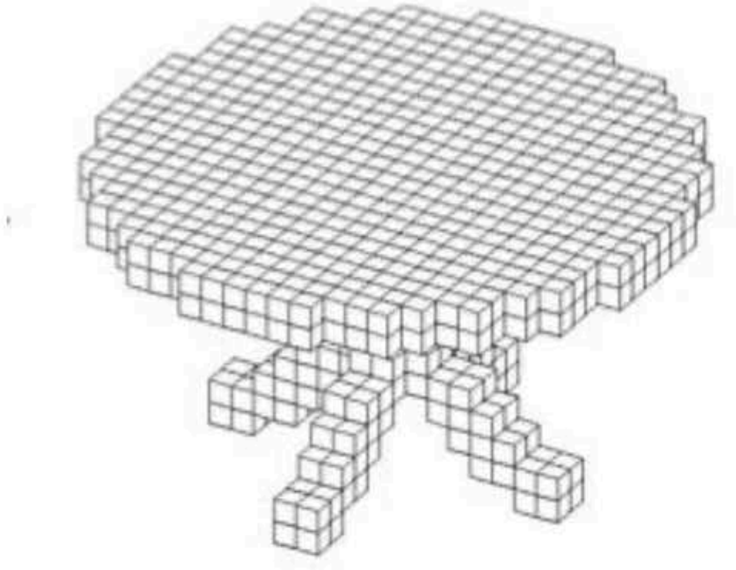
Deep Learning
for Geometry Processing

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

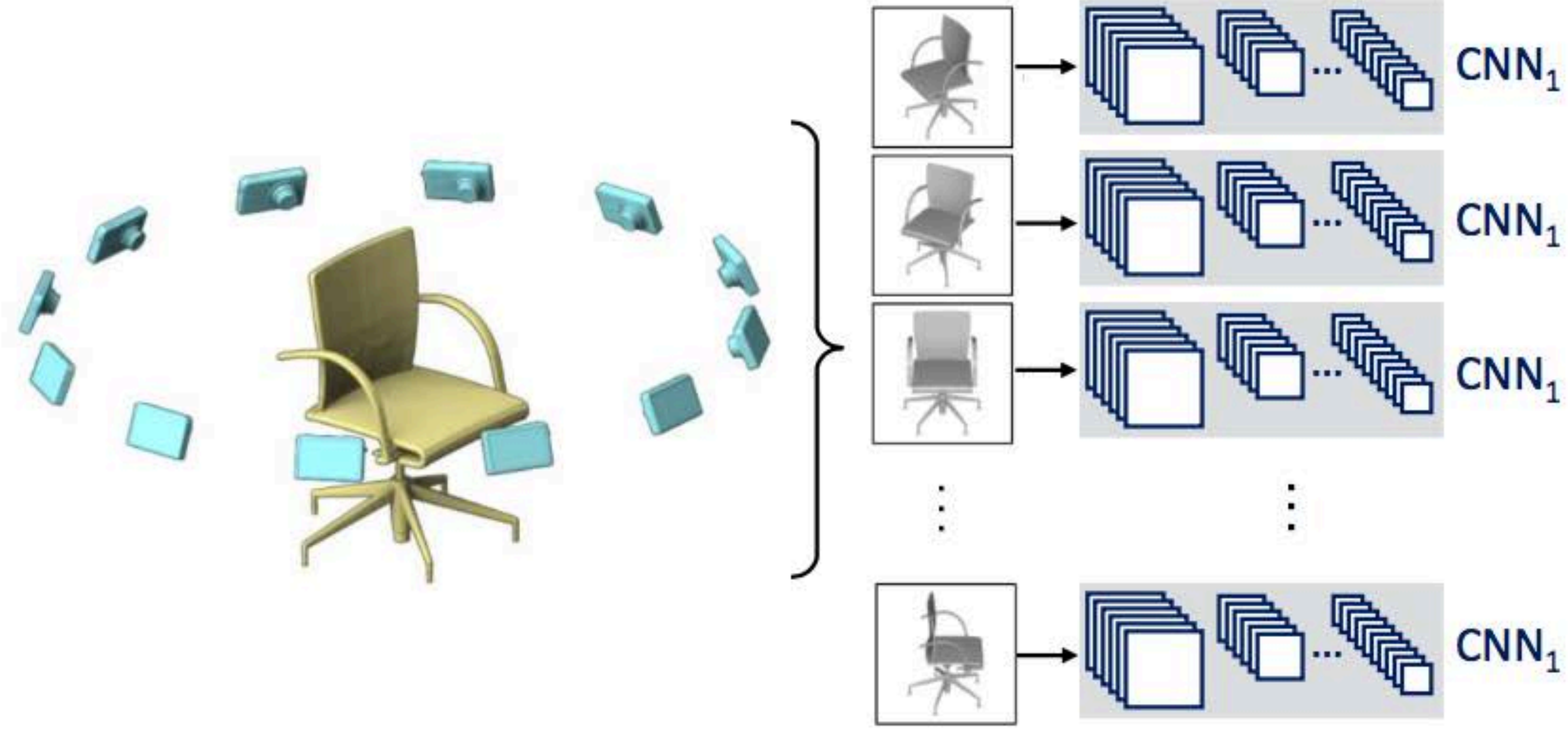
Multi-Views

MVCNN by H. Su et al. ICCV 15

DeepPano by B. Shi et al. IEEE/SPL 15

Multi-View CNNs

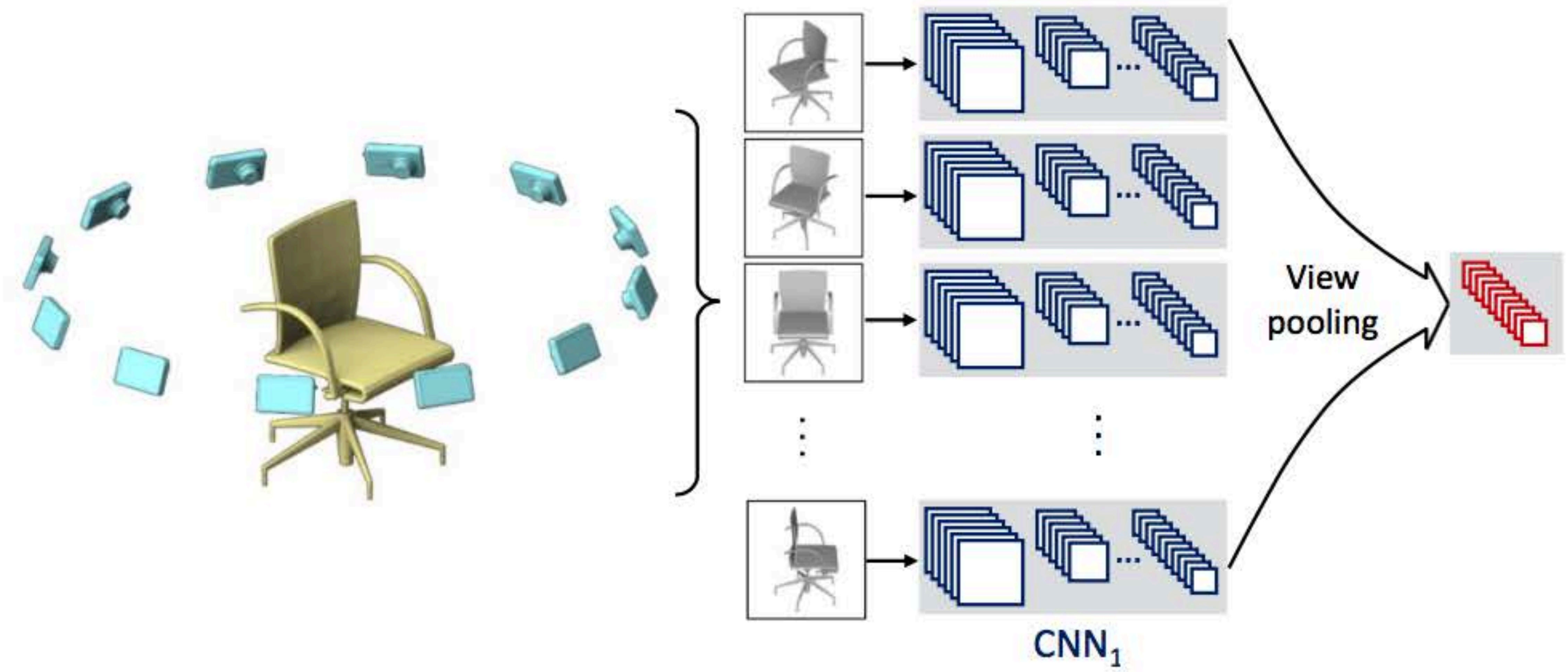
Su et al. 2015



- CNN1 extracts image features (parameters are shared across views)

Multi-View CNNs

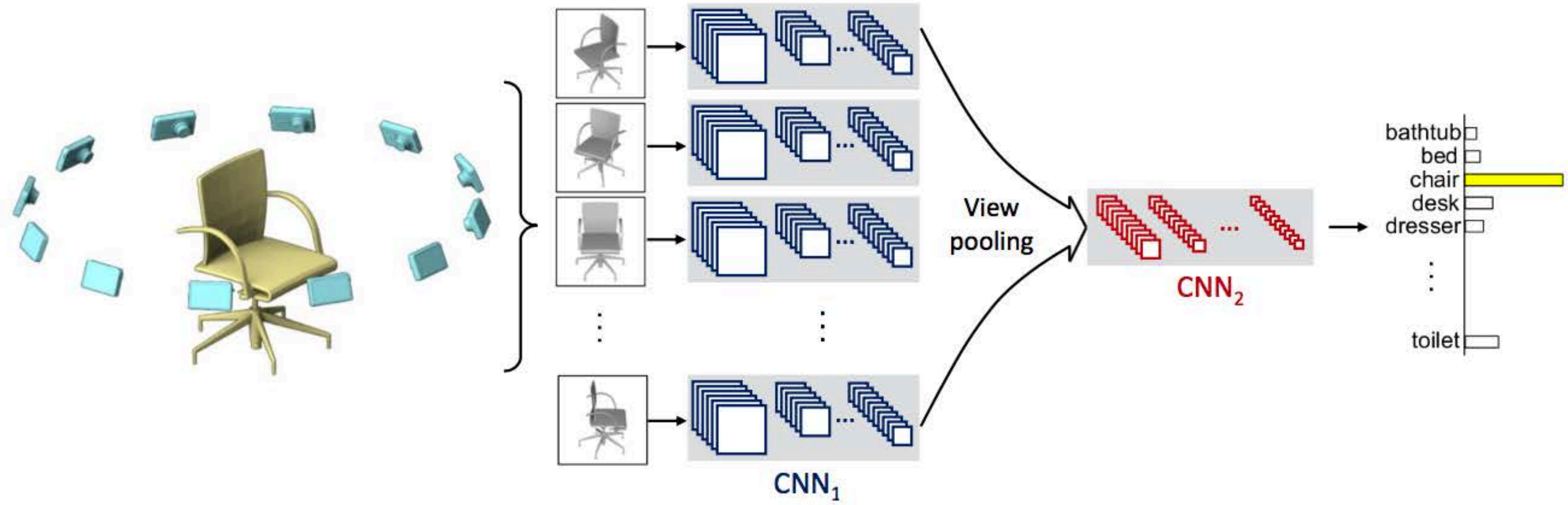
Su et al. 2015



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

Multi-View CNNs

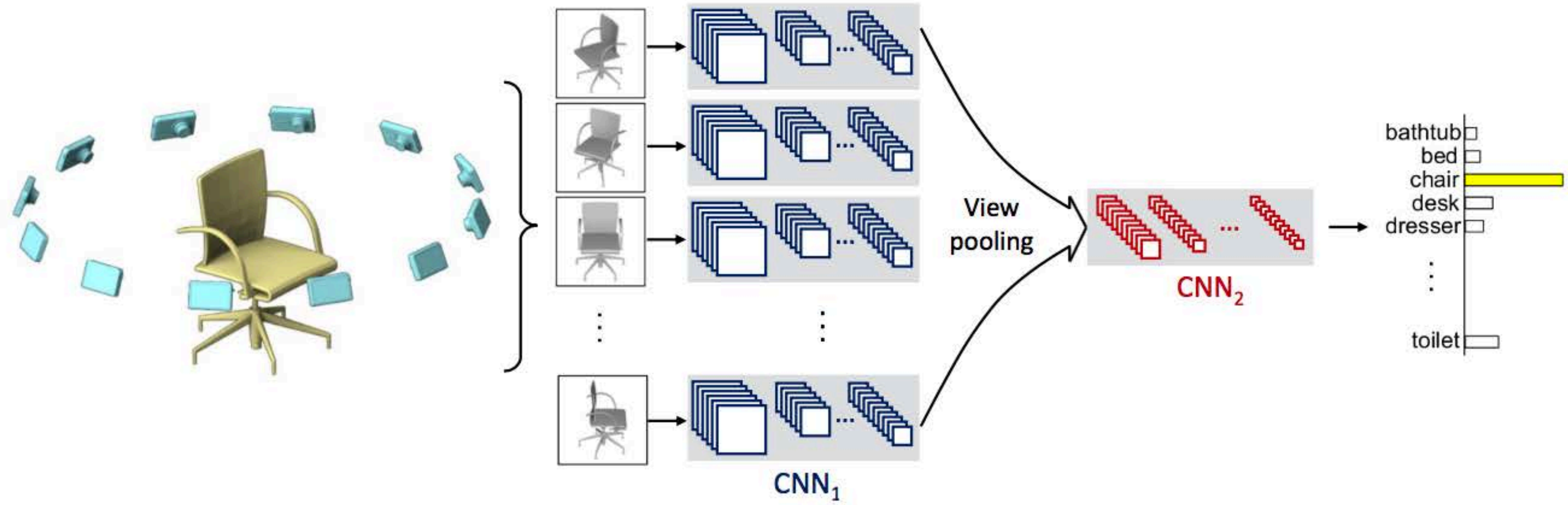
Su et al. 2015



- 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

Su et al. 2015



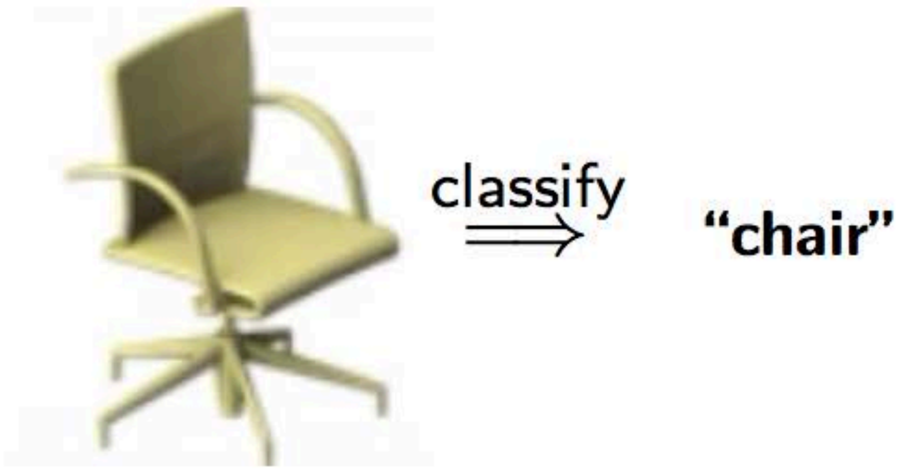
- CNN1 extracts image features (parameters are shared across views)
- Element-wise max pooling across all views
- CNN2 produces shape descriptors + final prediction

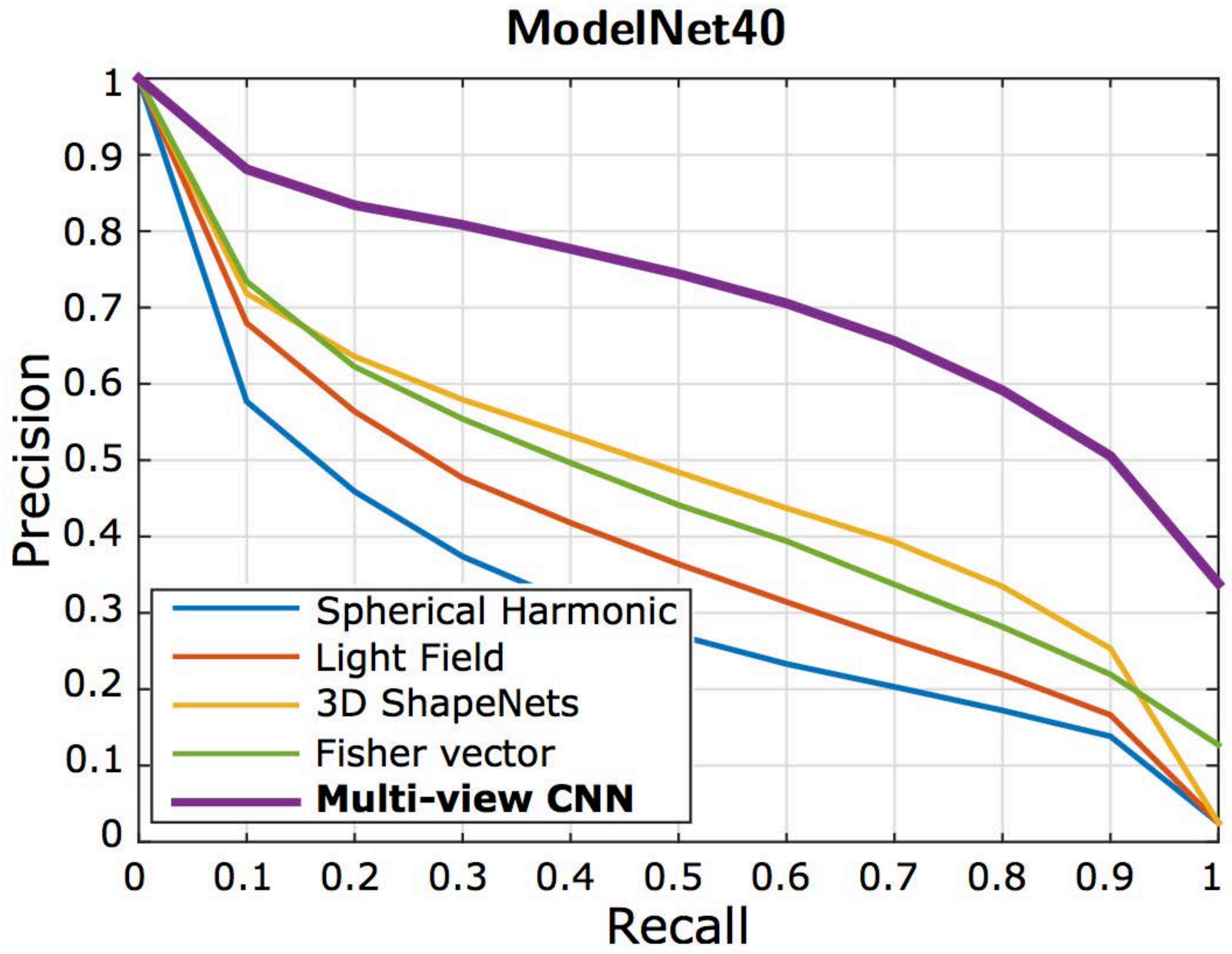
Multi-View CNNs

Su et al. 2015

3D shape classification and retrieval

- Pre-trained on ImageNet
- Fine-tuned on 2D views

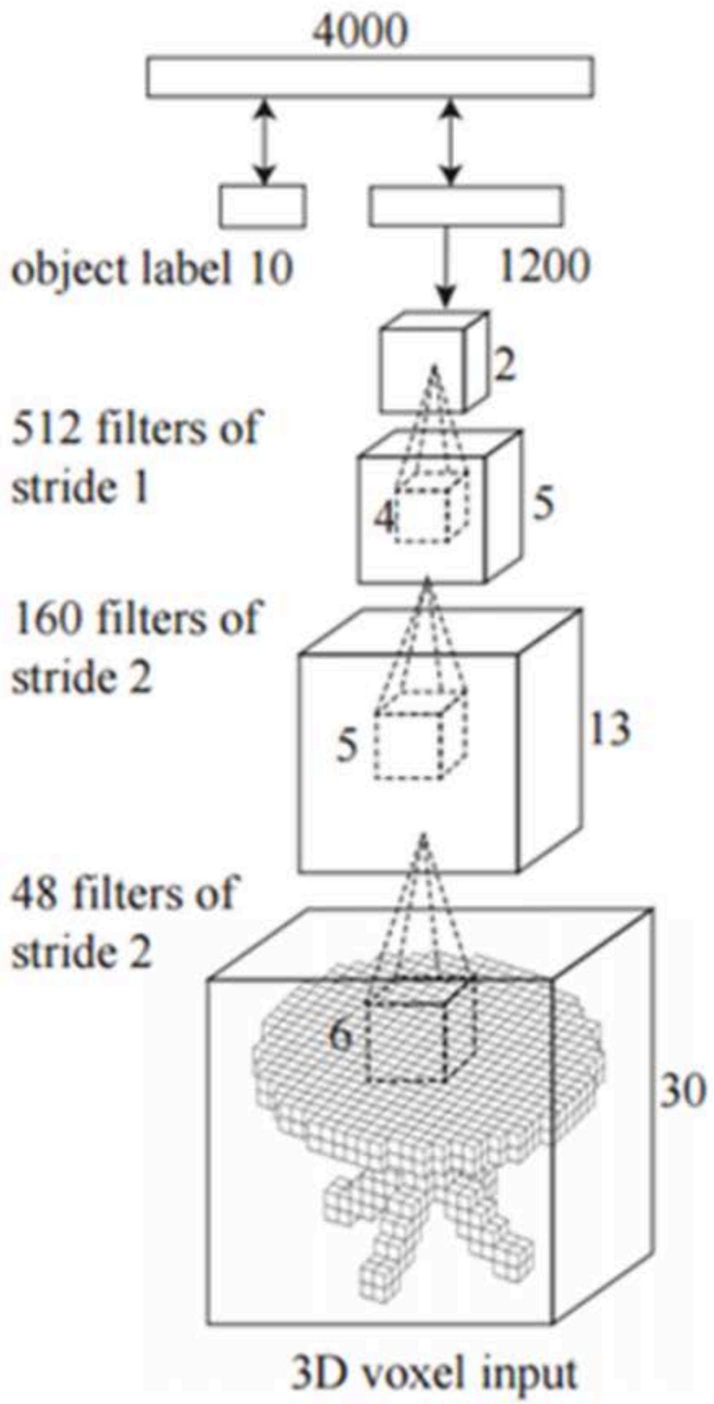




Volumetric CNNs

Wu et al. 2015

Volumetric representation: shapes as binary voxels in a 3D grid



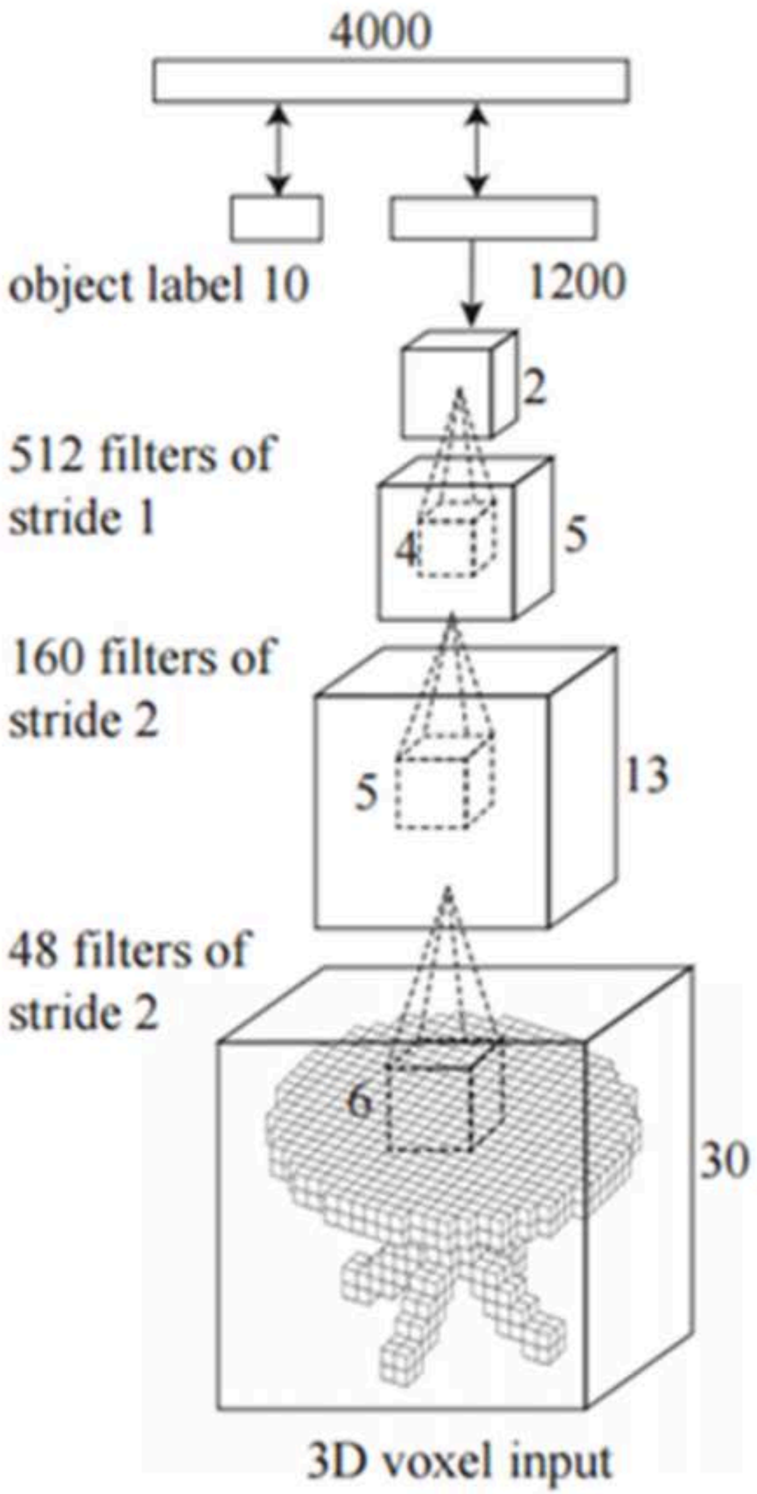
convolutional deep belief network

Volumetric CNNs

Wu et al. 2015

Volumetric representation: shapes as binary voxels in a 3D grid

Learn filters operating on these volumetric data

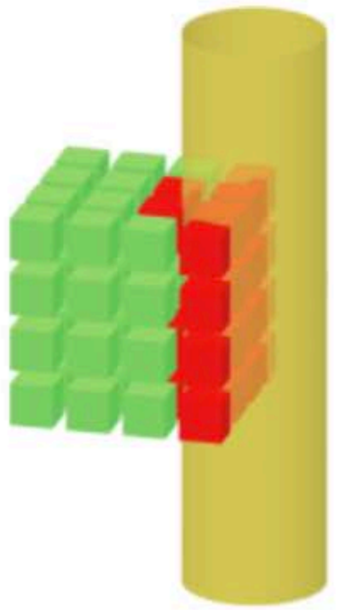


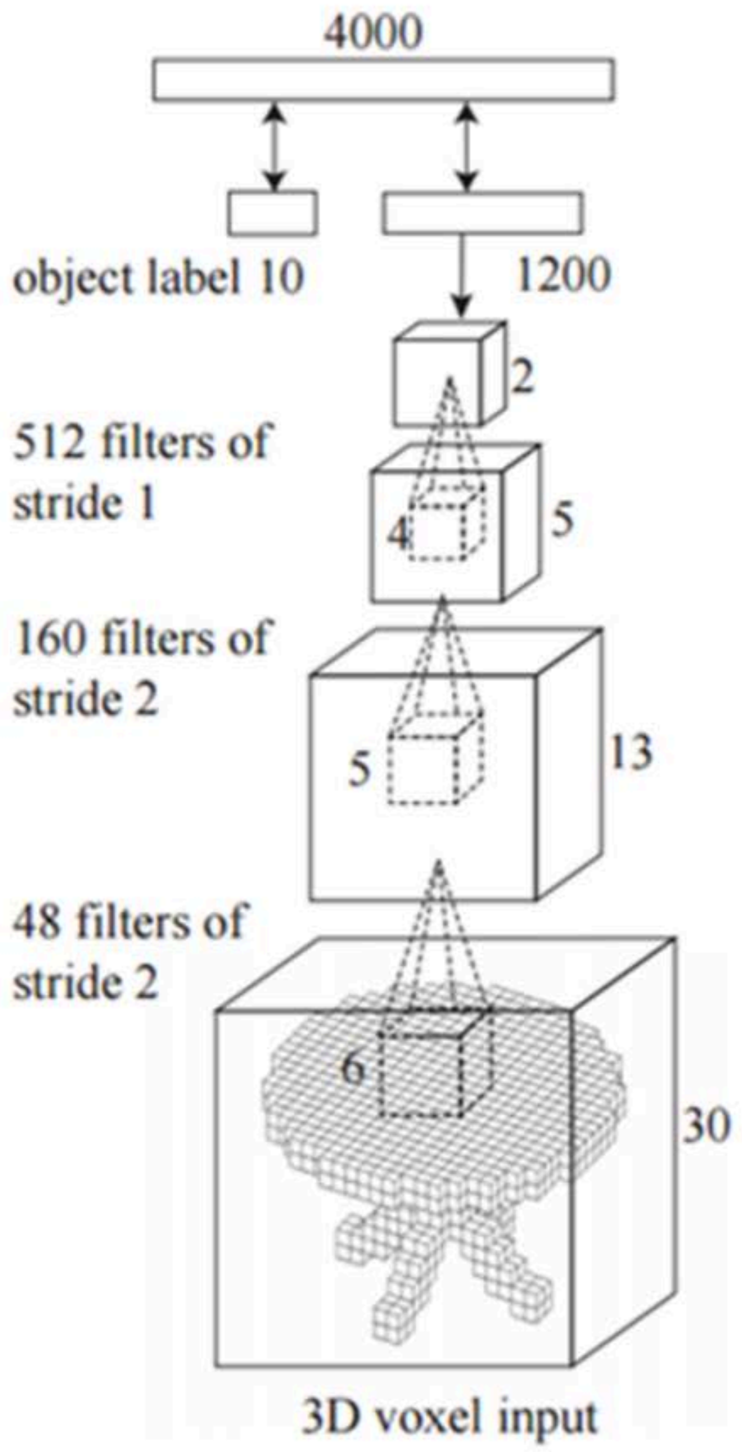
convolutional deep belief network

Volumetric CNNs

Wu et al. 2015

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

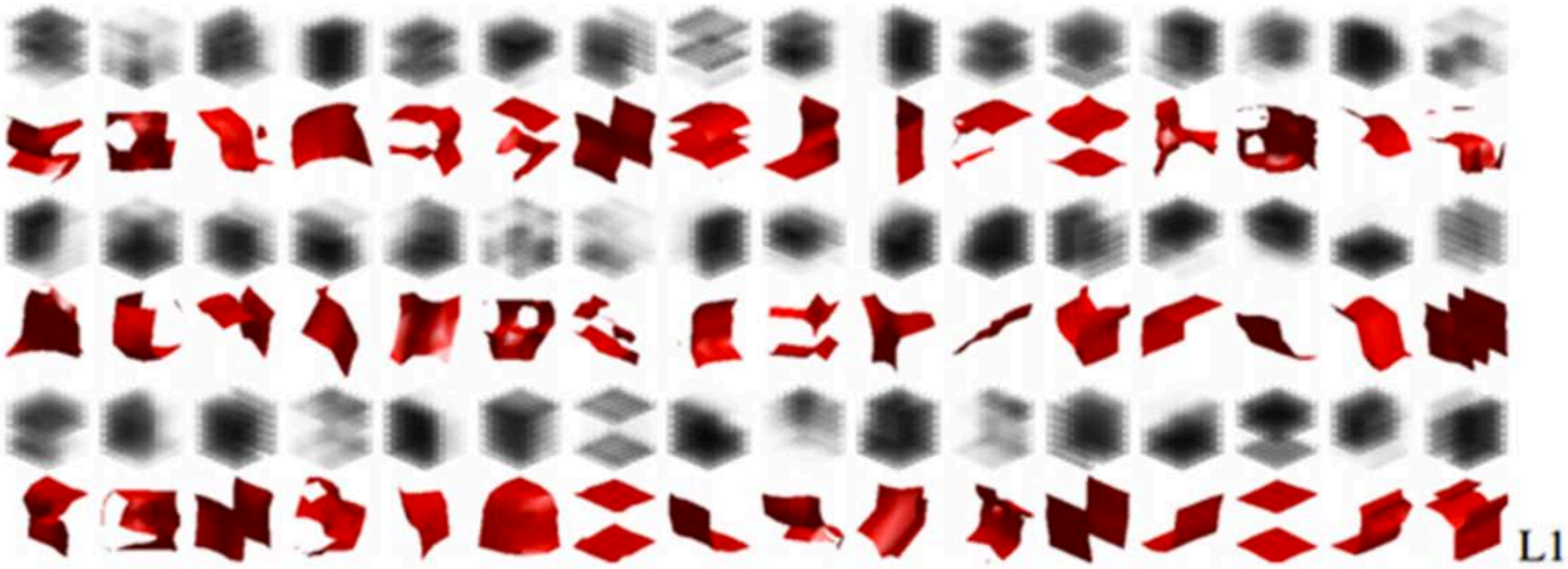




convolutional deep belief network

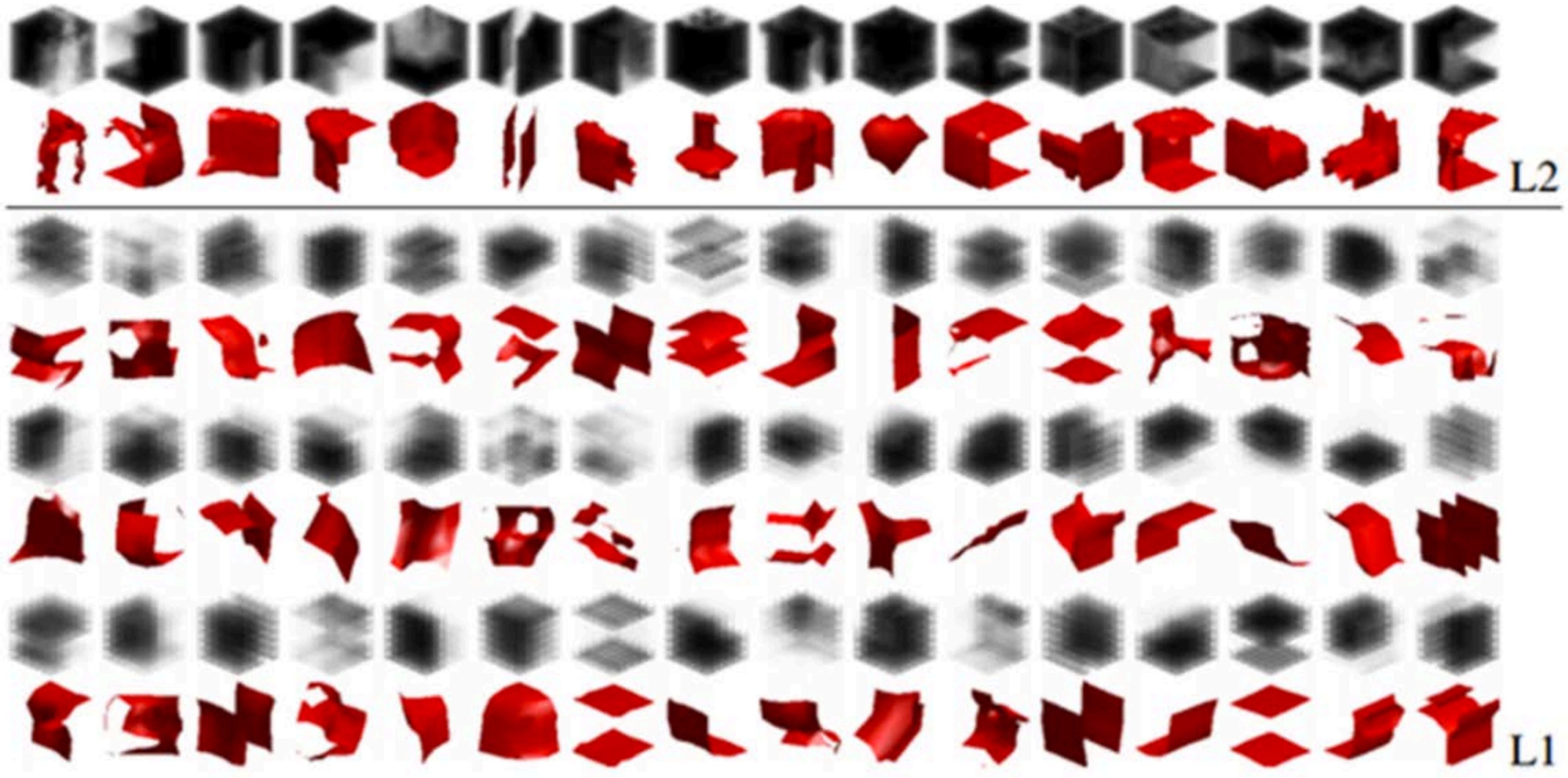
Learned Features: 3D Primitives / Filter Visualization

Wu et al. 2015



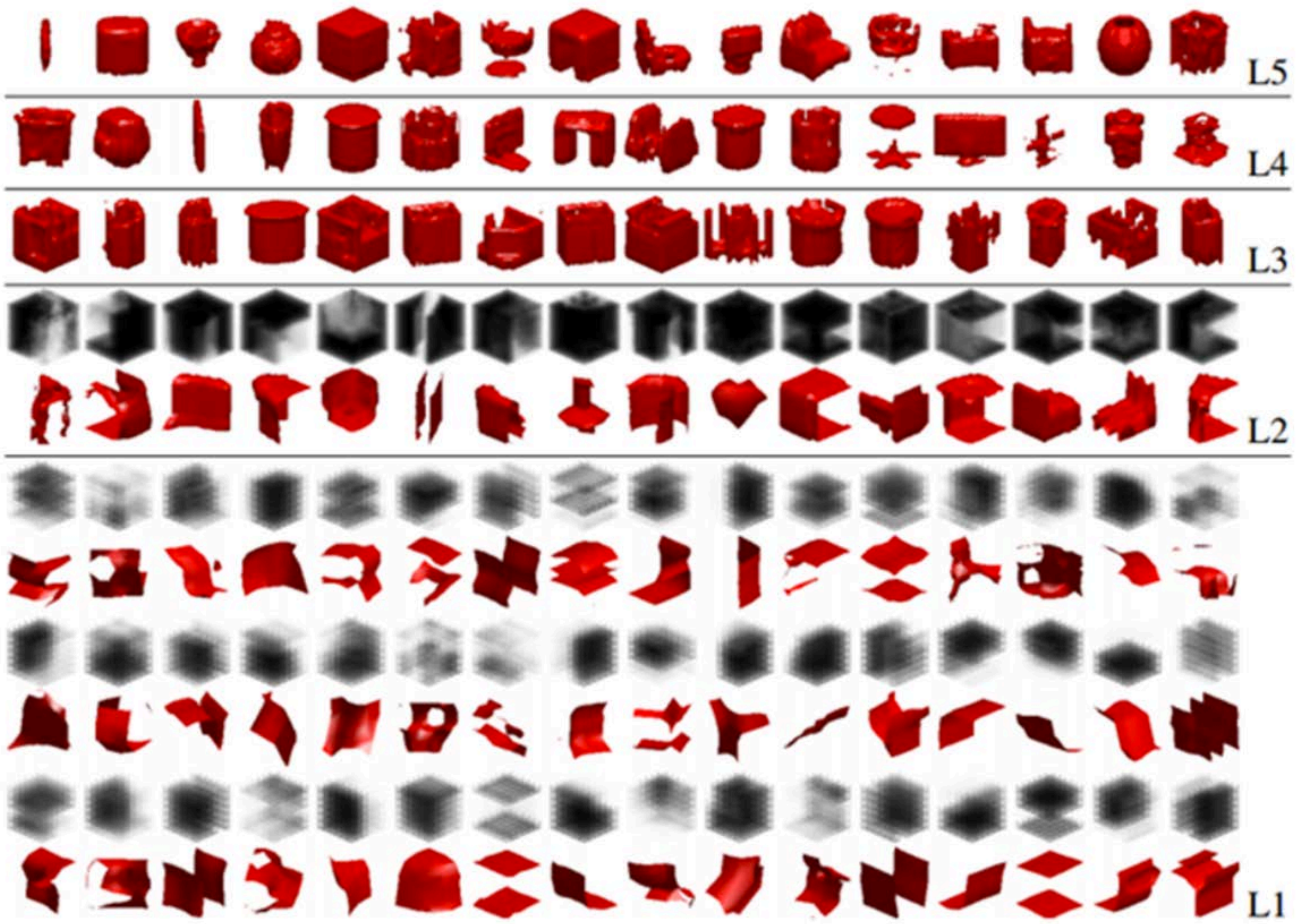
Learned Features: 3D Primitives / Filter Visualization

Wu et al. 2015



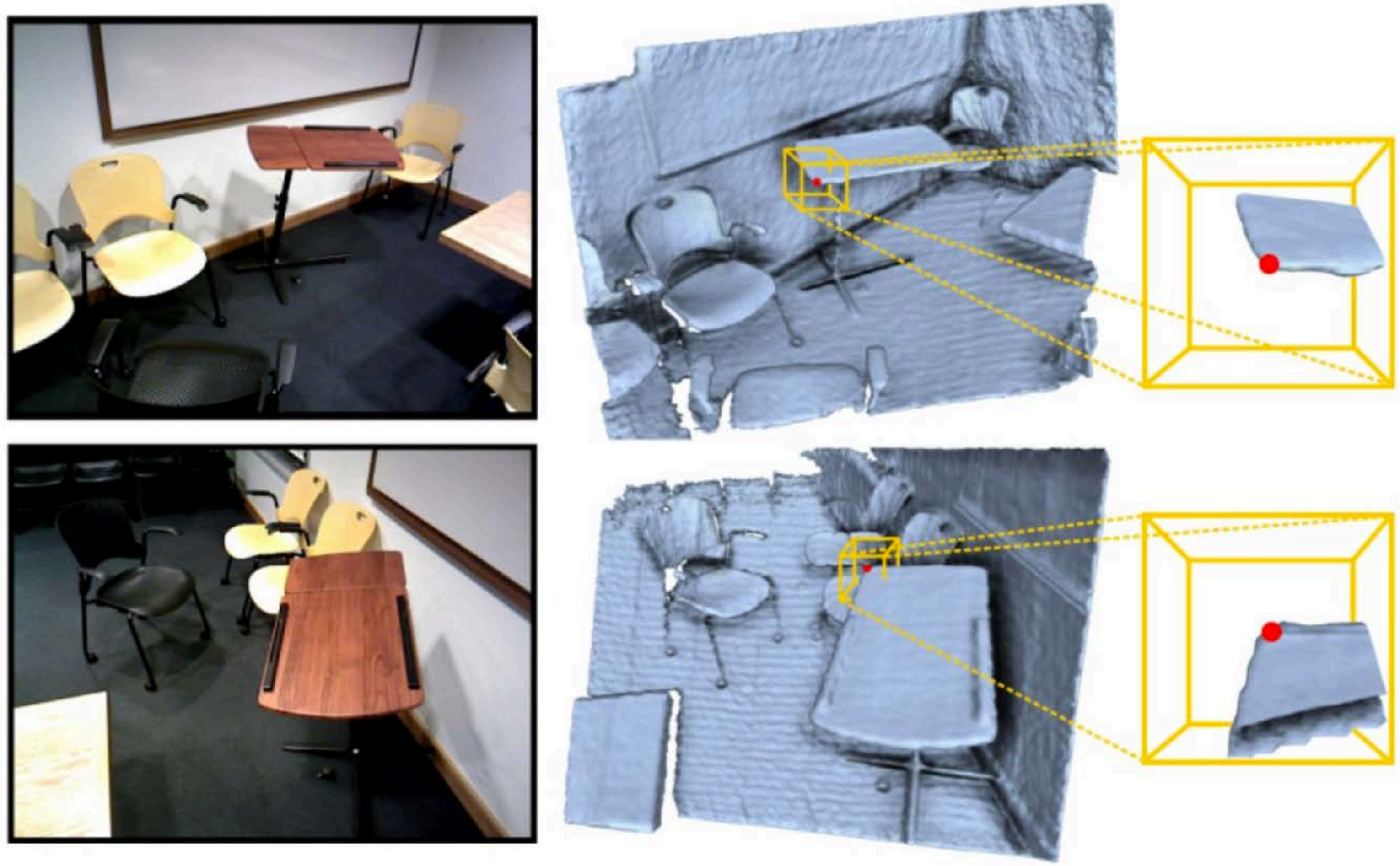
Learned Features: 3D Primitives / Filter Visualization

Wu et al. 2015



3DMatch

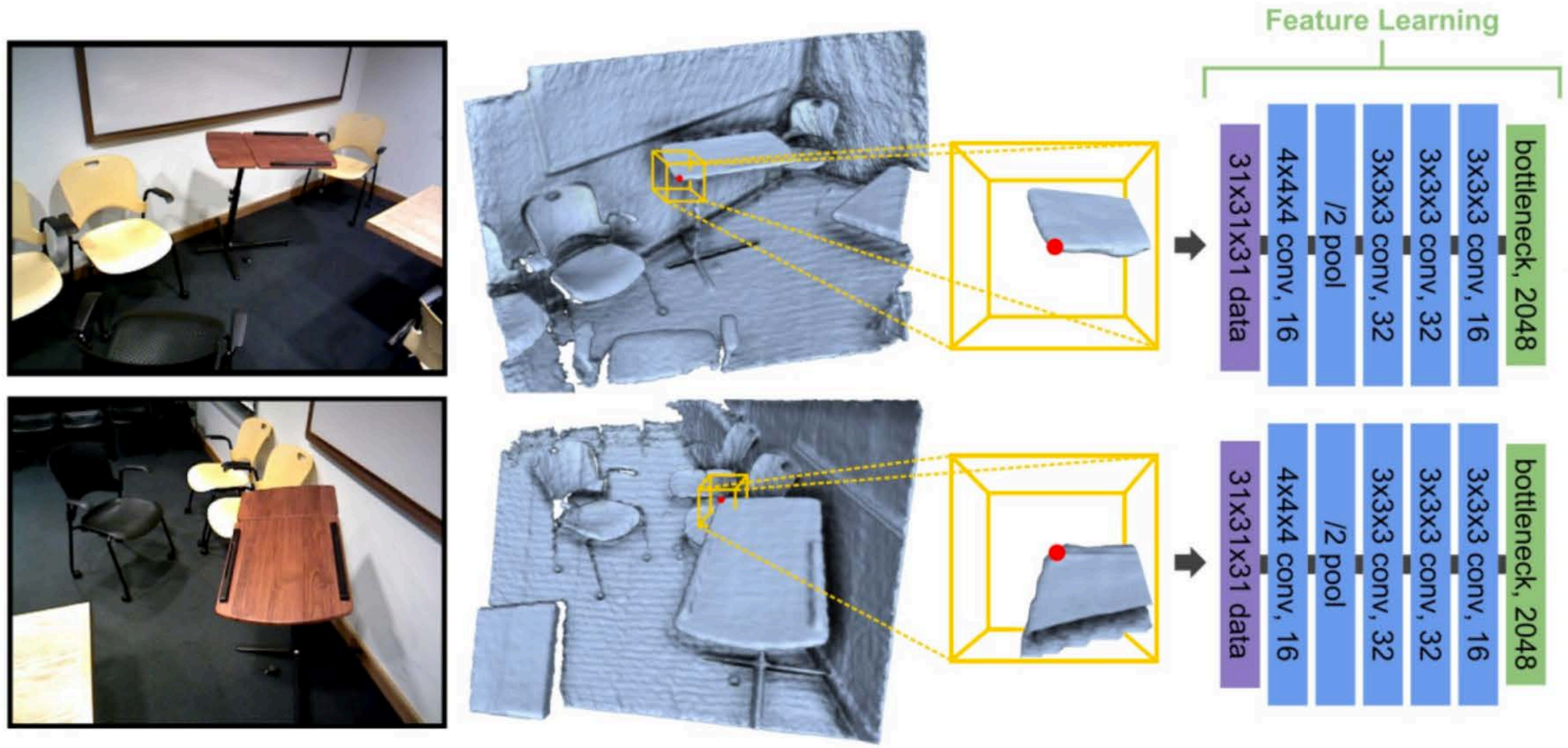
Zeng et al. 2016



- Extract local, volumetric patches from RGBD data

3DMatch

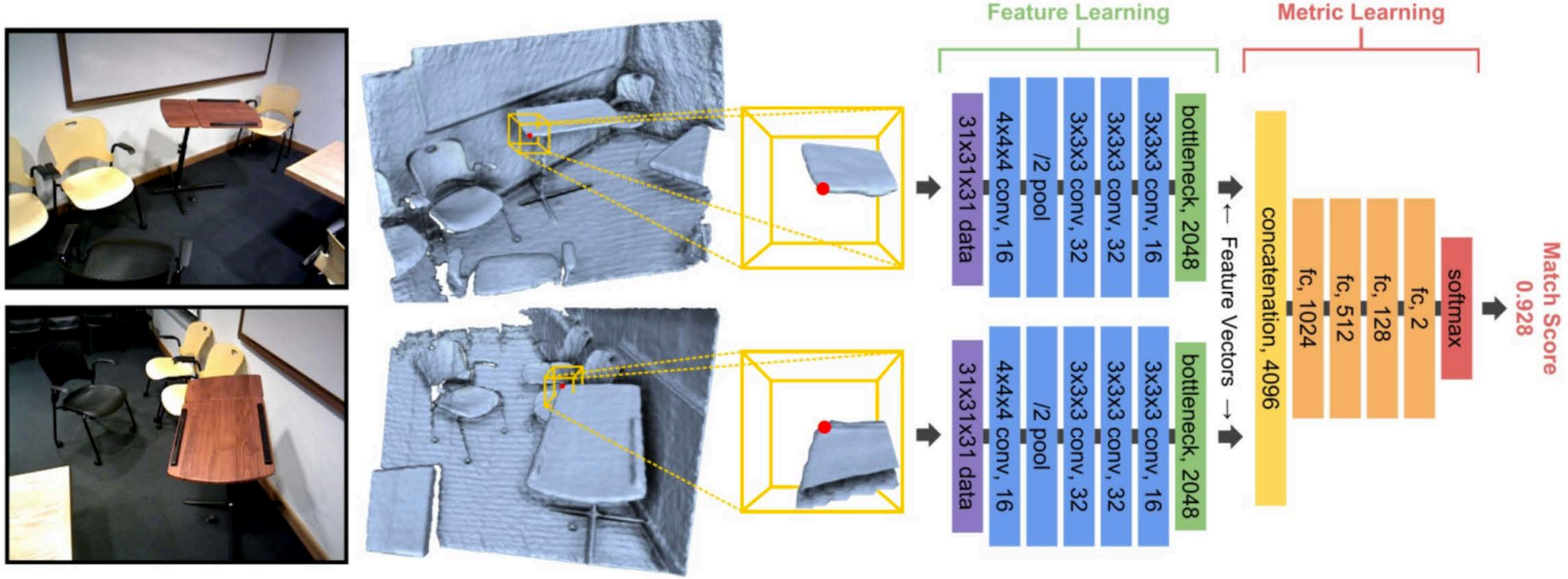
Zeng et al. 2016



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

3DMatch

Zeng et al. 2016

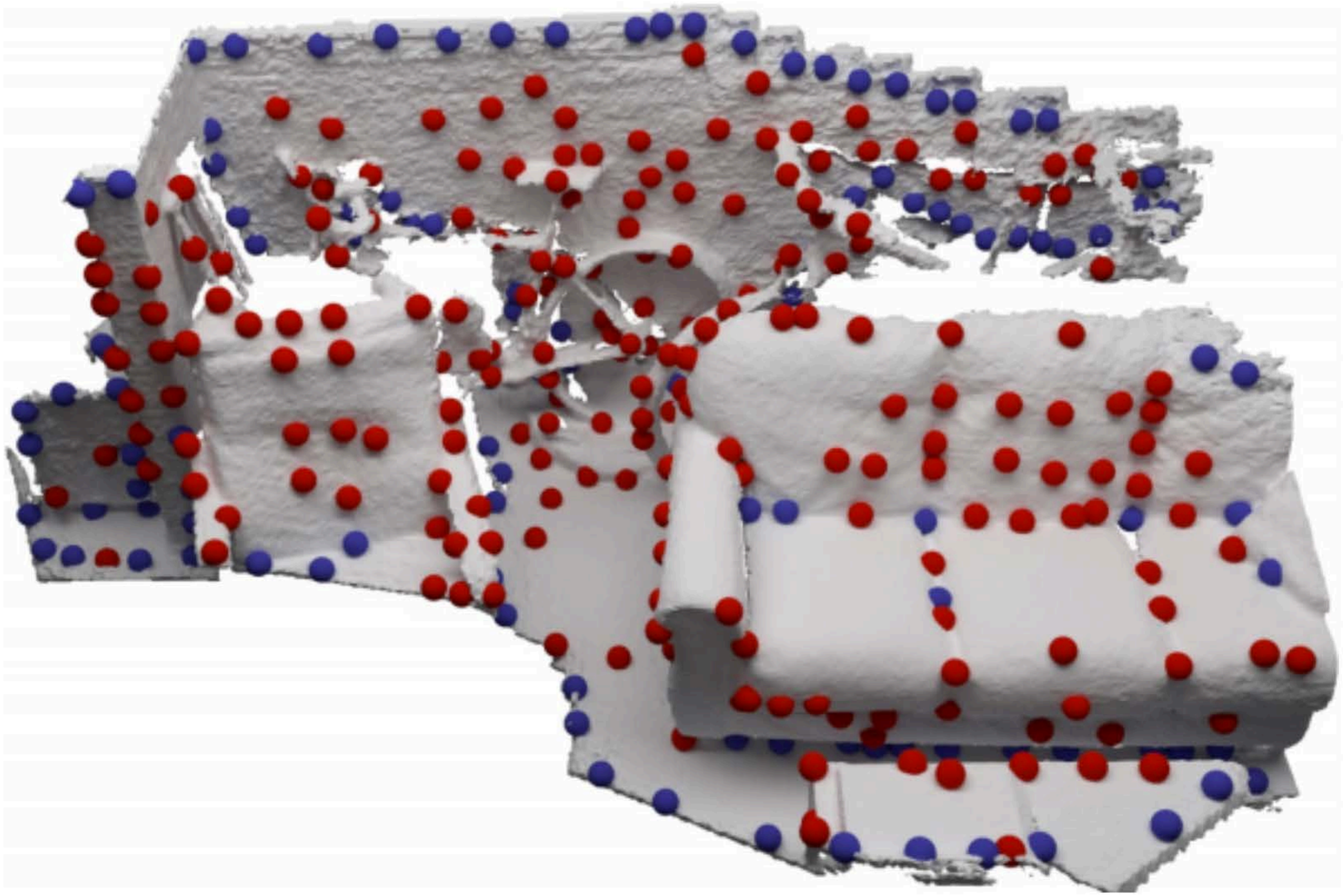


- Extract local, volumetric patches from RGBD data
- Use a pair of 3D CNNs (with shared params) to produce 2048-dim feature vectors
- Compare feature vectors via a fully connected NN

Training Data

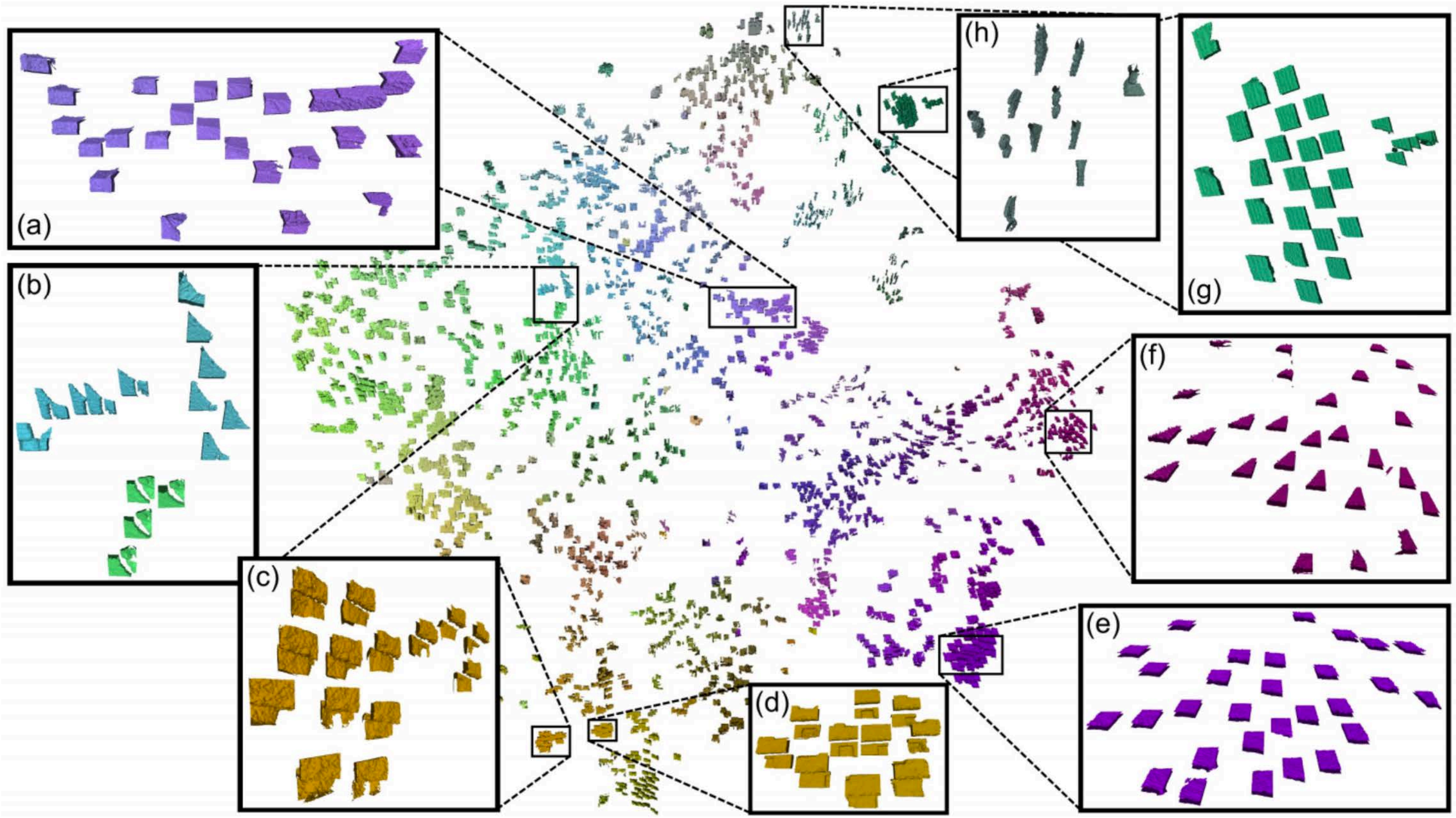
Zeng et al. 2016

- Real-world RGBD datasets are available
- Only well-observed patches around keypoints are used for training



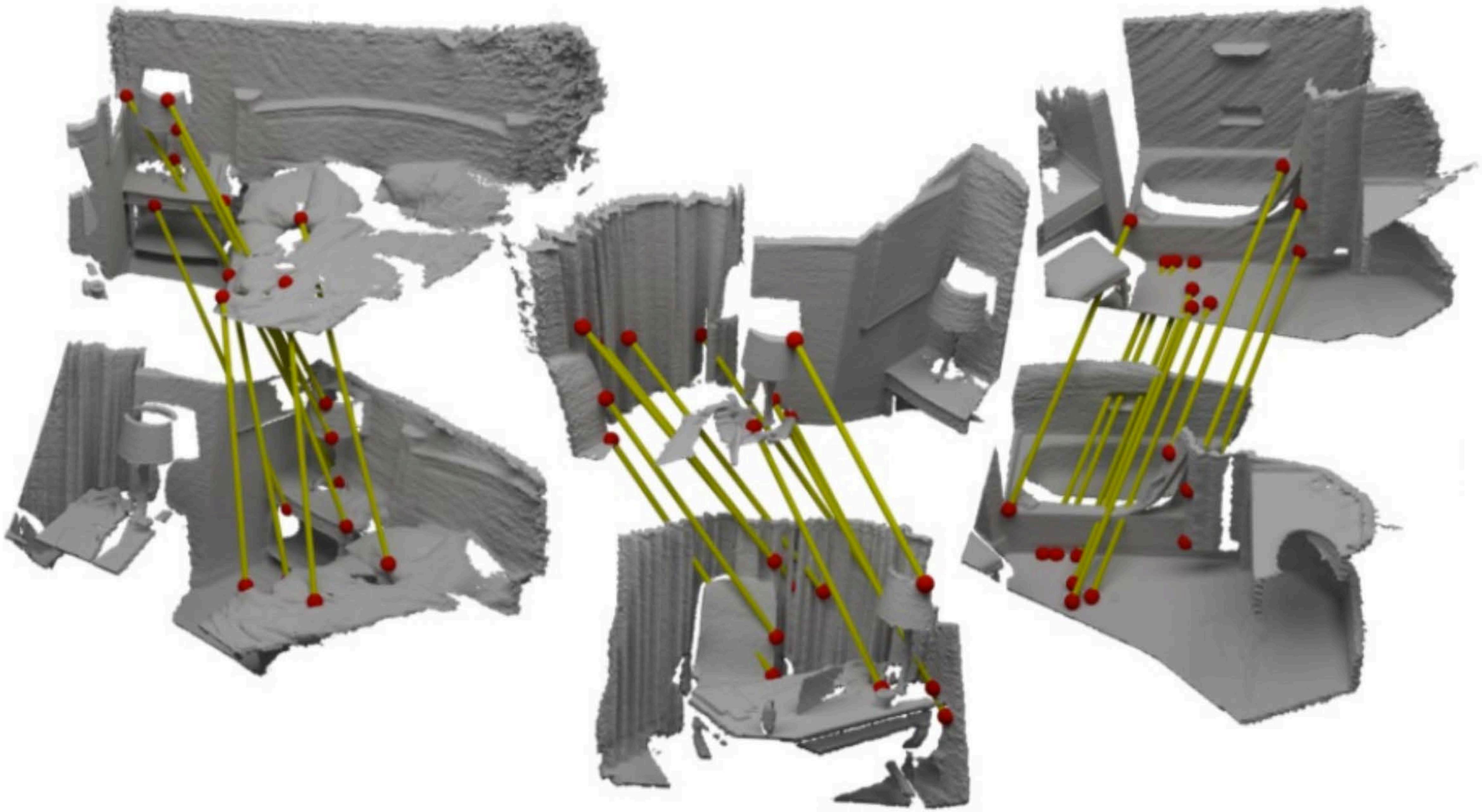
3DMatch Embedding

Zeng et al. 2016



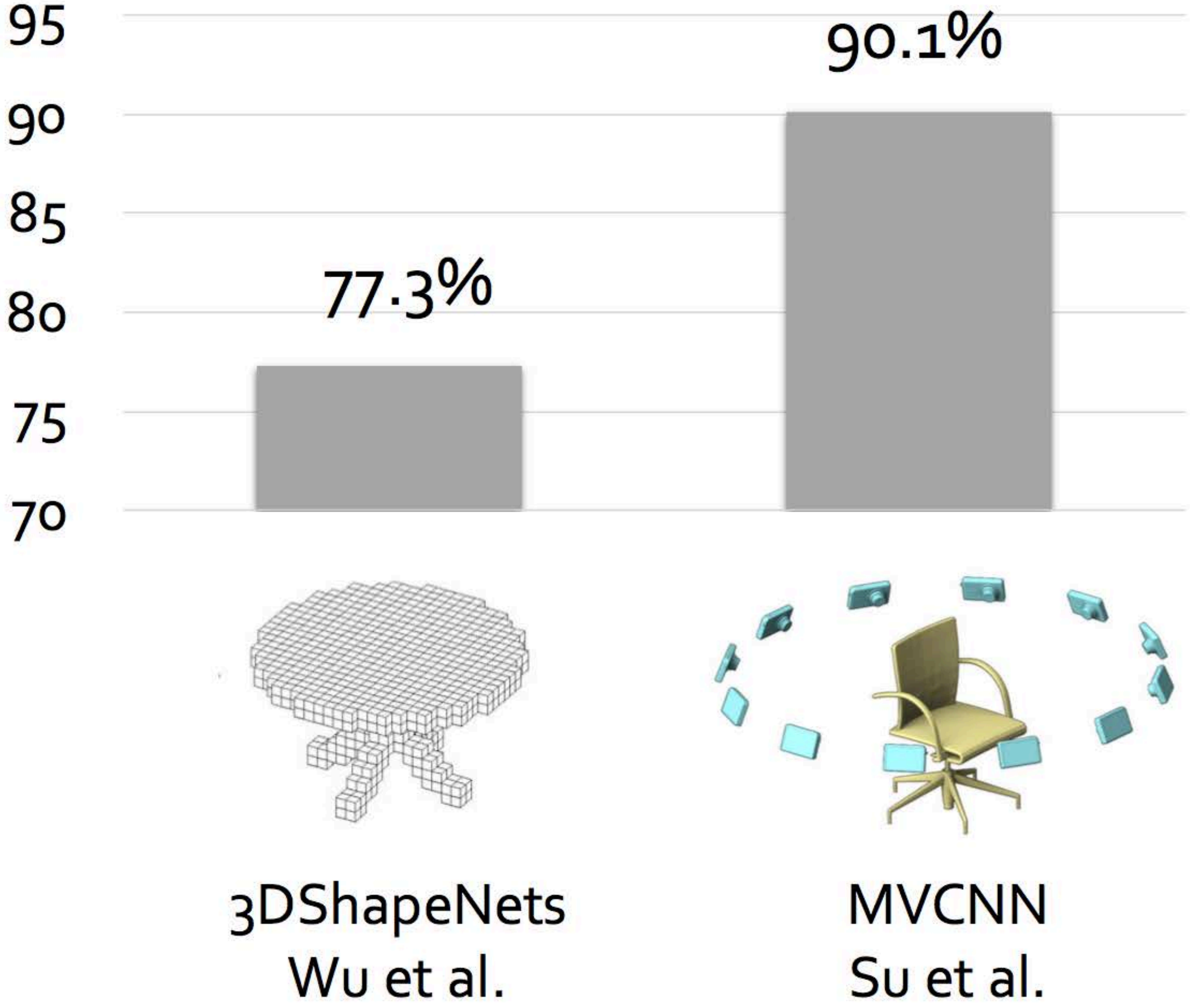
3DMatch Results

Zeng et al. 2016

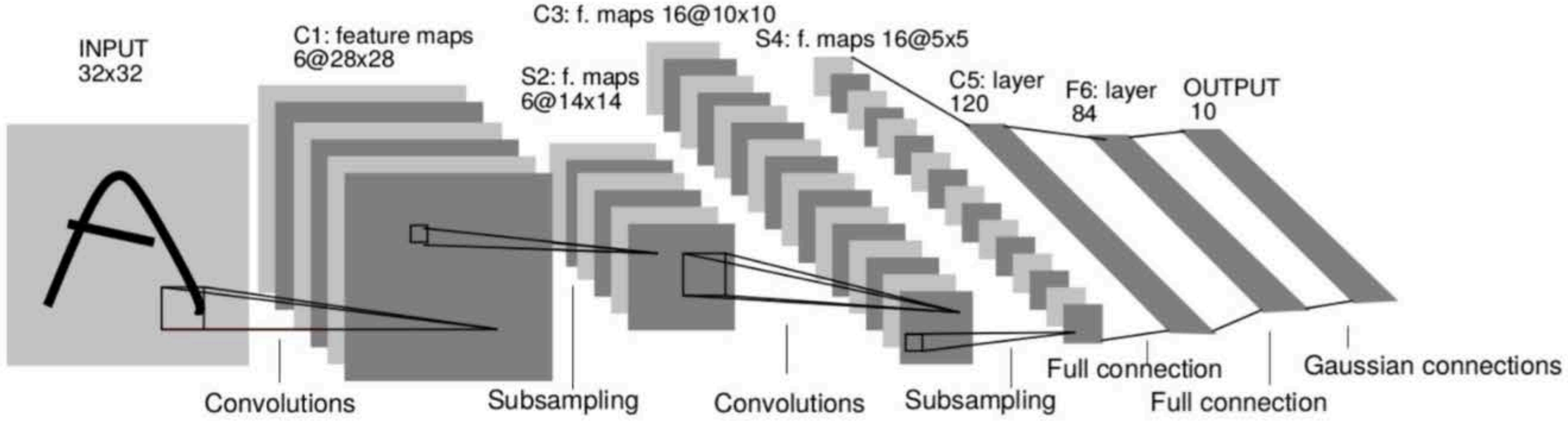


Shape Classification Results

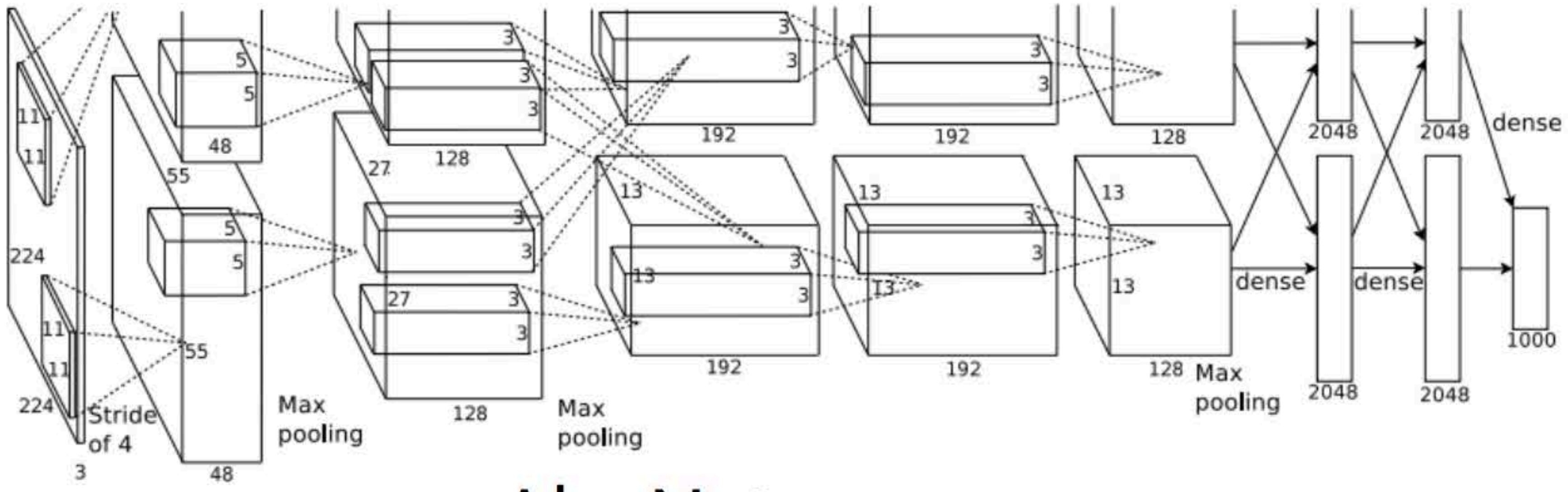
Qi et al. 2016



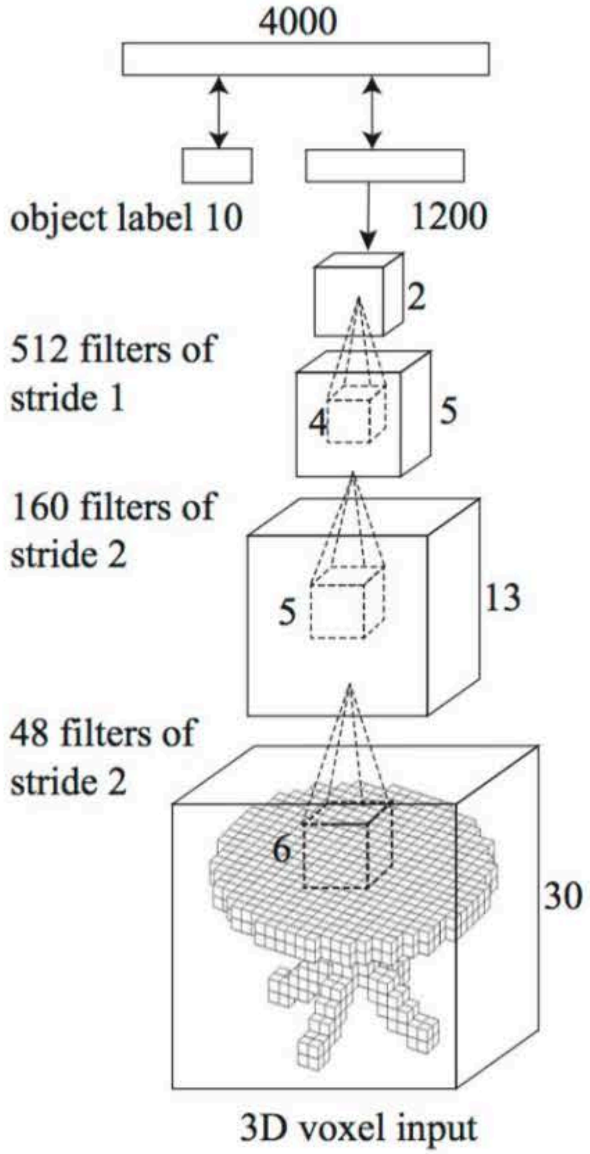
Cause 1: Architecture and Engineering



LeNet, **1998**



AlexNet, **2012**



3DShapeNets, **2015**

Cause 2: Resolution

Qi et al. 2016

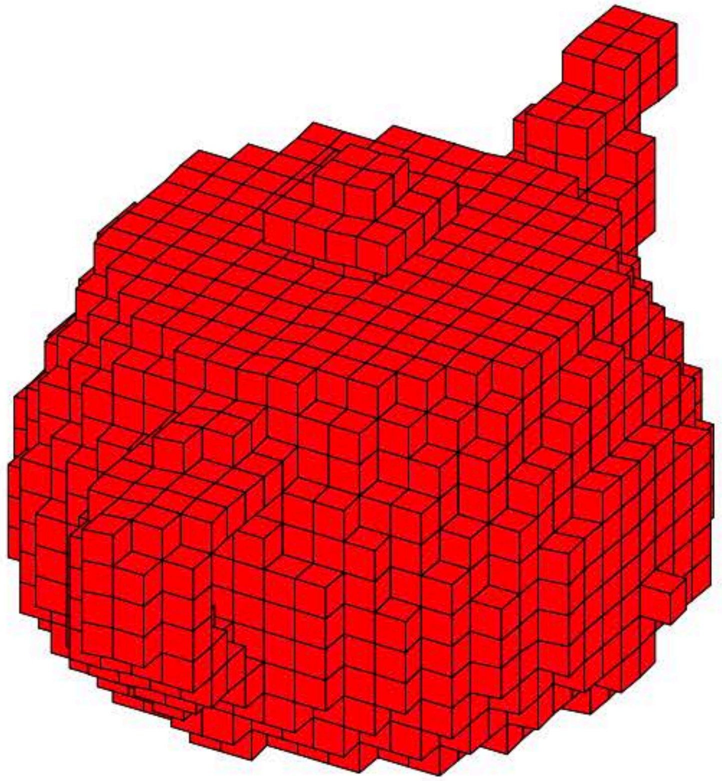
Multi-View CNNs

MVCNN Su et al.

224x224 Images

Volumetric CNNs

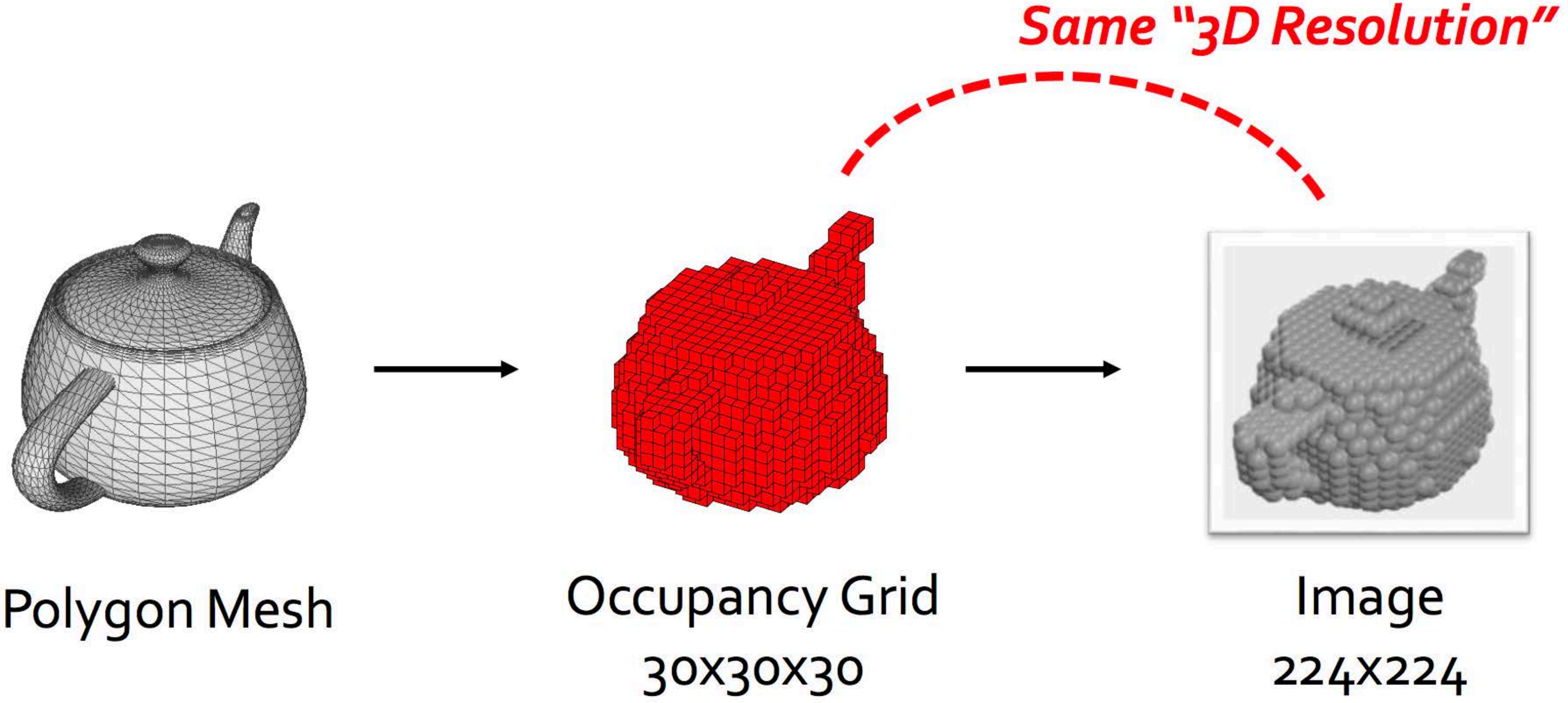
3DShapeNets Wu et al.



30x30x30 Volumes

Compatible Representation

Qi et al. 2016



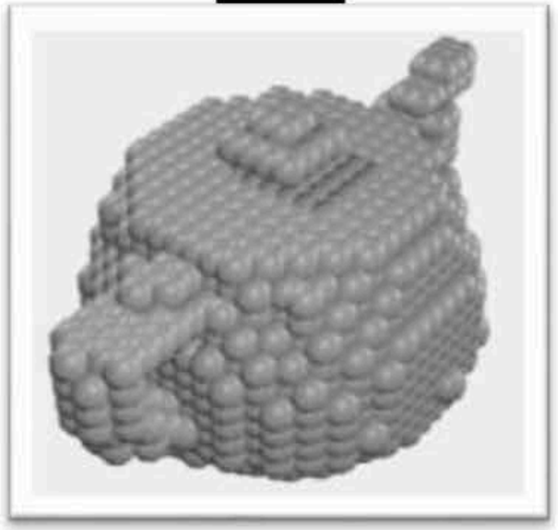
Investigating Architectures

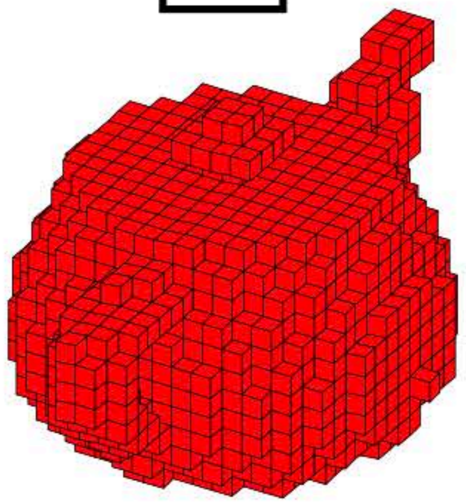
Qi et al. 2016

Different Architecture

Multi-View Image CNN

3D CNN





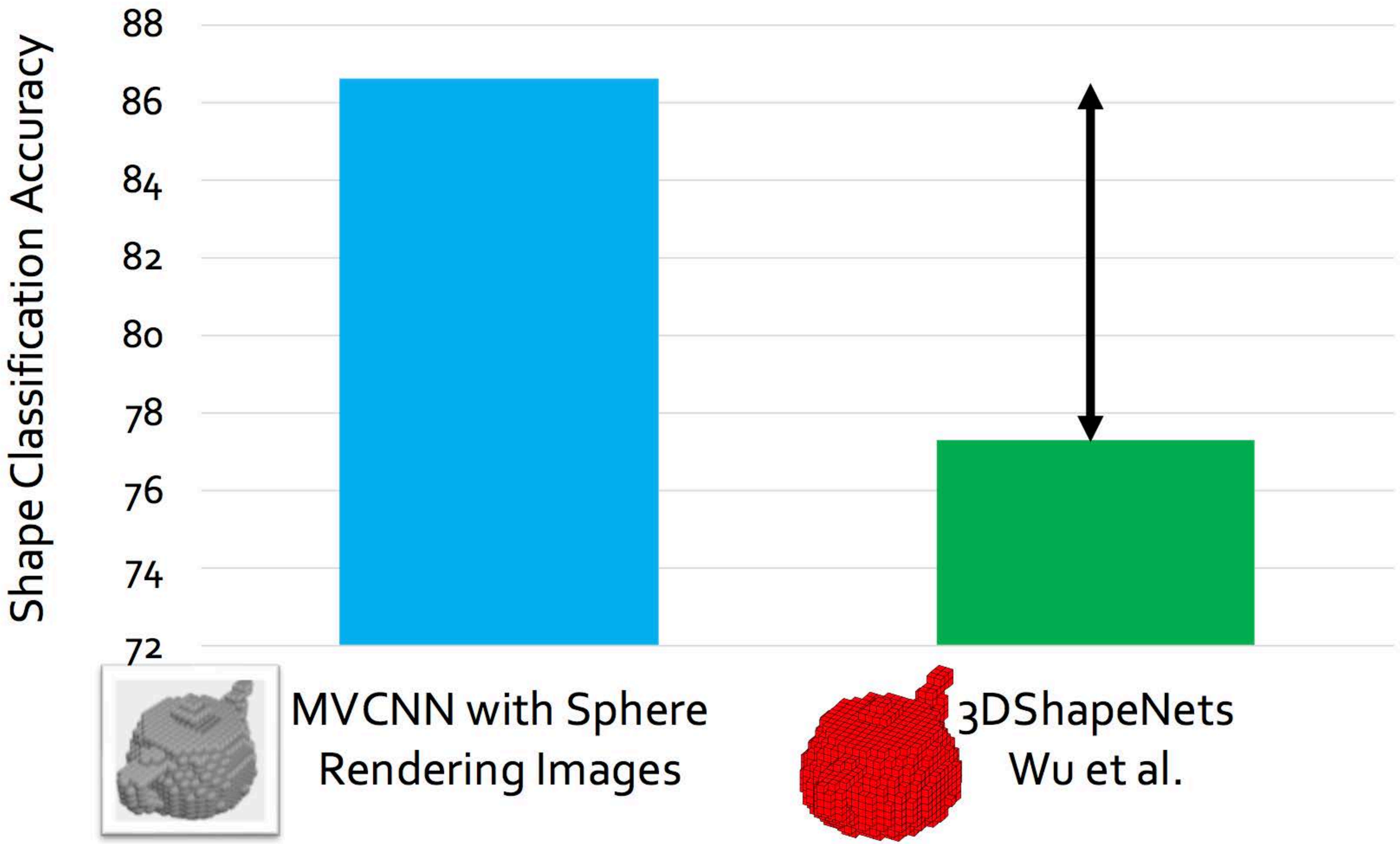
Same
3D Resolution
(30x30x30)

Sphere Rendering Images

Occupancy Grid Volumes

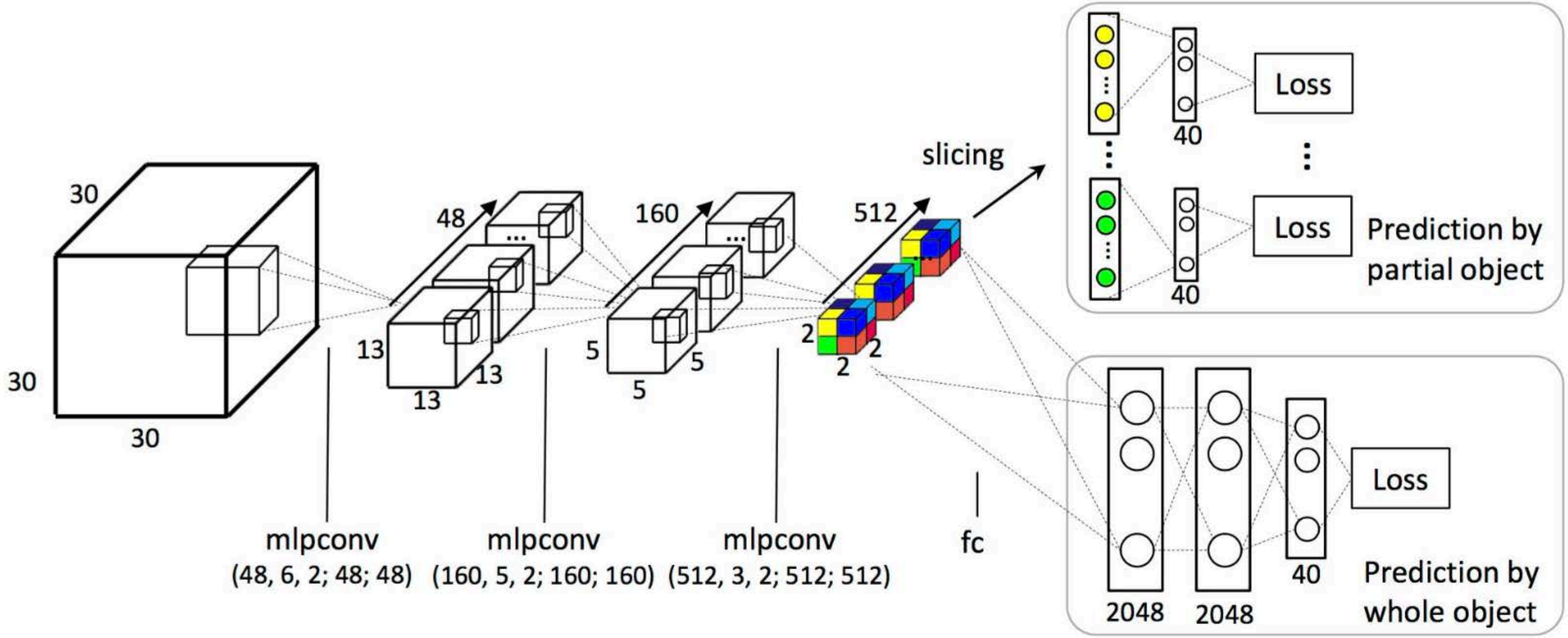
Different Architecture and Same Resolution

Qi et al. 2016



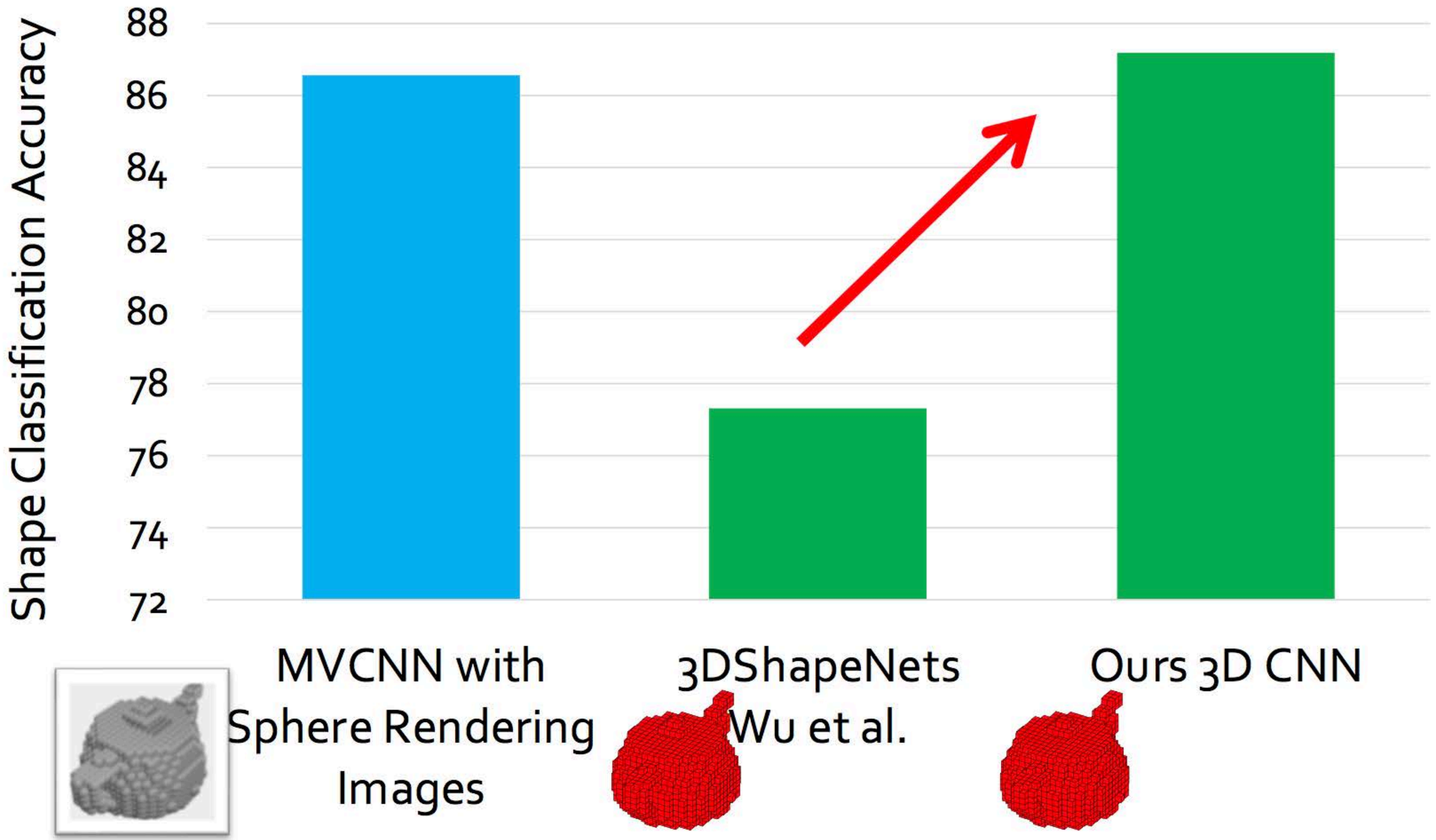
3D CNN with Micro-Neural Network

Qi et al. 2016



3D CNN with Micro-Neural Network

Qi et al. 2016



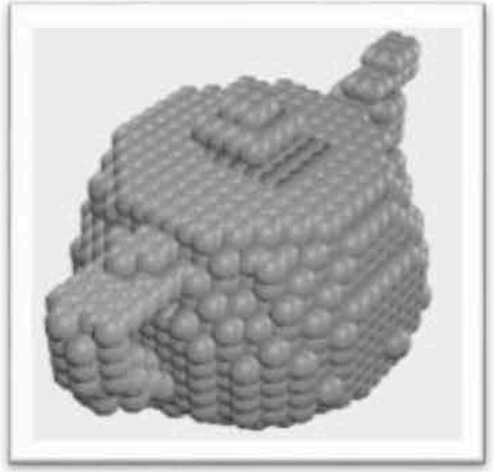
Investigating Resolution

Qi et al. 2016

Multi-View Image CNN

Multi-View Image CNN

Same Architecture



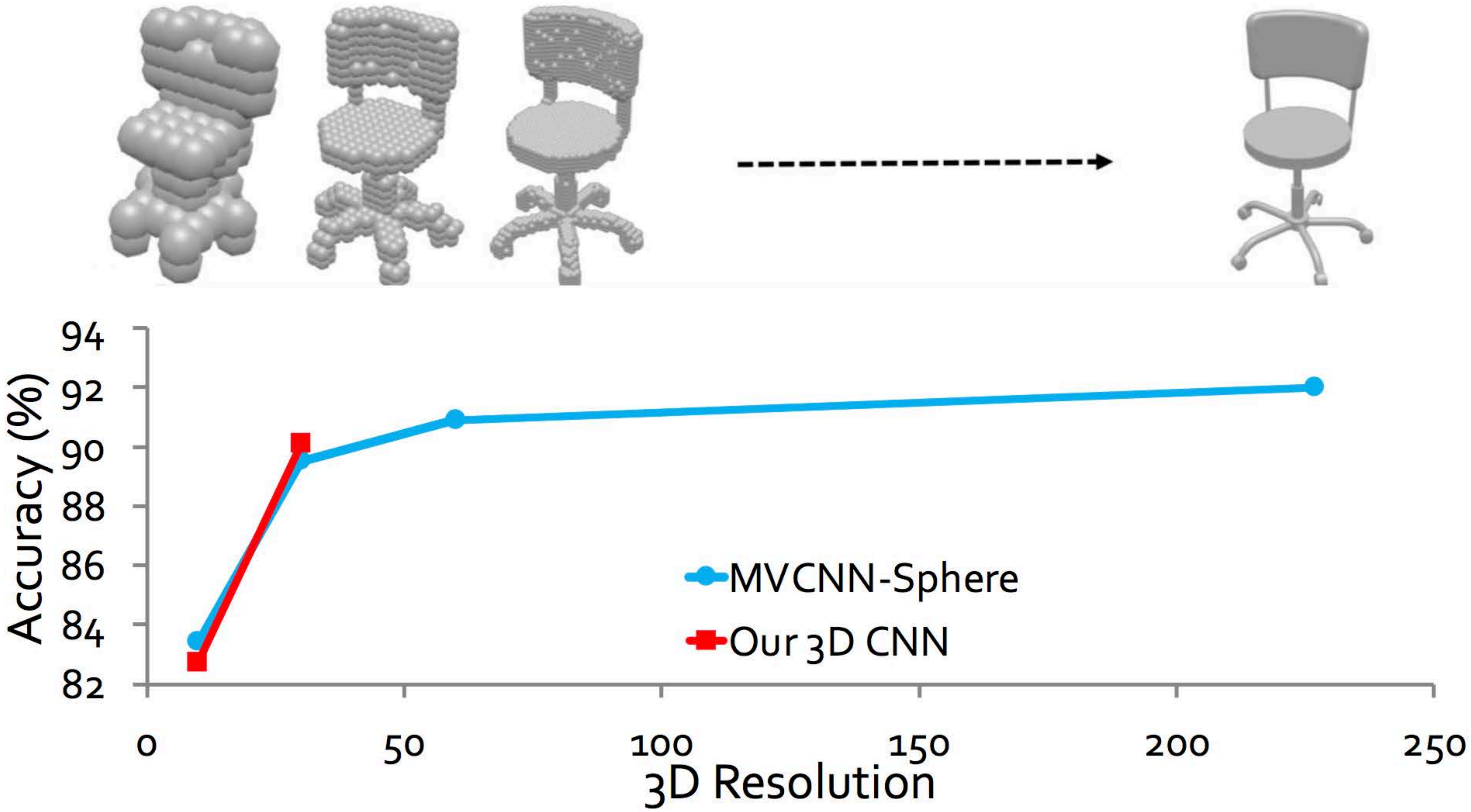
Standard Rendering Images

Sphere Rendering Images

Different 3D Resolution

Investigating Resolution

Qi et al. 2016

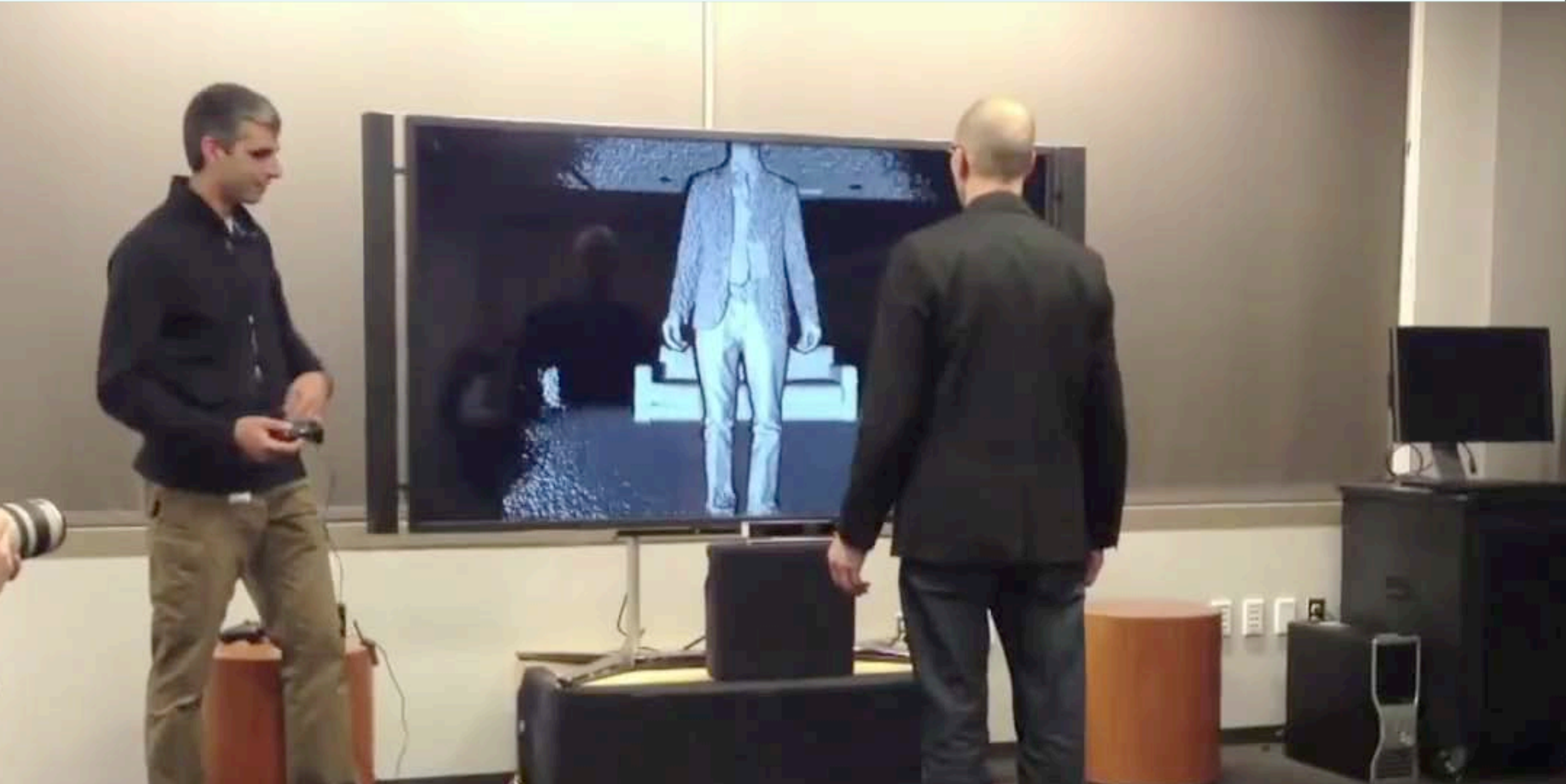


Application

Dense Correspondences of Clothed Humans

3D Human Capture

Microsoft 2013

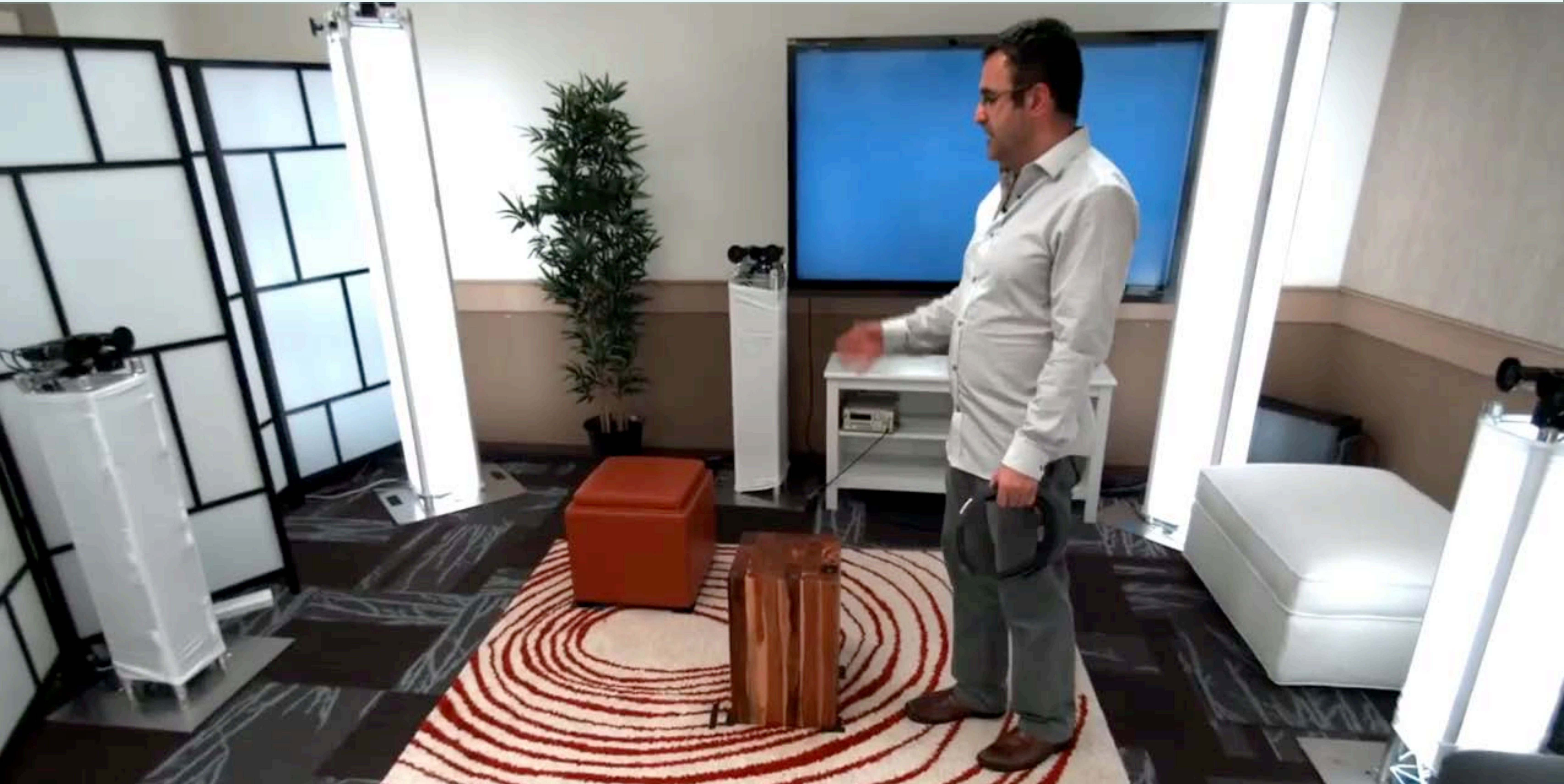


3D Human Capture

Microsoft 2015

3D Human Capture

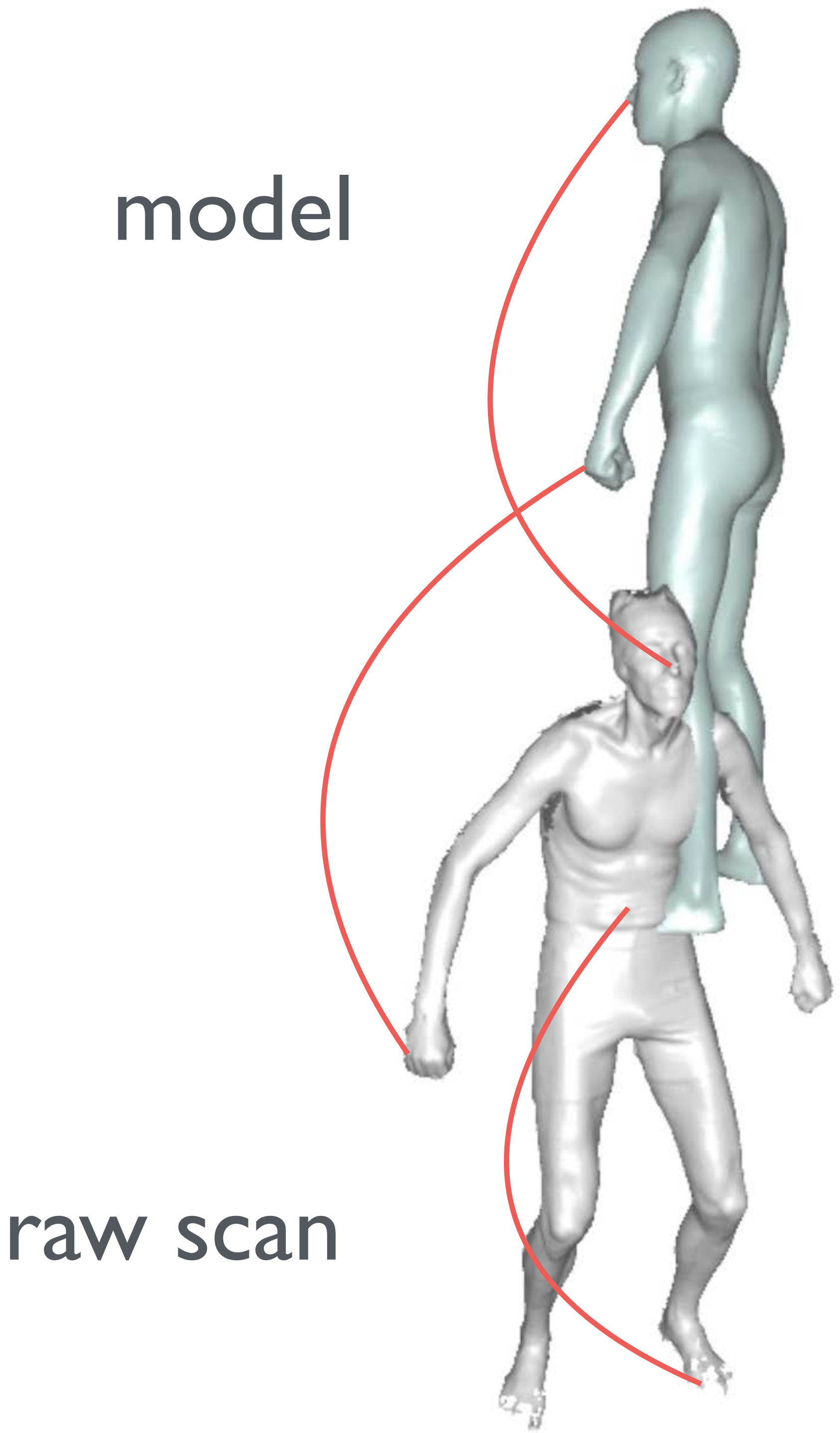
[Dou et al. ’16]

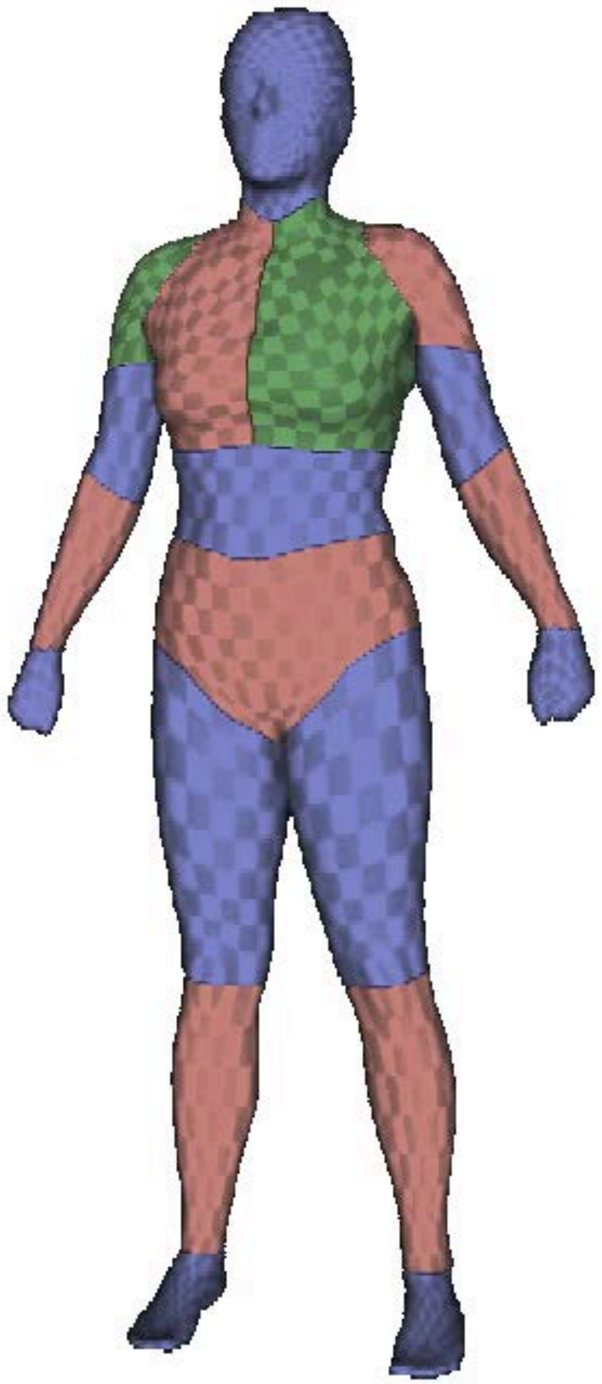


Analysis & Reasoning

Correspondences on Clothed Human Bodies

Shape Analysis

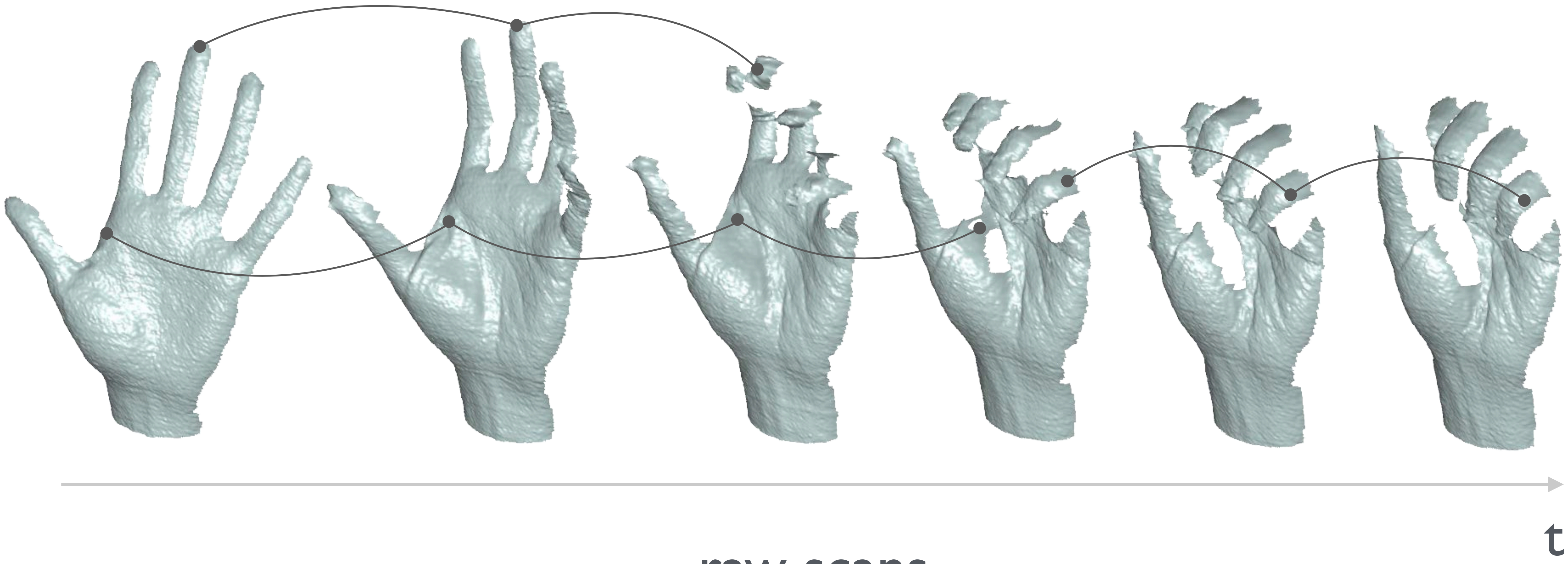




human body
body pose
gender
BMI
...

SCAPE model of Lee from Hirshberg et al. 2012

Motion Understanding



raw scans

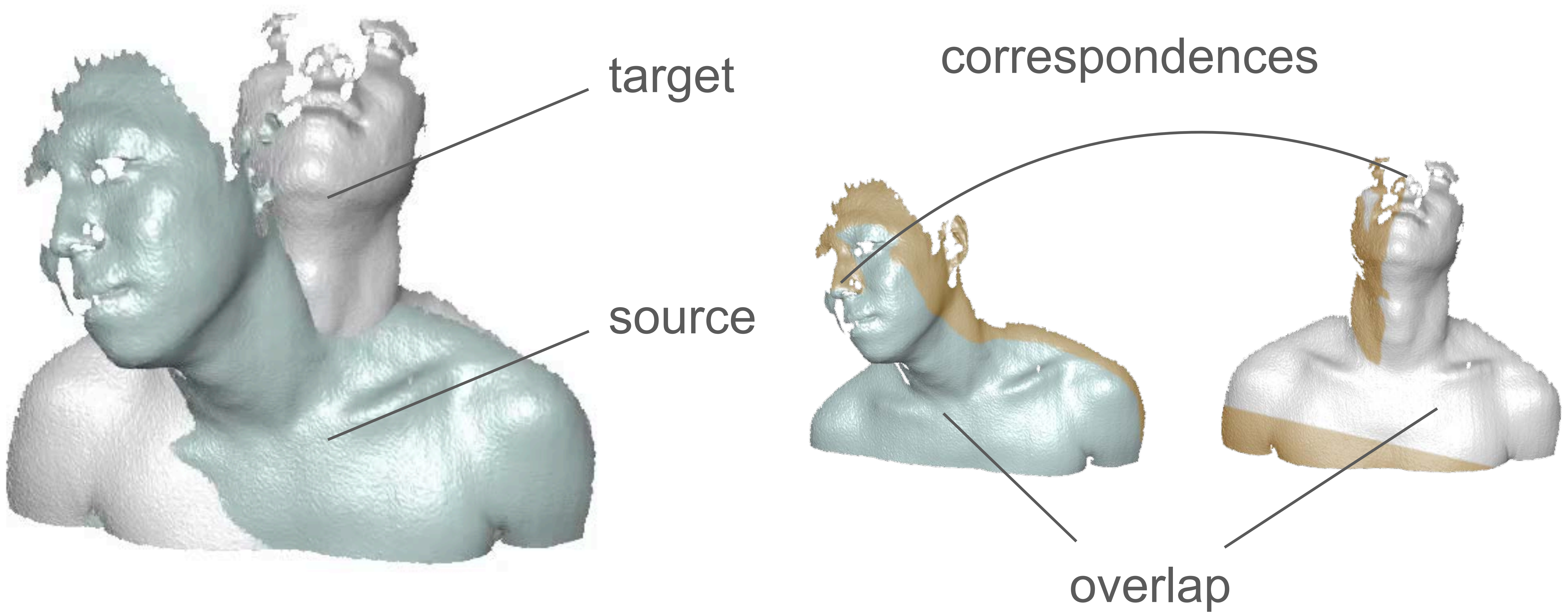
Motion Understanding

raw scans

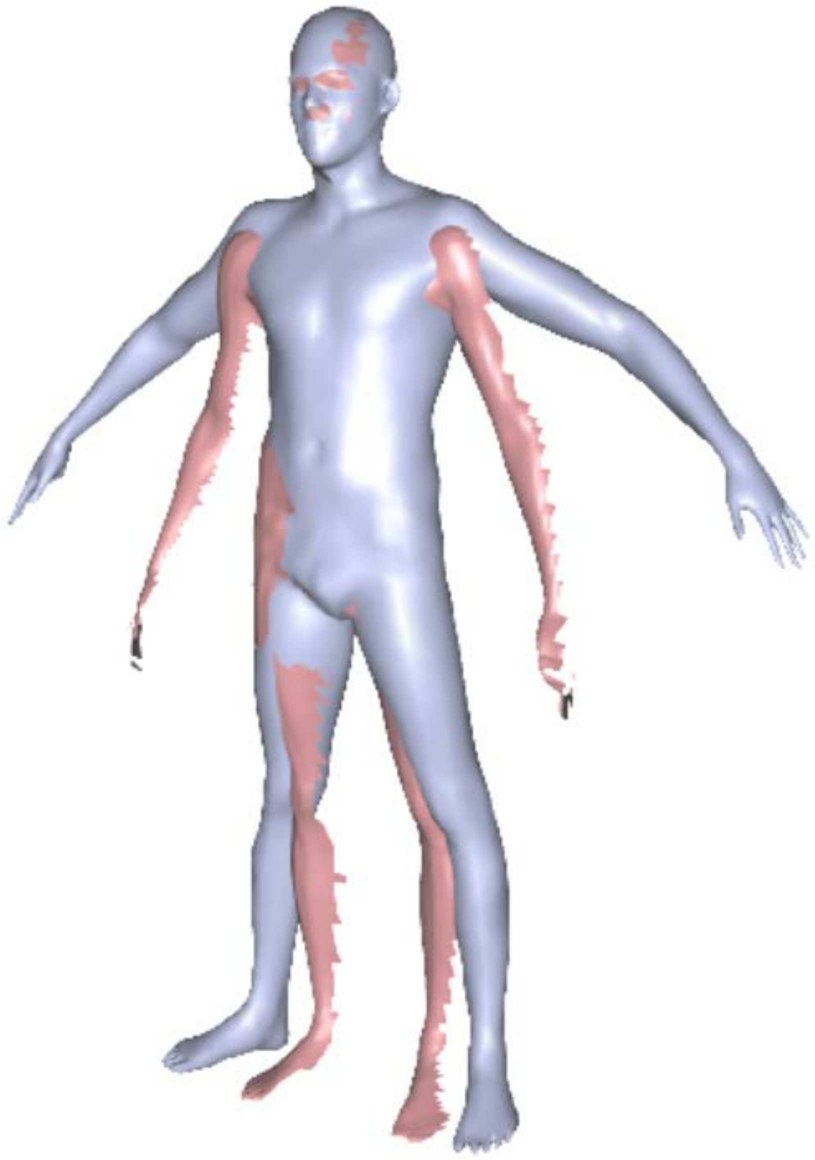
“grasping”

Correspondences?

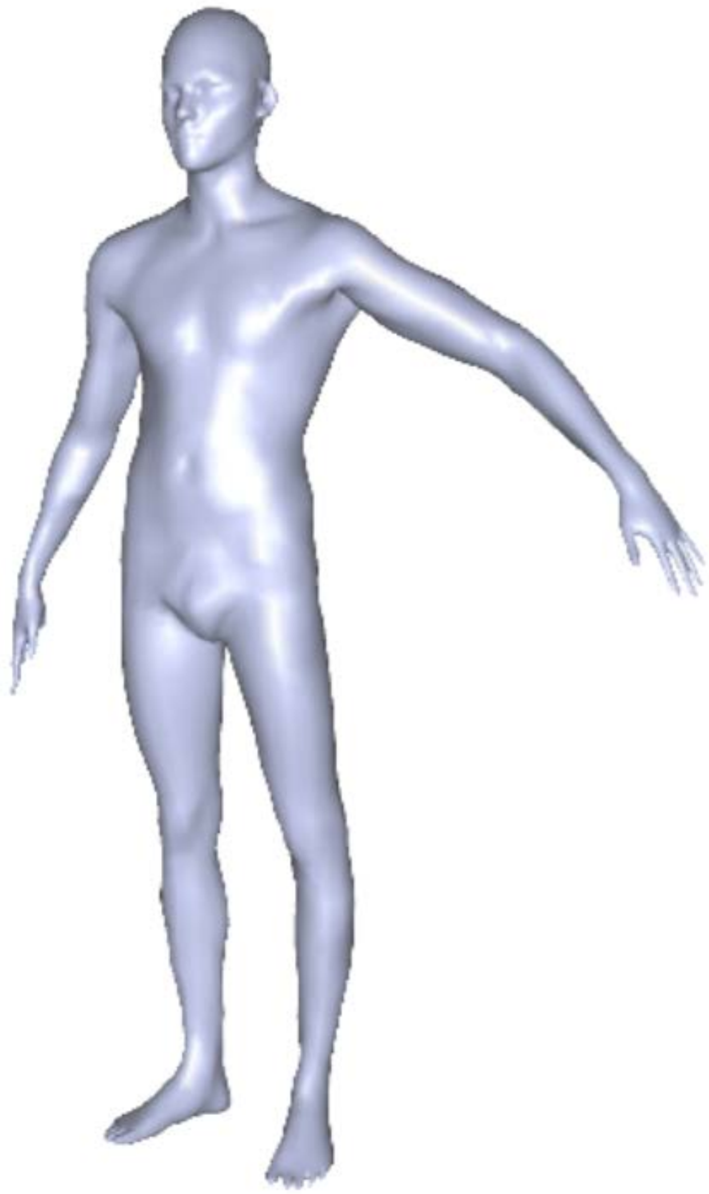
Non-Rigid Registration [Li et al. 2008]



Large Pose Changes



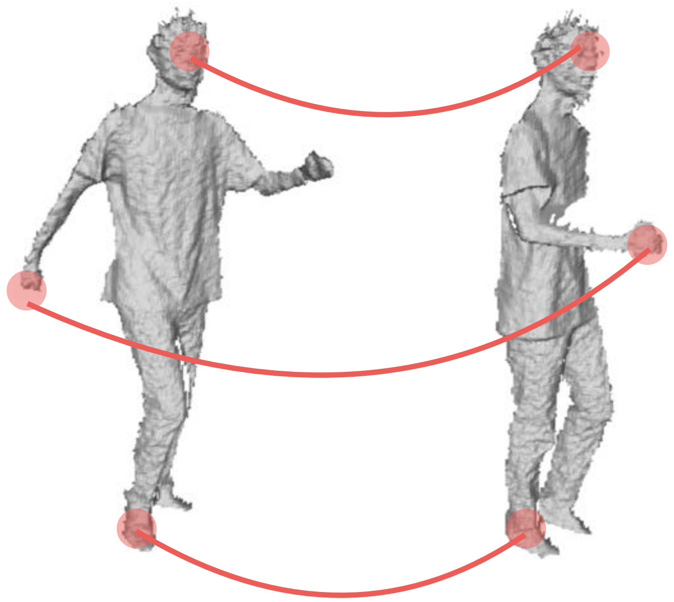
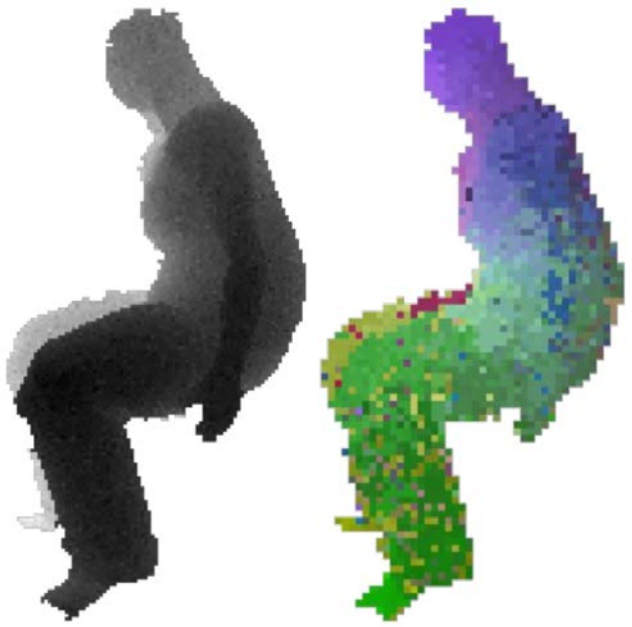
source & target



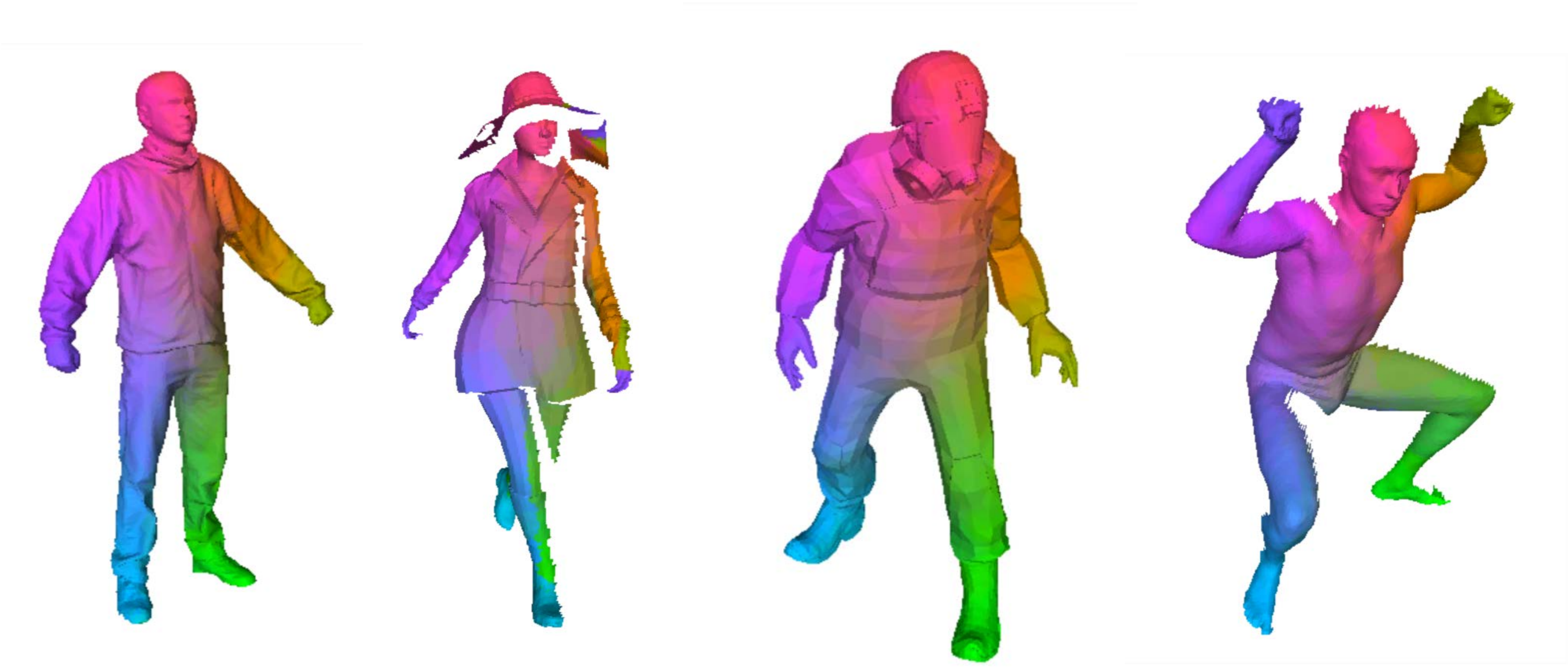
[Li et al. 09]

[Huang et al. 08]

Descriptors

	designed descriptor	learned descriptor
partial scans	[Hebert 99] [Bronstein et al. 06] … 	[Taylor et al. 12] [Pons-Moll et al. 15] … 
complete model (or small holes)	[Jain & Zhang 06] [Bronstein et al. 10] [Kim et al. 11] [Windheuser et al. 14] [Chen & Koltun 15] … 	[Litman & Bronstein 14] [Rodola et al. 14] [Windheuser et al. 14] [Macsi et al. 15] … 

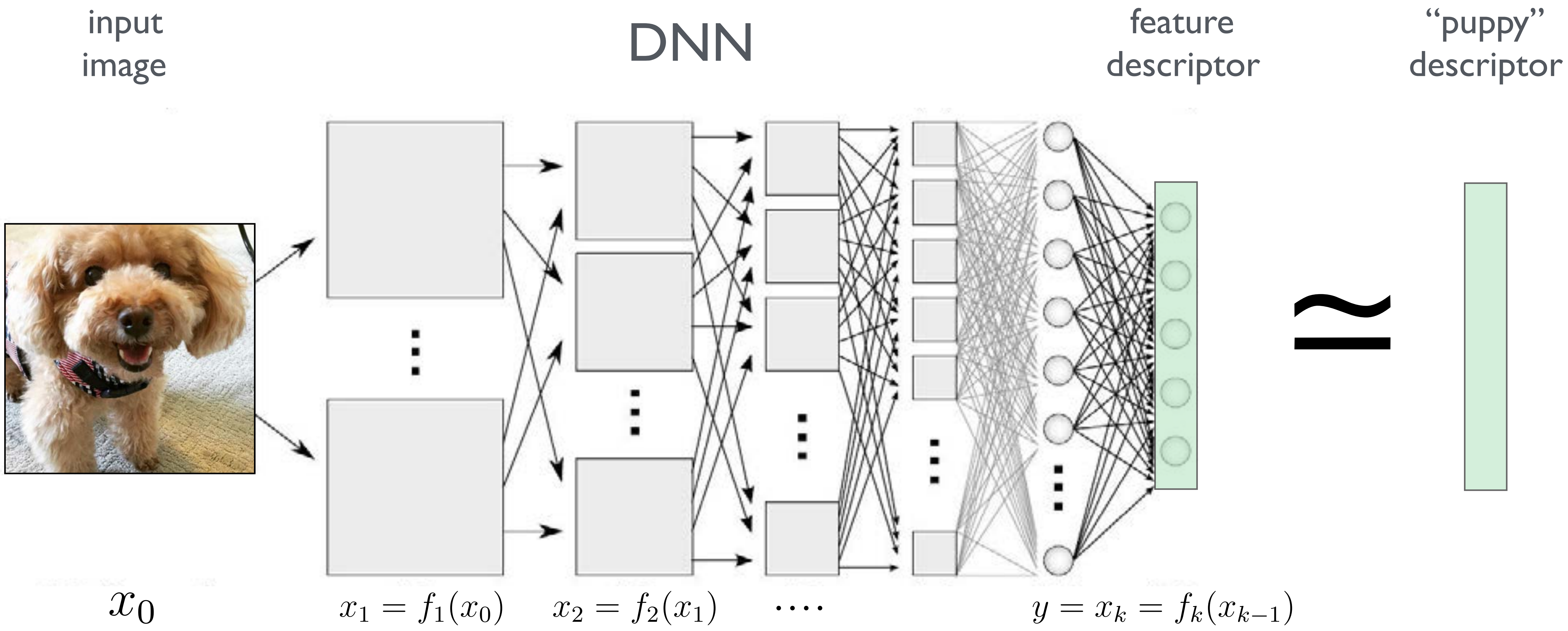
Clothed and Partial Data



immense space of variations

Classification Networks

Deep Convolutional Neural Network



classification network, e.g. AlexNet [Krizhevsky et al. 2012]

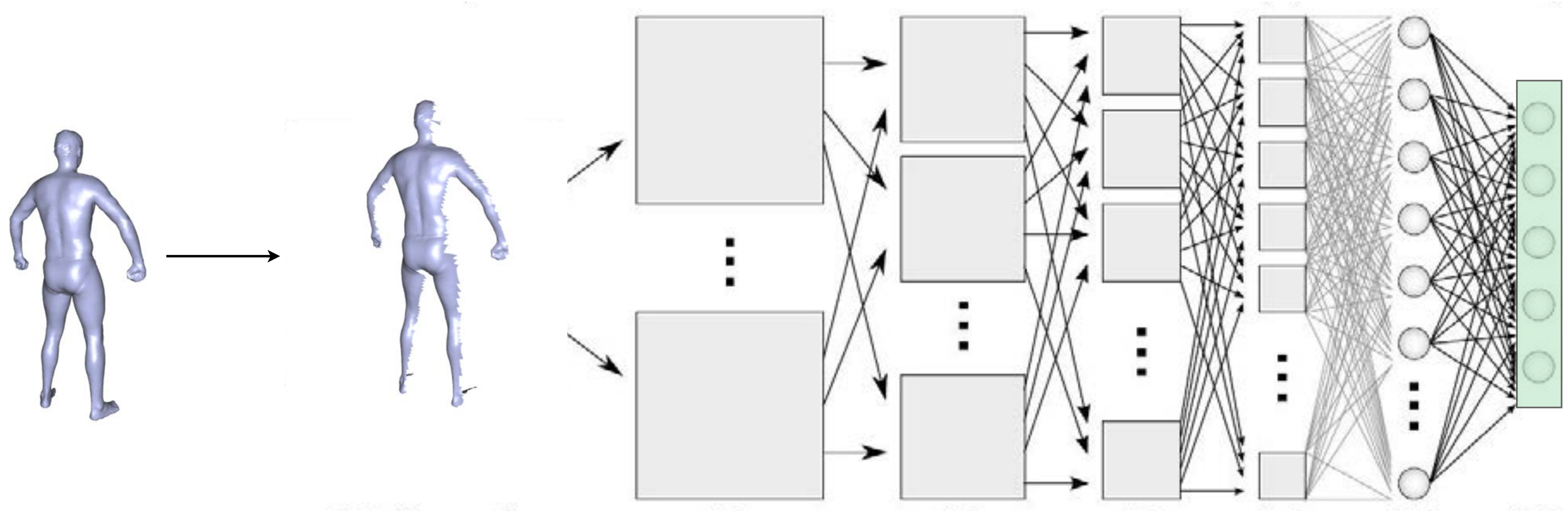
Deep Convolutional Neural Network

3D model

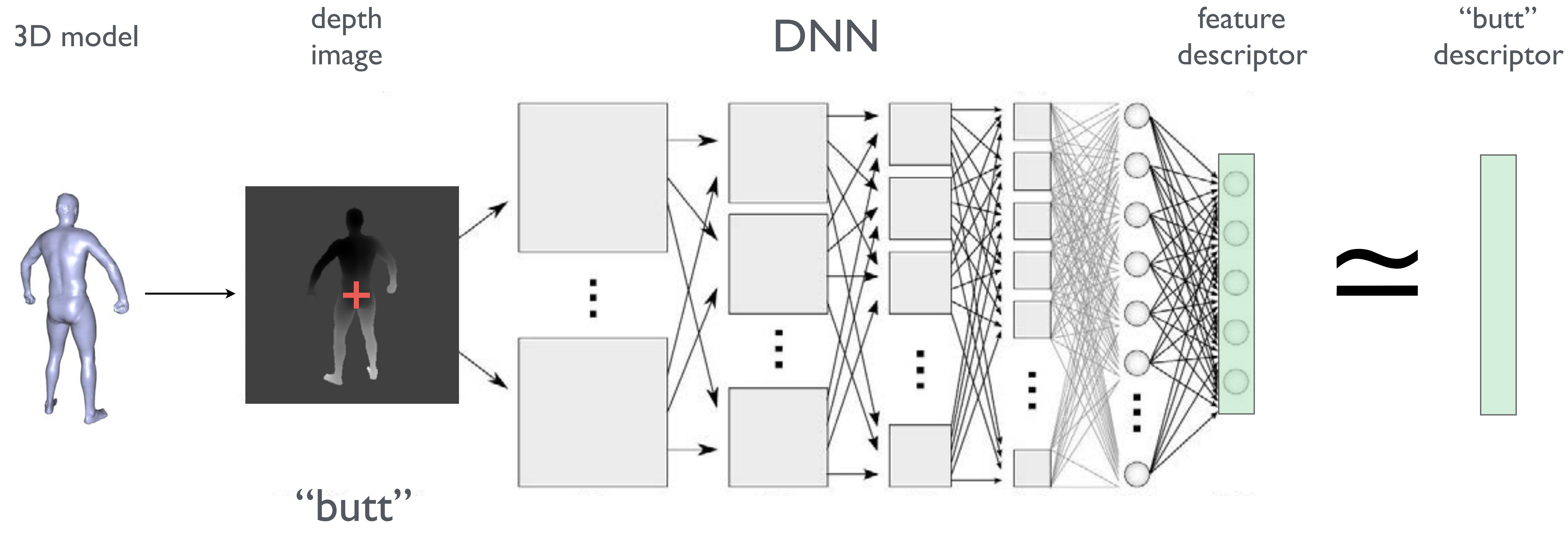
depth
image

DNN

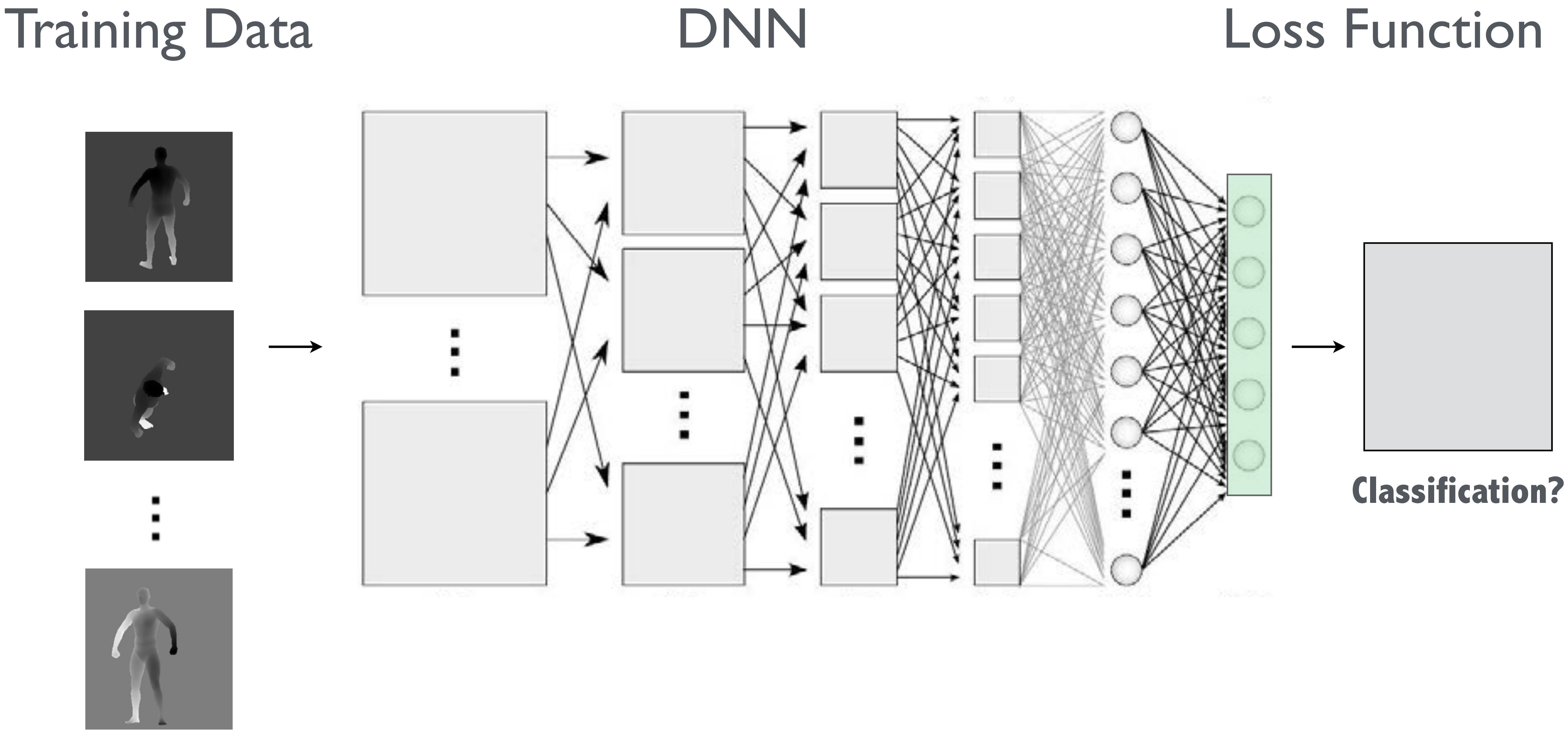
feature
descriptor



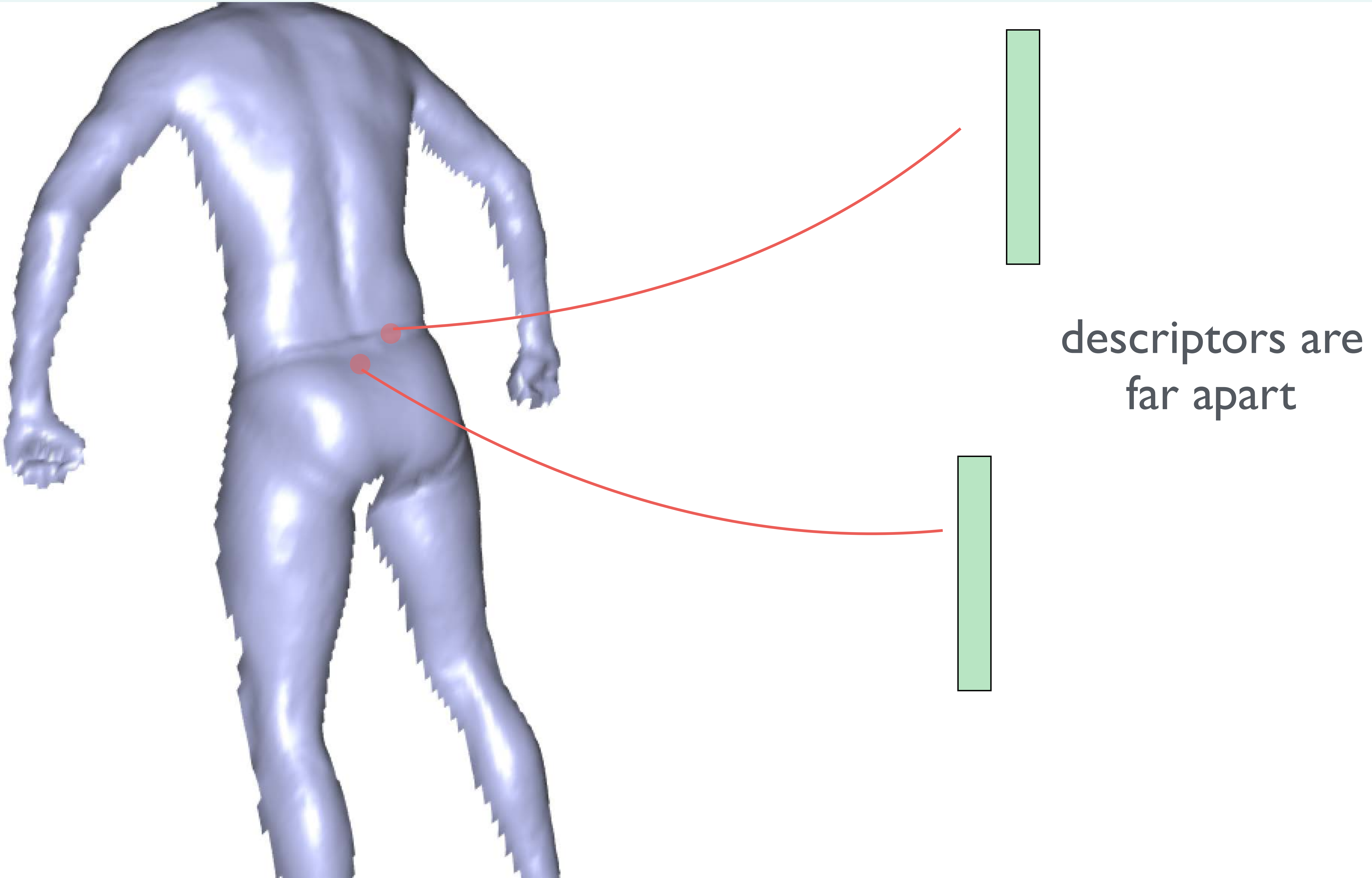
Deep Convolutional Neural Network



Loss Function

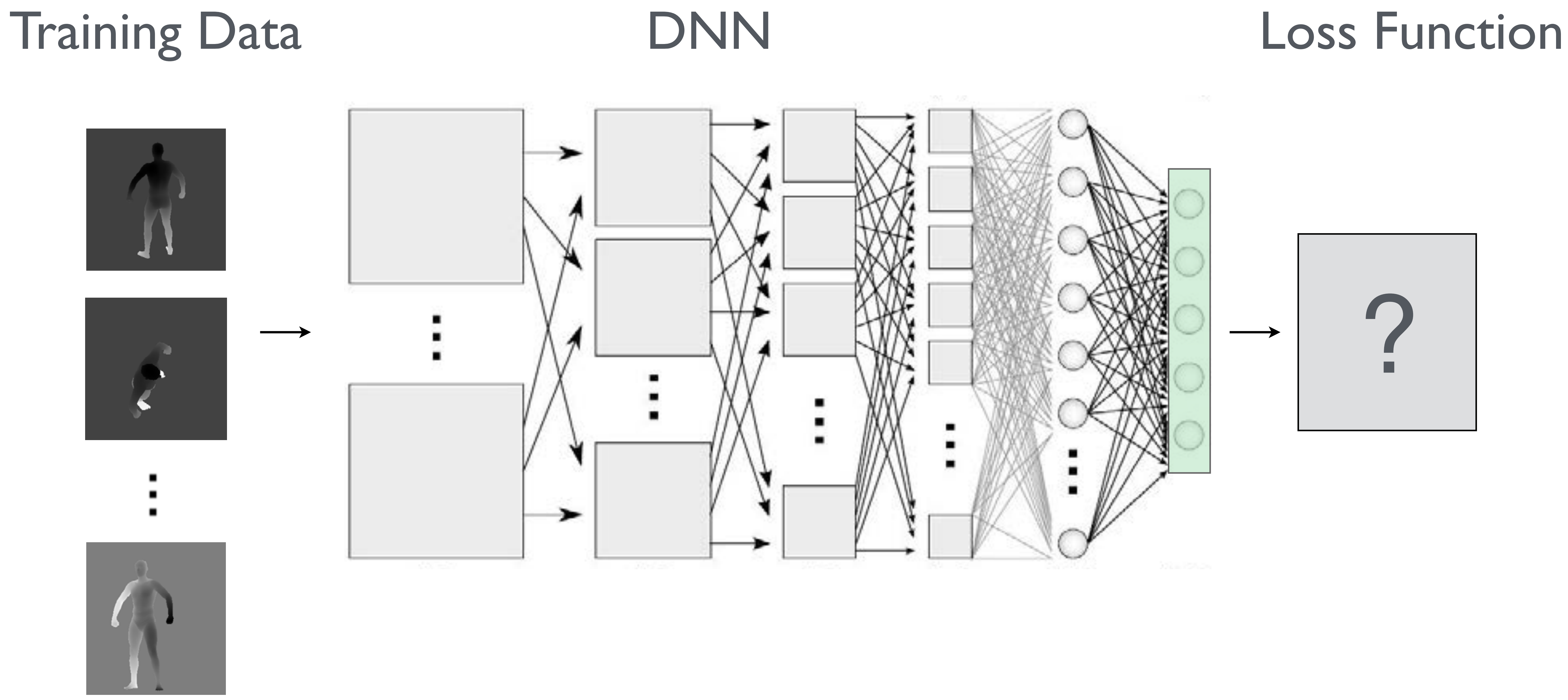


Classification Task

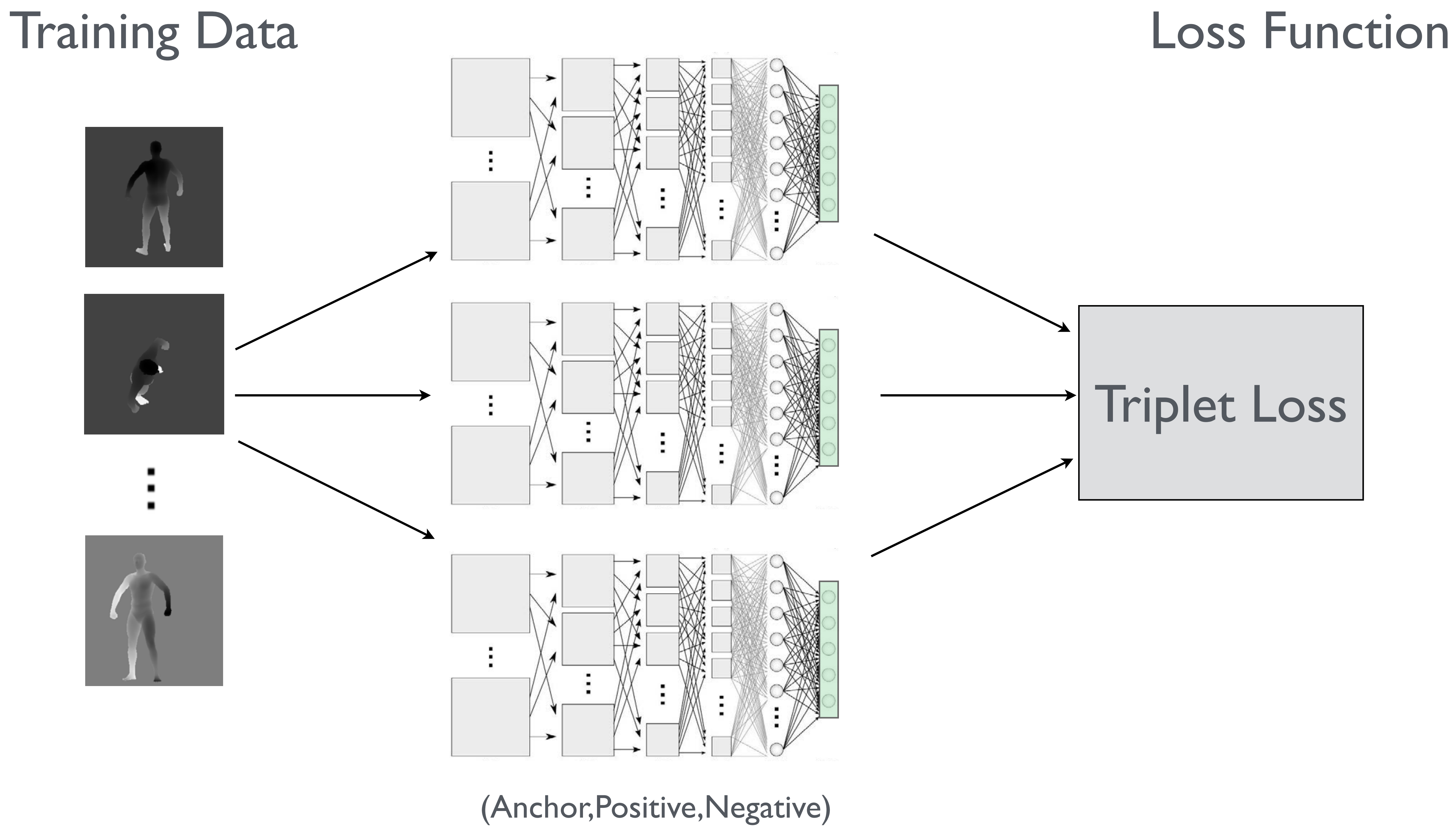


How to preserve distances?

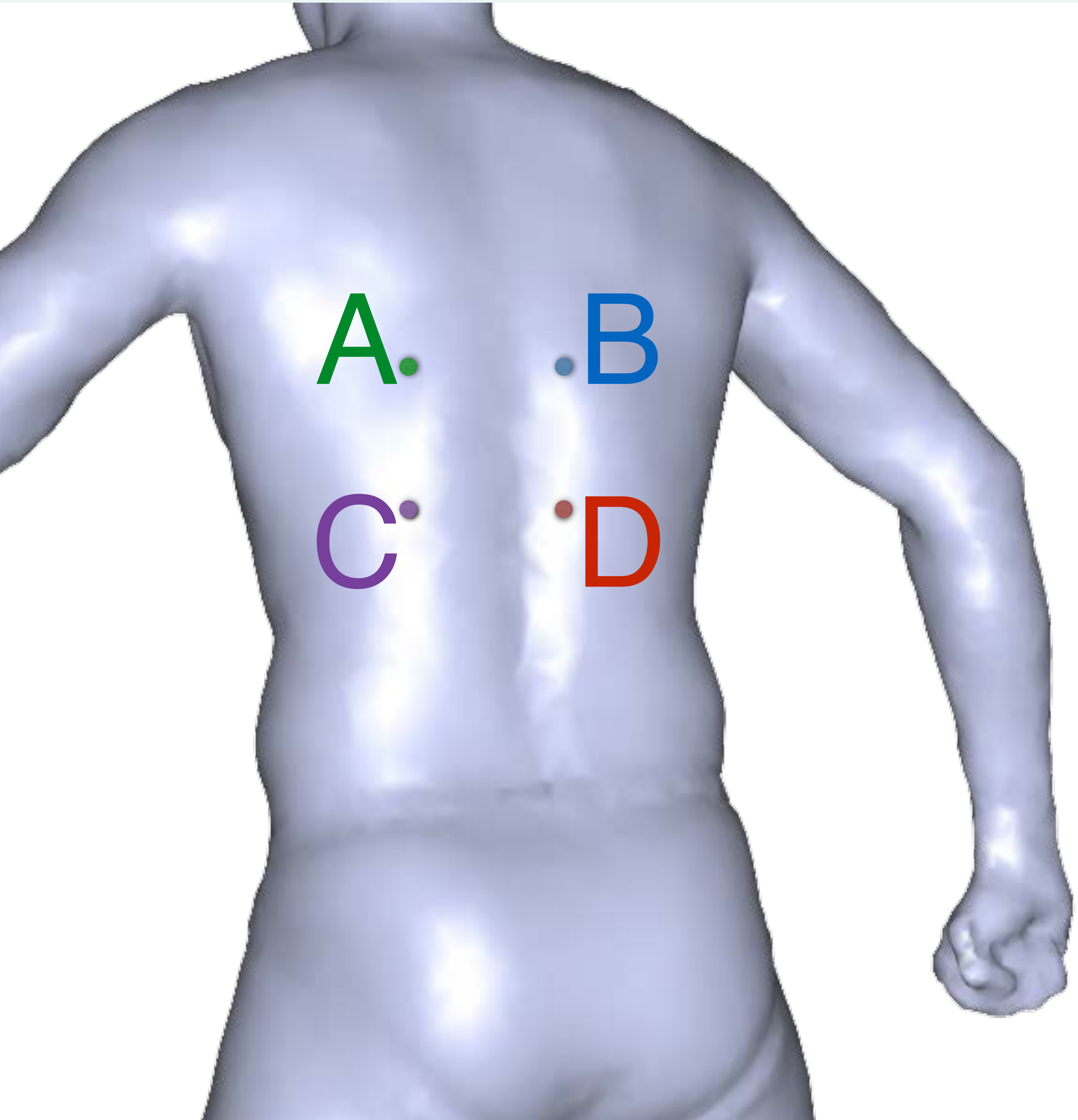
Deep Convolutional Neural Network

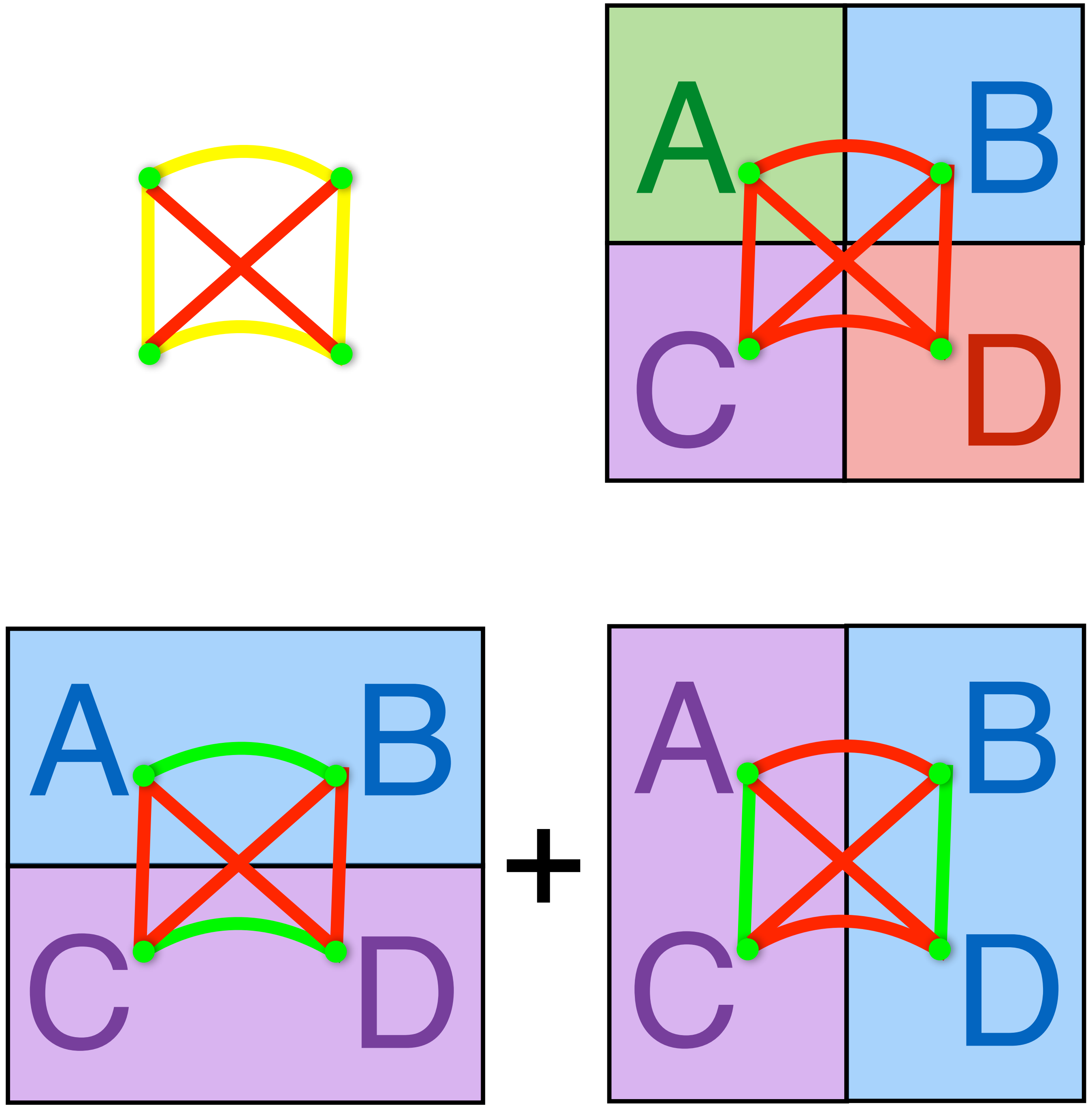


Loss Function

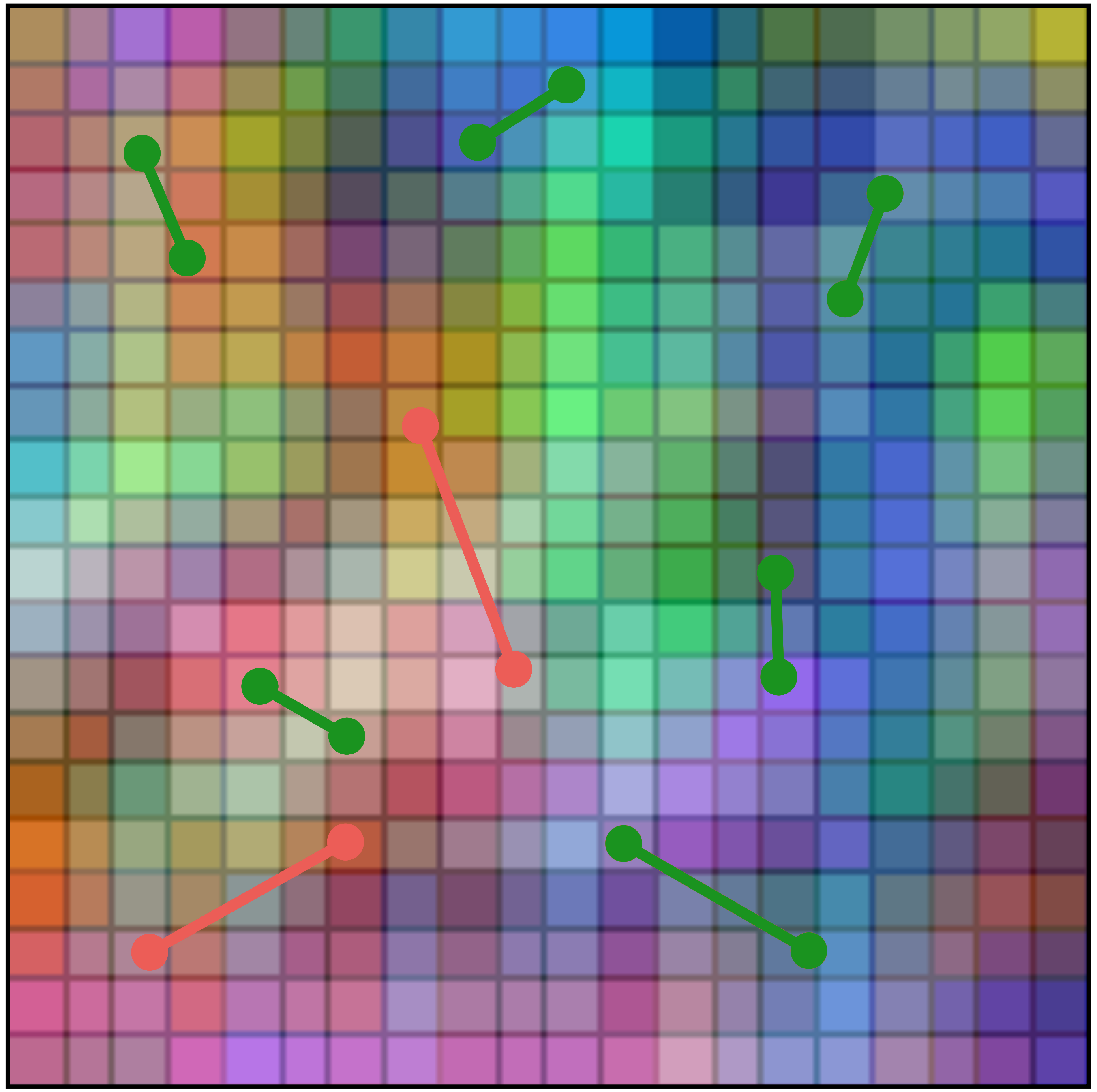


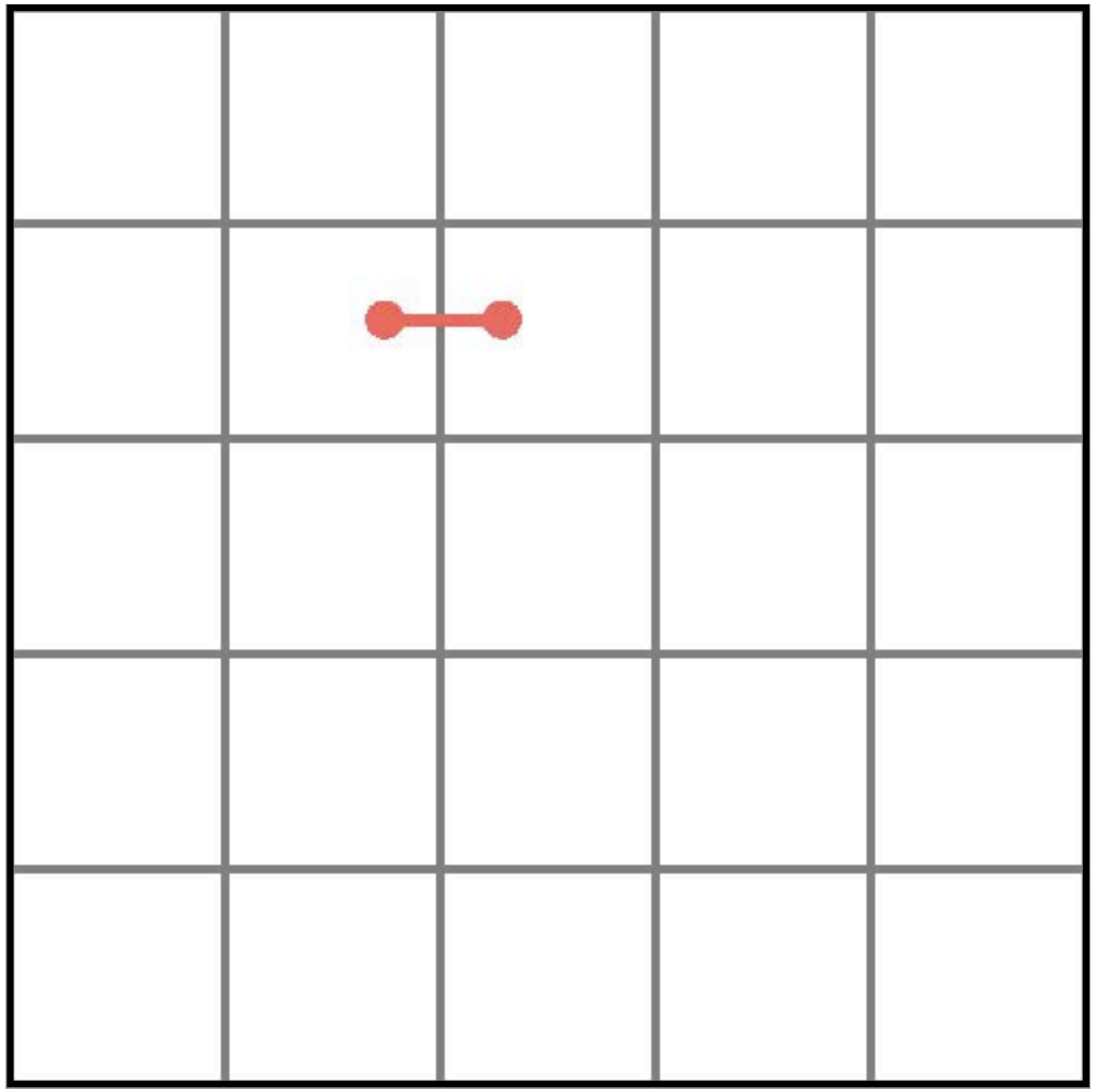
Multi-Segmentation



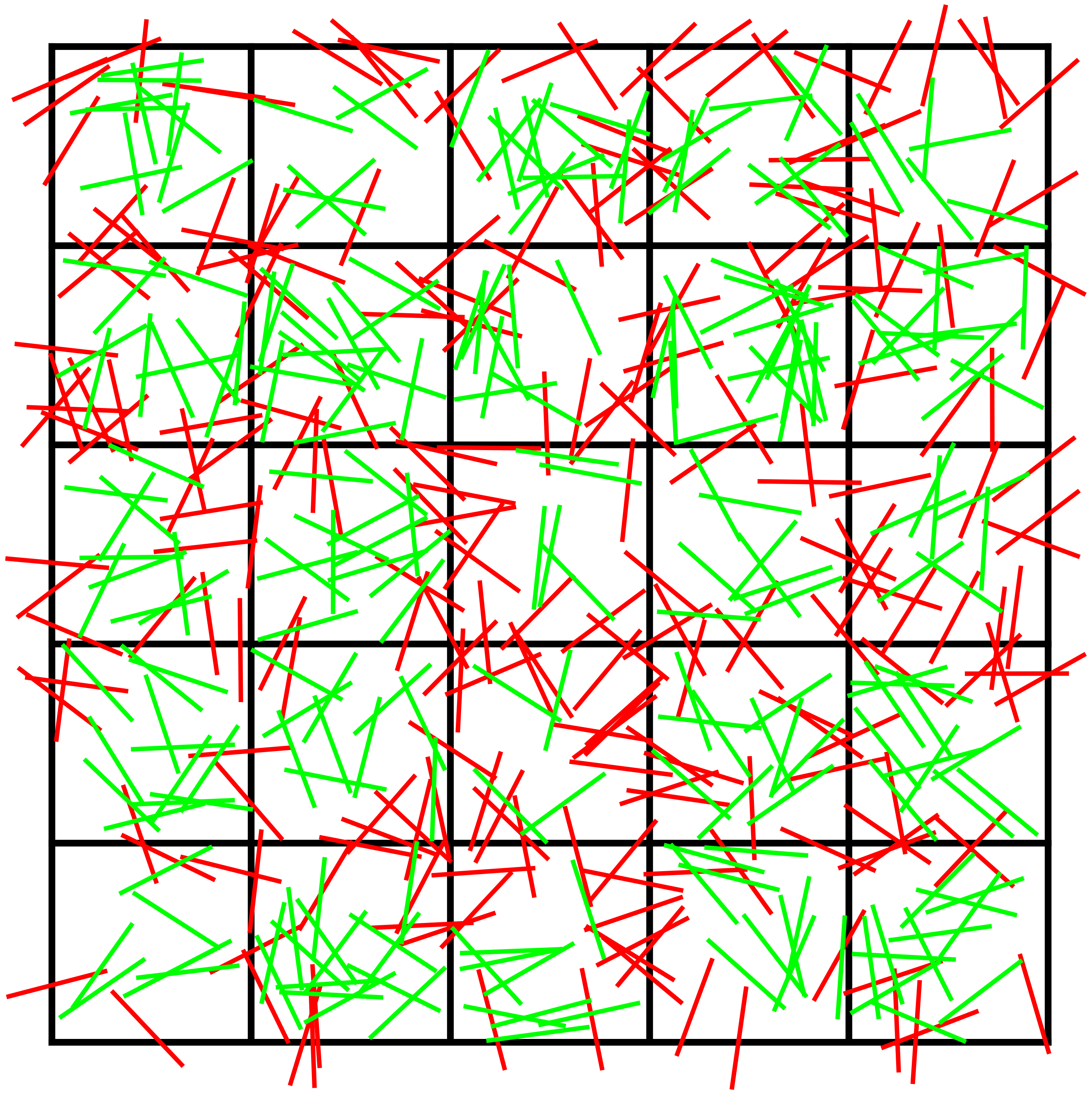


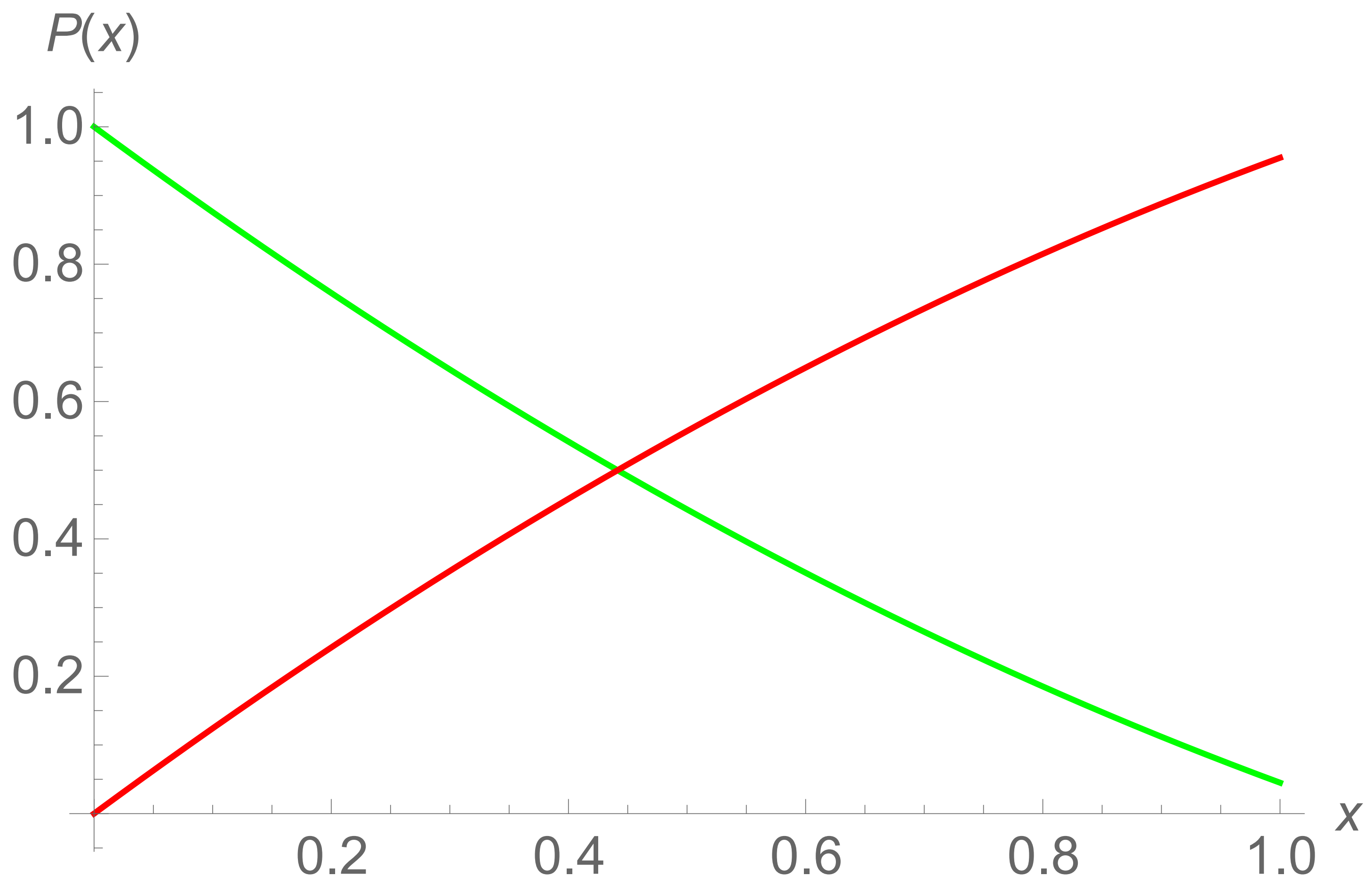
Multiple Segmentation





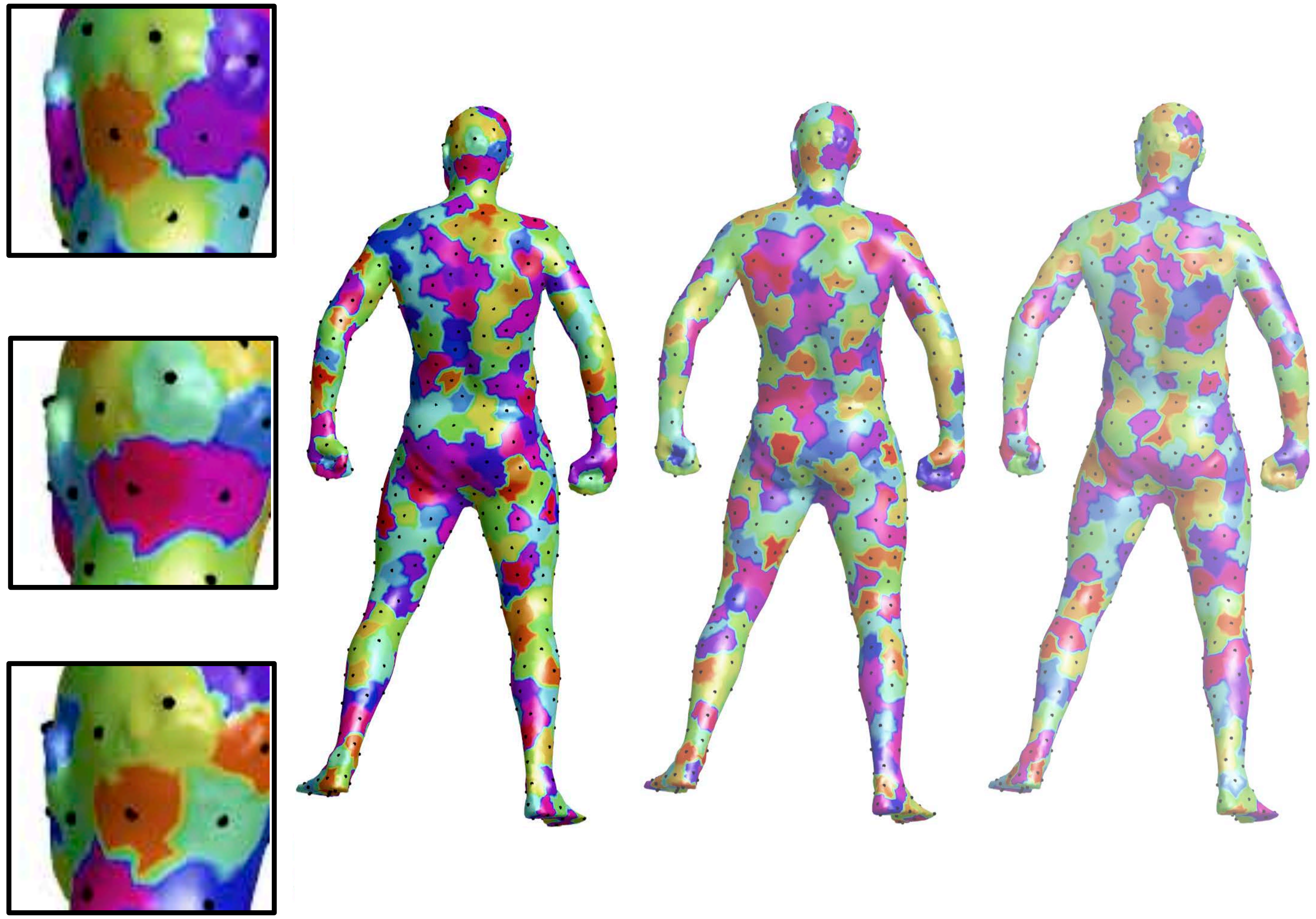
Buffon-Laplace Needle Problem (18th Century)





Distance Preserving Learning

500 classes



⋮

100 random segmentations

Distance Preserving Learning

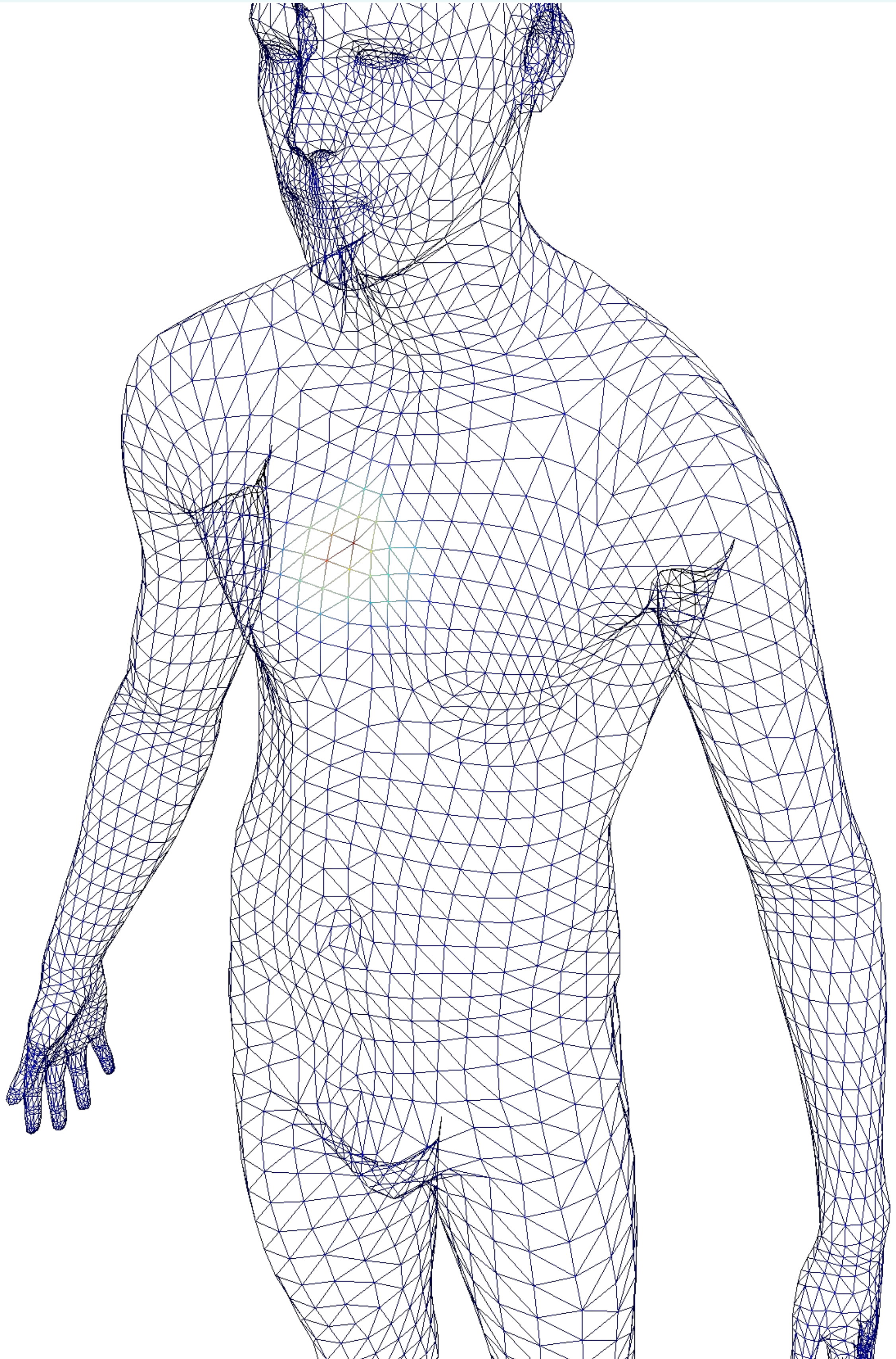
AlexNet

image → DNN → descriptor → Classification ⋮ Classification ⋮ Classification

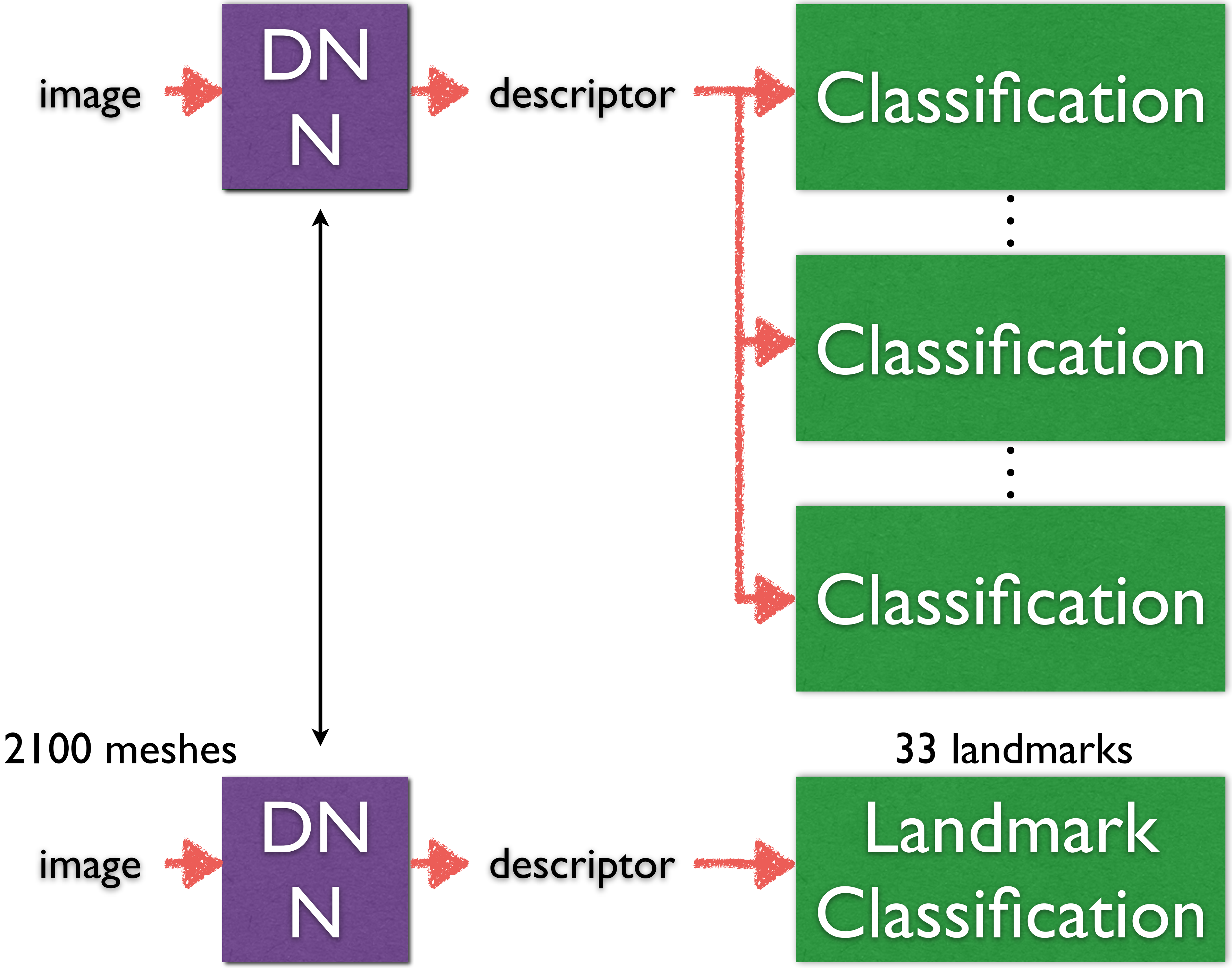
1x512x512

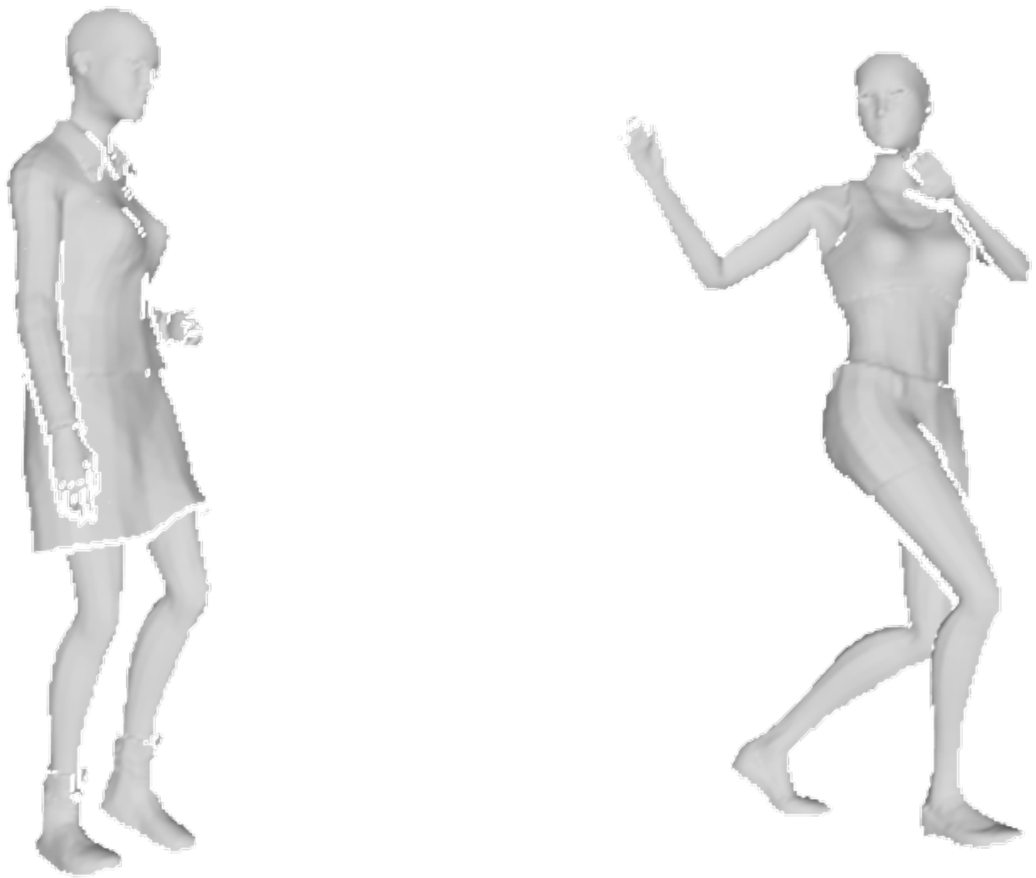
16x512x512

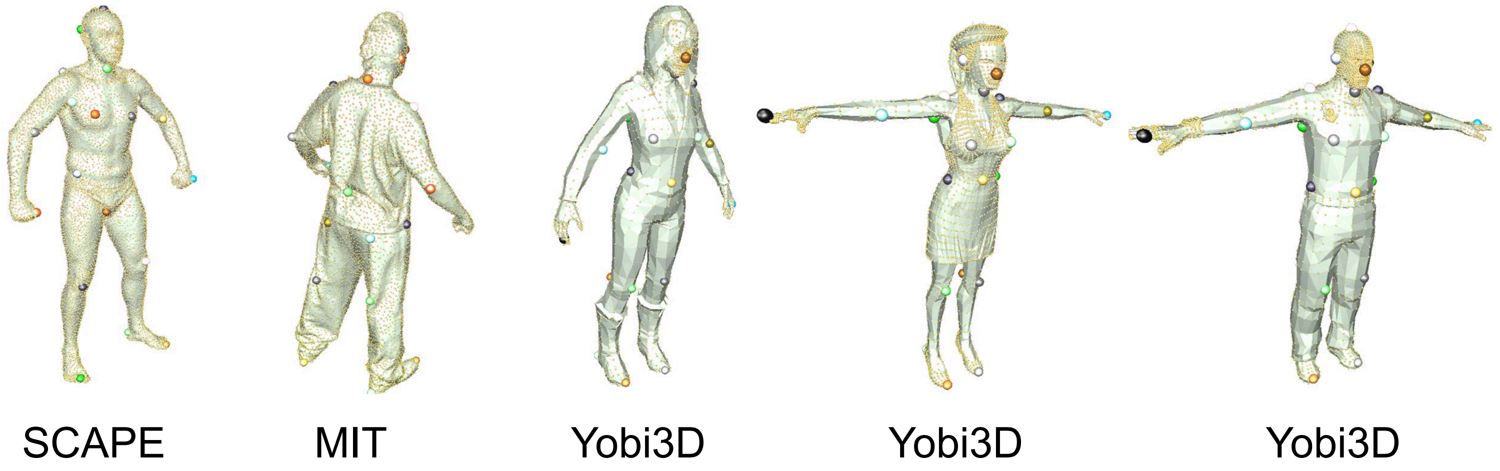
100 segmentations
500 classes



Variation on Clothing



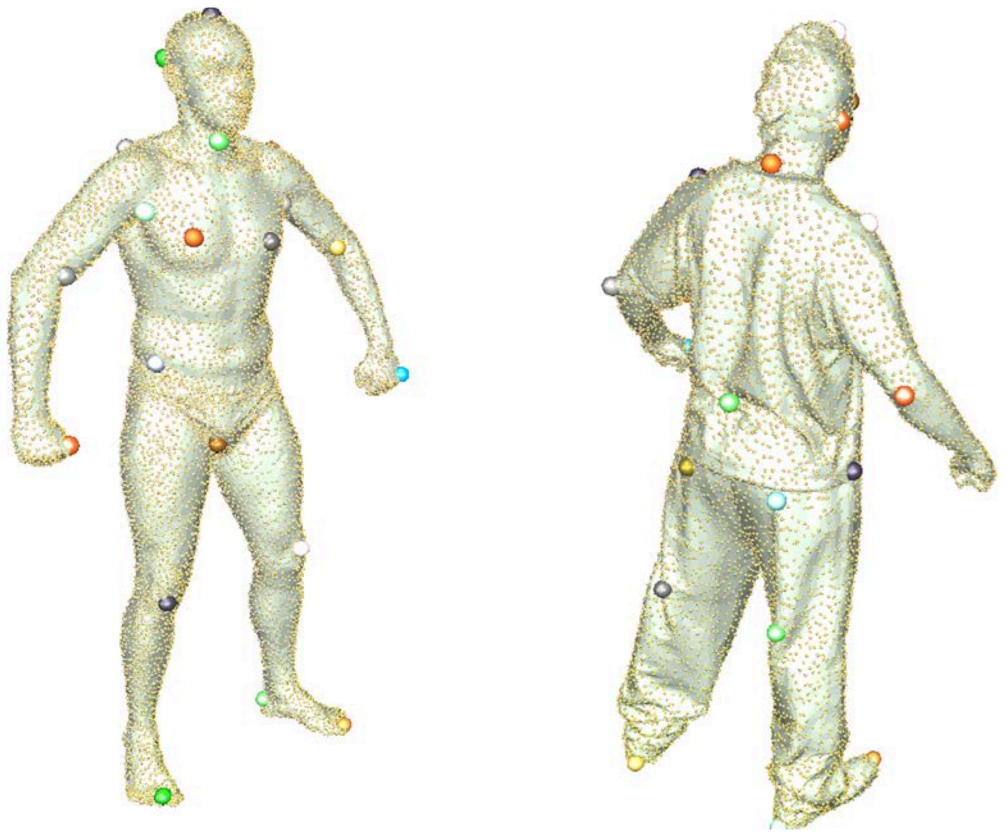


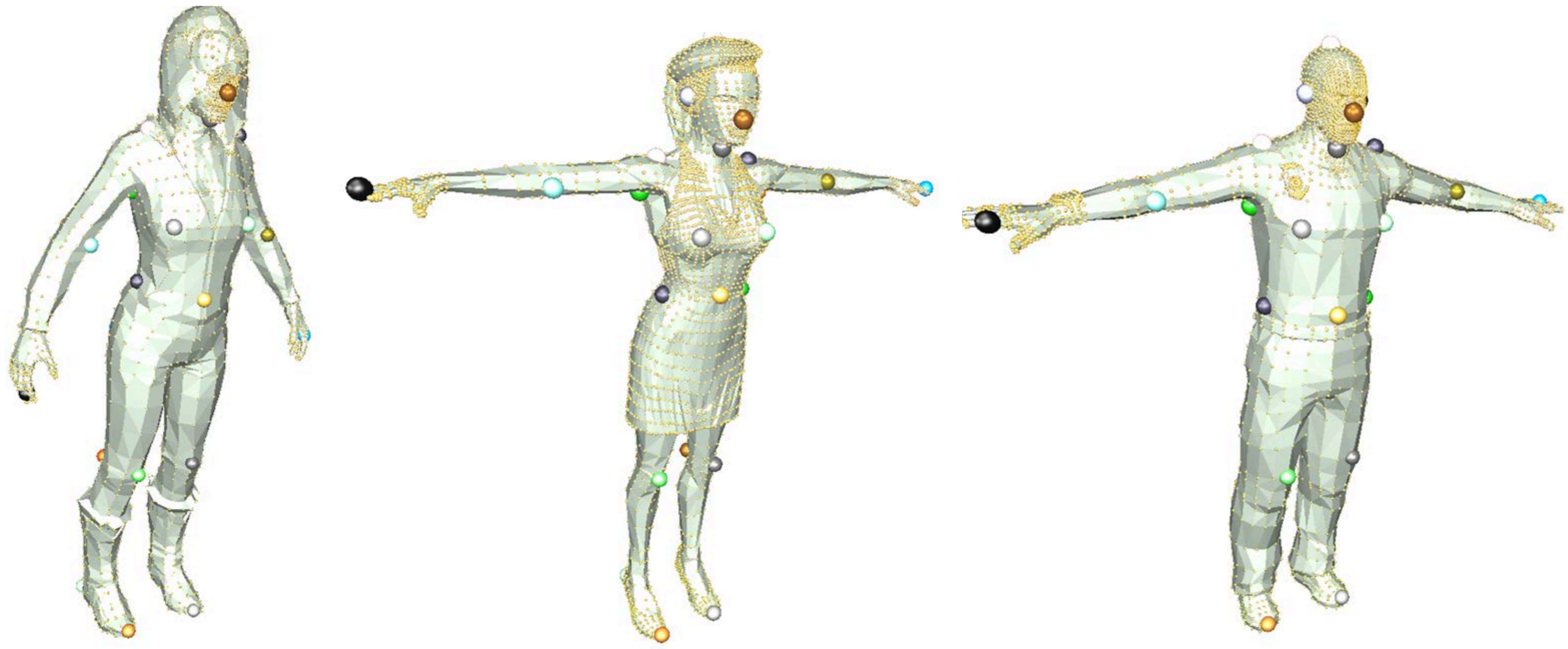


Training Data

Shape & Pose

Clothing

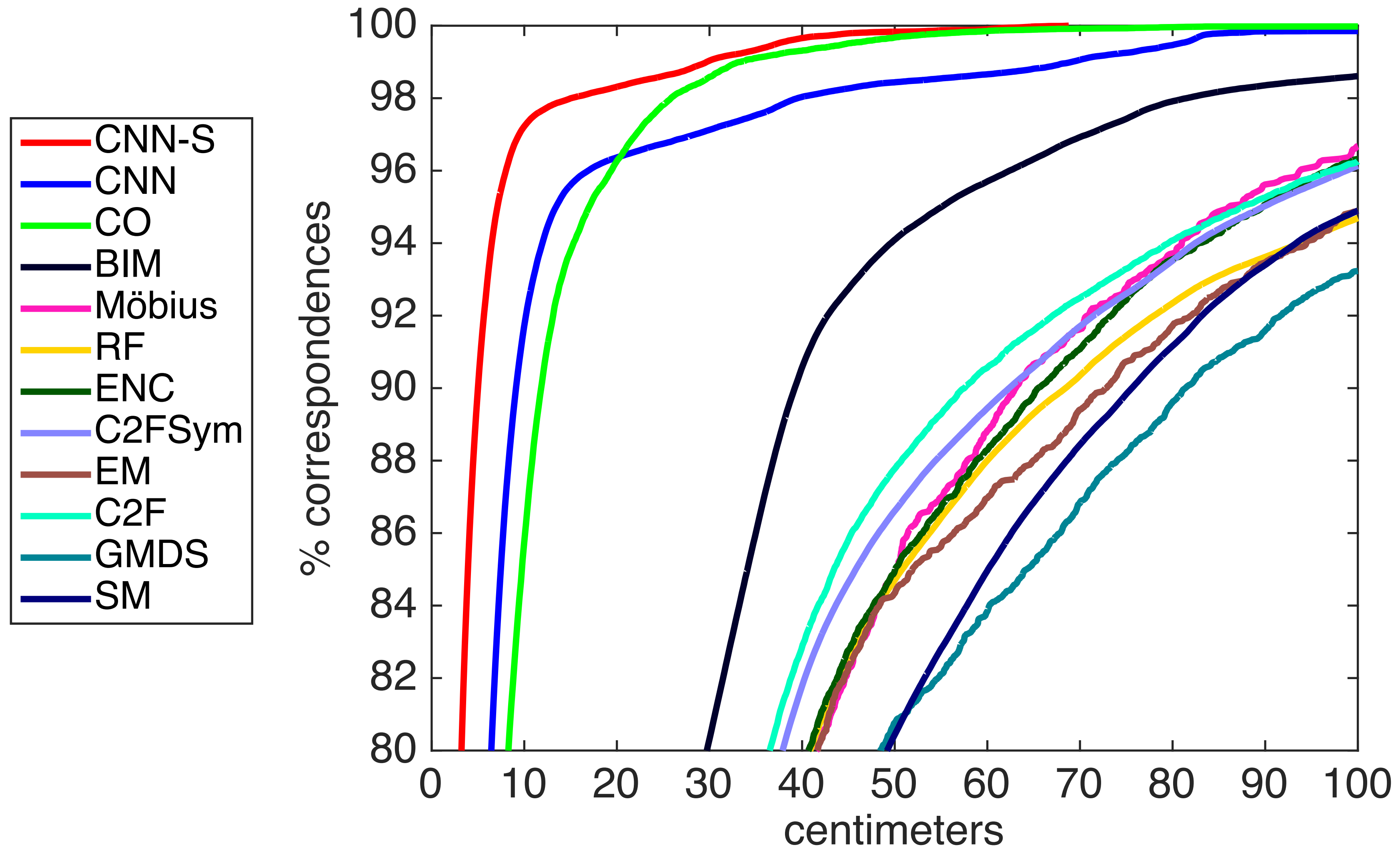




SCAPE MIT

Yobi3D Yobi3D Yobi3D

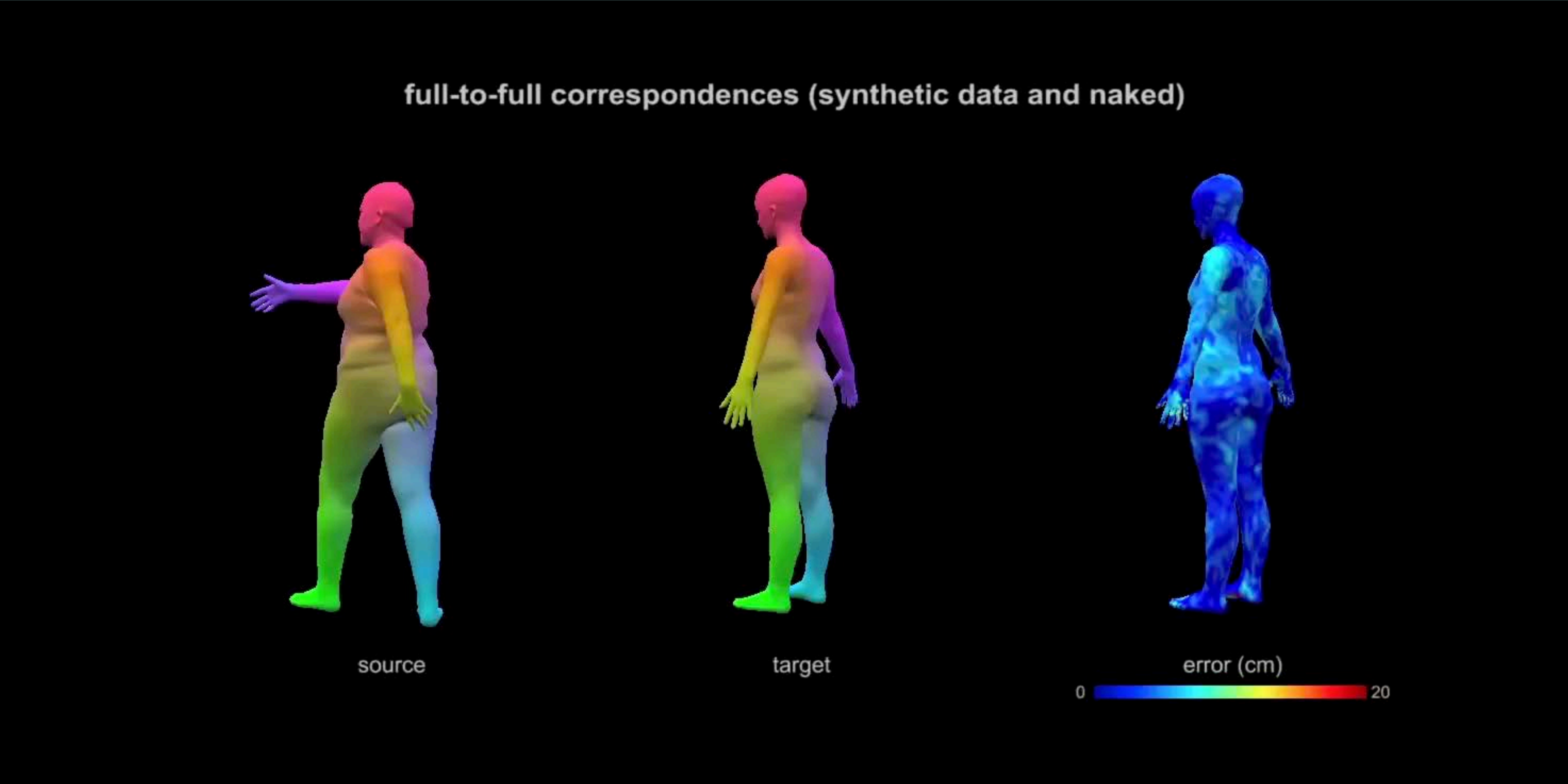
Evaluation



FAUST dataset

Results

Results: Static Shapes



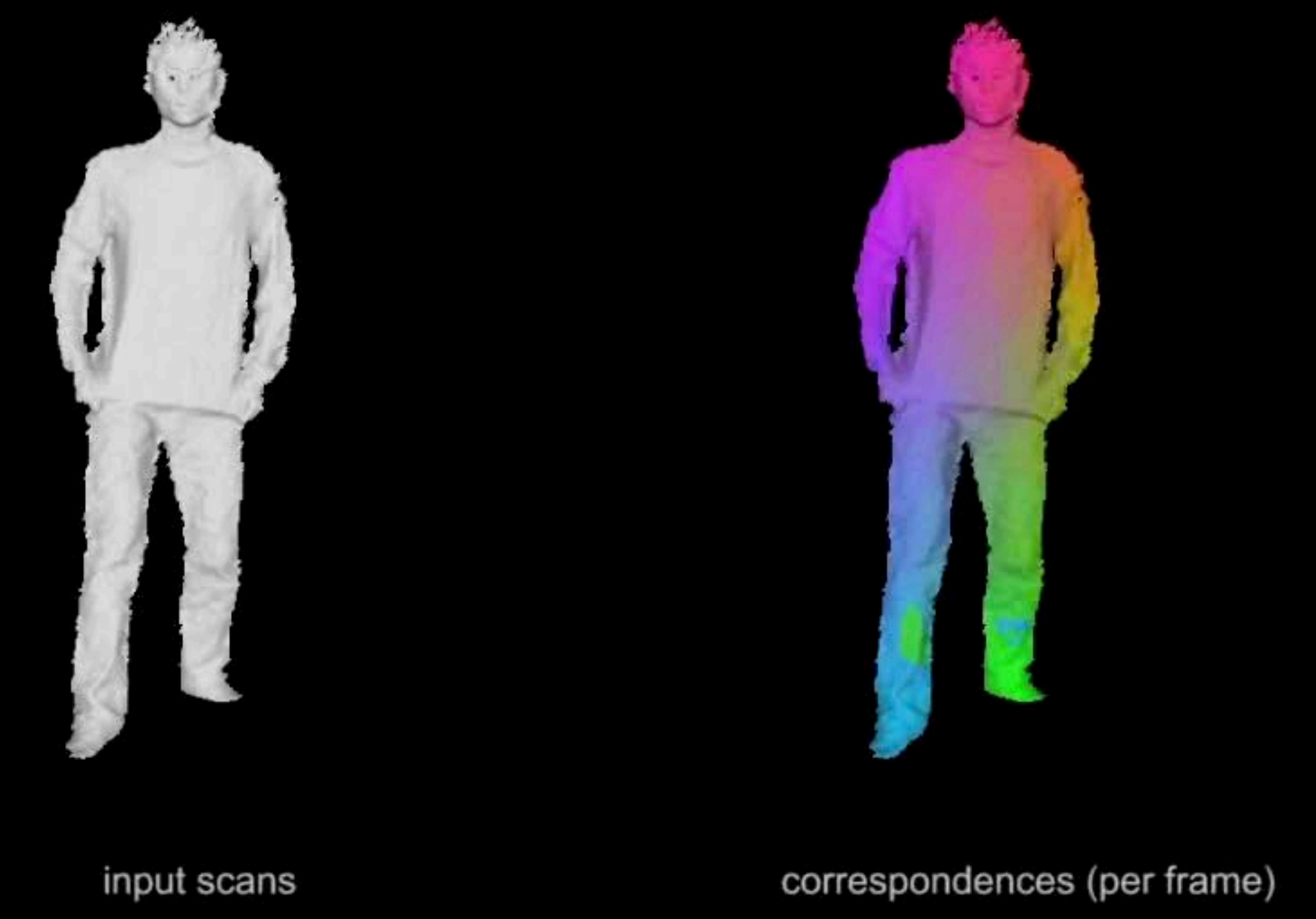
Results: Static Shapes

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

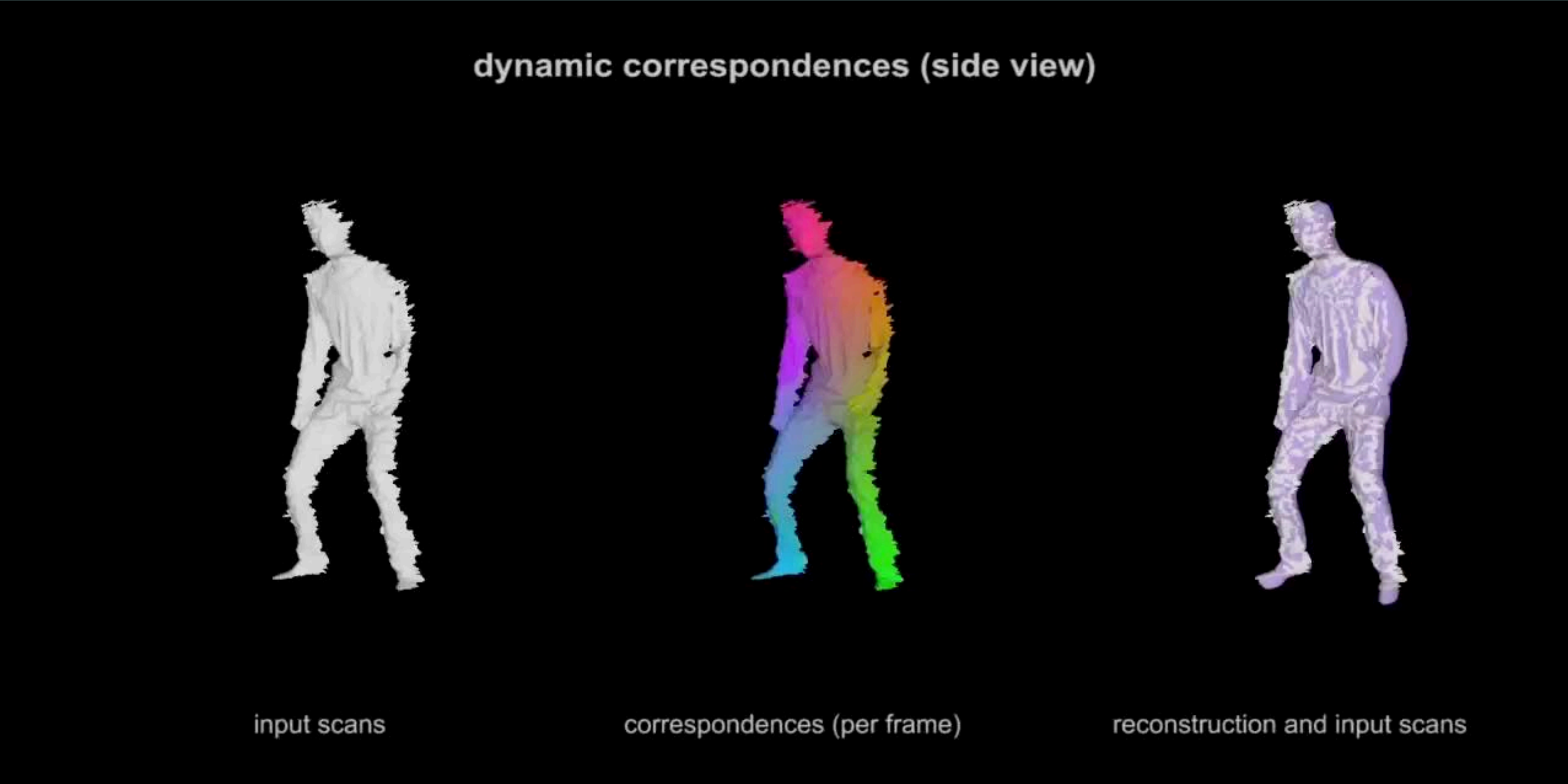
source

target

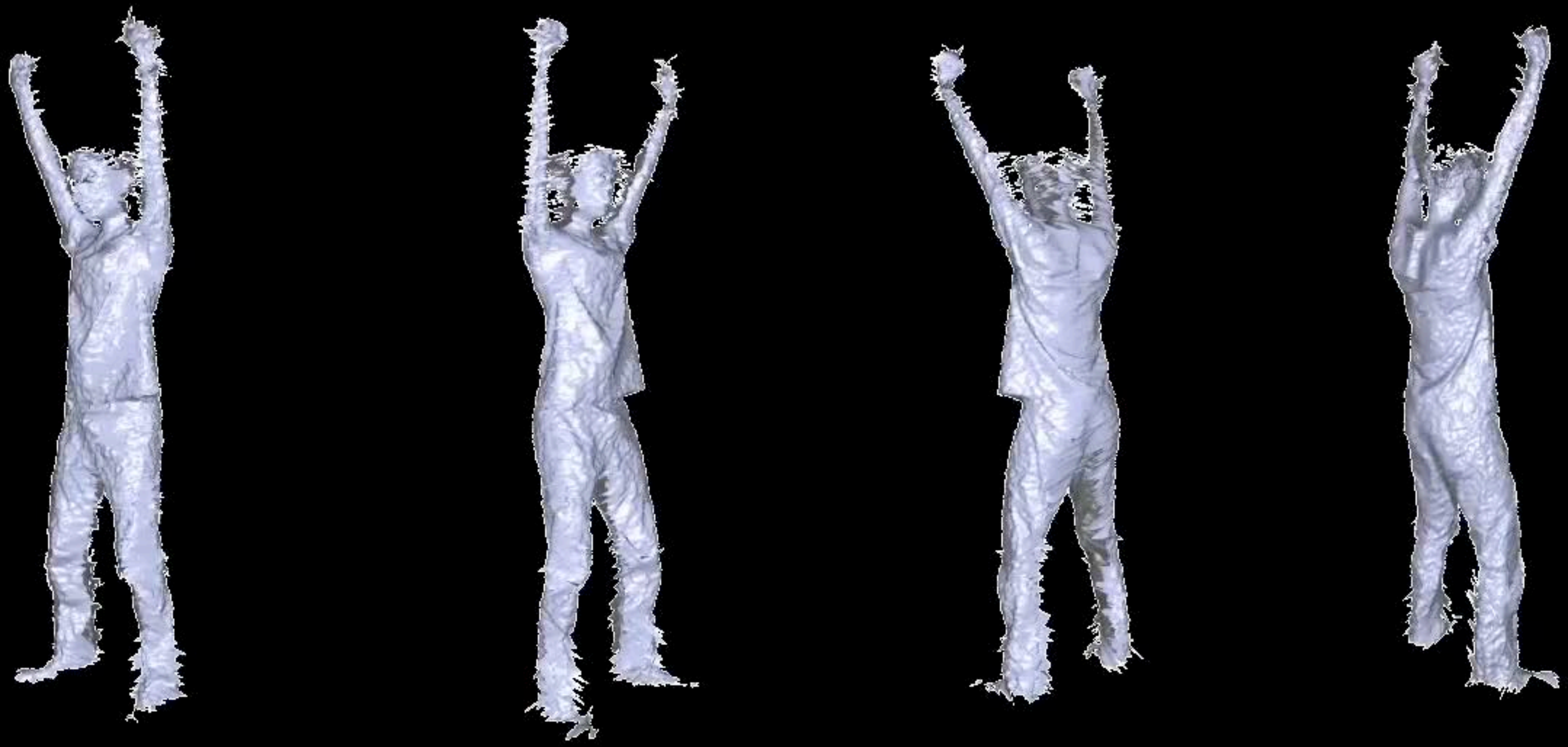
Results: Dynamic Shapes



Results: Dynamic Shape Reconstruction



4 Stationary Kinects



Dense Correspondences

Applications

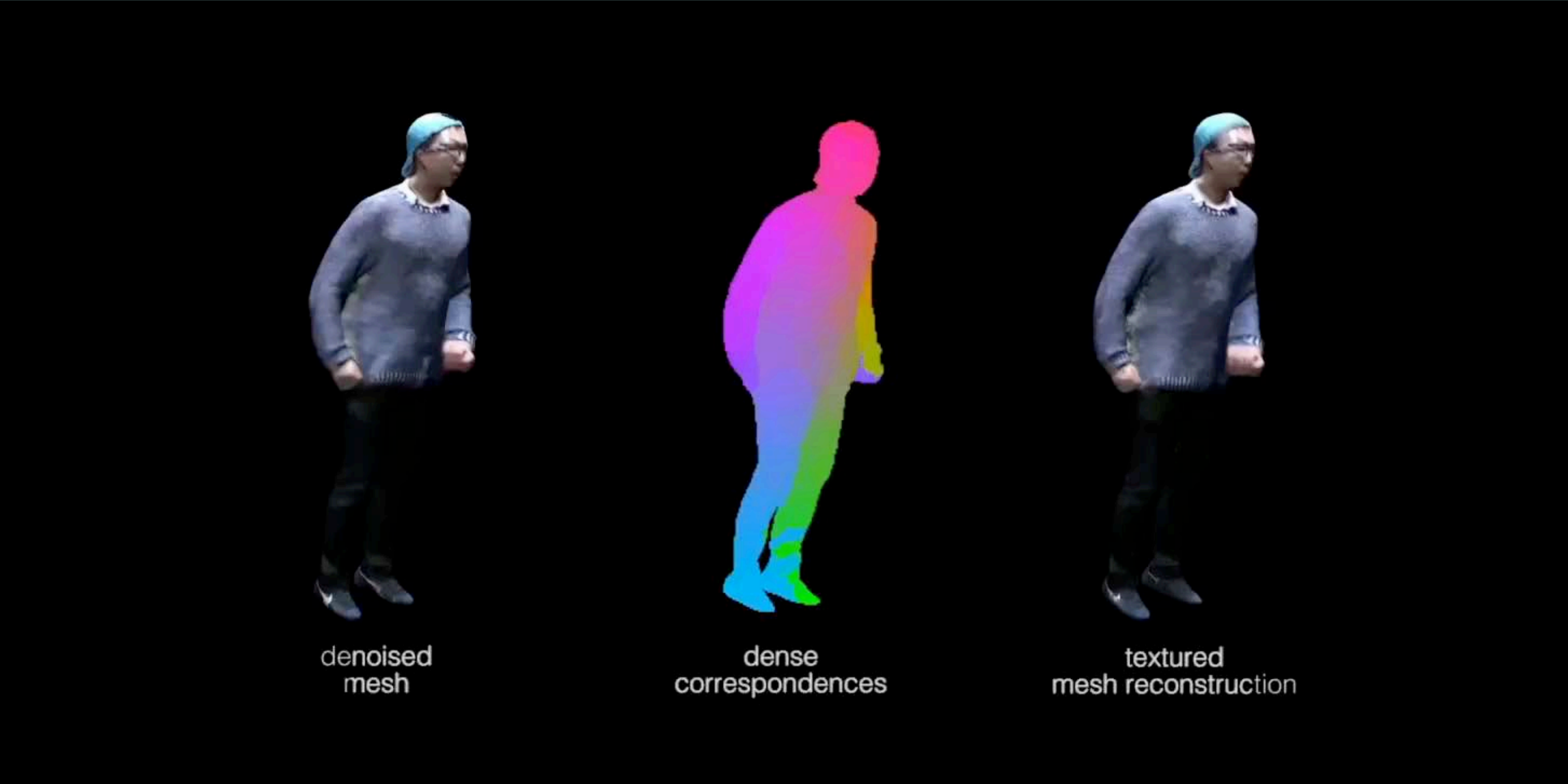
Low Cost Capture & Moving Target

Registration and Reconstruction

output scan alignment

output textured reconstruction

Filtering and Texture Reconstruction



Application

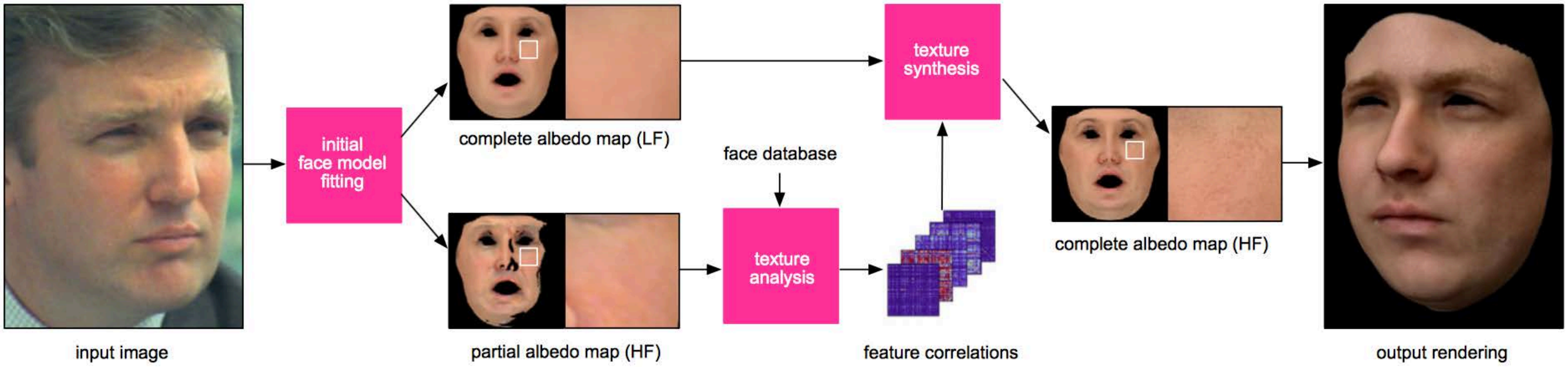
Photorealistic Texture Synthesis

Photo-Realistic Faces Using Deep Learning

input picture | output albedo map | rendering | rendering (zoom)

Inspiration: Style Transfer(Gatys et al. 2016)

Deep CNN-based Synthesis Approach



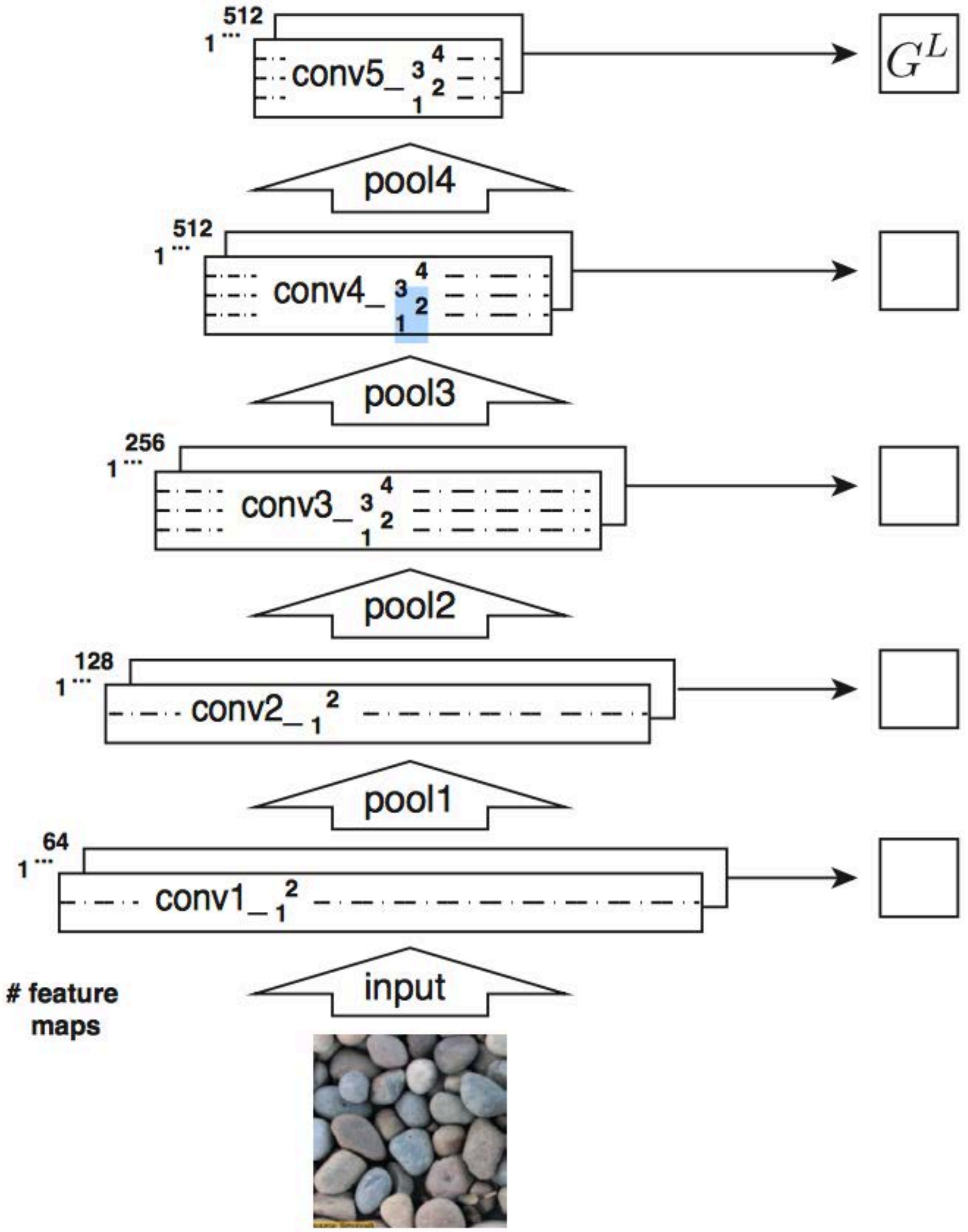
Feature Correlations (Gatys et al. 2015)

$$G^l(I) = \frac{1}{M_l} F^l(I) \left(F^l(I)\right)^T \in \mathbf{R}^{N_l \times N_l}$$

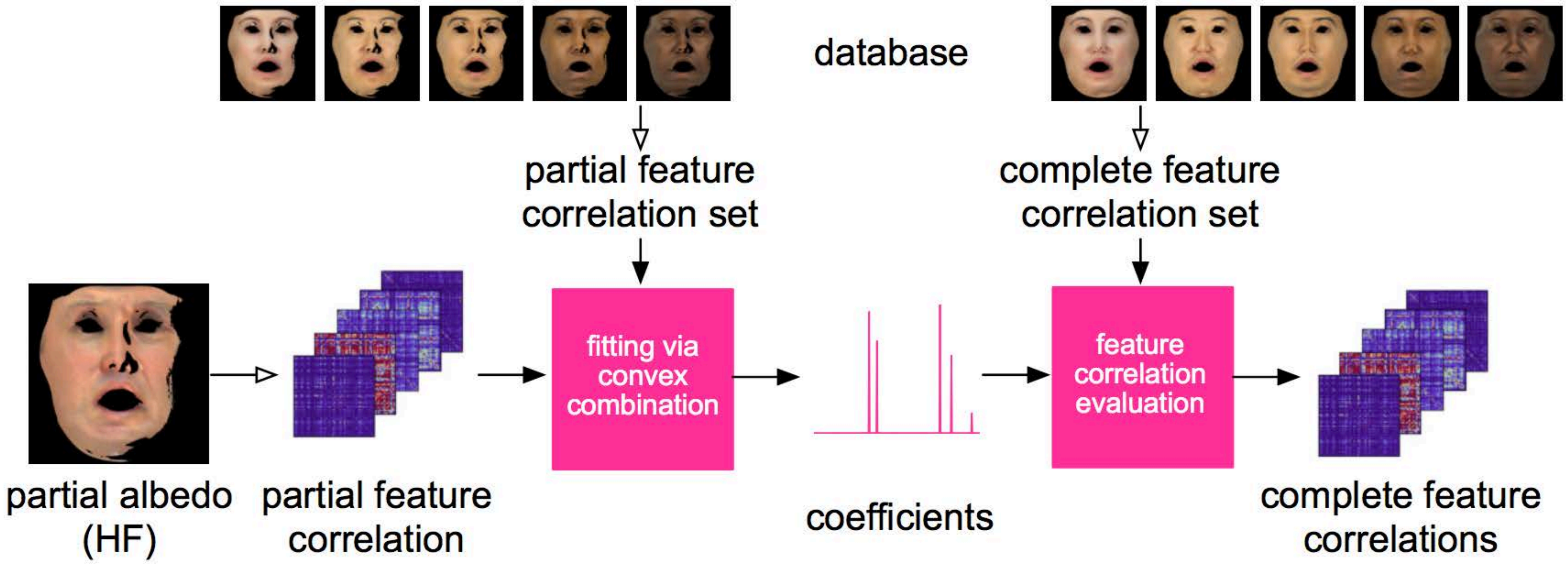
Feature correlation

$$F^l(I) \in \mathbf{R}^{N_l \times M_l}$$

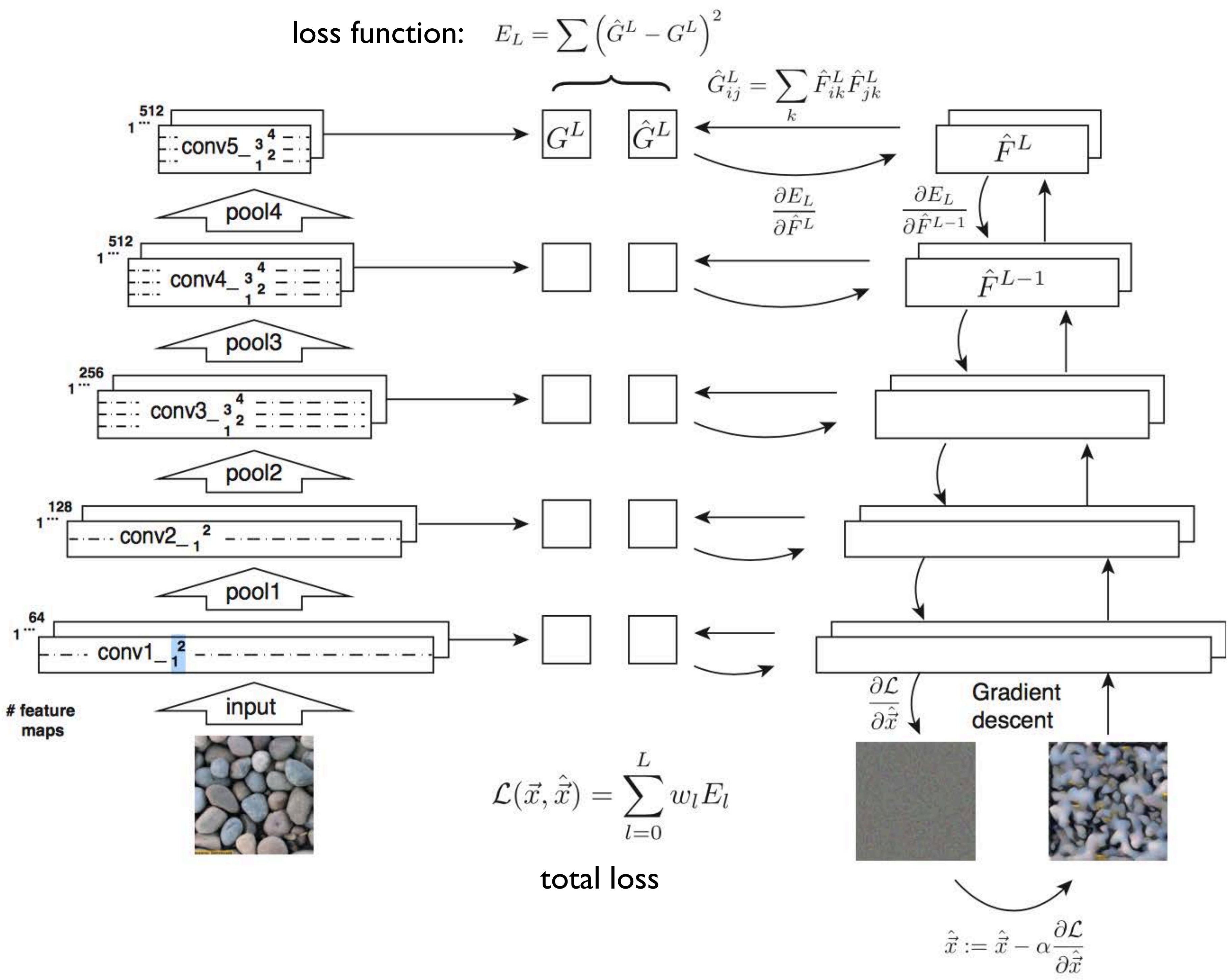
Feature response



Texture Analysis



Texture Synthesis (Gatys et al. 2015)

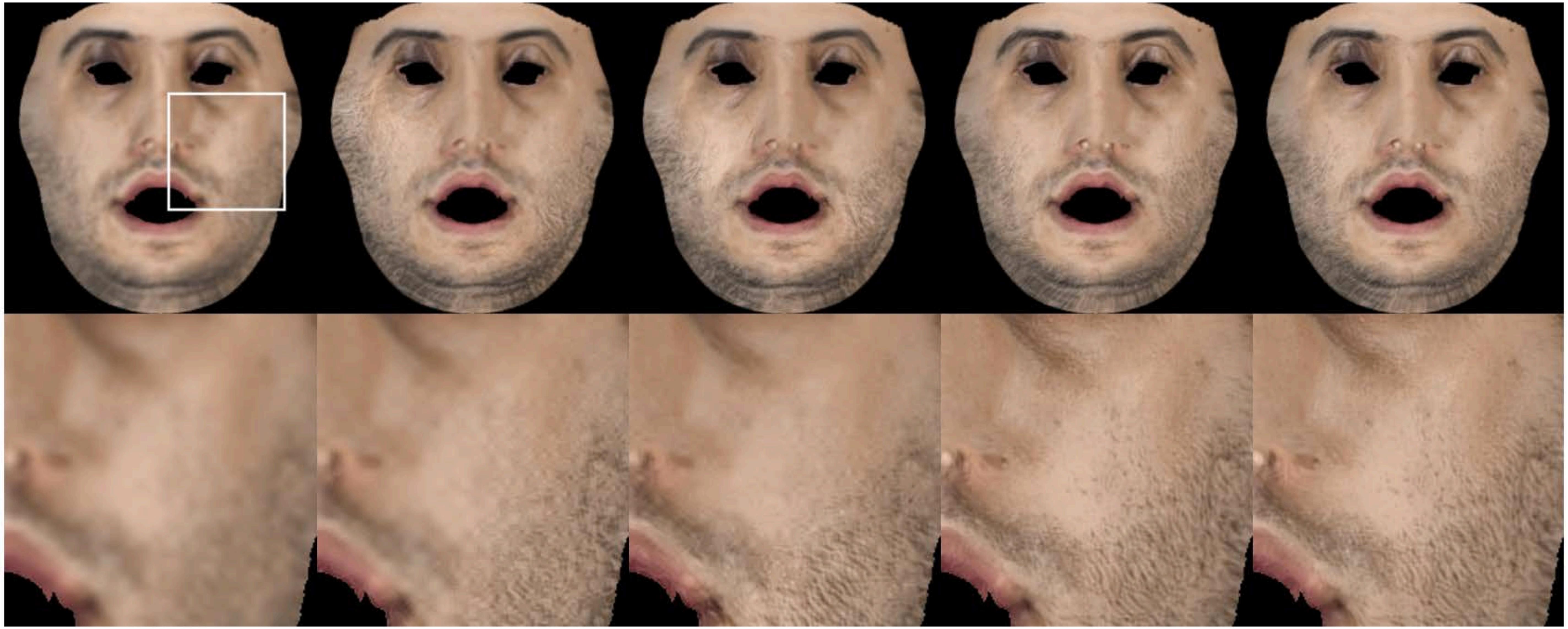


Texture Synthesis (Saito et al. 2016)

$$\min_{I} \sum_{l \in L_F} \left\| F^l(I) - \hat{F}^l(I_0) \right\|_F^2 + \alpha \sum_{l \in L_G} \left\| G^l(I) - \hat{G}^l(I_0) \right\|_F^2$$

input $\alpha = 0$ $\alpha = 2$ $\alpha = 20$ $\alpha = 2000$

Different Number of Mid-Layers



1 layer | 2 layers | 3 layers | 4 layers | 5 layers

Detail Preservation via Convex Combination

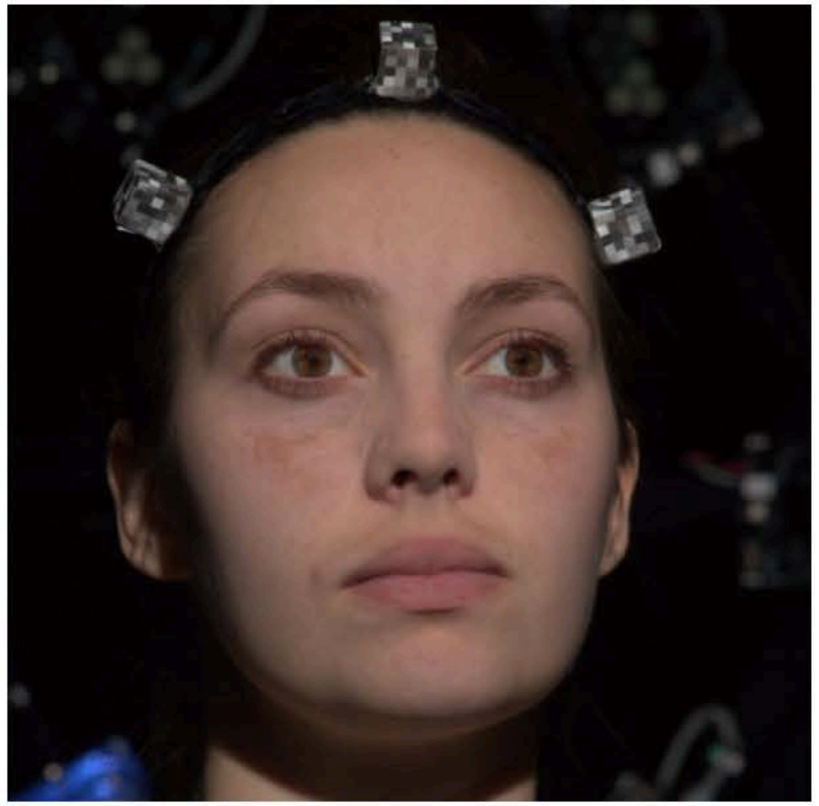
input

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

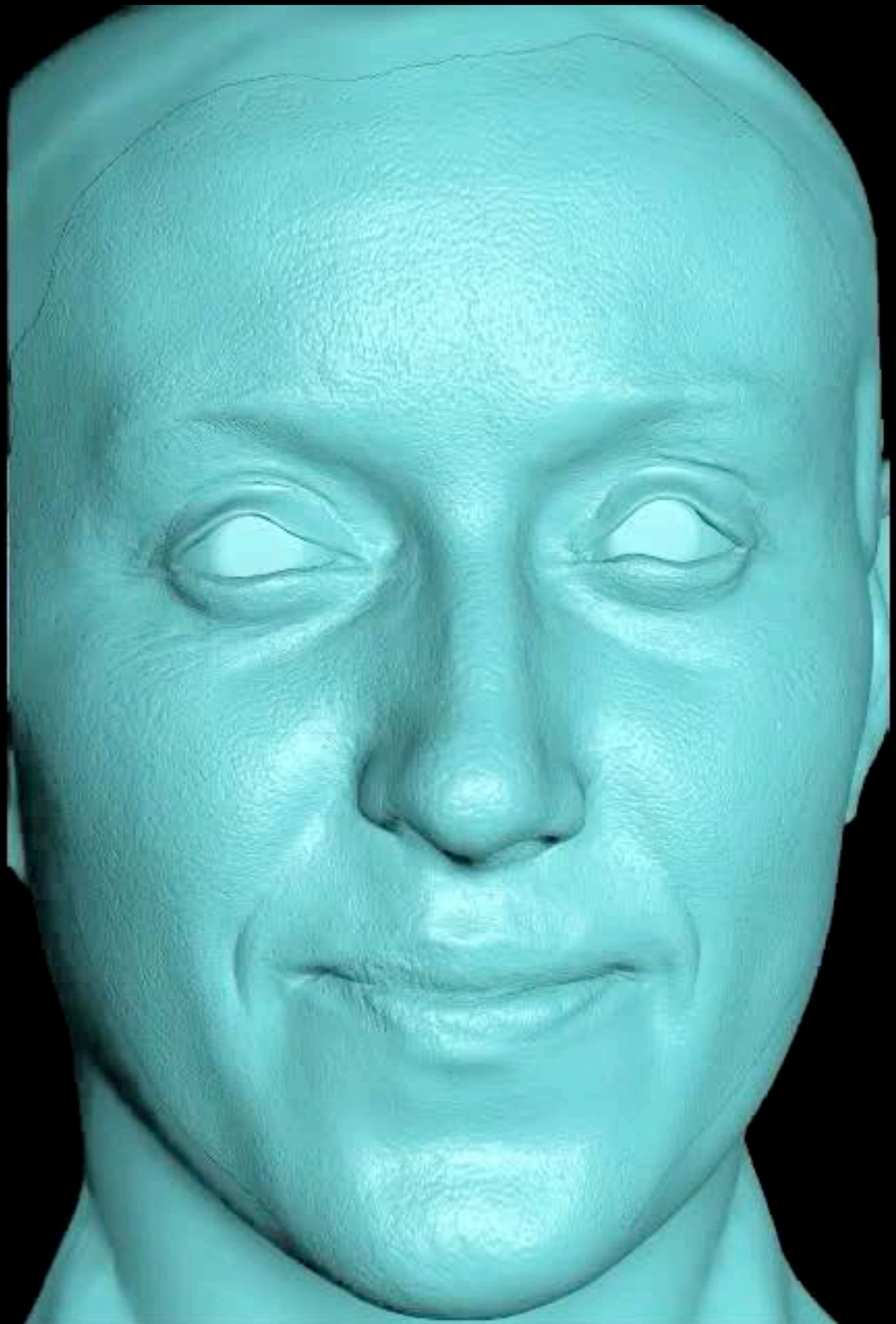
CVPR 2016

input 2D image

output textured 3D face (AFW)

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)

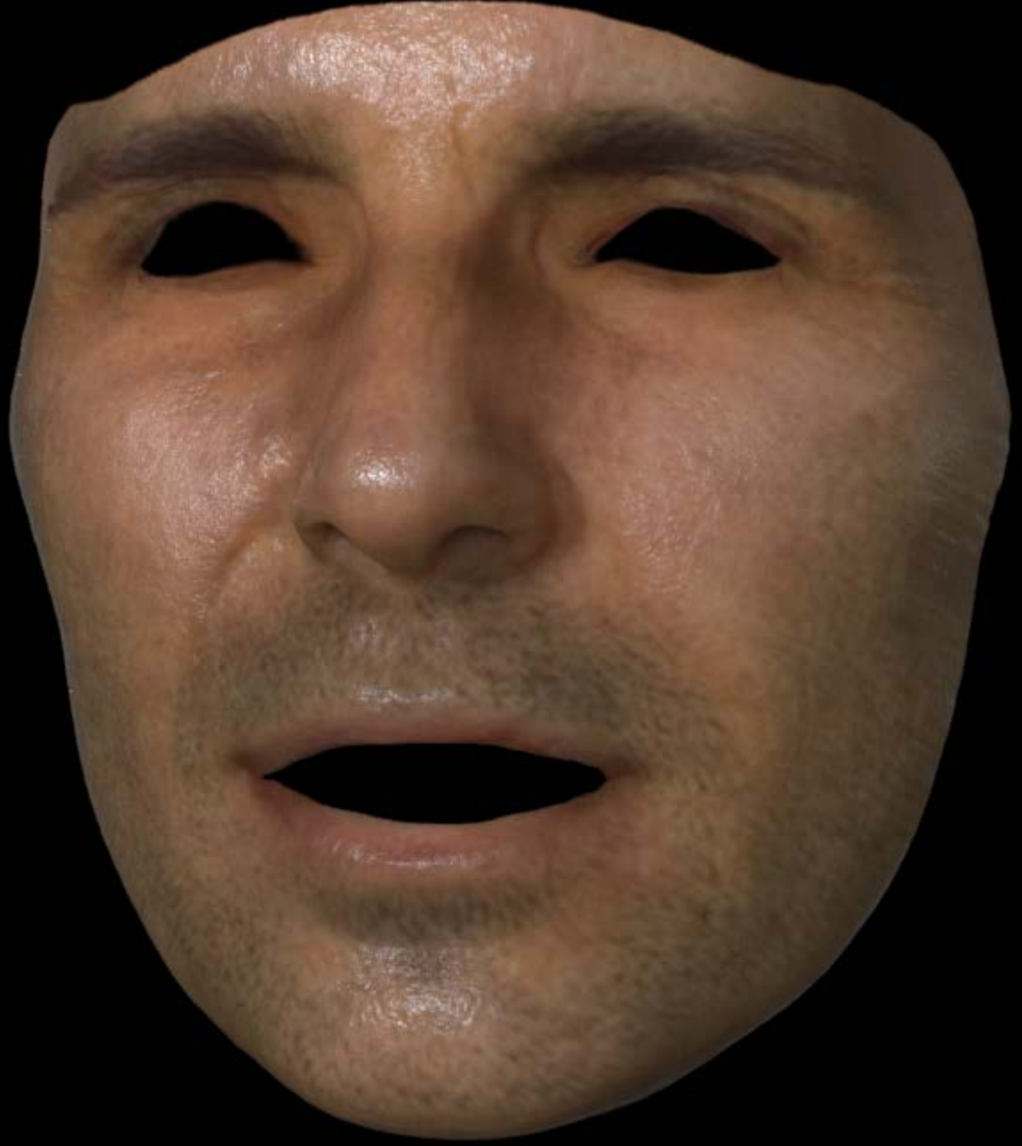
Full Geometry and Reflectance Inference

internet picture

high-fidelity reconstruction

Yamaguchi et al. (2018)

Full Geometry and Reflectance Inference



internet picture

high-fidelity reconstruction

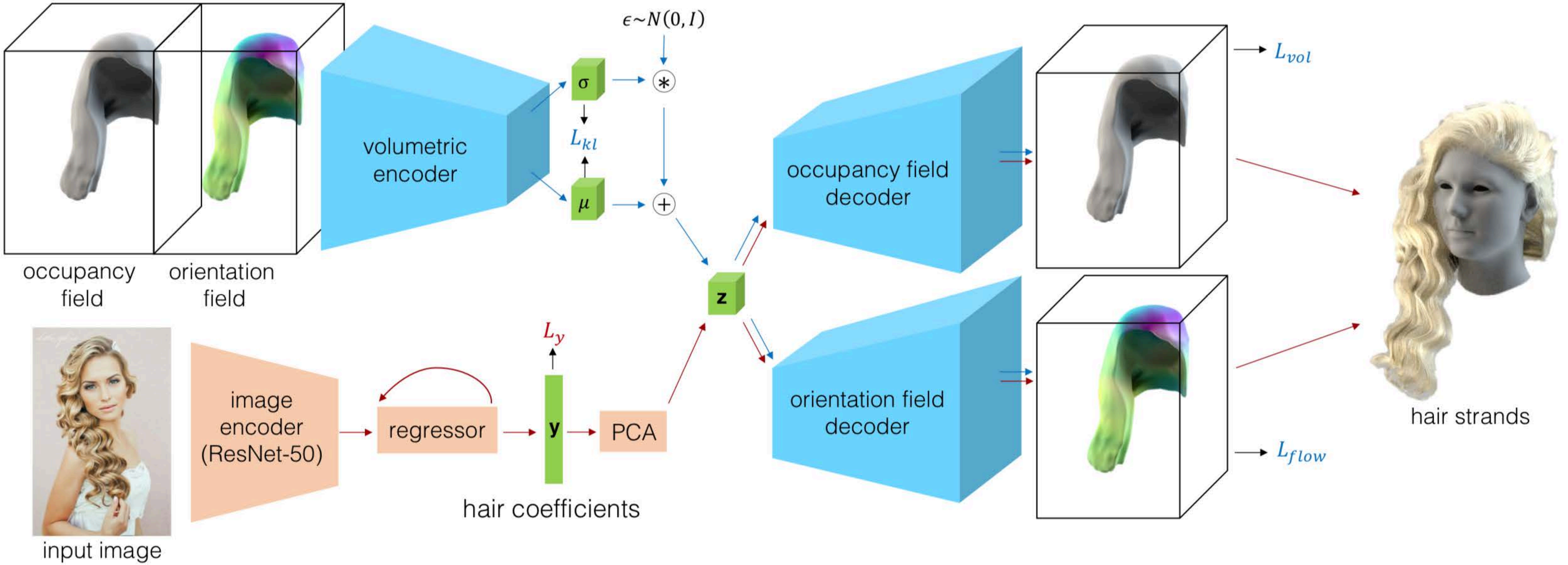
Hair Digitization

Five-strand Dutch braid

Reference photo

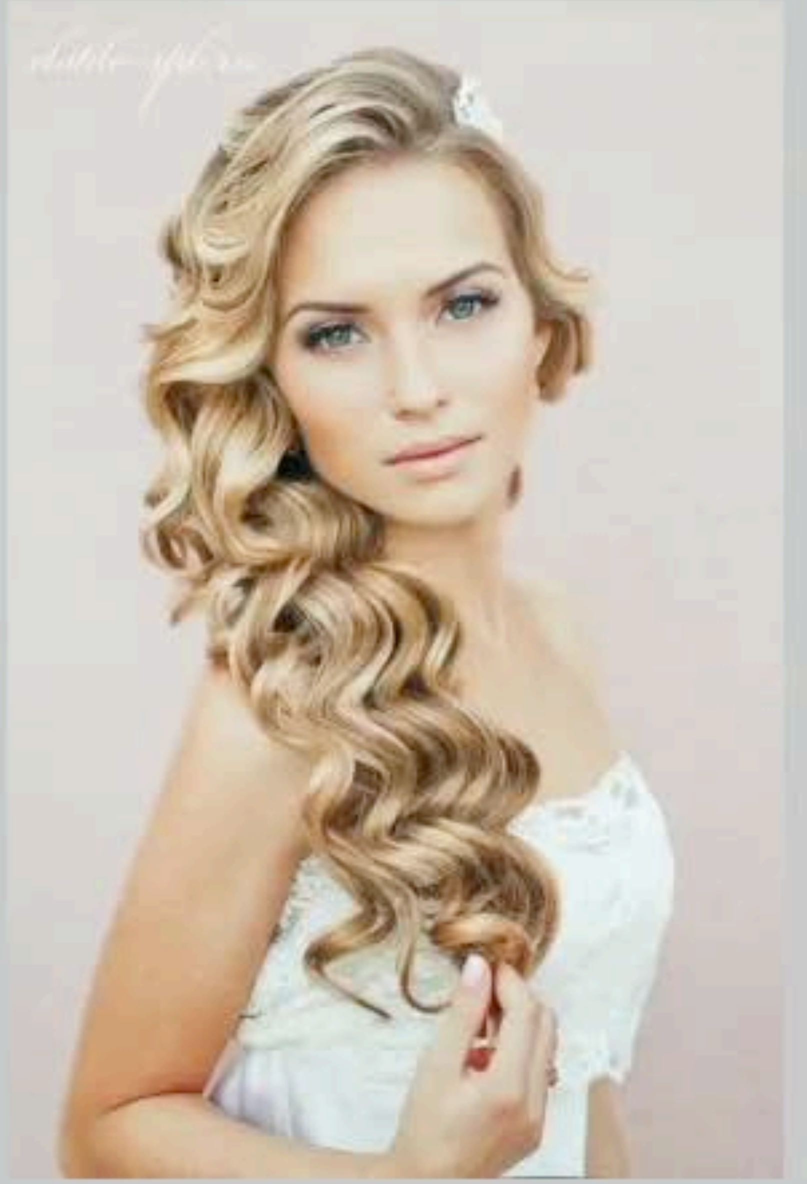
Our result

Deep Learning for Hair Modeling



Saito et al. (2018)

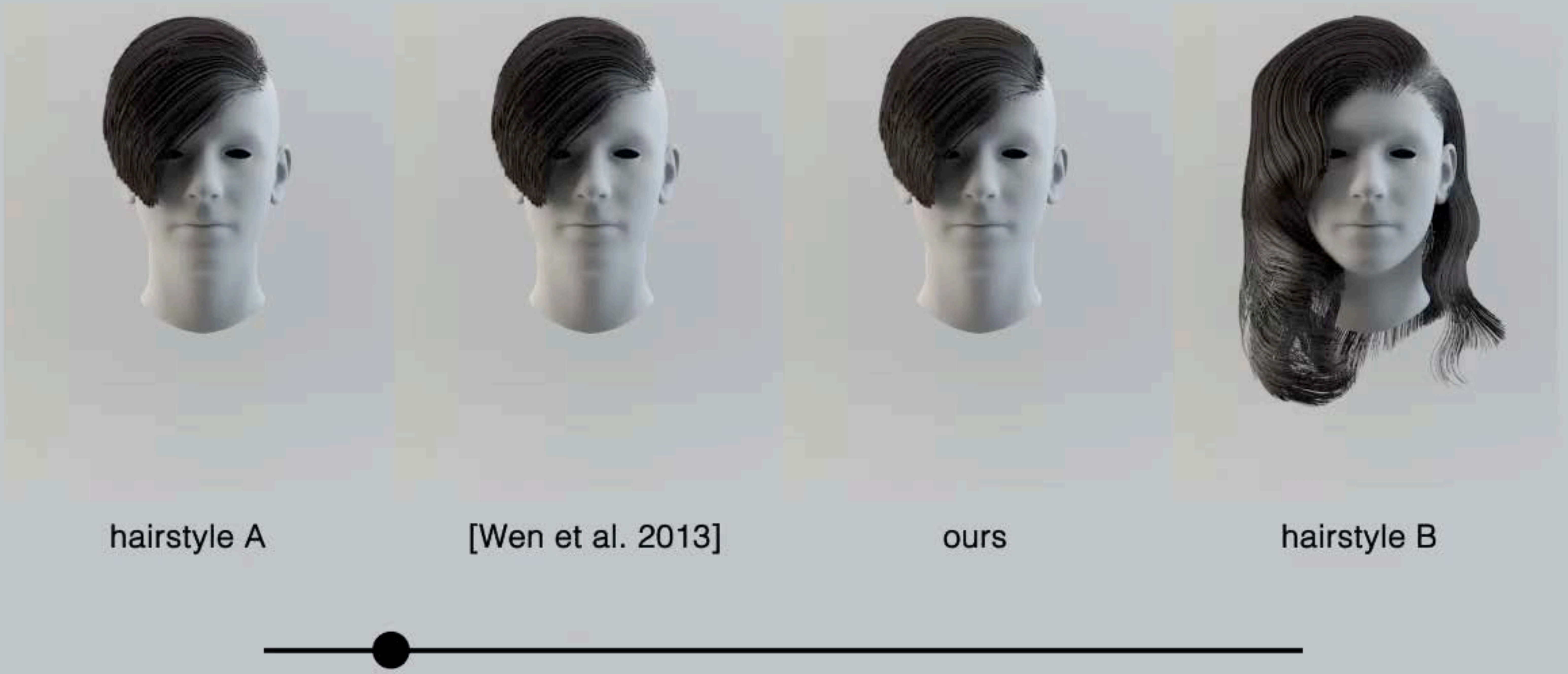
Deep Learning for Hair Modeling



input image

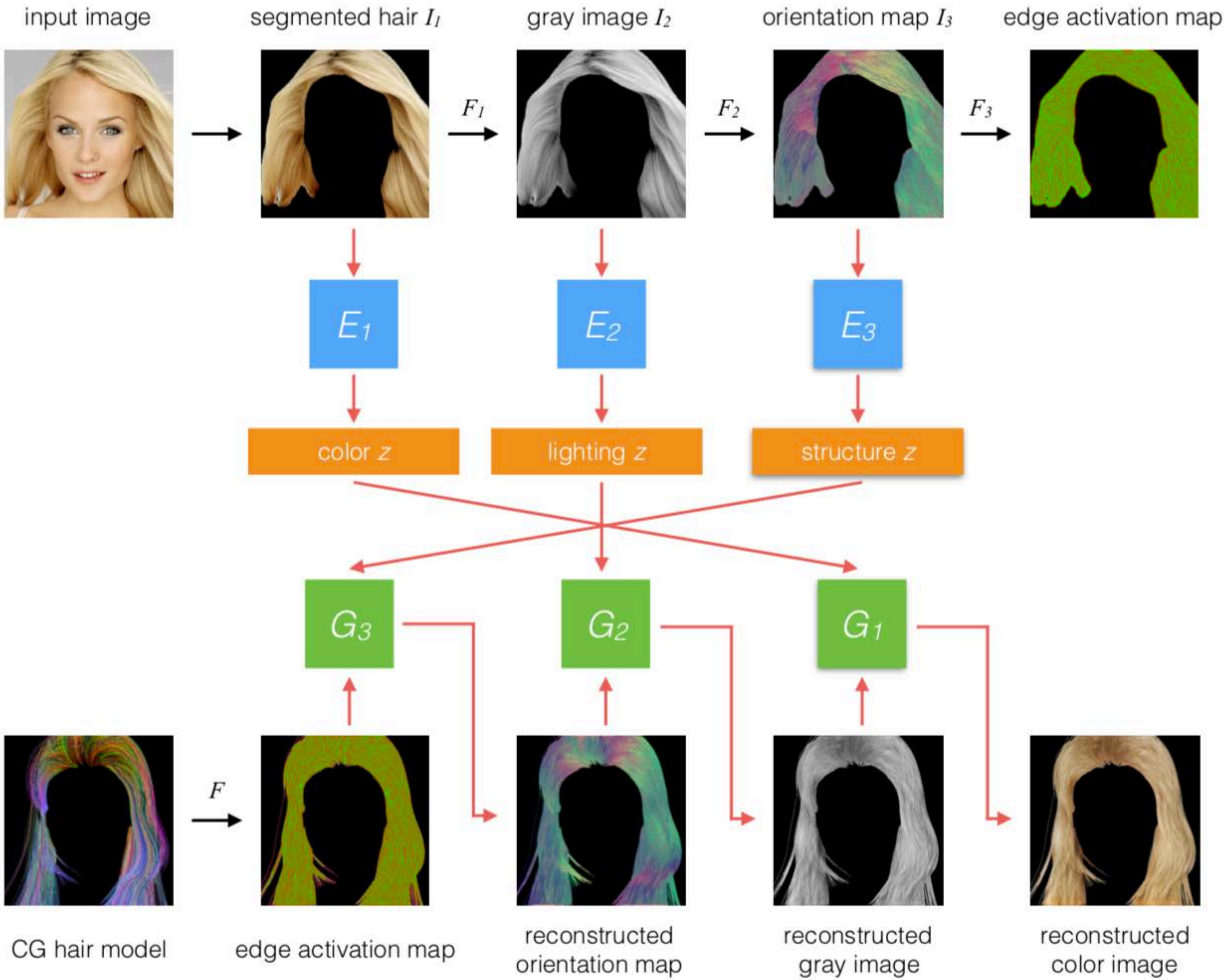
Saito et al. (2018)

Deep Learning for Hair Interpolation



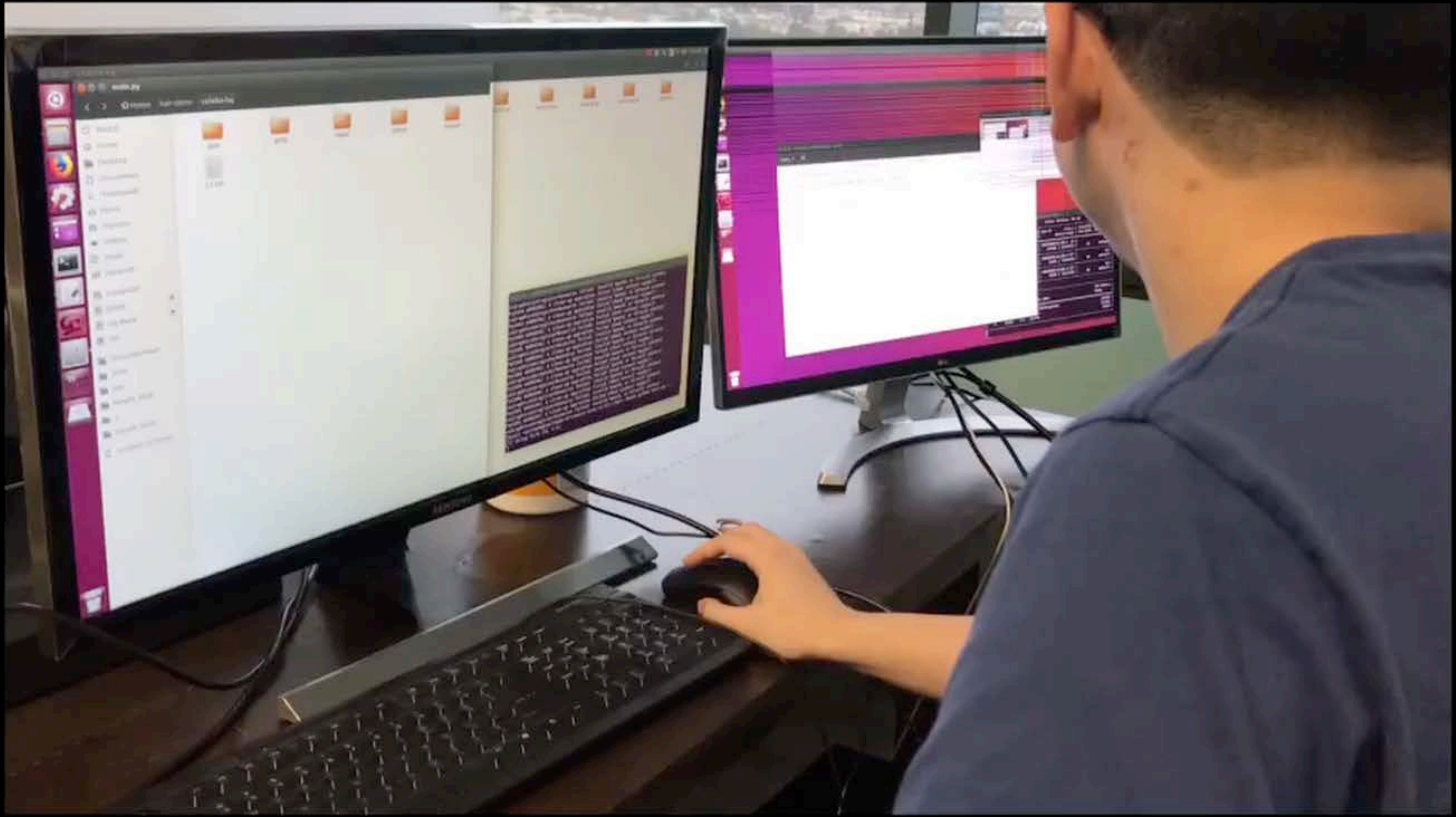
Saito et al. (2018)


Wei et al. (2018)



Wei et al. (2018)

Starting

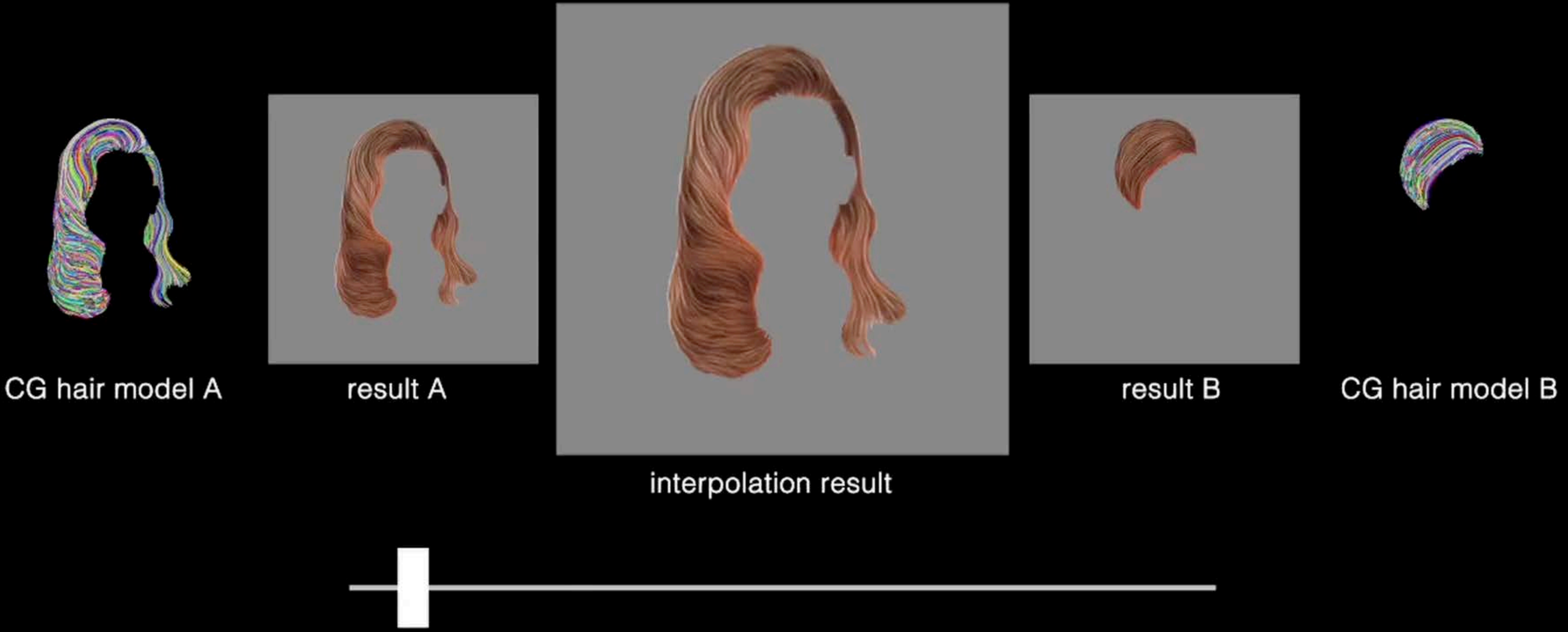


Wei et al. (2018)

Color Interpolation

Wei et al. (2018)

Hairstyles Interpolation



Wei et al. (2018)

What’s next?

AI-Driven Graphics


YouTube
Search

VFX-Level Augmented Reality

Thanks!