Yi Zhou 周易

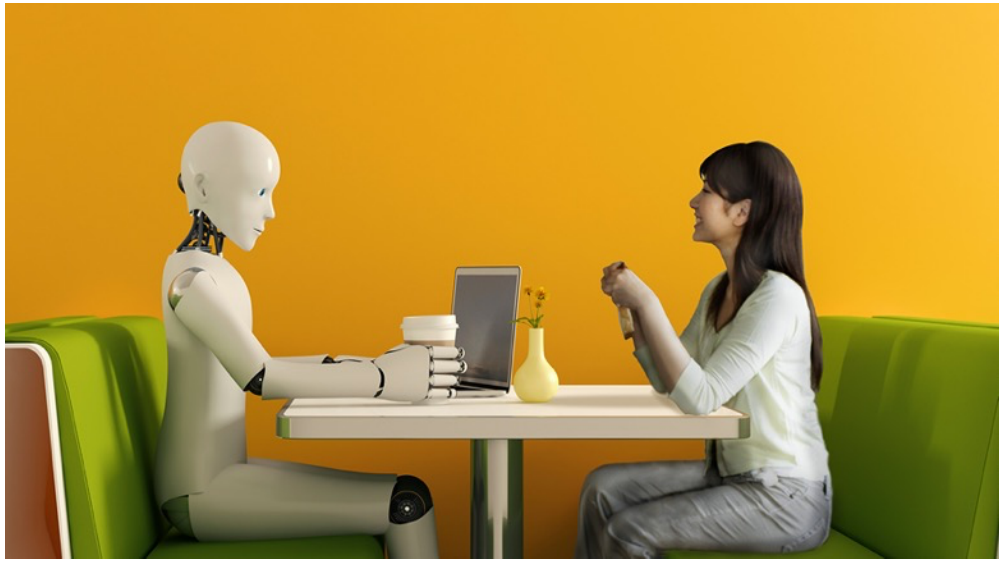
Research Scientist at Adobe

PhD from the University of Southern California.
Advisor: Dr. Hao Li

Ms and Bs from Shanghai Jiao Tong University.
Advisor: Dr. Shuangjiu Xiao

Research Area

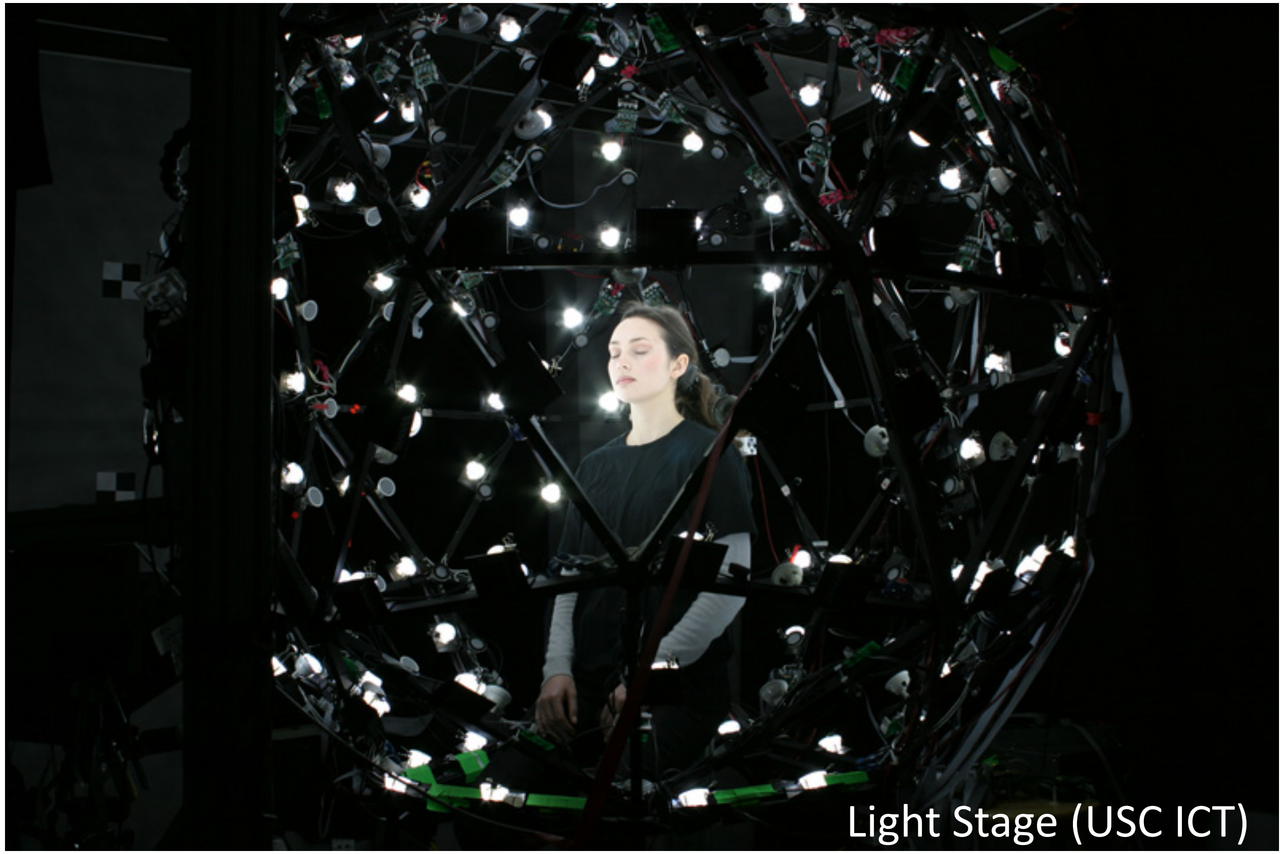
Autonomous 3D Avatar

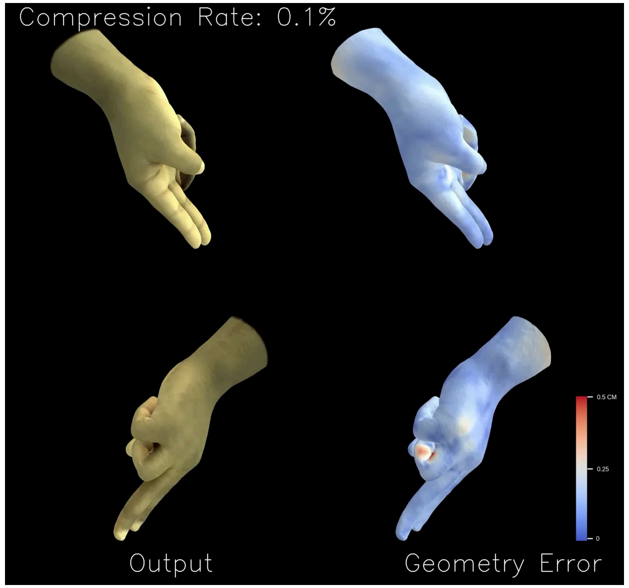


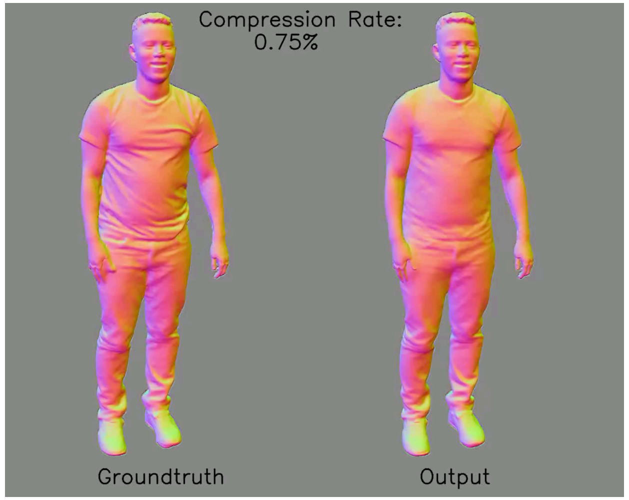
- Human Digitizing
- Motion Synthesis
- Deep Representation Learning
- AR & VR

Human Digitizing

Reconstruction from stereo systems



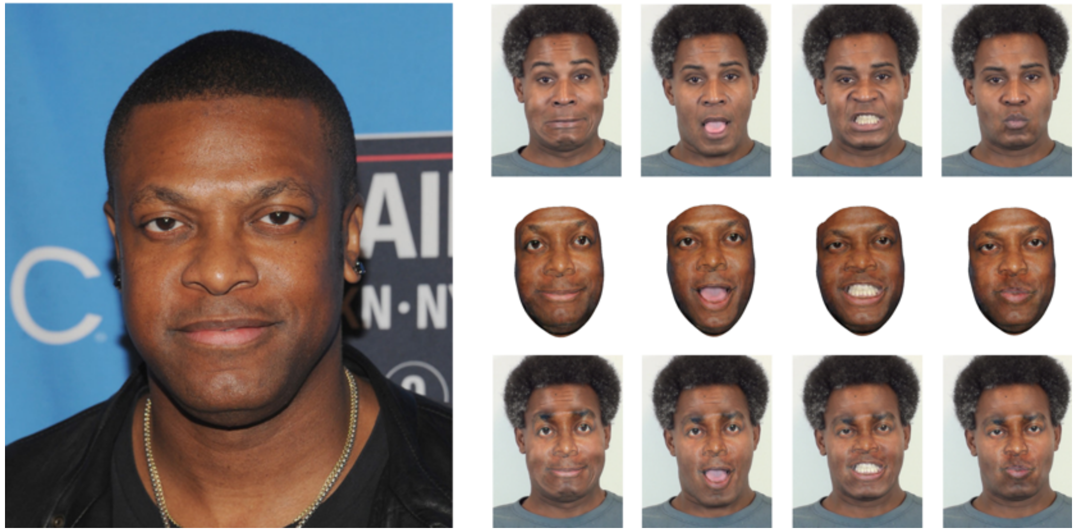




Neurips 2020

Human Digitizing

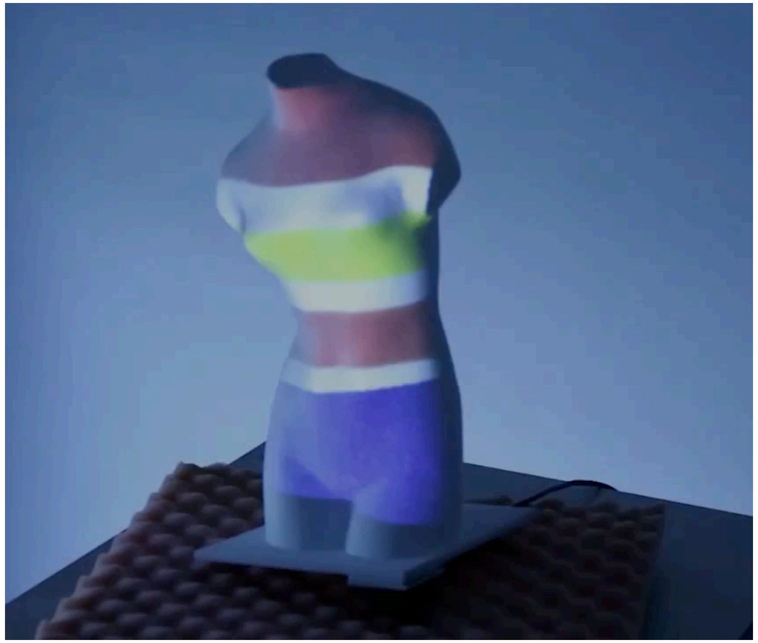
Reconstruction from single-view images

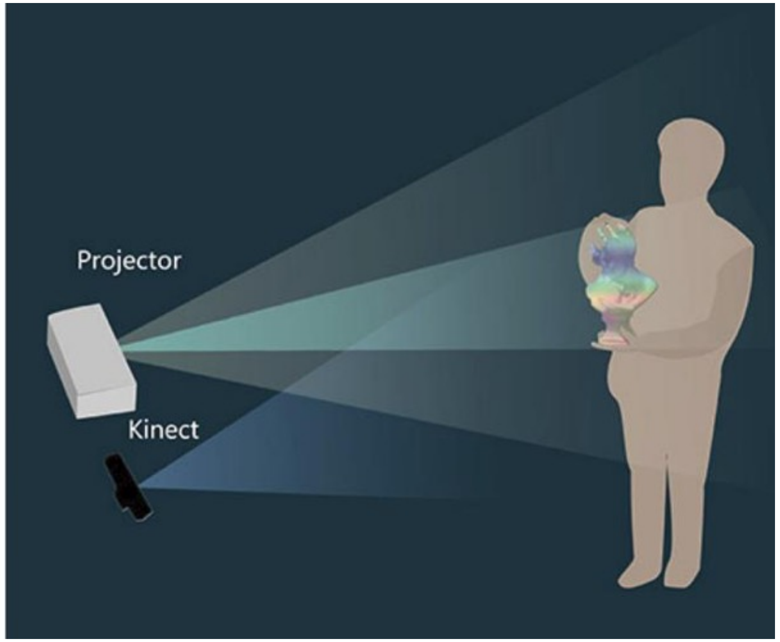


ECCV 2018

ICCV 2017

AR & VR

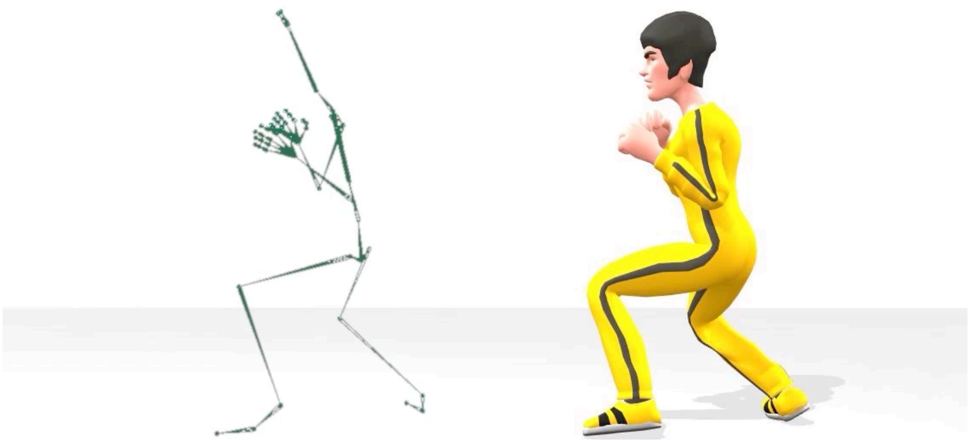




CHI 2016

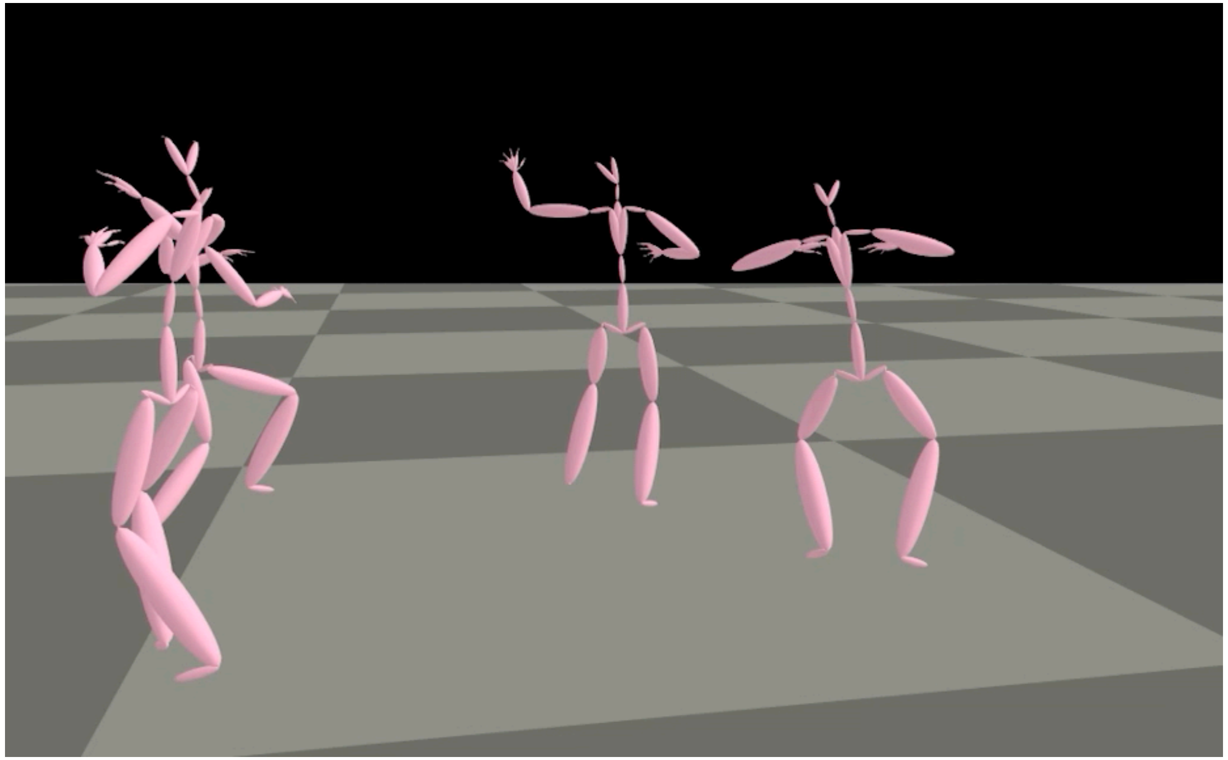
Motion Synthesis

Recurrently generate long motion sequence.



ICLR 2018

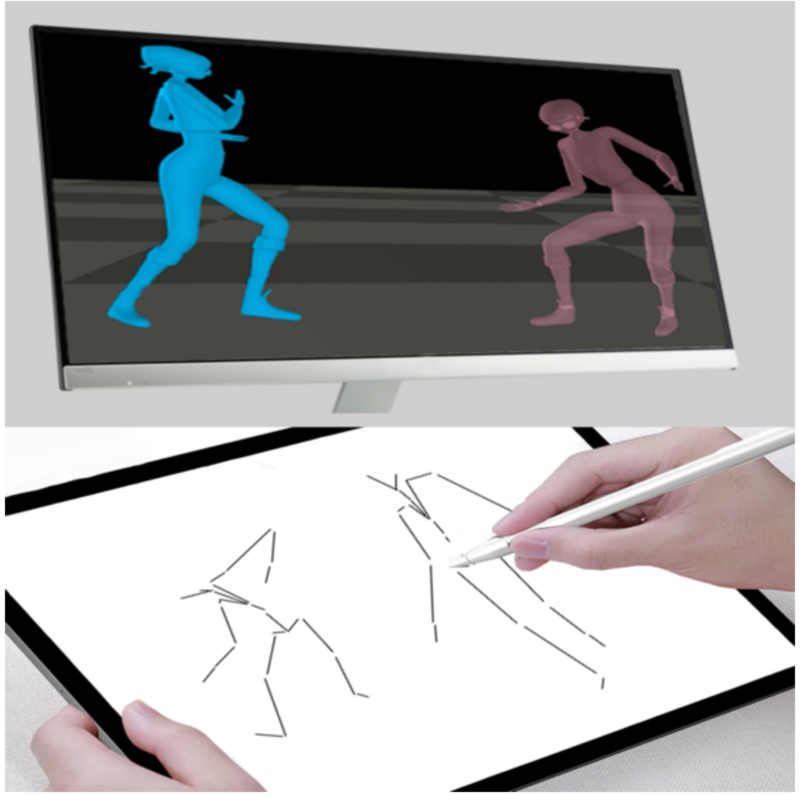
Long-term motion inbetweening.

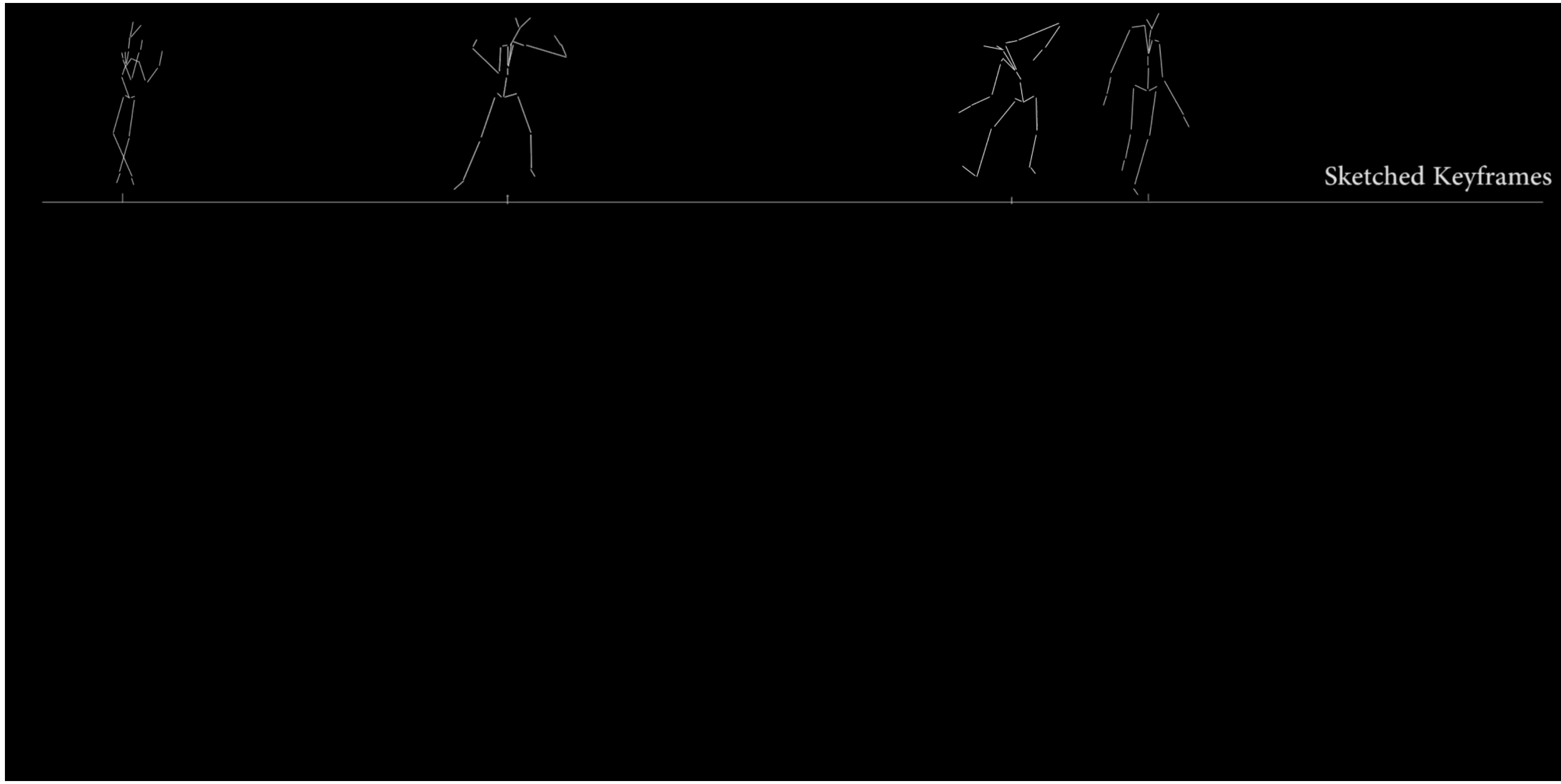


arxiv

Motion Synthesis

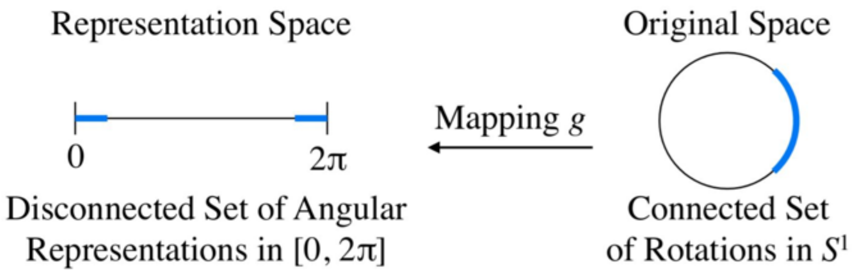
3D motion from 2D sketches

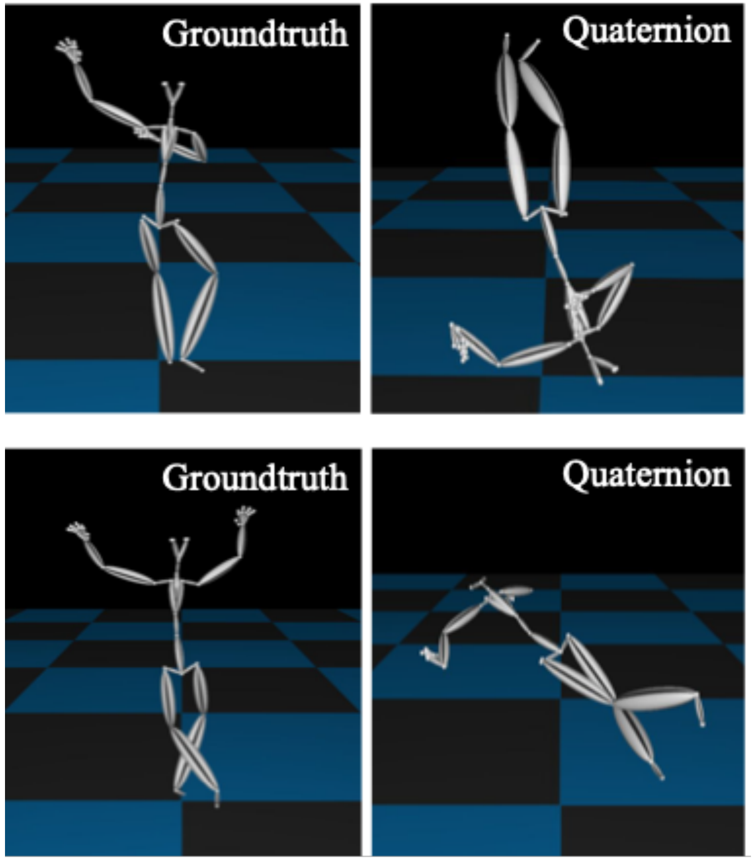




Deep Representation Learning

How to represent **rotation** in neural networks?





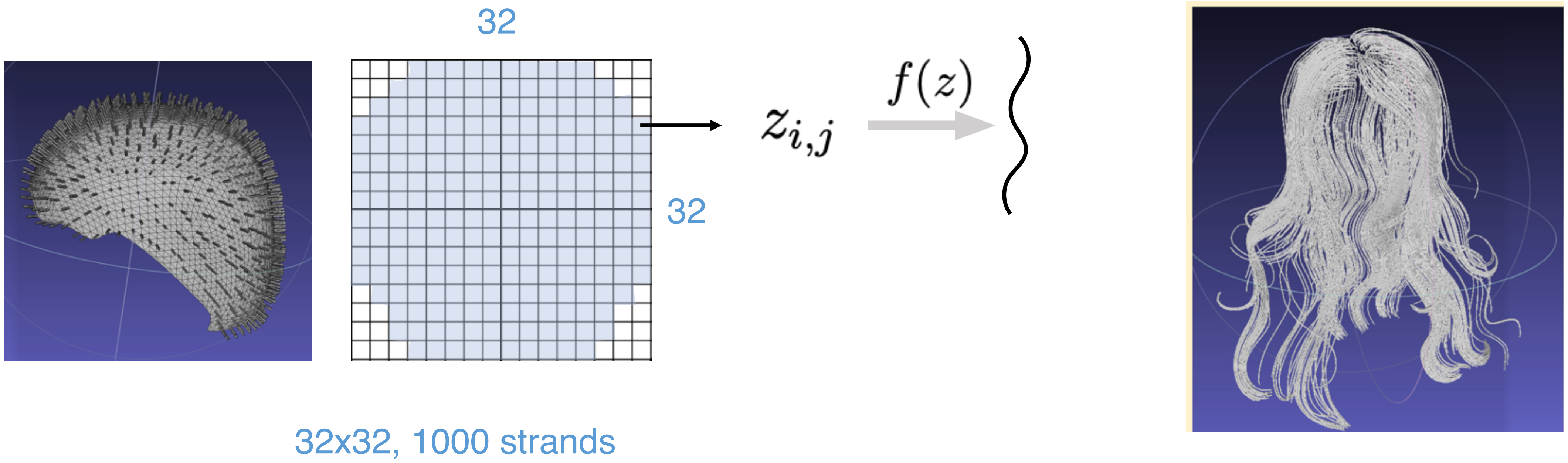
Quaternion
Axis Angle
Euler Angles
Rodriguez
...

3D Rotation
SO(3)

CVPR 2019

Deep Representation Learning

How to represent **3D hair** in neural networks?



Deep Representation Learning

How to represent **Mesh** in neural networks?

Fully Convolutional Mesh Autoencoder using Spatially Varying Kernels

Neurips 2020

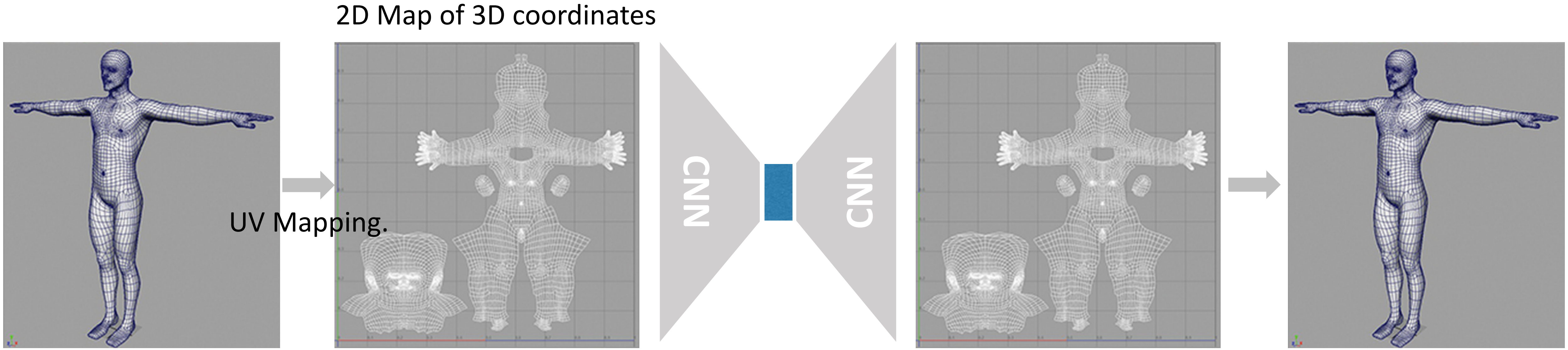
Yi Zhou, Chenglei Wu, Zimo Li, Chen Cao, Yuting Ye, Jason Saragih, Hao Li and Yaser Sheikh.

How to apply CNN on registered 3D meshes?

Registered Mesh: Mesh with the same number and order of vertices and edges.

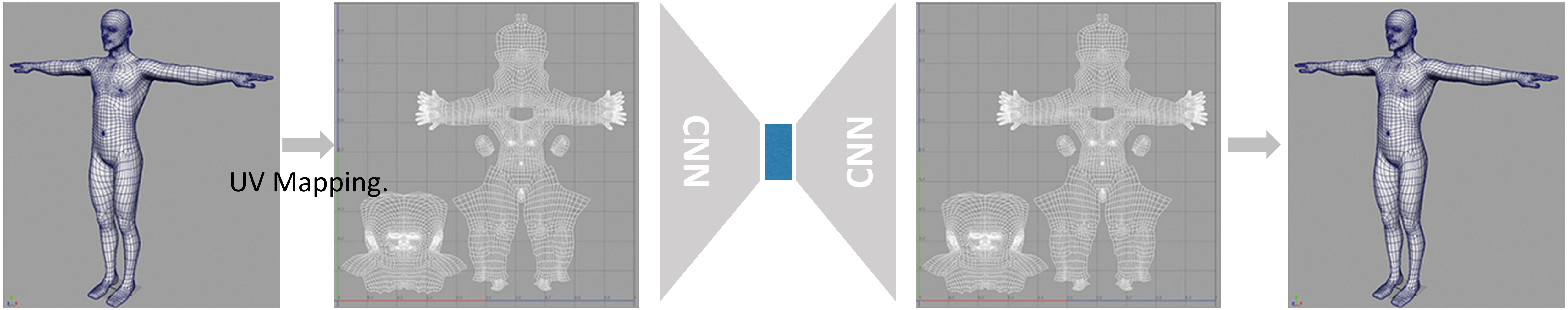
http://dfaust.is.tue.mpg.de/

Common practice: 2D CNN in UV space

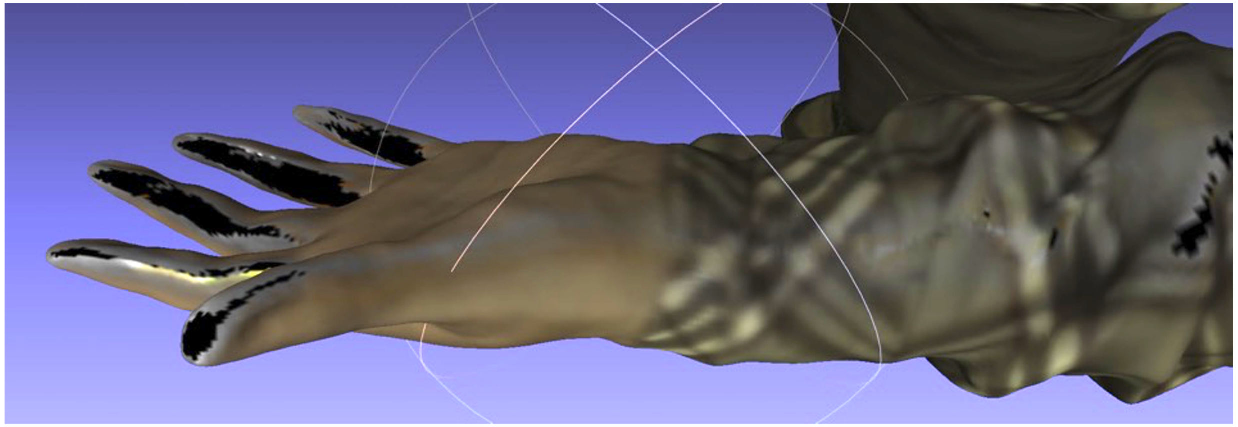


Common practice: 2D CNN in UV space

2D Map of 3D coordinates



UV Mapping.



Problems:

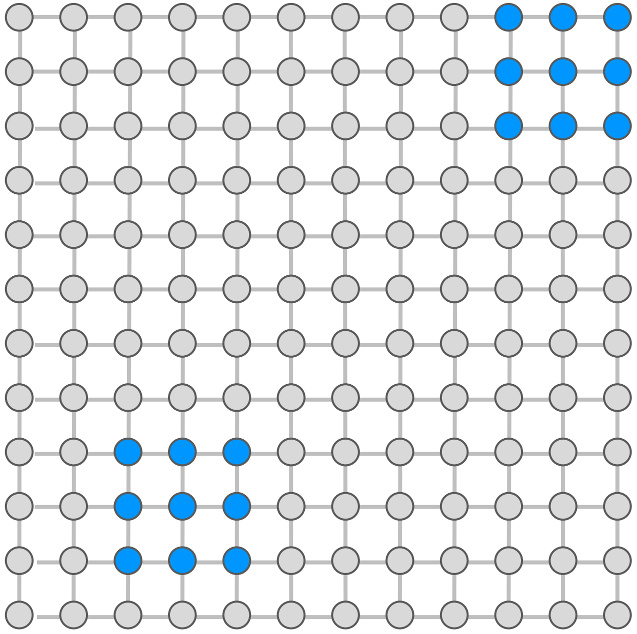
1. Artifacts along seam lines and from distortion
2. Poor performance when reconstructing global deformation

Directly apply CNN on Mesh?

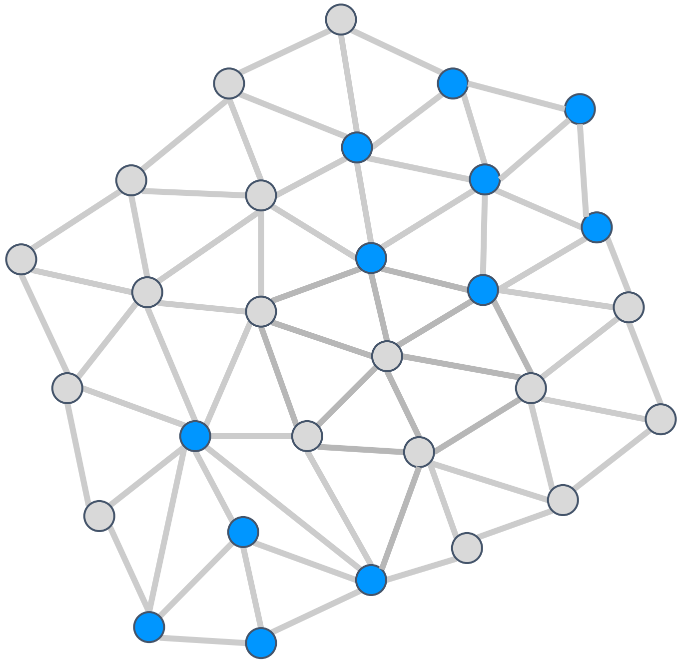
Challenge:

Cannot sample uniform kernels on a non-uniform mesh

Regular 2D CNN ? Mesh CNN

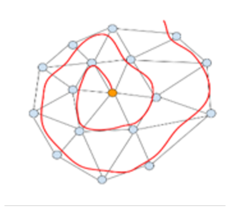


Shift-invariance Grid

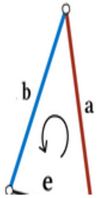


Shift-variance Mesh

Existing Methods:

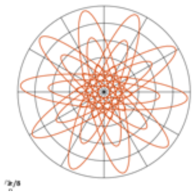


Spiral CNN

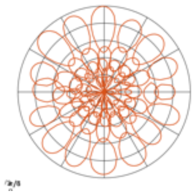


Edge CNN

GCNN



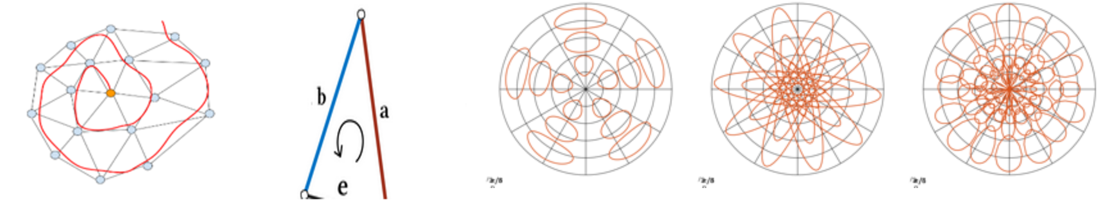
ACNN



MoNet

- **Low** Reconstruction Accuracy
- **Low** Generalizability

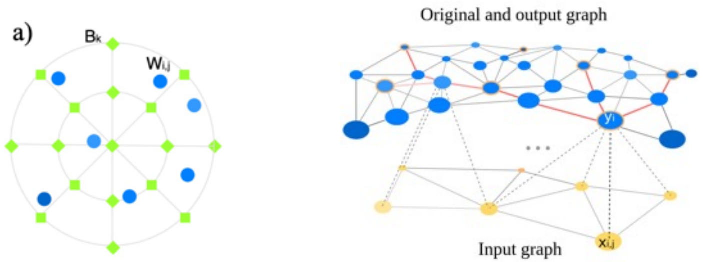
Existing Methods:



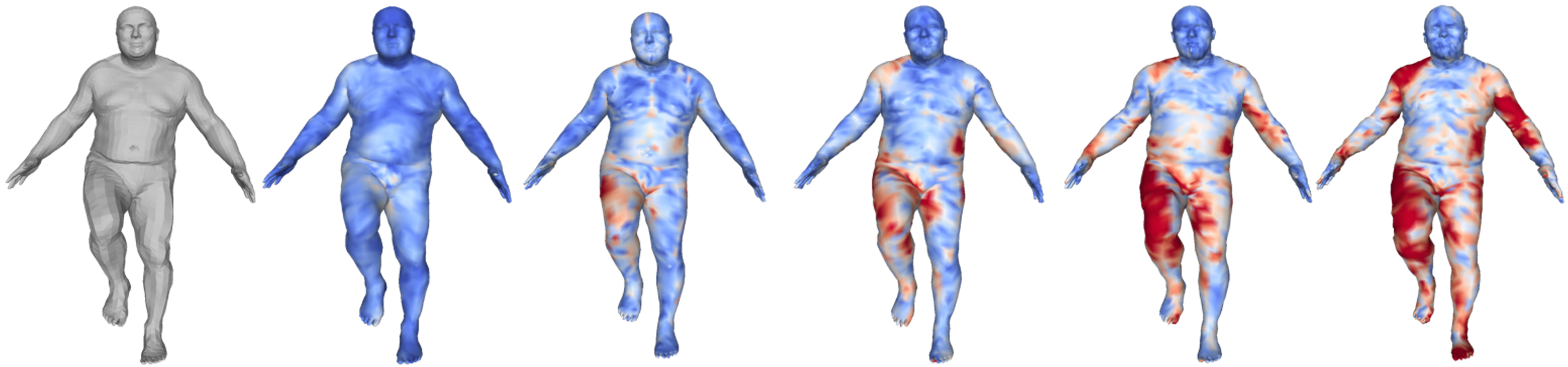
Spiral CNN Edge CNN GCNN ACNN MoNet

- **Low** Reconstruction Accuracy
- **Low** Generalizability

Our Method:

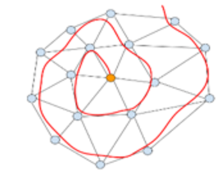


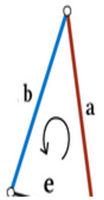
- **High** Reconstruction Accuracy



Groundtruth **Ours** Neural3DMM Spectral (Cheb) MoNet FeaST

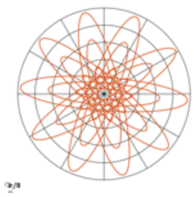
Existing Methods:

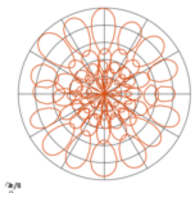

Spiral CNN



Edge CNN

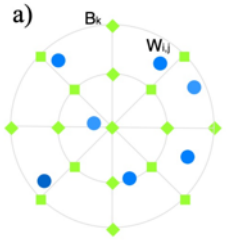

GCNN

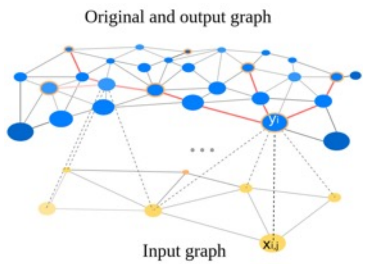

ACNN


MoNet

- **Low** Reconstruction Accuracy
- **Low** Generalizability

Our Method:





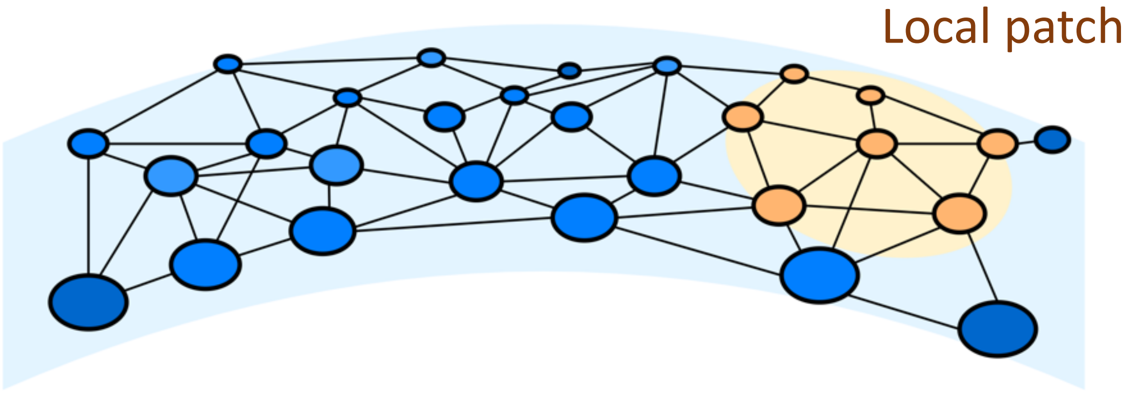
- **High** Reconstruction Accuracy
- **High** Generalizability
- Provides natural analogs of 2D operations:
  - Stride, receptive field, pooling, unpooling, etc

Insights

A mesh is a discretization of a continuous space.

A continuous kernel can be shared in a continuous space.

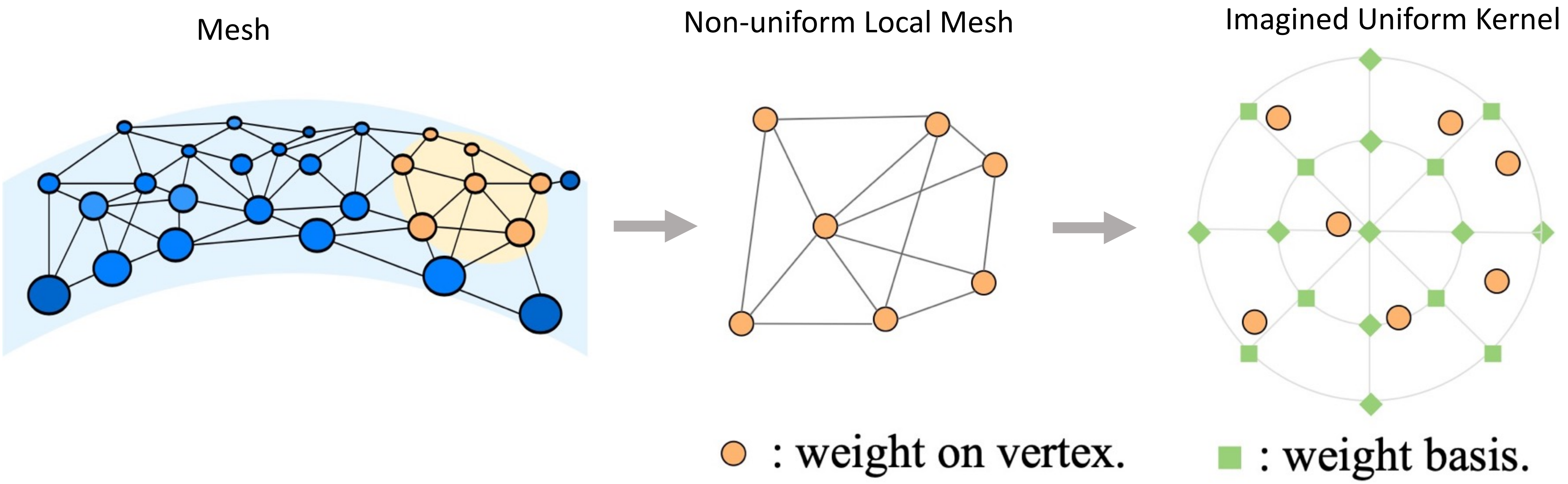
A discretized kernel can be sampled from a continuous kernel.



Core Idea:

1. Sample weight from a shared uniform kernel.

2. Learned sampling coefficients.



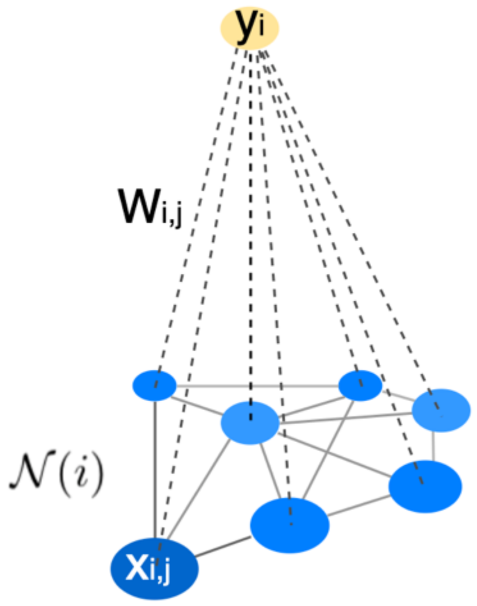
Convolution Operation:

Each Convolution Layer has one weight basis $B = \{\mathbf{B}_k\}_{k=1}^{M}, \mathbf{B}_k \in \mathbb{R}^{I \times O}$

Each edge j for a local vertex i has coefficients $A_{i,j} = \{\alpha_{i,j,k}\}_{k=1}^{M}, \alpha \in \mathbb{R}$:

The weight $\mathbf{W}_{i,j}$ on each edge is computed as $\mathbf{W}_{i,j} = \sum_{k=1}^{M} \alpha_{i,j,k}\mathbf{B}_k$

The output feature is computed as $\mathbf{y}_i = \sum_{x_{i,j} \in \mathcal{N}(i)} \mathbf{W}_{i,j}\mathbf{x}_{i,j} + \mathbf{b}$



B and $A_{i,j}$ are training parameters, shared across the dataset.

Pooling Operation

Observation: local density is non-uniform

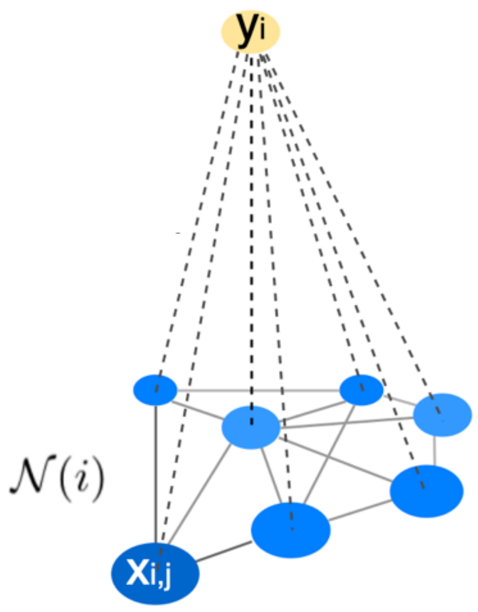
Solution: Monte Carlo Integration with learned density coefficients

Formulation:

Each local vertex j has a density coefficient $$\rho'_{i,j} = \frac{|\rho_{i,j}|}{\sum_{j=1}^{E_i} |\rho_{i,j}|}$$

The output feature is computed as $$\mathbf{y}_i = \sum_{j \in \mathcal{N}(i)} \rho'_{i,j} \mathbf{x}_{i,j}$$

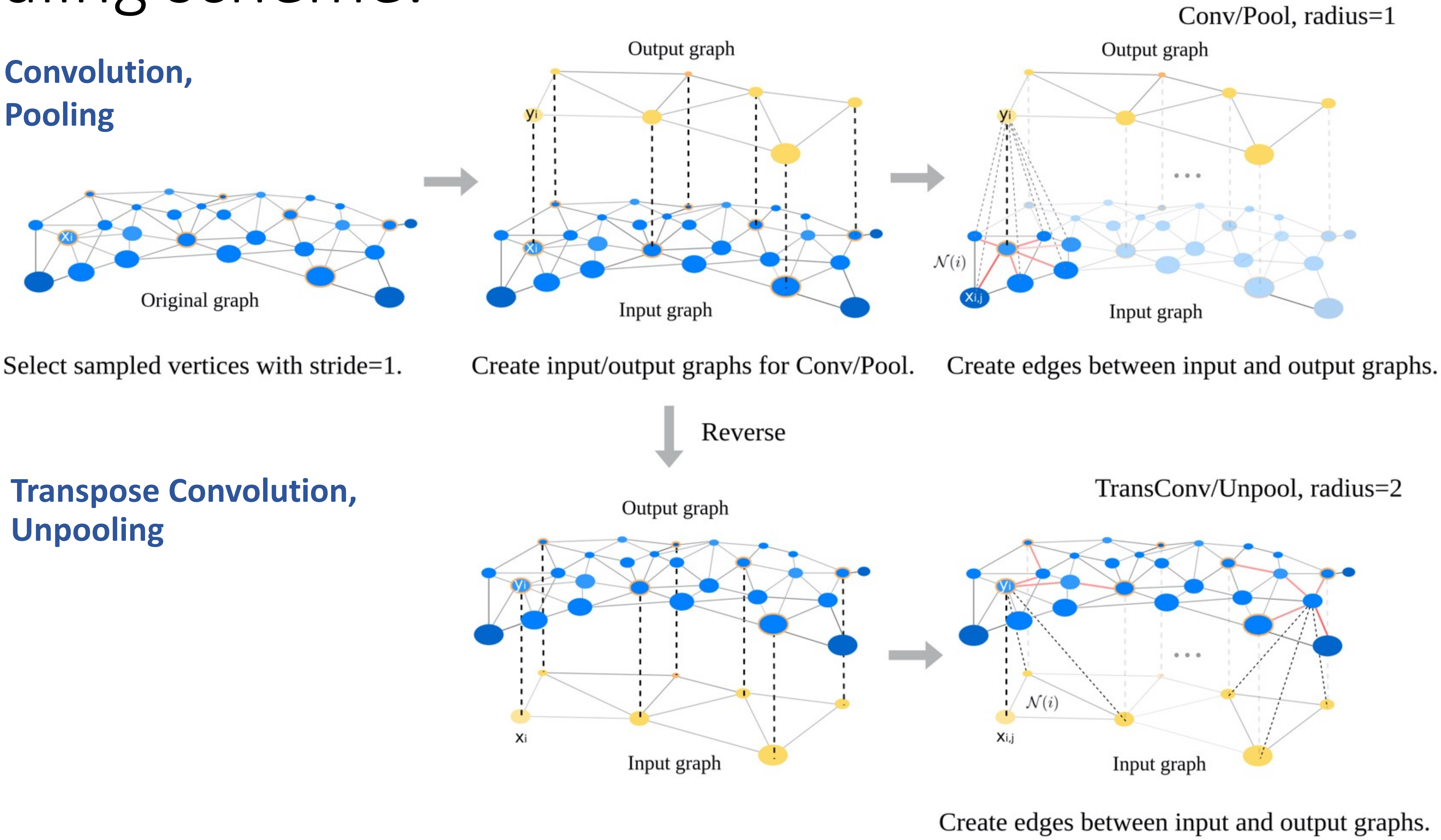
$\rho_{i,j}$ are training parameters, shared across the dataset.



Scaling scheme:

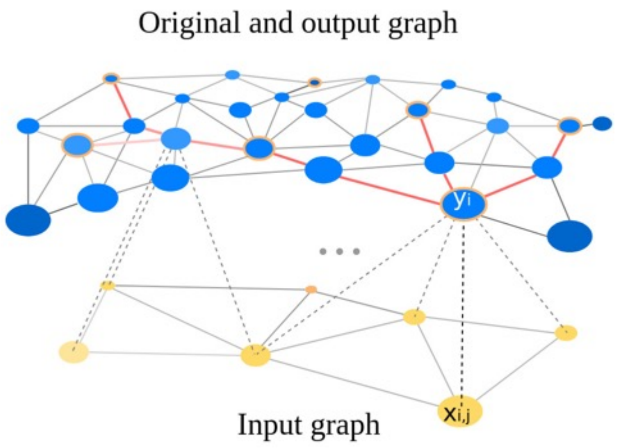
Convolution, Pooling

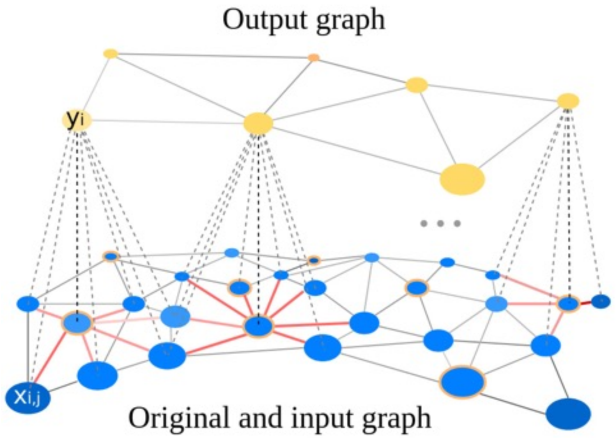
Transpose Convolution, Unpooling



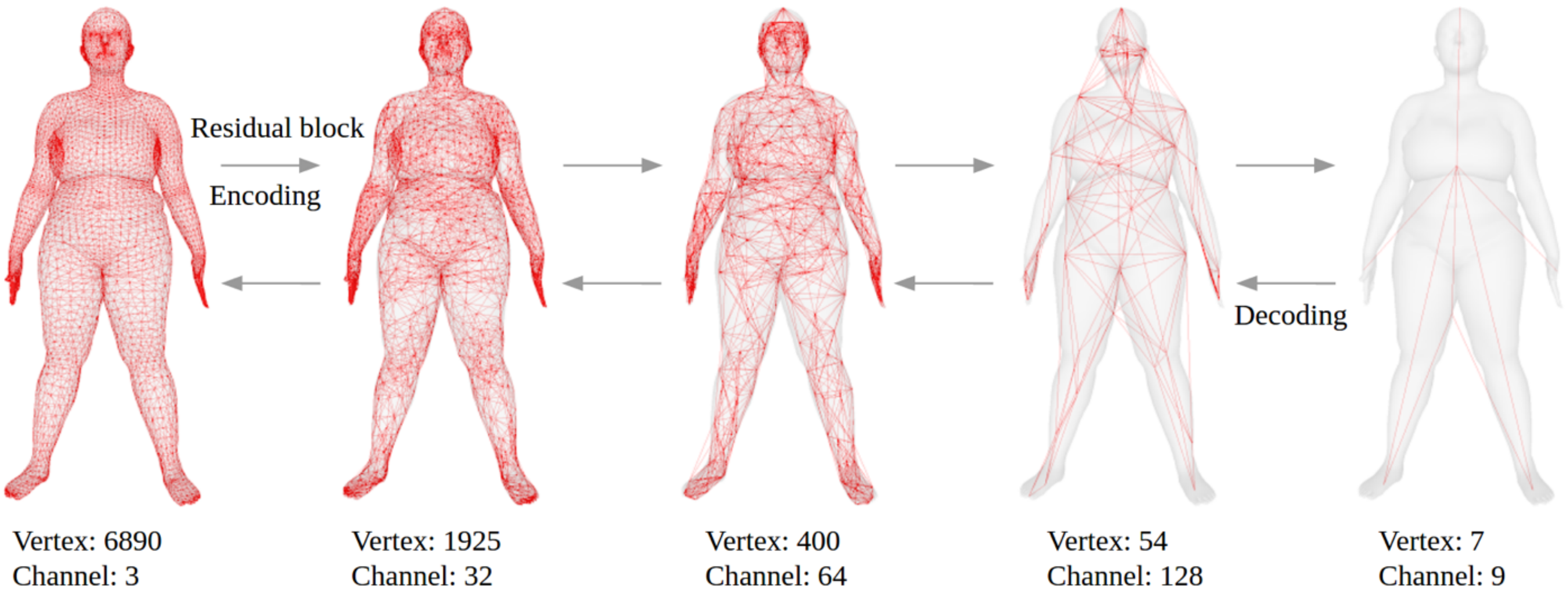
Operations Analog to Regular CNN

Down-sampling	Up-sampling	Attributes
vcConv	vcTransposeConv	(Stride, kernel radius, basis size, in_channel, out_channel, dilation)
vdPool	vdUnpool	(Stride)
vdDownResidual	vdUpResidual	(In_channel, out_channel)

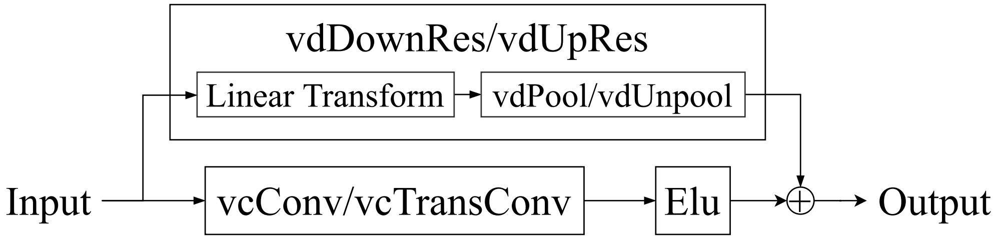




Fully Convolutional architecture

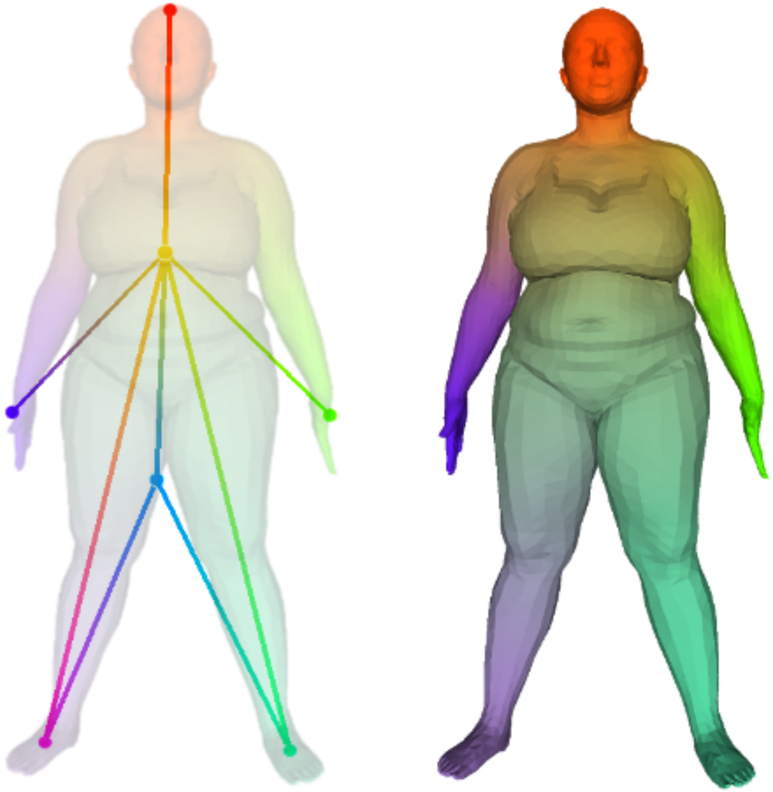


Residual block



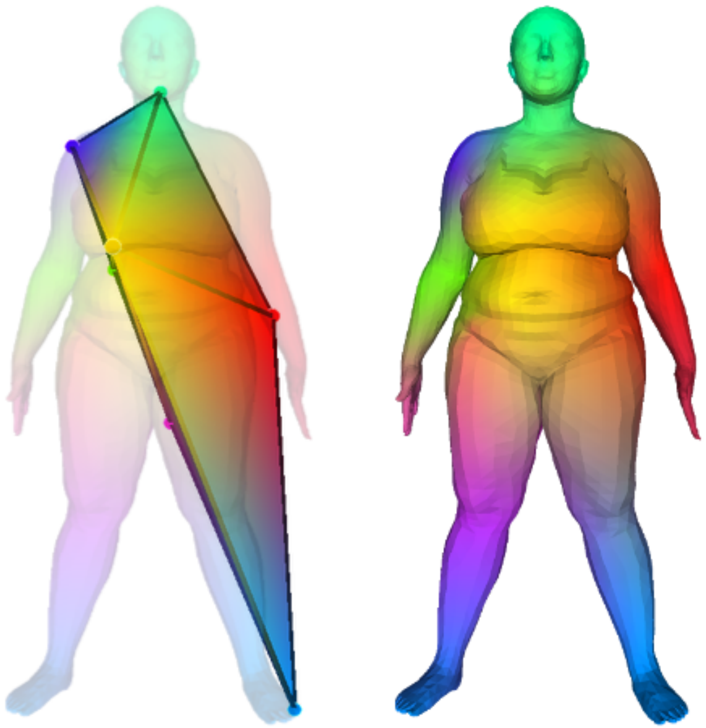
Localized Latent Feature

Ours



Middle layer graph Receptive field

Previous Methods

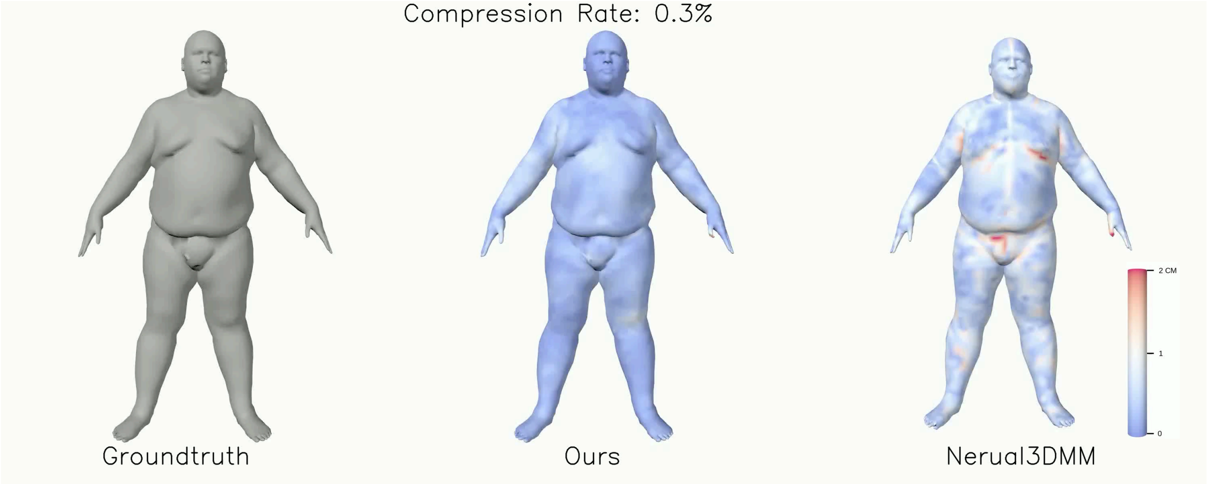


Middle layer graph Receptive field

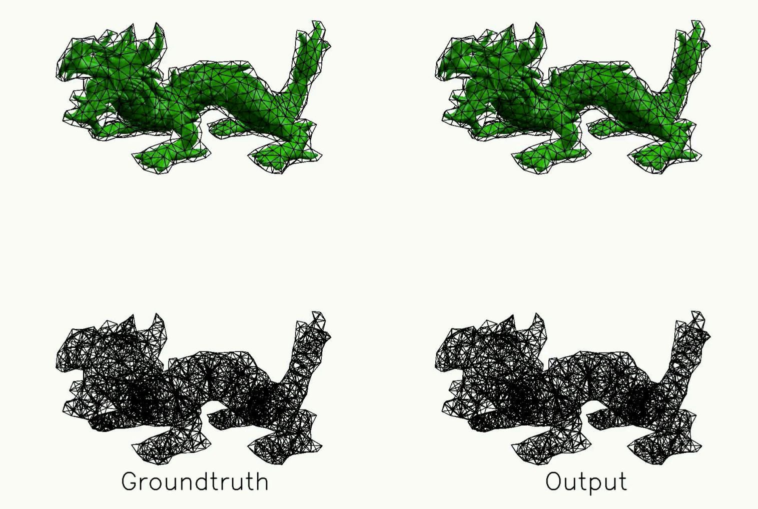
Quadric Mesh Simplification

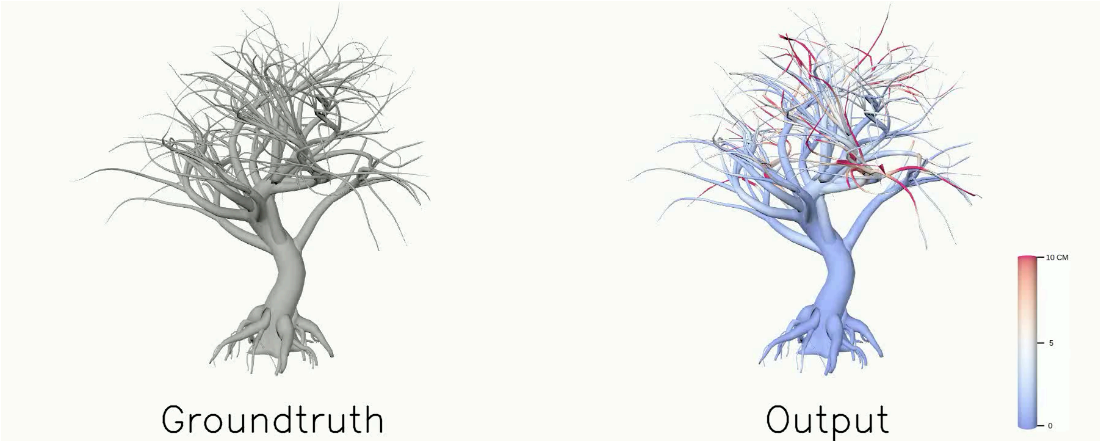
Results:

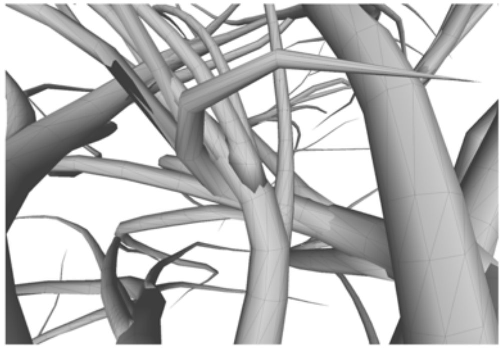
- High reconstruction accuracy



- Work for all types of mesh (Tetrahedrons, Non-Manifold Mesh)



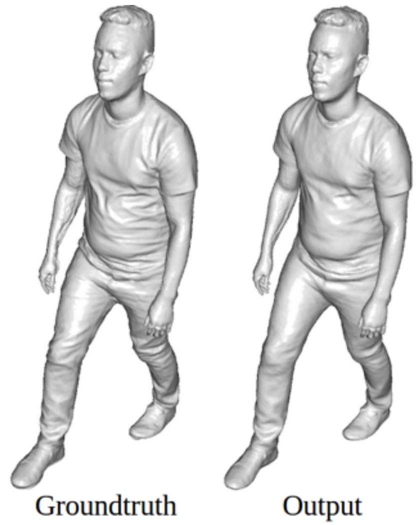




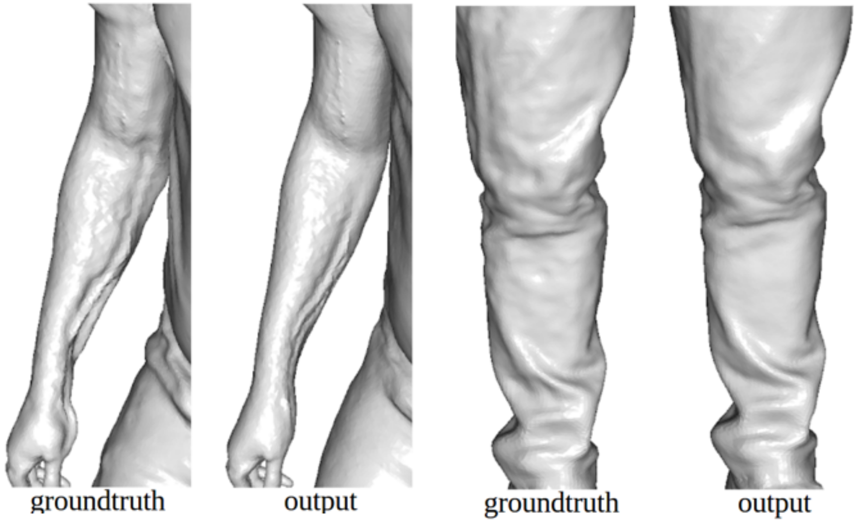
Zoom-in non-manifold structures

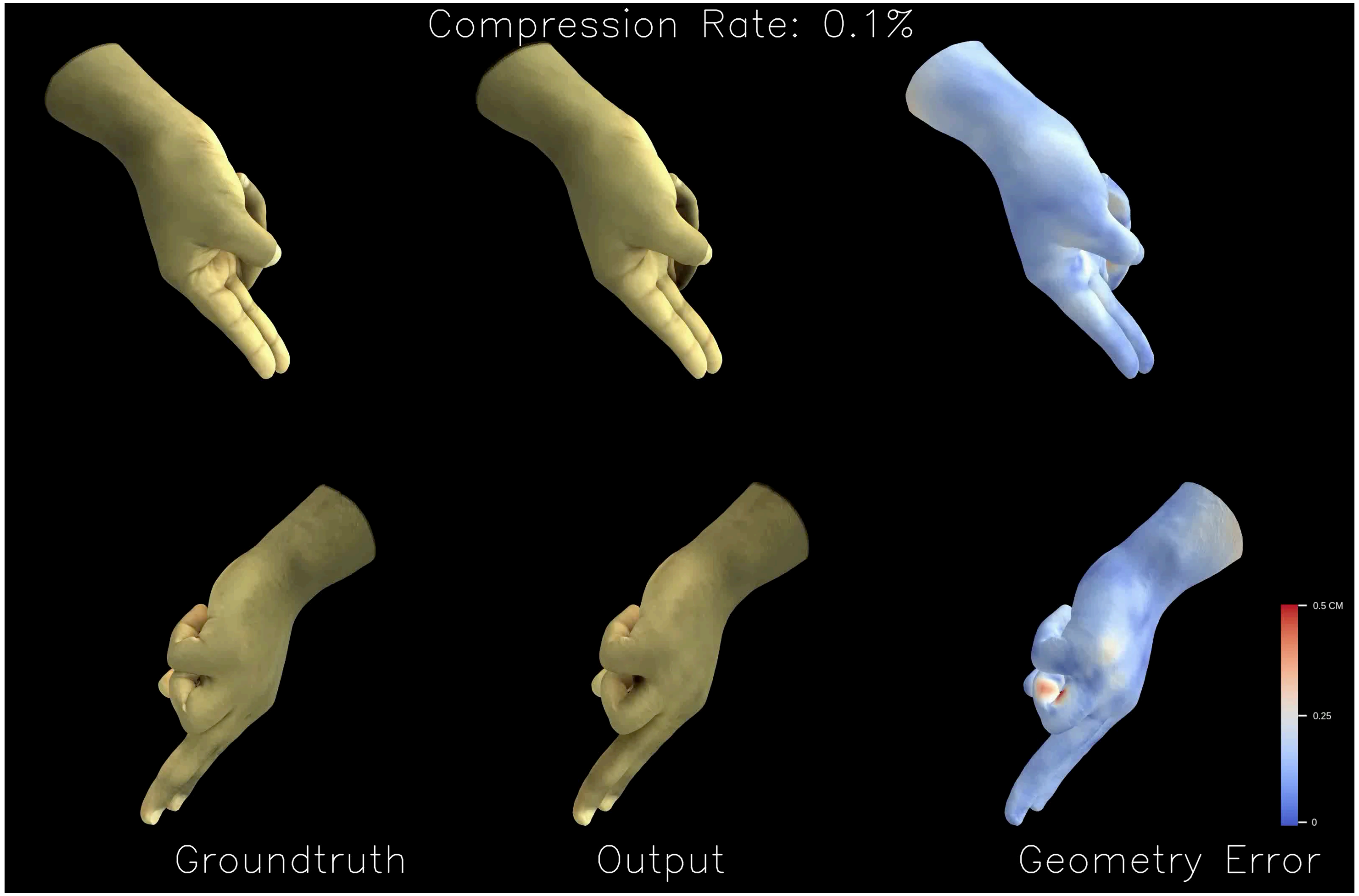
Results:

- Efficient for high-resolution mesh



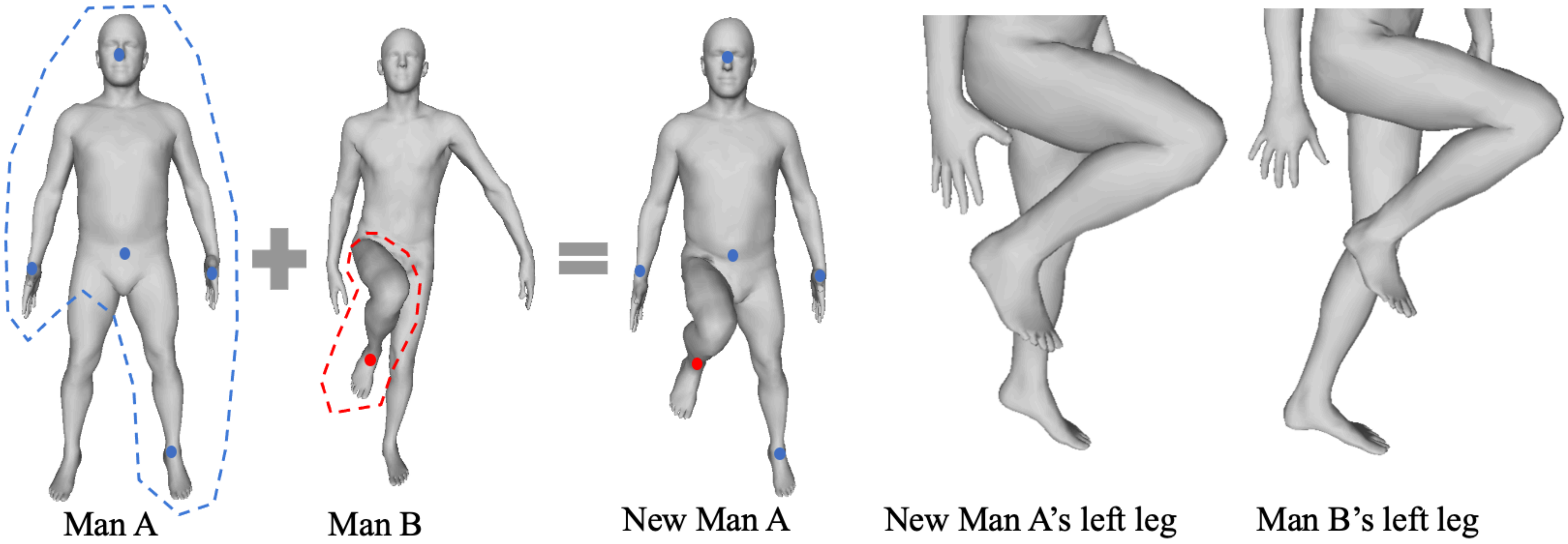
+ 150,000 vertices per mesh





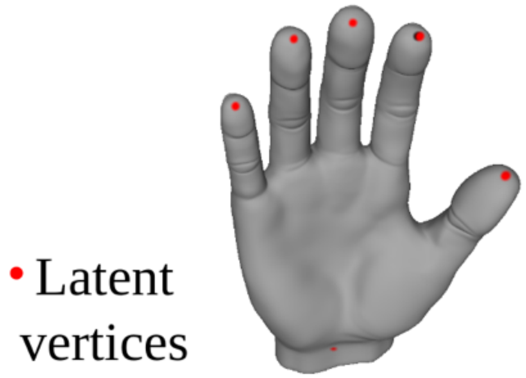
Results:

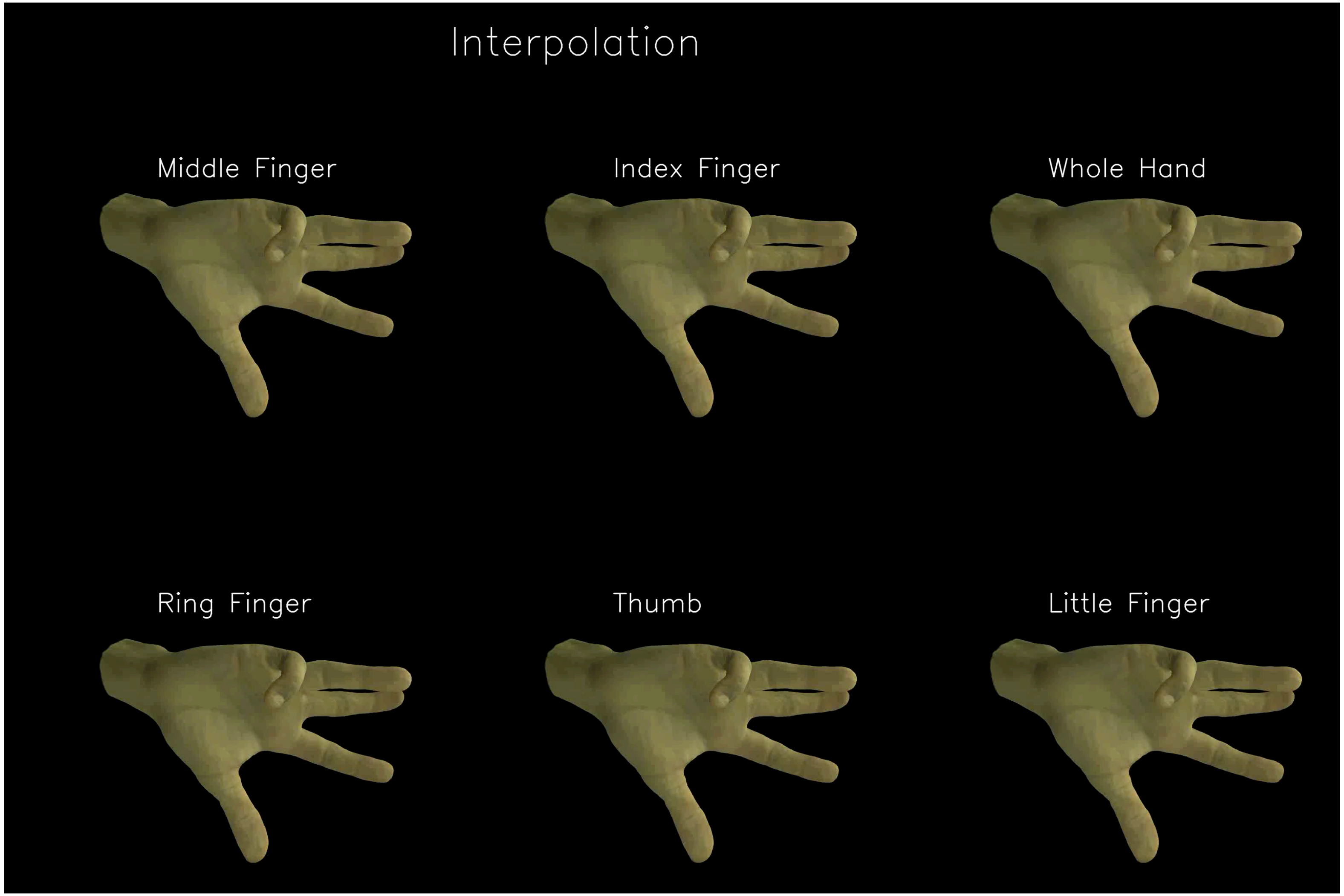
- Localized Latent Space Interpolation



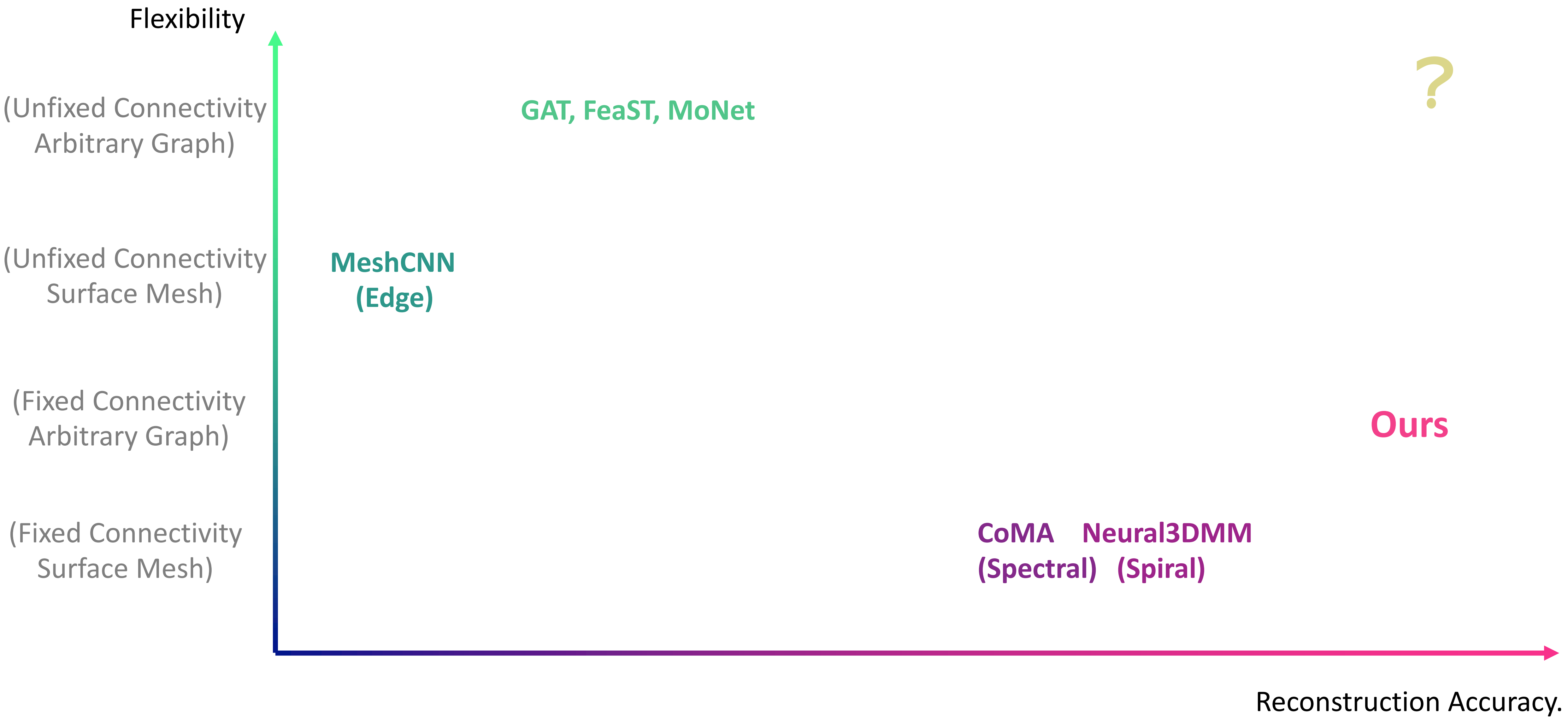
Results:

- Localized Latent Space Interpolation





Limitation and Future Works:



Conclusion

Autonomous 3D Avatar

- Human Digitizing

  Face, body, hair hand …

- Motion Synthesis

  Recurrent synthesis, motion inbetween,

  2D sketch to 3D animation,

  secondary motion

- Deep Representation Learning

  Mesh, hair, rotation

- AR & VR

  AR texture, VR conference, telepresence

Research Webpage:

https://zhouyisjtu.github.io

Questions?