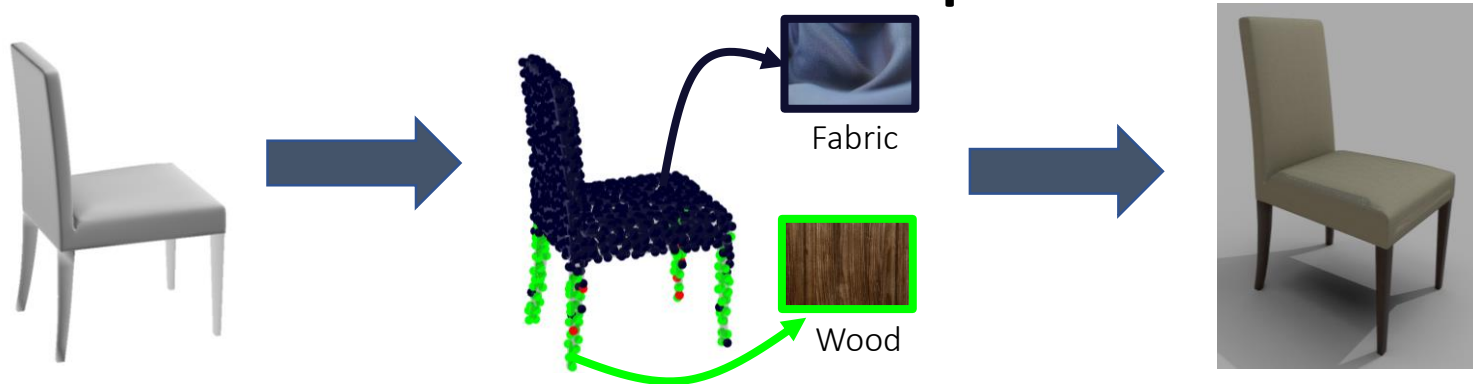


# Learning Material-Aware Local Descriptors for 3D Shapes



Hubert Lin<sup>1</sup> Melinos Averkiou<sup>2</sup> Evangelos Kalogerakis<sup>3</sup> Balazs Kovacs<sup>4</sup> Siddhant Ranade<sup>5</sup>  
Vladimir G. Kim<sup>6</sup> Siddhartha Chaudhuri<sup>6,7</sup> Kavita Bala<sup>1</sup>

<sup>1</sup>Cornell Univ. <sup>2</sup>Univ. of Cyprus <sup>3</sup>UMass Amherst <sup>4</sup>Zoox <sup>5</sup>Univ. of Utah <sup>6</sup>Adobe <sup>7</sup>IIT Bombay

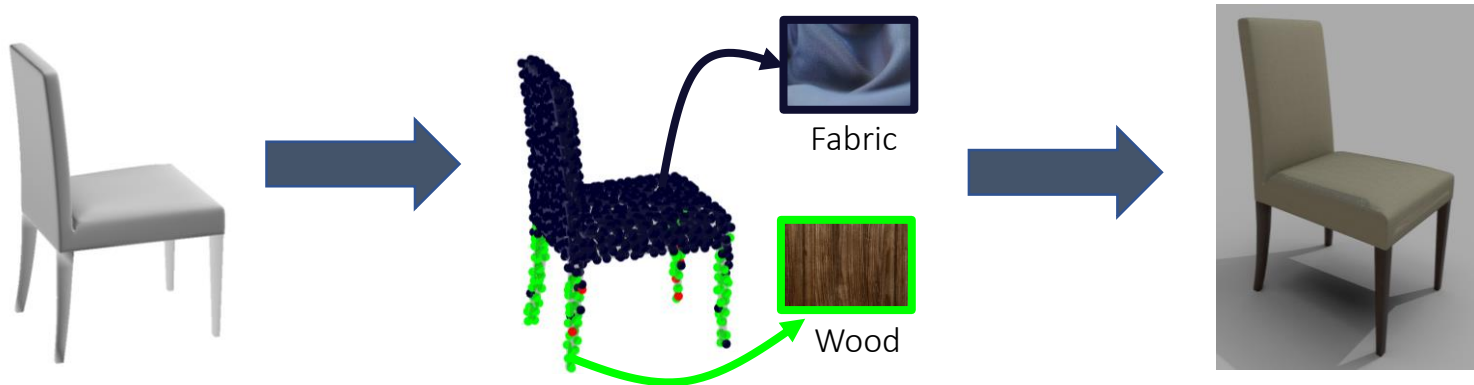
# Outline

1. Goal
2. Motivation
3. Related Work
4. Data Collection
5. Network Architecture and Training Pipeline
6. Post-Processing
7. Results
8. Future Directions

# Outline

1. Goal
2. Motivation
3. Related Work
4. Data Collection
5. Network Architecture and Training Pipeline
6. Post-Processing
7. Results
8. Future Directions

Goal: Learn local shape descriptors  
sensitive to physical material



# Outline

1. Goal
- 2. Motivation**
3. Related Work
4. Data Collection
5. Network Architecture and Training Pipeline
6. Post-Processing
7. Results
8. Future Directions

# Motivation

Understanding physical material properties from 3D geometry:

- Jointly reason about materials and geometry
- Interactive design tool
- Robotic perception
- ...



[Morrison et al 2018]

# Motivation

Jointly reason about materials and geometry



?



What material is typically used for an object part like this?

How can we retrieve objects that are composed of similar materials?

...

# Motivation

## Design and fabrication



Which material is suitable for fabrication?

Wood ✓

Metal ✓

Glass ✗



# Motivation

## Design and fabrication

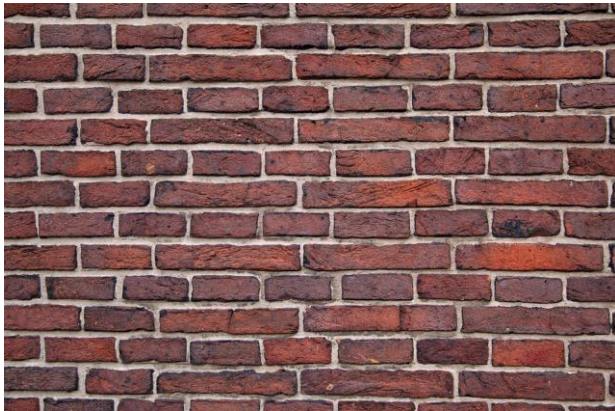


Suggested materials



# Motivation

## Robotic Perception



Which one is better for an emergency collision?



Which one requires more gentle handling?



# Outline

1. Goal
2. Motivation
- 3. Related Work**
4. Data Collection
5. Network Architecture and Training Pipeline
6. Post-Processing
7. Results
8. Future Directions

# Related Work

1. Shape databases
2. Deep learning for shape analysis
3. Material understanding for shapes
4. Material understanding for images

# Shape Databases

## ShapeNet

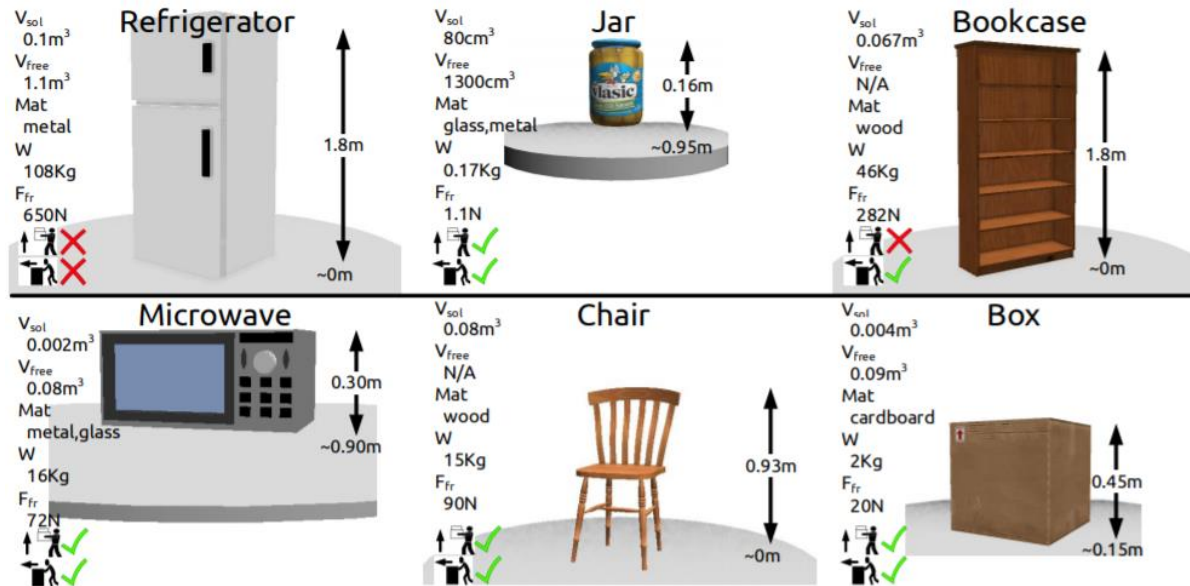
- Large-scale database with many object classes
- Some shapes are textured; part segmentation



# Shape Databases

## Semantically-Enriched 3D Models for Common-sense Knowledge

- Many different annotations, including category-level priors over material labels



# Shape Databases

## Text2Shape

- Natural language descriptions for 3D shapes
- Joint text / shape embedding

### a) 3D shapes and natural language descriptions



Circular glass coffee table with two sets of wooden legs that clasp over the round glass edge.

1



A brown wooden moon shaped table with three decorative legs with a wooden vine shaped decoration base connecting the legs.

2a

Wooden half round table.

2b

[Chen et al 2018]

# Deep Learning for Shape Analysis

Based on...

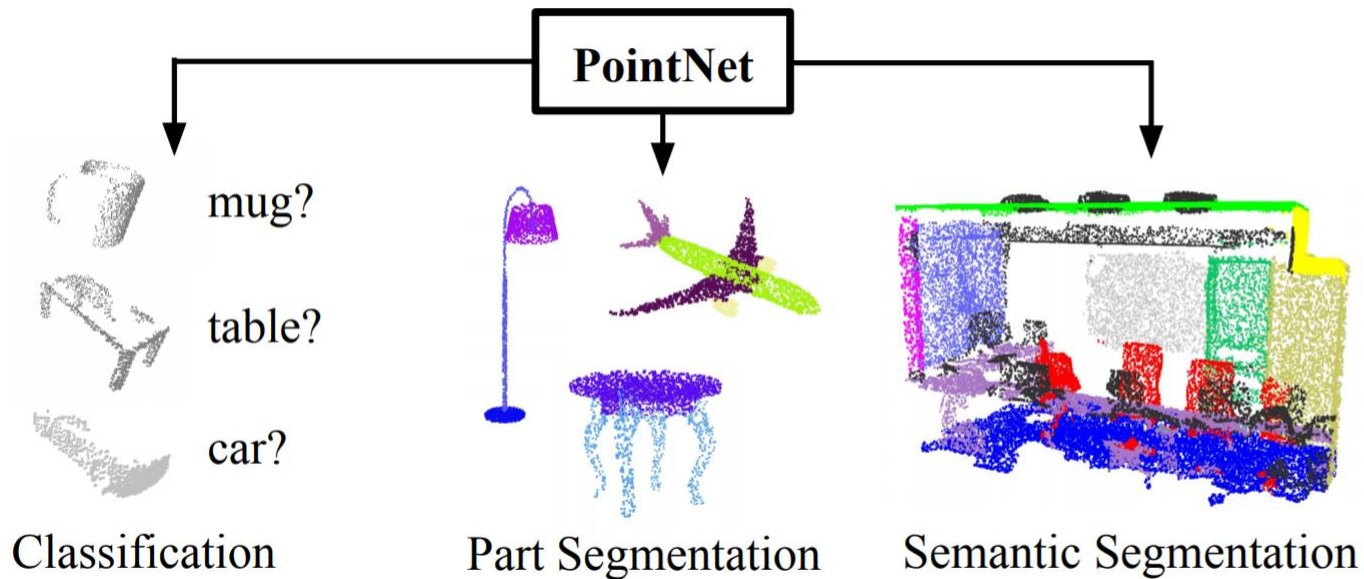
- Mesh
- Canonicalized meshes
- 2D renderings
- Point sets
- Dense Voxels
- Voxel octrees
- Spectral alignment
- Surface patch collection

And more...



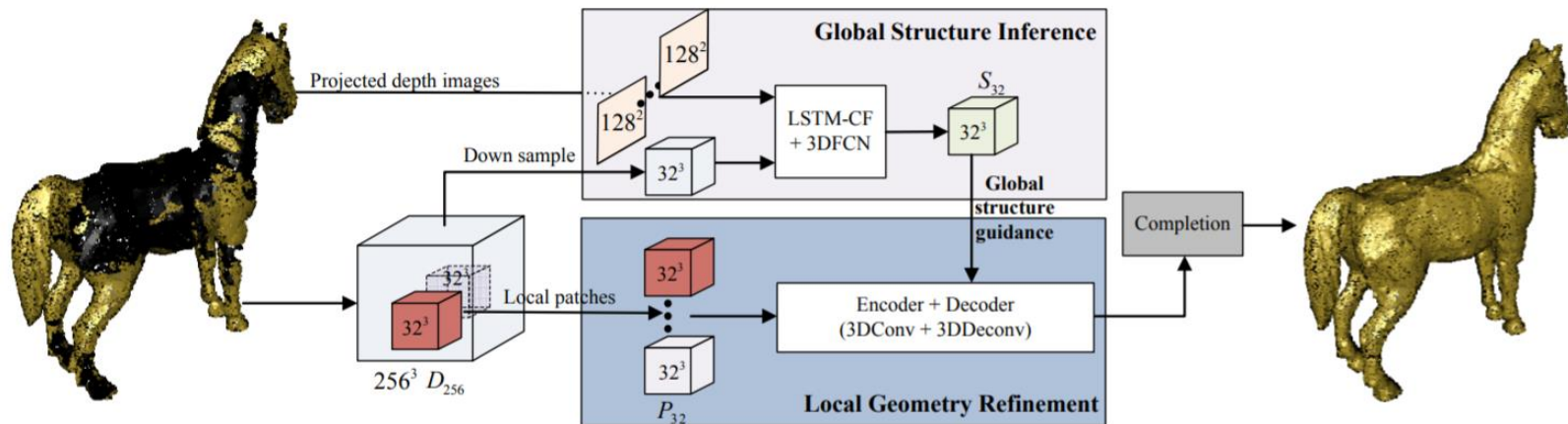
# Deep Learning for Shape Analysis

- Segmentation, classification



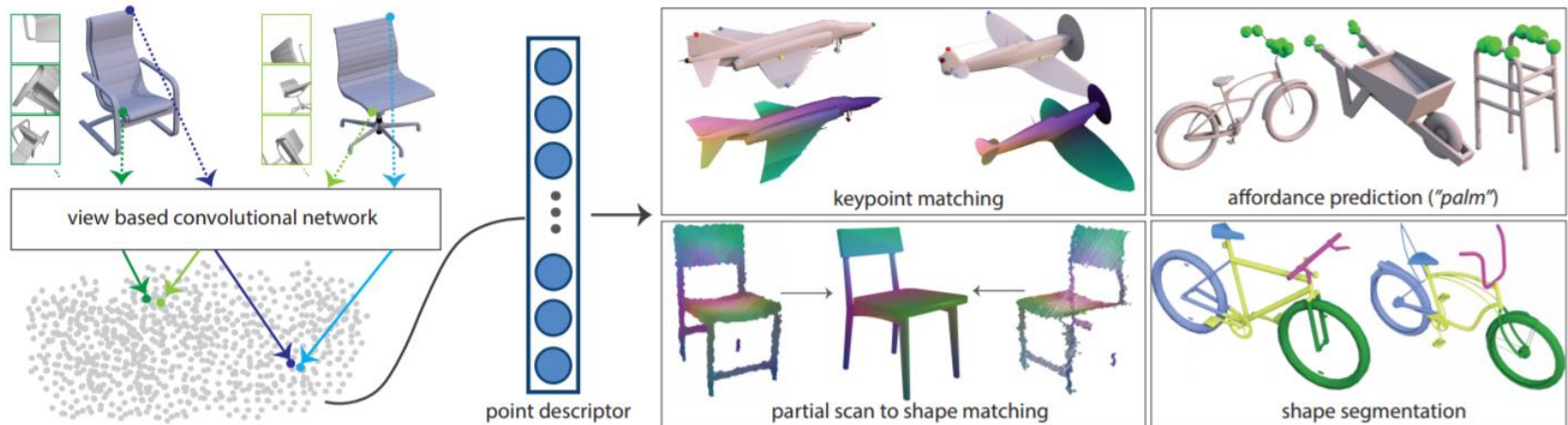
# Deep Learning for Shape Analysis

- Shape completion



# Deep Learning for Shape Analysis

- Geometric descriptors



# Material Understanding for Shapes

## Material Memex

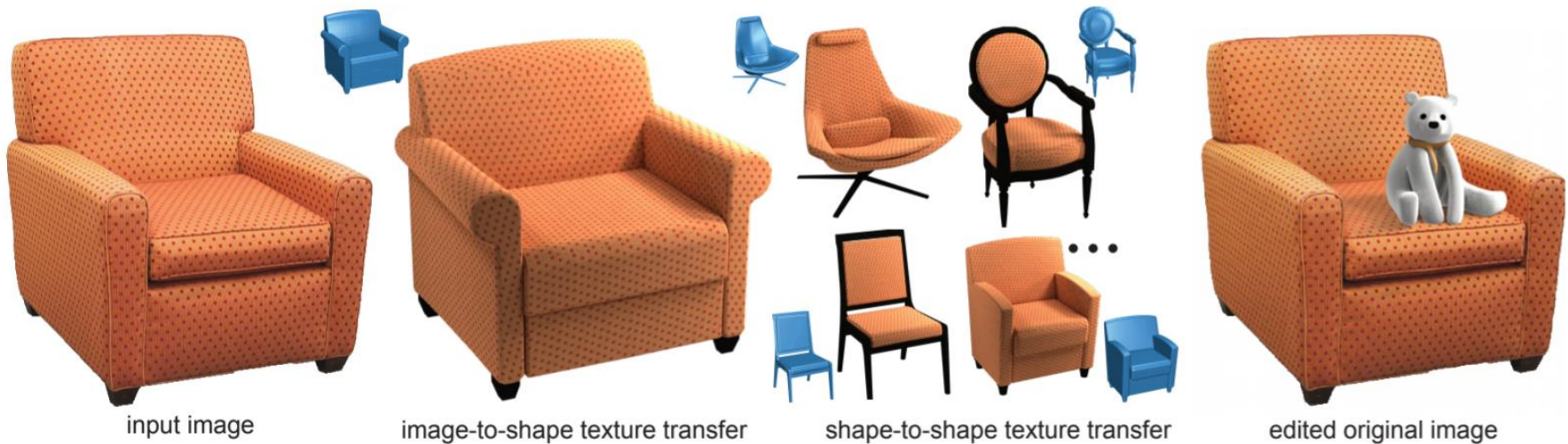
- Automatic material suggestion for parts
- Requires database of with known part properties



# Material Understanding for Shapes

## Unsupervised Texture Transfer from Images to Shapes

- Image-to-shape, shape-to-shape texture transfer
- Aligns user-specified image to shape



# Material Understanding for Shapes

## Magic Decorator: Indoor Material Suggestion

- Automatically suggest textures for indoor 3D scene
- Used color / texture statistics of 2D images
- Requires scene segmented and labeled



Input Scene

Suggestion 1

Suggestion 2

Suggestion 3





# Material Understanding for Images

## Describable Textures Dataset

- Textures described by attributes (“striped”, ...)
- Dataset of representative textures





# Material Understanding for Images

## OpenSurfaces

- Segmented surfaces from consumer photographs labelled with material and appearance properties



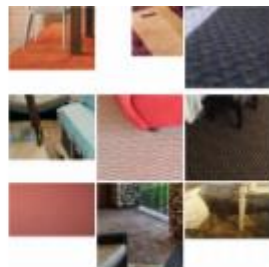
# Material Understanding for Images

## Materials in Context Database

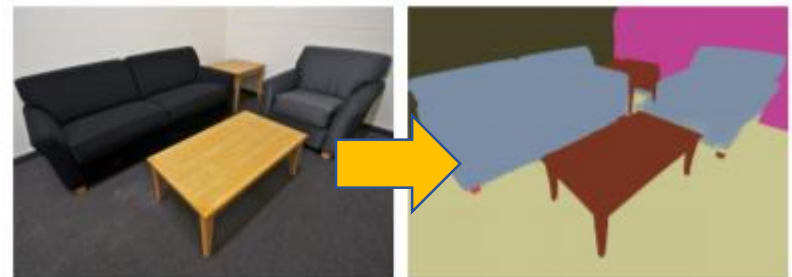
- Millions of material points in real-world images
- Strong material recognition performance with deep learning



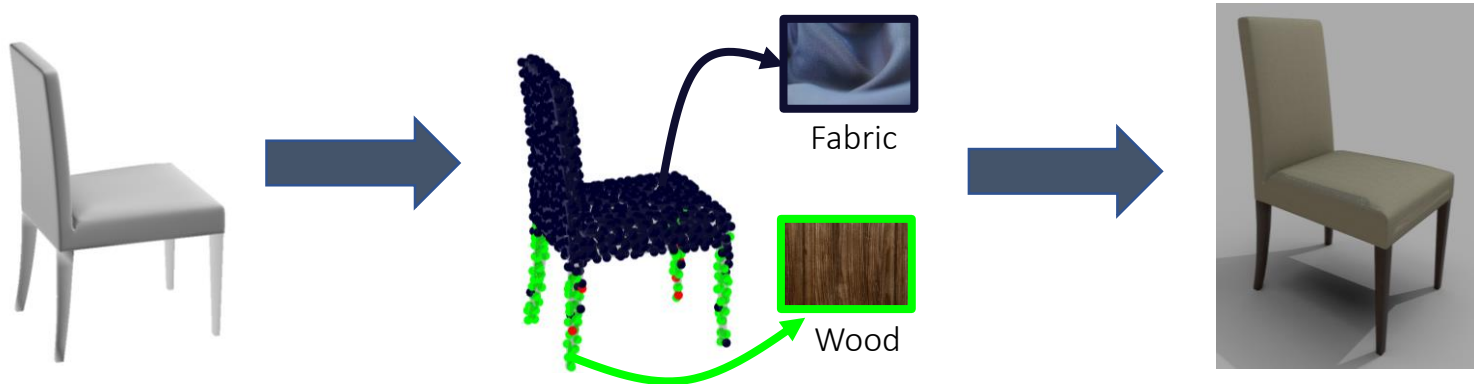
Brick



Carpet



Reminder: Learn local shape descriptors sensitive to physical material



## Our work :

- Focuses on physical material rather than appearance
- Does not strictly require additional input (such as semantic segmentation, image-to-shape matching, parts, ...)
- Only uses shape geometry as input
- Leverages existing deep learning approaches

# Outline

1. Goal
2. Motivation
3. Related Work
- 4. Data Collection**
5. Network Architecture and Training Pipeline
6. Post-Processing
7. Results
8. Future Directions

Challenge: Existing data  
is insufficient

# Crowdsourced Data

- Selected 17K chairs, tables, cabinets from ShapeNet
- Remove hard-to-label shapes for reliable crowdsourced annotations
- **Remaining shapes (17K)**



# Crowdsourced Data

- Selected 17K chairs, tables, cabinets from ShapeNet
- Remove hard-to-label shapes for reliable crowdsourced annotations
- **Remaining shapes (12K)**





# Crowdsourced Data

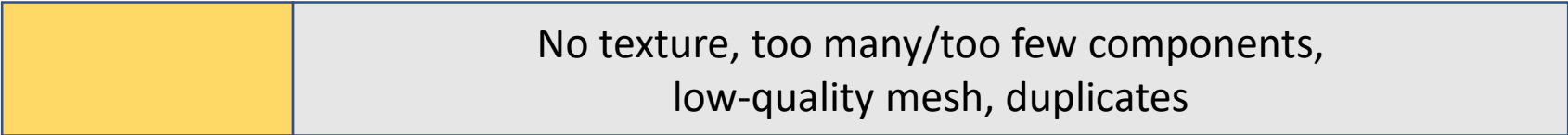
- Selected 17K chairs, tables, cabinets from ShapeNet
- Remove hard-to-label shapes for reliable crowdsourced annotations
- **Remaining shapes (8K)**



No texture, too many/too few components

# Crowdsourced Data

- Selected 17K chairs, tables, cabinets from ShapeNet
- Remove hard-to-label shapes for reliable crowdsourced annotations
- **Remaining shapes (3K)**



No texture, too many/too few components,  
low-quality mesh, duplicates

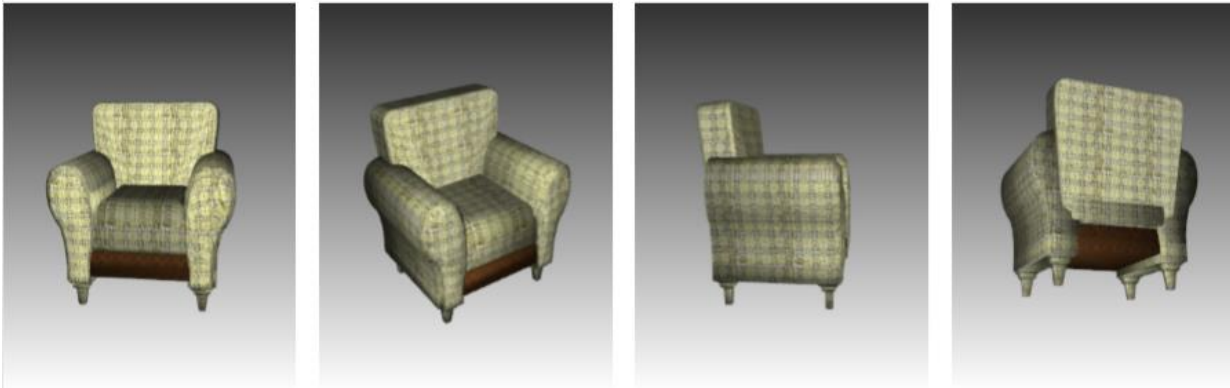
# Crowdsourced Data

Material categories (commonly found in furniture):

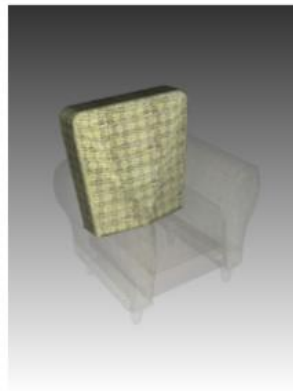
1. Wood
2. Plastic
3. Metal
4. Glass
5. Fabric (including leather)
- ~~6. Stone~~

# Crowdsourced Data

Here are a few views of a 3D object:



Now look carefully at the selected part of this 3D object below (rest of the object is faded):



What material is this part made of?

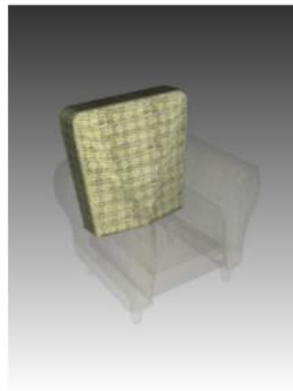
- Fabric / Leather
- Glass
- Metal
- Plastic
- Metal OR Plastic
- Stone
- Wood
- Can't tell / None of the above

# Crowdsourced Data

Here are a few views of a 3D object:



Now look carefully at the selected part of this 3D object below (rest of the object is faded):



What material is this part made of?

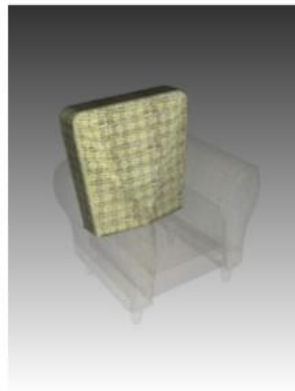
- Fabric / Leather
- Glass
- Metal
- Plastic
- Metal OR Plastic
- Stone
- Wood
- Can't tell / None of the above

# Crowdsourced Data

Here are a few views of a 3D object:



Now look carefully at the selected part of this 3D object below (rest of the object is faded):

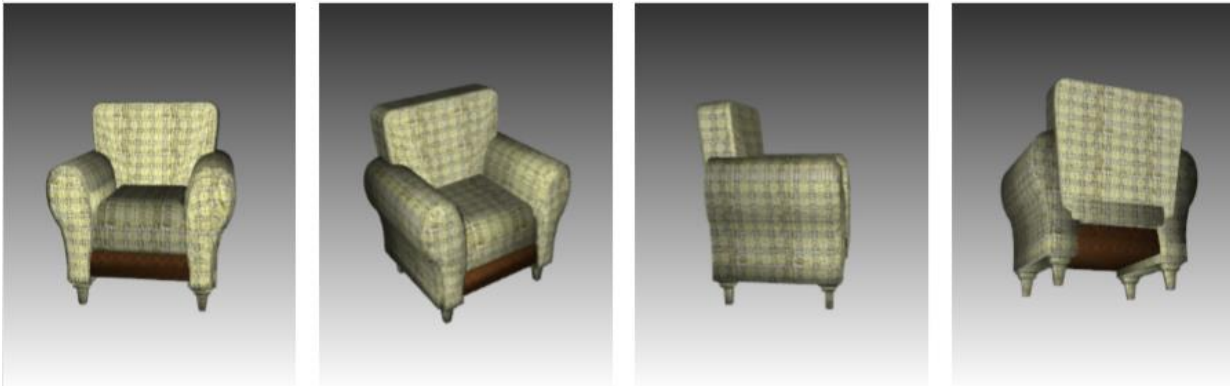


What material is this part made of?

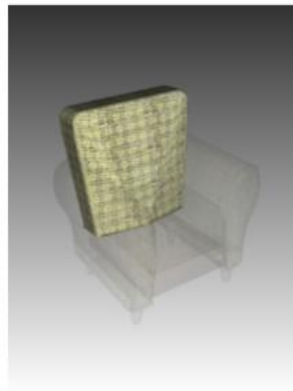
- Fabric / Leather
- Glass
- Metal
- Plastic
- Metal OR Plastic
- Stone
- Wood
- Can't tell / None of the above

# Crowdsourced Data

Here are a few views of a 3D object:



Now look carefully at the selected part of this 3D object below (rest of the object is faded):



What material is this part made of?

- Fabric / Leather
- Glass
- Metal
- Plastic
- Metal OR Plastic
- Stone
- Wood
- Can't tell / None of the above

# Crowdsourced Data

- 20 questions per task
- 3 sentinels per task
- Ignored labels from workers who incorrectly labeled sentinels or selected “Can’t tell” too often
- 5 votes per part, with 4+/5 considered reliable
- Parts with transparent textures labelled as glass (manually checked)



# Expert-Annotated Data

- Crowdsourced data is noisy
- Only one label assigned per part, but...



e.g. This seat body  
can be made of  
wood or plastic.

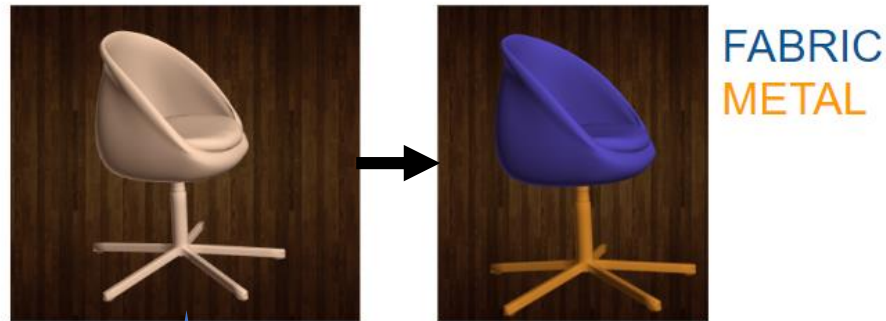
- Need high quality annotations for evaluation
- Selected 115 chairs, tables, cabinets from 3D Warehouse and Herman Miller

[<https://3dwarehouse.sketchup.com/>]

[<https://www.hermanmiller.com/resources/models/3d-models>]

# Expert-Annotated Data

Expert annotators reference product images and descriptions for accurate labelling

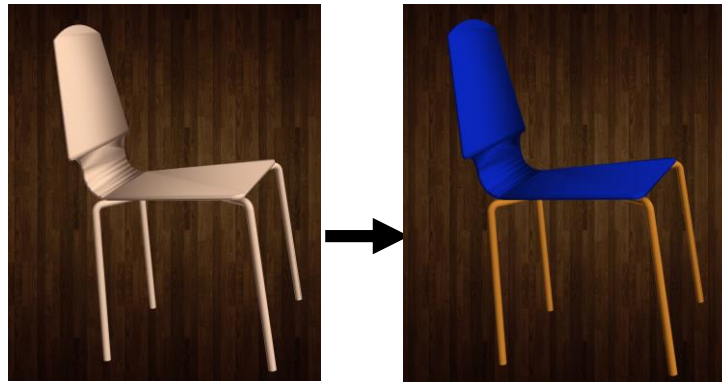


Manufacturer  
Product Images



# Expert-Annotated Data

Expert annotators reference product images and descriptions for accurate labelling



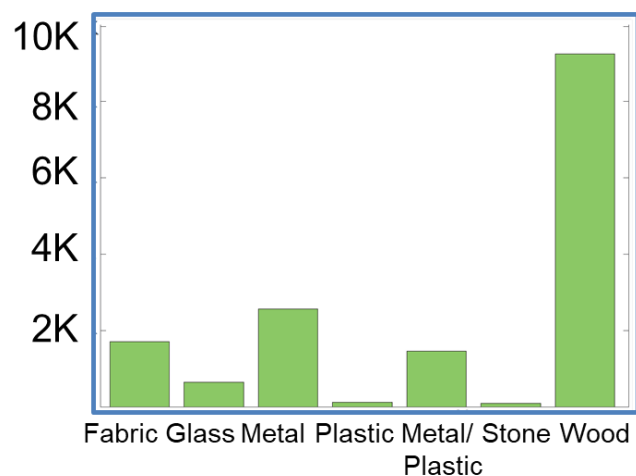
WOOD or PLASTIC  
METAL

Manufacturer  
Product Images

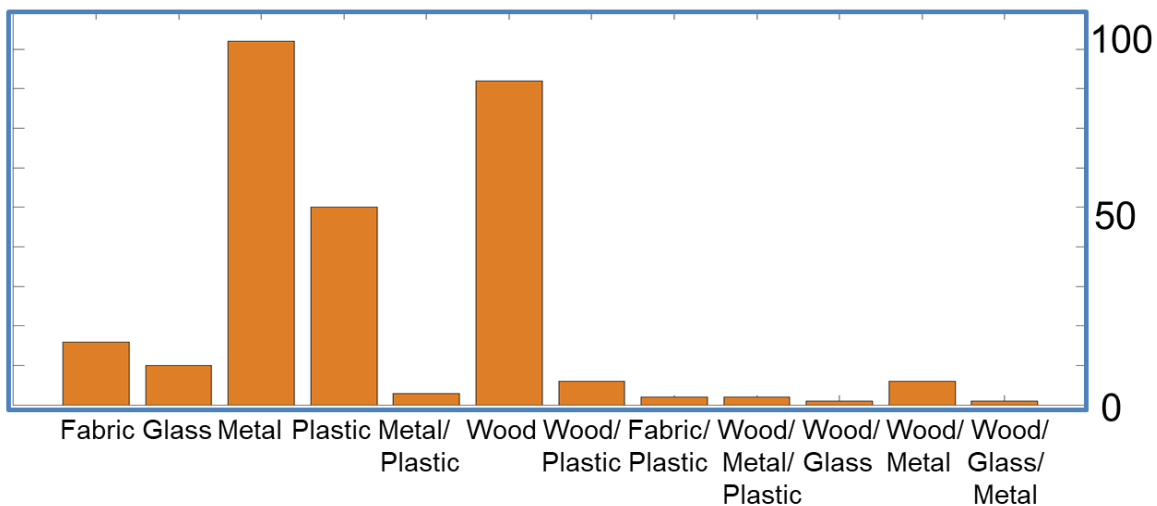


# Label Distribution (# Parts / Label)

(Left) Crowdsourced Dataset



(Right) Expert Labeled Dataset



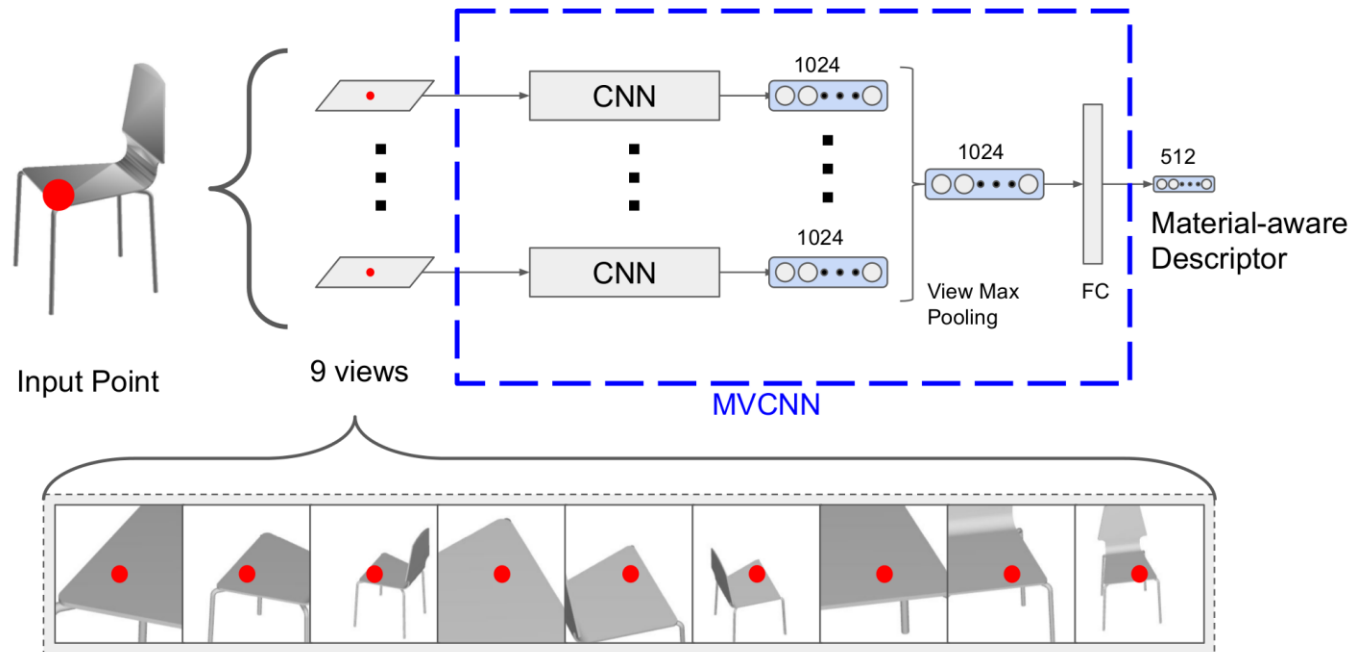
# Outline

1. Goal
2. Motivation
3. Related Work
4. Data Collection
- 5. Network Architecture and Training Pipeline**
6. Post-Processing
7. Results
8. Future Directions

# Challenge: Learning Pipeline

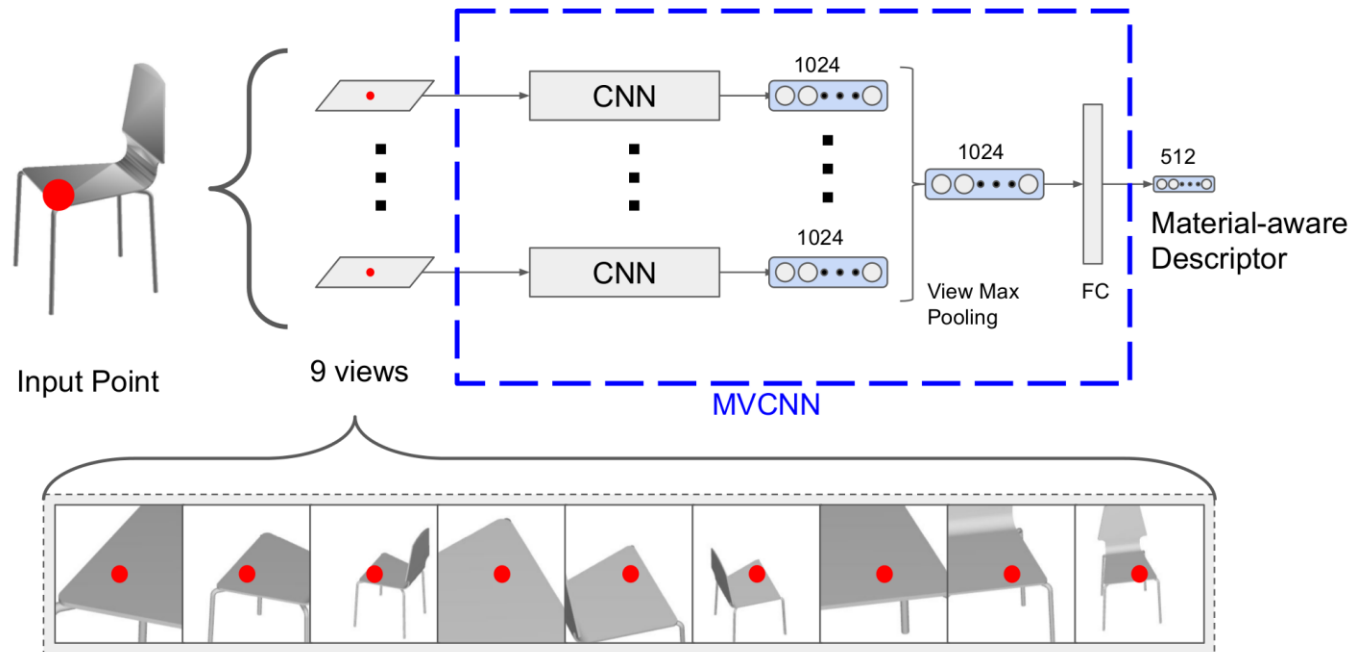
# Architecture

Based on MVCNN architecture [Huang et al. 2018]



# Architecture

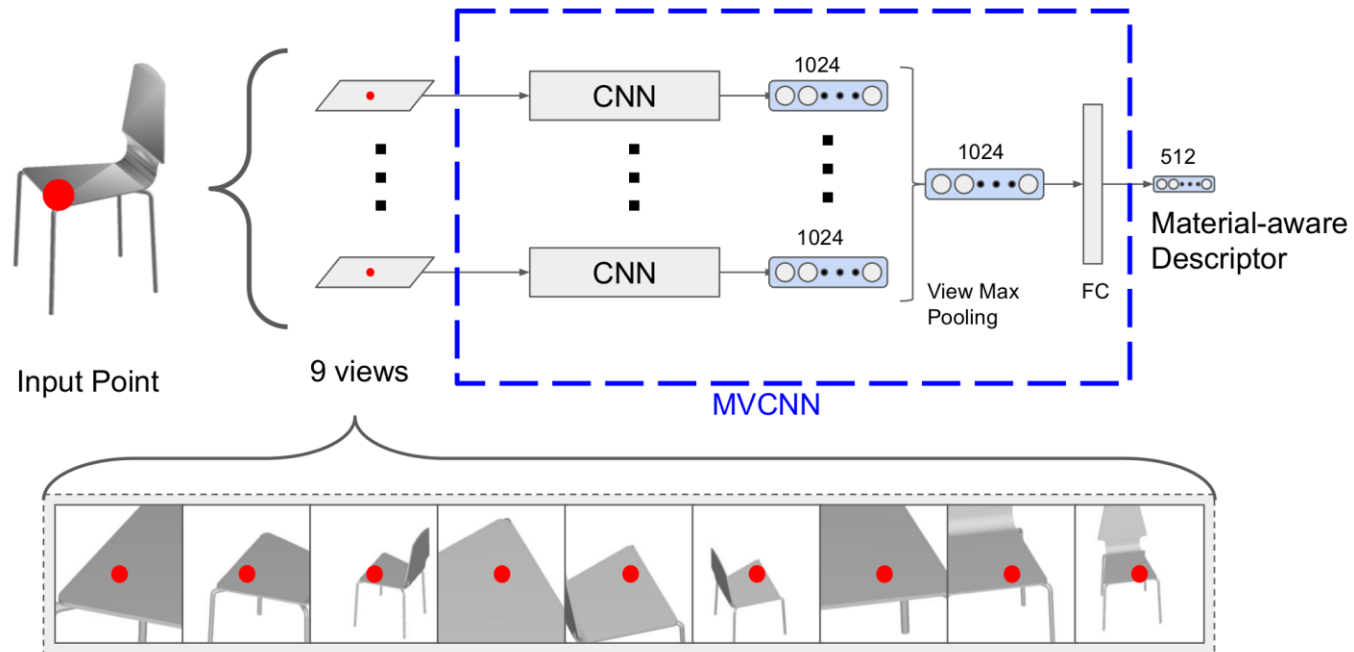
- CNN backbone is Googlenet (VGG etc also works)





# Architecture

- Input is 9 rendered views around surface point



# Architecture

- Input is 9 rendered views around surface point
- Views are selected to maximize surface coverage
- 3 viewing directions at 3 viewing distances
  
- Camera is oriented upright wrt shape
  
- Also tried 36 views

# Training

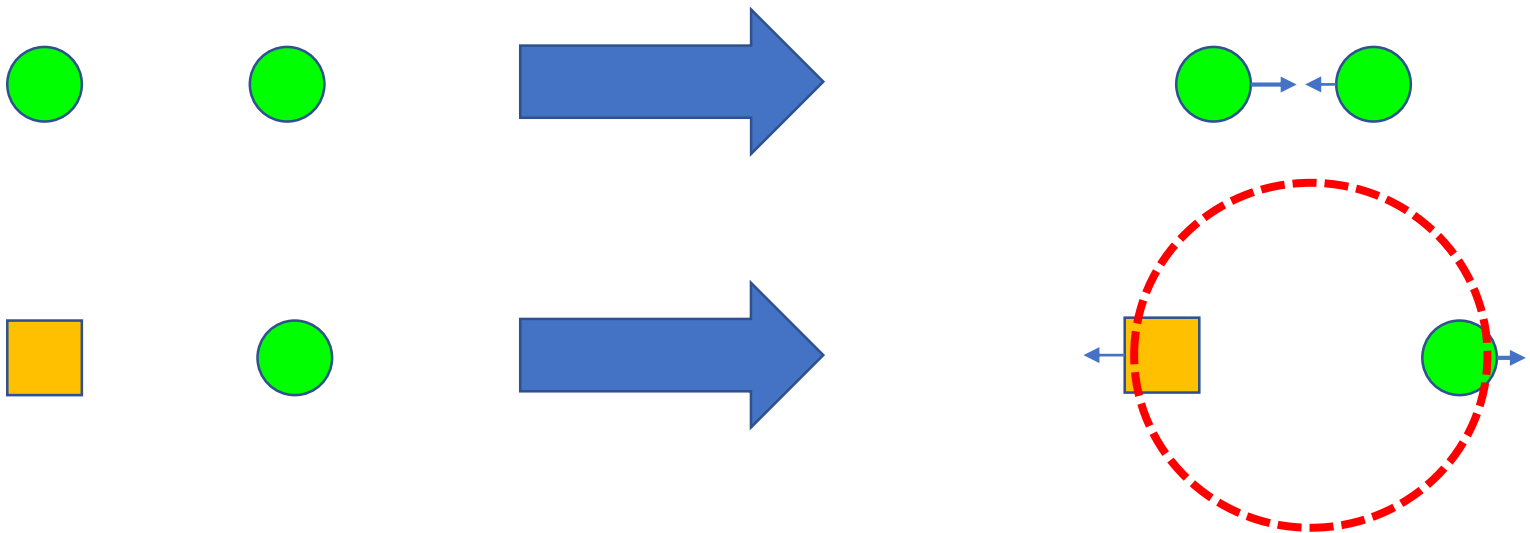
Loss function:

- 1) Contrastive loss [Hadsell et al. 2006] + classification loss
- 2) Classification loss only

# Training

Loss function:

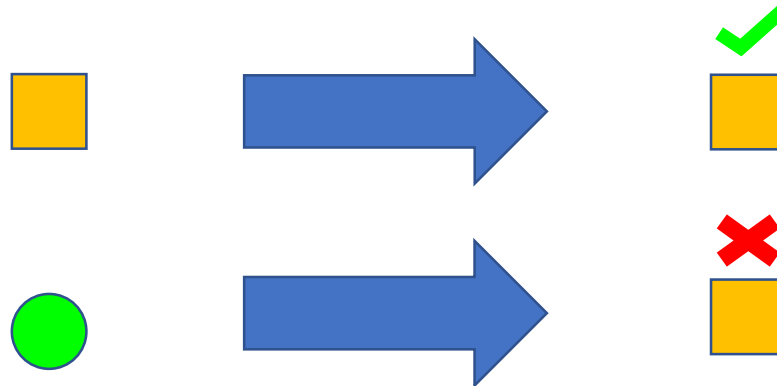
- 1) Contrastive loss [Hadsell et al. 2006] + classification loss
- 2) Classification loss only



# Training

Loss function:

- 1) Contrastive loss [Hadsell et al. 2006] + classification loss
- 2) Classification loss only



# Training

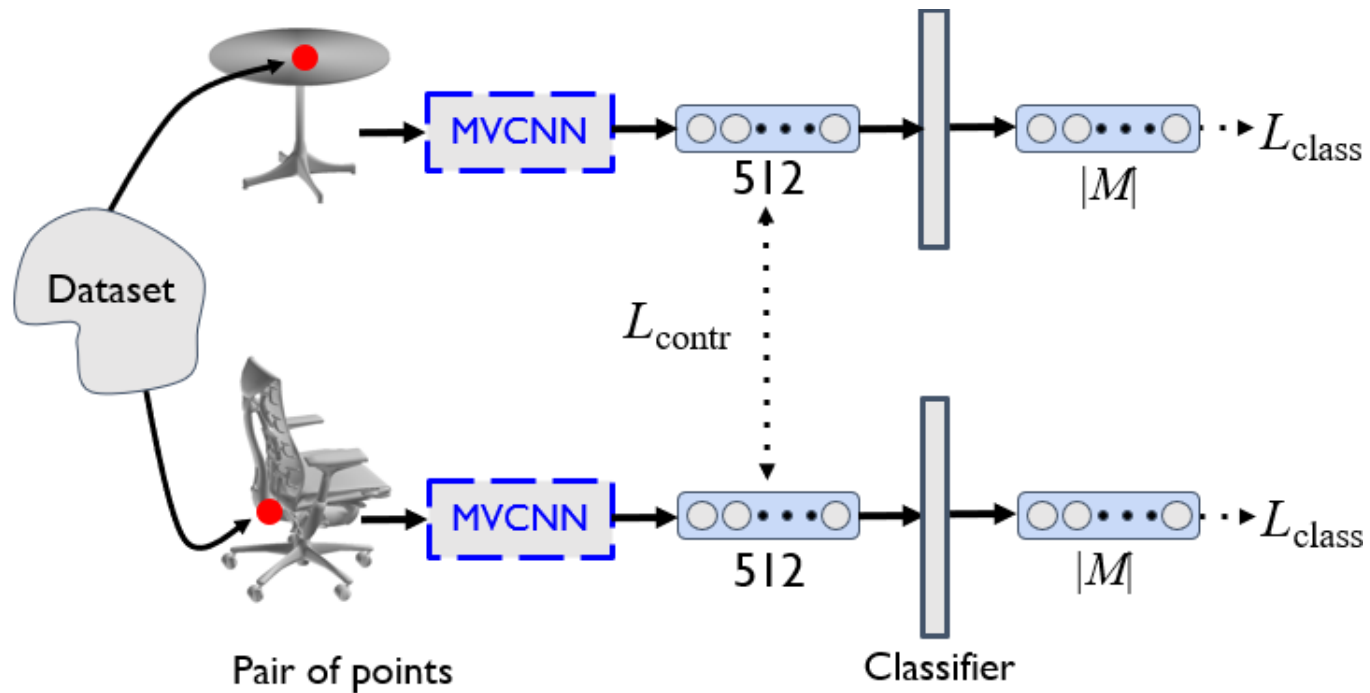
Loss function:

- 1) Contrastive loss [Hadsell et al. 2006] + classification loss
- 2) Classification loss only

These two variants produced the best results.

# Training

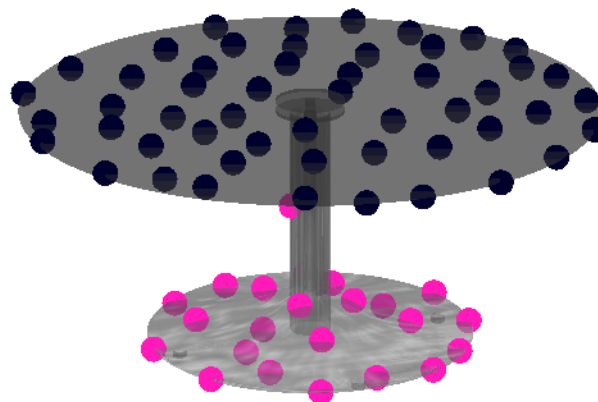
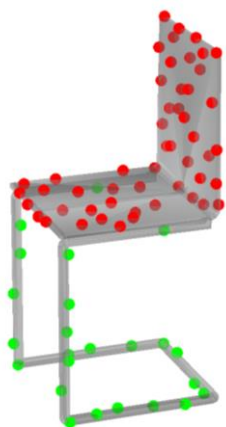
Trained in Siamese fashion



# Training

Training set is sampled from crowdsourced data (>50% parts labeled)

- 75 uniformly separate points are sampled from each shape (occluded points ignored)
- Final training set consists of ~150K points.





# Training

- Dataset is biased / imbalanced
- Class-balanced training – explicitly cycle through each combination of label pairs when sampling

e.g. (wood, wood)

(wood, metal)

(wood, fabric)

...

# Training

- Dataset is biased / imbalanced
- Class-balanced training – explicitly cycle through each combination of label pairs when sampling

e.g. ~~(wood, wood)~~

(wood, metal)

(wood, fabric)

...

# Training

- Dataset is biased / imbalanced
- Class-balanced training – explicitly cycle through each combination of label pairs when sampling

e.g. ~~(wood, wood)~~

~~(wood, metal)~~

(wood, fabric)

...

# Training

- Dataset is biased / imbalanced
- Class-balanced training – explicitly cycle through each combination of label pairs when sampling

e.g. ~~(wood, wood)~~

~~(wood, metal)~~

~~(wood, fabric)~~

...

# Training

- Dataset is biased / imbalanced
- Class-balanced training – explicitly cycle through each combination of label pairs when sampling

e.g. ~~(wood, wood)~~

~~(wood, metal)~~

~~(wood, fabric)~~

...

- Sample **same class pairs 20%** of time, sample **different class pairs 80%** of time

# Outline

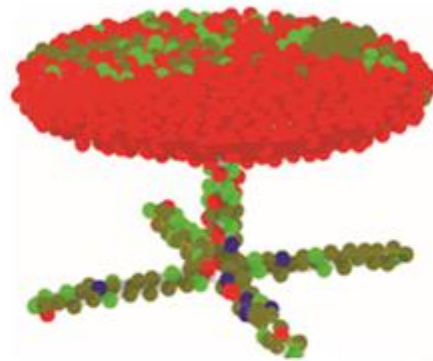
1. Goal
2. Motivation
3. Related Work
4. Data Collection
5. Network Architecture and Training Pipeline
- 6. Post-Processing**
7. Results
8. Future Directions

# Challenge: Global Reasoning

# Local Material Predictions



Ground-truth  
materials



Local Predictions

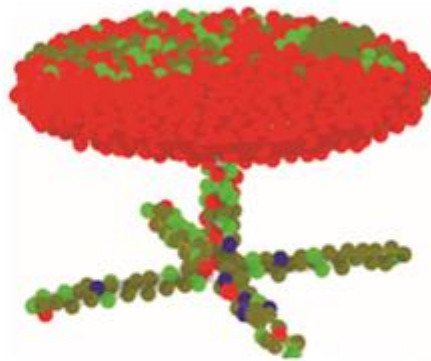
● Wood ● Metal ● Metal OR Plastic



# CRF



Ground-truth  
materials



Local Predictions



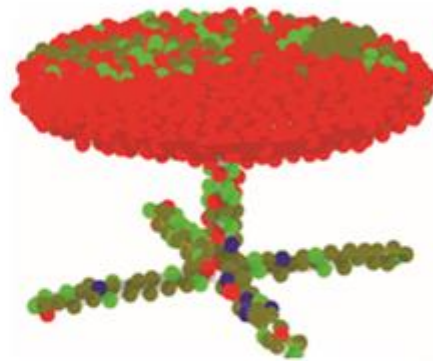
CRF  
(no symmetry)

● Wood ● Metal ● Metal OR Plastic

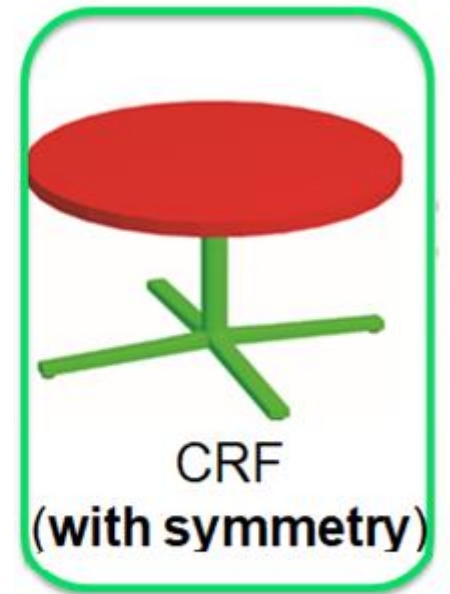
# CRF with symmetry



Ground-truth  
materials



Local Predictions



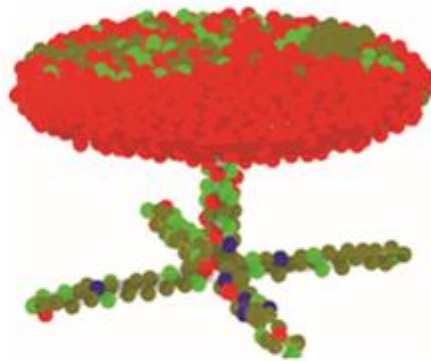
CRF  
(with symmetry)

● Wood ● Metal ● Metal OR Plastic

# Comparison



Ground-truth  
materials



Local Predictions



CRF  
(no symmetry)



CRF  
(with symmetry)

● Wood ● Metal ● Metal OR Plastic

# CRF

- Use CRF to smooth local material predictions
- Three pairwise factors between polygons:
  - Low dihedral angle → same material
  - Low geodesic distance → same material
  - Rotational / reflective symmetry → same material

# CRF

- Use CRF to smooth local material predictions
- Three pairwise factors between polygons:
  - Low dihedral angle → same material
  - Low geodesic distance → same material
  - Rotational / reflective symmetry → same material

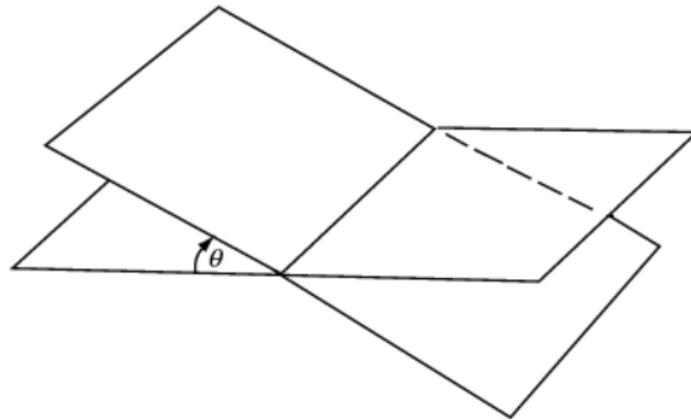


Fig from <http://mathworld.wolfram.com/DihedralAngle.html>

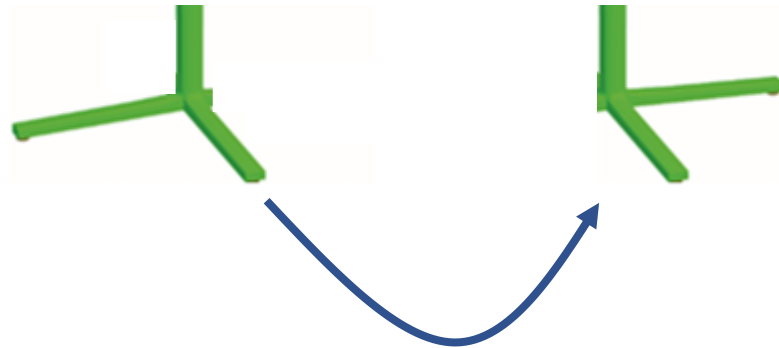
# CRF

- Use CRF to smooth local material predictions
- Three pairwise factors between polygons:
  - Low dihedral angle → same material
  - Low geodesic distance → same material
  - Rotational / reflective symmetry → same material



# CRF

- Use CRF to smooth local material predictions
- Three pairwise factors between polygons:
  - Low dihedral angle  $\rightarrow$  same material
  - Low geodesic distance  $\rightarrow$  same material
  - Rotational / reflective symmetry  $\rightarrow$  same material



# Outline

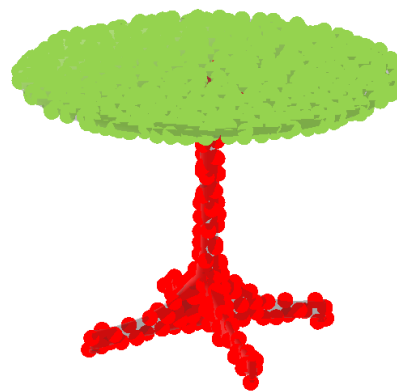
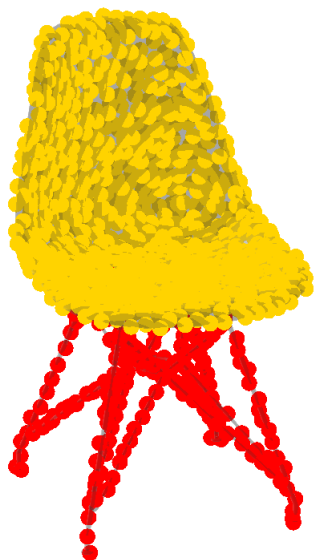
1. Goal
2. Motivation
3. Related Work
4. Data Collection
5. Network Architecture and Training Pipeline
6. Post-Processing
- 7. Results**
8. Future Directions



# Test Set

1024 uniformly separated points sampled from each benchmark shape:

- Occluded points are discarded
- Final test set consists of 117K points



# Material Prediction

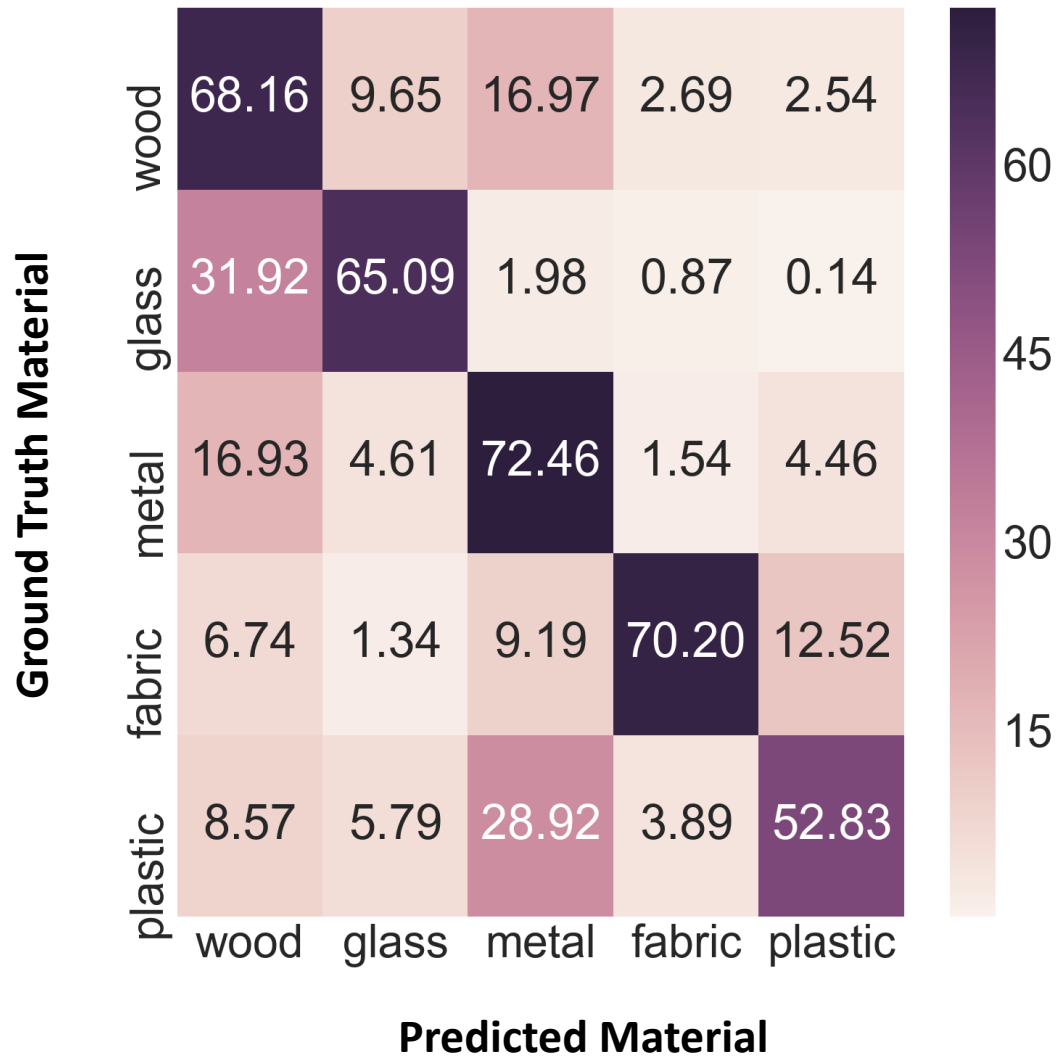
## Mean Class (Top 1) Accuracy

- Multitask has more balanced predictions and **highest mean accuracy**
- +CRF boosts performance across all categories except glass

| Network             | Mean | Wood | Glass | Metal | Fabric | Plastic |
|---------------------|------|------|-------|-------|--------|---------|
| Classification      | 65   | 82   | 53    | 72    | 62     | 55      |
| Classification +CRF | 66   | 85   | 36    | 77    | 66     | 65      |
| Multitask           | 66   | 68   | 65    | 72    | 70     | 53      |
| Multitask +CRF      | 71   | 75   | 64    | 74    | 74     | 68      |

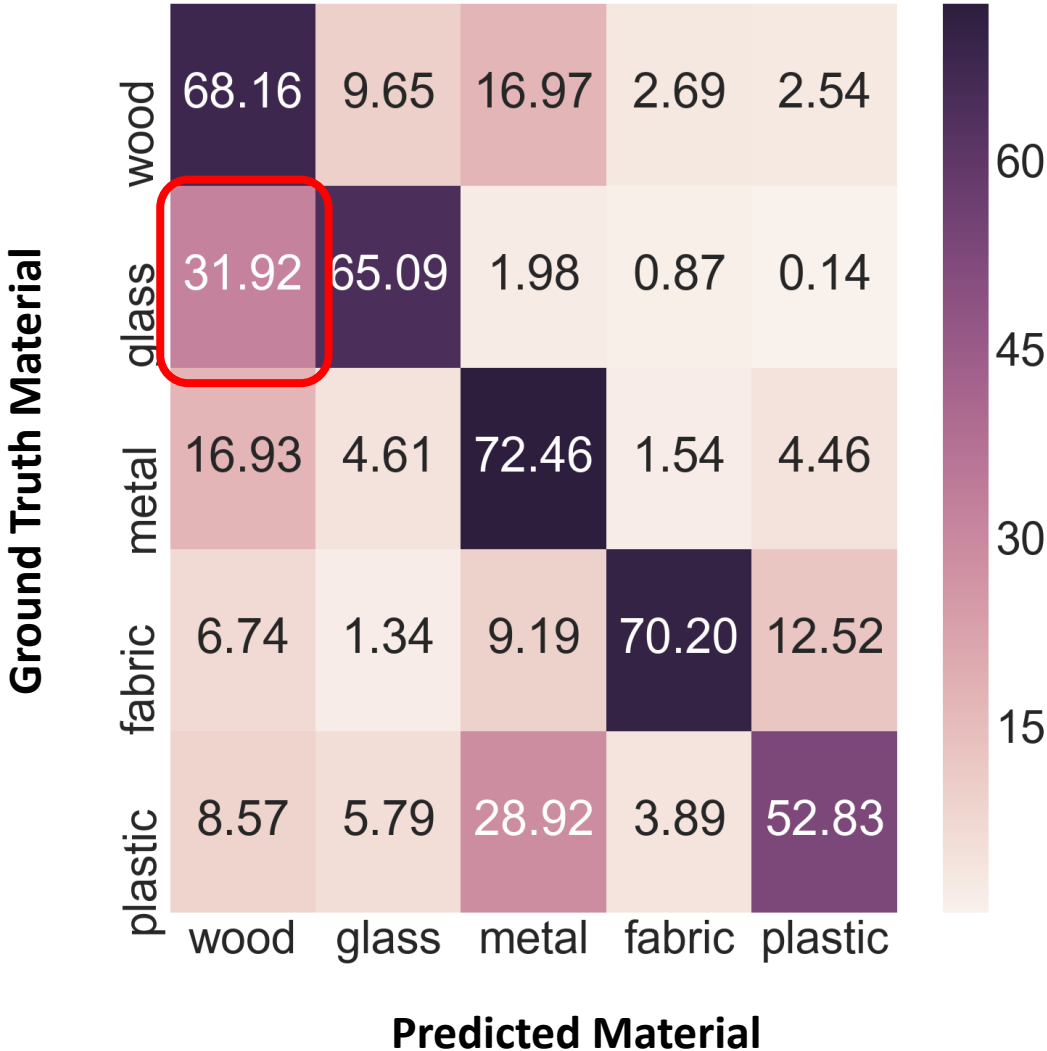
# Material Prediction

Multitask (No CRF)



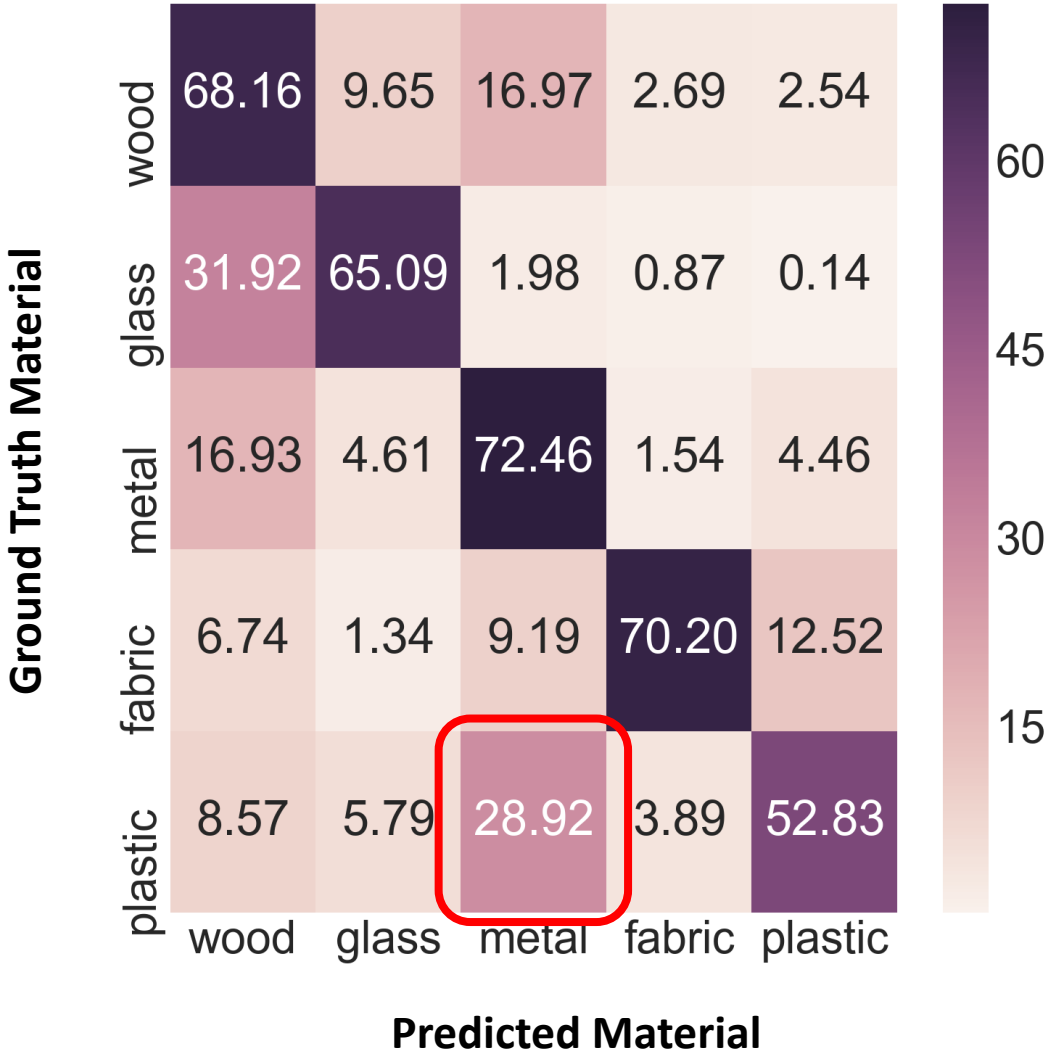
# Material Prediction

Multitask (No CRF)



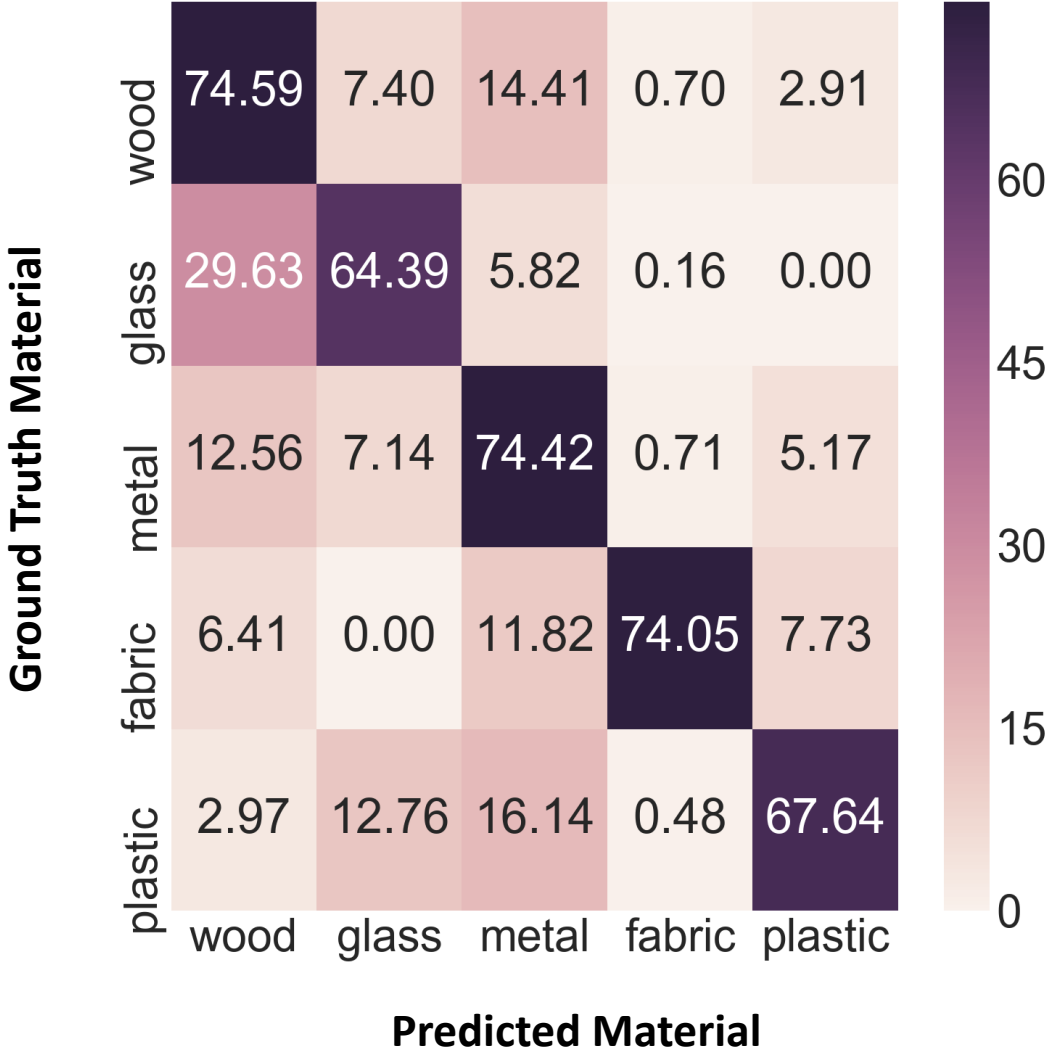
# Material Prediction

Multitask (No CRF)



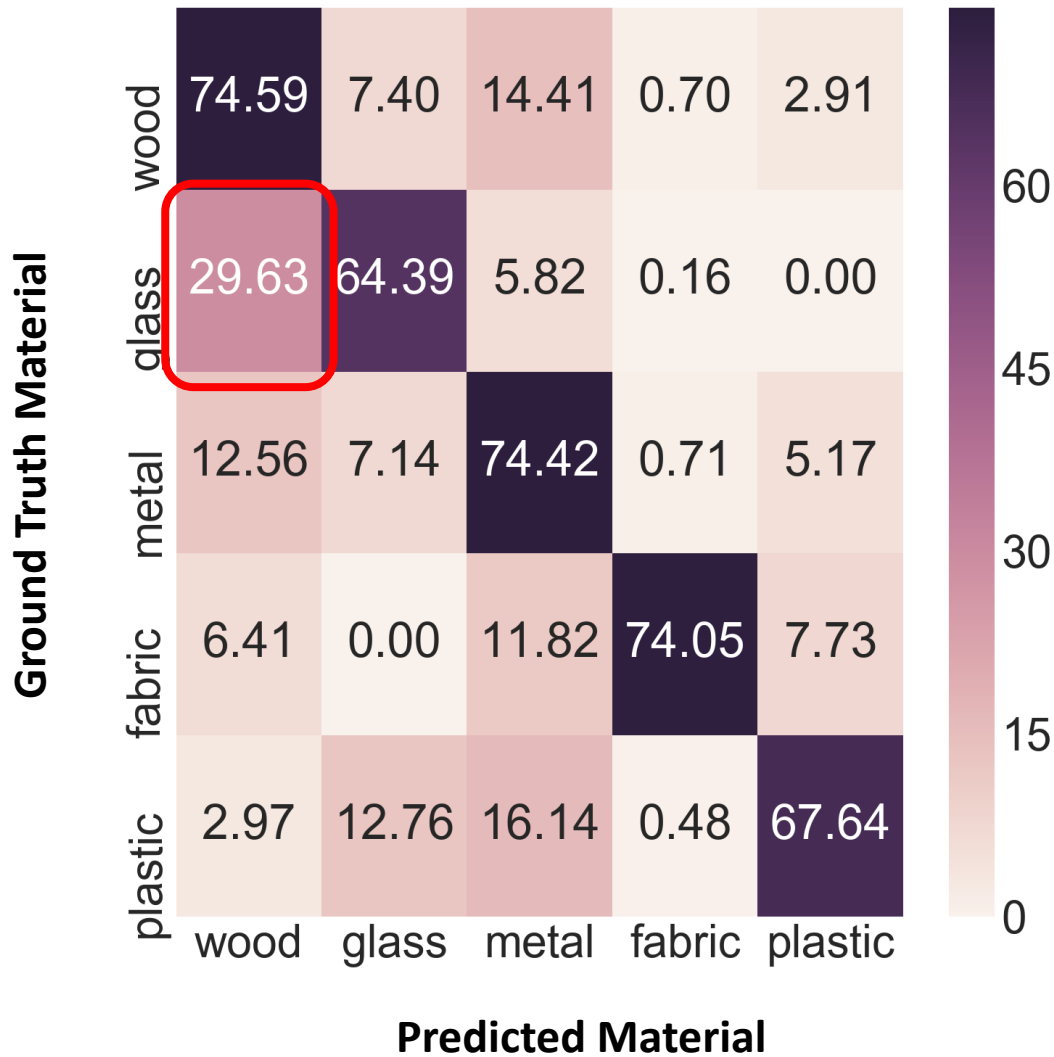
# Material Prediction

Multitask+CRF:



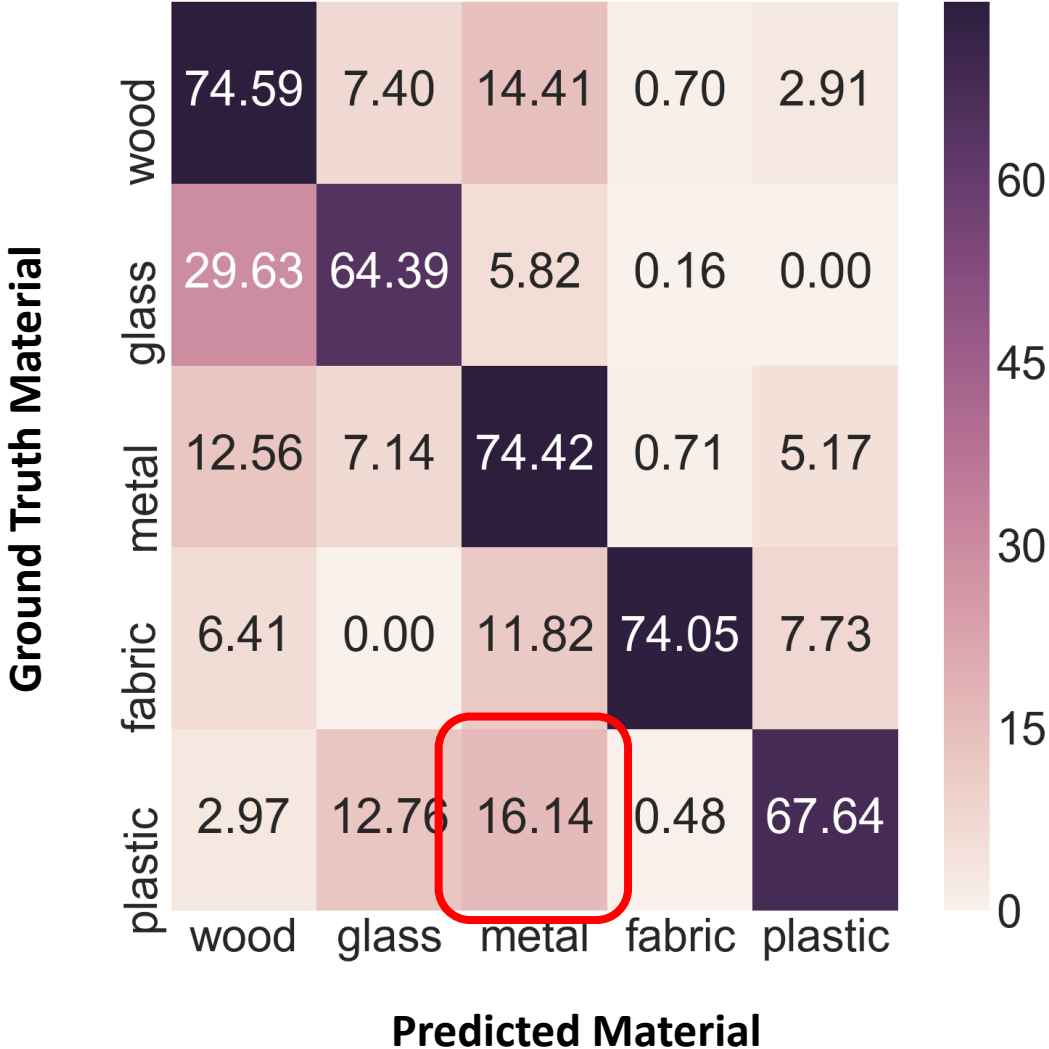
# Material Prediction

Multitask+CRF:



# Material Prediction

Multitask+CRF:





# Descriptor Retrieval

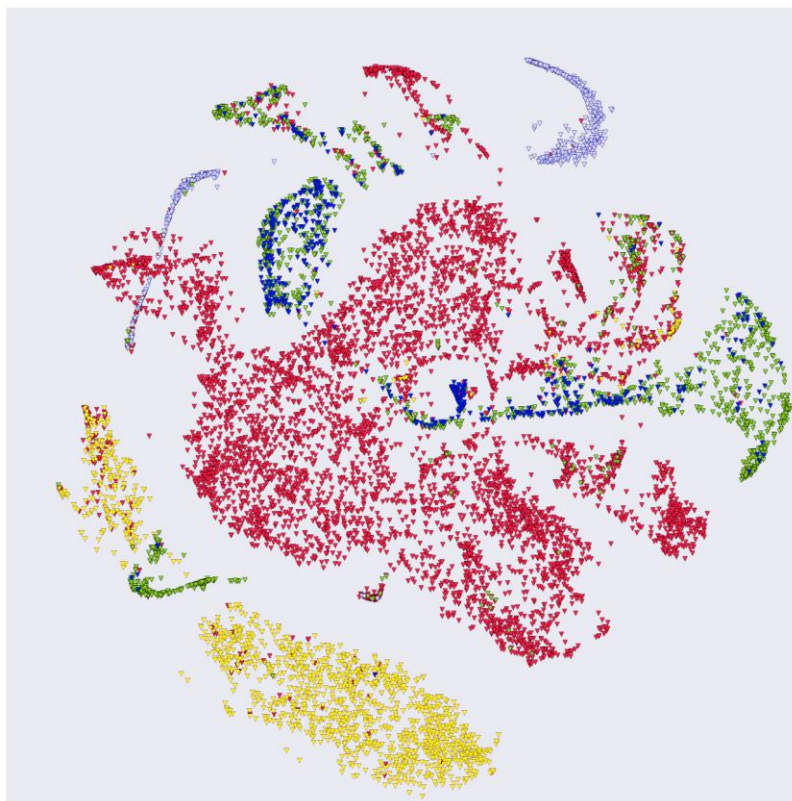
## Mean Class Precision

- Similar mean class performance
- Multitask outperforms Classification for all materials except wood

| Network        | Mean | Wood | Glass | Metal | Fabric | Plastic |
|----------------|------|------|-------|-------|--------|---------|
| Classification |      |      |       |       |        |         |
| k=1            | 55.7 | 76.4 | 34.3  | 65.0  | 56.1   | 46.7    |
| k=30           | 56.9 | 75.3 | 41.1  | 64.9  | 55.3   | 47.6    |
| k=100          | 57.3 | 75.1 | 43.0  | 64.9  | 55.5   | 48.0    |
| Multitask      |      |      |       |       |        |         |
| k=1            | 56.2 | 62.2 | 40.8  | 68.6  | 58.0   | 51.2    |
| k=30           | 56.2 | 61.0 | 42.6  | 68.9  | 57.4   | 51.1    |
| k=100          | 56.6 | 60.7 | 44.7  | 68.7  | 57.4   | 51.5    |

# Embedding Visualization (tSNE)

Multitask Descriptor Space



- ▼ Wood
- ▼ Fabric
- ▼ Glass
- ▼ Metal
- ▼ Plastic

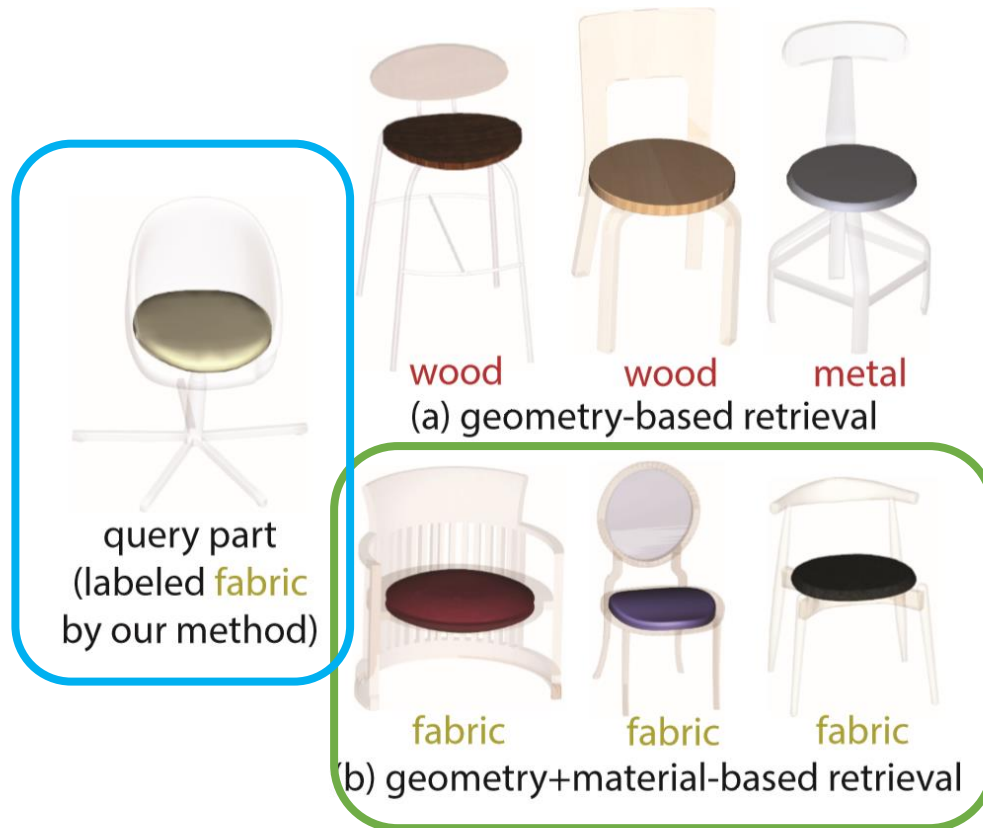
# Effect of # of Input Views

3 views (1 direction, 3 distances) vs 9 views (3, 3)

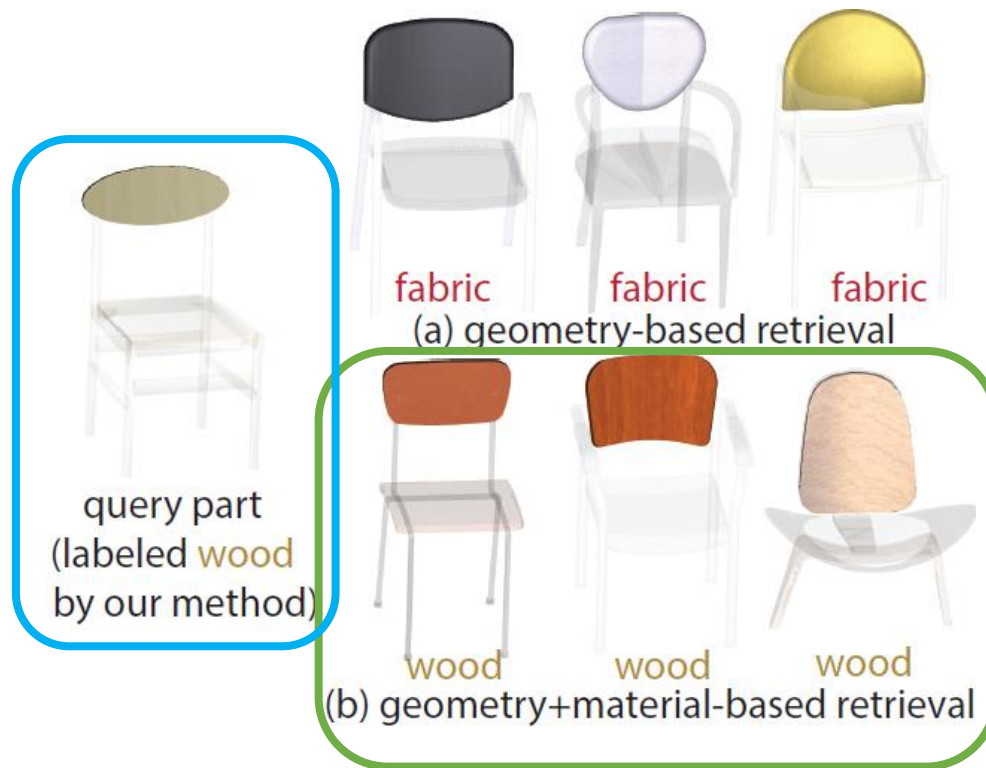
- Multiple view directions are advantageous
- Top 1 classification accuracy:

| Network                   | Mean | Wood | Glass | Metal | Fabric | Plastic |
|---------------------------|------|------|-------|-------|--------|---------|
| Classification<br>3 views | 59   | 81   | 41    | 71    | 60     | 40      |
| Classification<br>9 views | 65   | 82   | 53    | 72    | 62     | 55      |
| Multitask<br>3 views      | 56   | 45   | 71    | 85    | 65     | 15      |
| Multitask<br>9 views      | 66   | 68   | 65    | 72    | 70     | 53      |

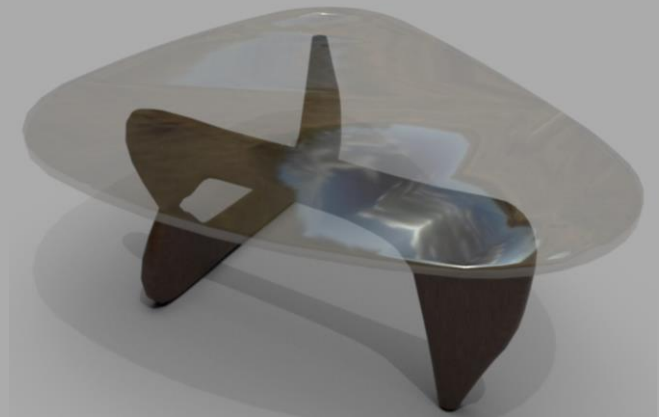
# Material-Aware Part Retrieval



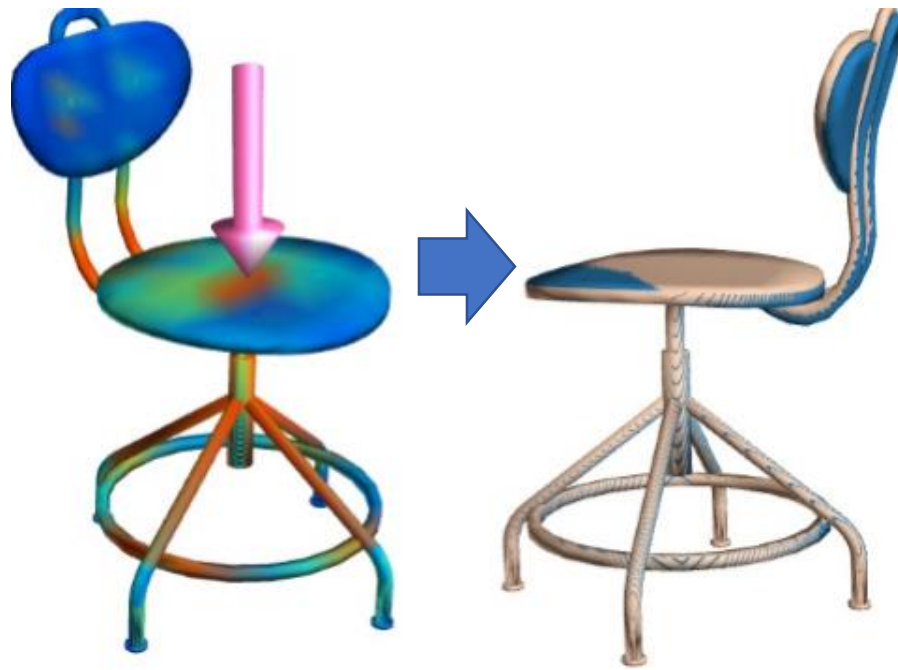
# Material-Aware Part Retrieval



# Material-Aware Automatic Texturing



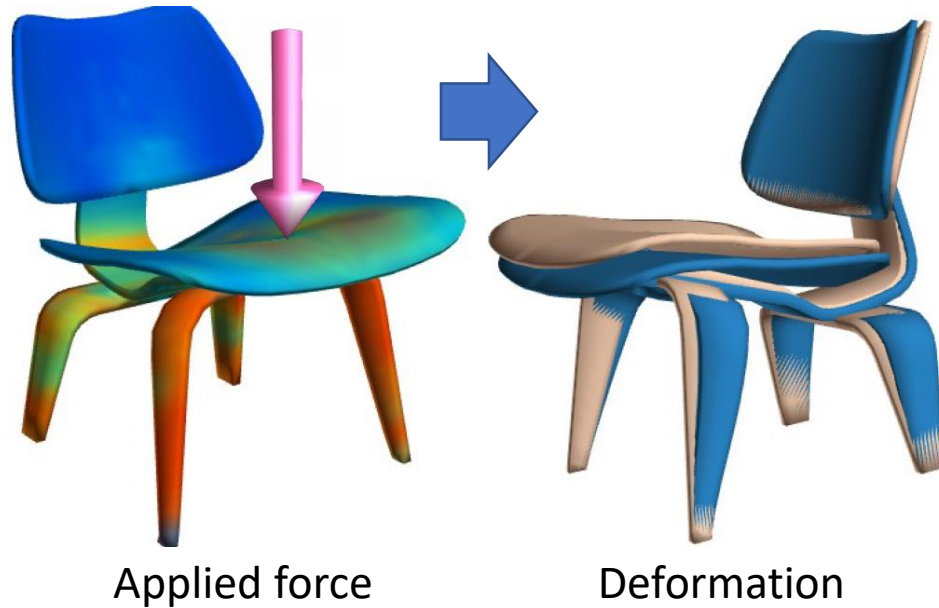
# Material-Aware Physics Simulation



Applied force

Deformation

# Material-Aware Physics Simulation





# Conclusion

- Two shape datasets with per-part material labels through crowdsourcing and expert-labelling
- Material-aware local descriptors computed through supervised learning pipeline
- Symmetry-aware CRF for global reasoning

# Future Directions

- Increase variety of shapes and materials
- Learn smooth predictions end-to-end without CRF
- Fine-grained materials
- 2D material classification has good performance. Leverage this to improve 3D understanding.

Thank you!