Deep Learning for Geometry Processing
3D Representations

View-Based and Volumetric CNNs
3D Representations for Object Classification

Volumetric

3DShapeNets by Z. Wu et al. CVPR 15

VoxNet by D. Maturana et al. IEEE/RSJ 15

Multi-Views

MVCNN by H. Su et al. ICCV 15

DeepPano by B. Shi et al. IEEE/SPL 15
Multi-View CNNs

- CNN1 extracts image features (parameters are shared across views)
- CNN1 extracts image features (parameters are shared across views)
- Element-wise max pooling across all views
Multi-View CNNs

- CNN1 extracts image features (parameters are shared across views)
- Element-wise max pooling across all views
- CNN2 produces shape descriptors + final prediction
Multi-View CNNs

- CNN1 extracts image features (parameters are shared across views)
- Element-wise max pooling across all views
- CNN2 produces shape descriptors + final prediction
Multi-View CNNs

3D shape classification and retrieval
- Pre-trained on ImageNet
- Fine-tuned on 2D views

ModelNet40

classify “chair”
Volumetric representation: shapes as binary voxels in a 3D grid
Volumetric representation: shapes as binary voxels in a 3D grid
Learn filters operating on these volumetric data
**Volumetric CNNs**

- **Volumetric** representation: shapes as binary voxels in a 3D grid
- Learn filters operating on these volumetric data
- Standard convolution in $\mathbb{R}^3$
Learned Features: 3D Primitives / Filter Visualization

Wu et al. 2015
Learned Features: 3D Primitives / Filter Visualization

Wu et al. 2015
Learned Features: 3D Primitives / Filter Visualization

Wu et al. 2015
Extract local, volumetric patches from RGBD data
- Extract **local, volumetric** patches from RGBD data
- Use a pair of 3D CNNs (with shared params) to produce 2048-dim feature vectors
Extract local, volumetric patches from RGBD data

Use a pair of 3D CNNs (with shared params) to produce 2048-dim feature vectors

Compare feature vectors via a fully connected NN
- Real-world RGBD datasets are available
- Only well-observed patches around keypoints are used for training
3DMatch Results

Zeng et al. 2016
Shape Classification Results

3DShapeNets (Wu et al.): 77.3%
MVCNN (Su et al.): 90.1%
Cause 1: Architecture and Engineering

LeNet, 1998

AlexNet, 2012

3DShapeNets, 2015
Cause 2: Resolution

**Multi-View CNNs**
MVCNN Su et al.

224x224 Images

**Volumetric CNNs**
3DShapeNets Wu et al.

30x30x30 Volumes
Compatible Representation

Polygon Mesh → Occupancy Grid 30x30x30 → Image 224x224

Same “3D Resolution”

Qi et al. 2016
Investigating Architectures

Different Architecture

Multi-View Image CNN

3D CNN

Same 3D Resolution (30x30x30)

Sphere Rendering Images

Occupancy Grid Volumes

Qi et al. 2016
Different Architecture and Same Resolution

MVCNN with Sphere Rendering Images

3DShapeNets Wu et al.
3D CNN with Micro-Neural Network

Qi et al. 2016
3D CNN with Micro-Neural Network

Shape Classification Accuracy

- MVCNN with Sphere Rendering Images
- 3DShapeNets Wu et al.
- Ours 3D CNN
Investigating Resolution

Multi-View Image CNN

Standard Rendering Images

Sphere Rendering Images

Same Architecture

Different 3D Resolution
Investigating Resolution

Qi et al. 2016

![Graph showing accuracy vs. 3D resolution with MVCNN-Sphere and Our 3D CNN metrics.](image-url)

- MVCNN-Sphere
- Our 3D CNN
Application

Dense Correspondences of Clothed Humans
3D Human Capture
3D Human Capture [Dou et al. '16]
Analysis & Reasoning
Shape Analysis

SCAPE model of Lee from Hirshberg et al. 2012
Motion Understanding

raw scans
Motion Understanding

raw scans

“grasping”
Correspondences?
Non-Rigid Registration [Li et al. 2008]

target

source

correspondences

overlap
Large Pose Changes

source & target

[Li et al. 09]

[Huang et al. 08]
Descriptors

<table>
<thead>
<tr>
<th>partial scans</th>
<th>complete model (or small holes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Hebert 99]</td>
<td>[Jain &amp; Zhang 06]</td>
</tr>
<tr>
<td>[Bronstein et al. 06]</td>
<td>[Bronstein et al. 10]</td>
</tr>
<tr>
<td></td>
<td>[Kim et al. 11]</td>
</tr>
<tr>
<td></td>
<td>[Windheuser et al. 14]</td>
</tr>
<tr>
<td></td>
<td>[Chen &amp; Koltun 15]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>[Taylor et al. 12]</td>
<td>[Litman &amp; Bronstein 14]</td>
</tr>
<tr>
<td>[Pons-Moll et al. 15]</td>
<td>[Rodola et al. 14]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Windheuser et al. 14]</td>
</tr>
<tr>
<td></td>
<td>[Macsi et al. 15]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

designed descriptor

learned descriptor
Clothed and Partial Data

immense space of variations
Classification Networks
Deep Convolutional Neural Network

classification network, e.g. AlexNet [Krizhevsky et al. 2012]
Deep Convolutional Neural Network
Deep Convolutional Neural Network

3D model

depth image

DNN

feature descriptor

“butt” descriptor

“butt”
Loss Function

Training Data → DNN → Loss Function → Classification?
Classification Task

descriptors are far apart
How to preserve distances?
Deep Convolutional Neural Network

Training Data → DNN → Loss Function
Loss Function

Training Data

(Anchor, Positive, Negative)

Loss Function

Triplet Loss
Multi-Segmentation
Distance Preserving Learning

500 classes

100 random segmentations
Distance Preserving Learning

AlexNet

image 1x512x512

DNN descriptor 16x512x512

100 segmentations 500 classes

Classification

Classification

Classification
Variation on Clothing

2100 meshes

33 landmarks

DNN

descriptor

Classification

Classification

Landmark Classification

SCAPE
MIT
Yobi3D
Yobi3D
Yobi3D
Training Data

Shape & Pose

SCAPE
MIT

Clothing

Yobi3D
Yobi3D
Yobi3D
Results
Results: Static Shapes

full-to-full correspondences (synthetic data and naked)
Results: Static Shapes

full-to-partial correspondences (real data and clothed)

source

target
Results: Dynamic Shapes

input scans

correspondences (per frame)
Results: Dynamic Shape Reconstruction

dynamic correspondences (side view)

input scans | correspondences (per frame) | reconstruction and input scans
Dense Correspondences
Applications
Low Cost Capture & Moving Target
Registration and Reconstruction

output scan alignment

output textured reconstruction
Filtering and Texture Reconstruction

denoised mesh

dense correspondences

textured mesh reconstruction
Application

Photorealistic Texture Synthesis
Photo-Realistic Faces Using Deep Learning

input picture  output albedo map  rendering  rendering (zoom)
Deep CNN-based Synthesis Approach

Initial face model fitting

- Input image
- Partial albedo map (HF)
- Complete albedo map (LF)

Texture analysis

- Face database
- Feature correlations

Texture synthesis

- Complete albedo map (HF)

Output rendering
Feature Correlations (Gatys et al. 2015)

\[ G^l(I) = \frac{1}{M_l} F^l(I) (F^l(I))^T \in \mathbb{R}^{N_l \times N_l} \]

Feature correlation

\[ F^l(I) \in \mathbb{R}^{N_l \times M_l} \]

Feature response
Texture Analysis

Database

Partial feature correlation set

Fitting via convex combination

Complete feature correlation set

Feature correlation evaluation

Coefficients

Partial albedo (HF)

Partial feature correlation

Complete feature correlations
Texture Synthesis (Gatys et al. 2015)

Loss function: $E_L = \sum (\hat{G}^L - G^L)^2$

Total loss: $\mathcal{L}(\tilde{x}, \hat{x}) = \sum_{l=0}^{L} w_l E_l$

Gradient descent: $\tilde{x} := \hat{x} - \alpha \frac{\partial \mathcal{L}}{\partial \tilde{x}}$
Texture Synthesis (Saito et al. 2016)

\[
\min_{I} \sum_{l \in L_F} \left\| F^l(I) - \hat{F}^l(I_0) \right\|^2_F + \alpha \sum_{l \in L_G} \left\| G^l(I) - \hat{G}^l(I_0) \right\|^2_F
\]
Different Number of Mid-Layers

1 layer  2 layers  3 layers  4 layers  5 layers
Detail Preservation via Convex Combination

input  visible texture  unconstrained least square  convex constraint
Consistent Reconstruction from Different Views

input image  albedo map  input image  albedo map
Comparison

ours

Light Stage

PCA
Huynh et al. (2018)

Geometry Synthesis
Yamaguchi et al. (2018)

**Full Geometry and Reflectance Inference**

- **iPhone picture**
- **reflectance map**
Yamaguchi et al. (2018)

Full Geometry and Reflectance Inference

iPhone picture

high-fidelity reconstruction
Yamaguchi et al. (2018)

Full Geometry and Reflectance Inference

internet picture

high-fidelity reconstruction
Yamaguchi et al. (2018)

Full Geometry and Reflectance Inference

internet picture
high-fidelity reconstruction
Hair Digitization
Five-strand Dutch braid

Reference photo

Our result
Deep Learning for Hair Modeling

Saito et al. (2018)
Deep Learning for Hair Modeling

Saito et al. (2018)
Deep Learning for Hair Interpolation


Saito et al. (2018)
Wei et al. (2018)
Starting

Wei et al. (2018)
Color Interpolation

reference image A  result A  interpolation result  result B  reference image B

Wei et al. (2018)
What’s next?
Interaction With 3D Avatars

Blade Runner 2049 (2017)
VFX-Level Augmented Reality
Thanks!