Learning Material-Aware Local Descriptors for 3D Shapes

Hubert Lin¹ Melinos Averkiou² Evangelos Kalogerakis³ Balazs Kovacs⁴ Siddhant Ranade⁵ Vladimir G. Kim⁶ Siddhartha Chaudhuri⁶,⁷ Kavita Bala¹

¹Cornell Univ. ²Univ. of Cyprus ³UMass Amherst ⁴Zoox ⁵Univ. of Utah ⁶Adobe ⁷IIT 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

Fabric

Wood
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

[https://www.shapenet.org]
Shape Databases

Semantically-Enriched 3D Models for Common-sense Knowledge

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

[Savva et al 2015]
Shape Databases

Text2Shape

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

![3D shapes and natural language descriptions](image)

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

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

[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

[Qi et al 2017]
Deep Learning for Shape Analysis

- Shape completion

[Han et al 2017]
Deep Learning for Shape Analysis

• Geometric descriptors

[Huang et al 2018]
Material Understanding for Shapes

Material Memex

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

[Jain et al 2012]
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

[Wang et al 2016]
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

[Chen et al 2015]
Material Understanding for Images

Flickr Material Database
• Surfaces of common materials; manually curated
• Relatively small dataset (100 per category)

[Sharan et al 2014]
Material Understanding for Images

Describable Textures Dataset

• Textures described by attributes ("striped", ...)  
• Dataset of representative textures

[Cimpoi et al 2014]
Material Understanding for Images

OpenSurfaces

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

[Bell et al 2013]
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]

[Bell et al 2015]
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)

No texture
Crowdsourced Data

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

- Remaining shapes (8K)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No texture, too many/too few components</td>
</tr>
</tbody>
</table>
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...

• 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.
Expert-Annotated Data

Expert annotators reference product images and descriptions for accurate labelling.
Label Distribution (# Parts / Label)

(Left) Crowdsourced Dataset    (Right) Expert Labeled Dataset

![Bar chart for Crowdsourced Dataset](image1)

![Bar chart for Expert Labeled Dataset](image2)
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. \( (\text{wood}, \text{wood}) \)
  \( (\text{wood}, \text{metal}) \)
  \( (\text{wood}, \text{fabric}) \)
  ...

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
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 \( \Rightarrow \) same material
  • Low geodesic distance \( \Rightarrow \) same material
  • Rotational / reflective symmetry \( \Rightarrow \) 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

<table>
<thead>
<tr>
<th>Network</th>
<th>Mean</th>
<th>Wood</th>
<th>Glass</th>
<th>Metal</th>
<th>Fabric</th>
<th>Plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>65</td>
<td>82</td>
<td>53</td>
<td>72</td>
<td>62</td>
<td>55</td>
</tr>
<tr>
<td>Classification +CRF</td>
<td>66</td>
<td>85</td>
<td>36</td>
<td>77</td>
<td>66</td>
<td>65</td>
</tr>
<tr>
<td>Multitask</td>
<td>66</td>
<td>68</td>
<td>65</td>
<td>72</td>
<td>70</td>
<td>53</td>
</tr>
<tr>
<td>Multitask +CRF</td>
<td><strong>71</strong></td>
<td>75</td>
<td>64</td>
<td>74</td>
<td>74</td>
<td>68</td>
</tr>
</tbody>
</table>
Material Prediction

Multitask (No CRF)

<table>
<thead>
<tr>
<th>Ground Truth Material</th>
<th>wood</th>
<th>glass</th>
<th>metal</th>
<th>fabric</th>
<th>plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wood</td>
<td>68.16</td>
<td>9.65</td>
<td>16.97</td>
<td>2.69</td>
<td>2.54</td>
</tr>
<tr>
<td>glass</td>
<td>31.92</td>
<td>65.09</td>
<td>1.98</td>
<td>0.87</td>
<td>0.14</td>
</tr>
<tr>
<td>metal</td>
<td>16.93</td>
<td>4.61</td>
<td>72.46</td>
<td>1.54</td>
<td>4.46</td>
</tr>
<tr>
<td>fabric</td>
<td>6.74</td>
<td>1.34</td>
<td>9.19</td>
<td>70.20</td>
<td>12.52</td>
</tr>
<tr>
<td>plastic</td>
<td>8.57</td>
<td>5.79</td>
<td>28.92</td>
<td>3.89</td>
<td>52.83</td>
</tr>
</tbody>
</table>
Material Prediction

Multitask (No CRF)

<table>
<thead>
<tr>
<th>Ground Truth Material</th>
<th>Predicted Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>wood</td>
<td>68.16 9.65 16.97 2.69 2.54</td>
</tr>
<tr>
<td>glass</td>
<td>31.92 65.09 1.98 0.87 0.14</td>
</tr>
<tr>
<td>metal</td>
<td>16.93 4.61 72.46 1.54 4.46</td>
</tr>
<tr>
<td>fabric</td>
<td>6.74 1.34 9.19 70.20 12.52</td>
</tr>
<tr>
<td>plastic</td>
<td>8.57 5.79 28.92 3.89 52.83</td>
</tr>
</tbody>
</table>

Predicted Material
Material Prediction

Multitask (No CRF)

<table>
<thead>
<tr>
<th>Ground Truth Material</th>
<th>wood</th>
<th>glass</th>
<th>metal</th>
<th>fabric</th>
<th>plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wood</td>
<td>68.16</td>
<td>9.65</td>
<td>16.97</td>
<td>2.69</td>
<td>2.54</td>
</tr>
<tr>
<td>glass</td>
<td>31.92</td>
<td>65.09</td>
<td>1.98</td>
<td>0.87</td>
<td>0.14</td>
</tr>
<tr>
<td>metal</td>
<td>16.93</td>
<td>4.61</td>
<td>72.46</td>
<td>1.54</td>
<td>4.46</td>
</tr>
<tr>
<td>fabric</td>
<td>6.74</td>
<td>1.34</td>
<td>9.19</td>
<td>70.20</td>
<td>12.52</td>
</tr>
<tr>
<td>plastic</td>
<td>8.57</td>
<td>5.79</td>
<td>28.92</td>
<td>3.89</td>
<td>52.83</td>
</tr>
</tbody>
</table>
Material Prediction

Multitask+CRF:

<table>
<thead>
<tr>
<th></th>
<th>wood</th>
<th>glass</th>
<th>metal</th>
<th>fabric</th>
<th>plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wood</td>
<td>74.59</td>
<td>7.40</td>
<td>14.41</td>
<td>0.70</td>
<td>2.91</td>
</tr>
<tr>
<td>glass</td>
<td>29.63</td>
<td>64.39</td>
<td>5.82</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>metal</td>
<td>12.56</td>
<td>7.14</td>
<td>74.42</td>
<td>0.71</td>
<td>5.17</td>
</tr>
<tr>
<td>fabric</td>
<td>6.41</td>
<td>0.00</td>
<td>11.82</td>
<td>74.05</td>
<td>7.73</td>
</tr>
<tr>
<td>plastic</td>
<td>2.97</td>
<td>12.76</td>
<td>16.14</td>
<td>0.48</td>
<td>67.64</td>
</tr>
</tbody>
</table>
Material Prediction

Multitask+CRF:

<table>
<thead>
<tr>
<th>Ground Truth Material</th>
<th>wood</th>
<th>glass</th>
<th>metal</th>
<th>fabric</th>
<th>plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wood</td>
<td>74.59</td>
<td>7.40</td>
<td>14.41</td>
<td>0.70</td>
<td>2.91</td>
</tr>
<tr>
<td>glass</td>
<td>29.63</td>
<td>64.39</td>
<td>5.82</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>metal</td>
<td>12.56</td>
<td>7.14</td>
<td>74.42</td>
<td>0.71</td>
<td>5.17</td>
</tr>
<tr>
<td>fabric</td>
<td>6.41</td>
<td>0.00</td>
<td>11.82</td>
<td>74.05</td>
<td>7.73</td>
</tr>
<tr>
<td>plastic</td>
<td>2.97</td>
<td>12.76</td>
<td>16.14</td>
<td>0.48</td>
<td>67.64</td>
</tr>
</tbody>
</table>

Predicted Material
### Material Prediction

**Multitask+CRF:**

<table>
<thead>
<tr>
<th>Ground Truth Material</th>
<th>wood</th>
<th>glass</th>
<th>metal</th>
<th>fabric</th>
<th>plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wood</td>
<td>74.59</td>
<td>7.40</td>
<td>14.41</td>
<td>0.70</td>
<td>2.91</td>
</tr>
<tr>
<td>glass</td>
<td>29.63</td>
<td>64.39</td>
<td>5.82</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>metal</td>
<td>12.56</td>
<td>7.14</td>
<td>74.42</td>
<td>0.71</td>
<td>5.17</td>
</tr>
<tr>
<td>fabric</td>
<td>6.41</td>
<td>0.00</td>
<td>11.82</td>
<td>74.05</td>
<td>7.73</td>
</tr>
<tr>
<td>plastic</td>
<td>2.97</td>
<td>12.76</td>
<td>16.14</td>
<td>0.48</td>
<td>67.64</td>
</tr>
</tbody>
</table>

The highlighted cell indicates a specific prediction comparison.
Descriptor Retrieval
Mean Class Precision

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

<table>
<thead>
<tr>
<th>Network</th>
<th>Mean</th>
<th>Wood</th>
<th>Glass</th>
<th>Metal</th>
<th>Fabric</th>
<th>Plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k=1</td>
<td>55.7</td>
<td>76.4</td>
<td>34.3</td>
<td>65.0</td>
<td>56.1</td>
<td>46.7</td>
</tr>
<tr>
<td>k=30</td>
<td>56.9</td>
<td>75.3</td>
<td>41.1</td>
<td>64.9</td>
<td>55.3</td>
<td>47.6</td>
</tr>
<tr>
<td>k=100</td>
<td>57.3</td>
<td>75.1</td>
<td>43.0</td>
<td>64.9</td>
<td>55.5</td>
<td>48.0</td>
</tr>
<tr>
<td>Multitask</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k=1</td>
<td>56.2</td>
<td>62.2</td>
<td>40.8</td>
<td>68.6</td>
<td>58.0</td>
<td>51.2</td>
</tr>
<tr>
<td>k=30</td>
<td>56.2</td>
<td>61.0</td>
<td>42.6</td>
<td>68.9</td>
<td>57.4</td>
<td>51.1</td>
</tr>
<tr>
<td>k=100</td>
<td>56.6</td>
<td>60.7</td>
<td>44.7</td>
<td>68.7</td>
<td>57.4</td>
<td>51.5</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Network</th>
<th>Mean</th>
<th>Wood</th>
<th>Glass</th>
<th>Metal</th>
<th>Fabric</th>
<th>Plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification 3 views</td>
<td>59</td>
<td>81</td>
<td>41</td>
<td>71</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Classification 9 views</td>
<td>65</td>
<td>82</td>
<td>53</td>
<td>72</td>
<td>62</td>
<td>55</td>
</tr>
<tr>
<td>Multitask 3 views</td>
<td>56</td>
<td>45</td>
<td>71</td>
<td>85</td>
<td>65</td>
<td>15</td>
</tr>
<tr>
<td>Multitask 9 views</td>
<td>66</td>
<td>68</td>
<td>65</td>
<td>72</td>
<td>70</td>
<td>53</td>
</tr>
</tbody>
</table>
Material-Aware Part Retrieval

- wood (a) geometry-based retrieval
- metal
- fabric (b) geometry+material-based retrieval
- query part (labeled fabric by our method)
Material-Aware Part Retrieval

- (a) geometry-based retrieval
- (b) geometry + material-based retrieval

query part (labeled wood by our method)
Material-Aware Automatic Texturing
Material-Aware Physics Simulation

Applied force  Deformation
Material-Aware Physics Simulation

Applied force → Deformation
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!