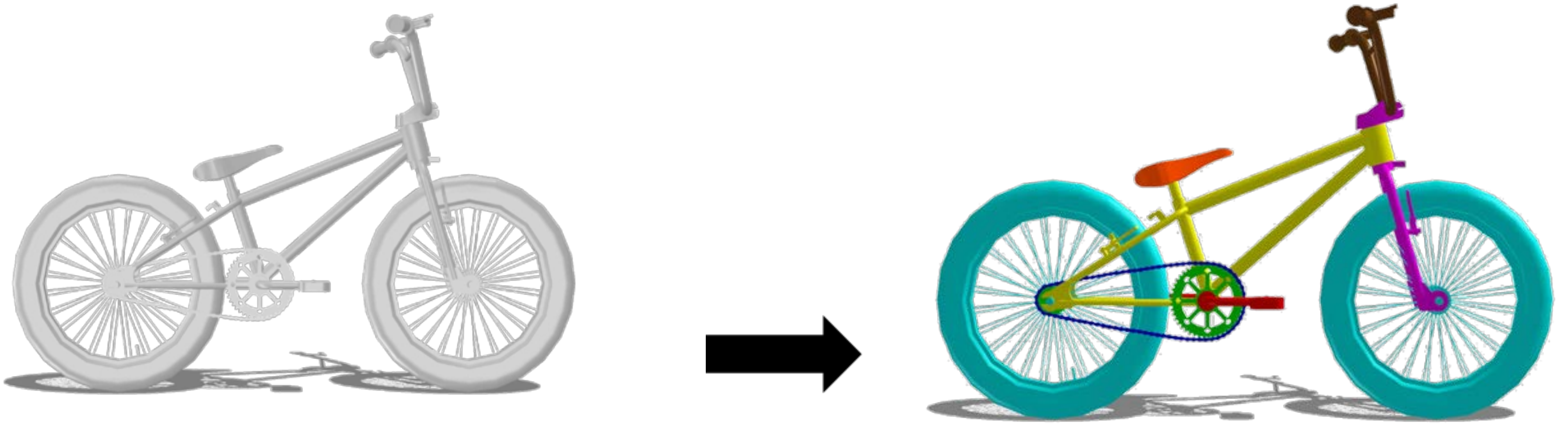
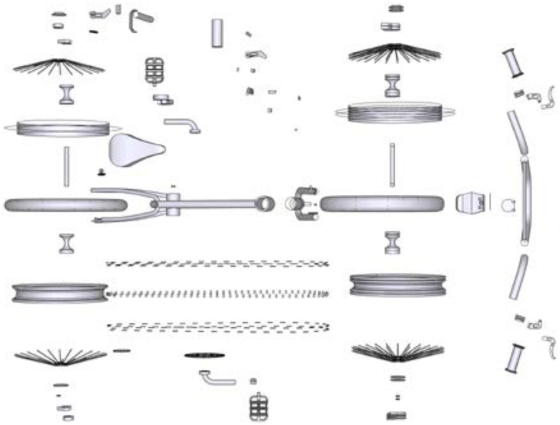
Learning to Group and Label Fine-Grained Shape Components

Xiaogang Wang, *Bin Zhou, Haiyue Fang, Xiaowu Chen, Qinping Zhao, Kai Xu*

Beihang University National University of Defense Technology Princeton University

Motivation

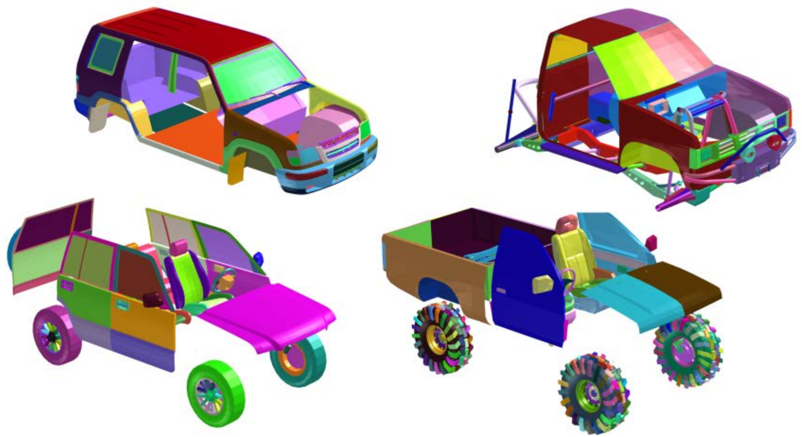




Handlebar	Pedal
Front fork	Chain
Frame	Fender
Wheel	Gear
Seat	Chainguard

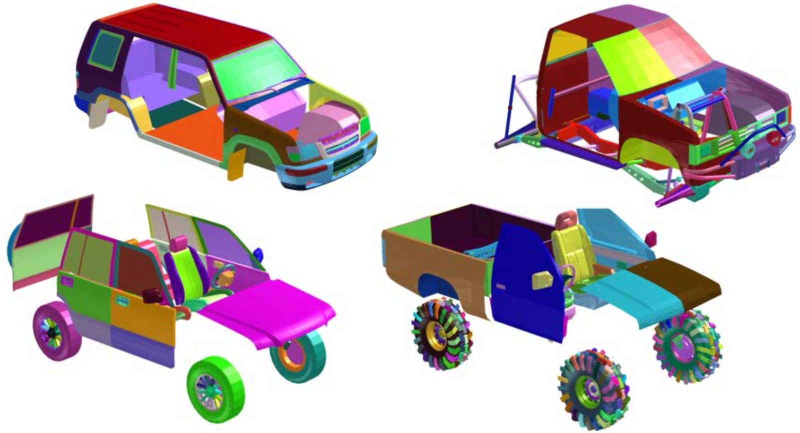
Challenges

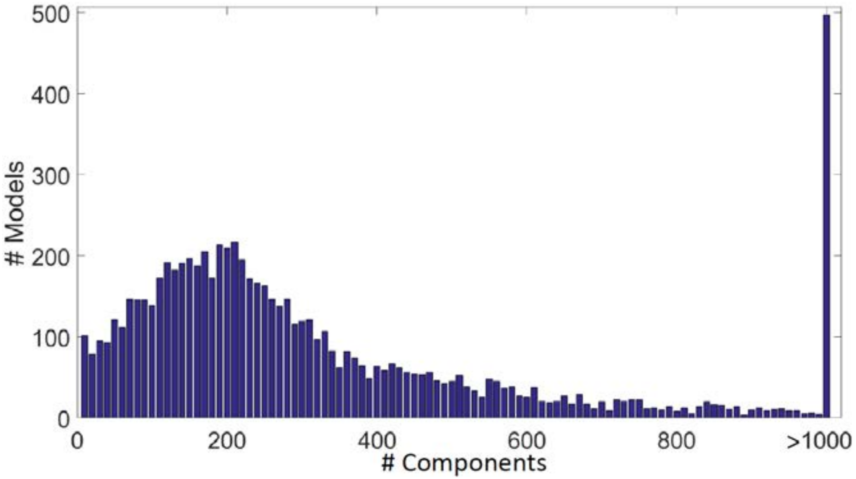
- Highly fine-grained
- The size of components varies significantly
- Highly inconsistent across different shapes



Challenges

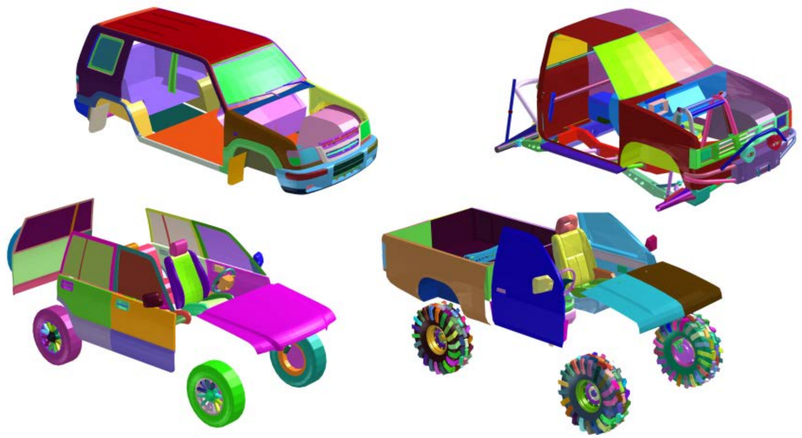
- Highly fine-grained
- The size of components varies significantly
- Highly inconsistent across different shapes

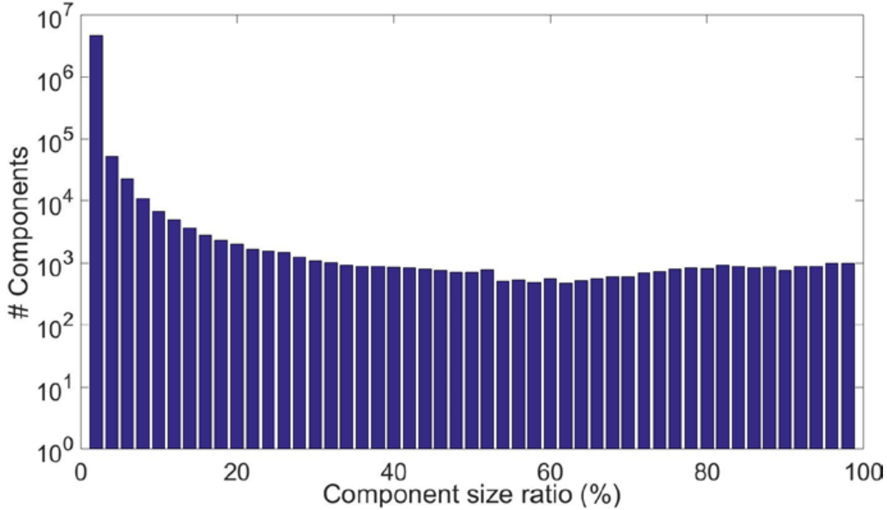




Challenges

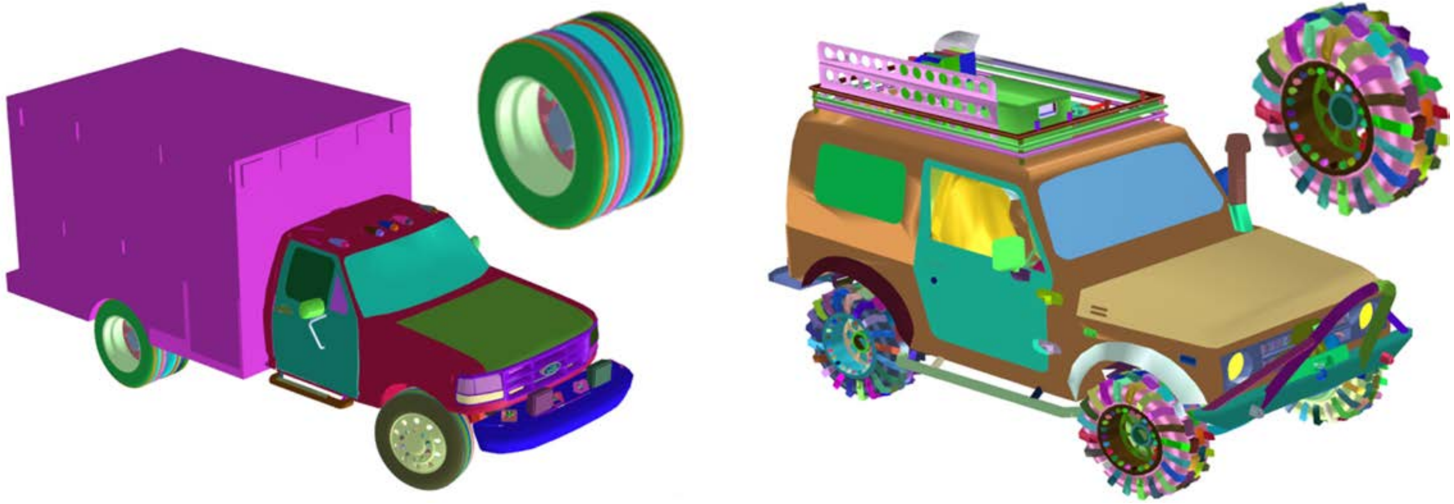
- Highly fine-grained
- The size of components varies significantly
- Highly inconsistent across different shapes





Challenges

- Highly fine-grained
- The size of components varies significantly
- Highly inconsistent across different shapes



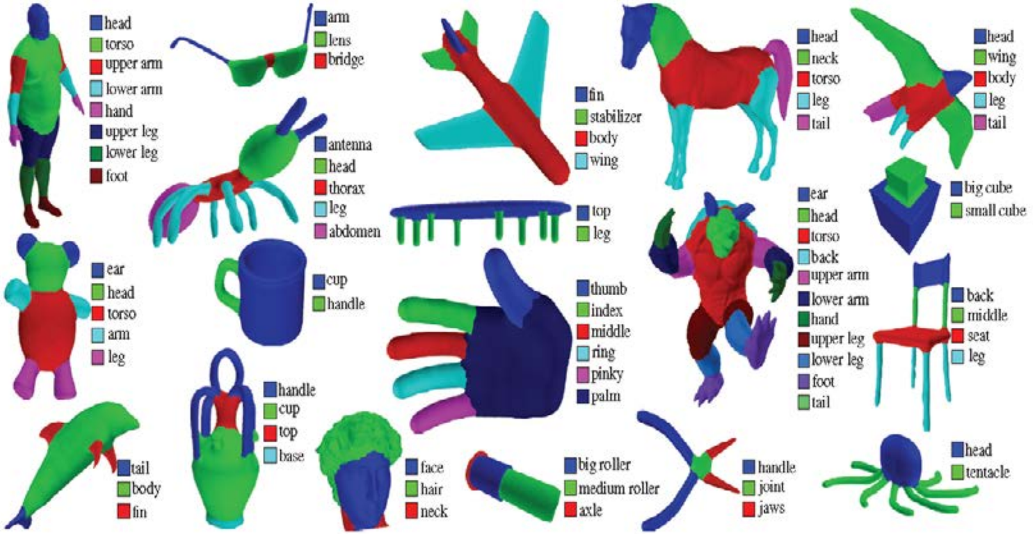
Contributions

- A new problem of segmentation of stock 3D models with pre-existing, highly fine-grained components
- A novel solution of part hypothesis generation and characterization
- A benchmark for multi-component labeling with component-wise ground-truth labels

Related Work

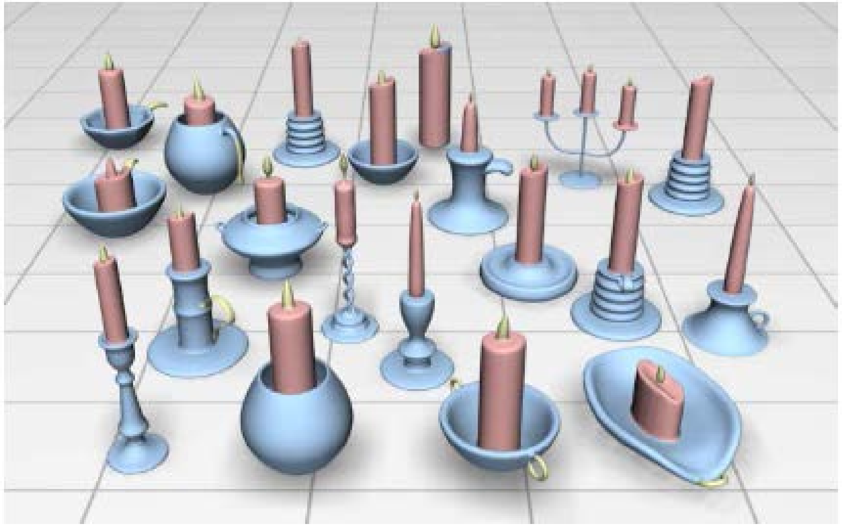
Mesh segmentation

Limited by hand designed features !



Learning 3D Mesh Segmentation and Labeling.

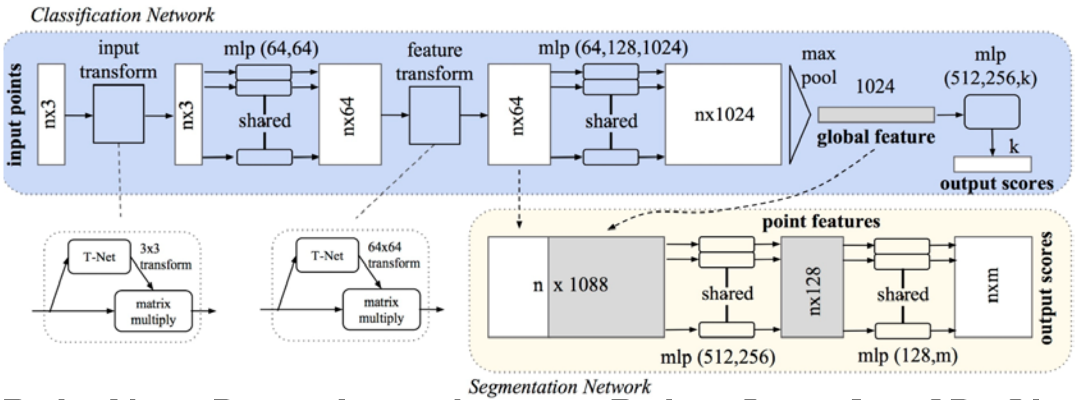
Kalogerakis et al. SIGGRAPH 2010.



Co-Segmentation of 3D Shapes via Subspace Clustering.

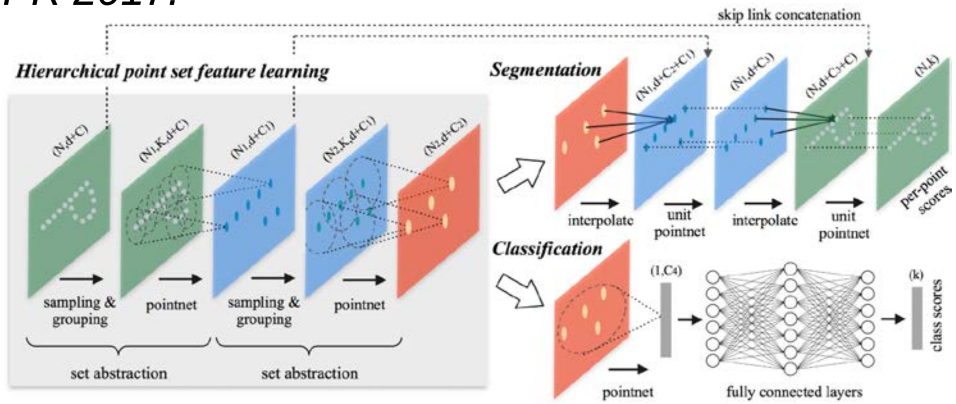
Hu et al. CGF 2012.

Point clouds segmentation



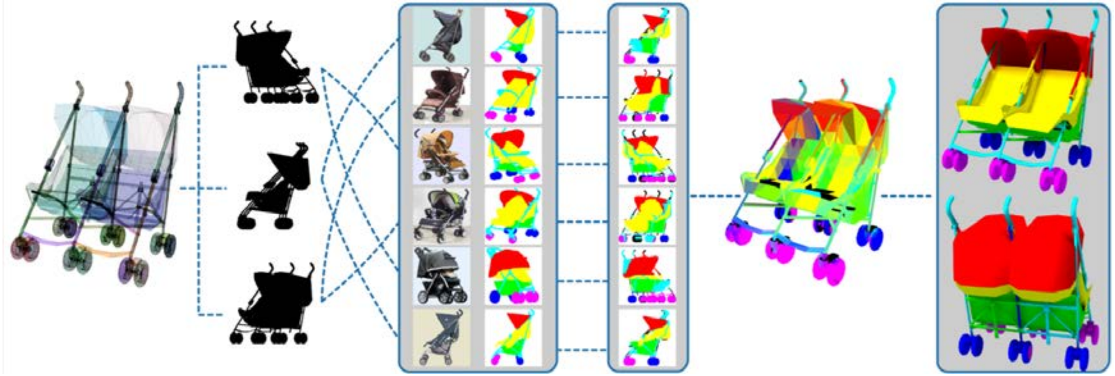
Cannot Handle Fine-grained parts

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Su et al. CVPR 2017.*



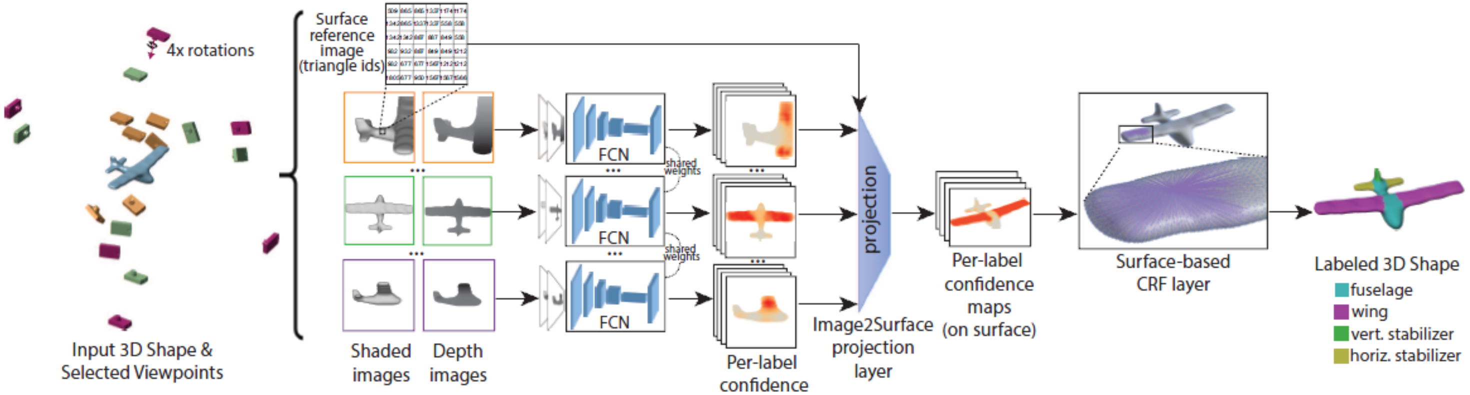
PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Qi et al. Nips 2017.*

Multi-view projective segmentation



Self-occlusion !

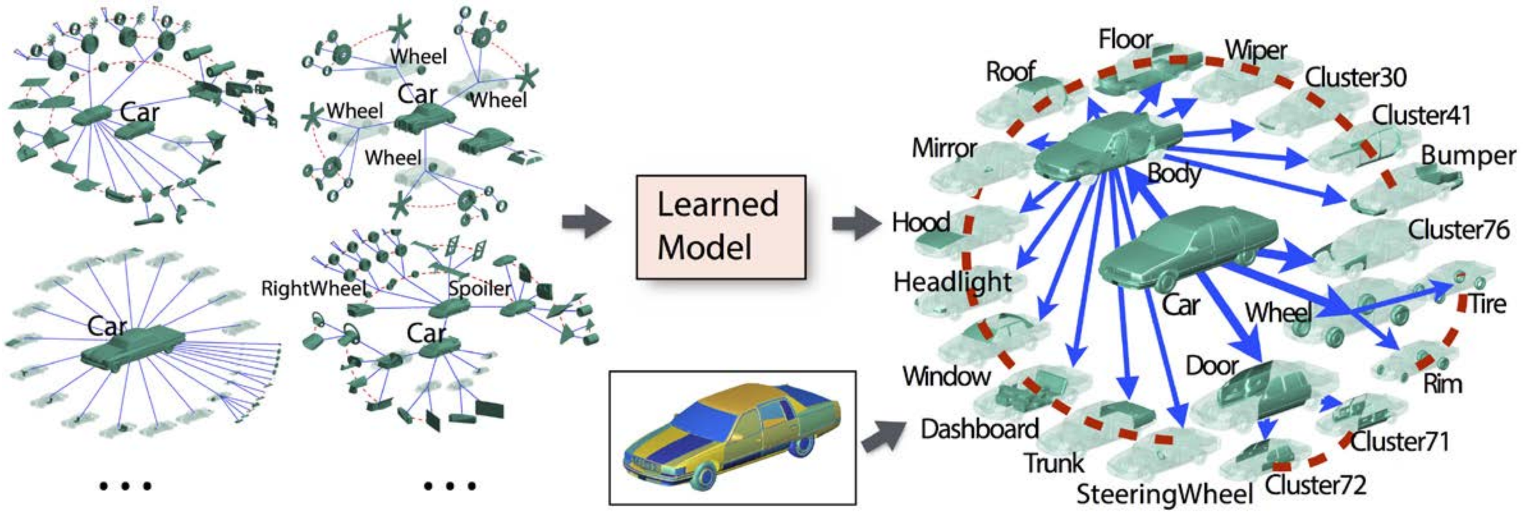
Projective Analysis for 3D Shape Segmentation. *Wang et al. Siggraph 2013.*



3D Shape Segmentation with Projective Convolutional Networks. Kalogerakis *et al. CVPR 2017.*

segmentation of multi-component models

Need scene graph !

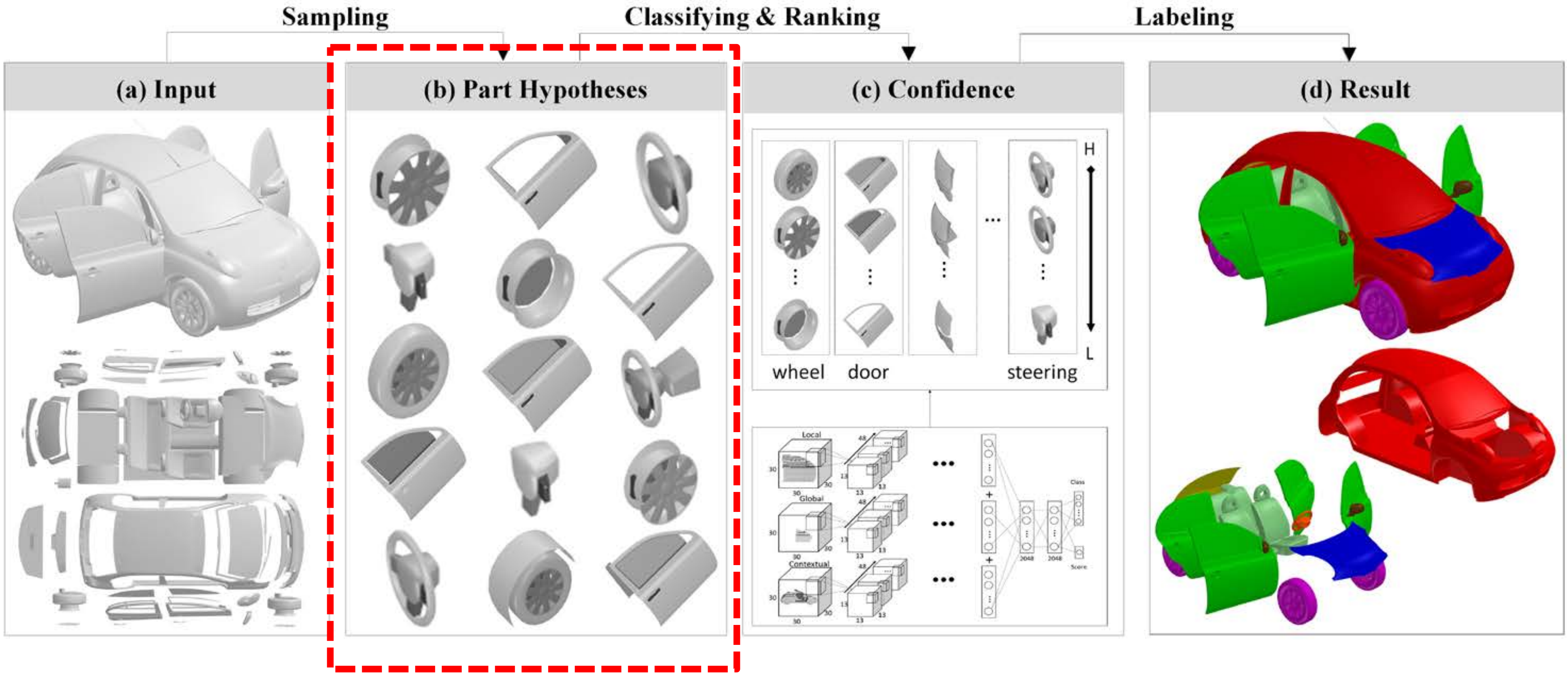


(a) Input Noisy Scene Graphs (b) Novel Instance (c) Hierarchical Shape Segmentation

Learning Hierarchical Shape Segmentation and Labeling from Online Repositories.
Yi et al. Siggraph 2017.

Method

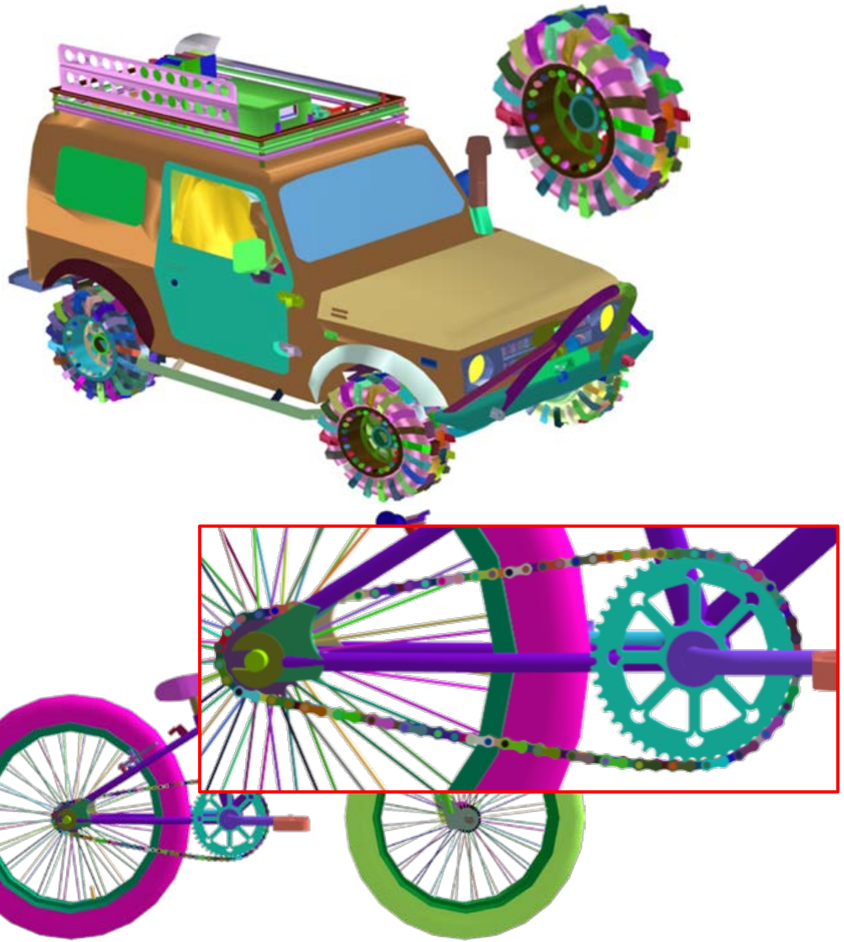
Pipeline



Sampling
Classifying & Ranking
Labeling
(a) Input
(b) Part Hypotheses
(c) Confidence
(d) Result
H
L
wheel
door
steering
Local
Global
Contextual

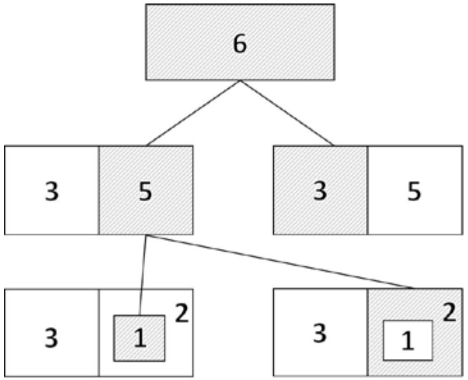
Grouping Strategy

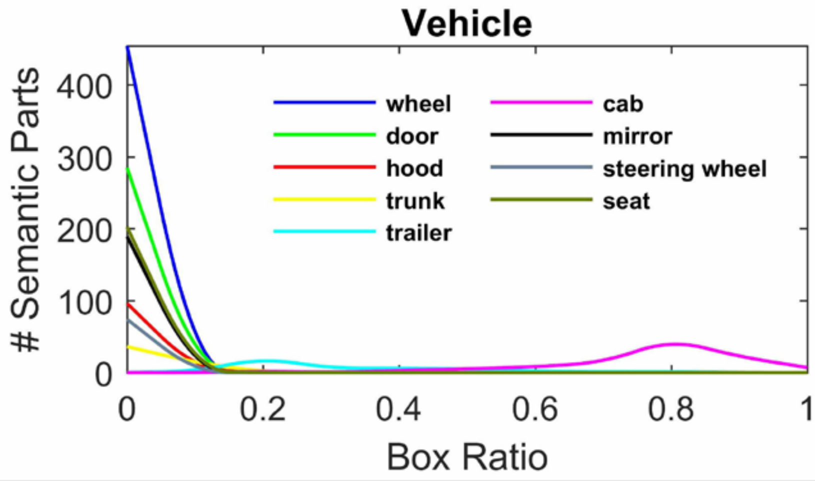
- Center Distance
- Group Size
- Geometric Contact



Grouping Strategy

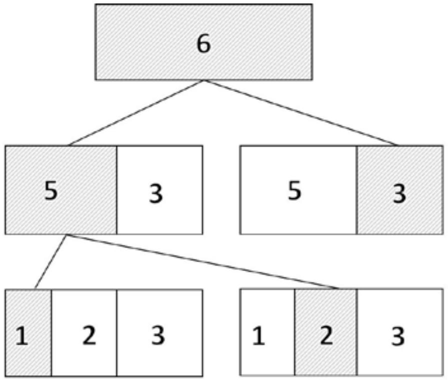
- Center Distance
- Group Size
- Geometric Contact

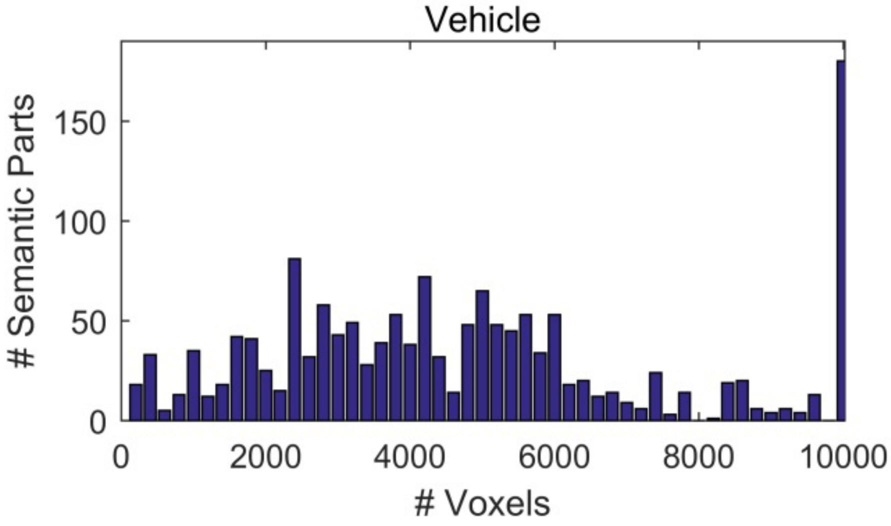




Grouping Strategy

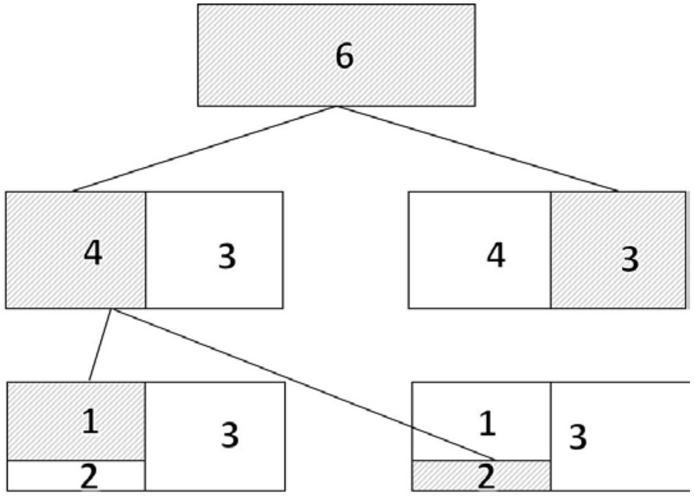
- Center Distance
- Group Size
- Geometric Contact





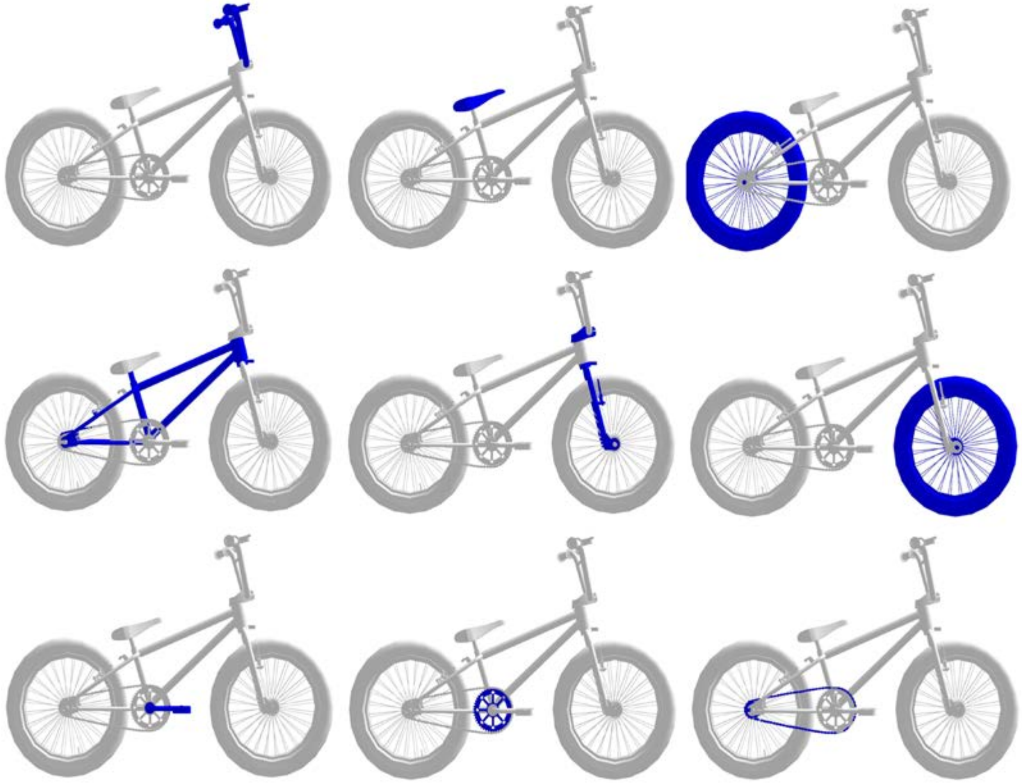
Grouping Strategy

- Center Distance
- Group Size
- Geometric Contact

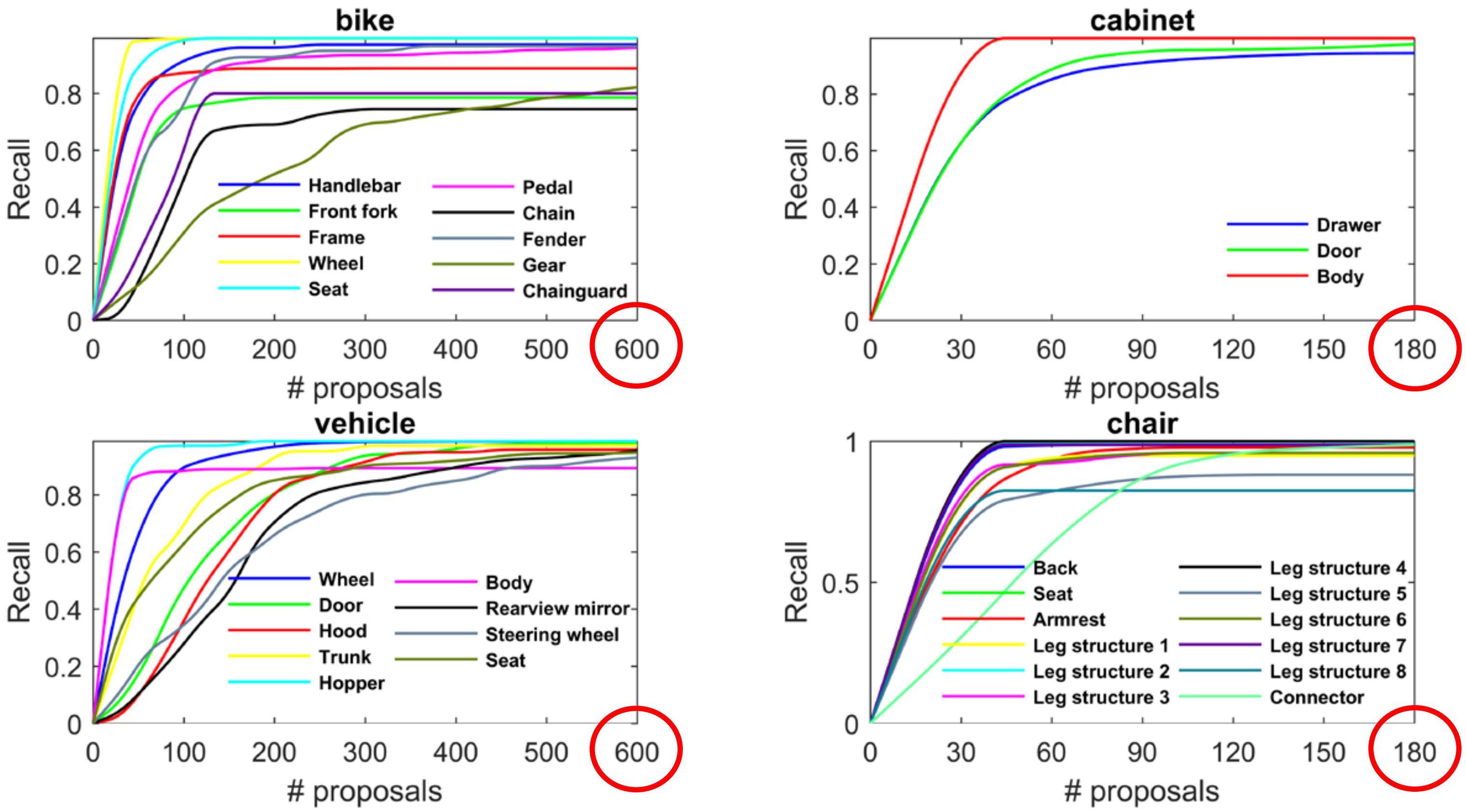


$$C_{\text{contact}}(a, b) = \max\{C_{ab}/V_a, C_{ab}/V_b\}$$

Sampling Results

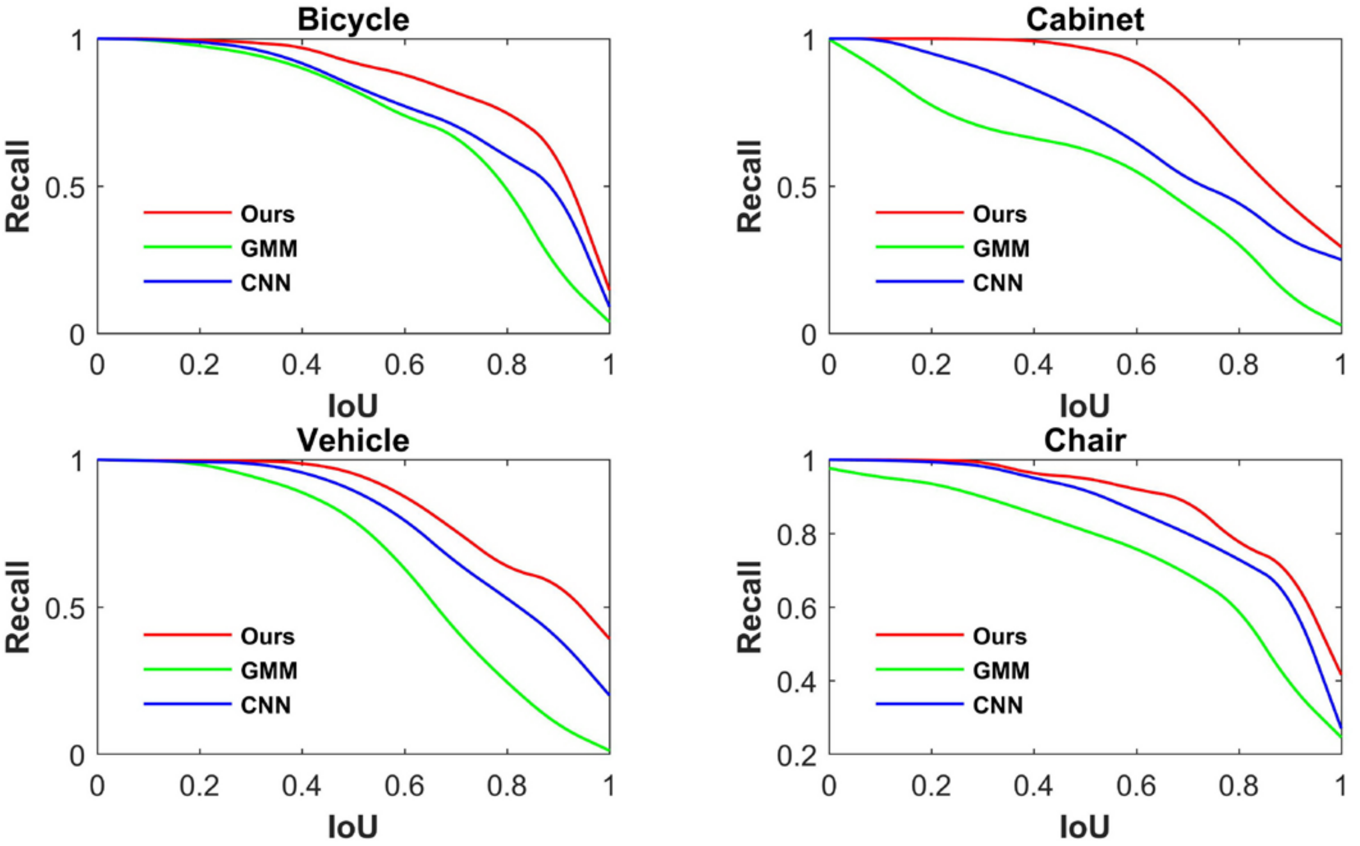


Sampling Results



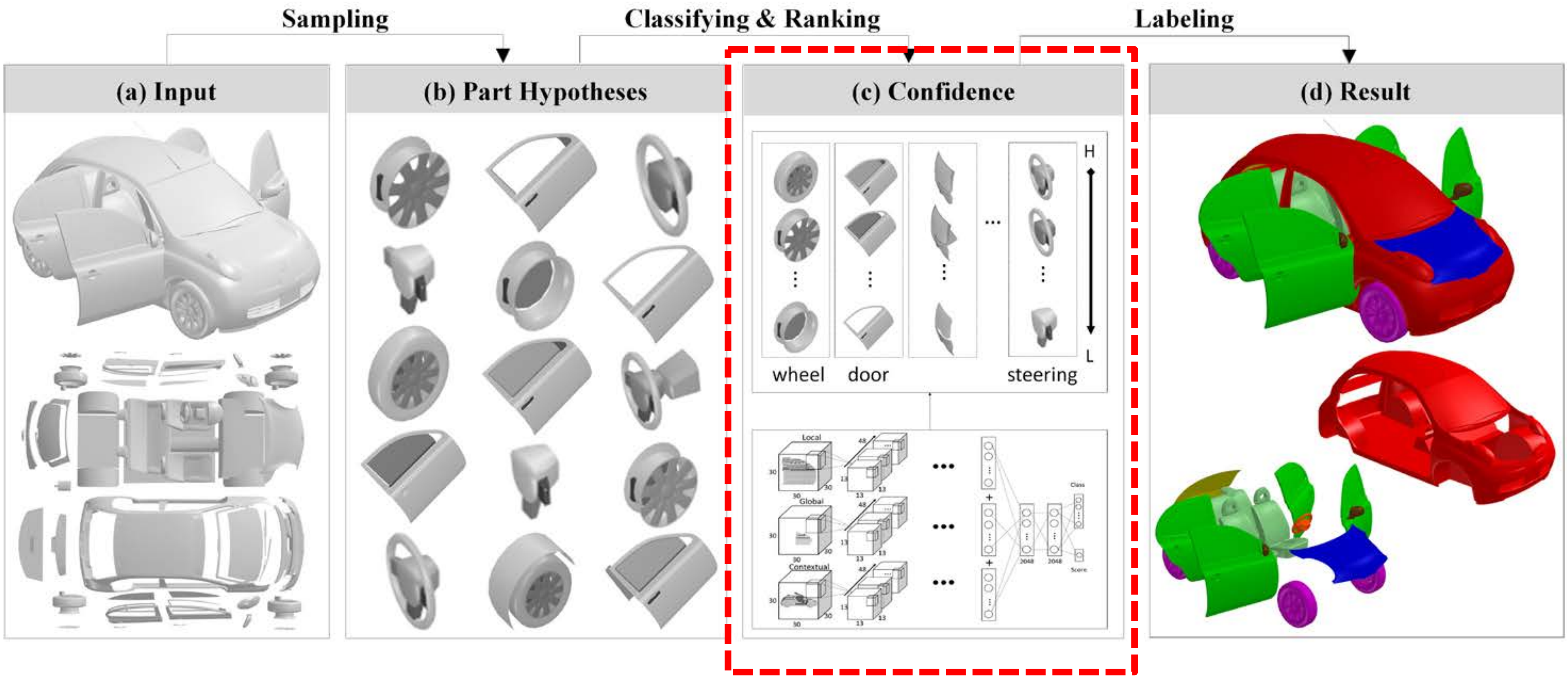
Part hypothesis quality vs. hypothesis count.

Sampling Results

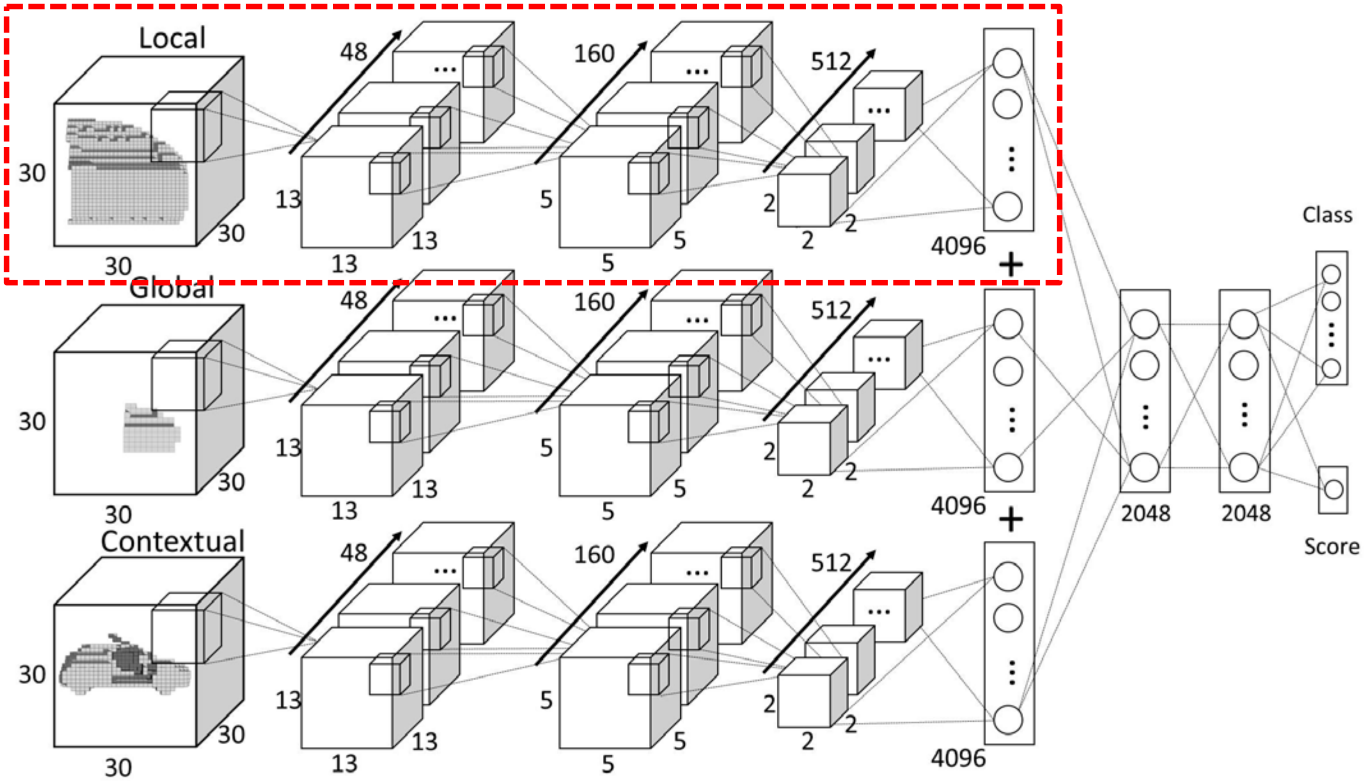


Comparison to Baseline (GMM and CNN-based).

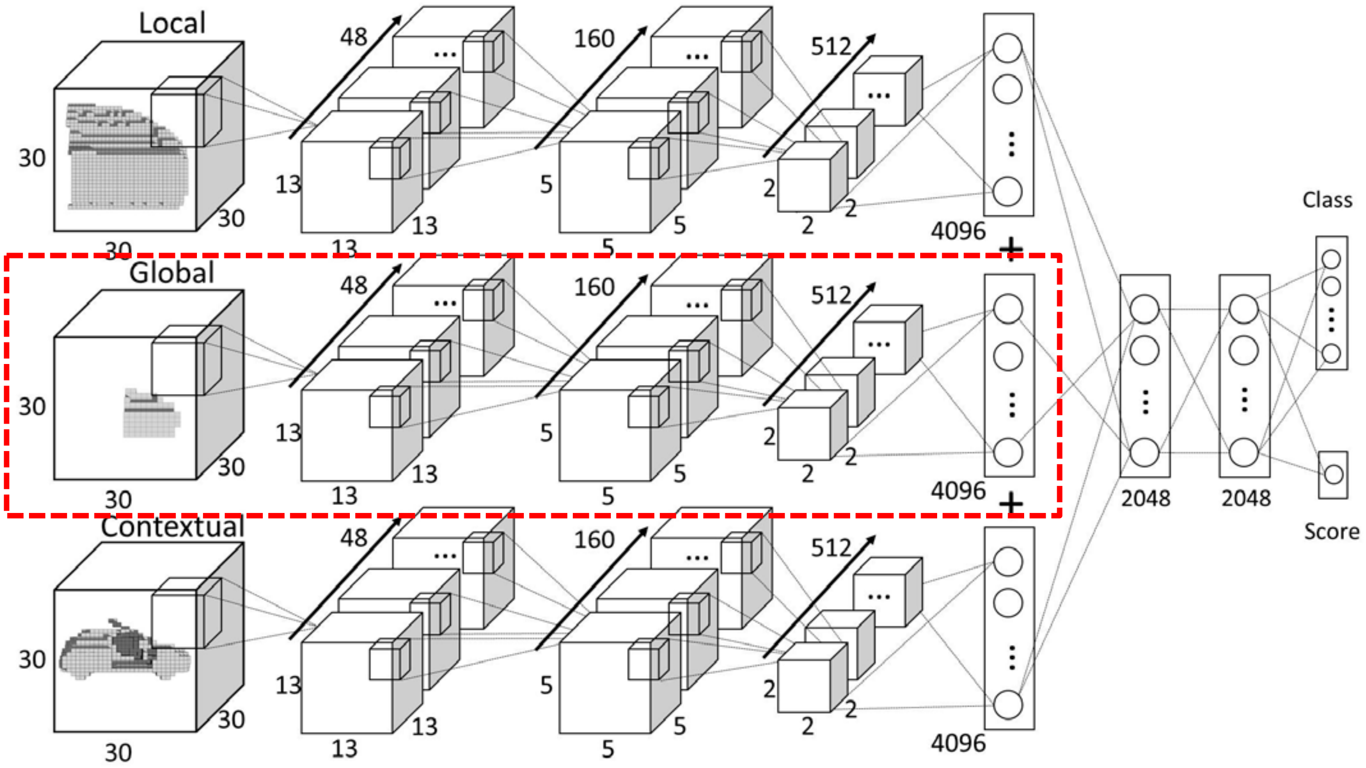
Pipeline


Sampling
Classifying & Ranking
Labeling
(a) Input
(b) Part Hypotheses
(c) Confidence
(d) Result
H
L
wheel
door
steering
Local
Global
Contextual

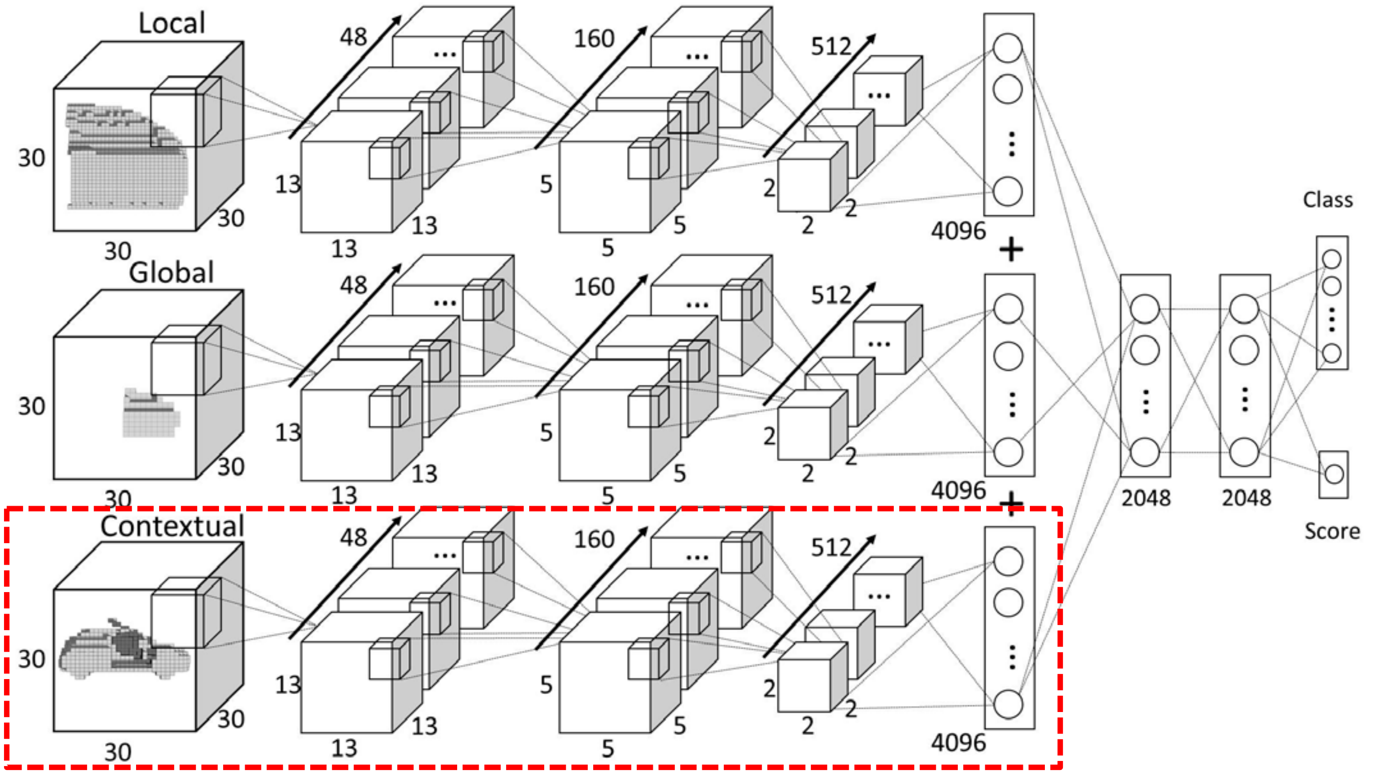
Classifiying and Ranking



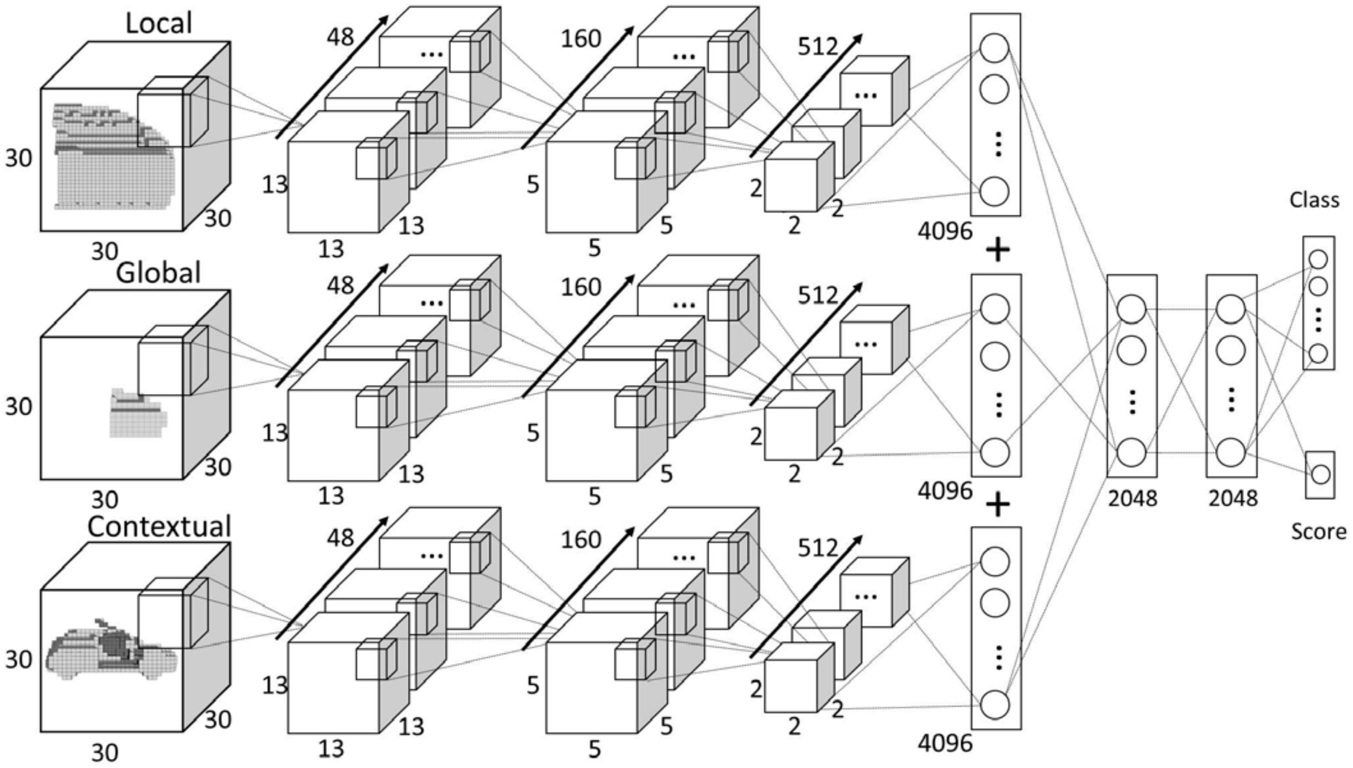
Classifiying and Ranking



Classifiying and Ranking

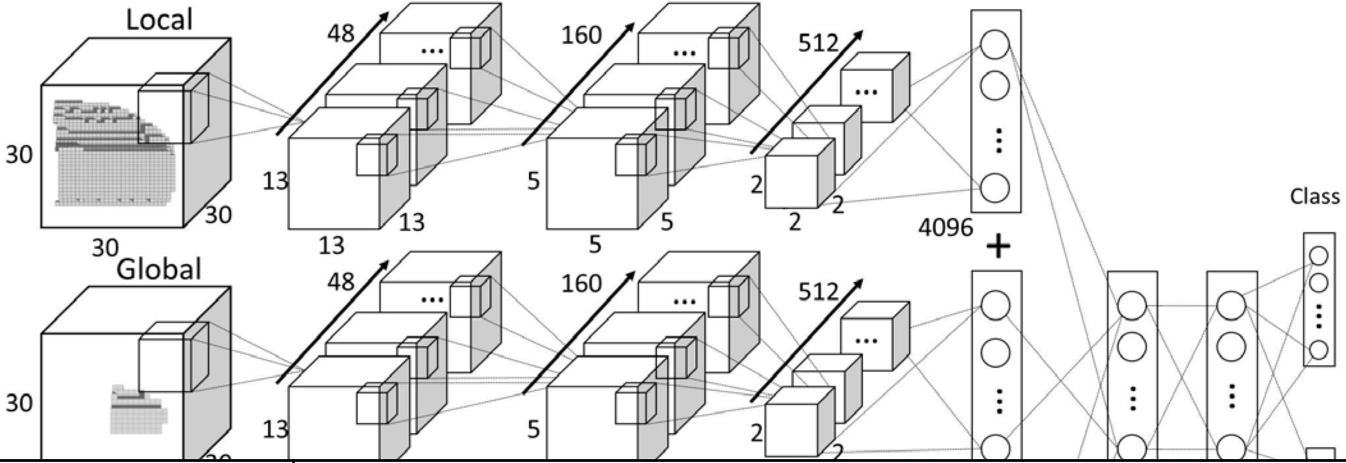


Classifiying and Ranking



$$L(p, r, c, s) = L_{\text{cls}}(p, c) + L_{\text{reg}}(r, s)$$

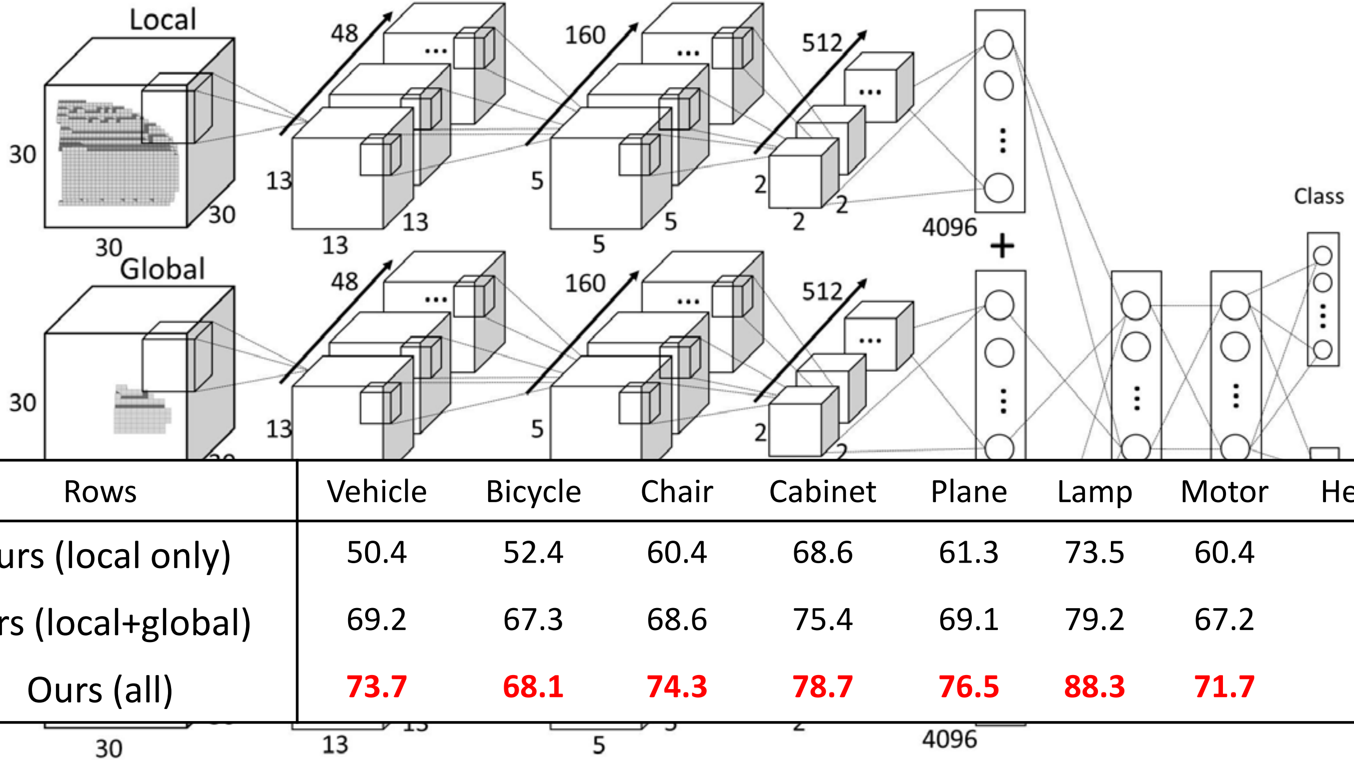
Classifiying and Ranking



Rows	Vehicle	Bicycle	Chair	Cabinet	Plane	Lamp	Motor	Helicopter	Living room	Office
Ours (local only)	50.4	52.4	60.4	68.6	61.3	73.5	60.4	78.5	62.7	54.8
Ours (local+global)	69.2	67.3	68.6	75.4	69.1	79.2	67.2	82.6	**68.3**	**76.4**
Ours (all)	**73.7**	**68.1**	**74.3**	**78.7**	**76.5**	**88.3**	**71.7**	**83.3**	66.1	65.4

$$L(p, r, c, s) = L_{\text{cls}}(p, c) + L_{\text{reg}}(r, s)$$

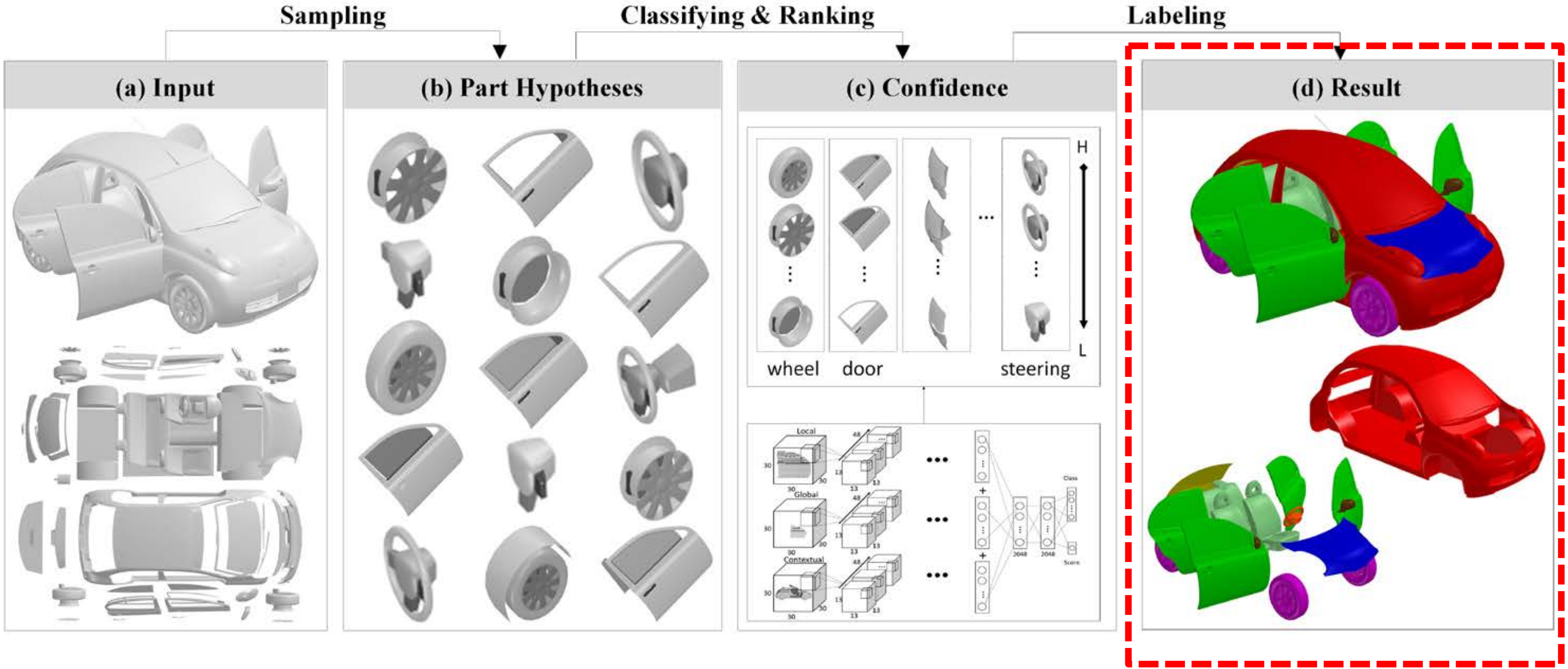
Classifiying and Ranking



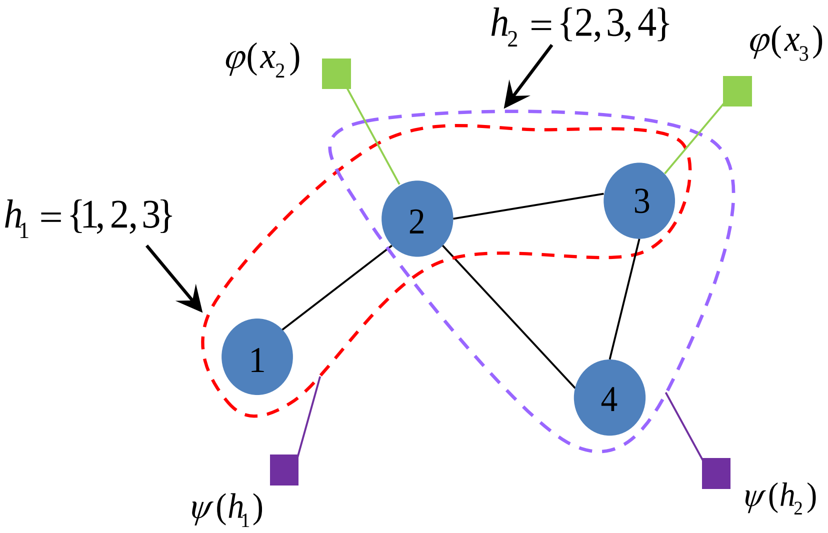
Rows	Vehicle	Bicycle	Chair	Cabinet	Plane	Lamp	Motor	Helicopter	Living room	Office
Ours (local only)	50.4	52.4	60.4	68.6	61.3	73.5	60.4	78.5	62.7	54.8
Ours (local+global)	69.2	67.3	68.6	75.4	69.1	79.2	67.2	82.6	**68.3**	**76.4**
Ours (all)	**73.7**	**68.1**	**74.3**	**78.7**	**76.5**	**88.3**	**71.7**	**83.3**	66.1	65.4

$$L(p, r, c, s) = L_{\text{cls}}(p, c) + L_{\text{reg}}(r, s)$$

Pipeline


Sampling
Classifying & Ranking
Labeling
(a) Input
(b) Part Hypotheses
(c) Confidence
(d) Result
H
L
wheel
door
steering
Local
Global
Contextual

Labeling via Higher-order CRF



$$E(L) = \sum_{c \in C} \underline{\varphi(x_c)} + \lambda \sum_{h \in \mathcal{H}} \psi(\mathbf{x}_h)$$

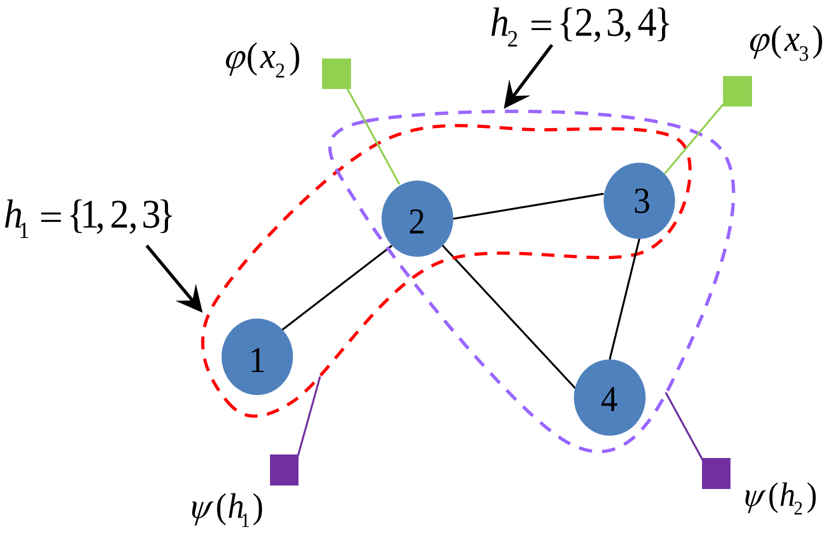
$$\Downarrow$$

$$\varphi(x_c) = -\log \underline{P(x_c = l_k)}$$

$$\Downarrow$$

$$P(x_c = l_k) = \frac{\sum_{i=1}^{K^c} e^{w_i^c s_i^c p(l_k | h_i^c)}}{\sum_{k=1}^{K} \sum_{j=1}^{K^c} e^{w_j^c s_j^c p(l_k | h_j^c)}}$$

Labeling via Higher-order CRF



$$E(L) = \sum_{c \in C} \varphi(x_c) + \lambda \sum_{h \in \mathcal{H}} \psi(\mathbf{x}_h)$$

$$\psi(\mathbf{x}_h) = \begin{cases} N(\mathbf{x}_h)\frac{1}{\eta}\gamma_{\max}, & \text{if } N(\mathbf{x}_h) \leq \eta \\ \gamma_{\max}, & \text{otherwise} \end{cases}$$

$$\gamma_{max} = e^{-G(h)/C^h}$$

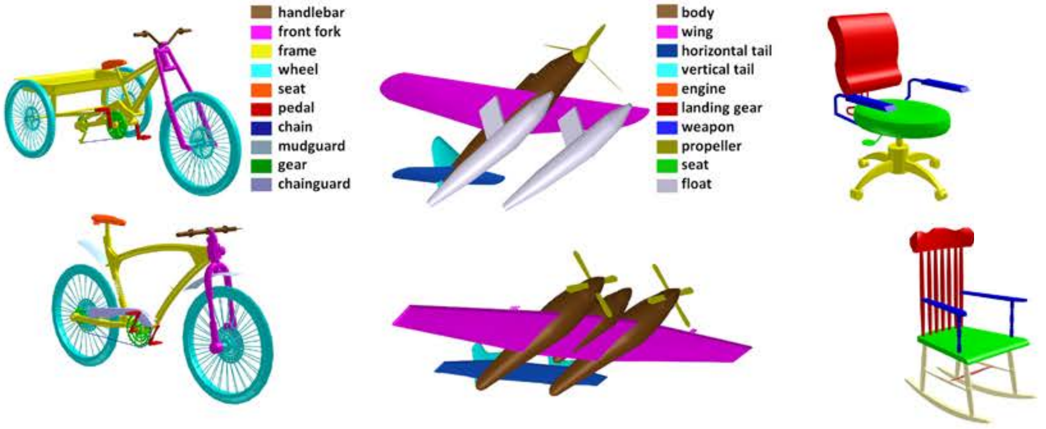
Experiments

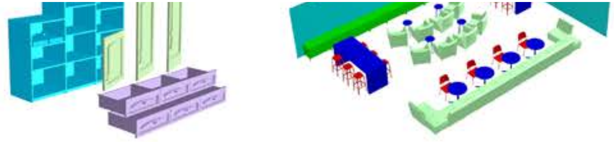
Experiments

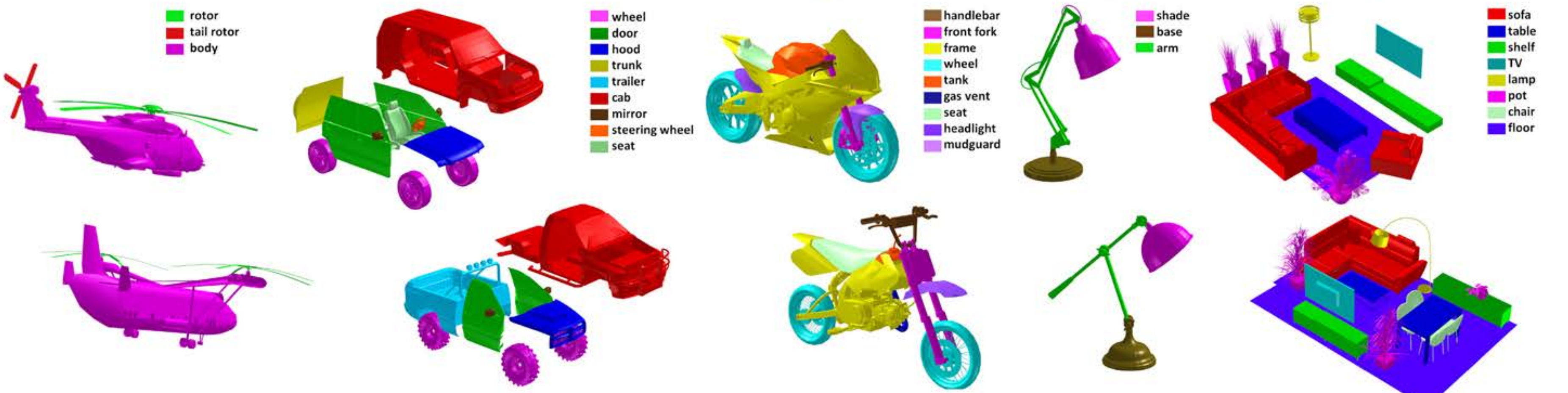
- Benchmark dataset
- Labeling results
- Labeling performance
- Parameter analyses

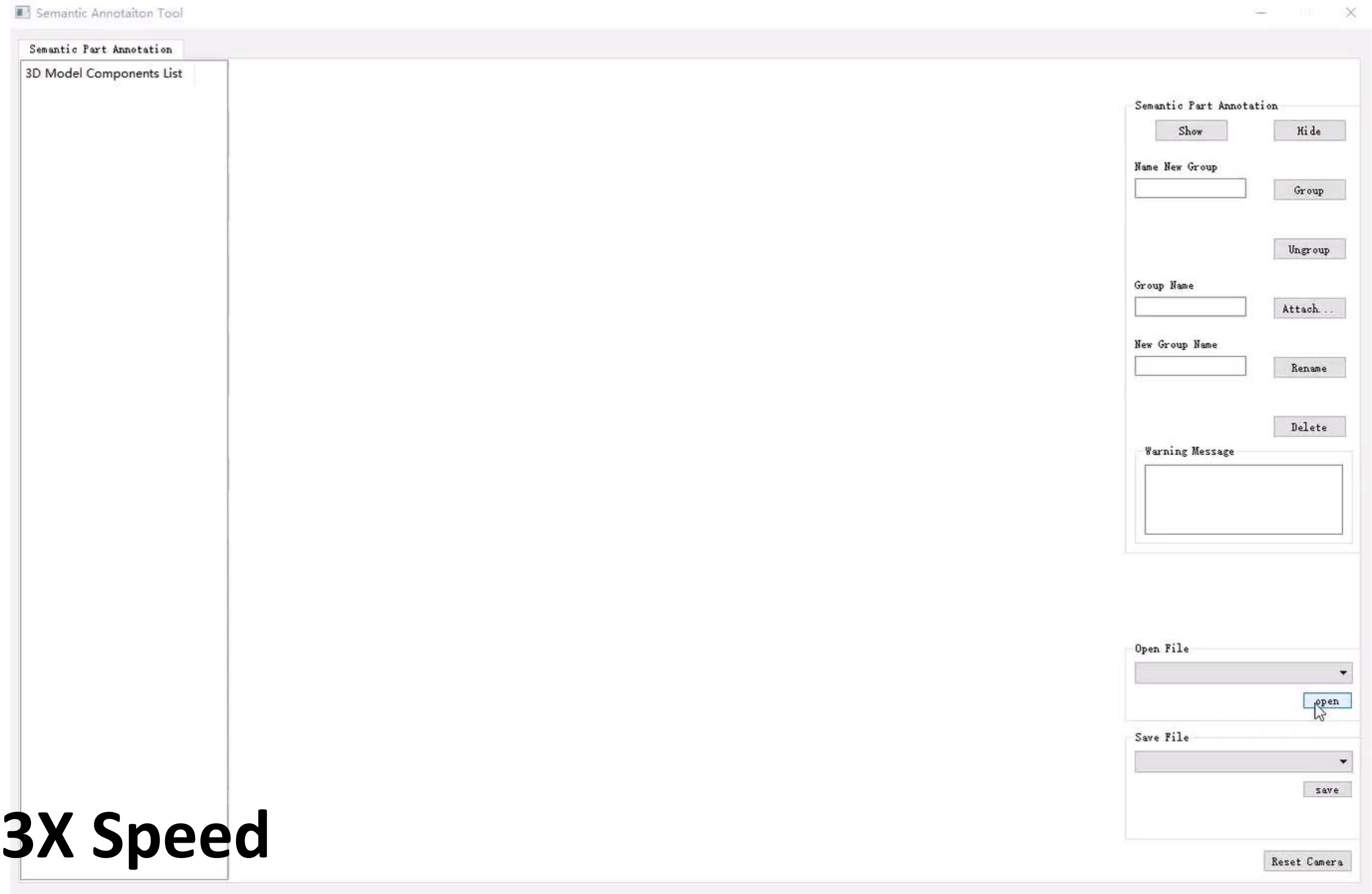
Benchmark Dataset

1019 models
8 object categories
2 scene categories





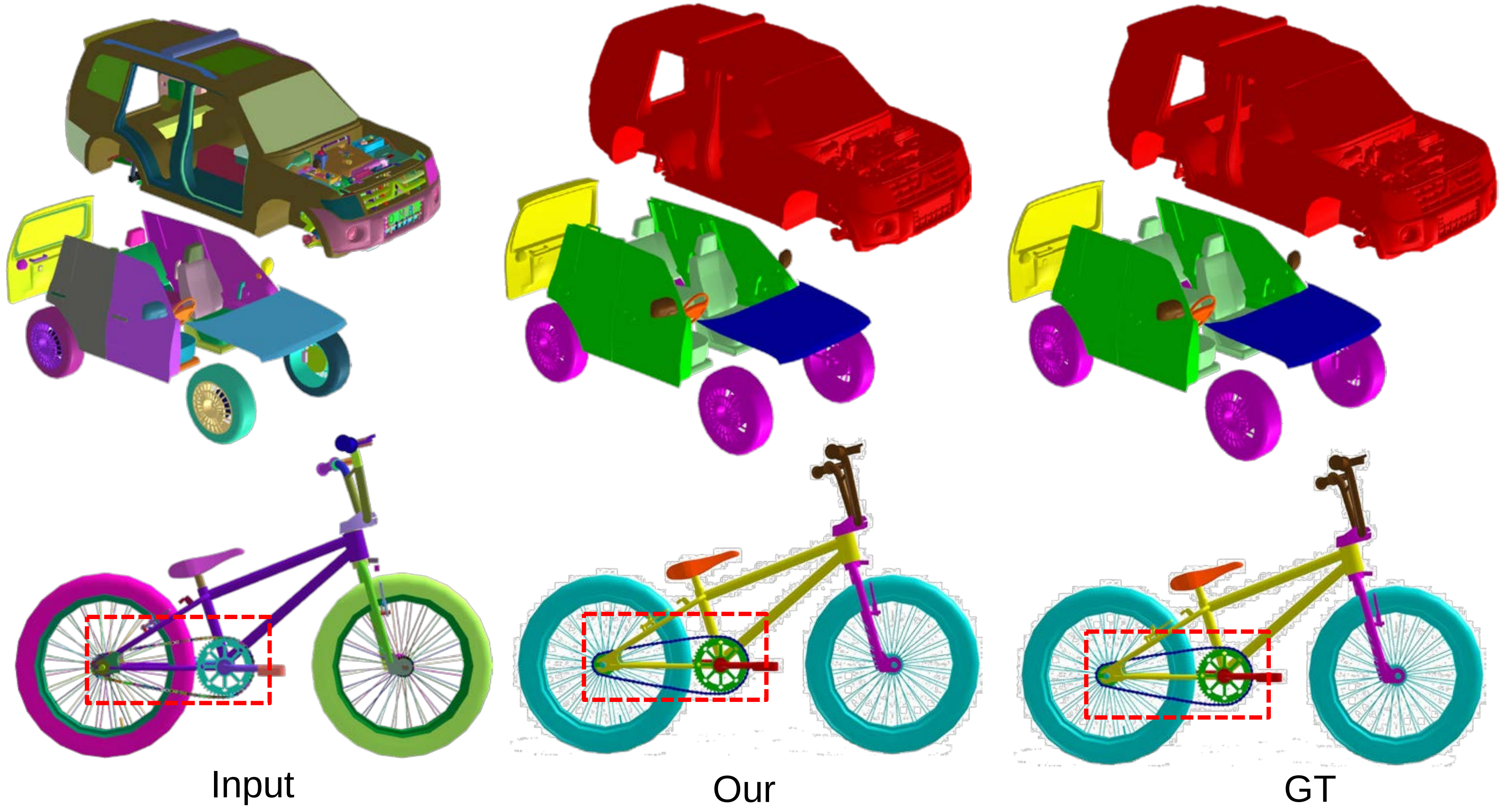


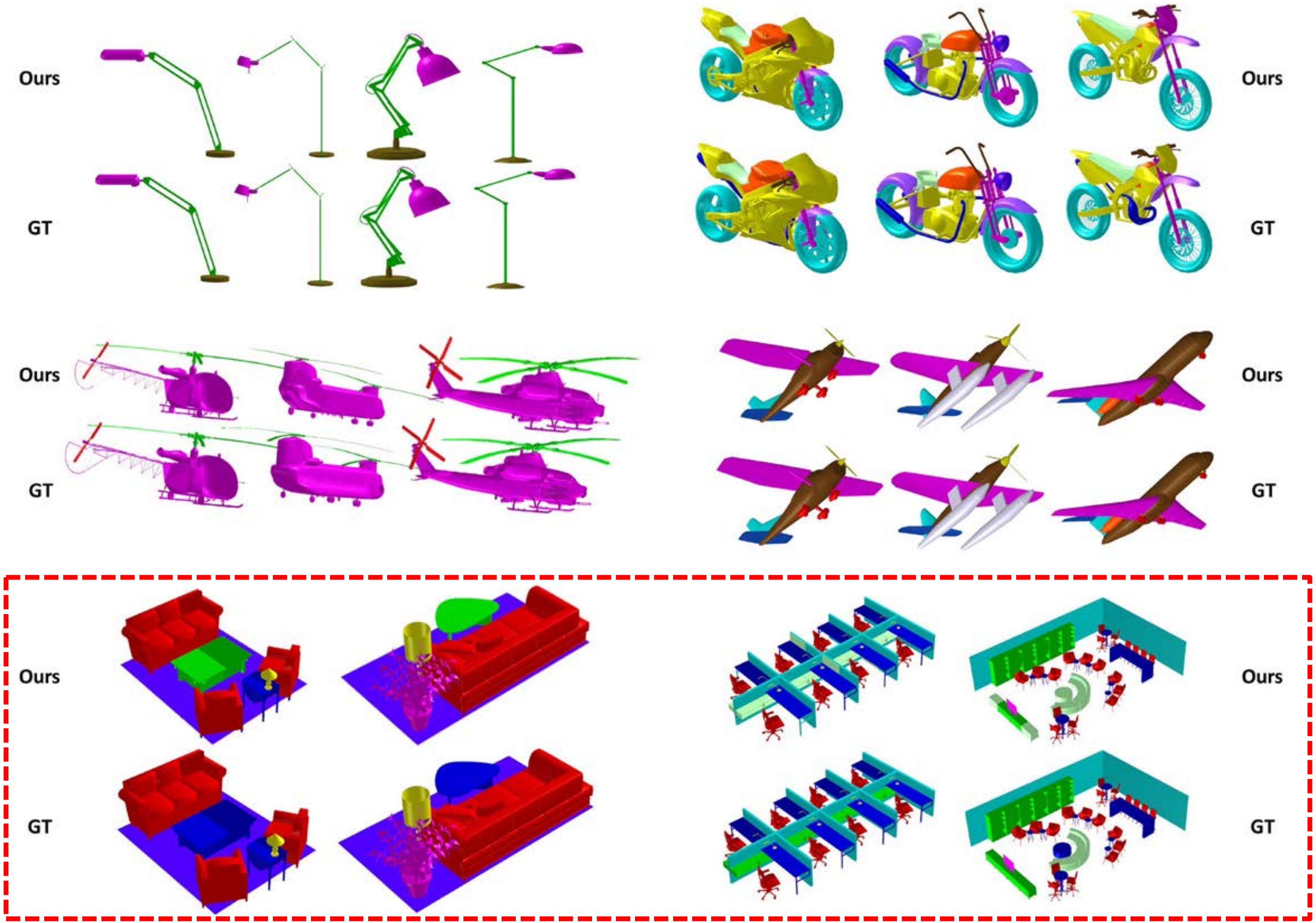


Semantic Annotaiton Tool
Semantic Part Annotation
3D Model Components List
Semantic Part Annotation
Show
Hide
Name New Group
Group
Ungroup
Group Name
Attach...
New Group Name
Rename
Delete
Warning Message
Open File
open
Save File
save
Reset Camera
3X Speed

Experiments

- Benchmark dataset
- Labeling Results
- Labeling performance
- Parameter analyses


Input
Our
GT

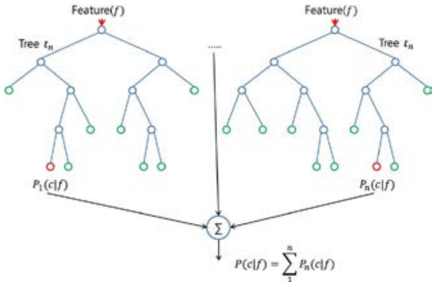

Ours
GT
Ours
GT
Ours
GT
Ours
GT
Ours
GT
Ours
GT

Experments

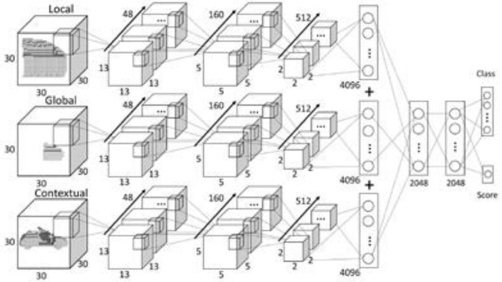
- Benchmark dataset
- Labeling results
- Labeling performance
- Parameter analysis

Experiment Results

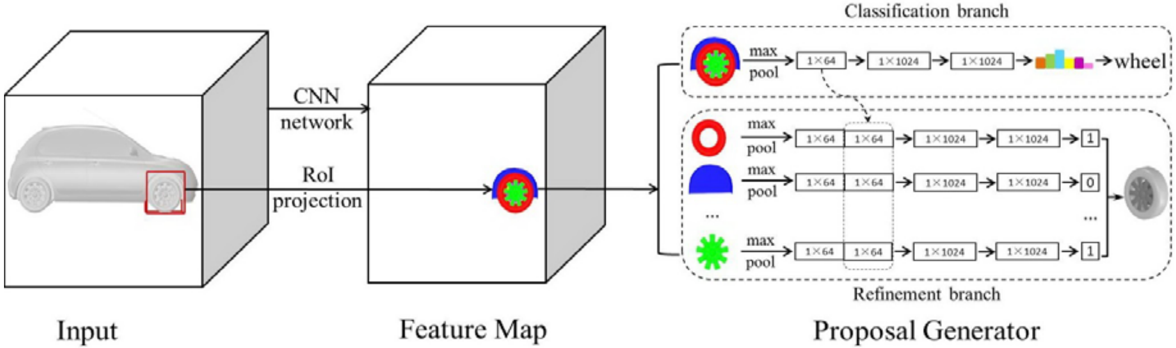
Comparison with three baseline methods



Random forest



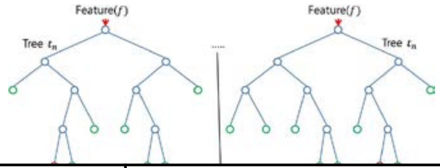
CNN-based component classification

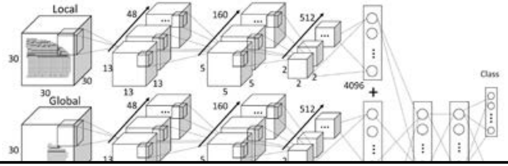


CNN-based hypothesis generation

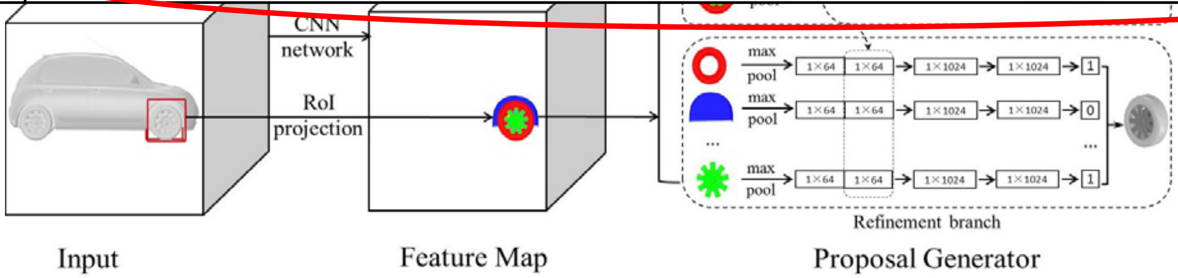
Experiment Results

Comparison with three baseline methods





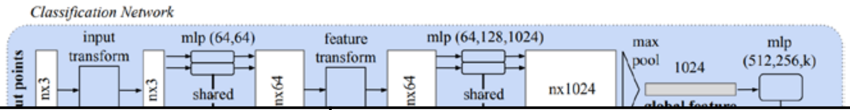
Rows	Vehicle	Bicycle	Chair	Cabinet	Plane	Lamp	Motor	Helicopter	Living room	Office
Baseline (Random Forest)	54.7	58.9	62.4	65.9	53.5	63.3	65.9	52.8	47.7	63.5
Baseline (CNN Classifier)	48.9	63.8	70.75	63.3	68.9	81.2	67.4	78.5	51.2	63.9
Baseline (CNN Hypo. Gen.)	56.3	51.9	68.5	45.7	58.5	71.1	53.1	72.2	58.6	65.1
Ours (all)	**73.7**	**68.1**	**74.3**	**78.7**	**76.5**	**88.3**	**71.7**	**83.3**	**66.1**	**65.4**



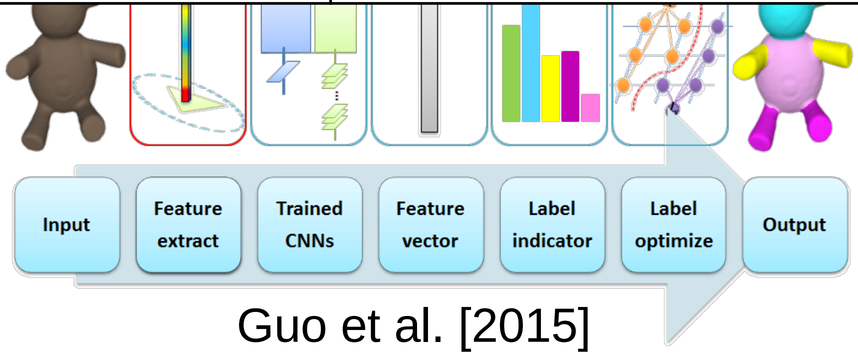
CNN-based hypothesis generation

Experiment Results

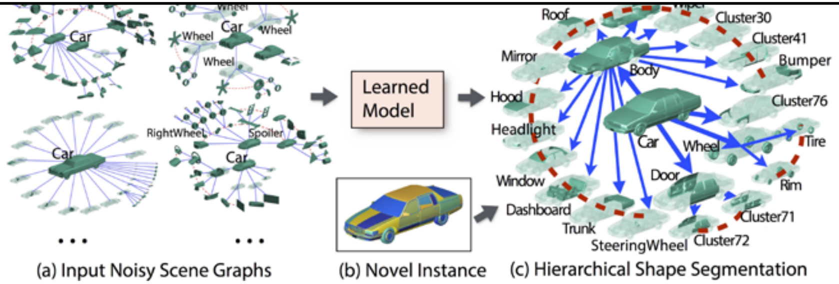
Comparison with 4 state-of-the-art methods



Rows	Vehicle	Bicycle	Chair	Cabinet	Plane	Lamp	Motor	Helicopter	Living room	Office
PointNet [Su et at. 2017]	24.3	30.6	68.6	21.0	47.2	46.3	35.8	32.6	-	-
PointNet++ [Qi et at. 2017]	51.7	53.8	69.3	62.0	53.9	79.8	62.2	79.3	-	-
Guo et al. [2015]	27.1	25.2	34.2	68.8	38.6	79.1	41.6	80.1	33.7	28.5
Yi et al. [2017a]	65.2	63.0	61.9	70.6	59.3	82.2	67.5	78.9	56.6	**68.6**
Ours (all)	**73.7**	**68.1**	**74.3**	**78.7**	**76.5**	**88.3**	**71.7**	**83.3**	**66.1**	65.4



Guo et al. [2015]

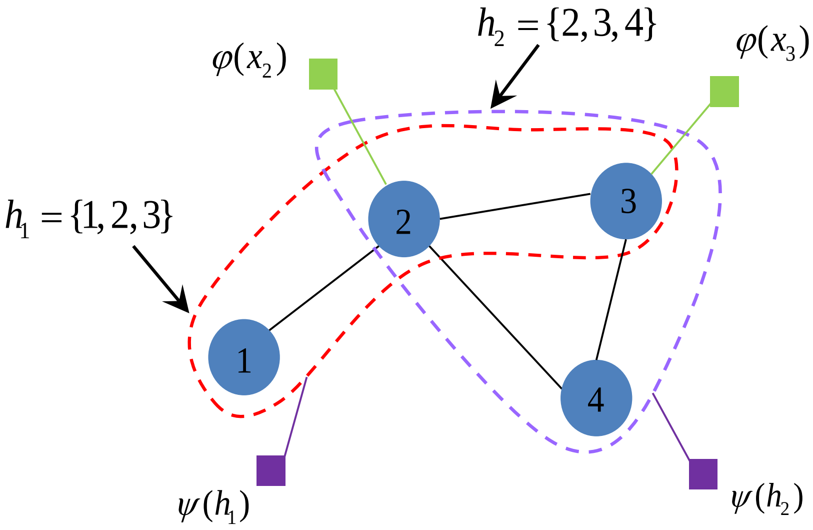


Yi et al. [2017]

Experments

- Benchmark dataset
- Labeling results
- Labeling performance
- Parameter analysis

Labeling performance without confidence score

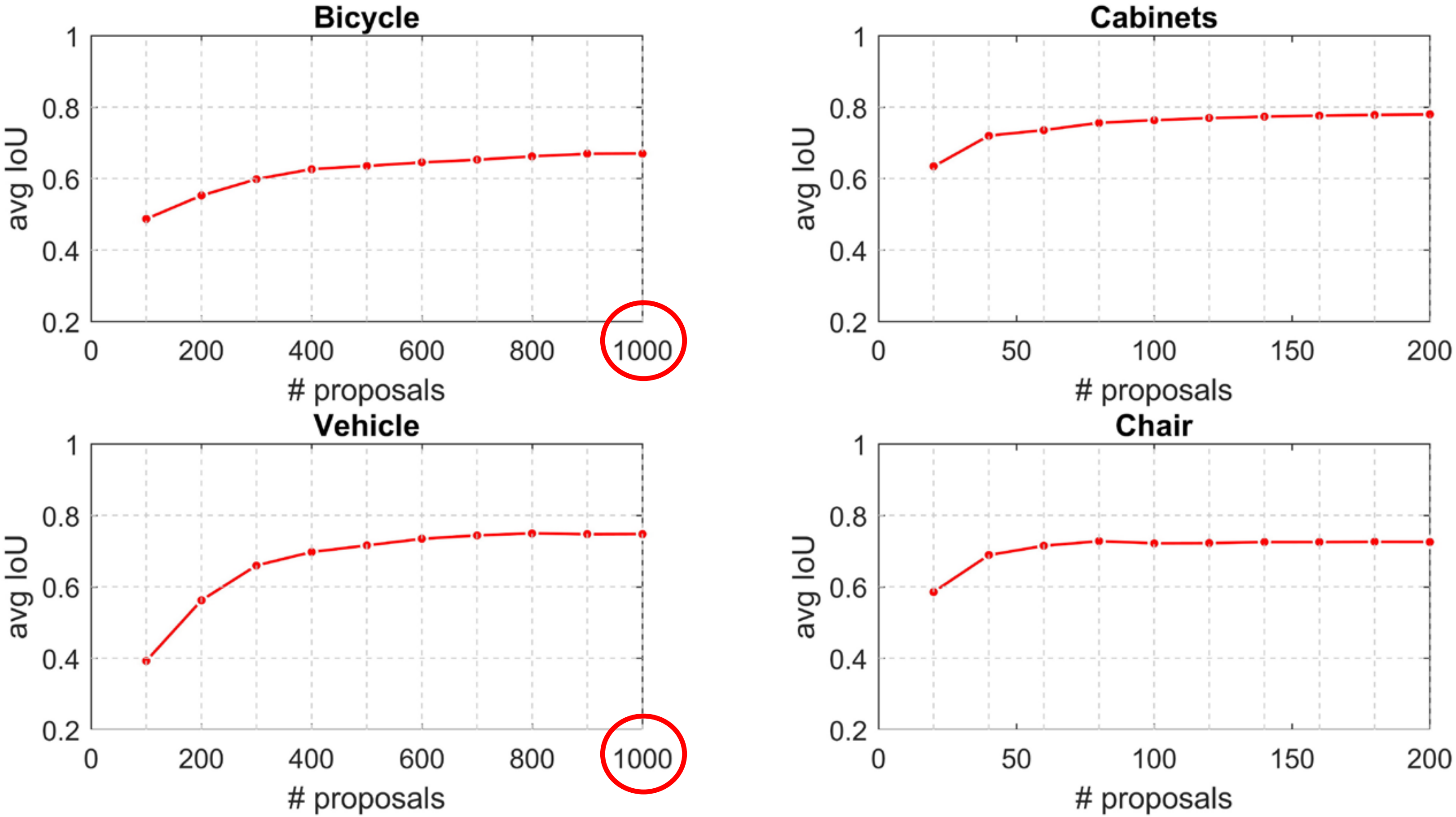


$$E(L) = \sum_{c \in C} \varphi(x_c) + \lambda \sum_{h \in \mathcal{H}} \psi(\mathbf{x}_h)$$

$$\varphi(x_c) = -\log P(x_c = l_k)$$

$$P(x_c = l_k) = \frac{\sum_{i=1}^{K^c} e^{w_i^c s_i^c p(l_k | h_i^c)}}{\sum_{k=1}^{K} \sum_{j=1}^{K^c} e^{w_j^c s_j^c p(l_k | h_j^c)}}$$

Rows	Vehicle	Bicycle	Chair	Cabinet	Plane	Lamp	Motor	Helicopter	Living room	Office
Ours (w/o score)	71.5	66.8	72.5	76.5	71.4	87.6	70.7	81.2	63.3	60.1
Ours (all)	**73.7**	**68.1**	**74.3**	**78.7**	**76.5**	**88.3**	**71.7**	**83.3**	**66.1**	**65.4**



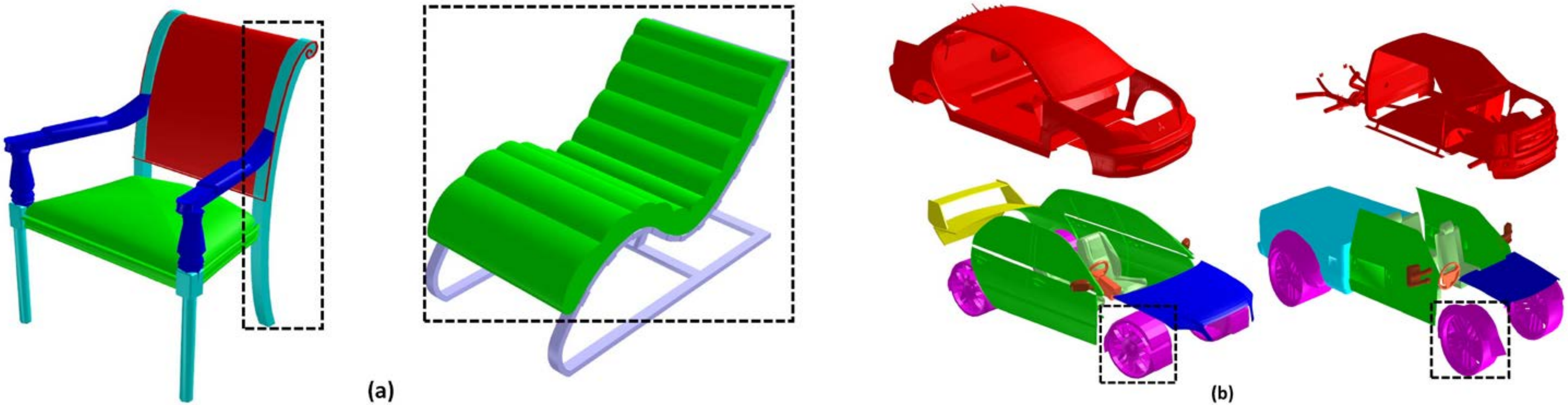
Labeling performance vs. part hypothesis count

Conclusion

- A new problem of segmentation of off-the-shelf 3D models with highly fine-grained components. And a benchmark with component-wise ground-truth labels
- A novel solution of part hypothesis generation based on a bottom-up hierarchical grouping process
- A deep neural network is trained to encode part hypothesis, rather than components
- A higher order potential adopts a soft constraint, providing more degree of freedom in optimal labeling search.

Limitations and Future Work

- Only groups the components but NOT segment
- Part hypotheses overlap significantly (shape concavity)
- Extend hypothesis for hierarchical segmentation, and Integrate CRF into the deep neural networks



(a) (b)

Q & A

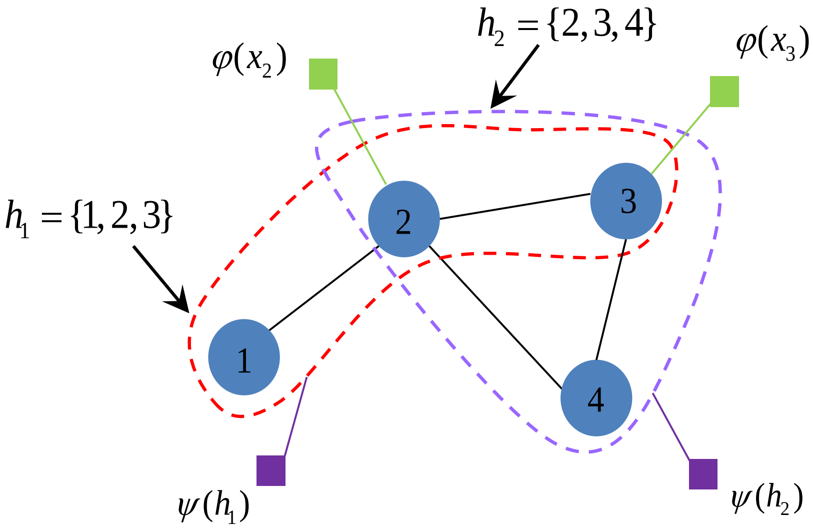
E-mail:

wangxiaogang@buaa.com.cn

Code&Dataset:

https://github.com/wangxiaogang866/fglabel

Parameter K^c



$$E(L) = \sum_{c \in C} \varphi(x_c) + \lambda \sum_{h \in \mathcal{H}} \psi(\mathbf{x}_h)$$

$$\varphi(x_c) = -\log P(x_c = l_k)$$

$$P(x_c = l_k) = \frac{\sum_{i=1}^{K^c} e^{w_i^c s_i^c p(l_k | h_i^c)}}{\sum_{k=1}^{K} \sum_{j=1}^{K^c} e^{w_j^c s_j^c p(l_k | h_j^c)}}$$

Rows	Vehicle	Bicycle	Chair	Cabinet	Plane	Lamp	Motor	Helicopter	Living room	Office
Ours (K^c = 1)	52.0	43.2	63.5	62.0	47.6	76.5	41.7	42.4	54.6	**70.7**
Ours (K^c = 3)	56.5	49.9	67.0	66.6	55.4	84.0	51.7	43.4	63.1	70.1
Ours (K^c = 5)	59.3	54.9	70.5	69.6	59.8	86.3	55.3	50.7	64.7	68.9
Ours (K^c = 10)	62.0	61.9	72.6	74.1	68.6	86.9	62.4	75.6	**66.6**	66.1
Ours (all)	**73.7**	**68.1**	**74.3**	**78.7**	**76.5**	**88.3**	**71.7**	**83.3**	66.1	65.4