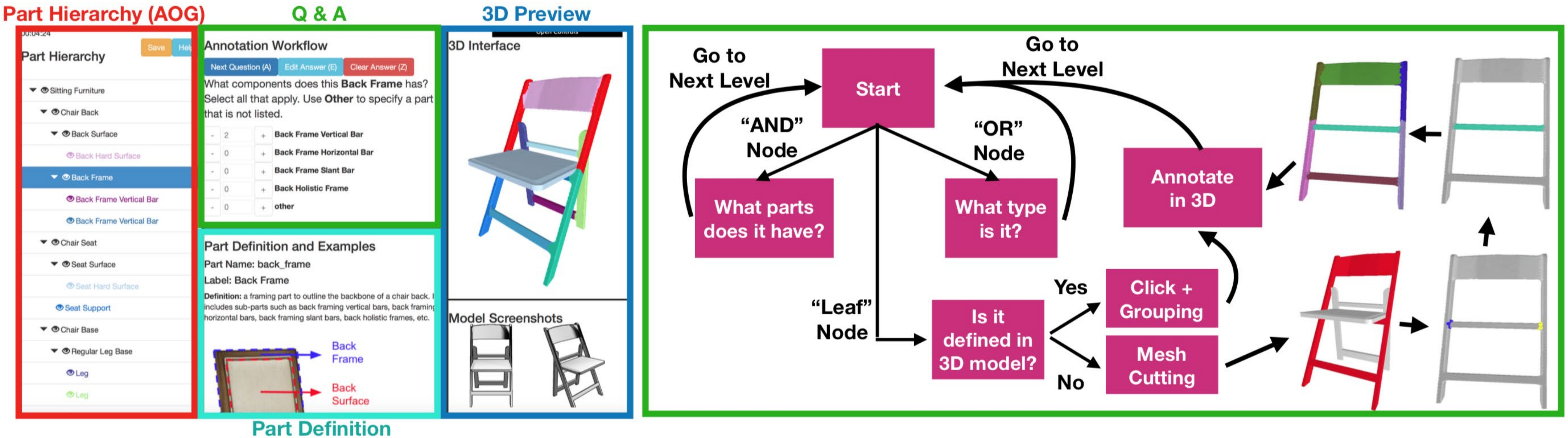
Part Template

One Floor Lamp Example

PartNet: And-Or-Graph Semantic Part-Template



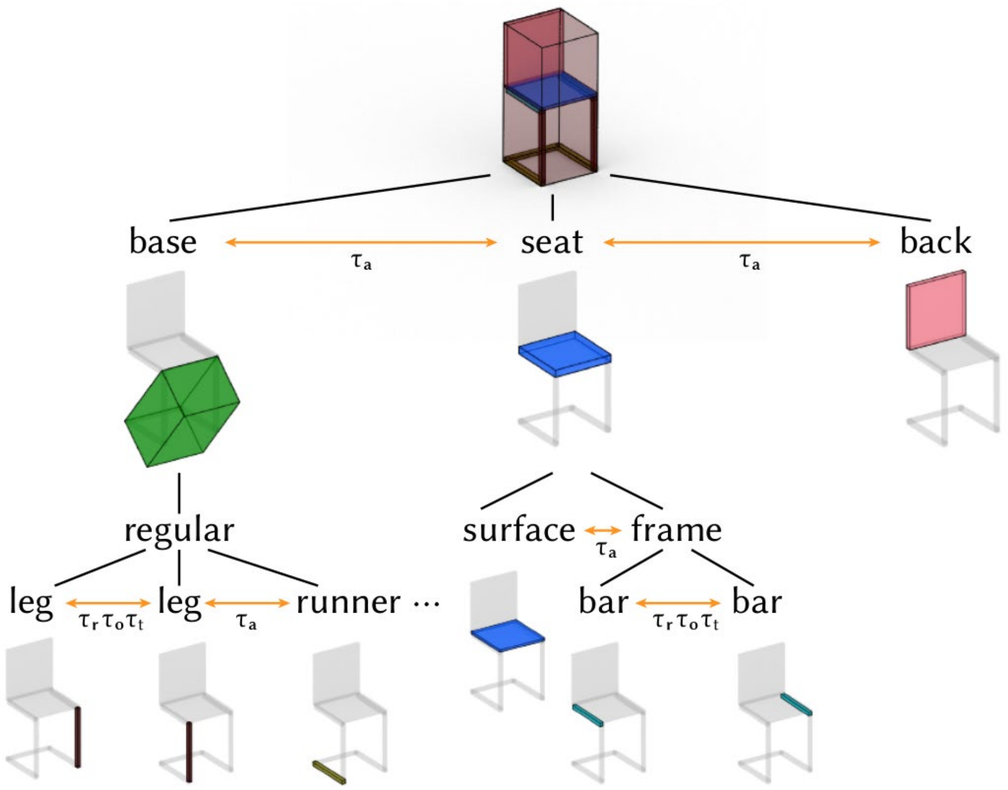
PartNet: Annotation Interface



(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

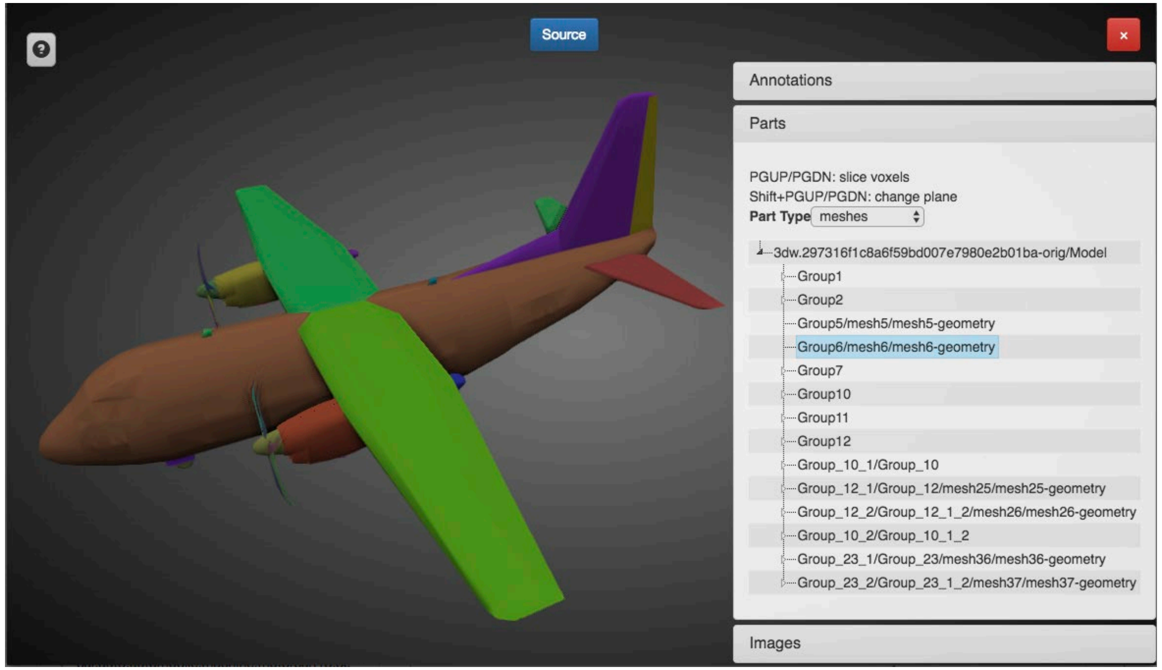
PartNet-Mobility: Annotate Part Motions

[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]

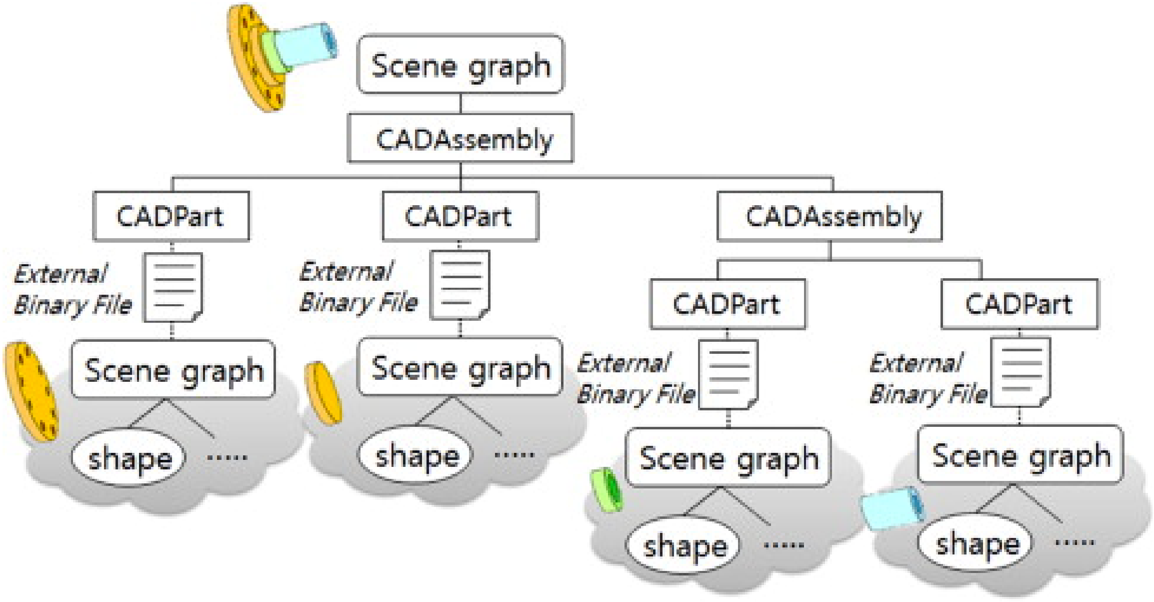
Learning from Noisy Web Data

Observations

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

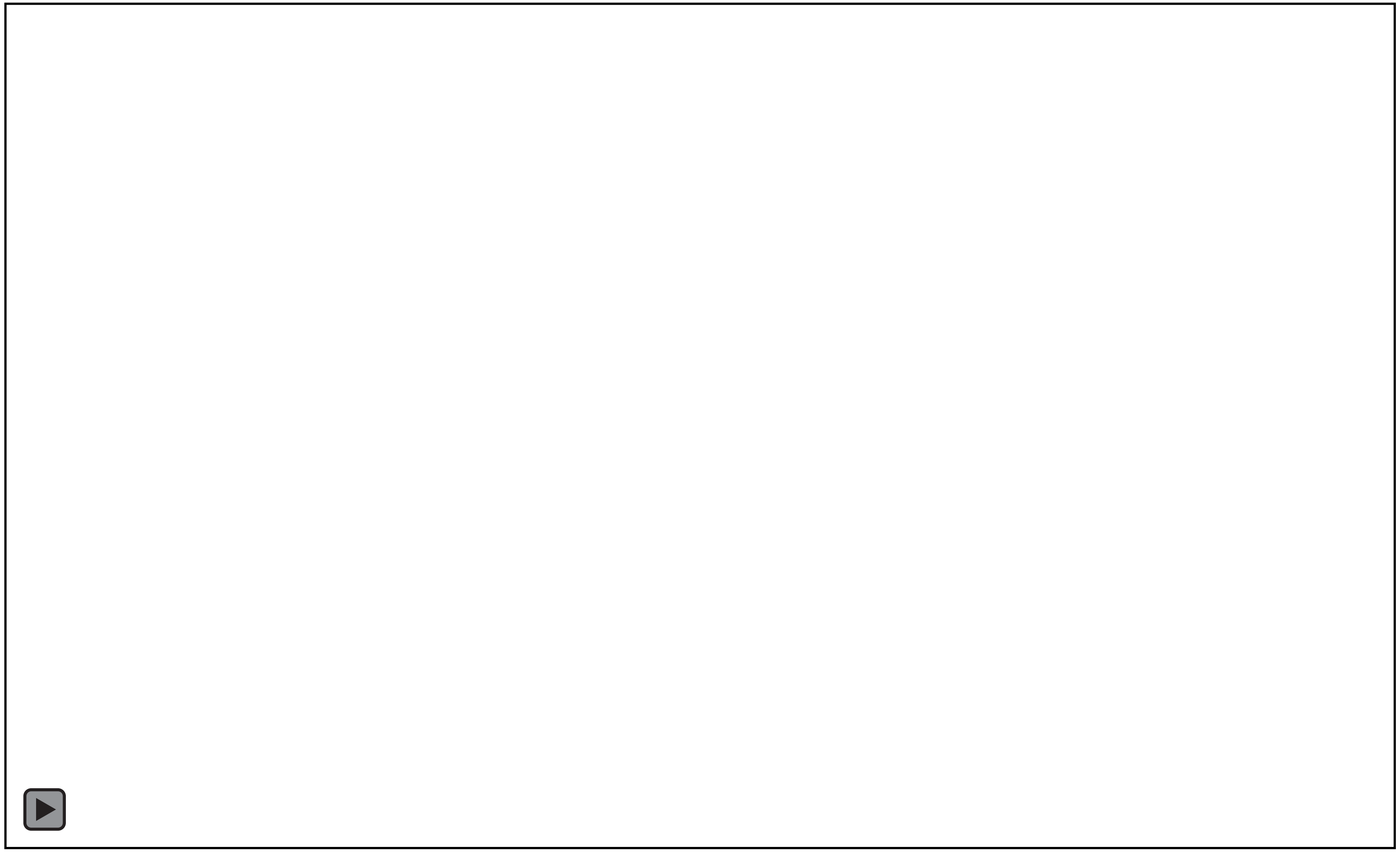


ShapeNet



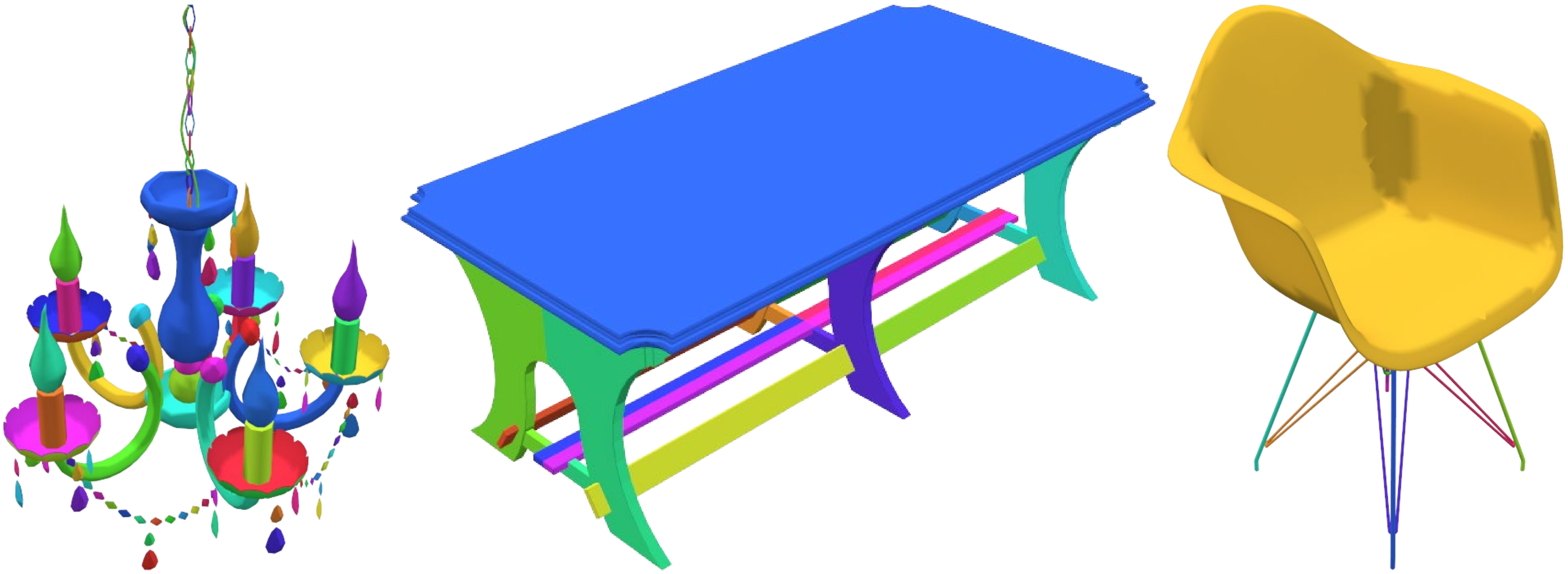
Kim et al., 2015

Distill Knowledge from Object Graphs



Observations

(+) Provide natural part segmentations.



Observations

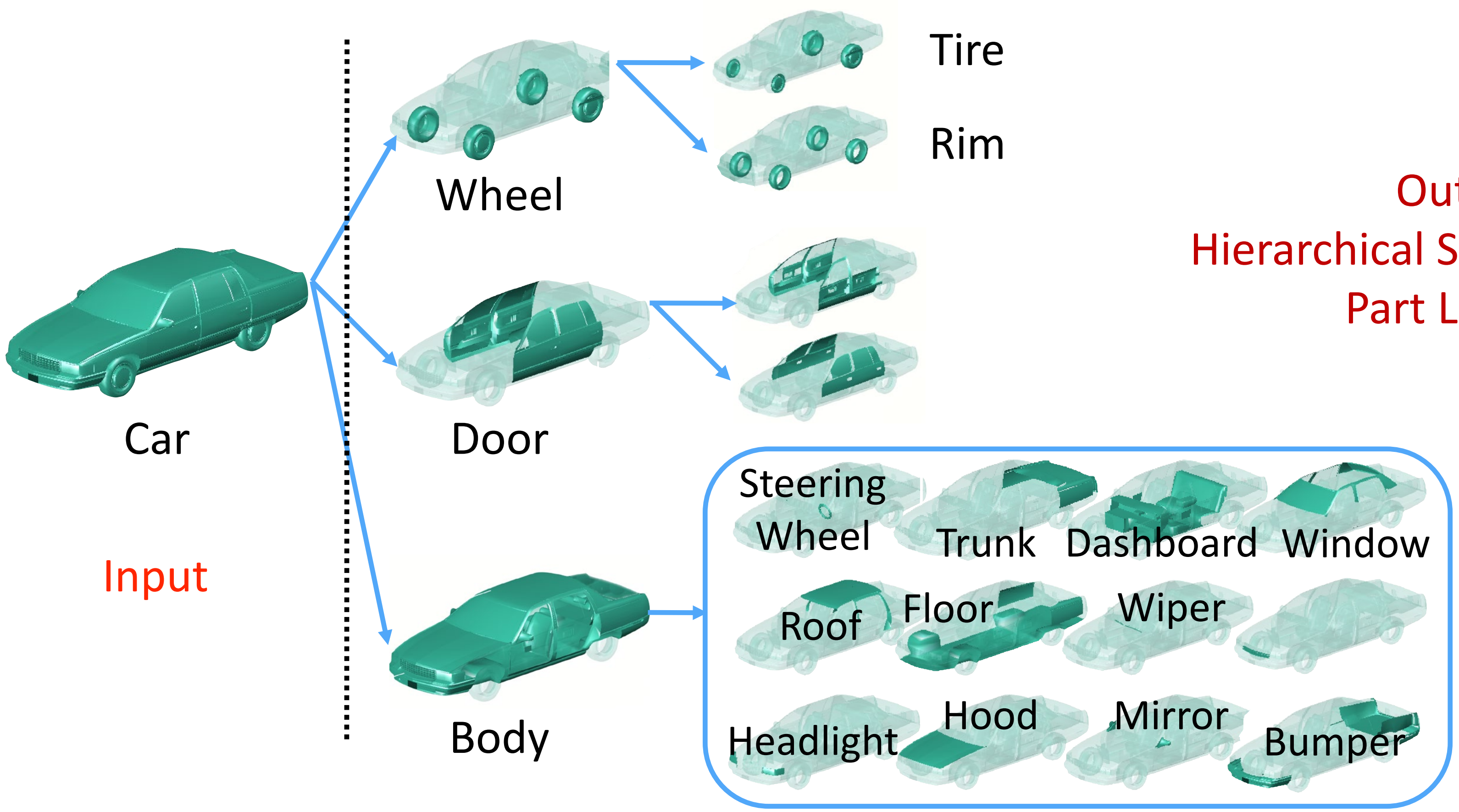
(+) Provides natural part segmentations.

(–) Inconsistent -- and often unlabeled.

An Application of Horizontal Networks

- A scalable active framework for part annotation in ShapeNet
- Based on “horizontal” information diffusion
- Learning hierarchical shape segmentation and labeling in a weakly supervised manner from online repositories

Problem Definition

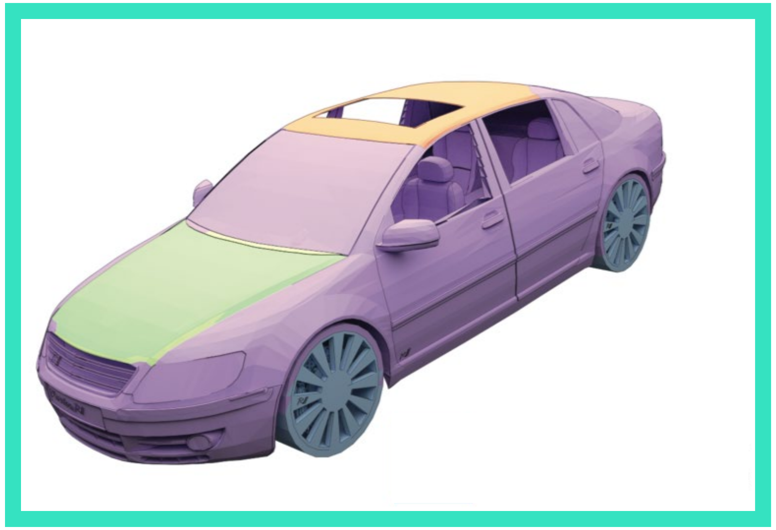


Output:
Hierarchical Segmentation &
Part Labeling

Observations — Abundant Shapes

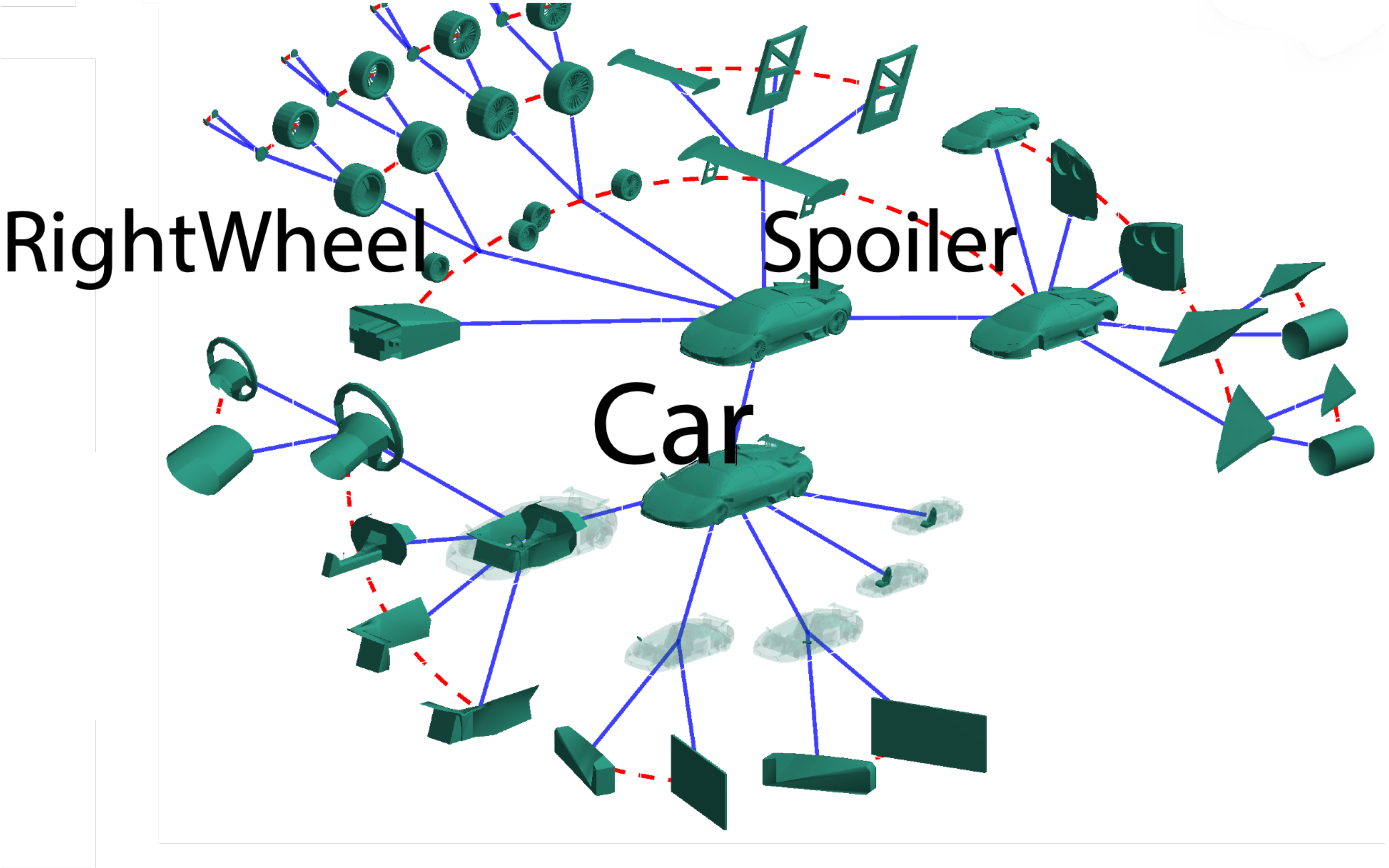
How to Define Parts?

- Previously expert defined

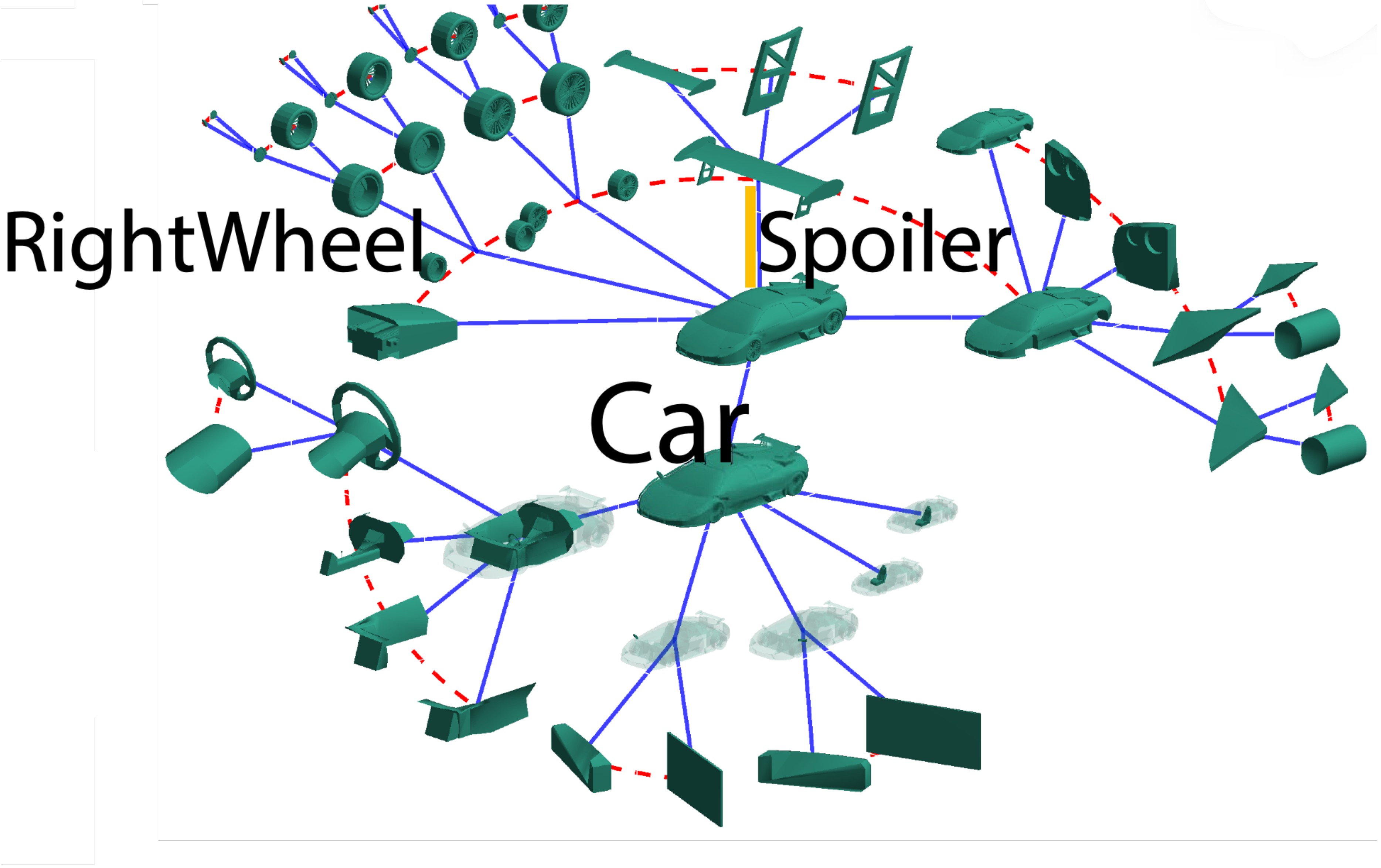


- Can consistent part knowledge emerge from multiple noisy hierarchies talking to each other
- Can we exploit this freely-available metadata to analyze and annotate new geometry?

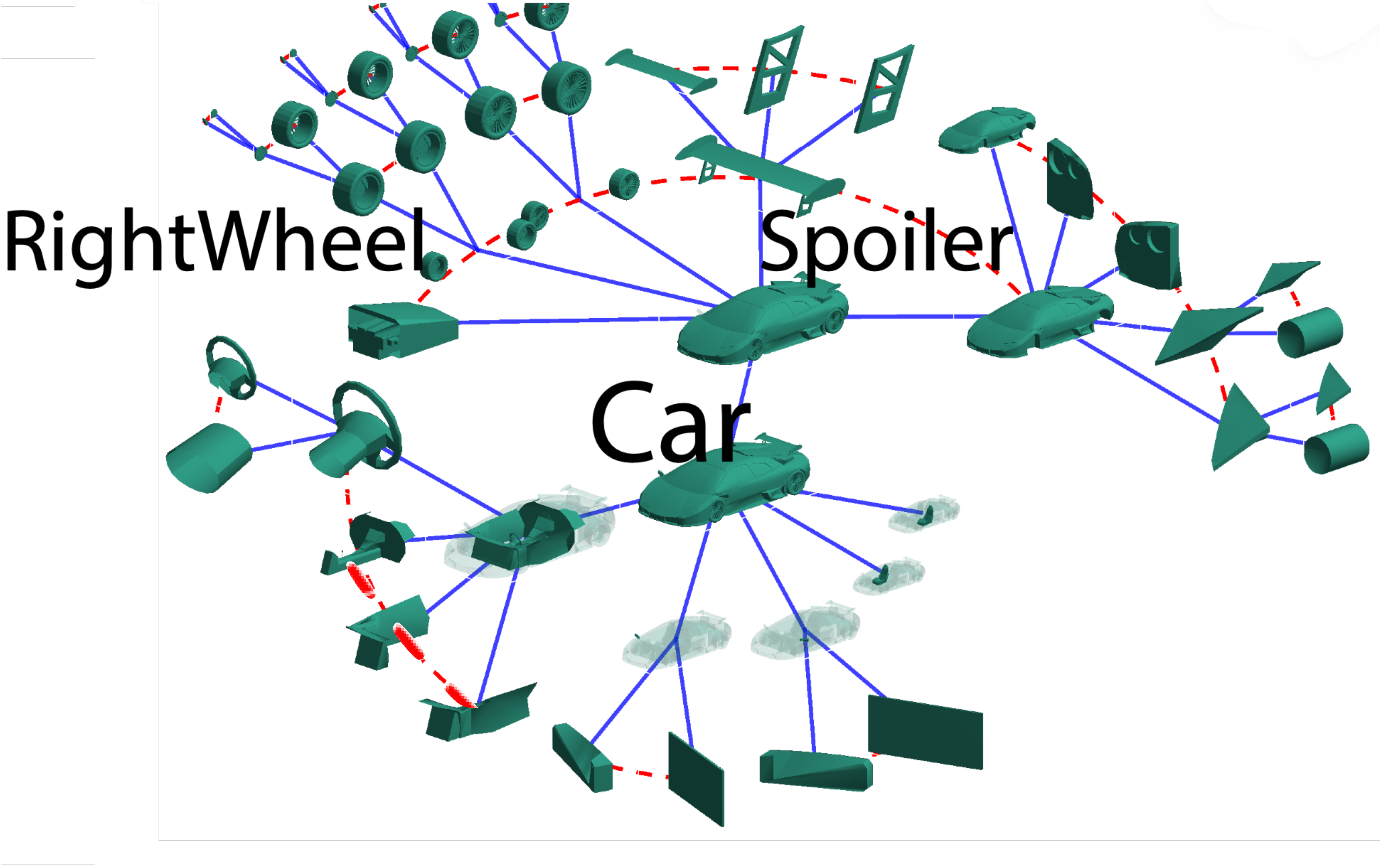
Object Graph



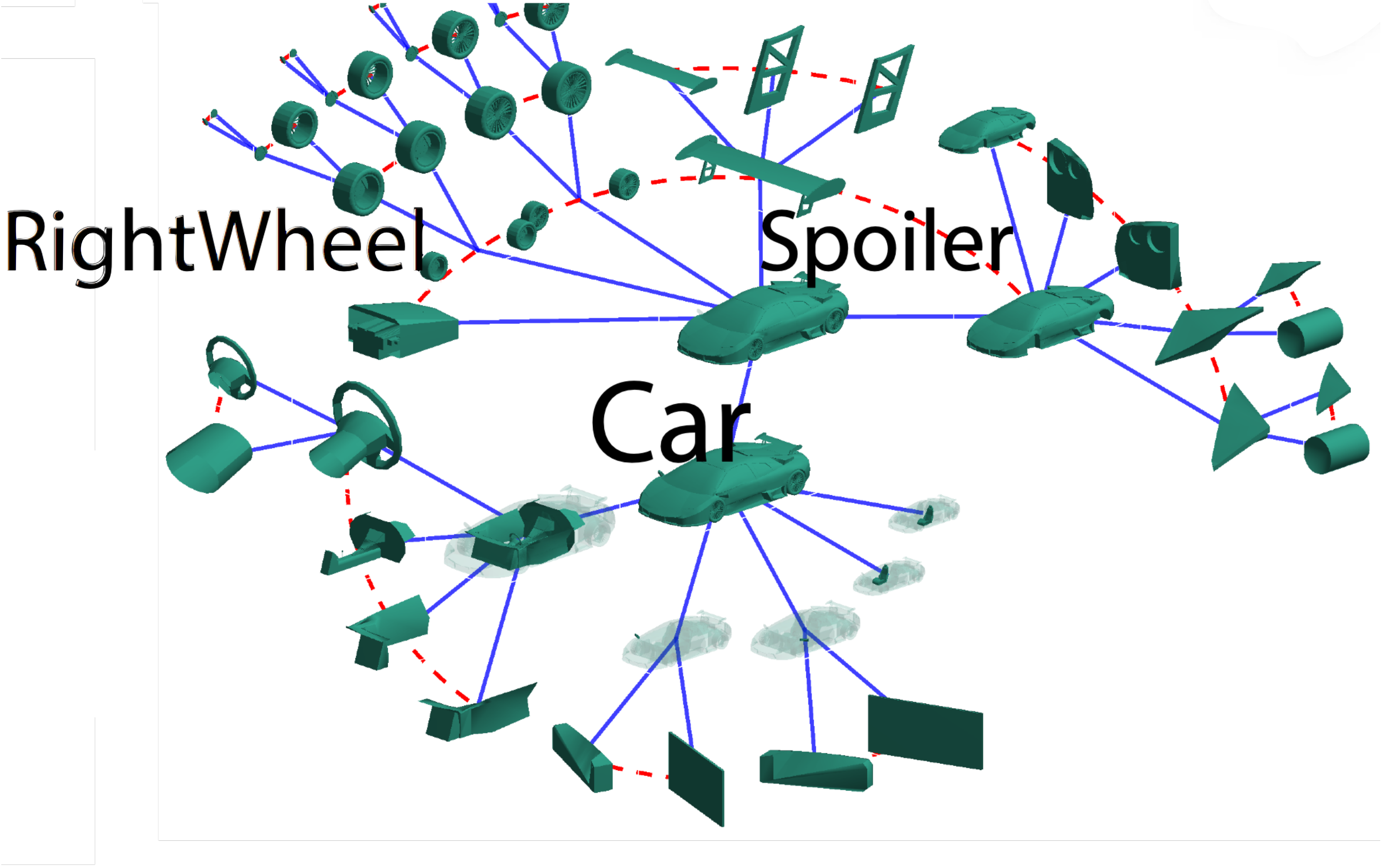
Object Graph



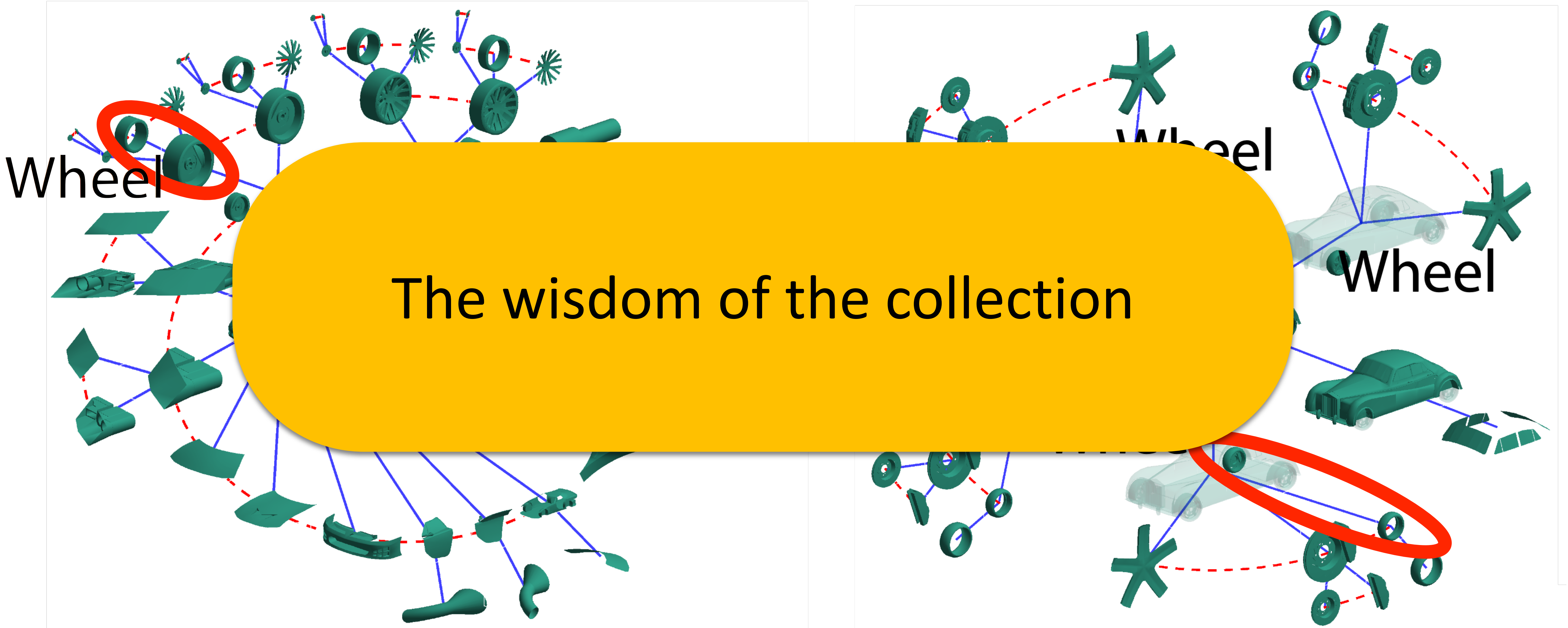
Object Graph



Object Graph



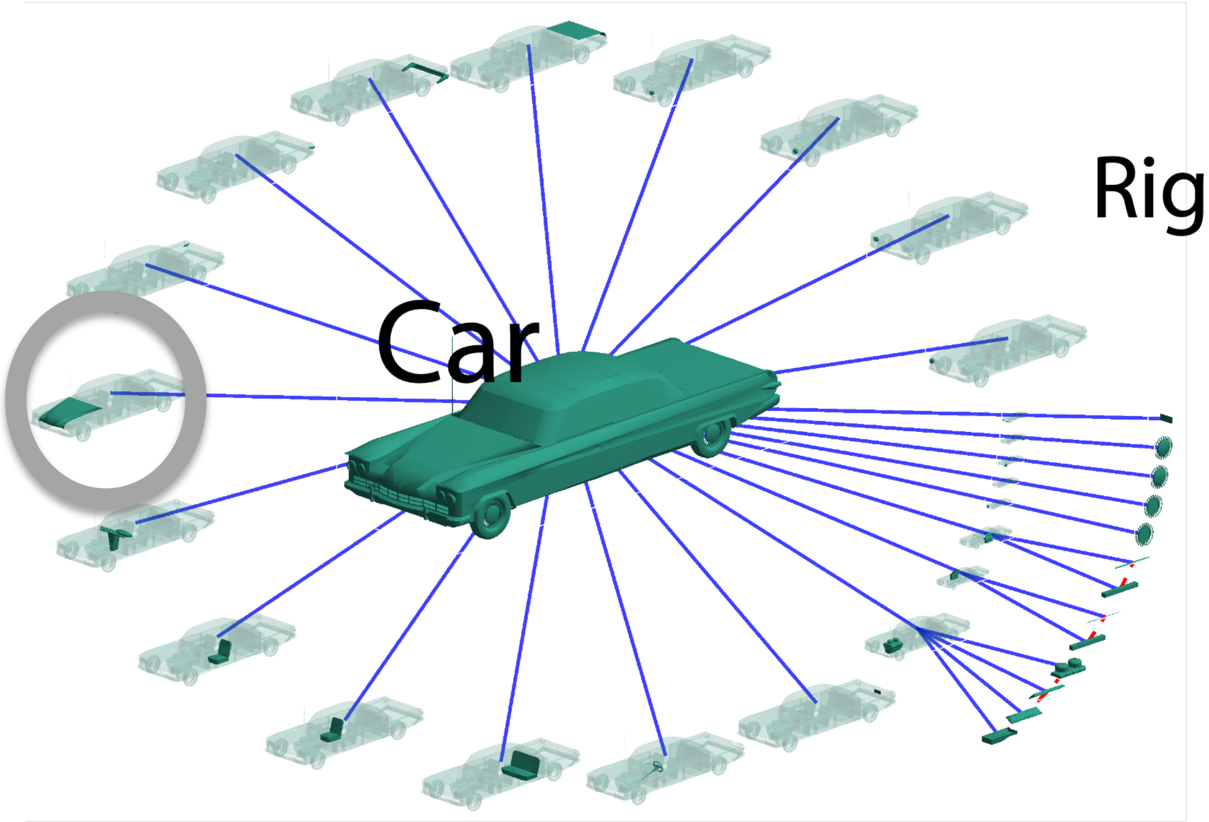
Common Structures in Object Graphs

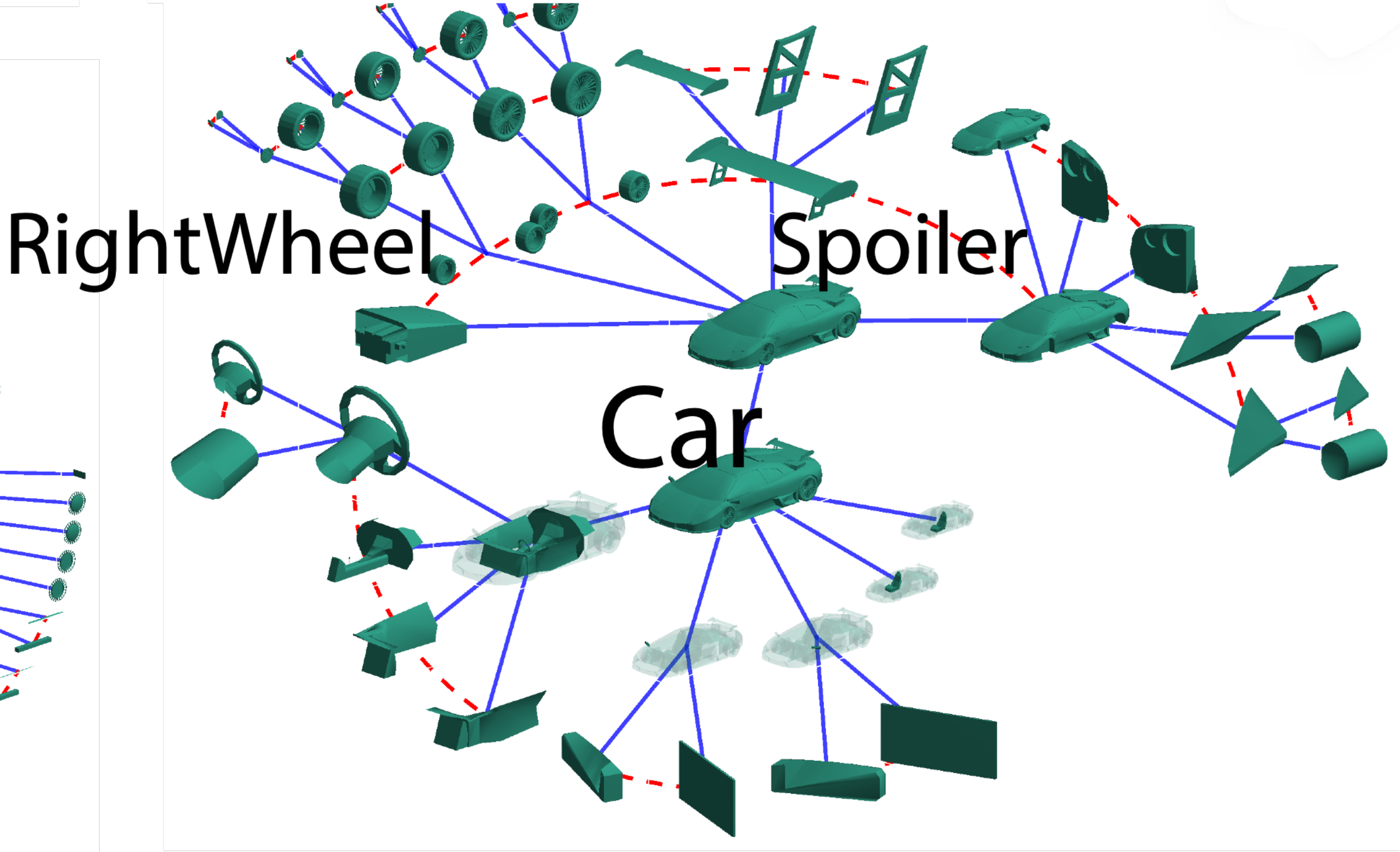


The wisdom of the collection

Challenges - Heterogeneous Data

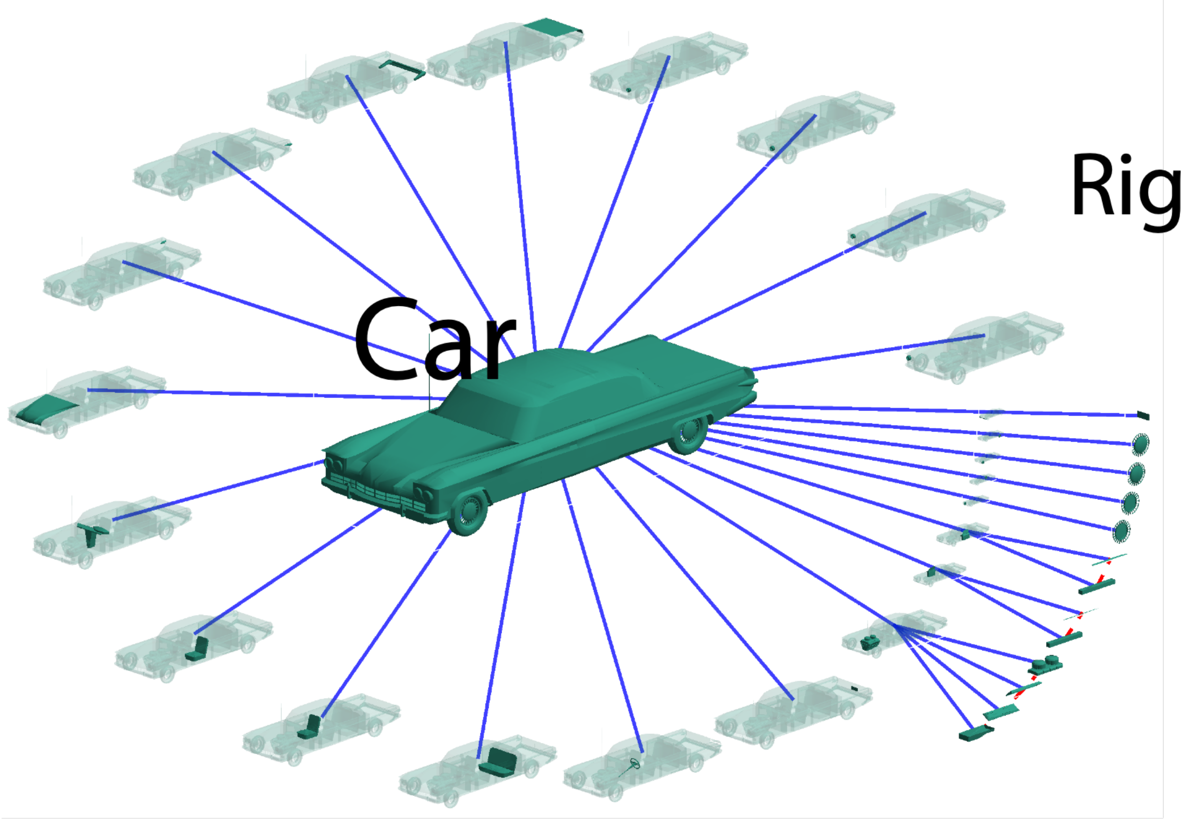
Different Parts

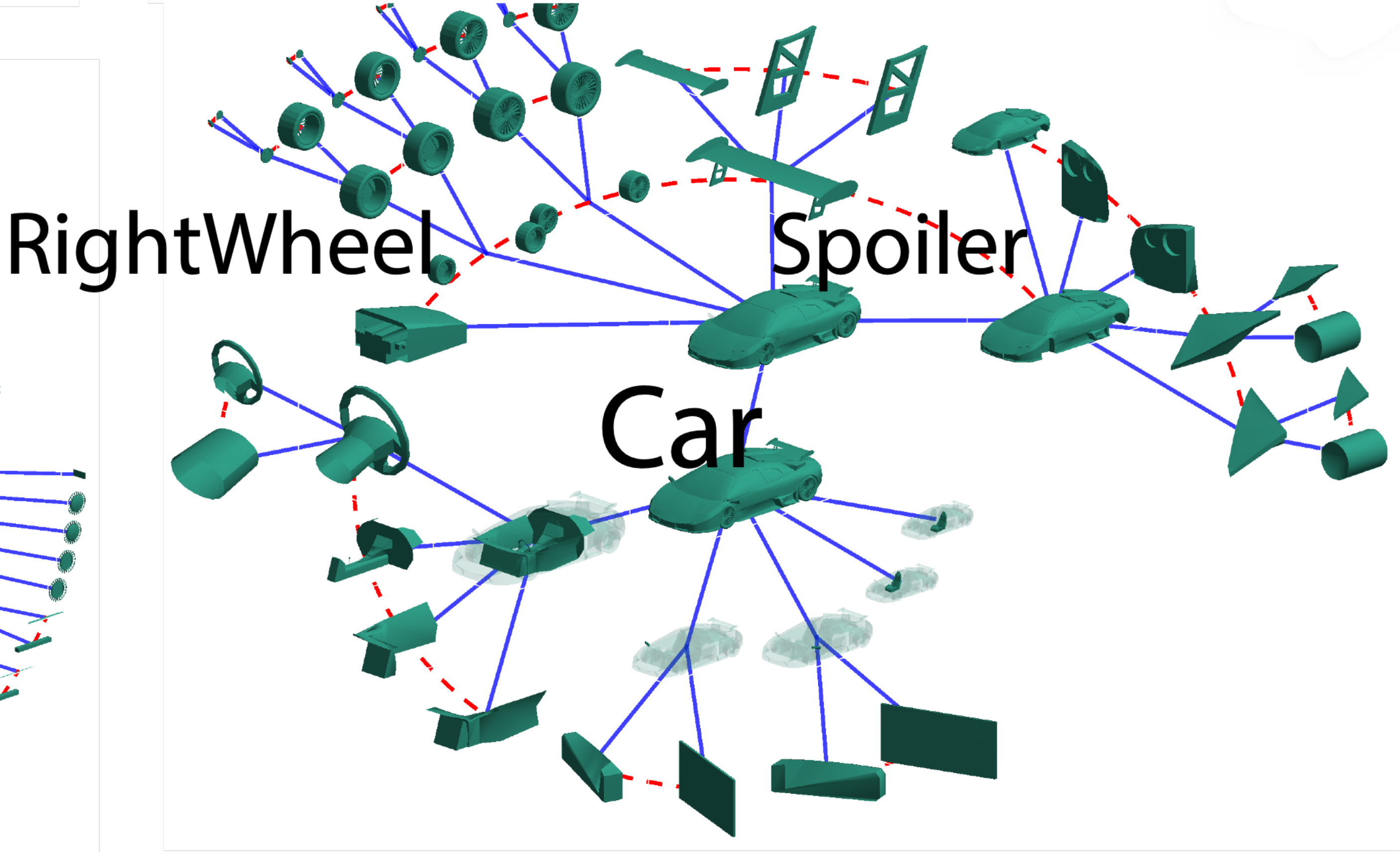




Challenges - Heterogeneous Data

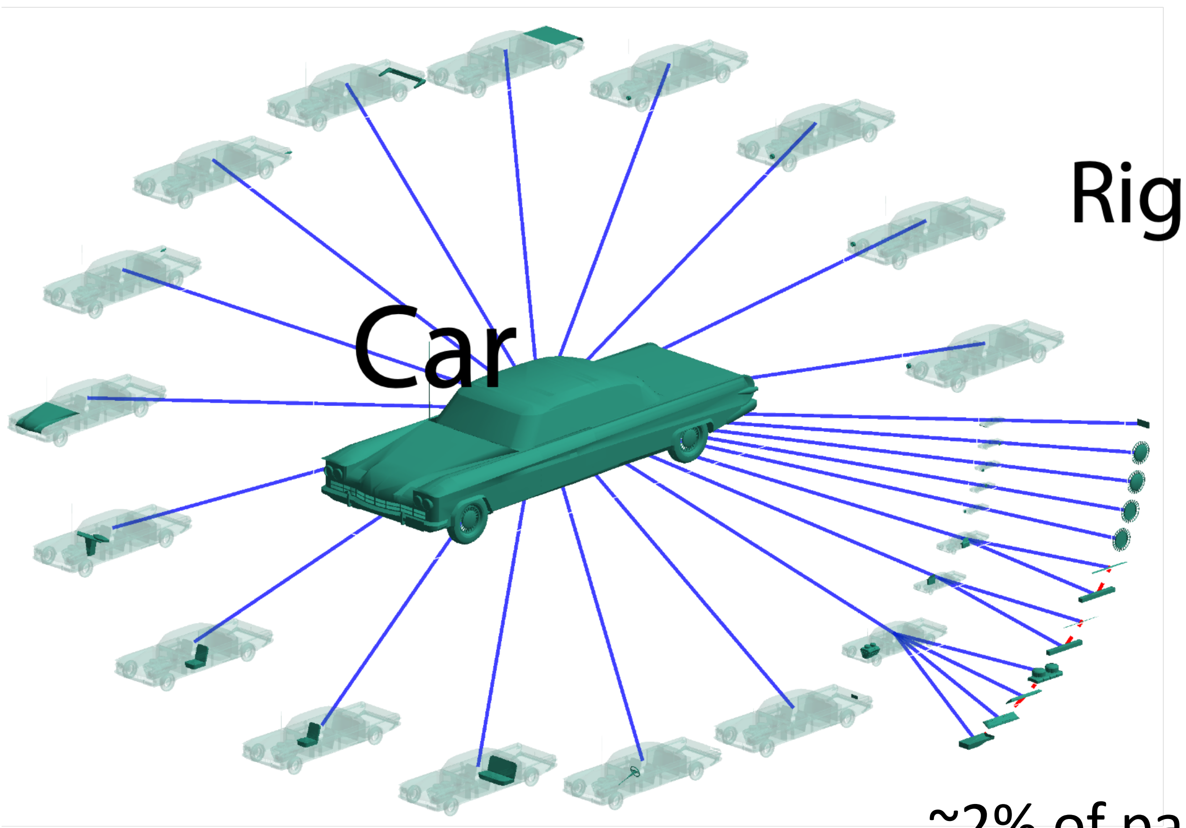
Different Parts, Different Hierarchy

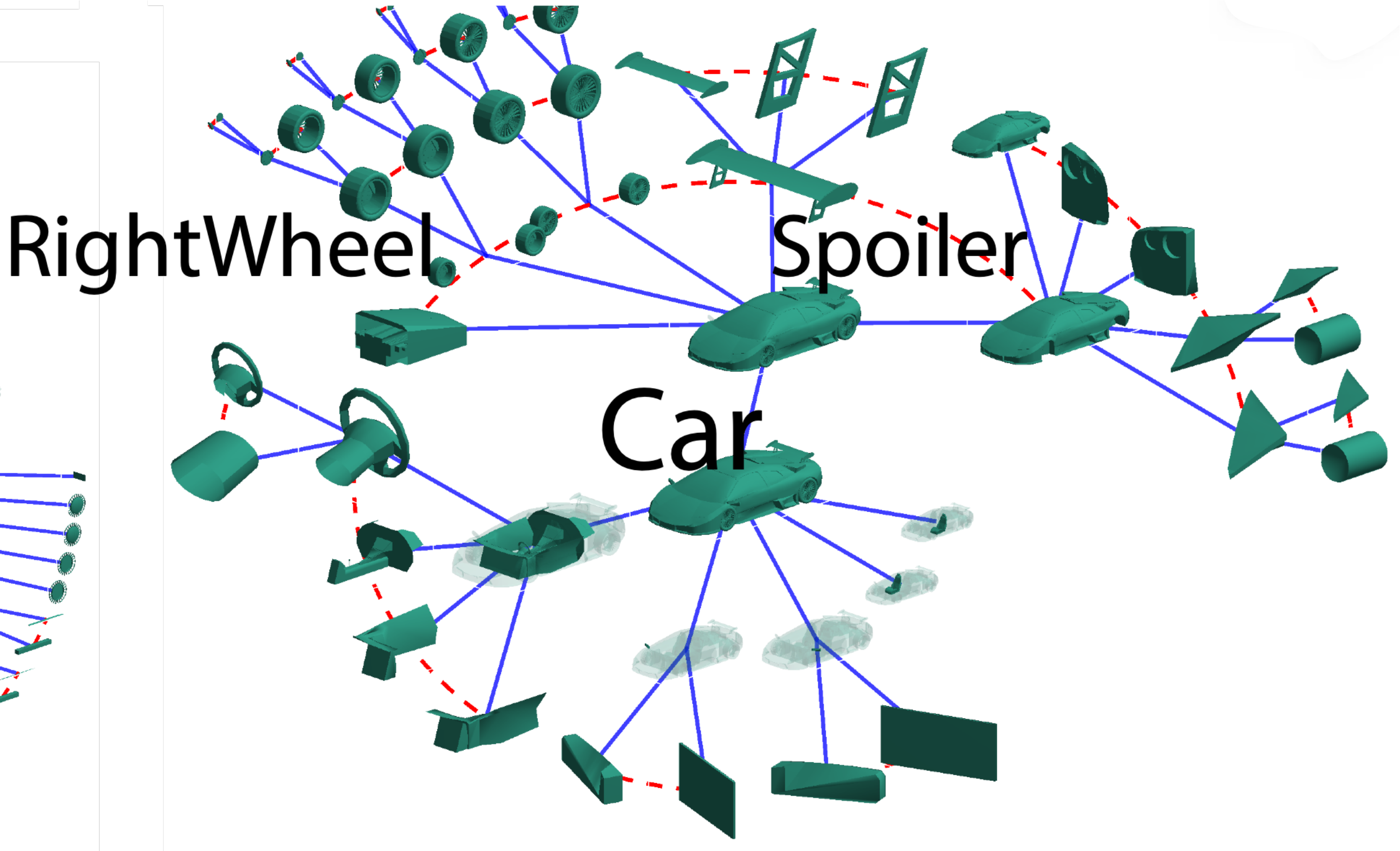




Challenges - Heterogeneous Data

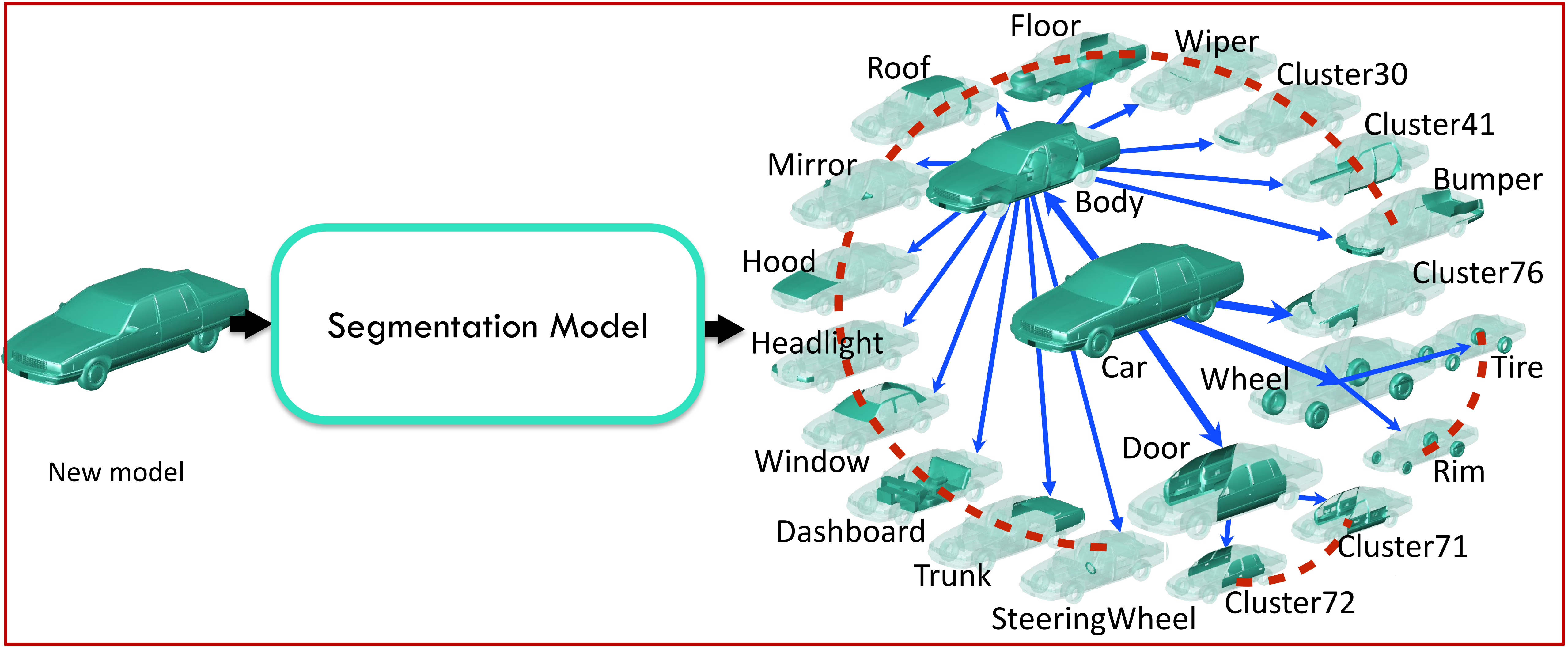
Different Parts, Different Hierarchy, Sparse Tagging





~2% of parts are tagged

Train a Network that Can Segment and Label

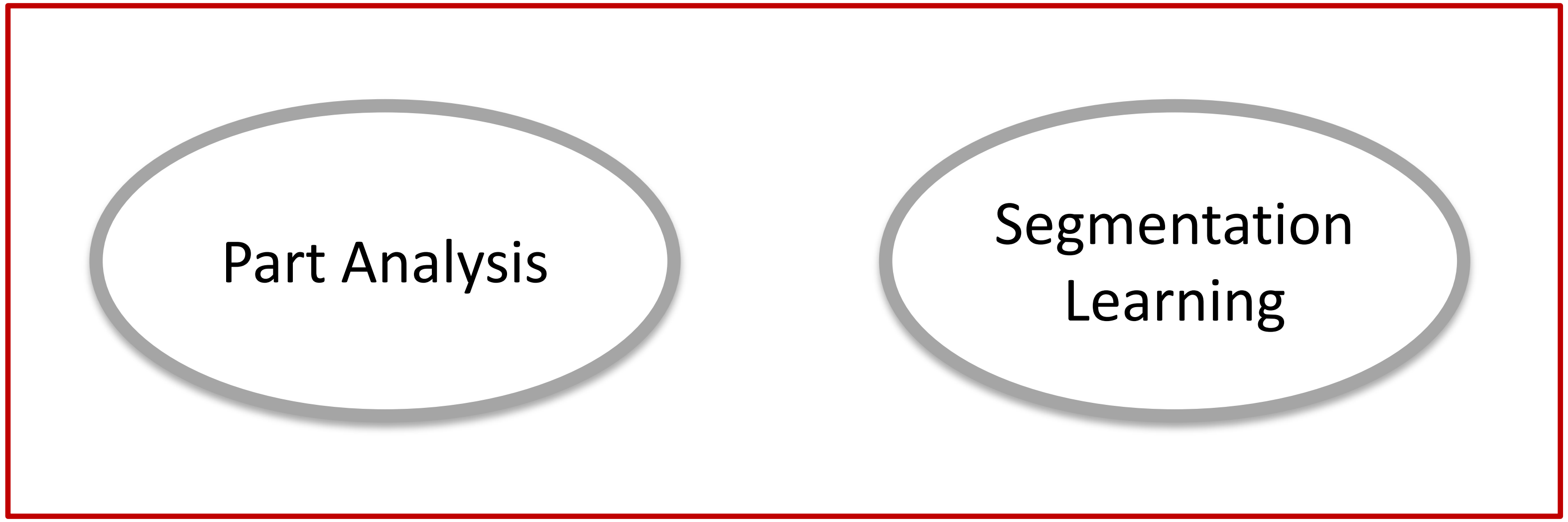


Approach Overview

Training Stage

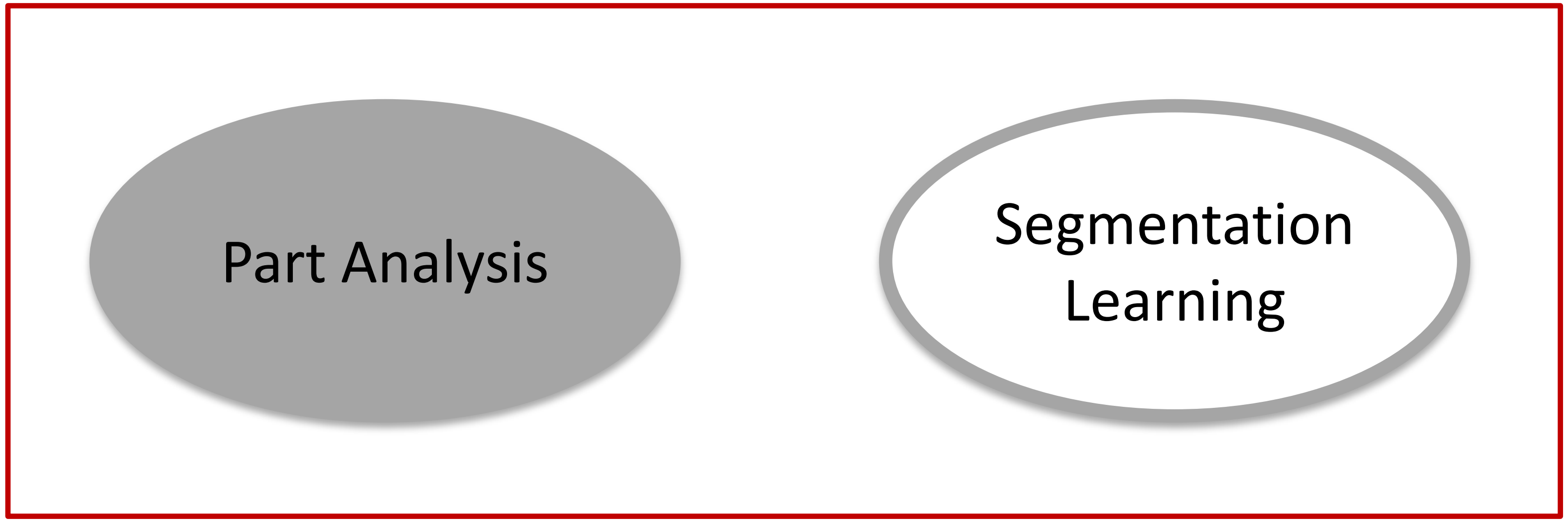
Inference Stage

Approach Overview



Training Stage

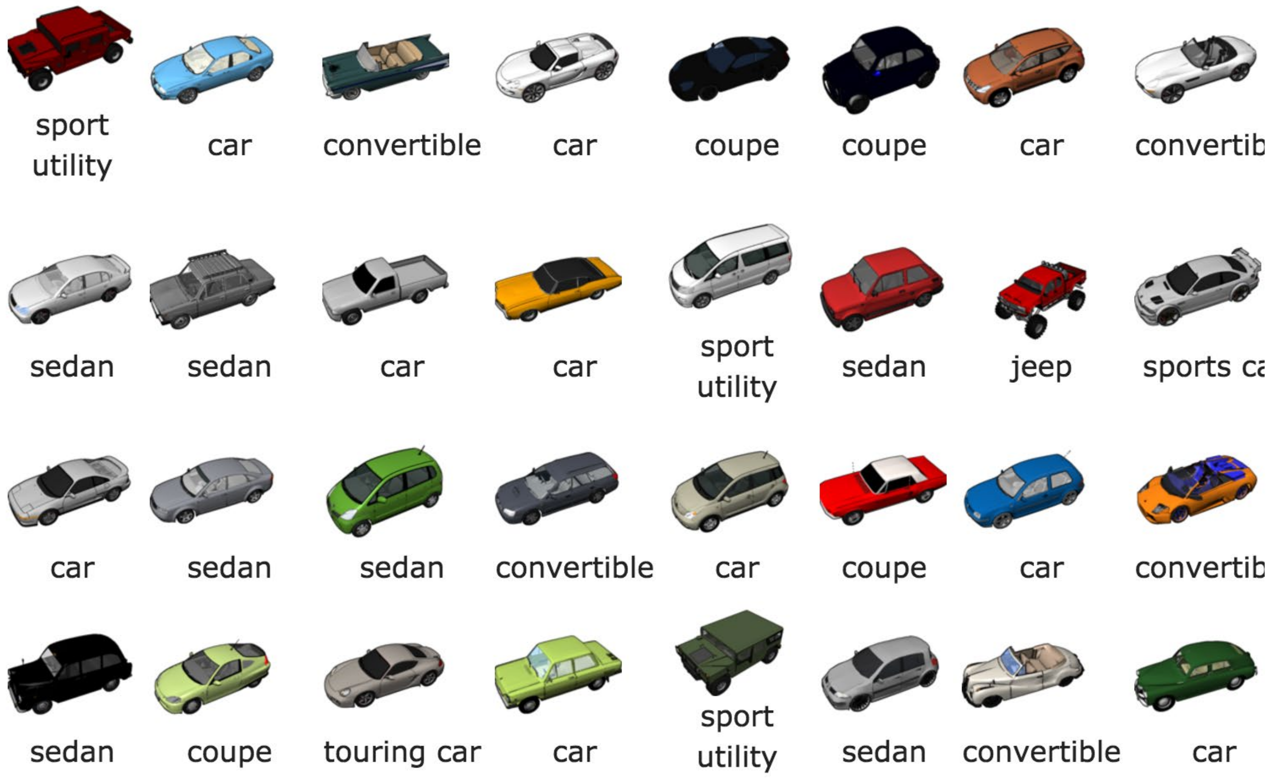
Approach Overview



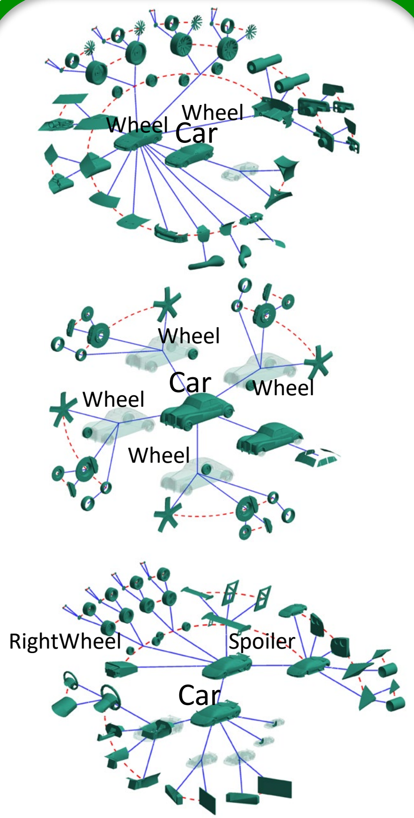
Training Stage

Observations — Abundant Shapes

SHAPENET



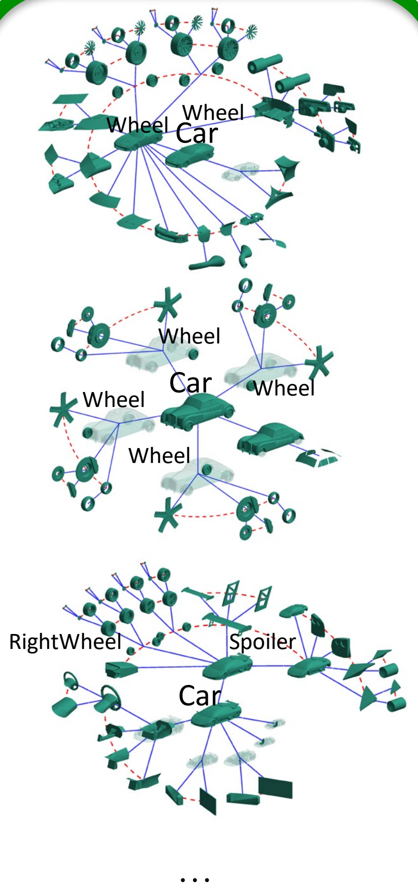
Part Analysis



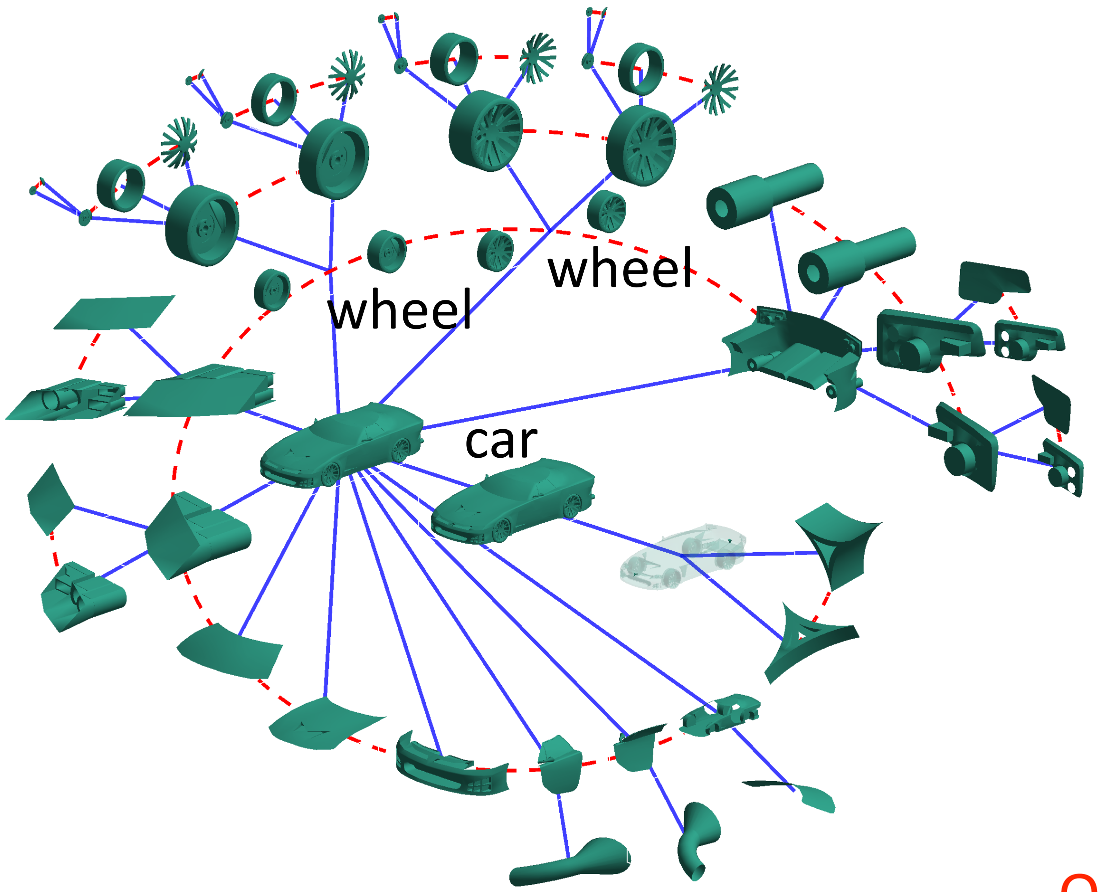
…

Input

Part Analysis

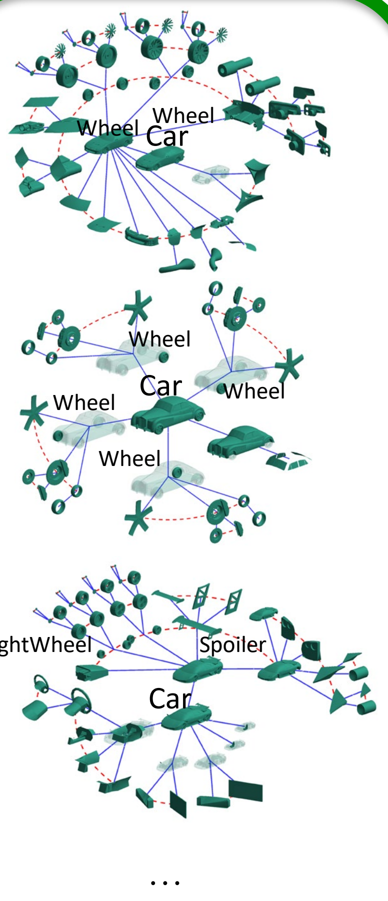


Input

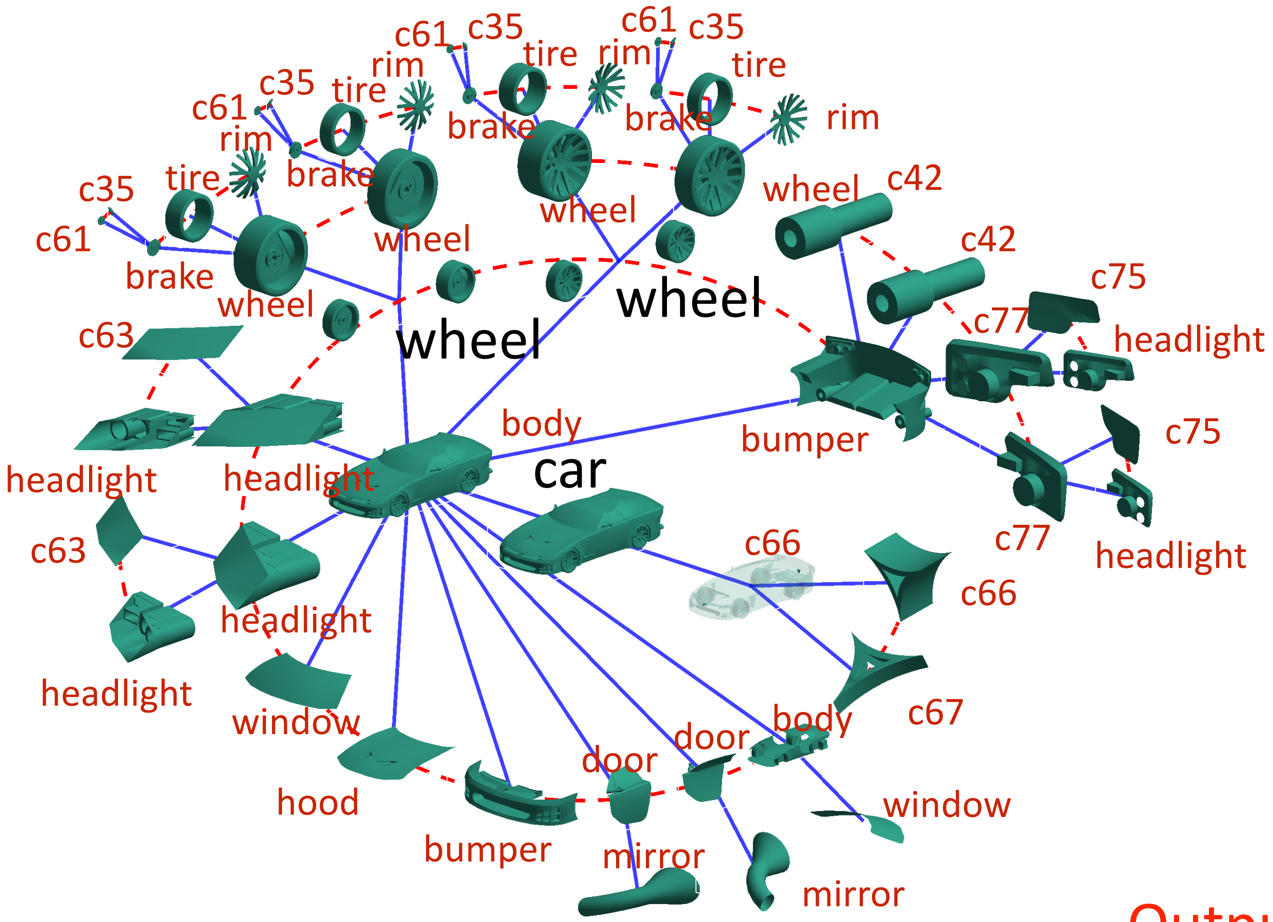


Output

Part Analysis

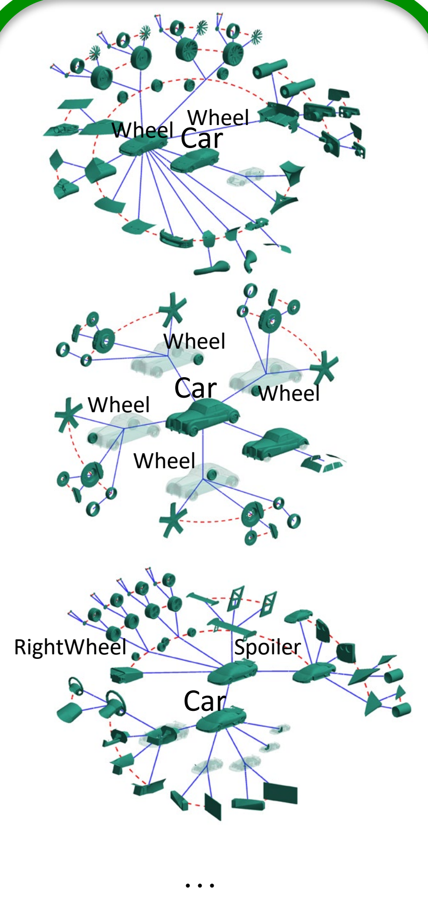


Input

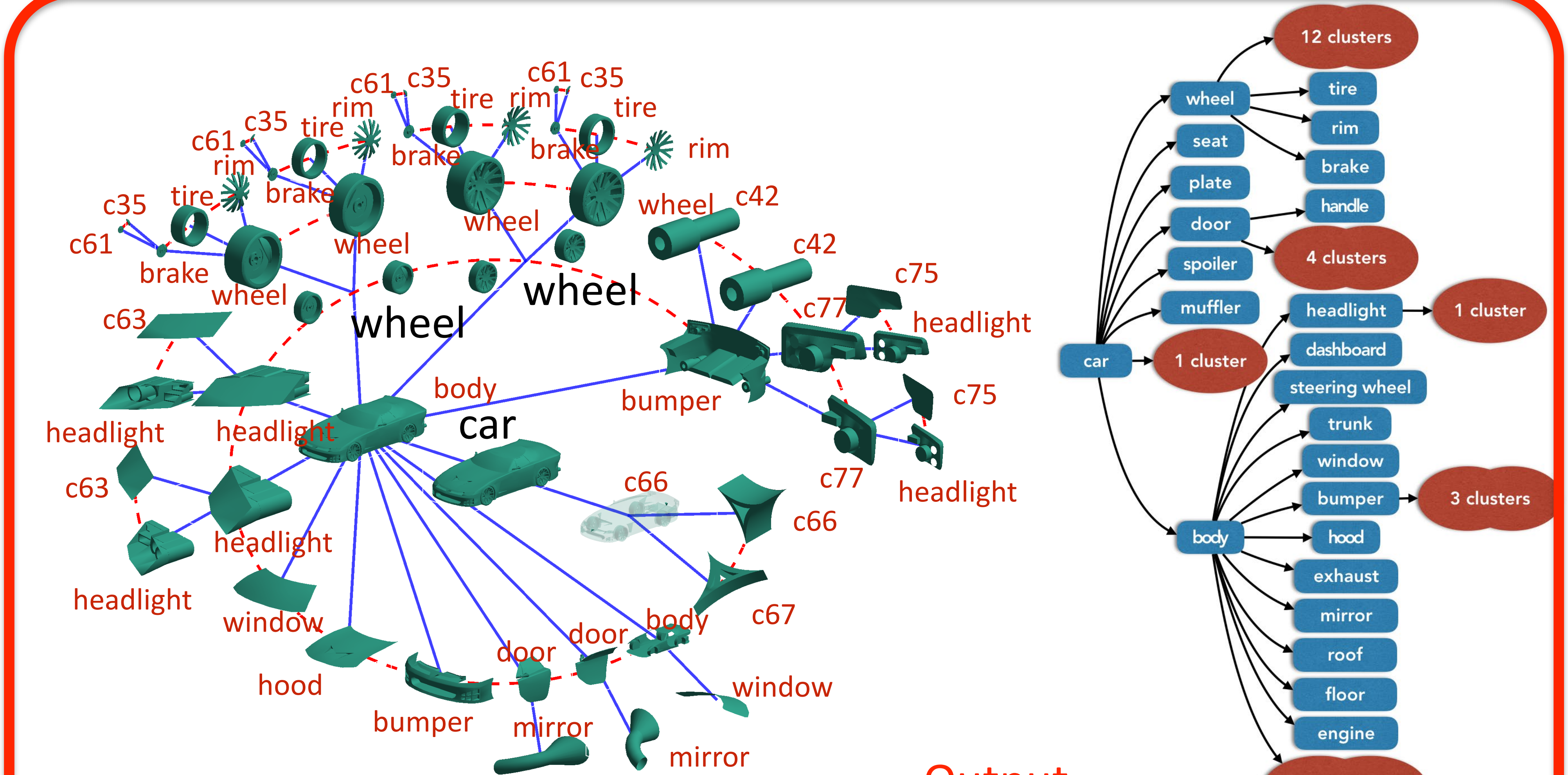


Output

Part Analysis



Input



Output

Preprocessing

- Gather all models from one category (e.g., “cars”)

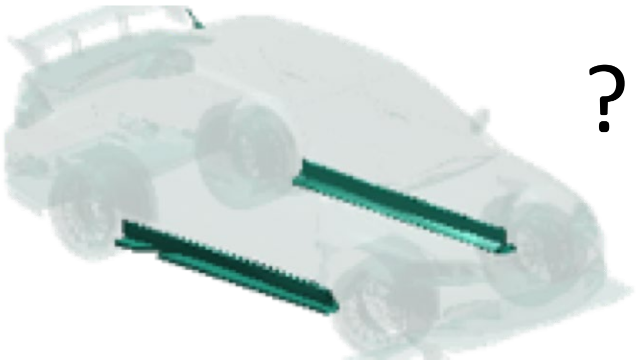
ShapeNet   Chang et al. 2015

Preprocessing

- Gather all meshes from one category (e.g., “cars”)
- Build manual vocabulary from common names
  e.g., `left_wheel18, rueda, wheeel` -> `wheel`

Preprocessing

- Gather all meshes from one category (e.g., “cars”)
- Build manual vocabulary from common names
  e.g., `left_wheel18, rueda, wheeel` -> `wheel`
- Notice the vocabulary is not comprehensive



?

Key Idea — Semi-Supervised Clustering

part geometry feature

$$\boxed{\mathbf{y} = f(\mathbf{x})}$$

part embedding

Key Idea — Semi-Supervised Clustering

part geometry feature

$$\mathbf{y} = f(\mathbf{x})$$

part embedding

cluster centroid

$$\text{clustering: } \min_i \| \mathbf{c}_i - f(\mathbf{x}) \|$$

Key Idea — Semi-Supervised Clustering

part geometry feature

$$\mathbf{y} = f(\mathbf{x})$$

part embedding

cluster centroid

$$\text{clustering: } \min_i \| \mathbf{c}_i - f(\mathbf{x}) \|$$

Key Idea — Semi-Supervised Clustering

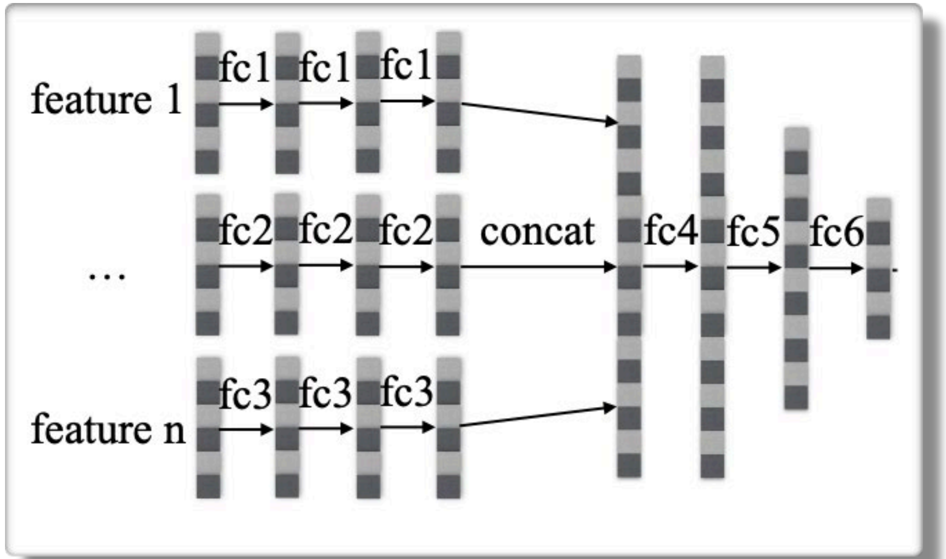
part geometry feature

$$\mathbf{y} = f(\mathbf{x})$$

part embedding

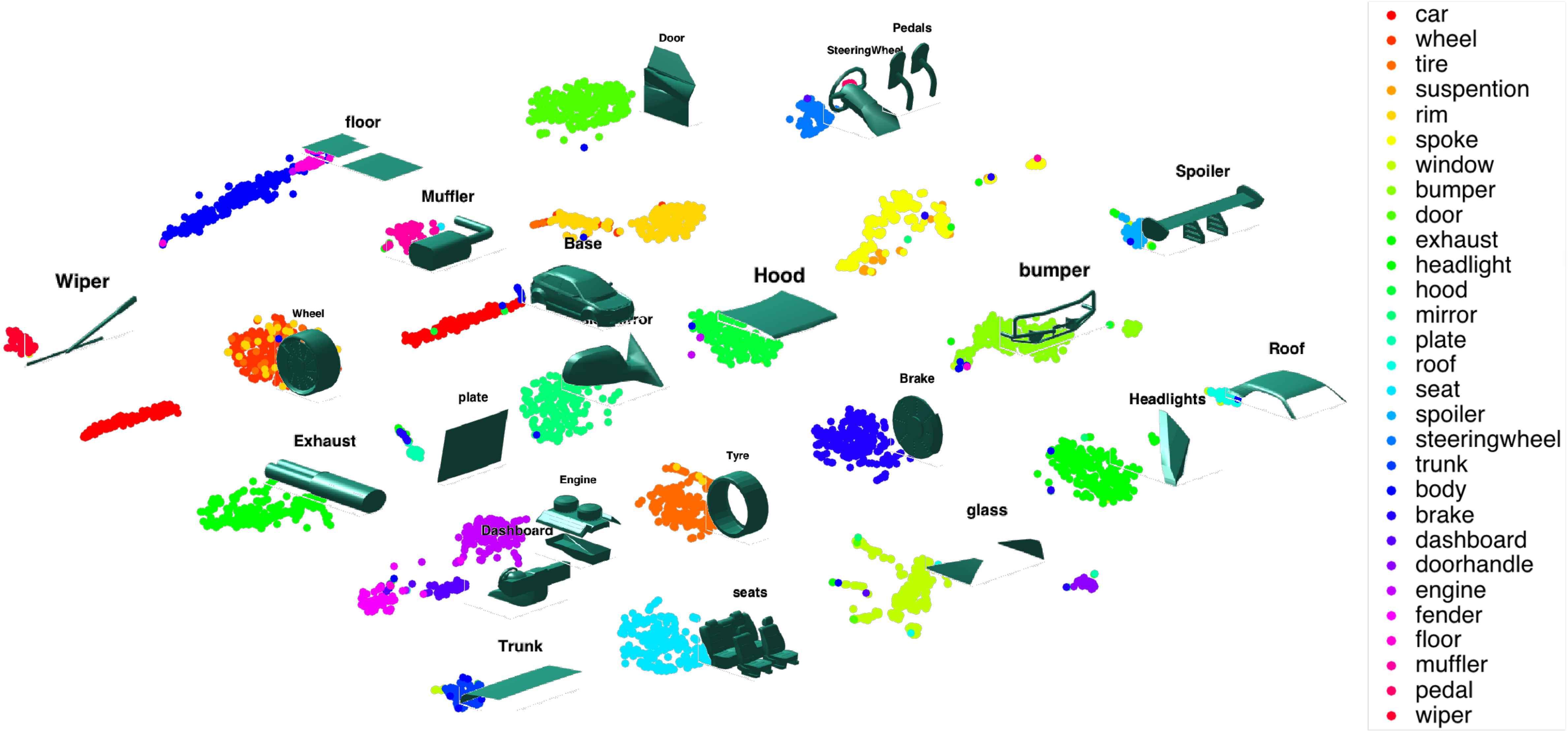
cluster centroid

clustering: $\min_i \| \mathbf{c}_i - f(\mathbf{x}) \|$



Supervision: sparse tags, inconsistent hierarchies

Key Idea — Semi-Supervised Clustering



Objective Function

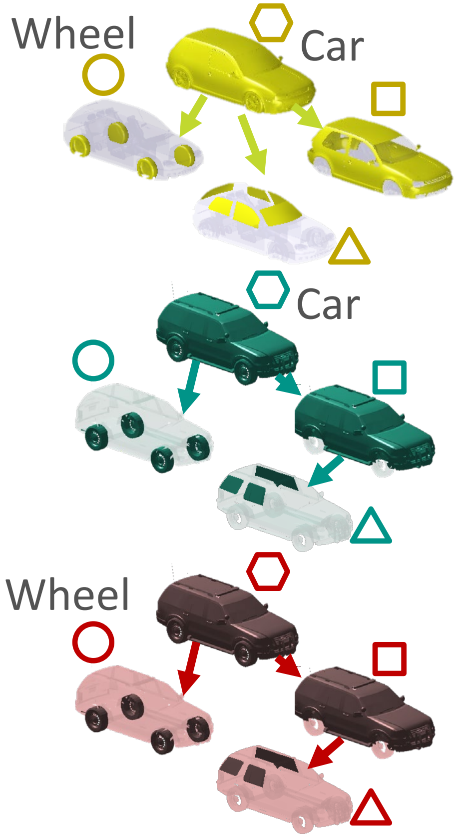
$$E(\theta, p, c, M) = \lambda_c E_c + \lambda_s E_s + \lambda_d E_d + \lambda_m E_m - H$$

Use an EM algorithm

c.f. Basu et al. KDD 2004

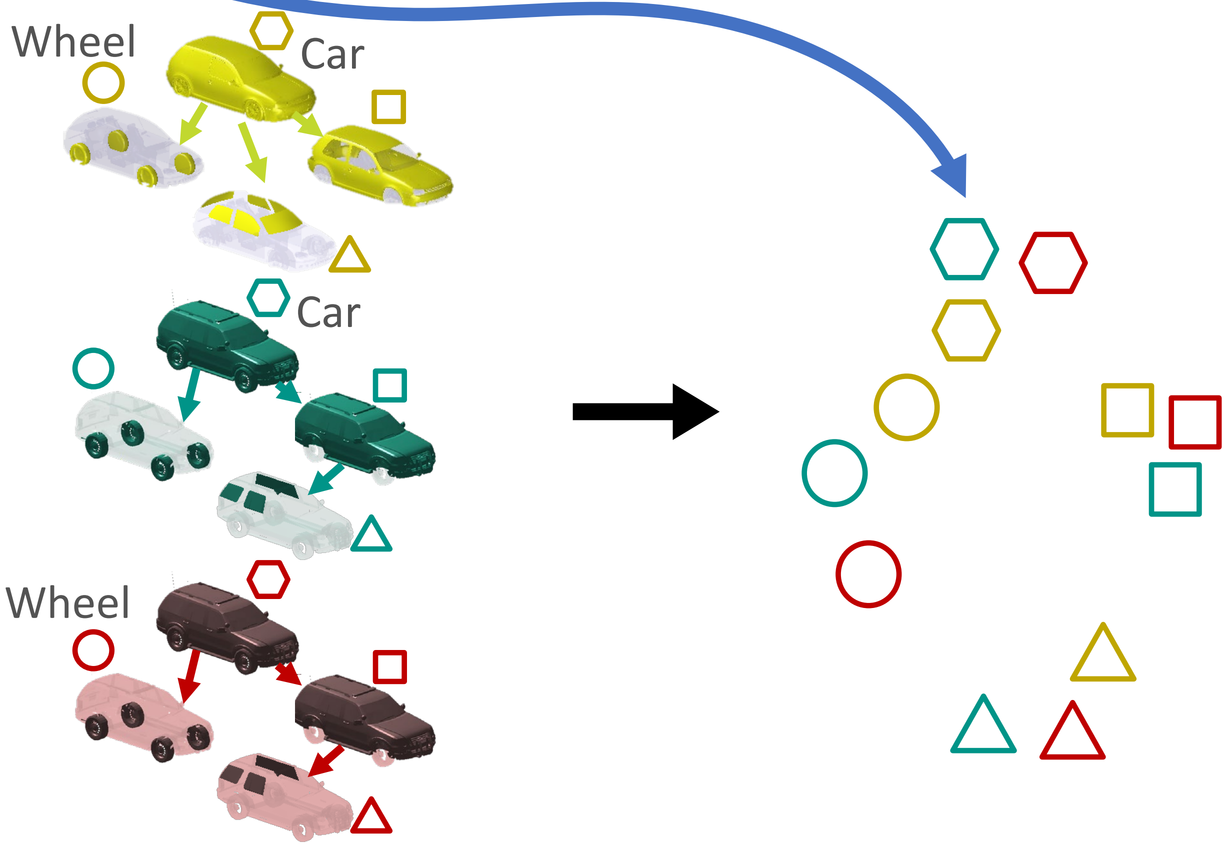
Objective Function

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



Objective Function

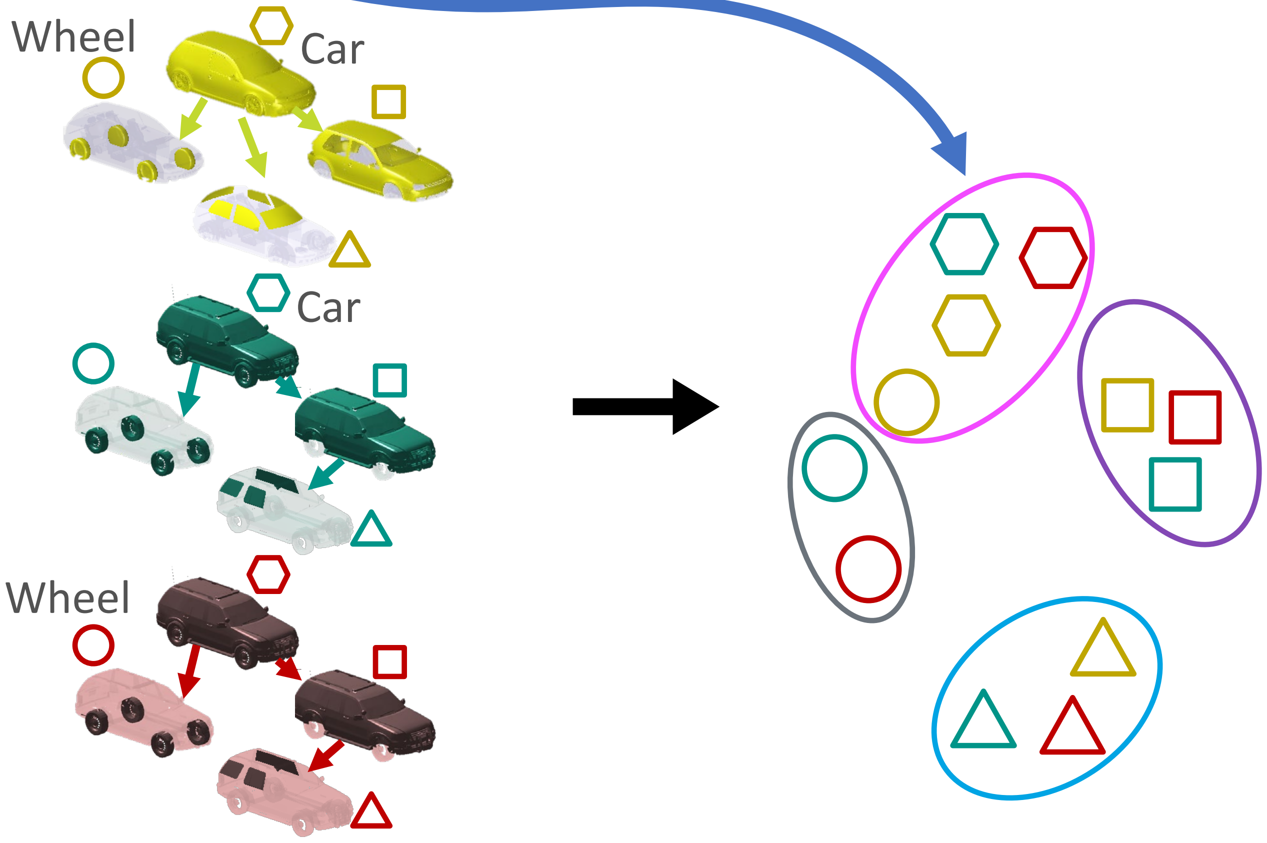
$$E(\theta, p, c, M) = \lambda_c E_c + \lambda_s E_s + \lambda_d E_d + \lambda_m E_m - H$$



Embedding parameters

Objective Function

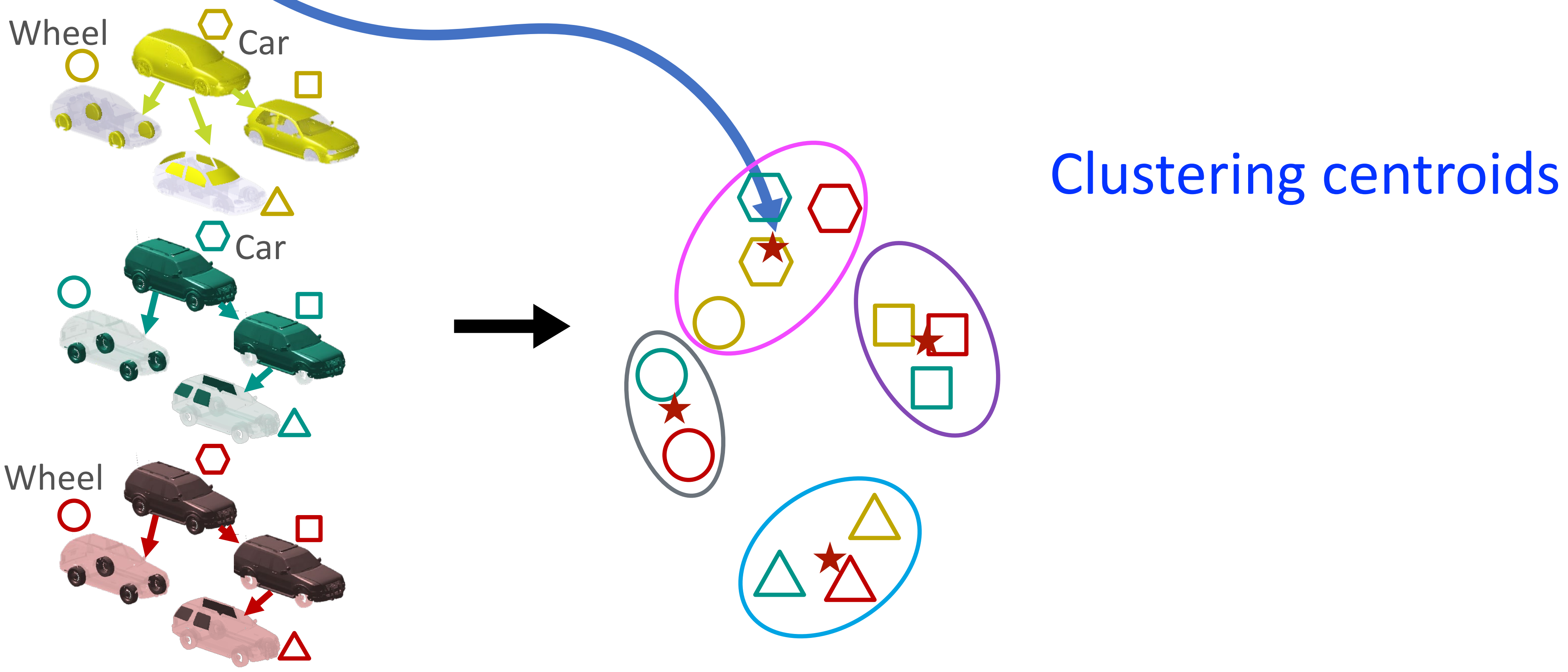
$$E(\theta, p, c, M) = \lambda_c E_c + \lambda_s E_s + \lambda_d E_d + \lambda_m E_m - H$$



Clustering labels

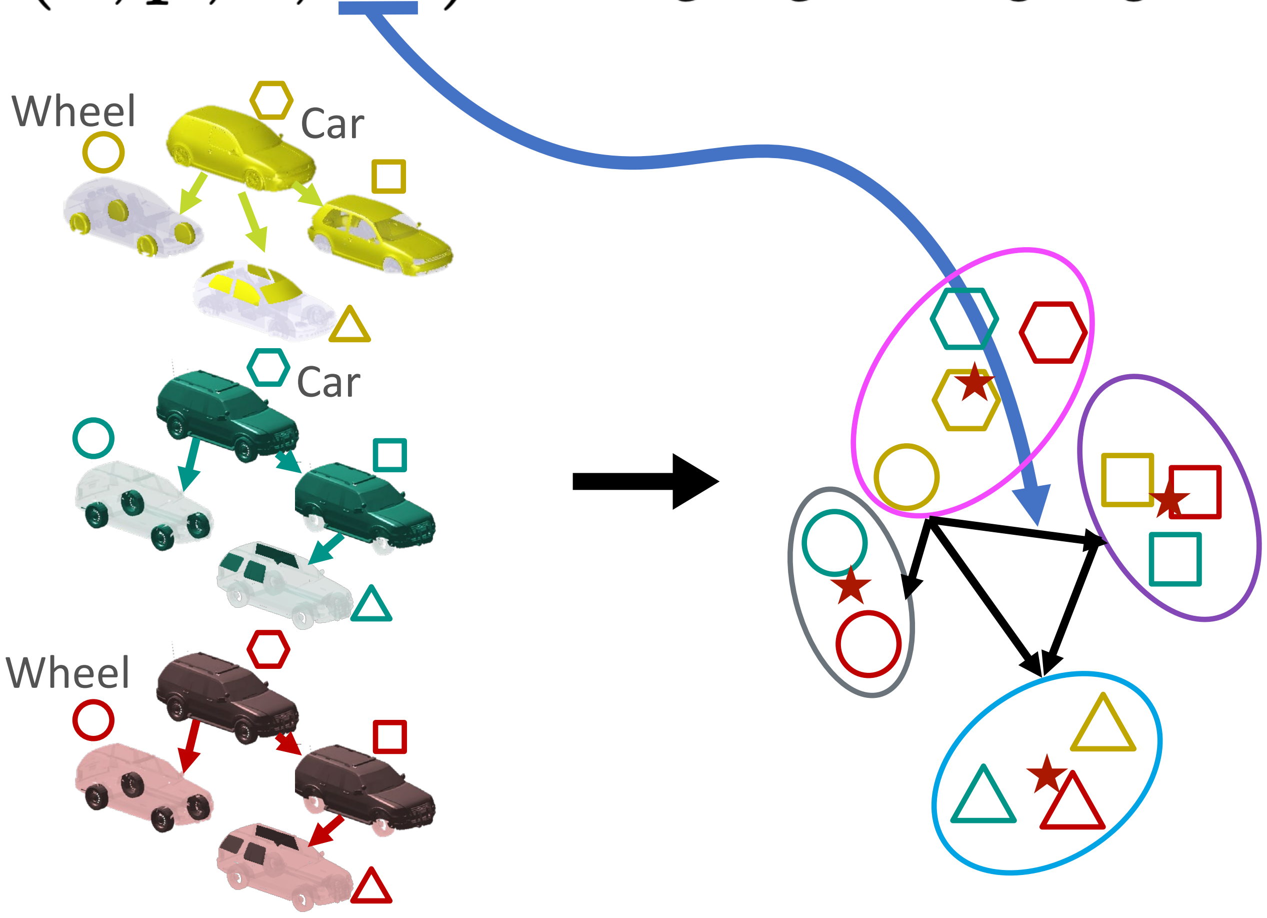
Objective Function

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



Objective Function

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

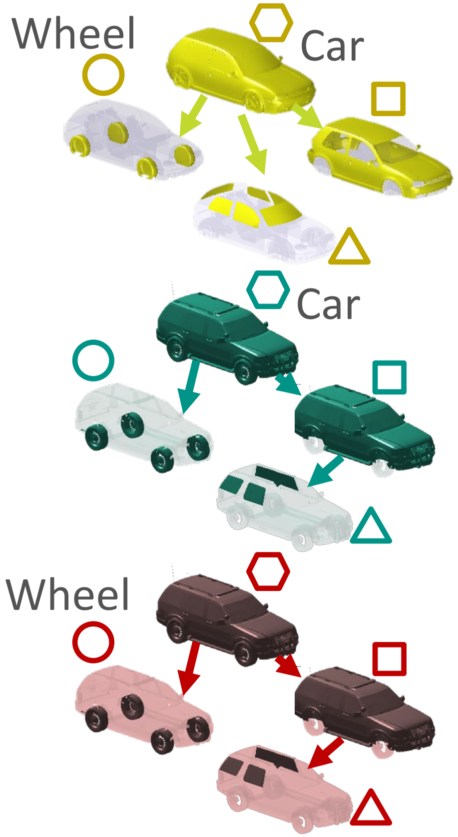


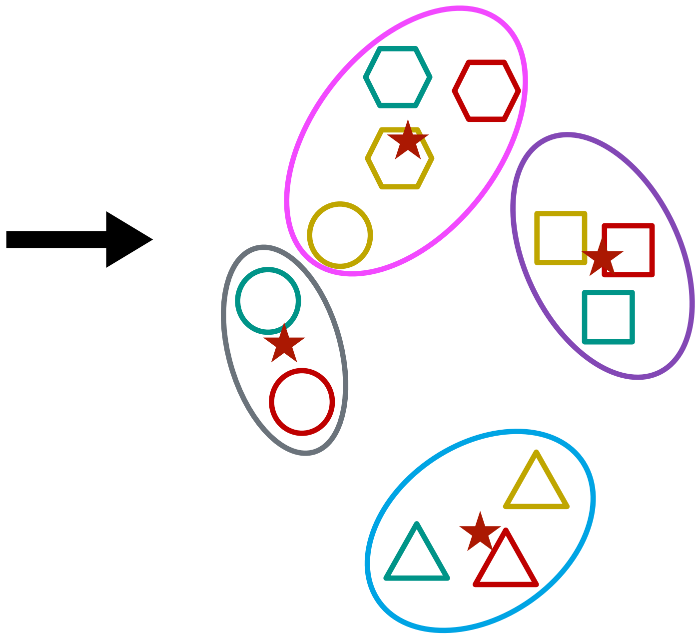
Soft hierarchy graph

Objective function

Clustering

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





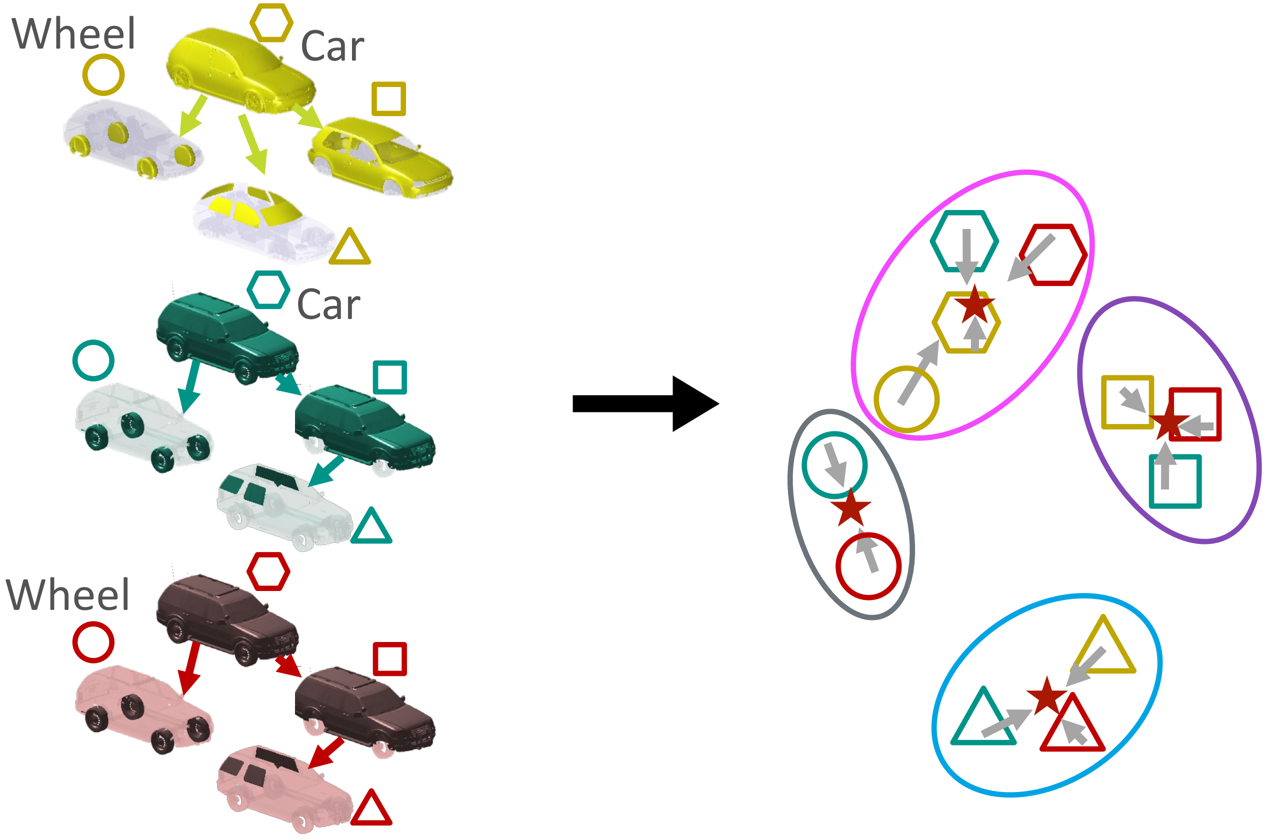
Clustering term:

Encourage parts to form clusters

Objective Function

Clustering

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



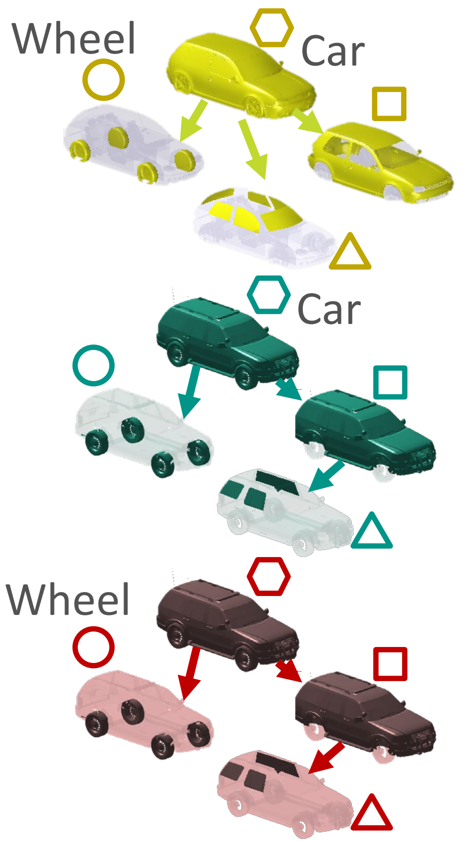
Clustering term:

Encourage parts to form clusters

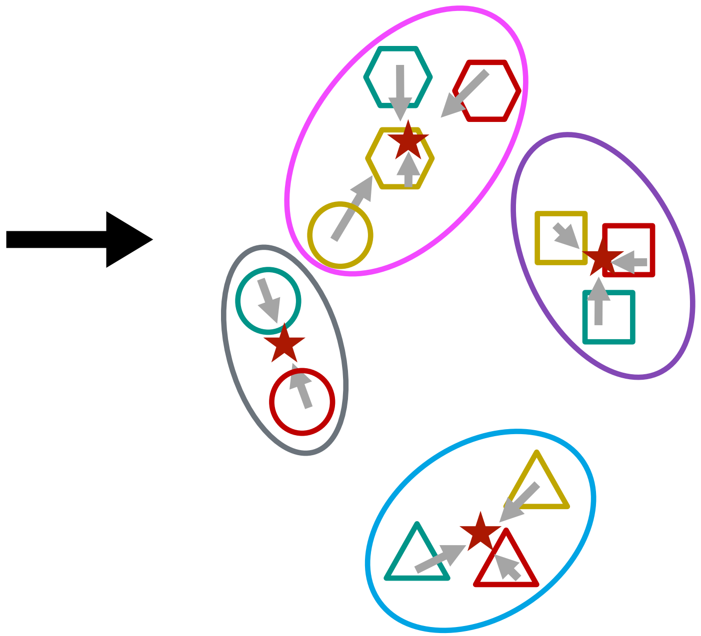
Objective Function

Clustering

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



$$E_c = \sum_{\text{parts}} \sum_{\text{centroids}} p_{\text{part,centroid}} ||f(\text{part}) - \text{centroid}||$$



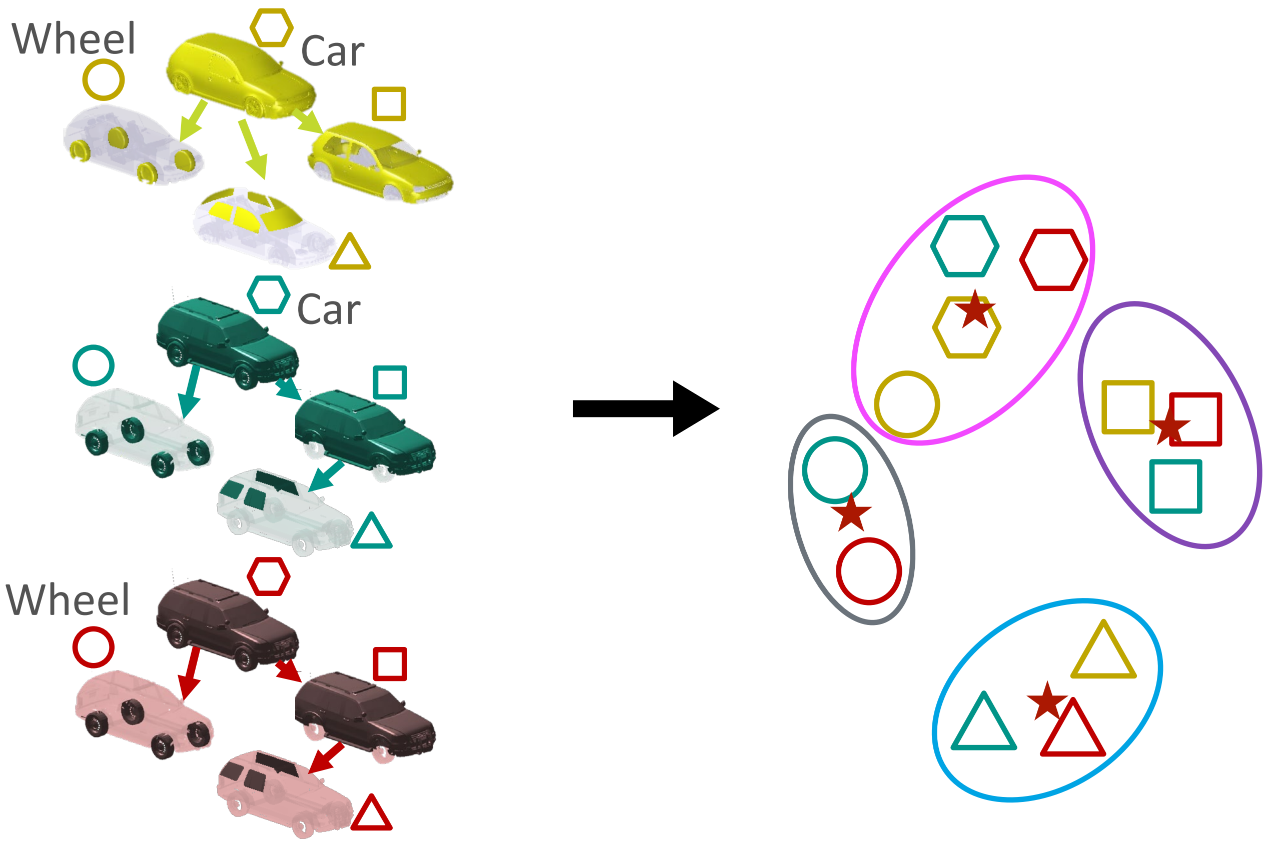
Clustering term:

Encourage parts to form clusters

Objective Function

Clustering Similarity

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



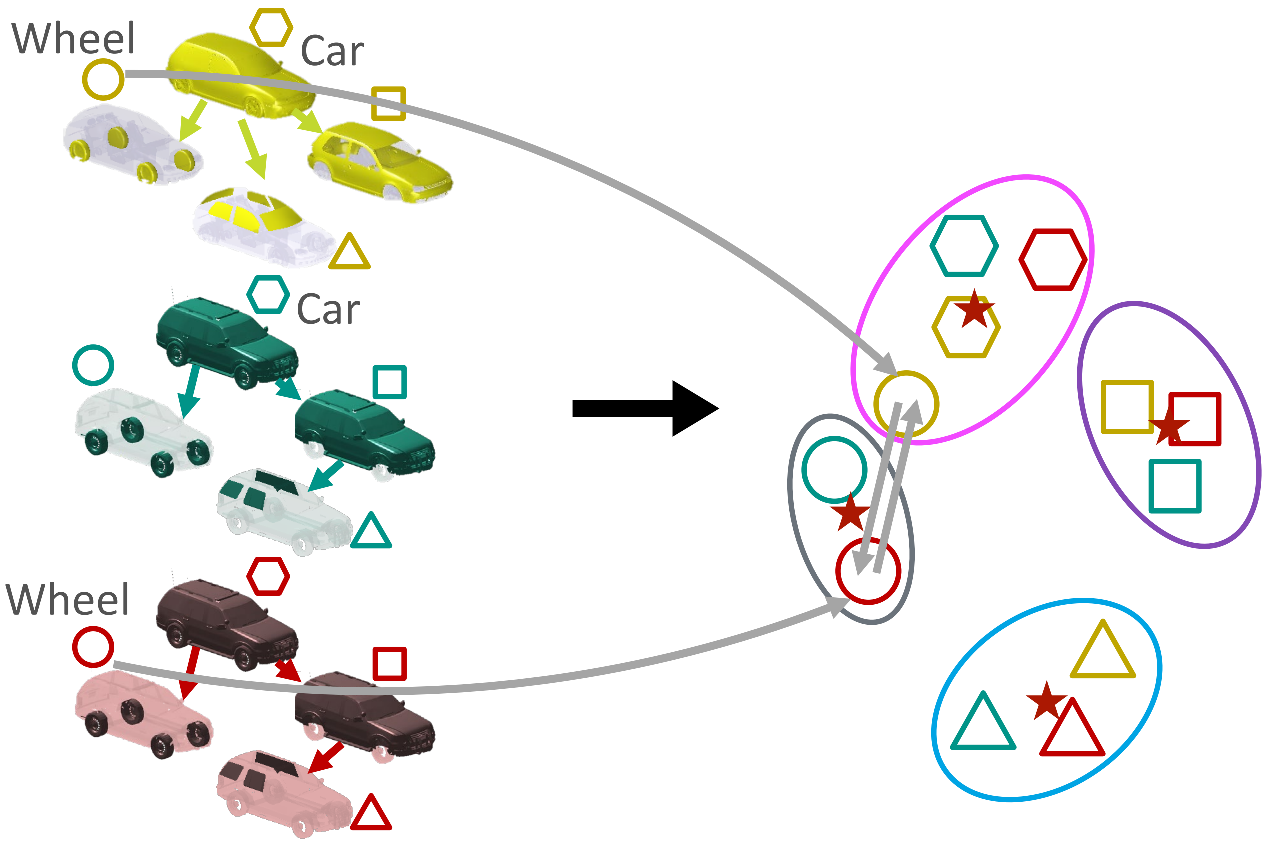
Similarity term:

Group parts with the same tag or geometry

Objective Function

Clustering Similarity

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



Similarity term:

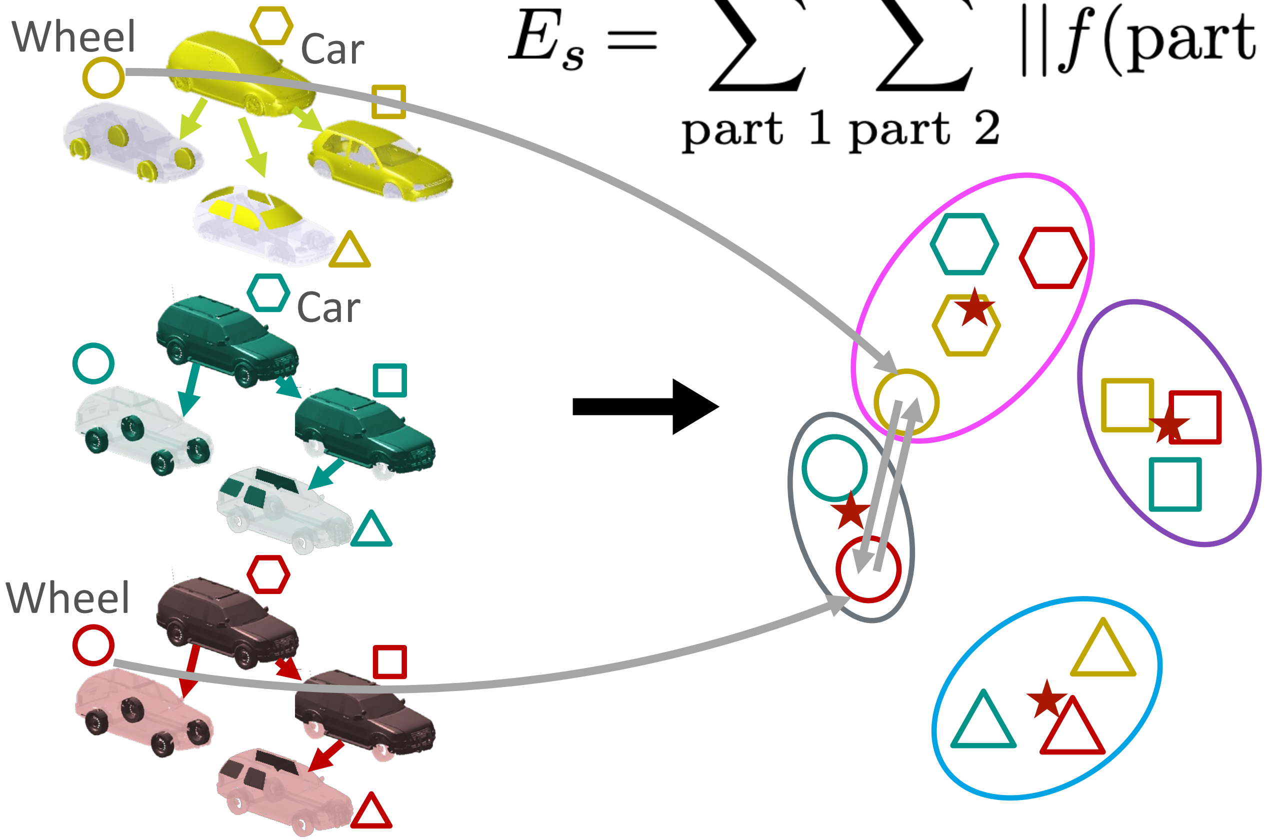
Group parts with the same tag or geometry

Objective Function

Clustering Similarity

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

$$E_s = \sum_{\text{part 1}} \sum_{\text{part 2}} ||f(\text{part 1}) - f(\text{part 2})||$$ iff almost identical or tags are the same



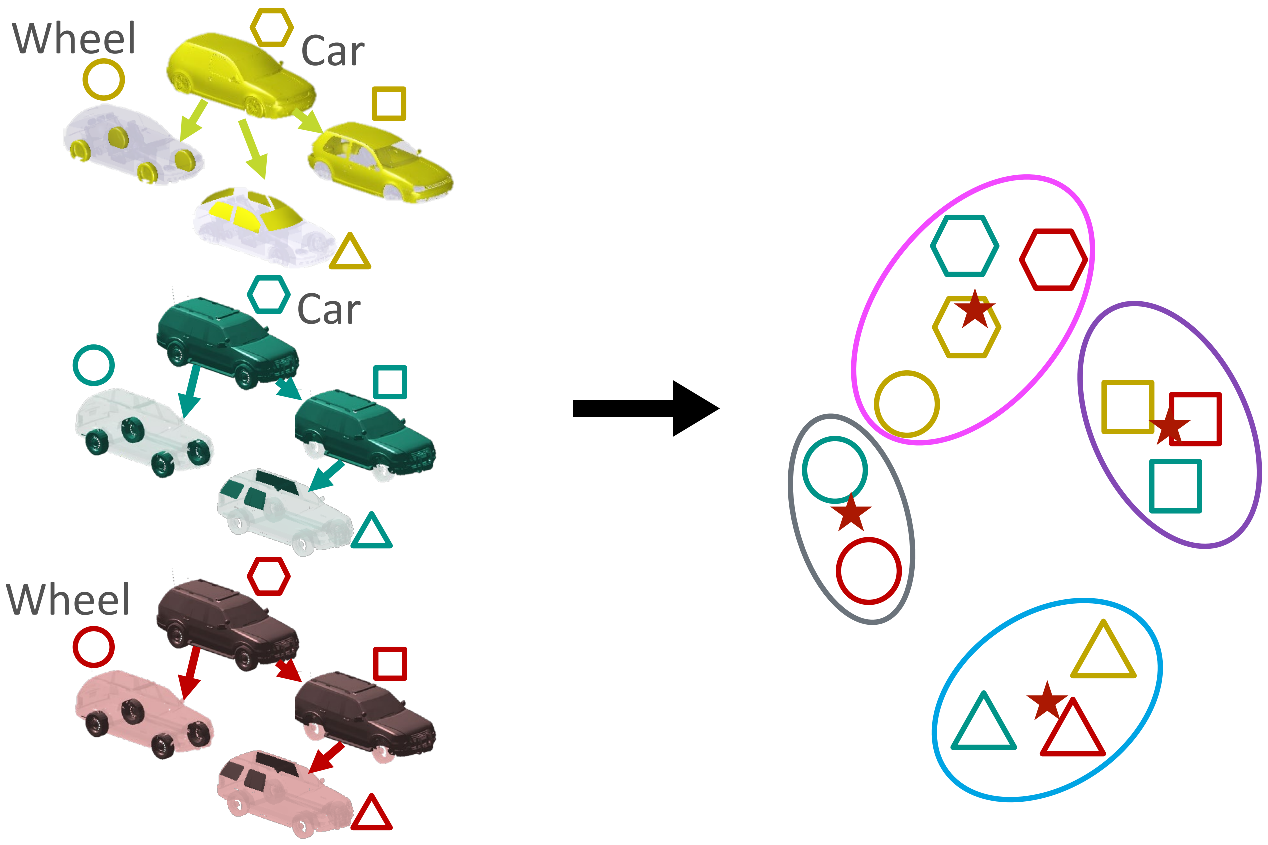
Similarity term:

Group parts with the same tag or geometry

Objective Function

Clustering Similarity Dissimilarity

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



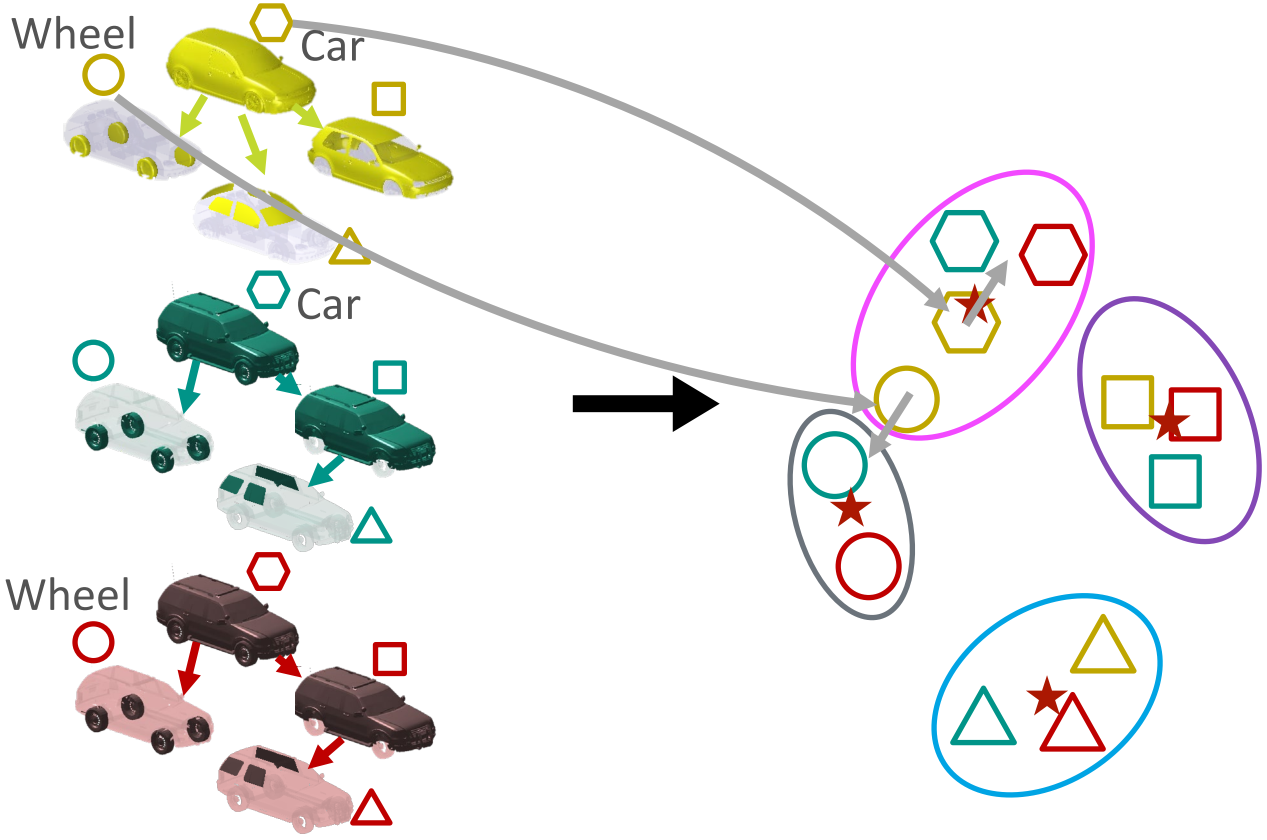
Dissimilarity term:

Separate parts which should not have the same label

Objective Function

Clustering Similarity Dissimilarity

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



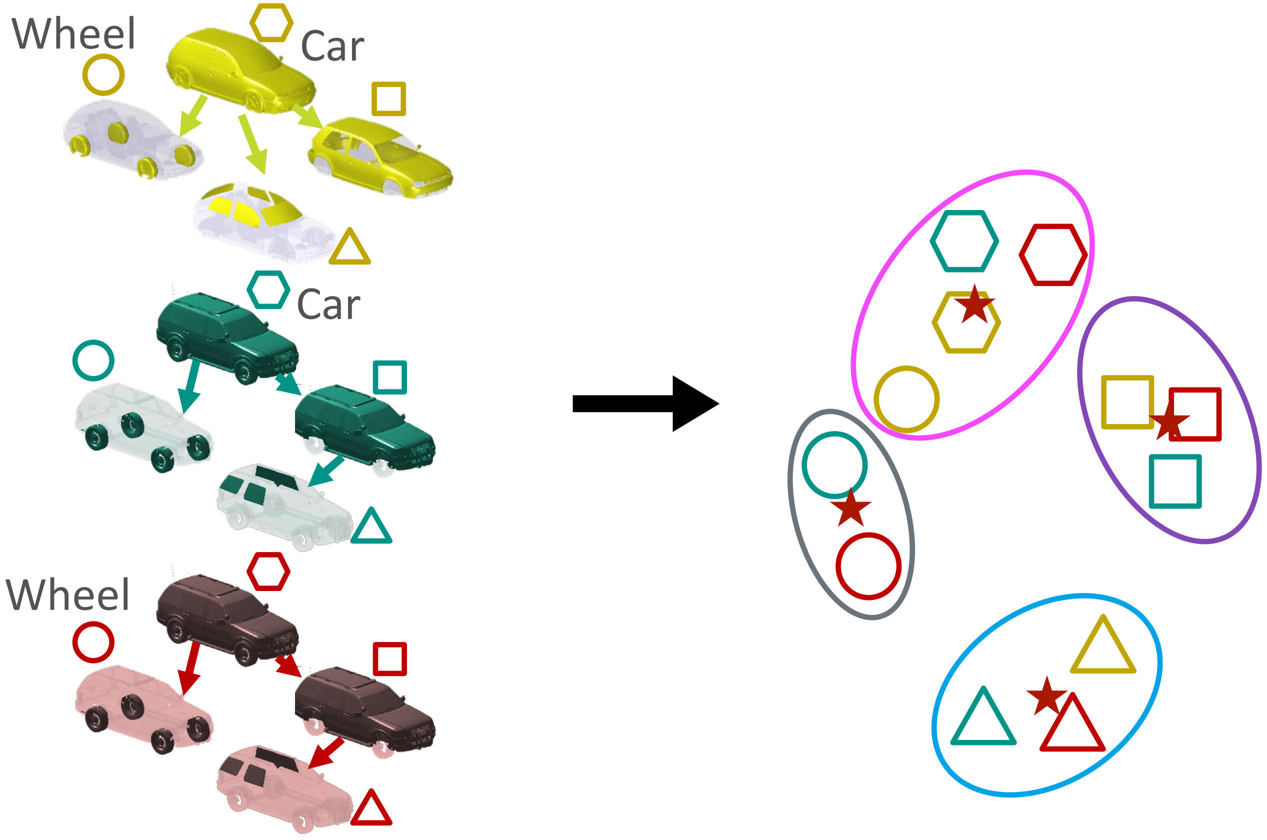
Dissimilarity term:

Separate parts which should not have the same label

Objective Function

Clustering Similarity Dissimilarity Structure

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



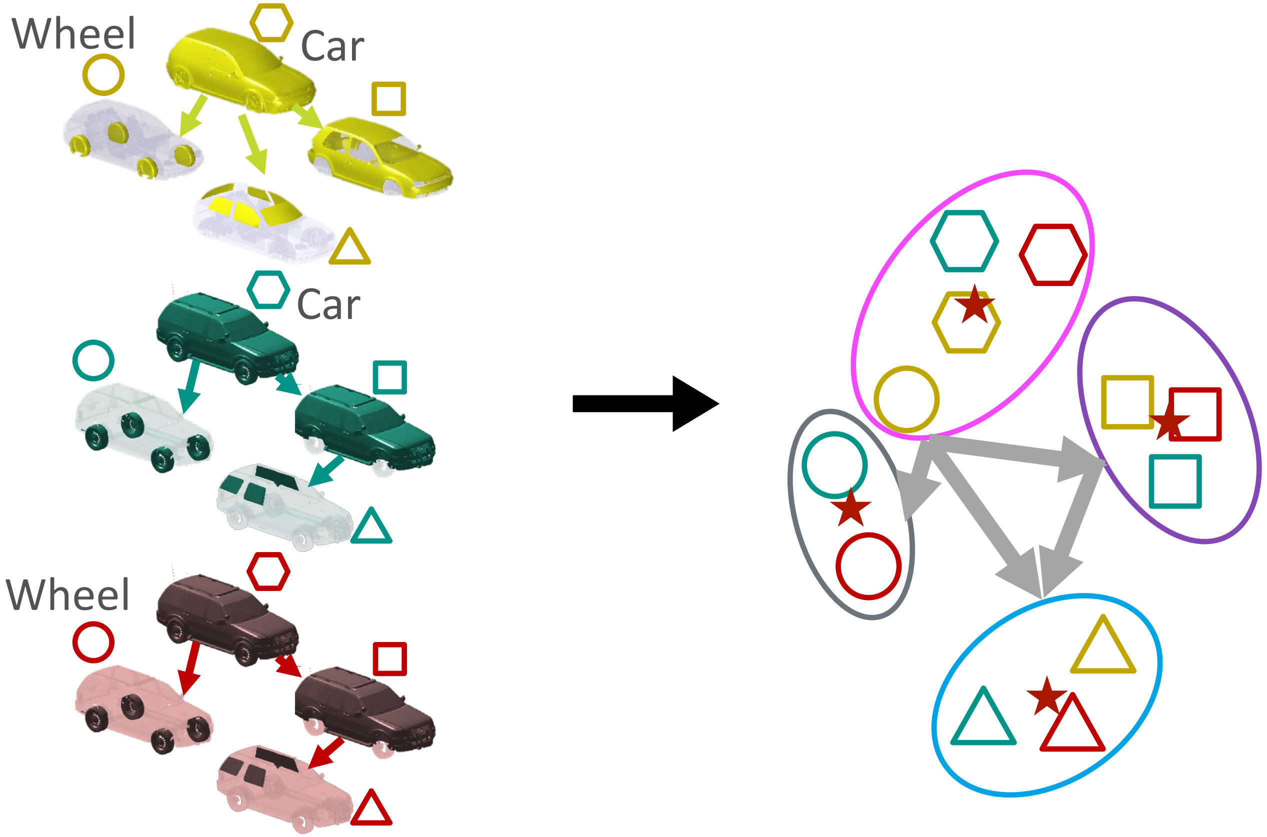
Structure term:

Encourage cluster labels to follow overall parent child relationship

Objective Function

Clustering Similarity Dissimilarity Structure

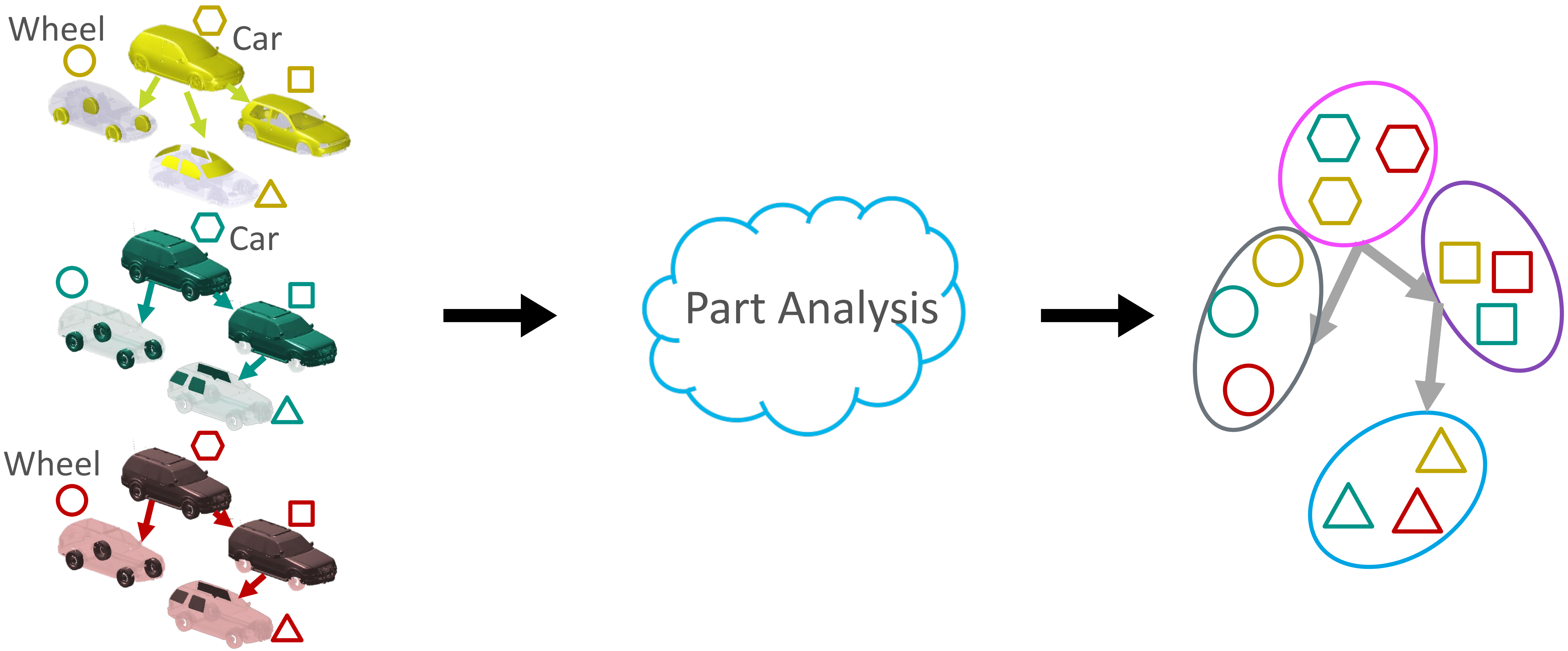
$$E(\theta, p, c, M) = \lambda_c E_c + \lambda_s E_s + \lambda_d E_d + \boxed{\lambda_m E_m - H}$$



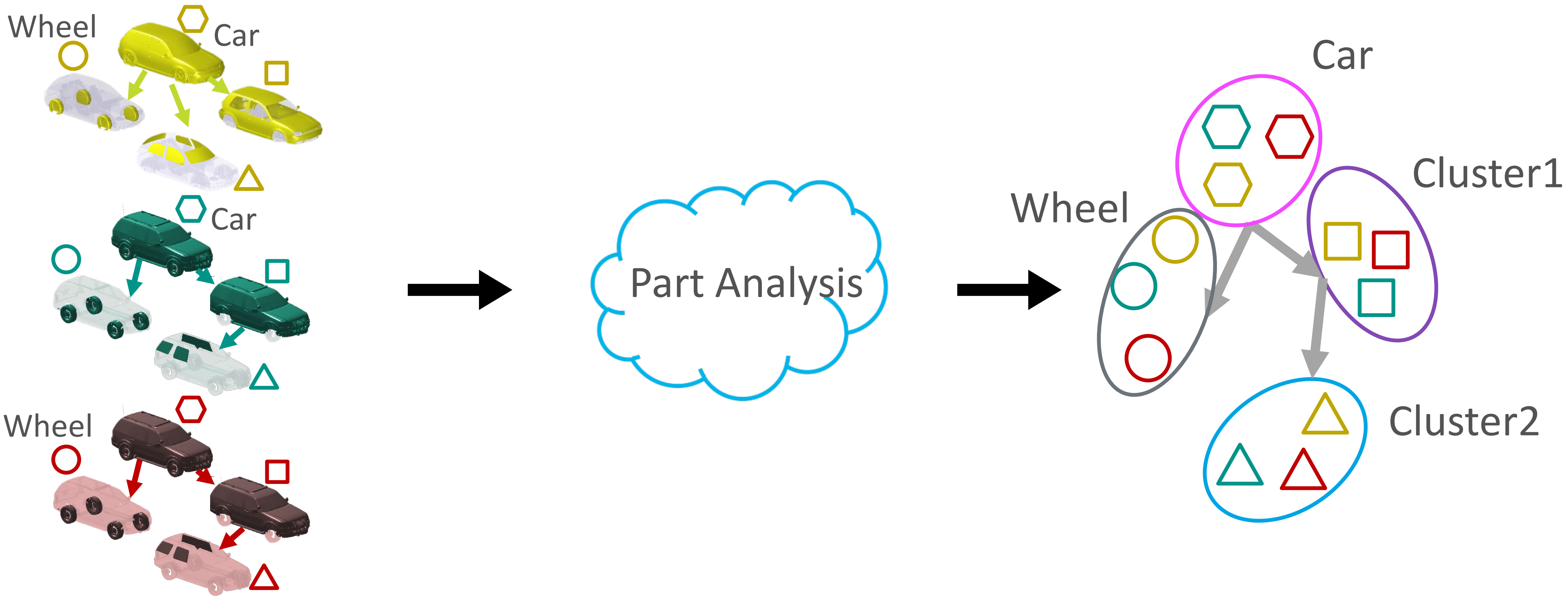
Structure term:

Encourage cluster labels to follow overall parent child relationship

Outputs

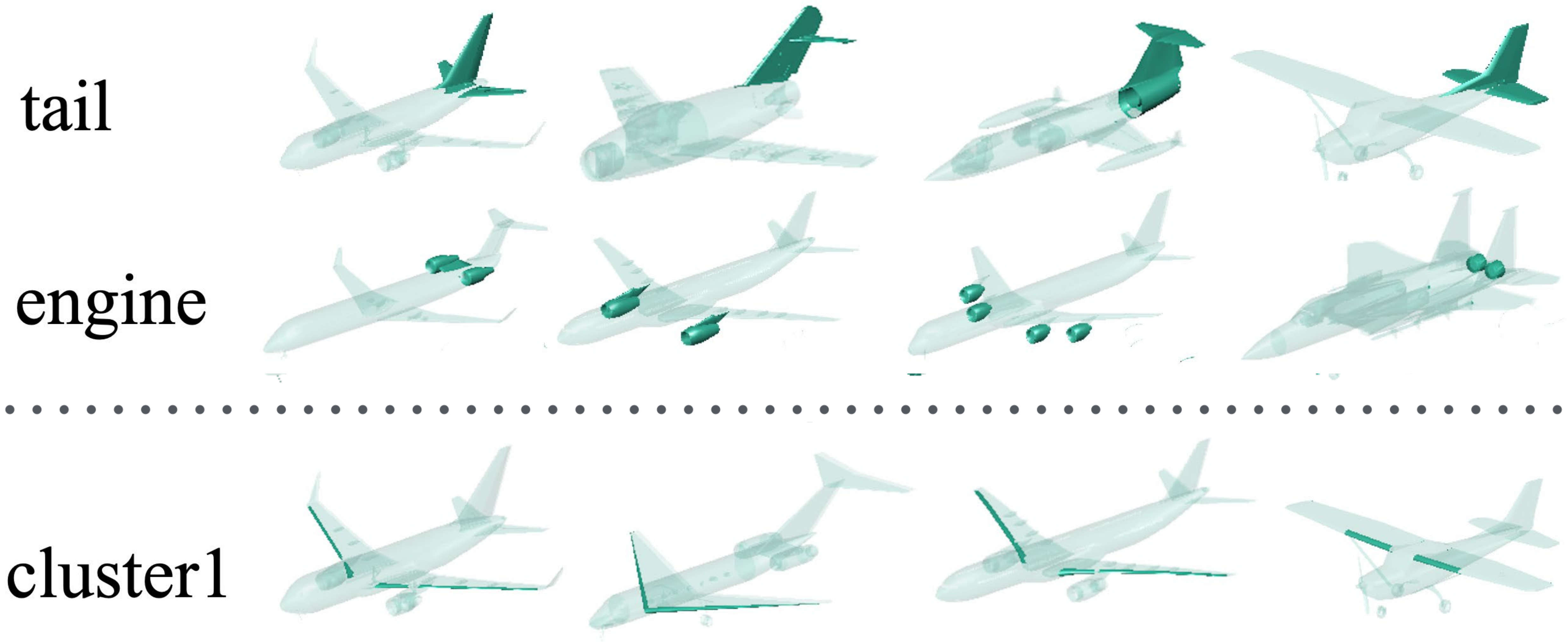


Outputs

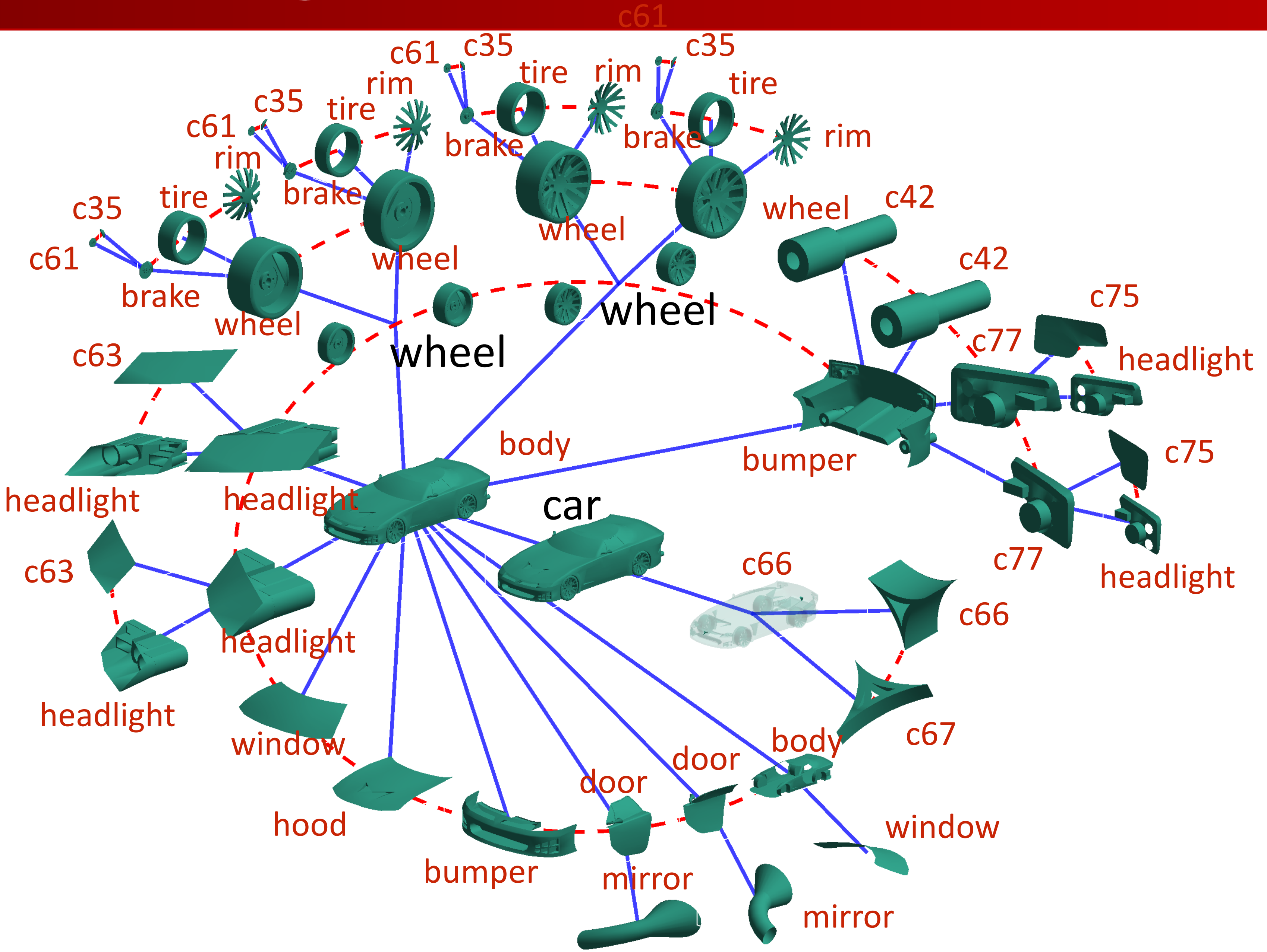


Wheel
Car
Car
Wheel
Part Analysis
Car
Cluster1
Wheel
Cluster2

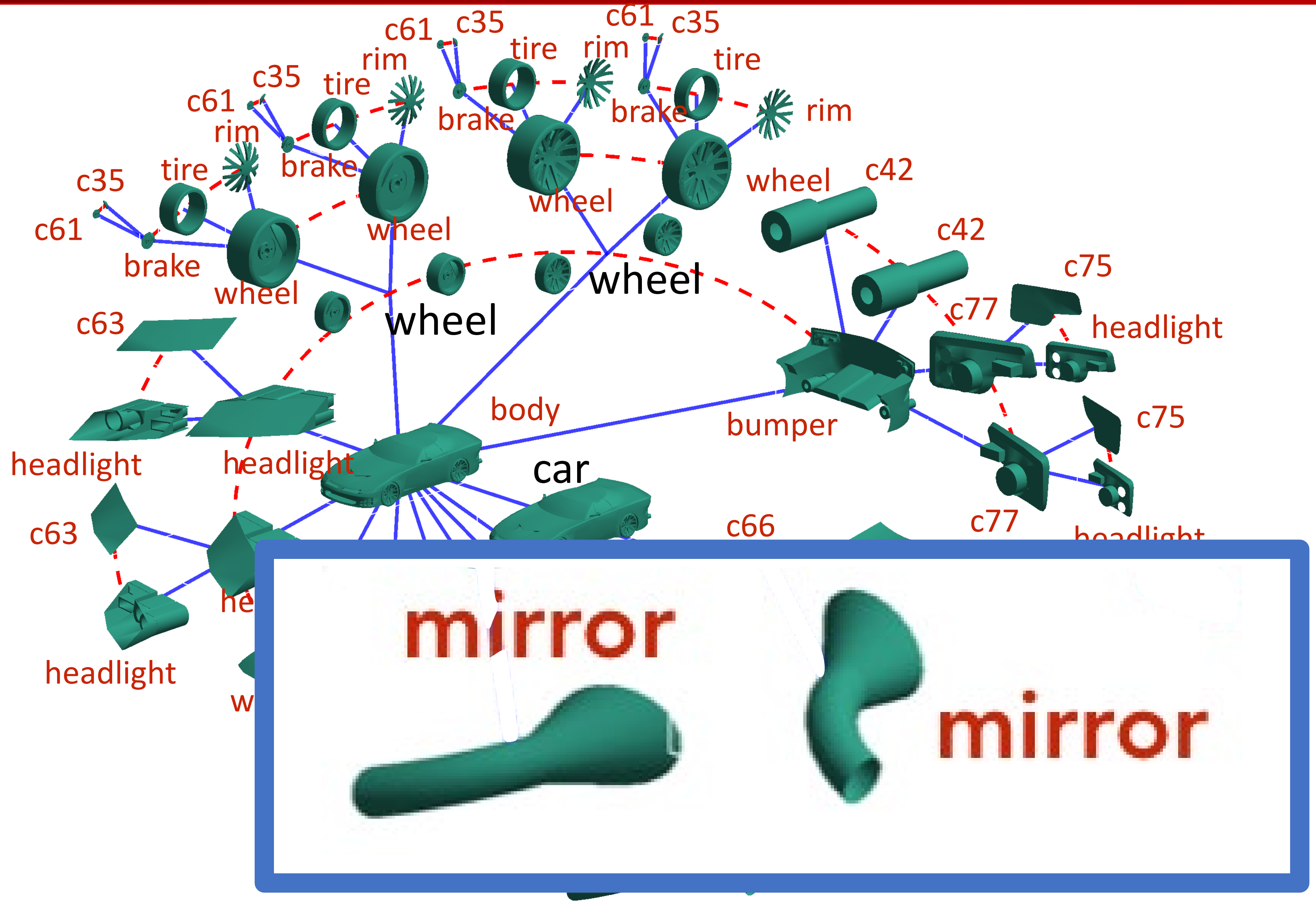
Sample Clusters



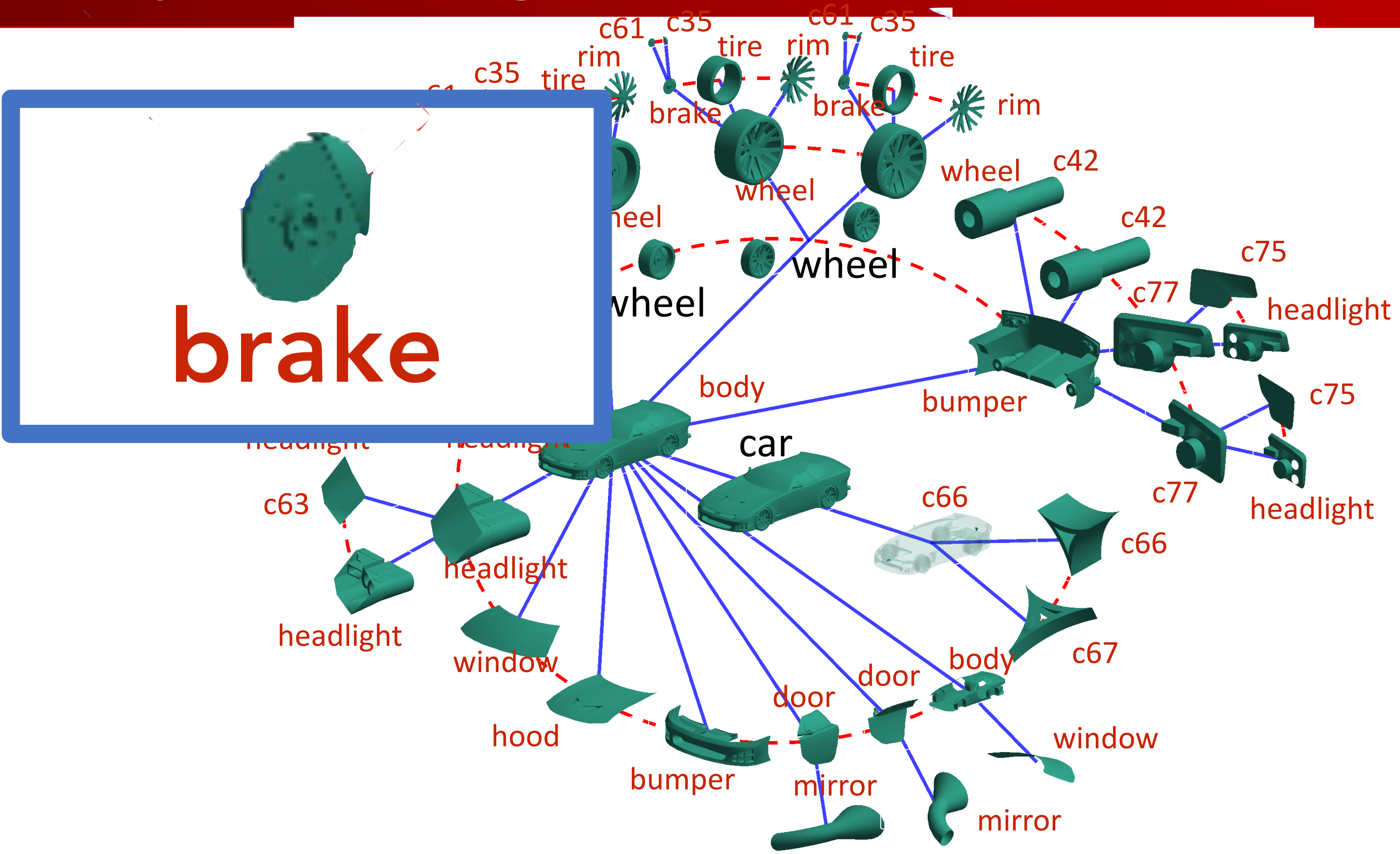
Sample Labeling



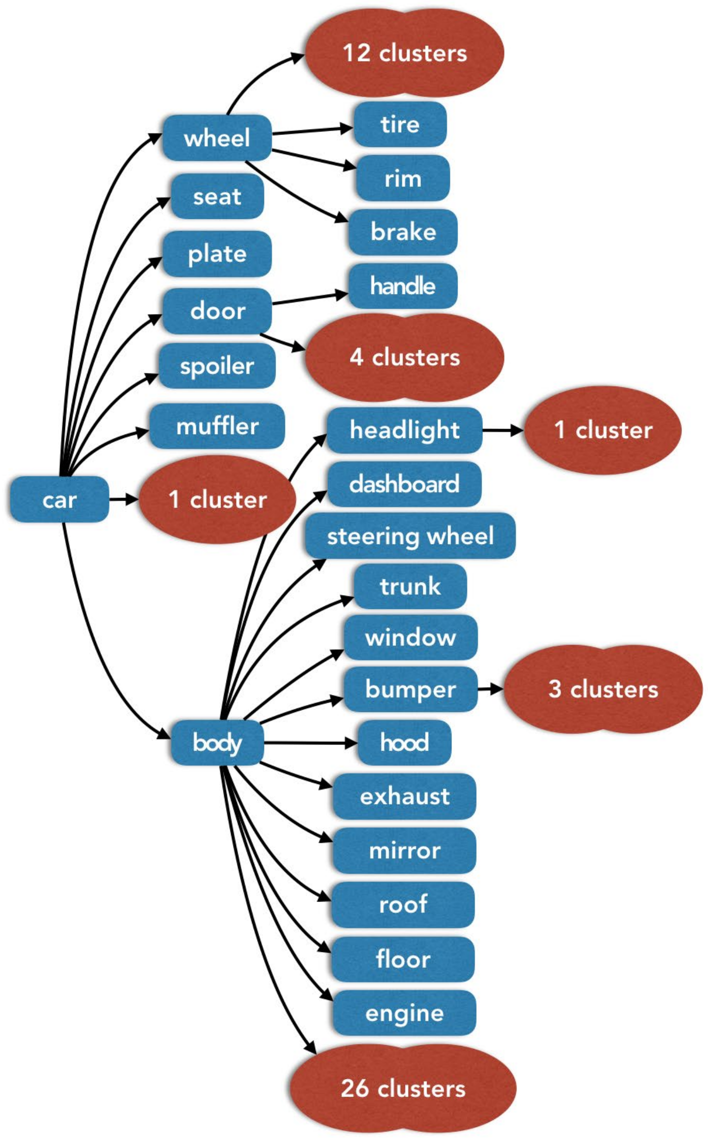
Sample Labeling



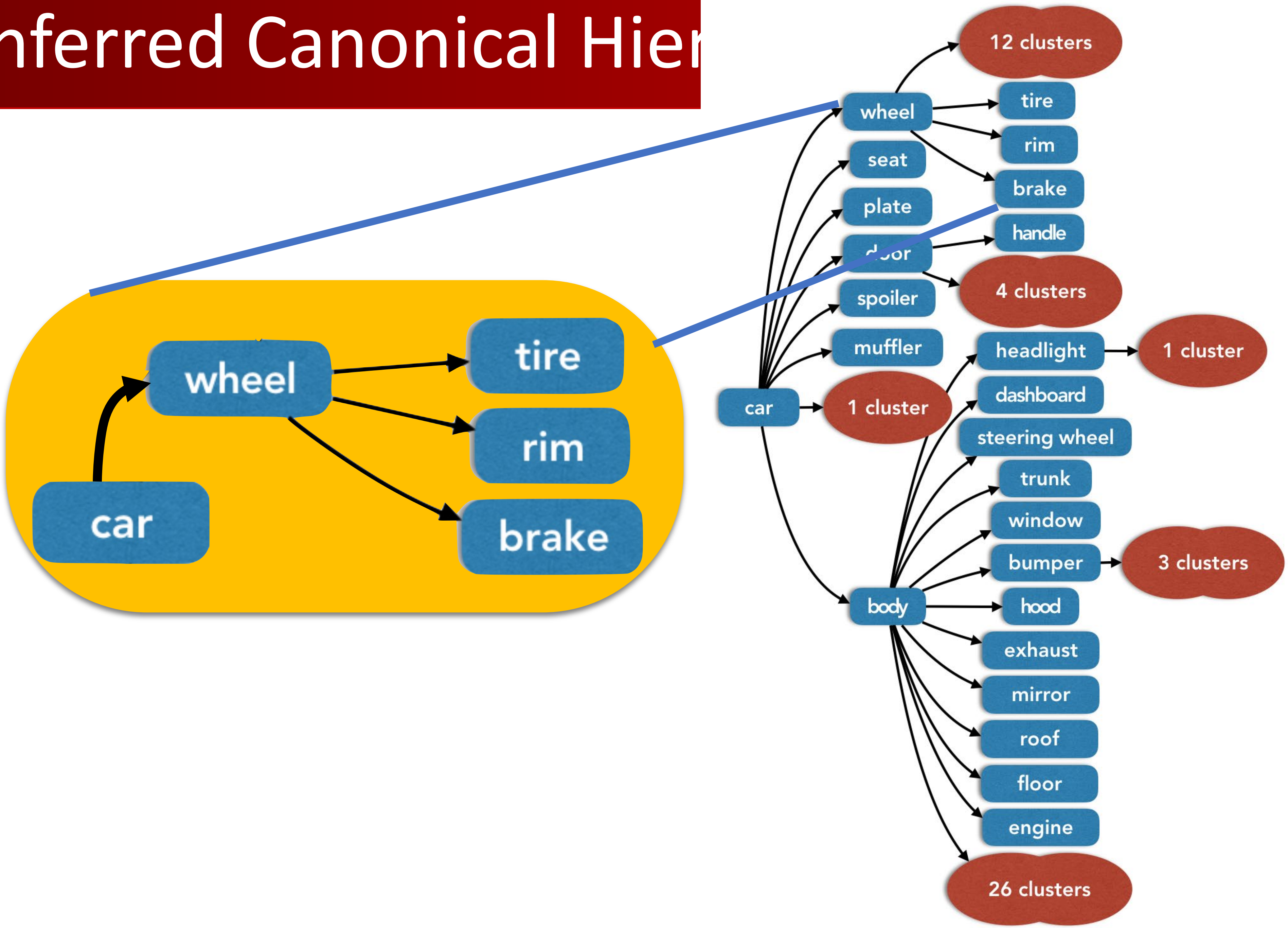
Sample Labeling



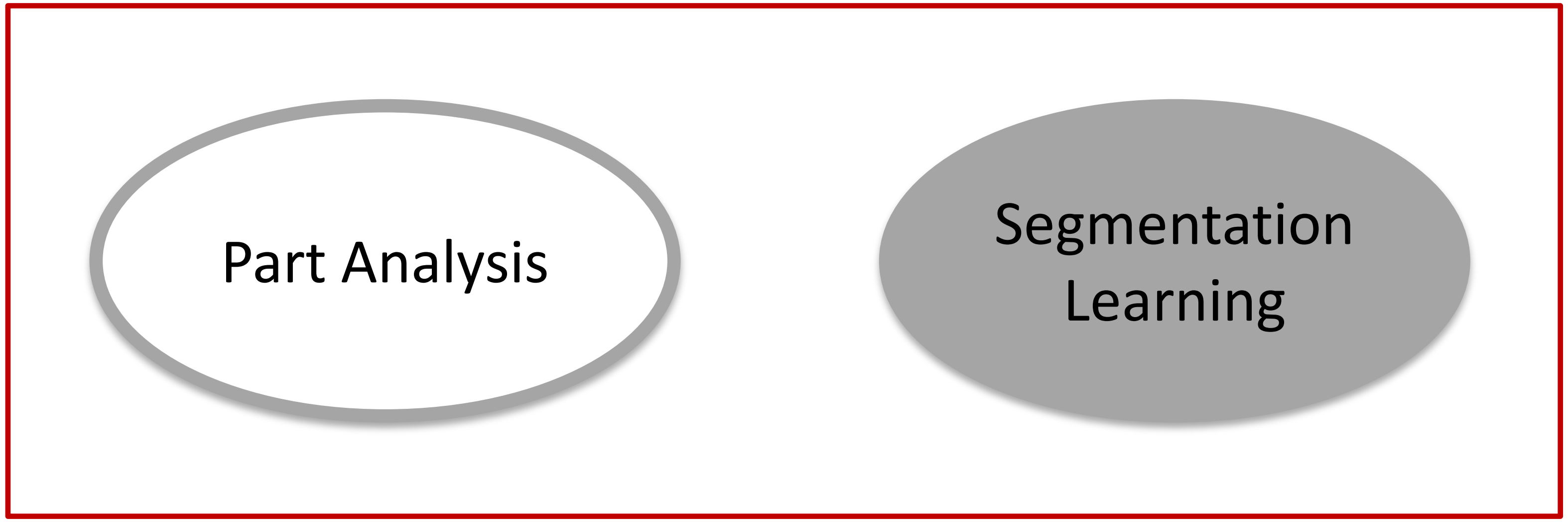
Inferred Canonical Hierarchy



Inferred Canonical Hier

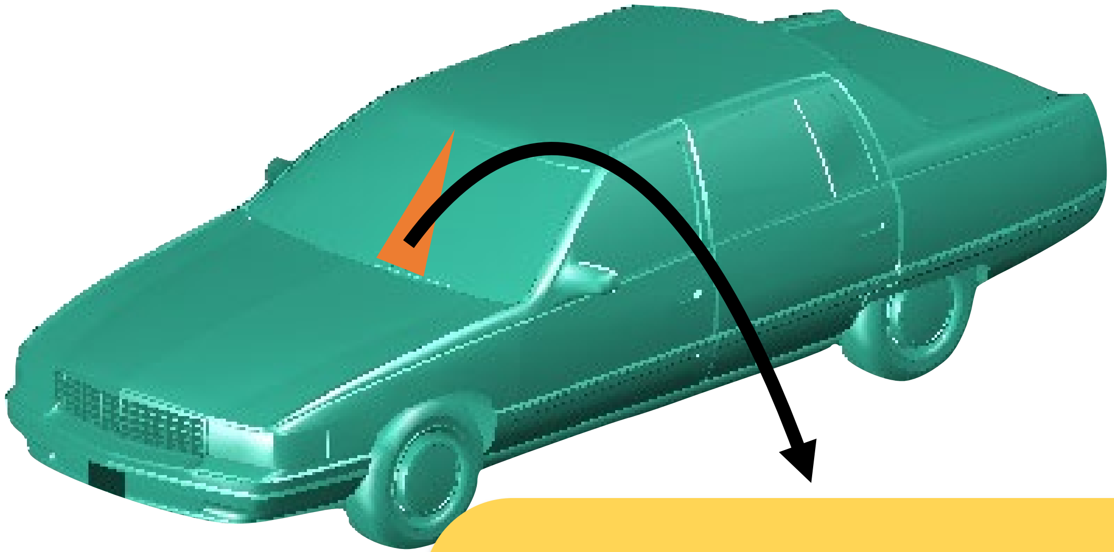


Approach Overview



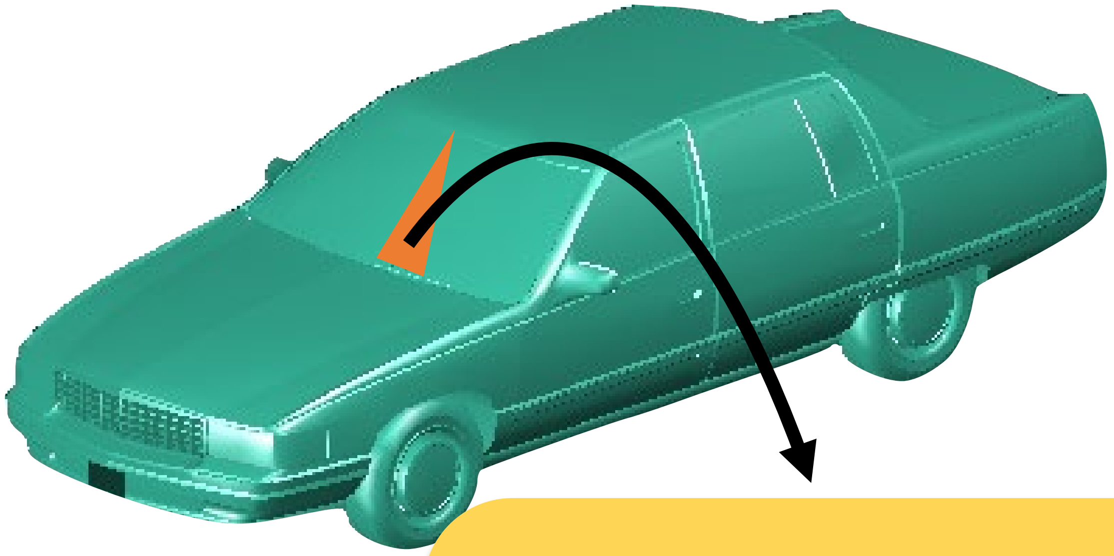
Training Stage

Learning Hierarchical Mesh Segmentation

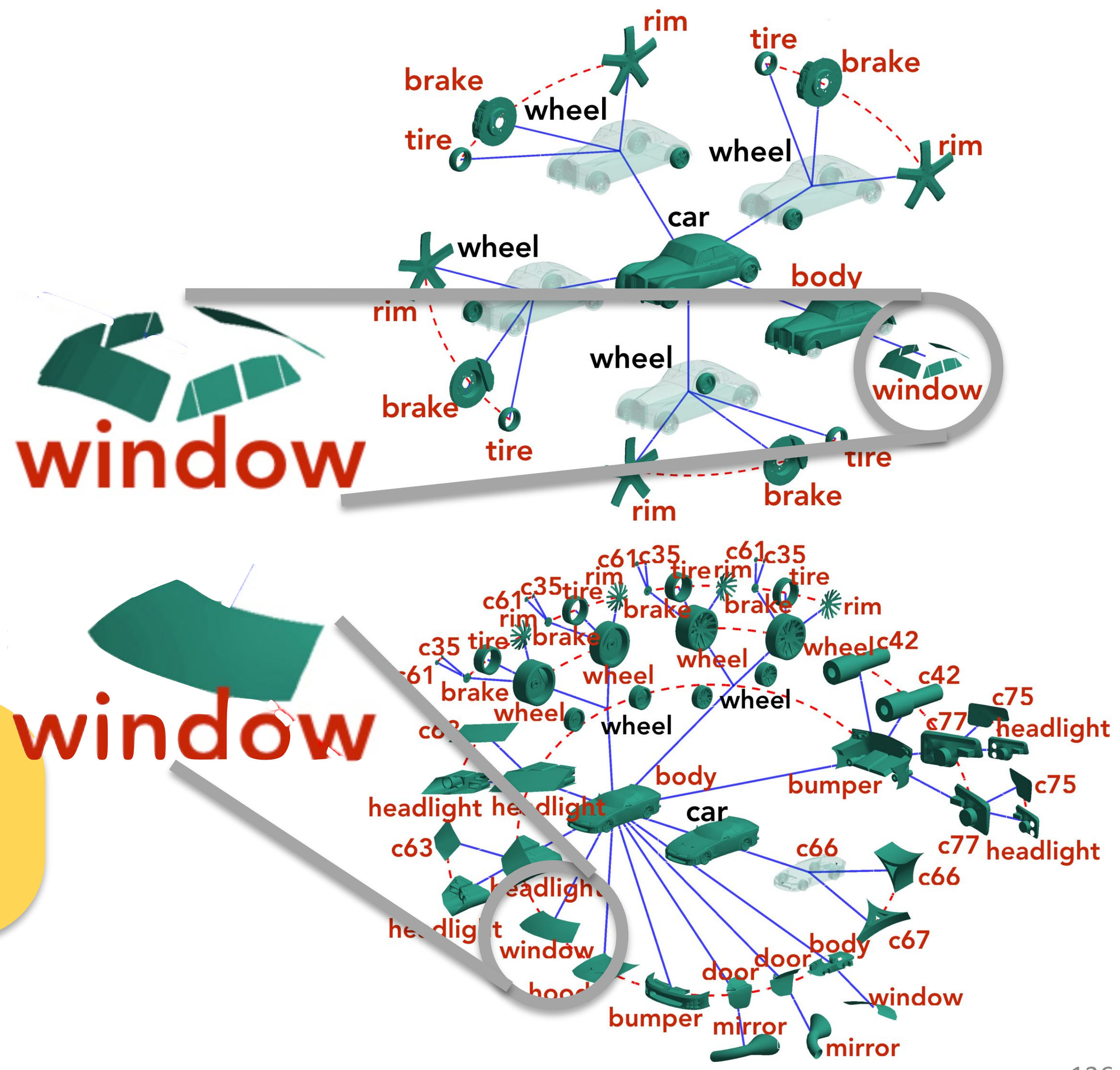


What's the label of this triangle face?

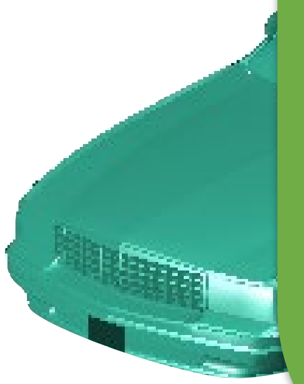
Learning Hierarchical Mesh Segmentation



What's the label of this triangle face?



Learning Hierarchical Mesh Segmentation



Supervised Labeling

triangle face?

Learning Hierarchical Mesh Segmentation

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

A similar MRF formulation as Kalogerakis et al. 2010

Learning Hierarchical Mesh Segmentation

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

Learning Hierarchical Mesh Segmentation

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

Learned through a fully connected neural network

Learning Hierarchical Mesh Segmentation

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

Learned through a fully connected neural network

Learning Hierarchical Mesh Segmentation

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

Learned through a fully connected neural network

Encodes label structures

Learning Hierarchical Mesh Segmentation

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

Learned through a fully connected neural network

Encodes label structures

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

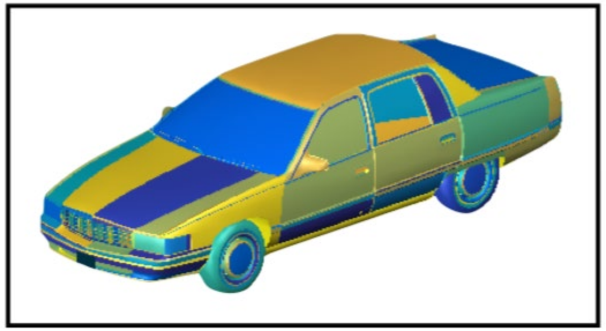
Learning Hierarchical Mesh Segmentation

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

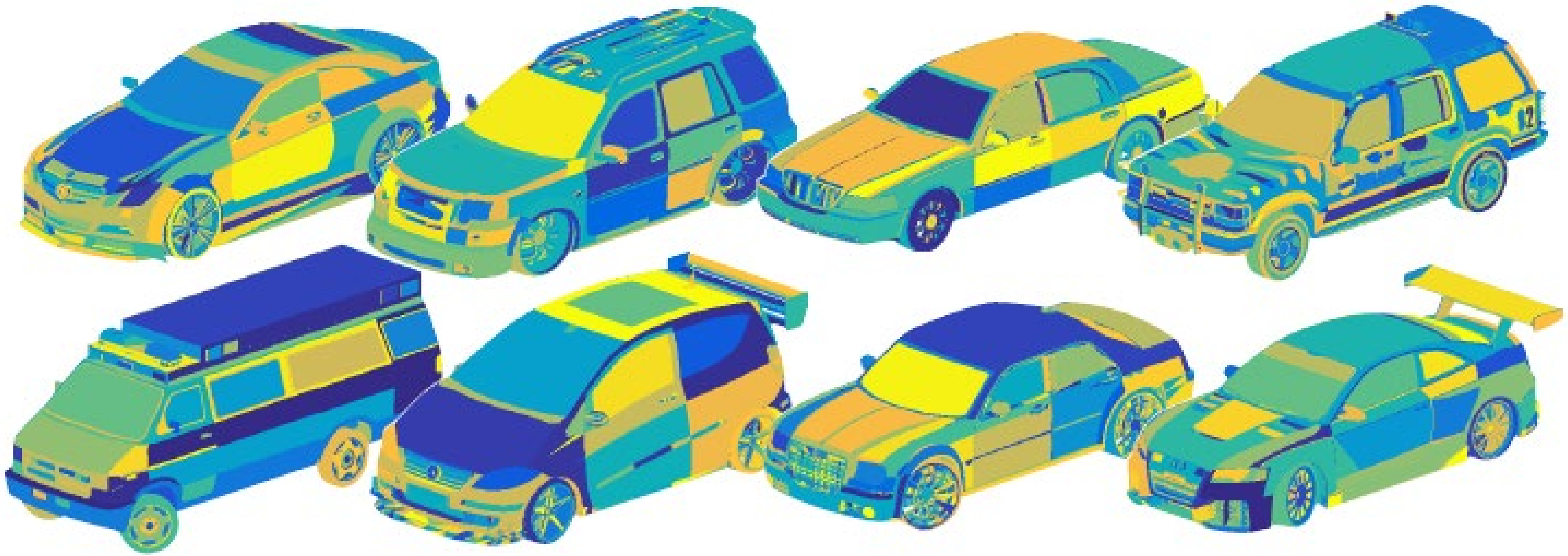
Learned through a fully connected neural network

Encodes label structures

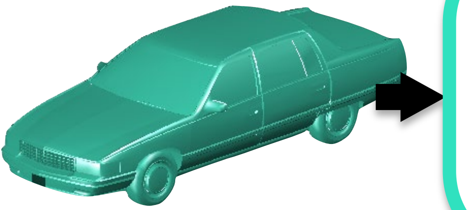
- infer hierarchy,
- handle incomplete training segmentations,
- handle disconnected surfaces
- exploit connected components



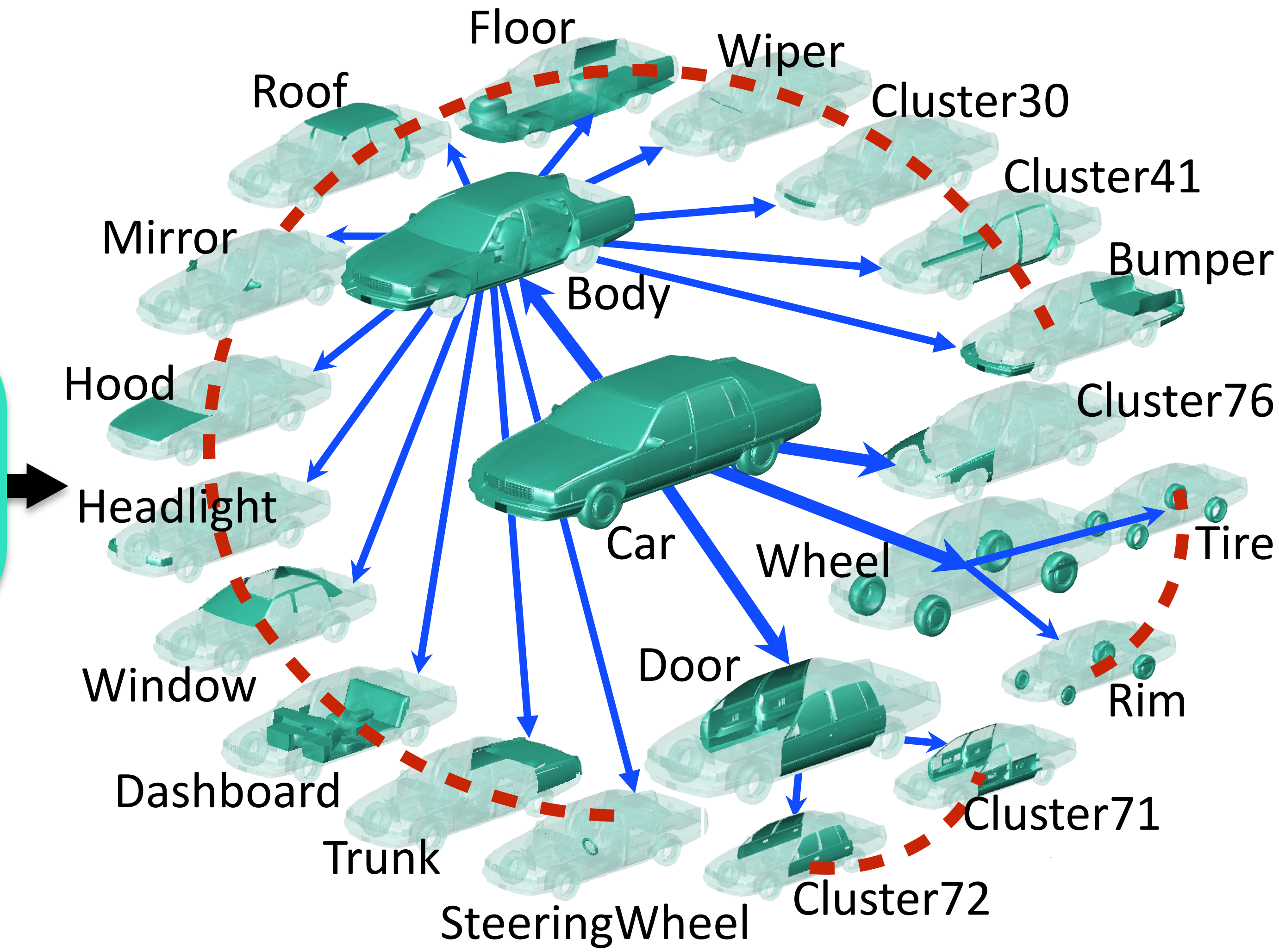
ShapeNet Connected Components



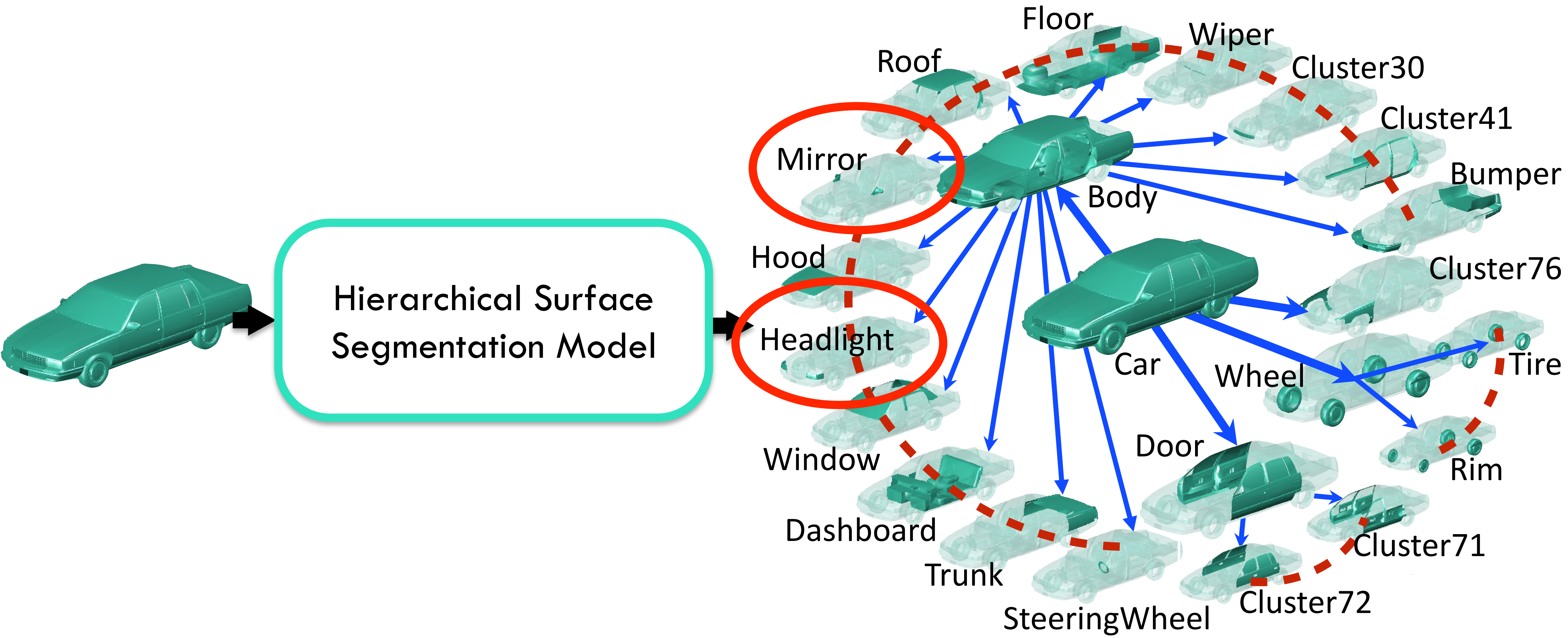
Segmentation Output



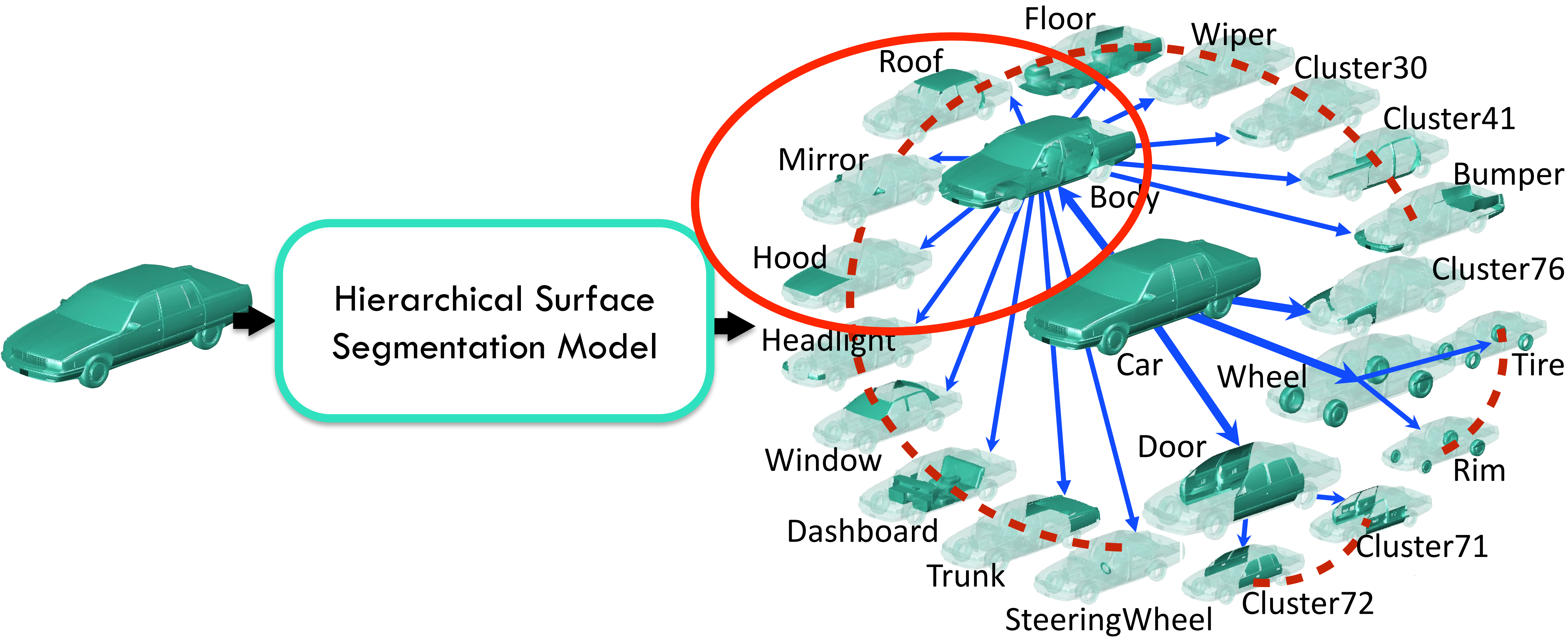
Hierarchical Surface Segmentation Model



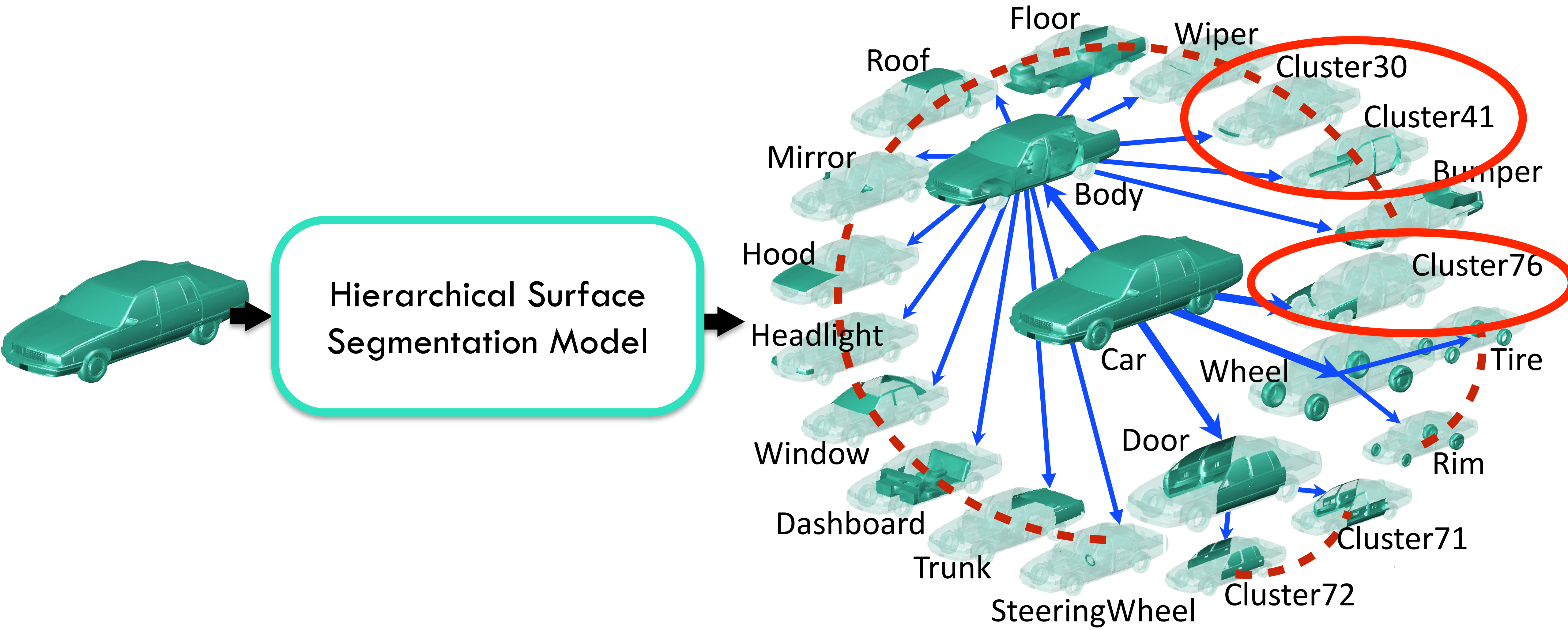
Segmentation Output

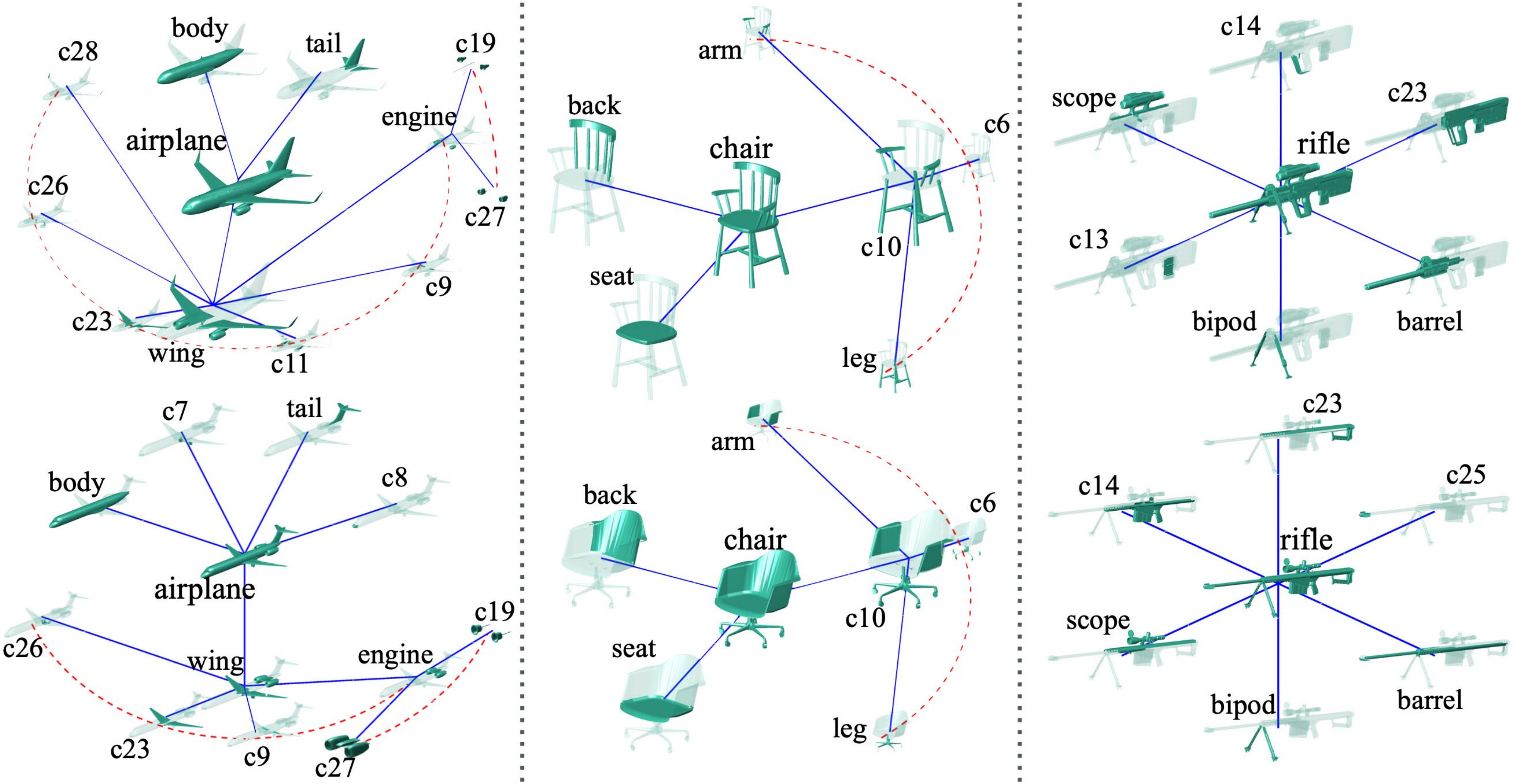


Segmentation output

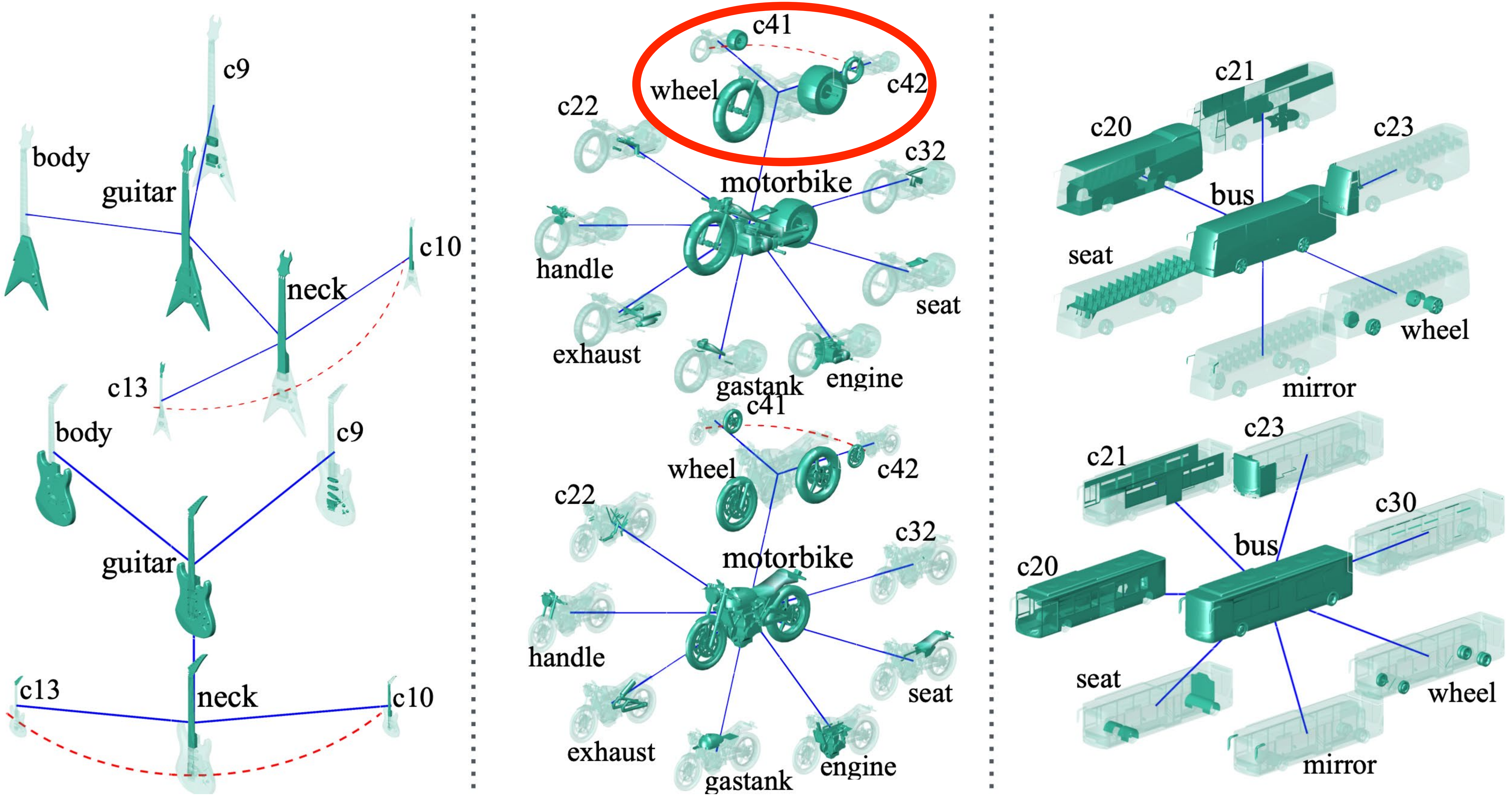


Segmentation Output





body
tail
c28
c19
engine
airplane
c26
c27
c9
c23
wing
c11
c7
tail
body
c8
airplane
c26
c19
wing
engine
c23
c9
c27
arm
back
chair
c6
c10
seat
leg
arm
back
c6
chair
c10
seat
leg
c14
scope
c23
rifle
c13
bipod
barrel
c23
c14
c25
rifle
scope
bipod
barrel


c9
body
guitar
neck
c10
c13
body
c9
guitar
c13
neck
c10
c41
wheel
c42
c22
c32
motorbike
handle
seat
exhaust
gastank
engine
c41
wheel
c42
c22
c32
motorbike
handle
seat
exhaust
gastank
engine
c21
c20
c23
bus
seat
wheel
mirror
c23
c21
c30
bus
c20
seat
wheel
mirror

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”

The End