EUROGRAPHICS 2018

Deep Learning for Graphics

Beyond Image Data

Niloy Mitra	Iasonas Kokkinos	Paul Guerrero	Vladimir Kim	Kostas Rematas	Tobias Ritschel
UCL	UCL/Facebook	UCL	Adobe Research	U Washington	UCL

Course Overview

- Part I: Introduction and ML Basics
- Part II: Supervised Neural Networks: Theory and Applications
- Part III: Unsupervised Neural Networks: Theory and Applications
- **Part IV: Beyond Image Data**

Motivating Applications

Motivating Applications

- 3D modeling, retrieval, classification for AR and VR

Motivating Applications

- 3D modeling, retrieval, classification for AR and VR

- Joint multi-modal understanding

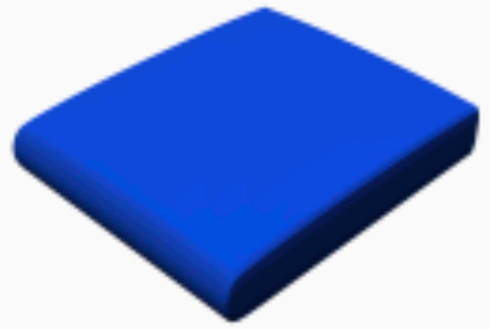
Motivating Applications

- 3D modeling, retrieval, classification for AR and VR
- Joint multi-modal understanding
- Semantic 3D reconstruction

Motivating Applications

- 3D modeling, retrieval, classification for AR and VR
- Joint multi-modal understanding
- Semantic 3D reconstruction
- Animation, rendering, …

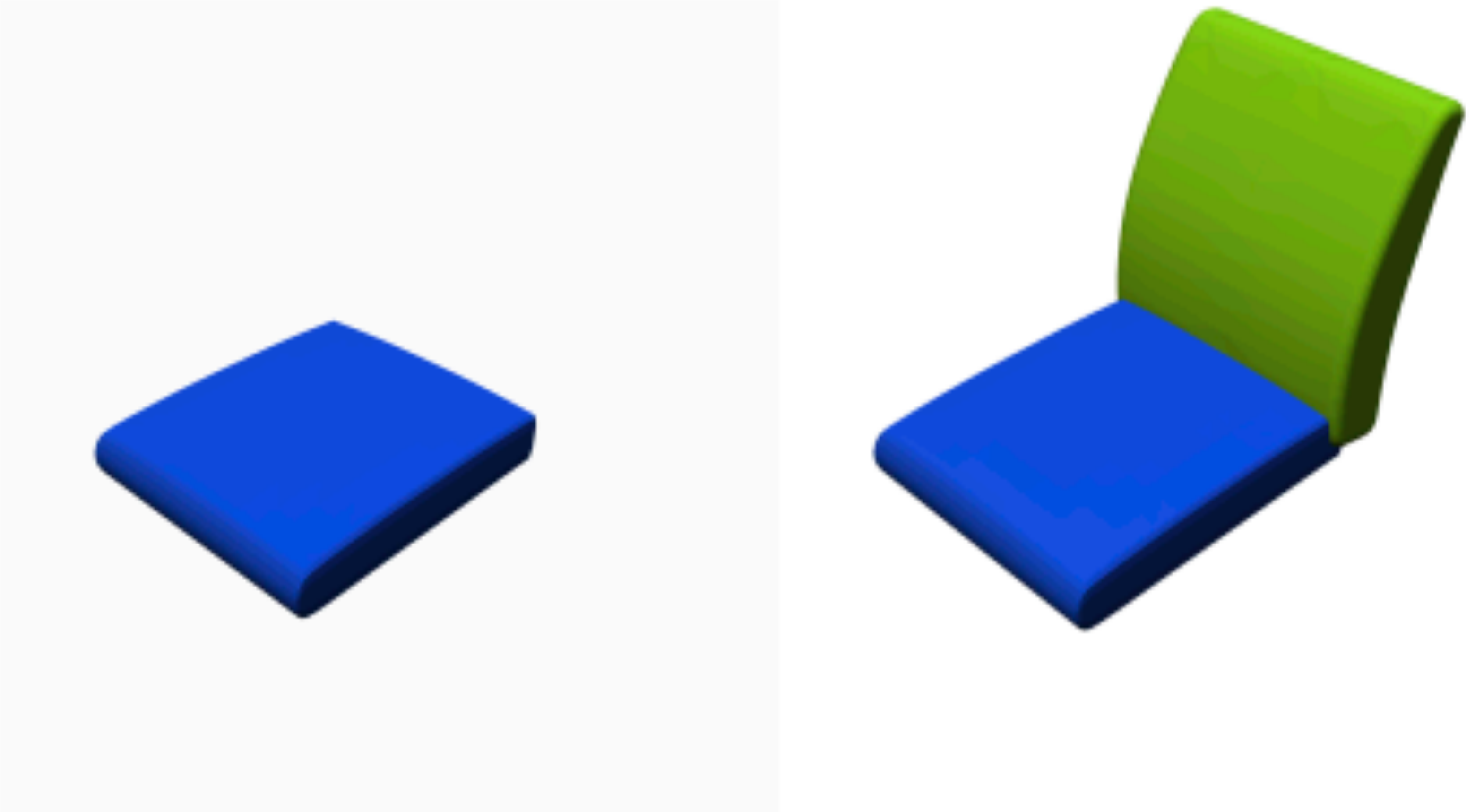
Motivating Applications



Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

[Sung et al. 2017]

Motivating Applications



Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

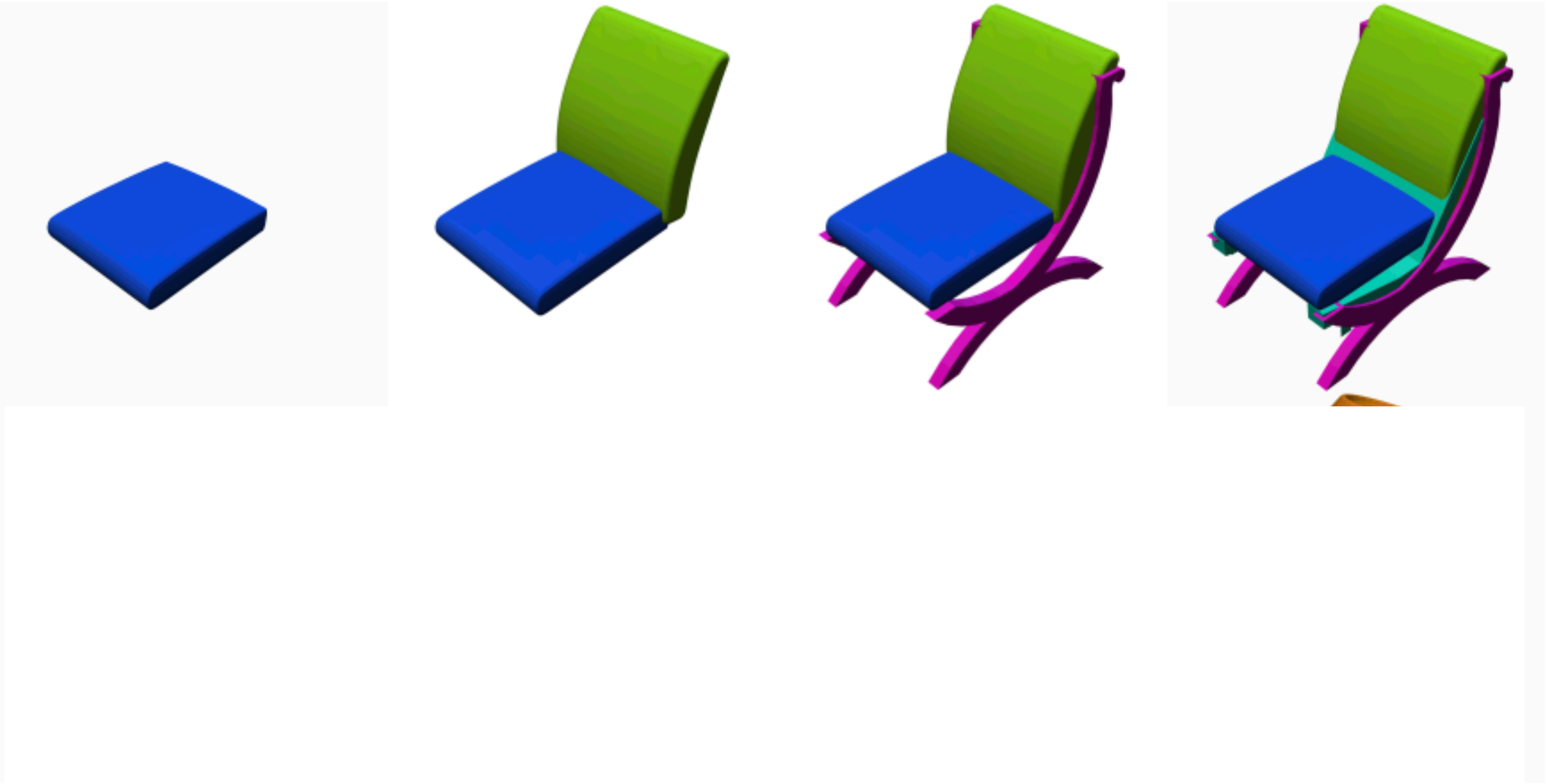
[Sung et al. 2017]

Motivating Applications

Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

[Sung et al. 2017]

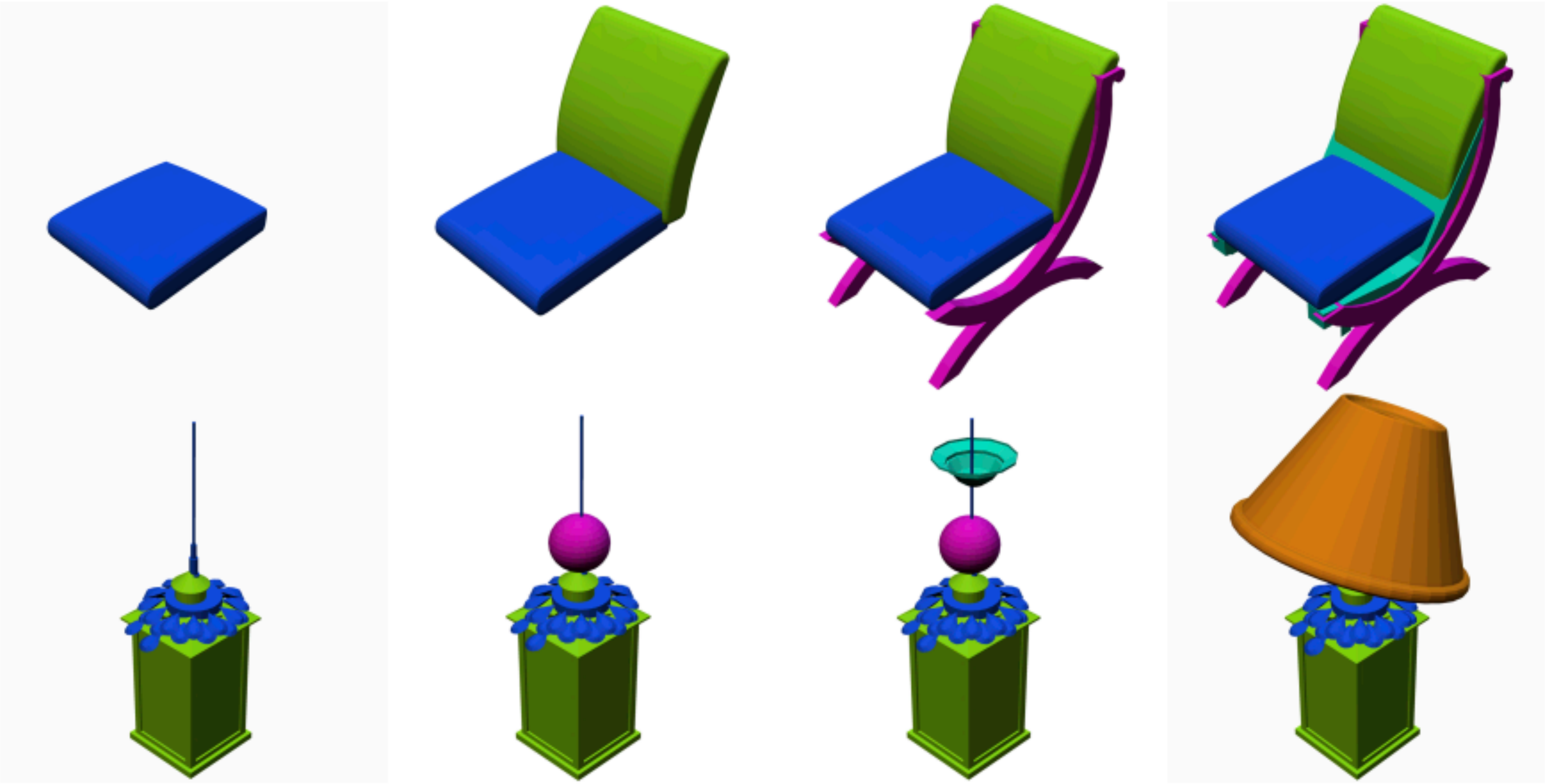
Motivating Applications



Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

[Sung et al. 2017]

Motivating Applications



Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

[Sung et al. 2017]

CrossLink: Linking Images and 3D Models

[Heuting et al. 2015]

CrossLink: Linking Images and 3D Models

[Heuting et al. 2015]

Motivating Applications

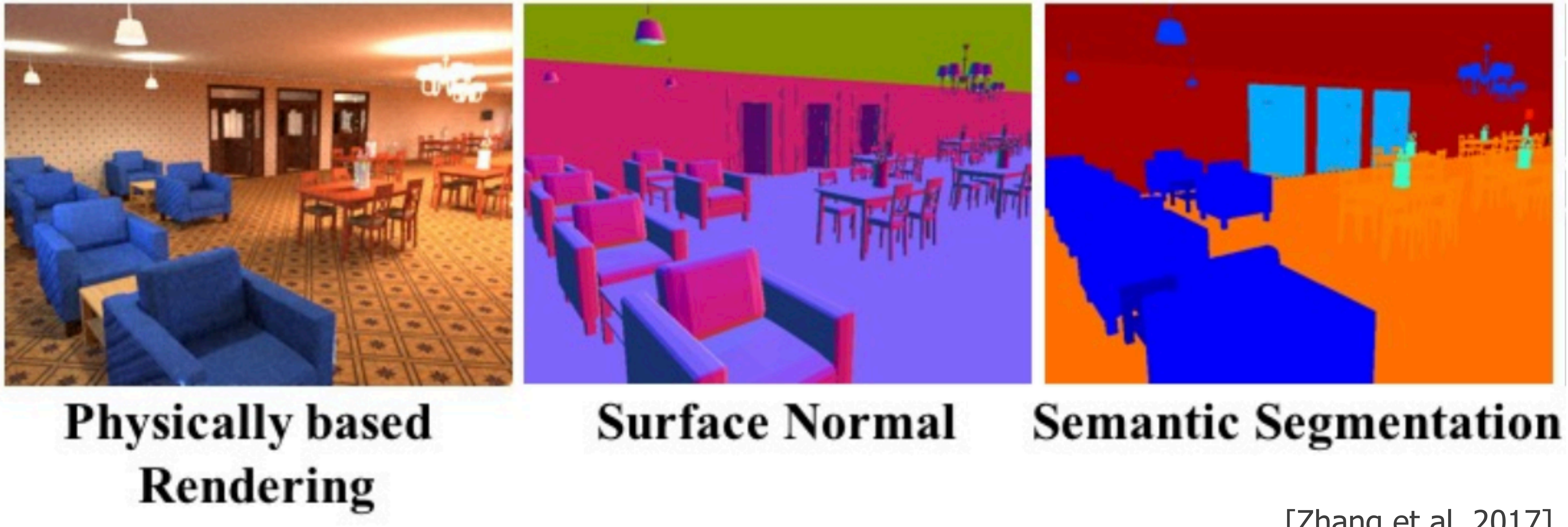
understanding 3D shapes can benefit image understanding

Physically based Rendering

[Zhang et al. 2017]

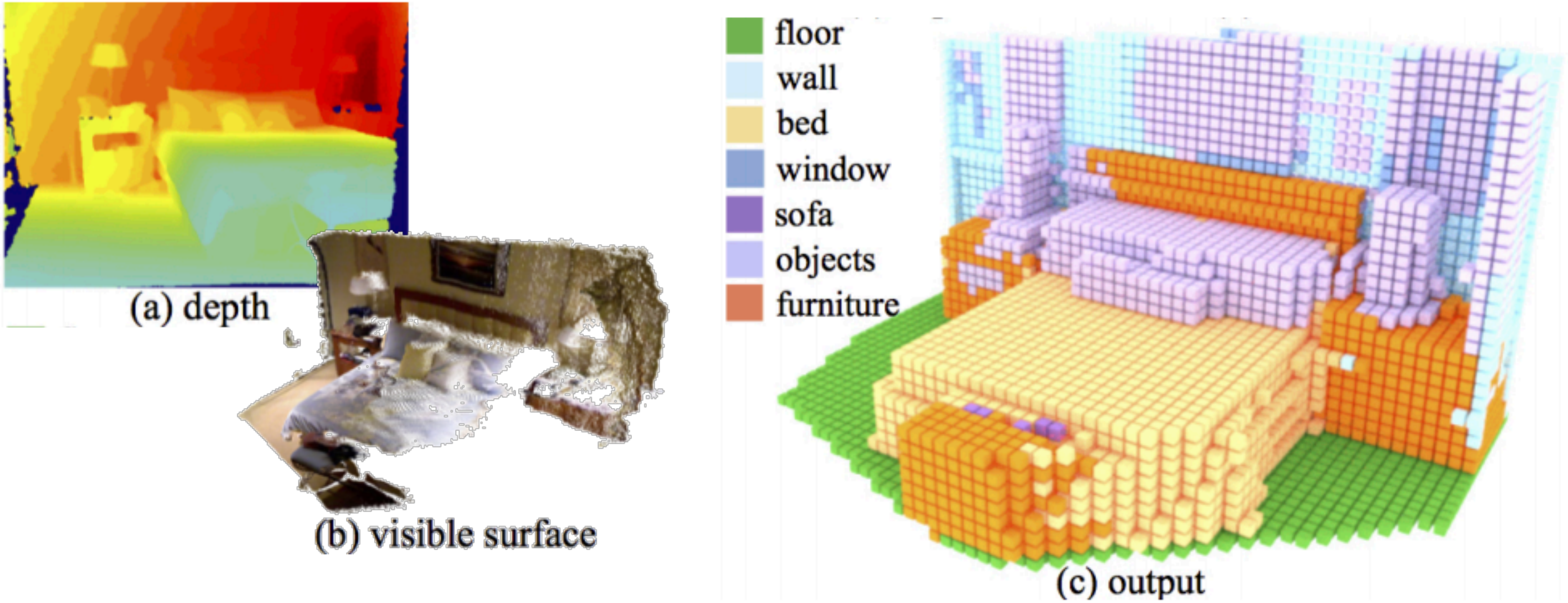
Motivating Applications

understanding 3D shapes can benefit image understanding



[Zhang et al. 2017]

Motivating Applications: Semantic Scene Understanding



[Song et al. 2017]

Motivating Applications: Semantic Scene Understanding

[Kelly et al. 2017]

Motivating Applications: Semantic Scene Understanding

[Kelly et al. 2017]

Motivating Applications: Semantic Scene Understanding

[Kelly et al. 2017]

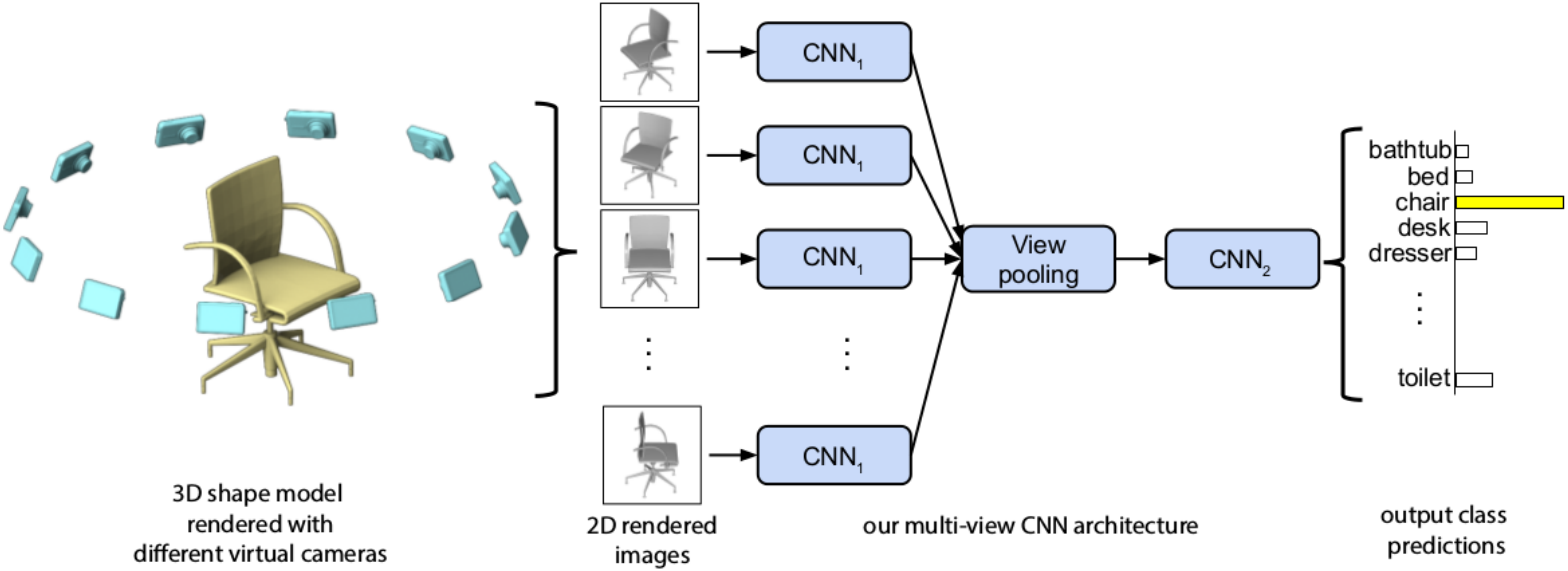
Representation for 3D

- Image-based
- Volumetric
- Point-based
- Surface-based

Representation for 3D

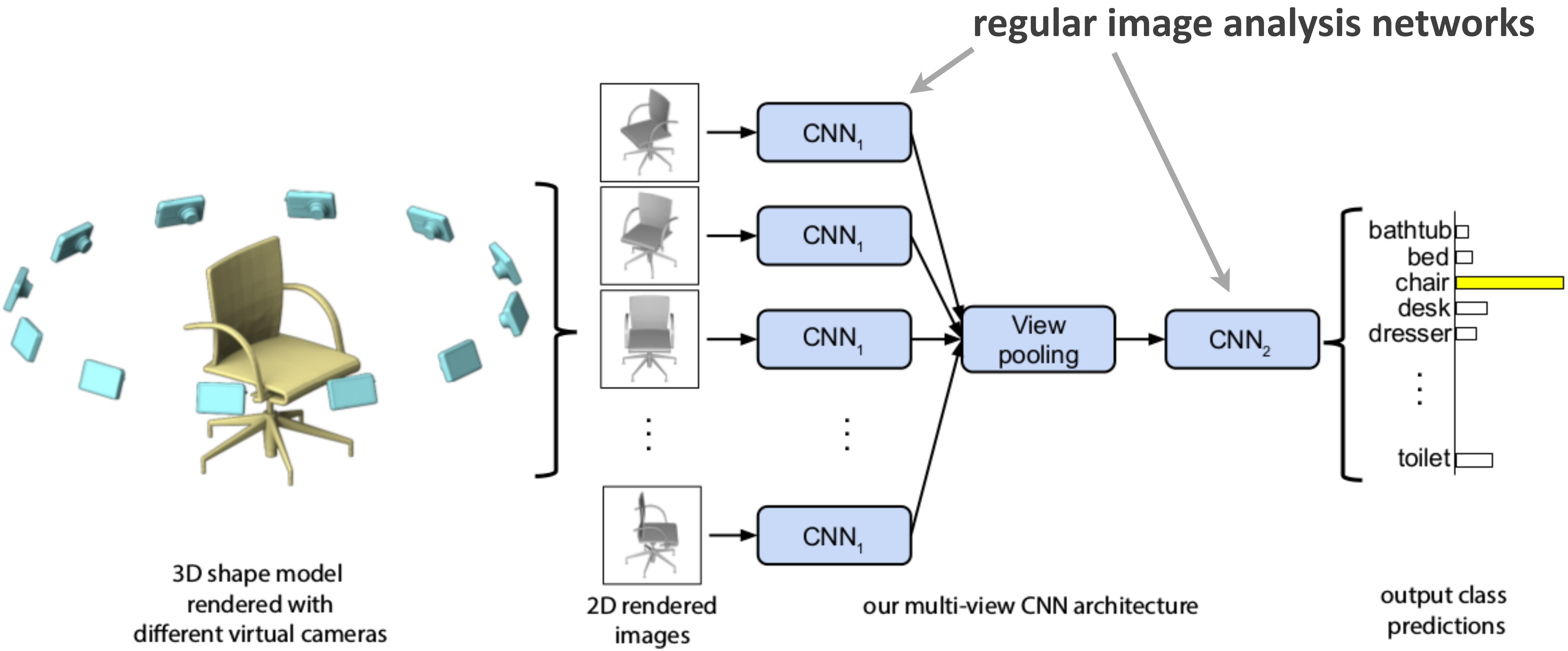
- **Image-based**
- Volumetric
- Point-based
- Surface-based

Representation for 3D: Multi-view CNN



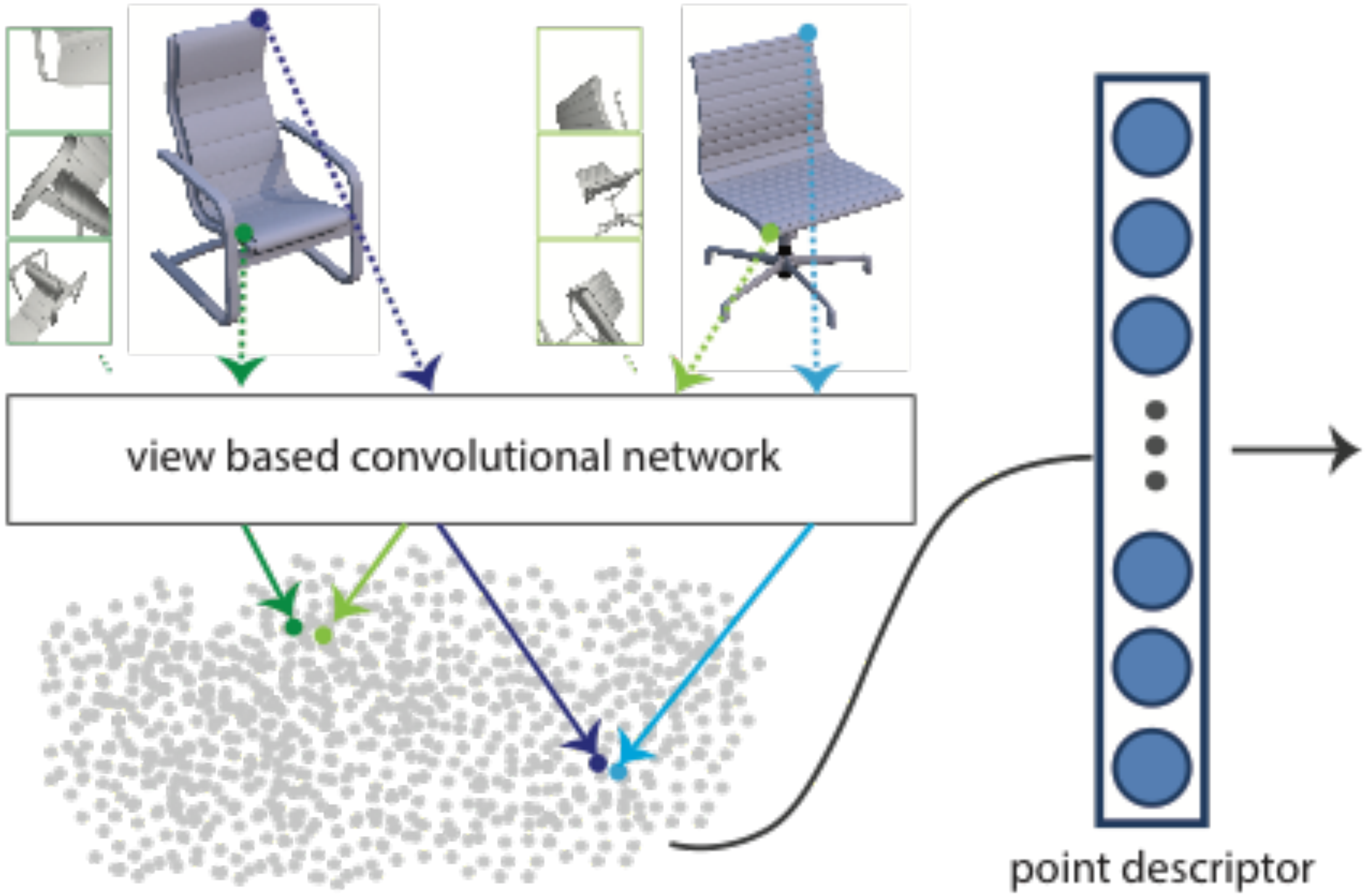
[Kalogerakis et al. 2015]

Representation for 3D: Multi-view CNN



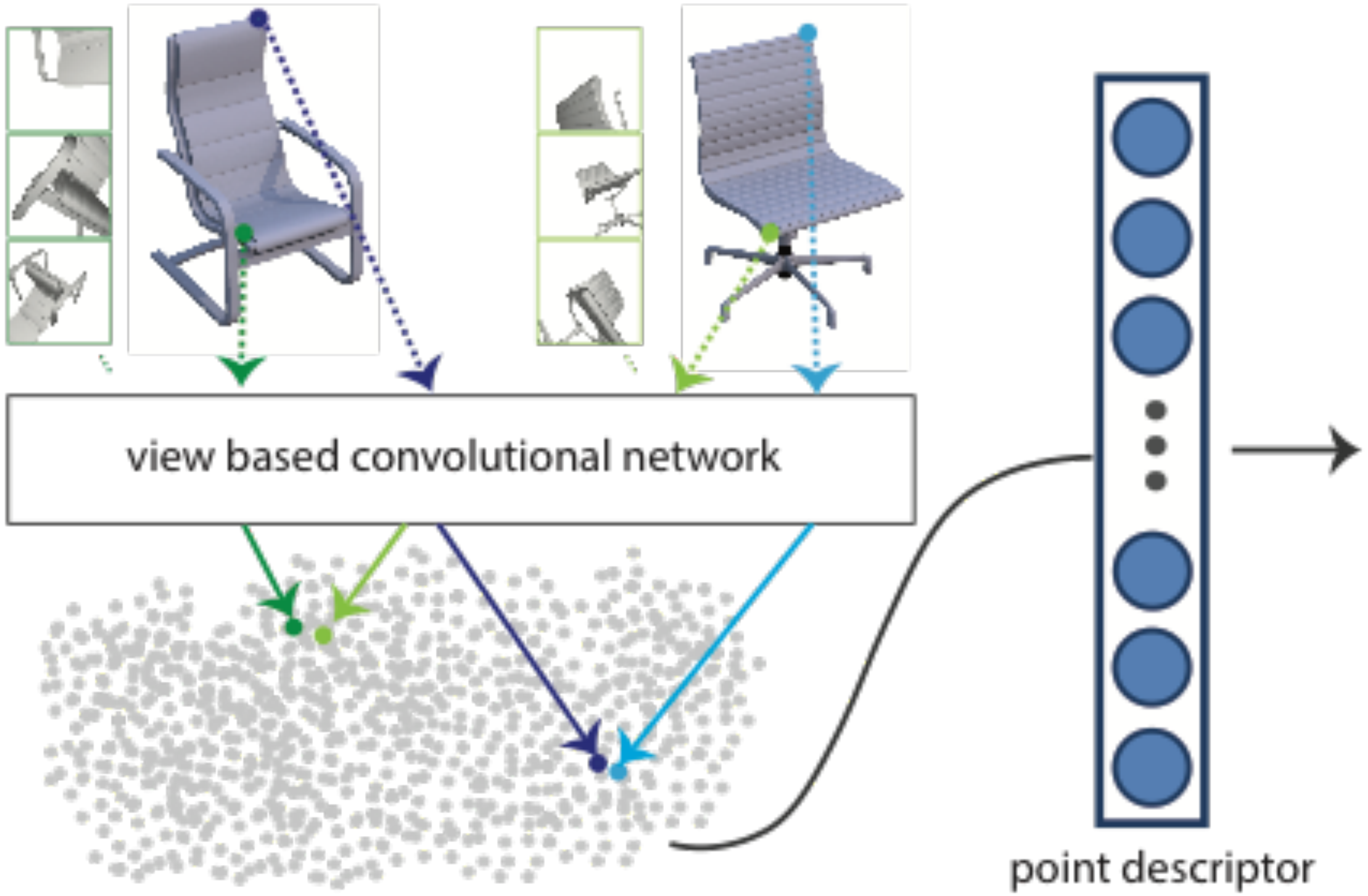
[Kalogerakis et al. 2015]

Representation for 3D: Local Multi-view CNN



[Huang et al. 2018]

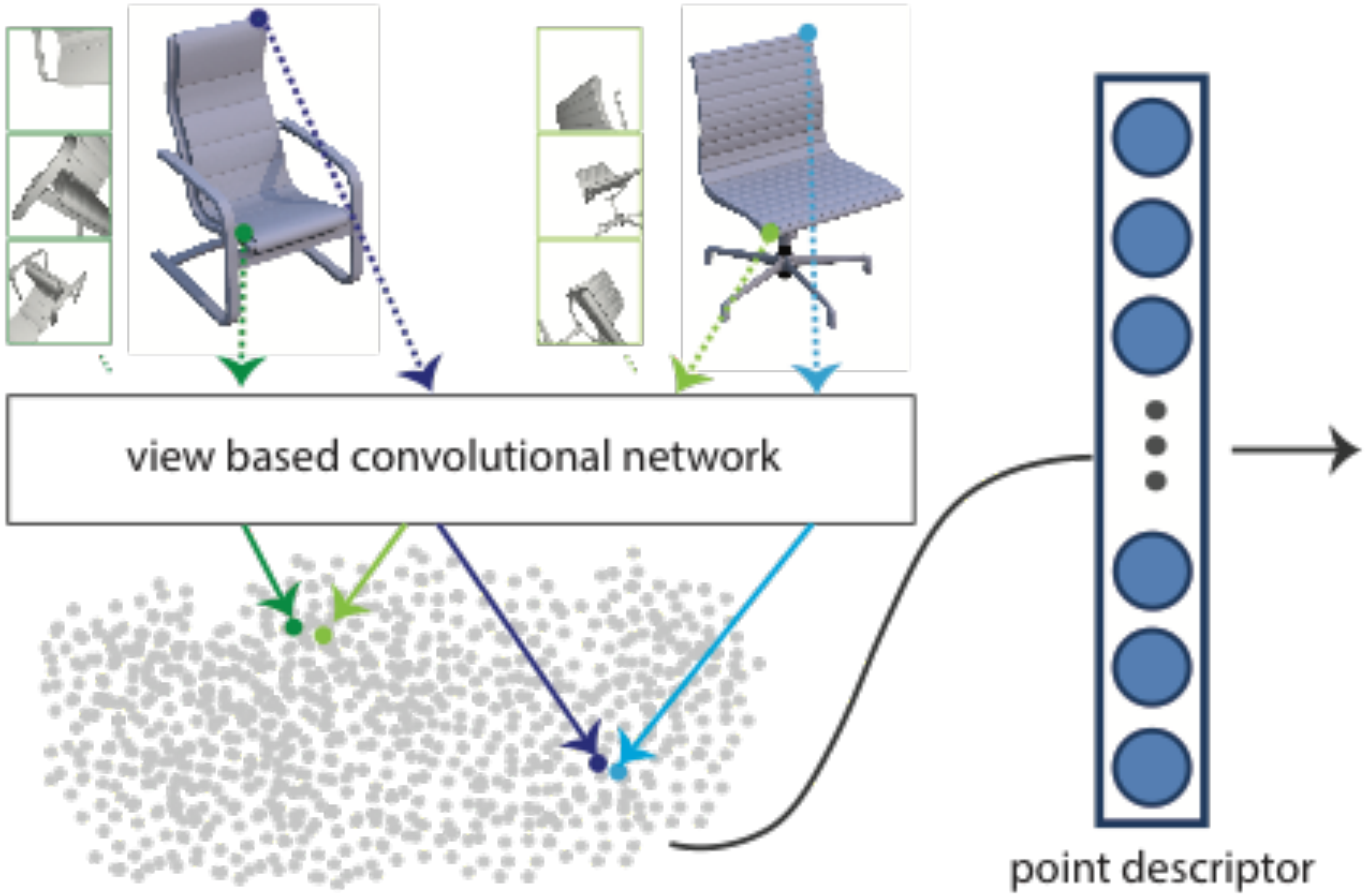
Representation for 3D: Local Multi-view CNN



Segmentation
Correspondence
Feature matching
Predicting semantic functions

[Huang et al. 2018]

Representation for 3D: Local Multi-view CNN



Segmentation
Correspondence
Feature matching
Predicting semantic functions

localized renderings for point-wise features

[Huang et al. 2018]

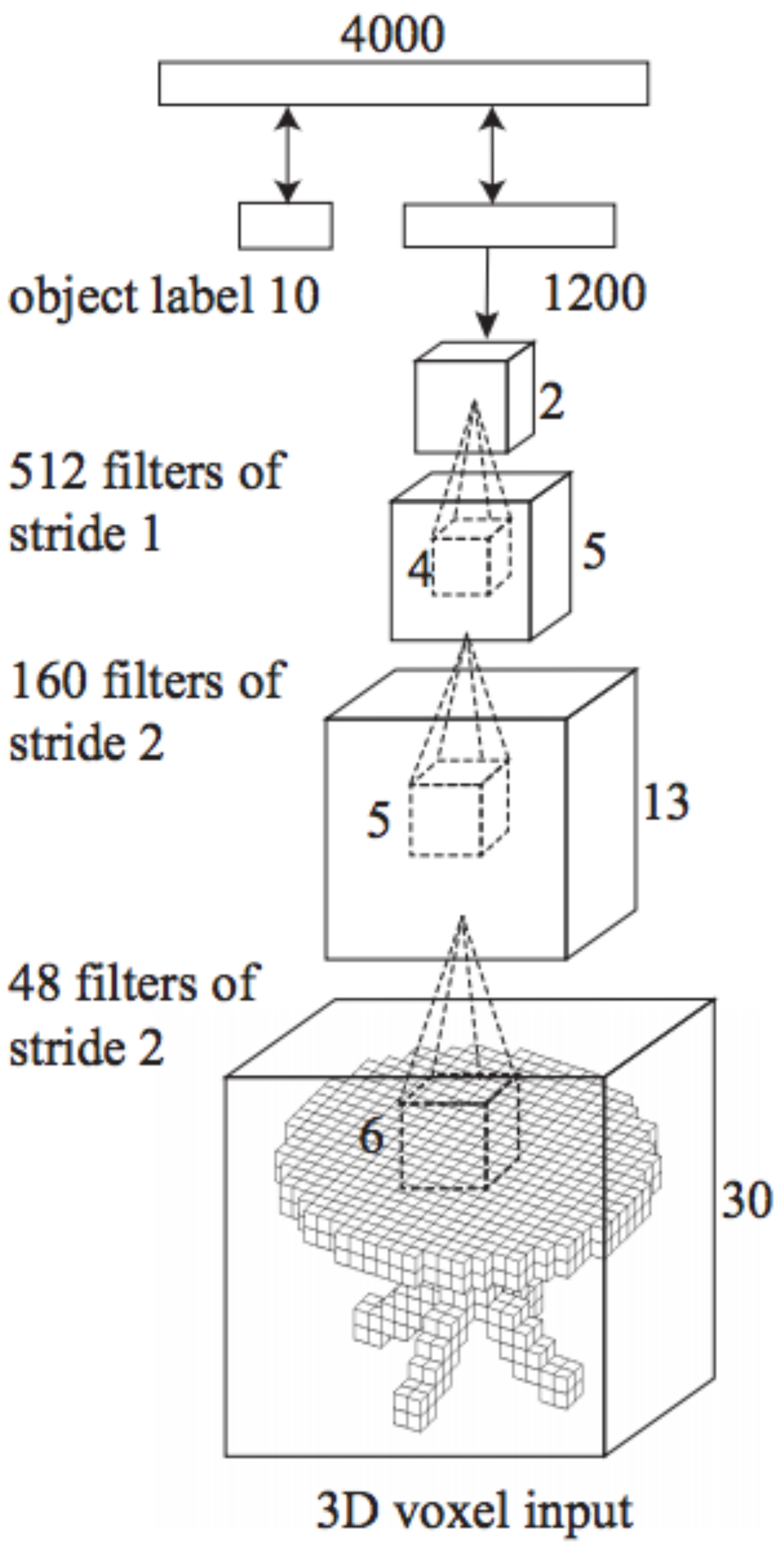
Representation for 3D

- **Image-based**
  - **PROS**: directly use image networks, good performance
  - **CONS**: rendering is slow and memory-heavy, not very geometric
- Volumetric
- Point-based
- Surface-based

Representation for 3D

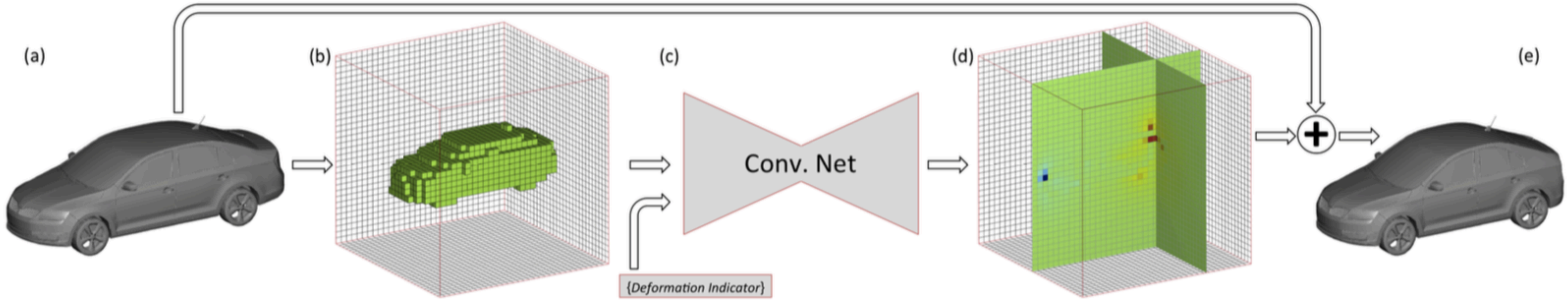
- Image-based
- **Volumetric**
- Point-based
- Surface-based

Representation for 3D: Volumetric



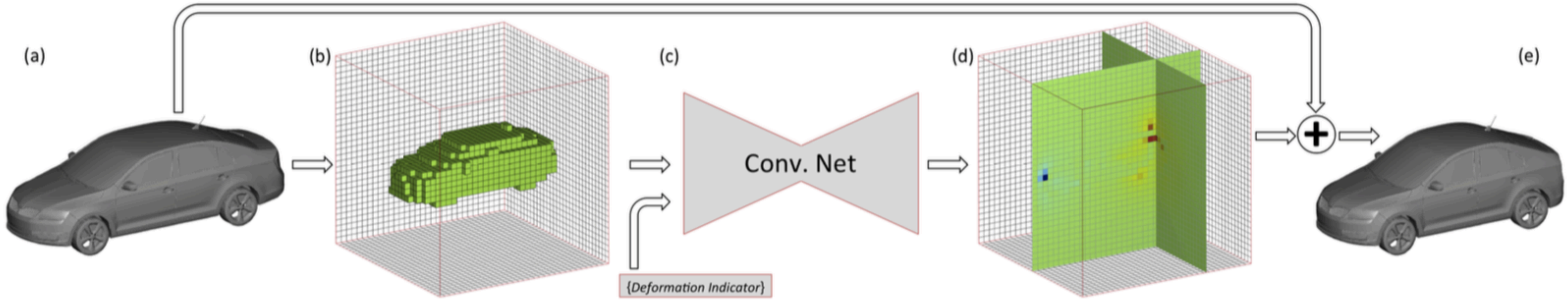
[Xiao et al. 2014]

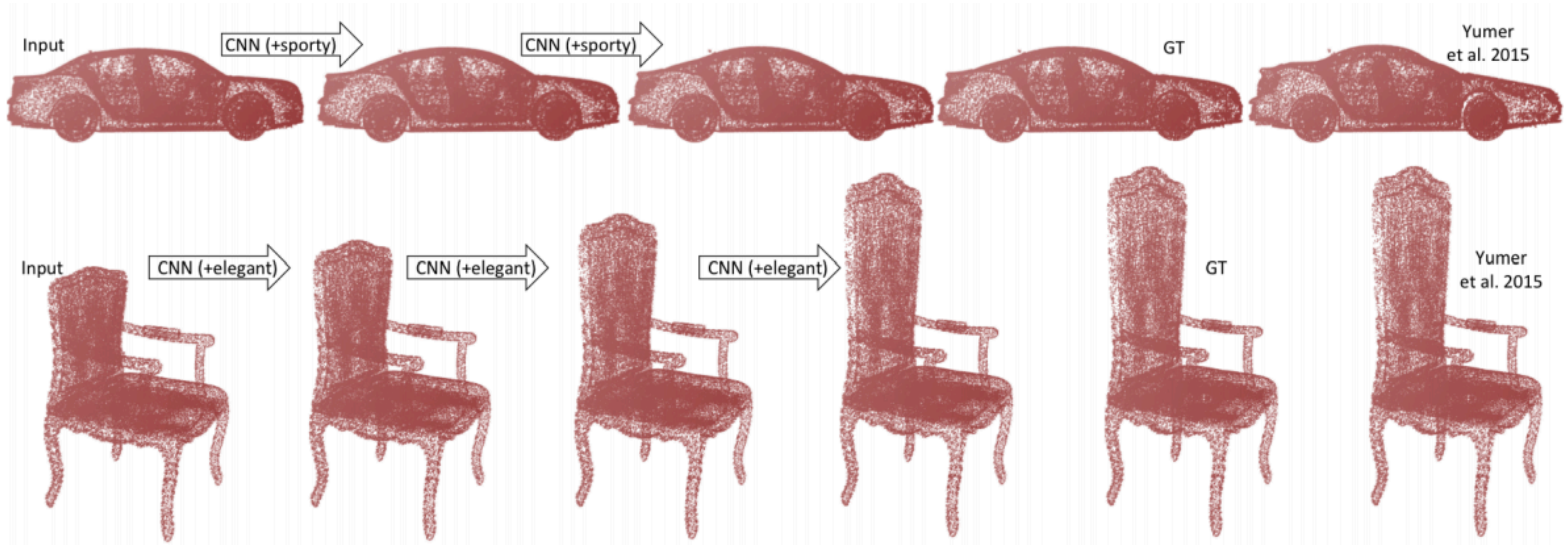
Representation for 3D: Volumetric Deformation



[Yumer et al. 2014]

Representation for 3D: Volumetric Deformation





[Yumer et al. 2014]

Efficient Volumetric Datastructures

[Wang et al. 2017]

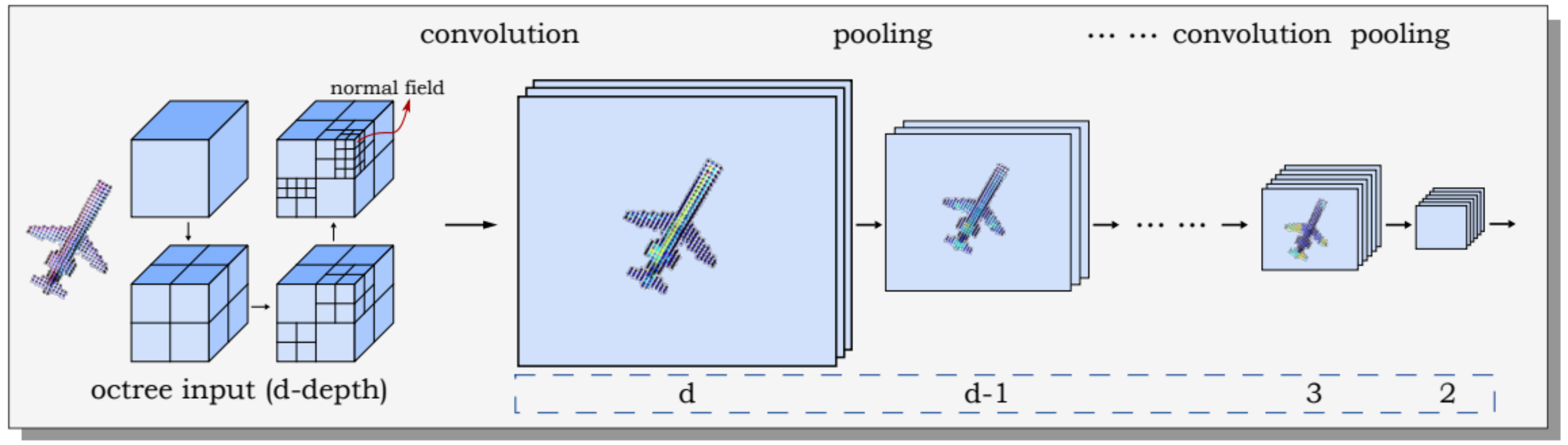
Efficient Volumetric Datastructures

Generator / Decoder

Volumetric (Up-) Convolutions

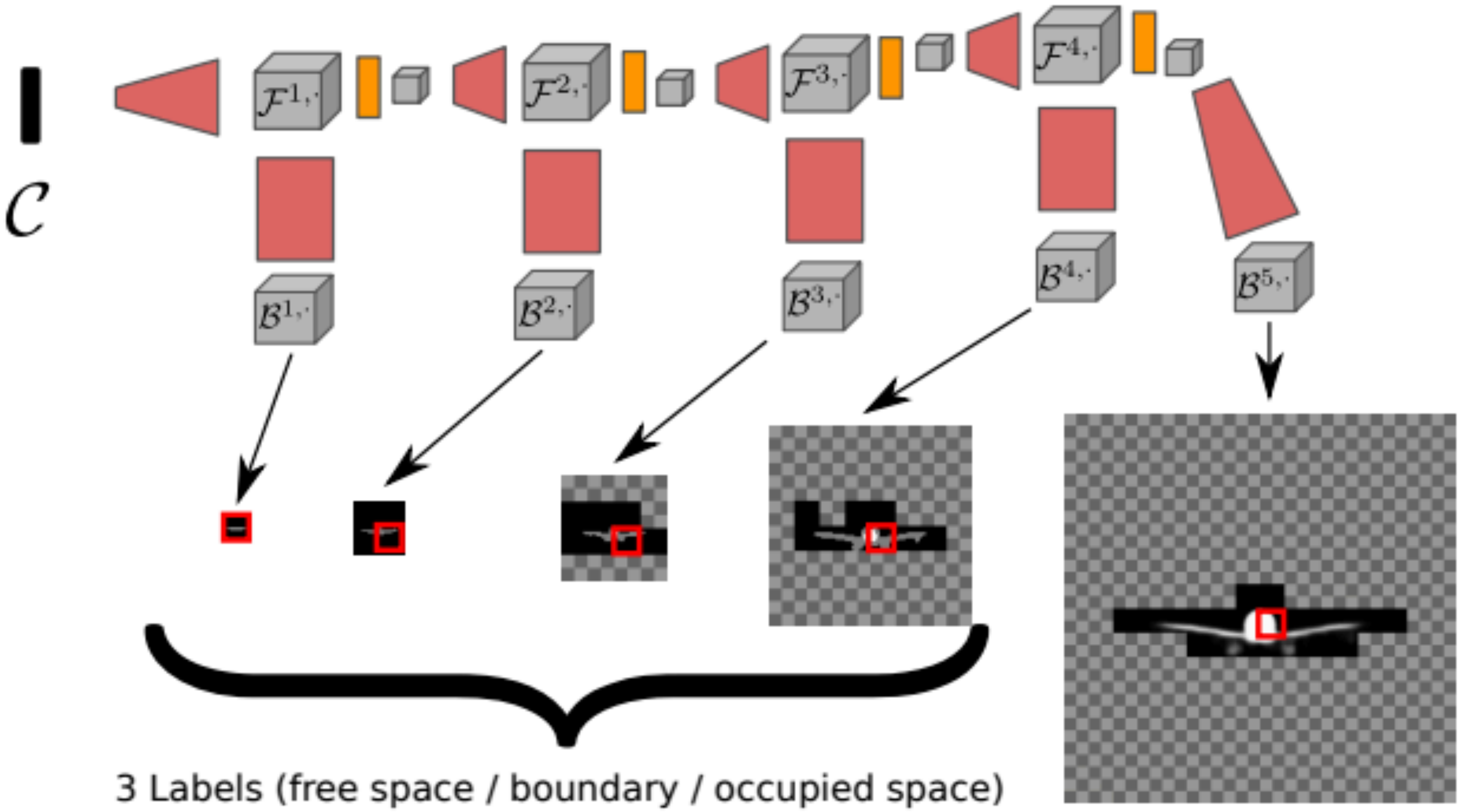
Cropping

Encoder



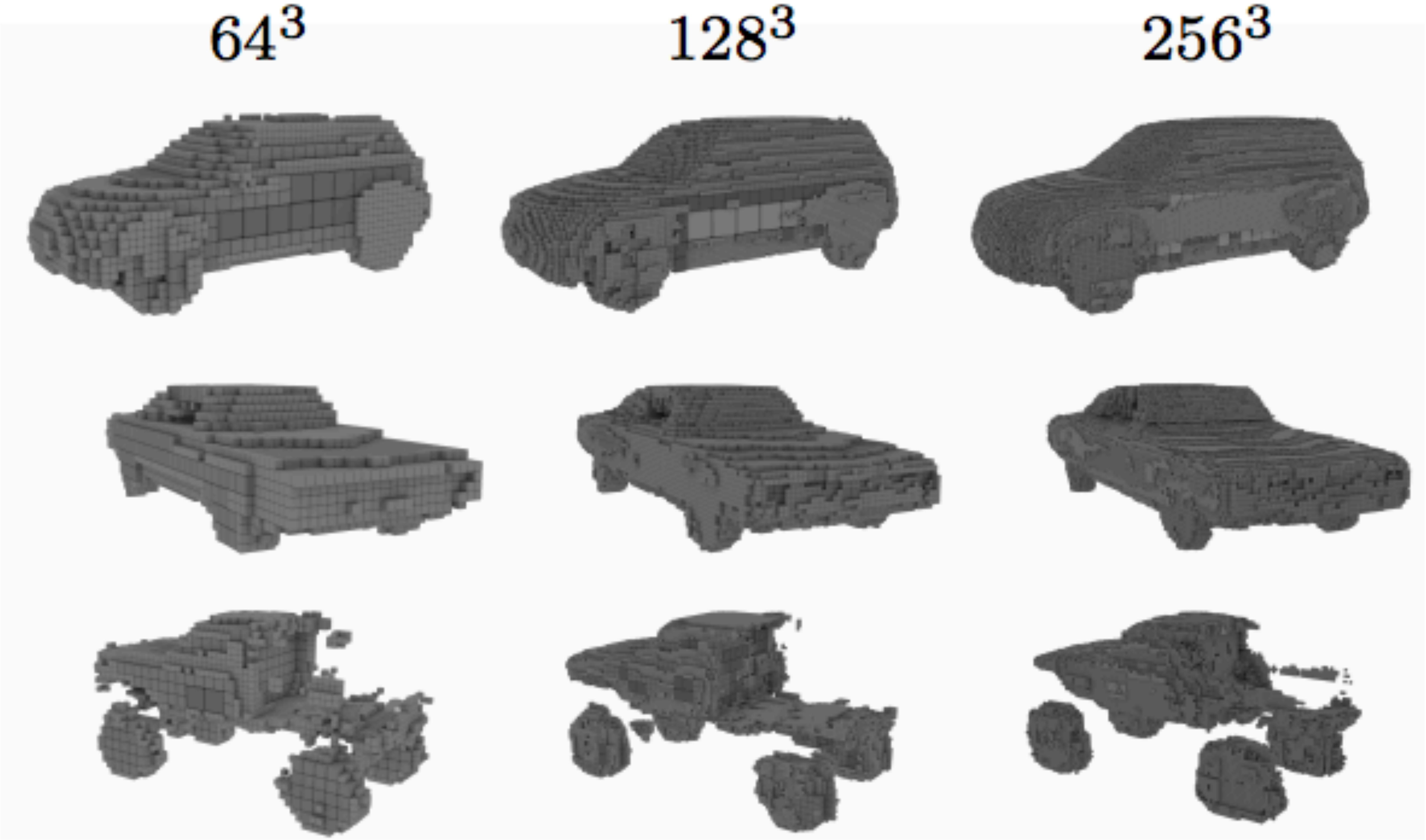
Wang et al. 2017

only generate non-empty voxels



[Hane et al. 2018]

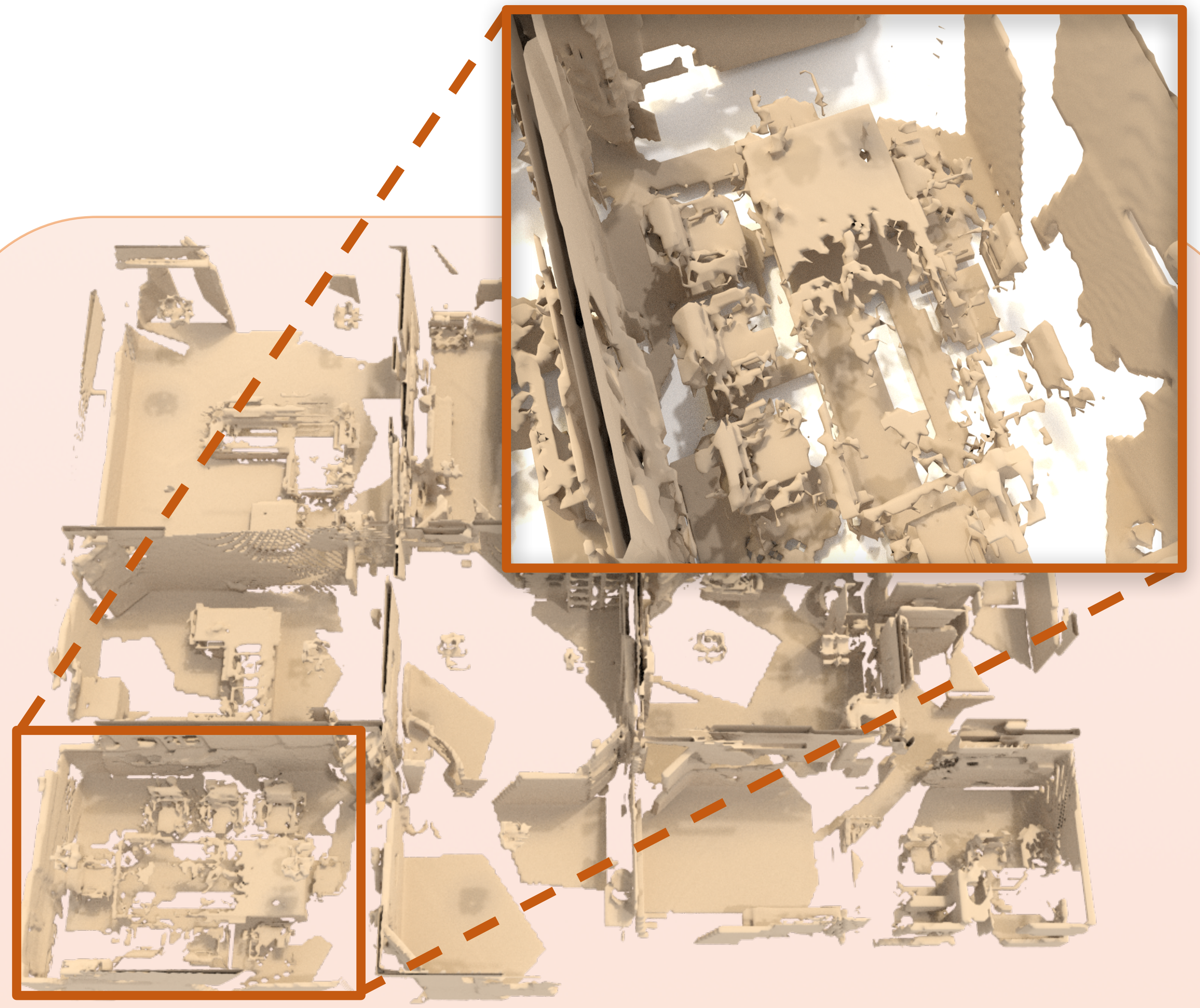
Efficient Volumetric Datastructures



[Hane et al. 2018]

Learning to Complete 3D Scans

(slide credit: Matthias Niessner)

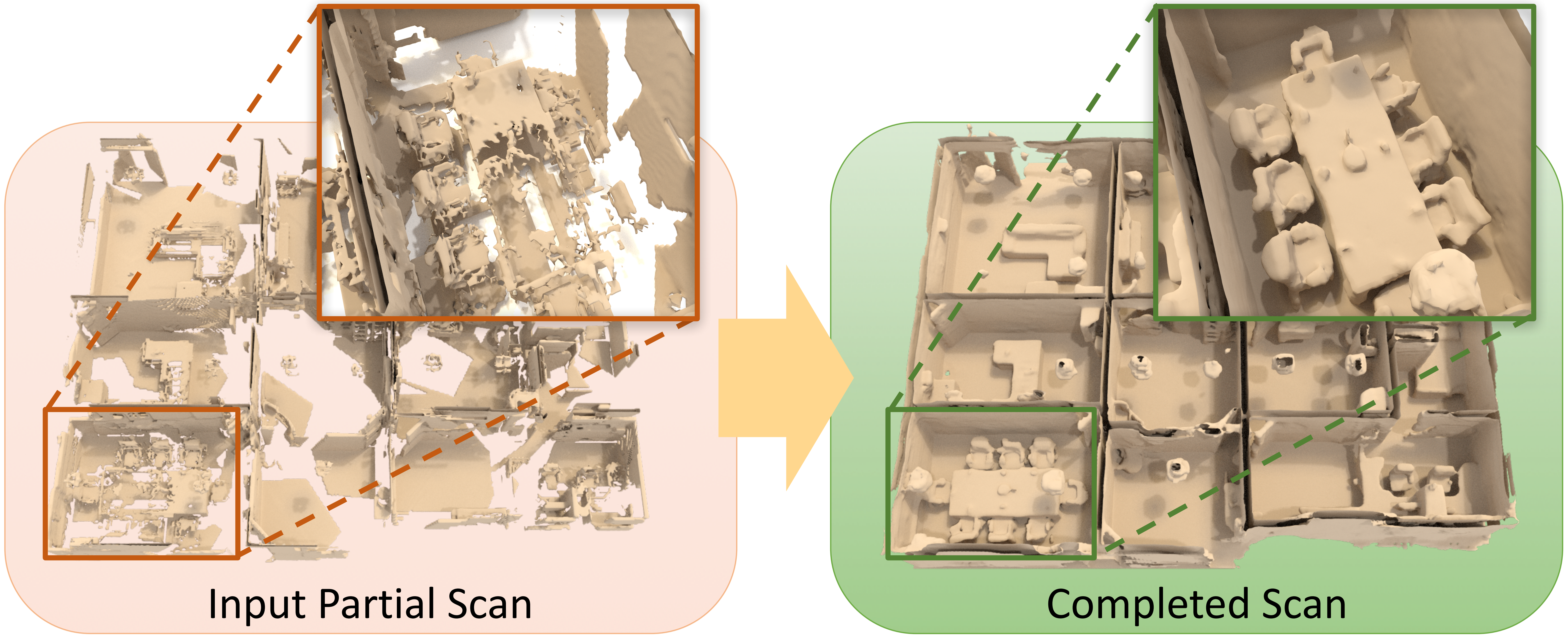


Input Partial Scan

[Dai et al. 2018]

Learning to Complete 3D Scans

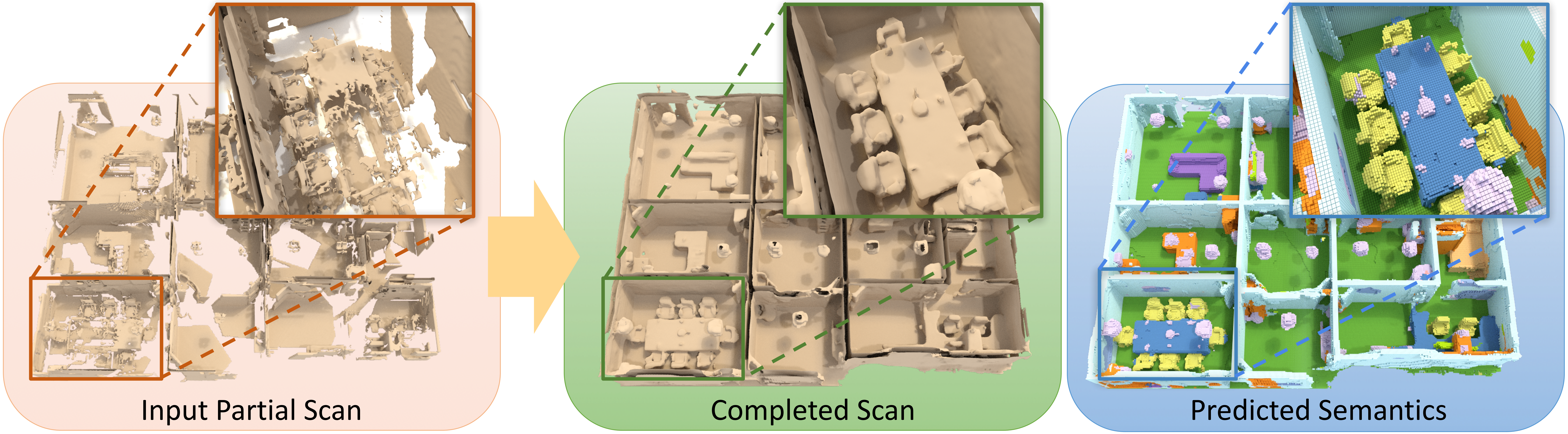
(slide credit: Matthias Niessner)



[Dai et al. 2018]

Learning to Complete 3D Scans

(slide credit: Matthias Niessner)



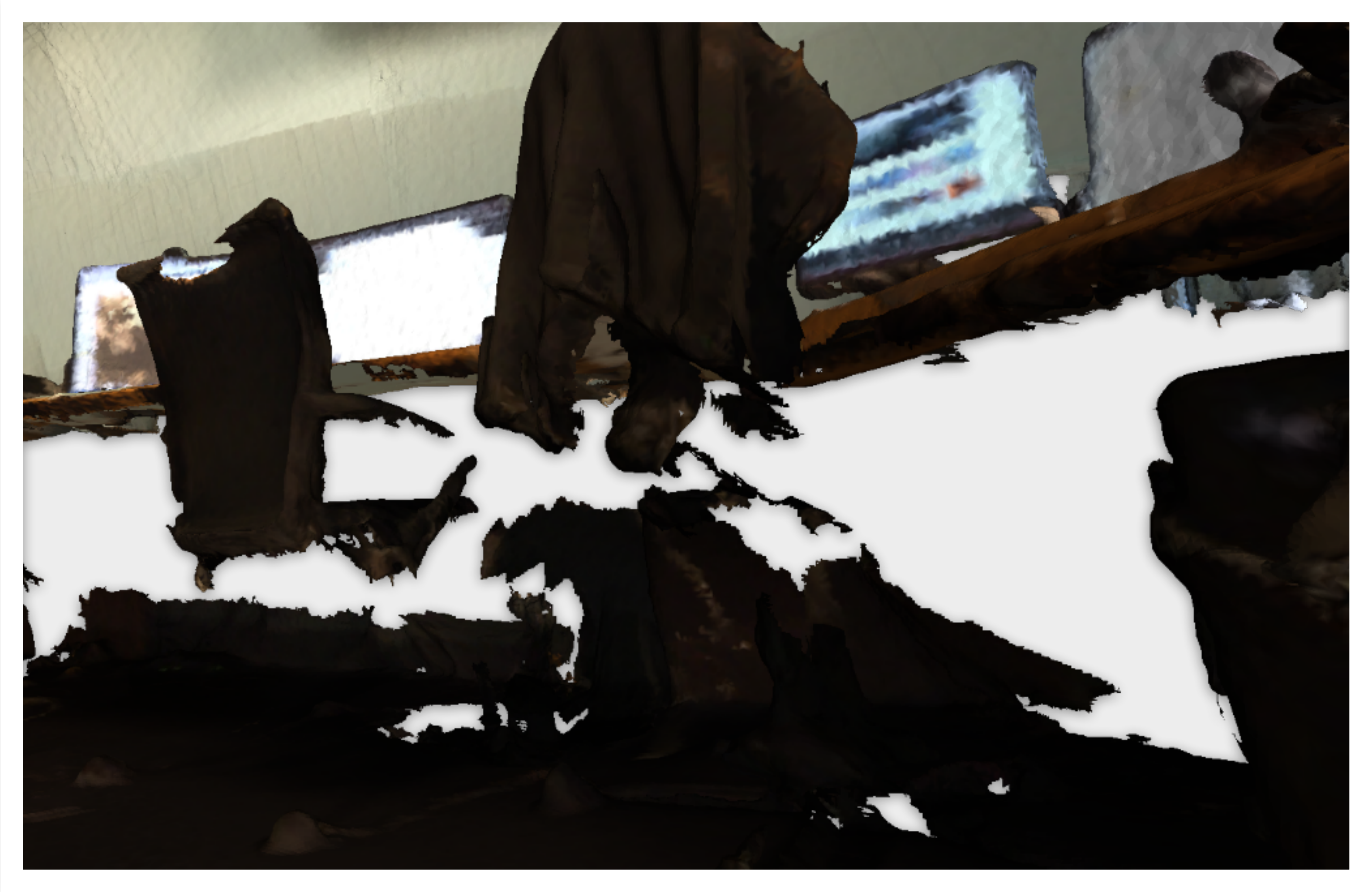
[Dai et al. 2018]

(slide credit: Matthias Niessner)

State-of-the-art 3D Reconstructions

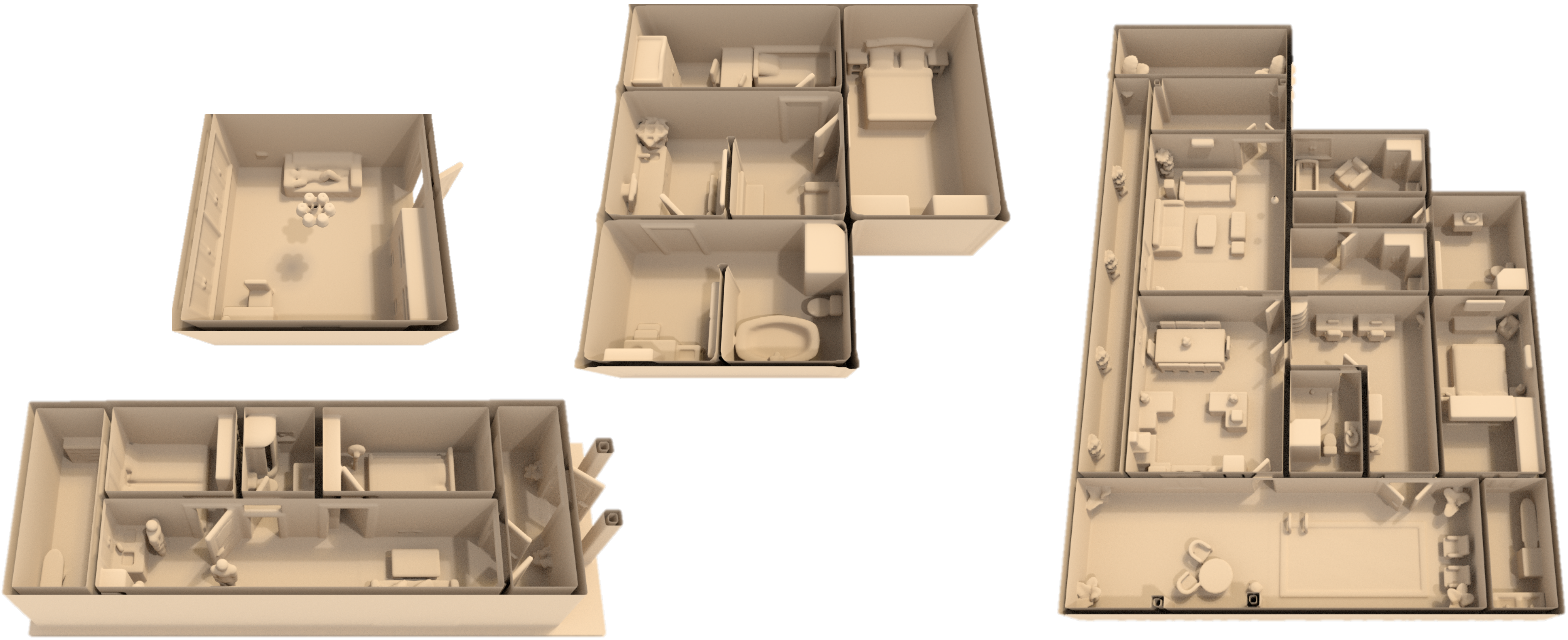
TOG'17 [Dai et al.]: BundleFusion

Problem: Incomplete Scan Geometry

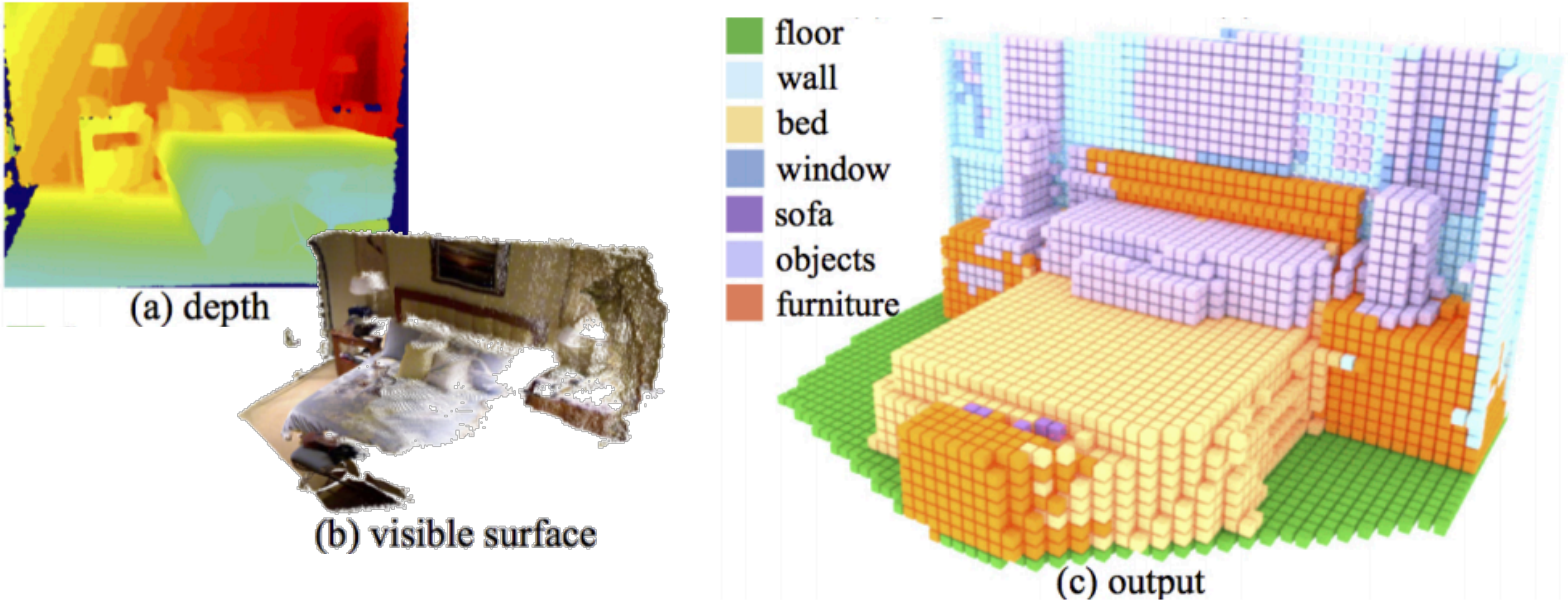


Problem: Incomplete Scan Geometry

Learning from Synthetic Data



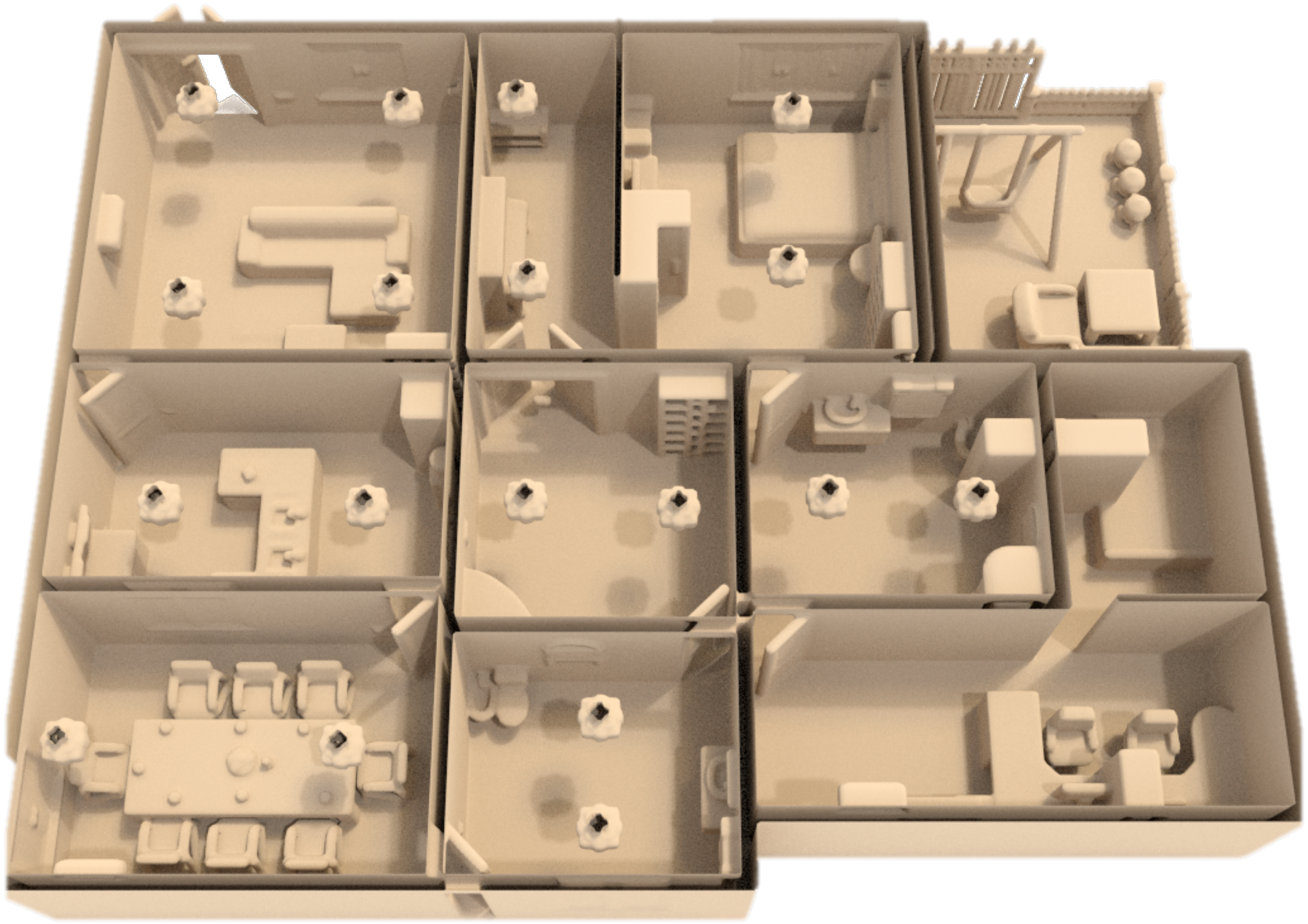
Recall: Semantic Scene Understanding



[Song et al. 2017]

Learning to Complete 3D Scans

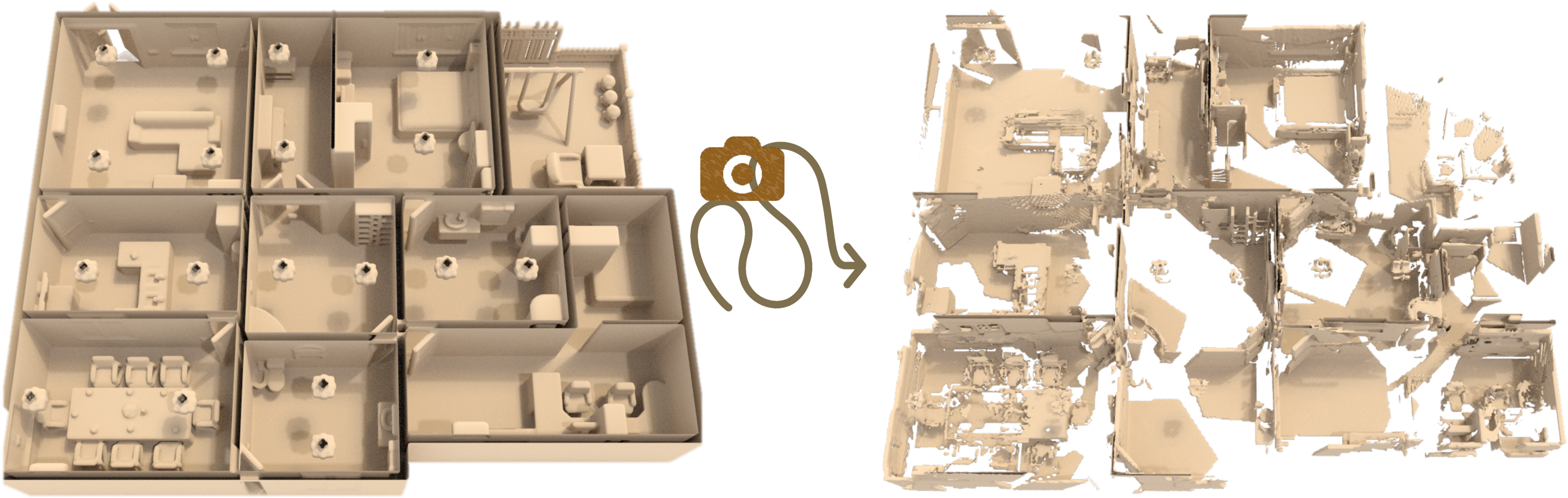
(slide credit: Matthias Niessner)



Scenes from SUNCG [Song et al. 17]

Learning to Complete 3D Scans

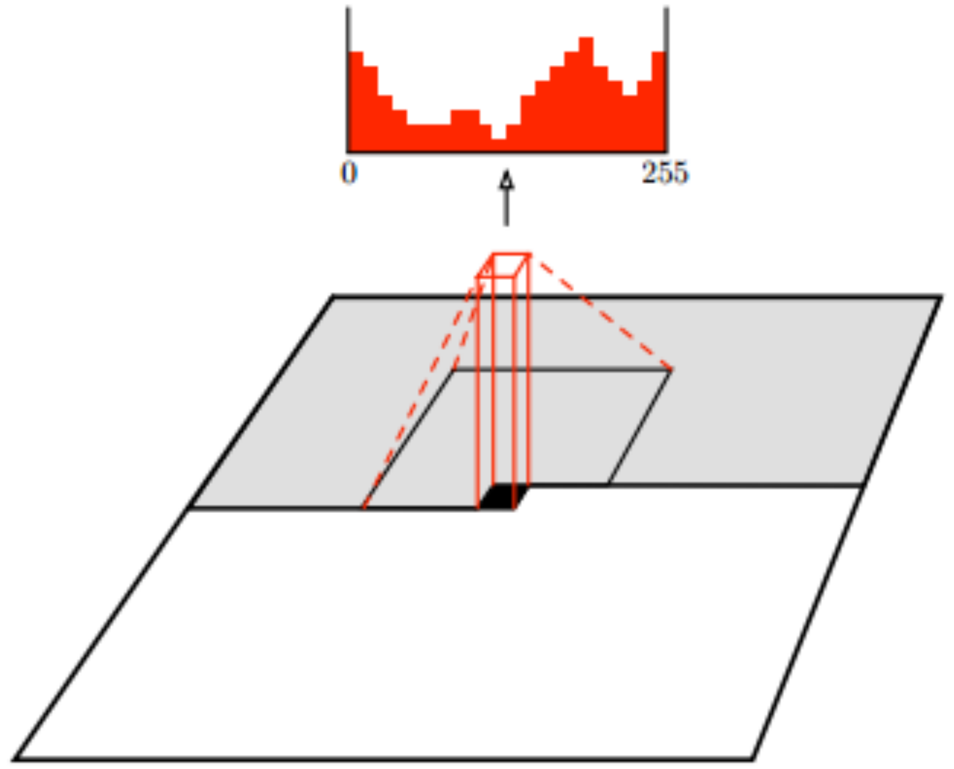
(slide credit: Matthias Niessner)

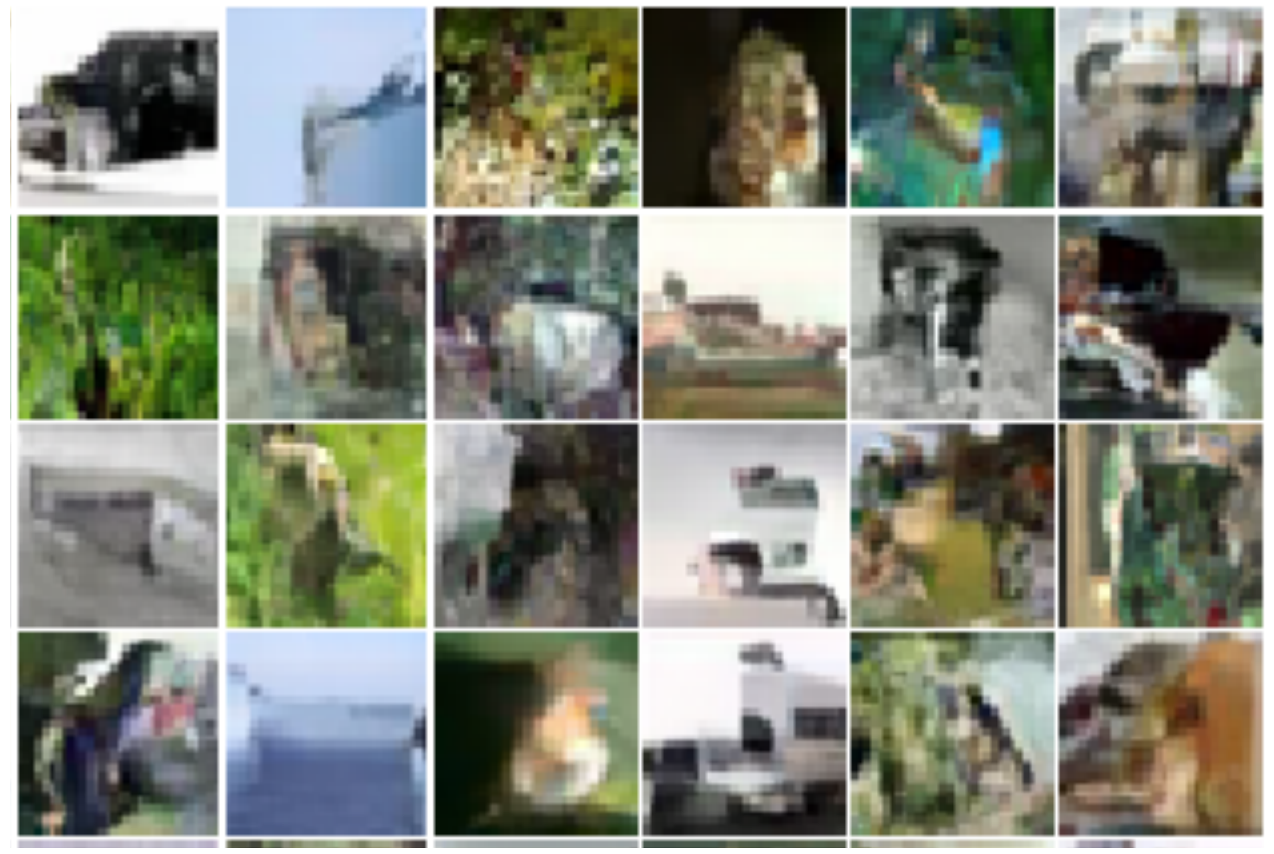


Scenes from SUNCG [Song et al. 17]

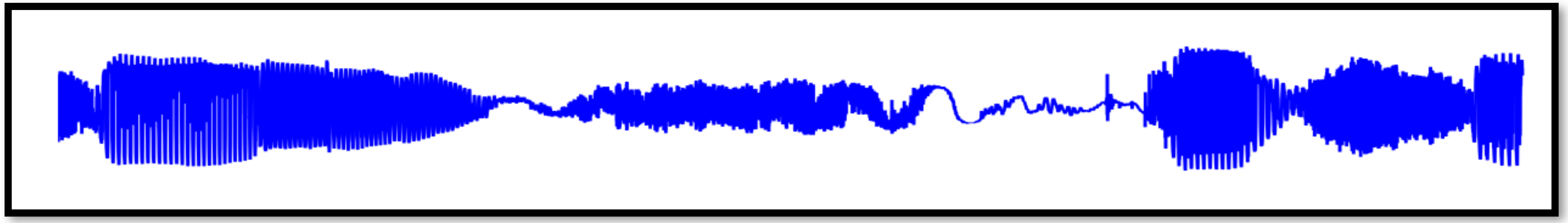
Dependent Predictions: Autoregressive Neural Networks

- PixelCNN [van den Oord 2015, van den Oord 2016, Reed 2017]

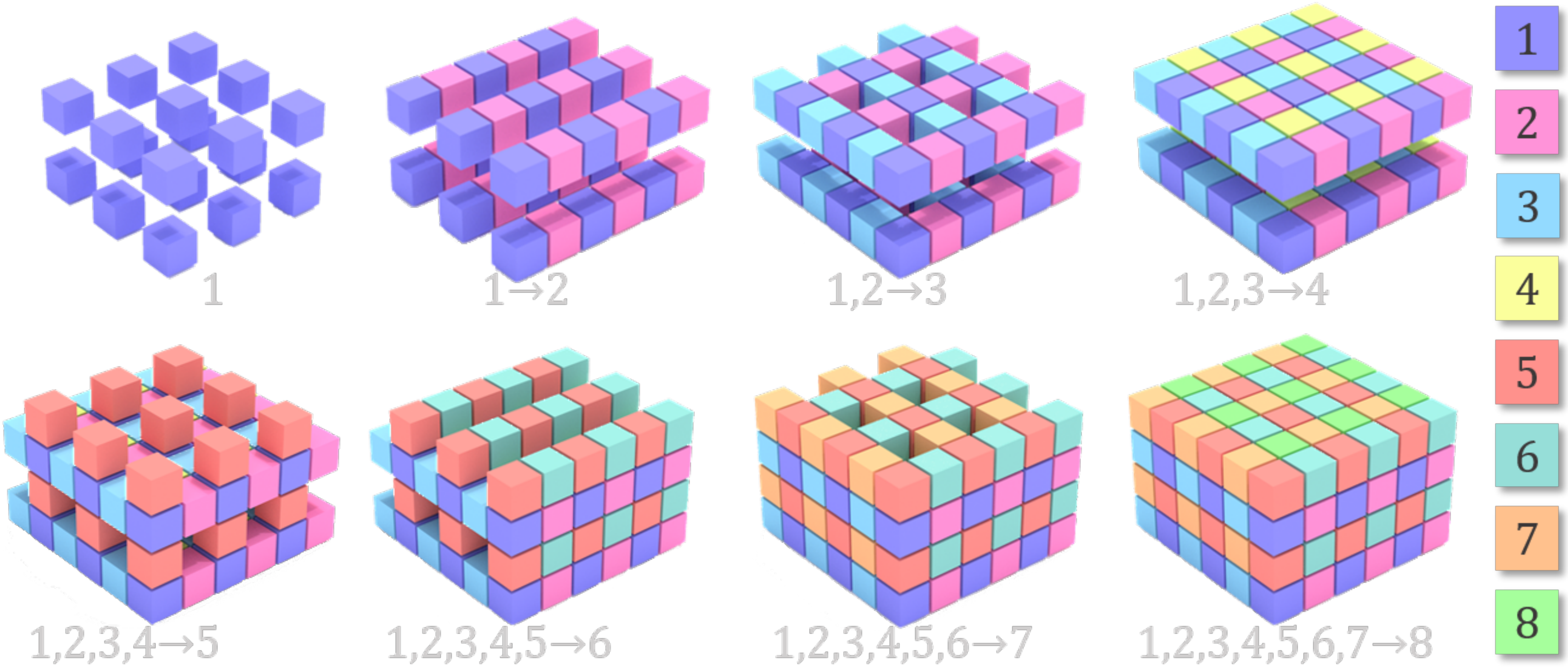




- WaveNet [van den Oord 2016]



Dependent Predictions: Autoregressive Neural Networks



ScanComplete: Completing 3D Scans

Input

[Dai et al. 2018]

ScanComplete: Completing 3D Scans

Input

Completion

[Dai et al. 2018]

ScanComplete: Completing 3D Scans

[Dai et al. 2018]

ScanComplete: Completing 3D Scans

[Dai et al. 2018]

ScanComplete: Completing 3D Scans

[Dai et al. 2018]

Representation for 3D

- Image-based
- Volumetric
  - **PROS**: modify image networks
  - **CONS**: special layers for hierarchical datastructures, still too coarse
- Point-based
- Surface-based

Representation for 3D

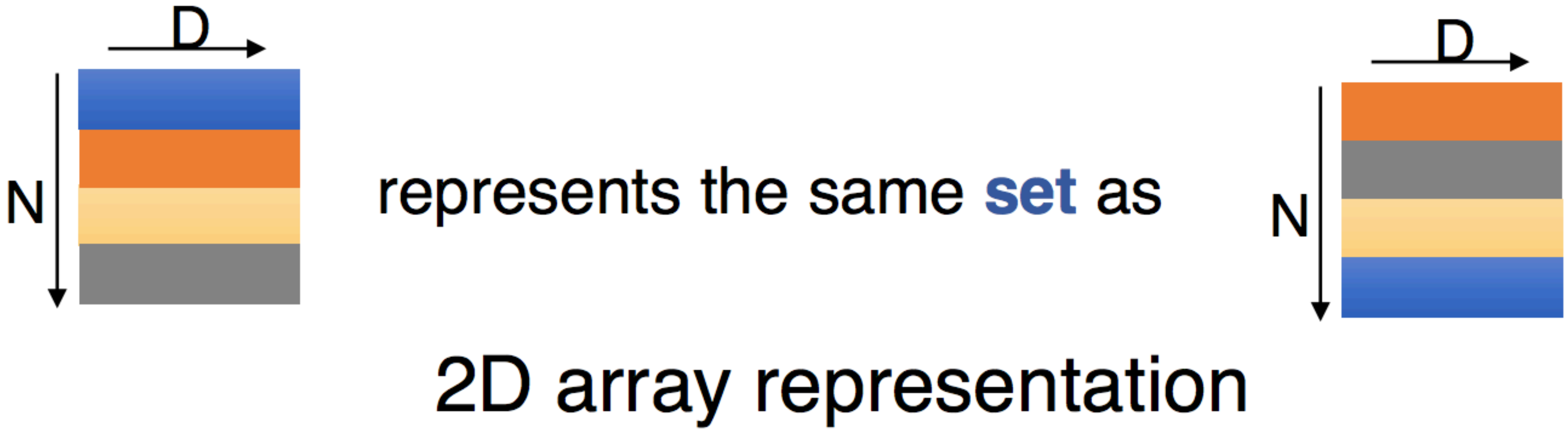
- Image-based
- Volumetric
- **Point-based**
- Surface-based

Representation for 3D: Point-based

- Common representation
- Easy to obtain from meshes, depth scans, laser scans

Representation for 3D

- Common representation
- Easy to obtain from meshes, depth scans, laser scans
- Unstructured (e.g., any permutation of points gives same shape!)



[Su et al. 2017]

PointNet for Point Cloud Analysis

$$f(x_1, x_2, \ldots, x_n) \equiv f(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}), \quad x_i \in \mathbb{R}^D$$

permutation-invariant functions

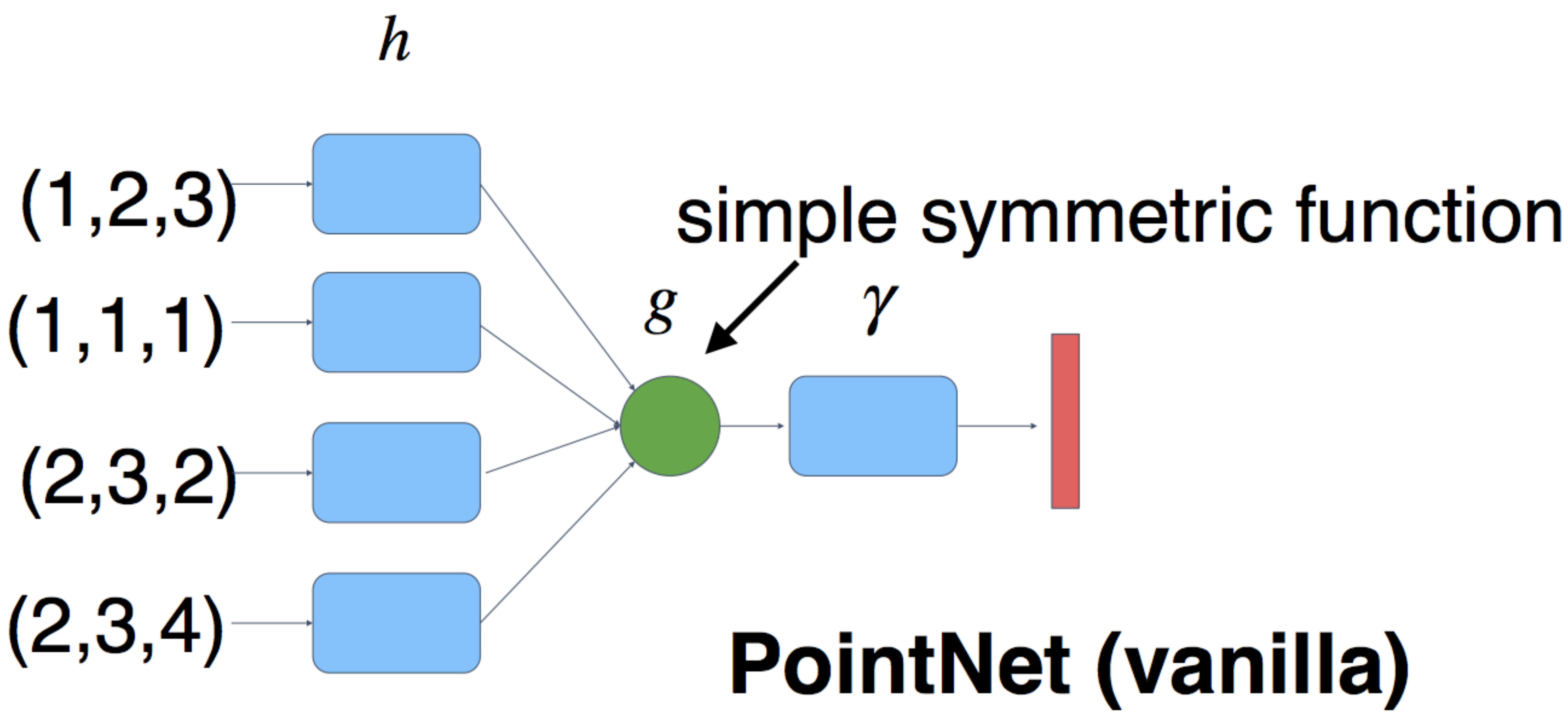
[Su et al. 2017]

PointNet for Point Cloud Analysis

Use MLPs (h) and max-pooling (g) as simple symmetric functions

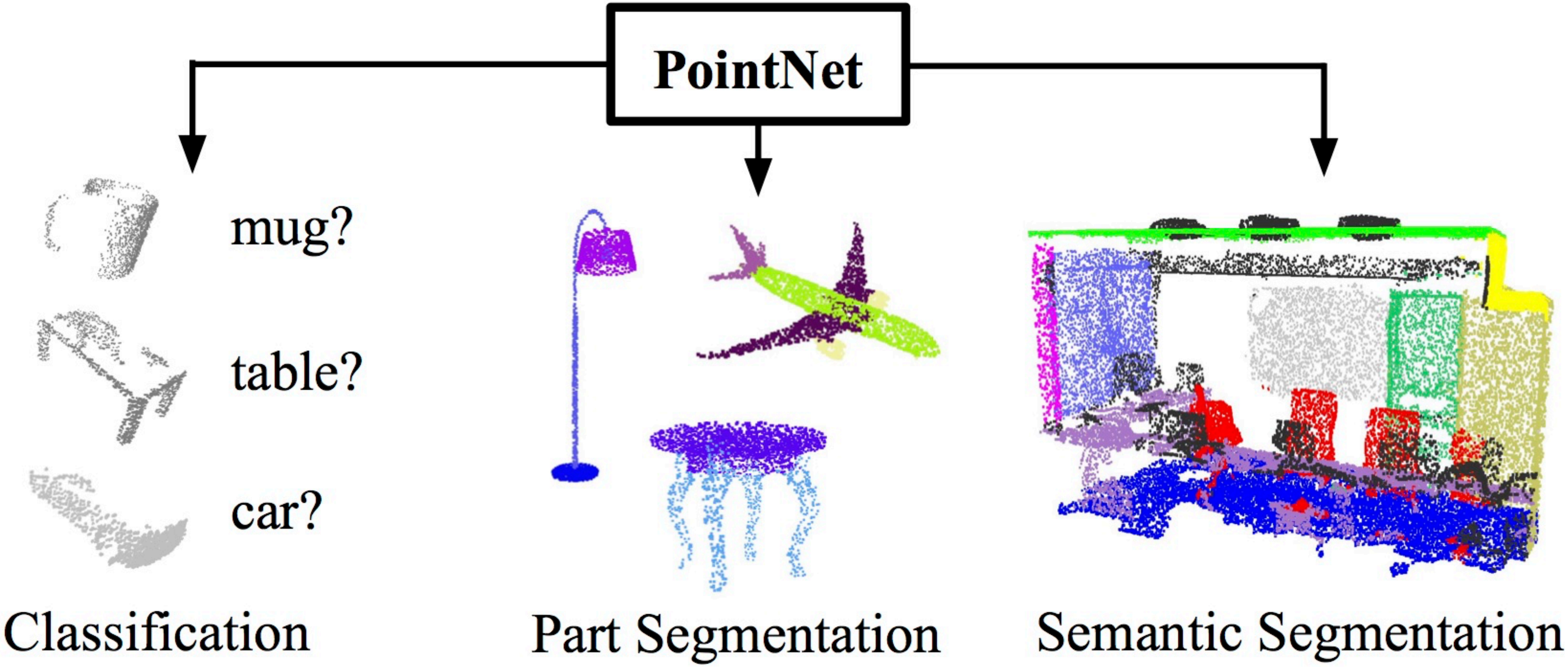
$$f(x_1, x_2, \ldots, x_n) \equiv f(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}), \quad x_i \in \mathbb{R}^D$$

$$f(x_1, x_2, \ldots, x_n) = \gamma \circ g(h(x_1), \ldots, h(x_n))$$

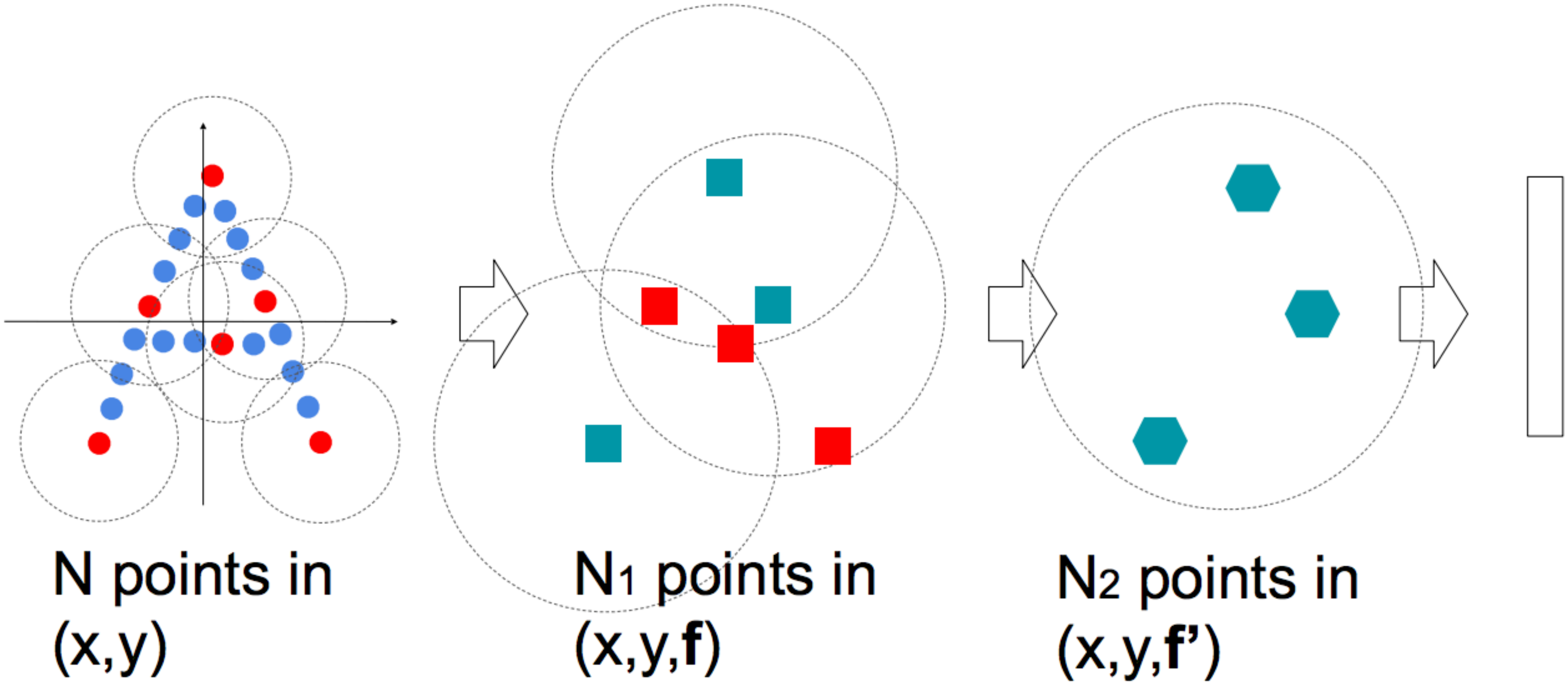


[Su et al. 2017]

PointNet for Point Cloud Analysis

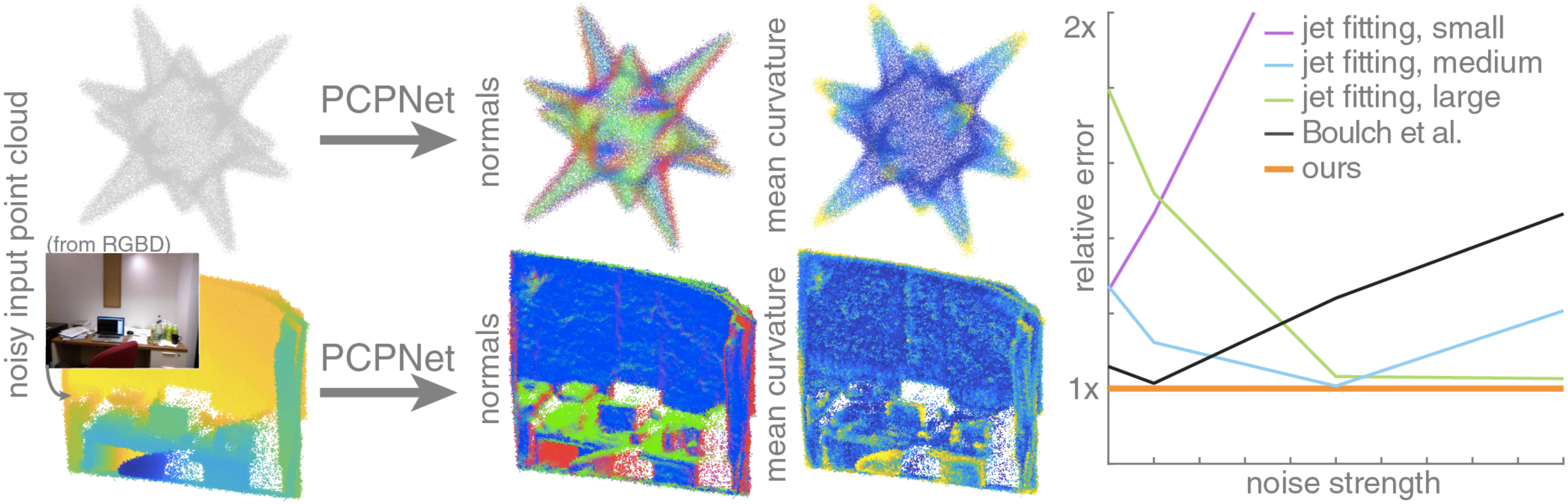


PointNet for Point Cloud Analysis: PointNet++



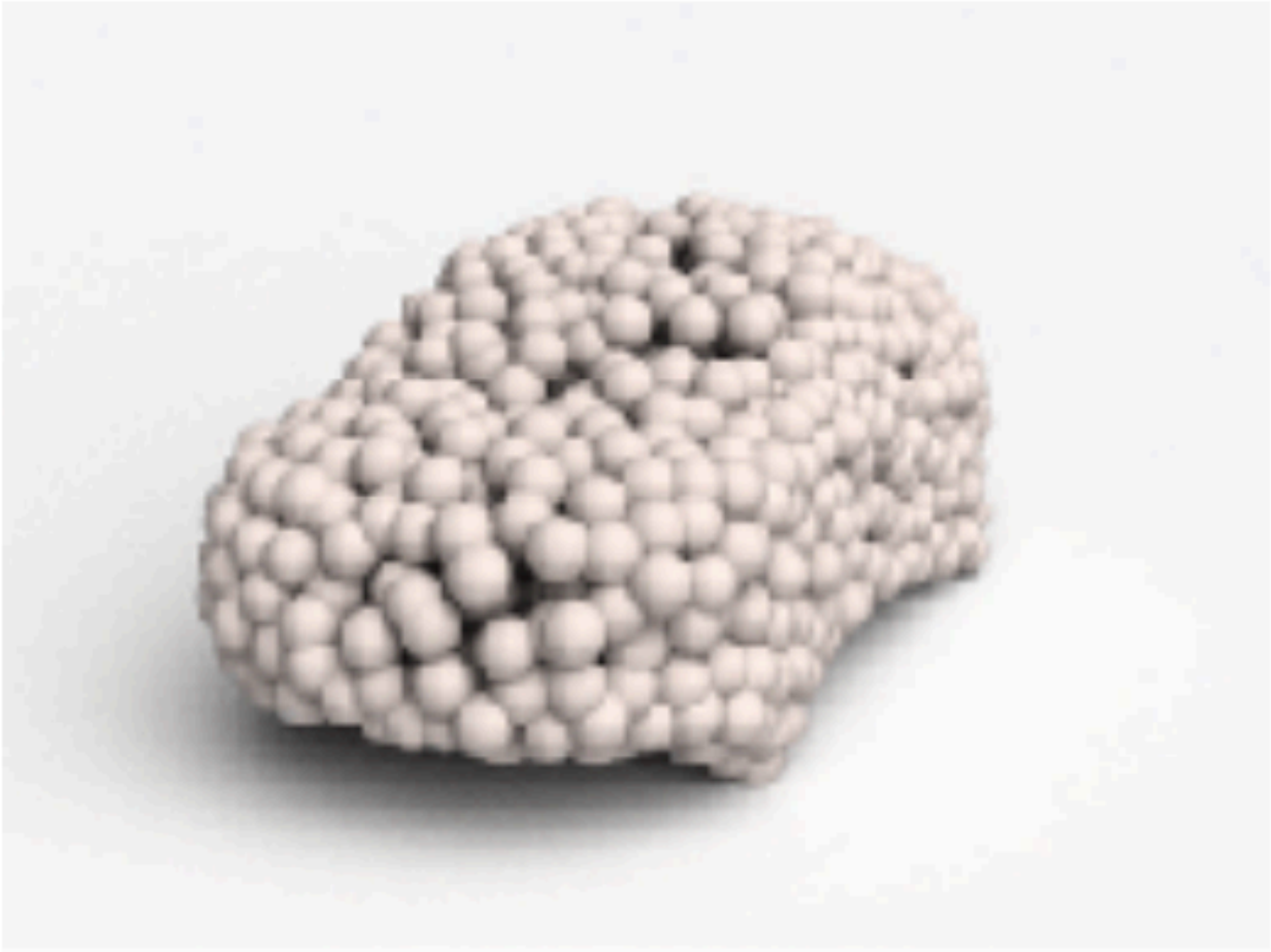
[Qi et al. 2018]

PointNet for Local Point Cloud Analysis



[Guerrero et al. 2018]

PointNet for Point Cloud Synthesis



Input

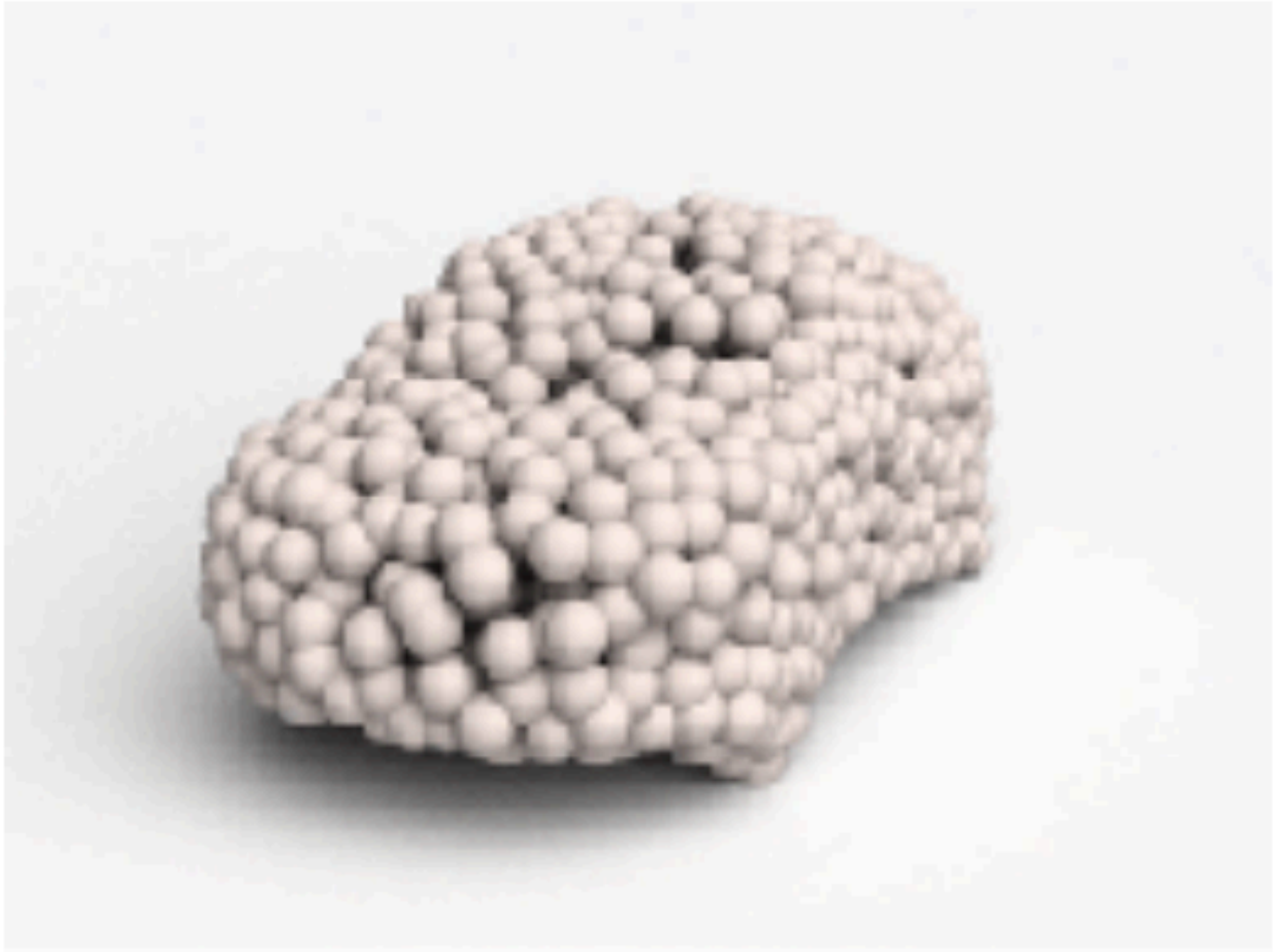
Reconstructed 3D point cloud

[Su et al. 2017]

PointNet for Point Cloud Synthesis

generated output needs to be compare to some true shape

Input

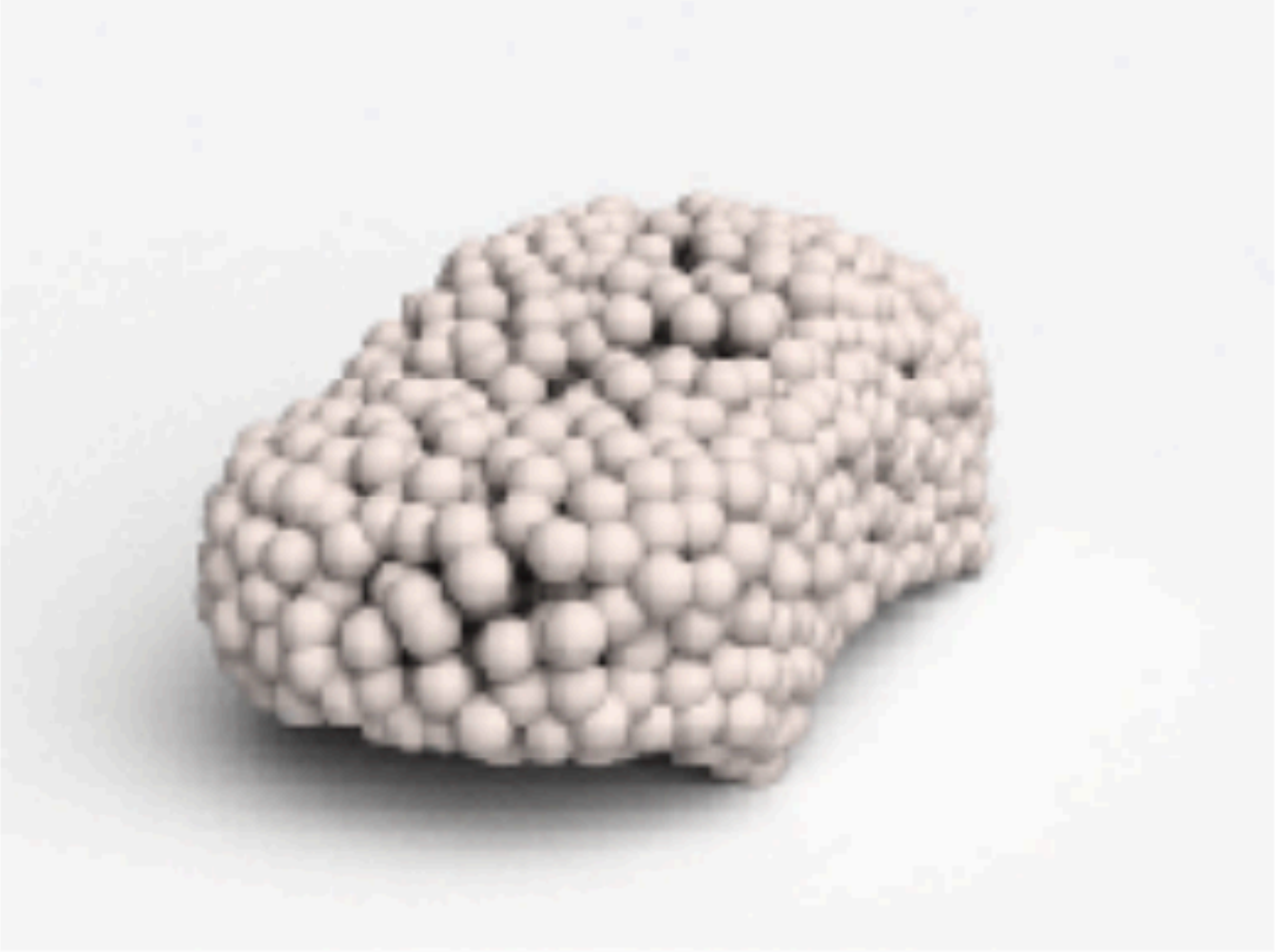


Reconstructed 3D point cloud

[Su et al. 2017]

PointNet for Point Cloud Synthesis

generated output needs to be compare to some true shape



Input

Earth Mover Distance as loss function

$$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \rightarrow S_2} \sum_{x \in S_1} \|x - \phi(x)\|_2$$

[Su et al. 2017]

Representation for 3D

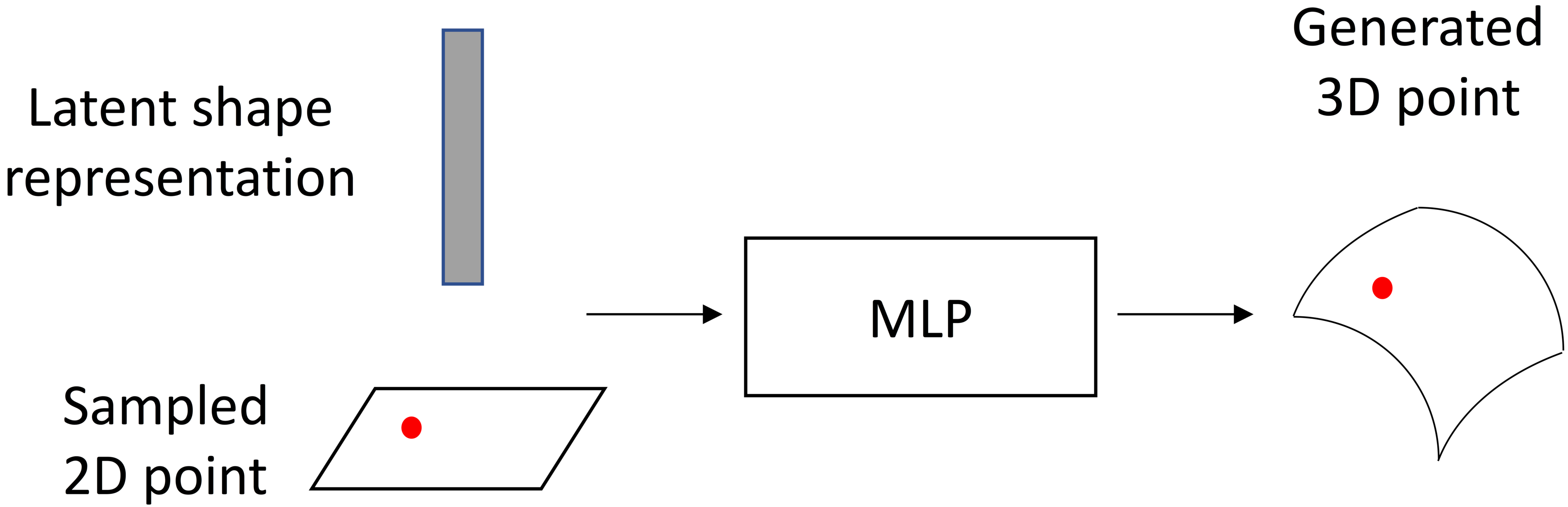
- Image-based
- Volumetric
- Point-based
- **Surface-based**

Surface models used in engineering (i.e., CAD) and computer graphics (i.e., meshes)

Image | Generated Volume | Generated Points | Generated Surface

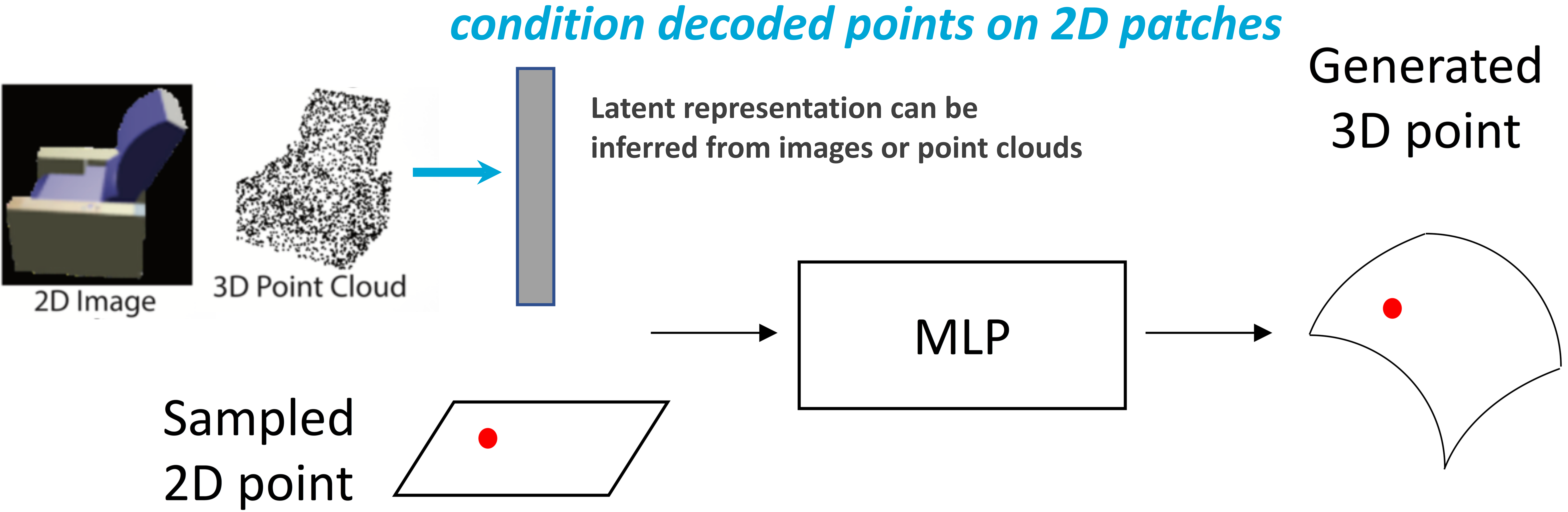
AtlasNet for Surface Generation

condition decoded points on 2D patches



[Groueix et al. 2018]

AtlasNet for Surface Generation



[Groueix et al. 2018]

AtlasNet for Surface Generation

Quad Mesh is generated by mapping a regular grid in 2D domain to 3D points

AtlasNet for Surface Generation

BONUS: natural space to store textures for CG

2D Image

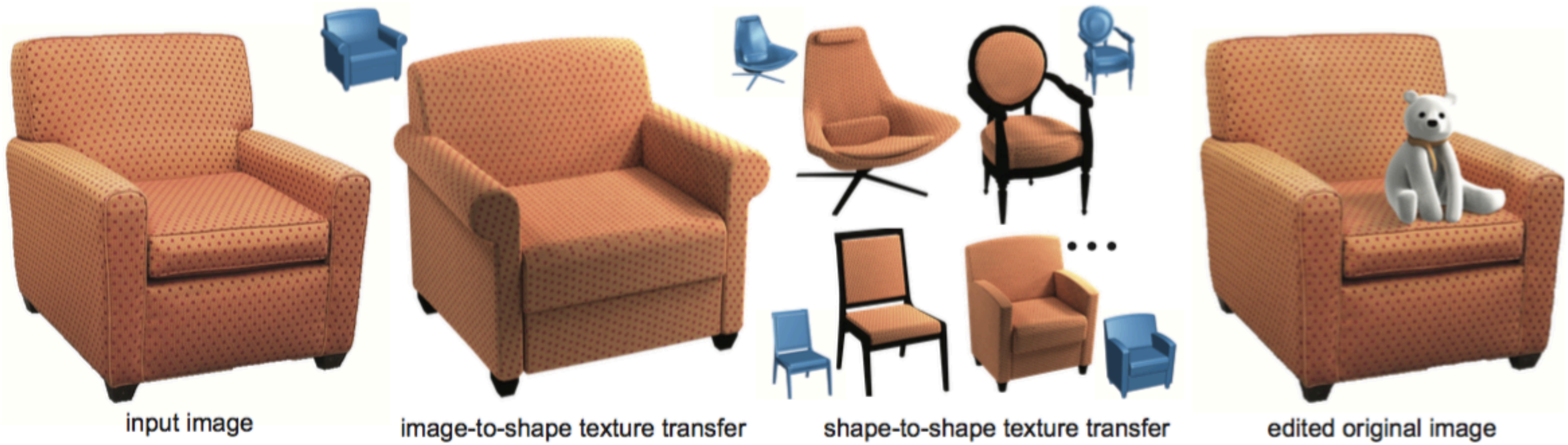
3D Point Cloud

Latent representation can be inferred from images or point clouds

Sampled 2D point

MLP

Texture Transfer



[Wang et al. 2016]

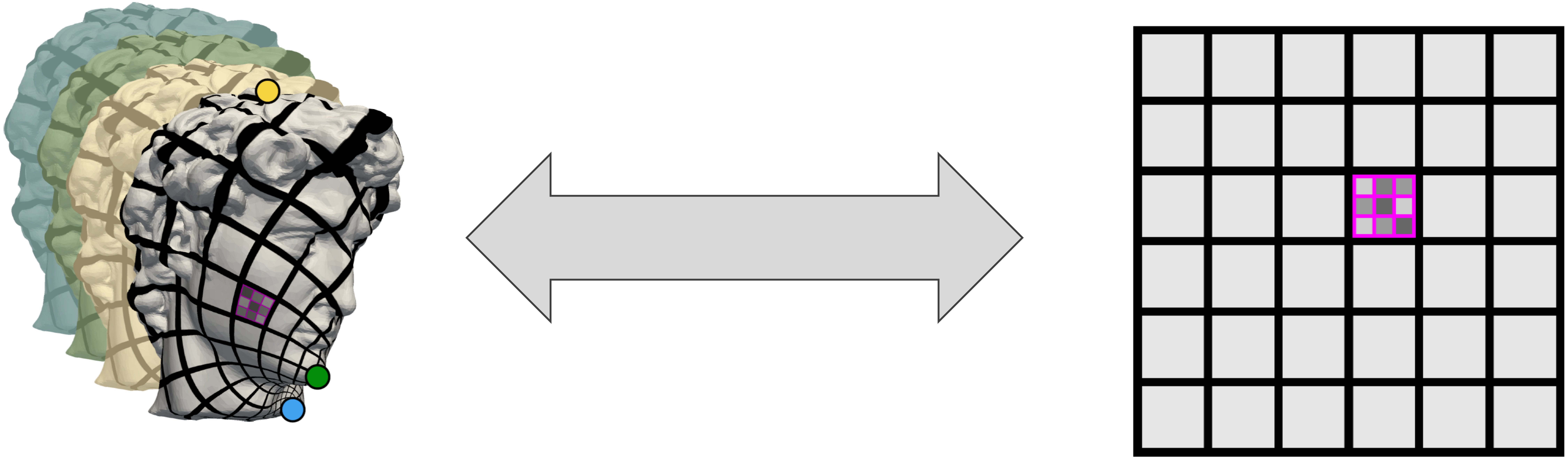
Parameterization for Surface Analysis

map 3D surface to 2D domain

[Maron et al. 2017]

Parameterization for Surface Analysis

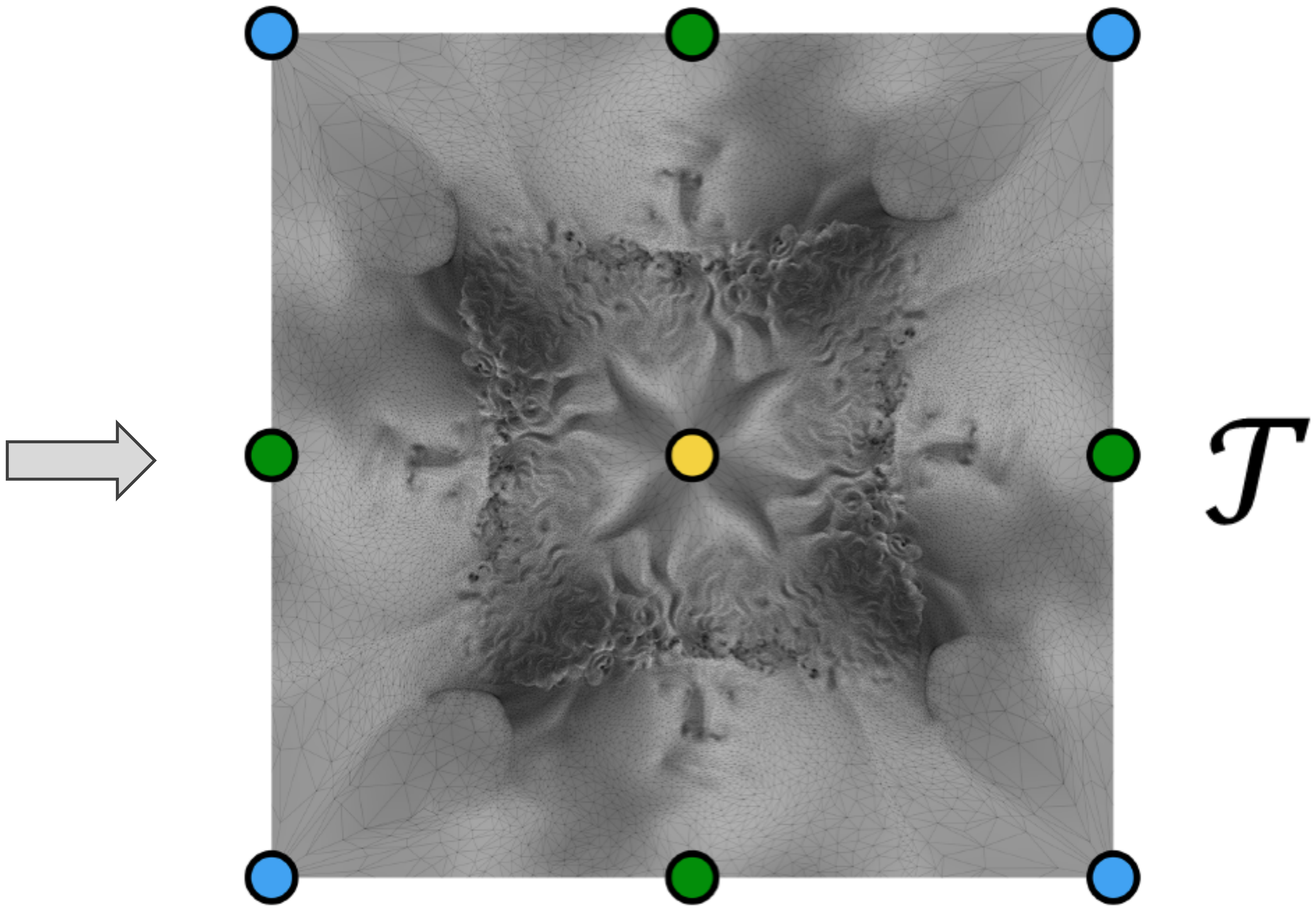
map 3D surface to 2D domain



[Maron et al. 2017]

Parameterization for Surface Analysis

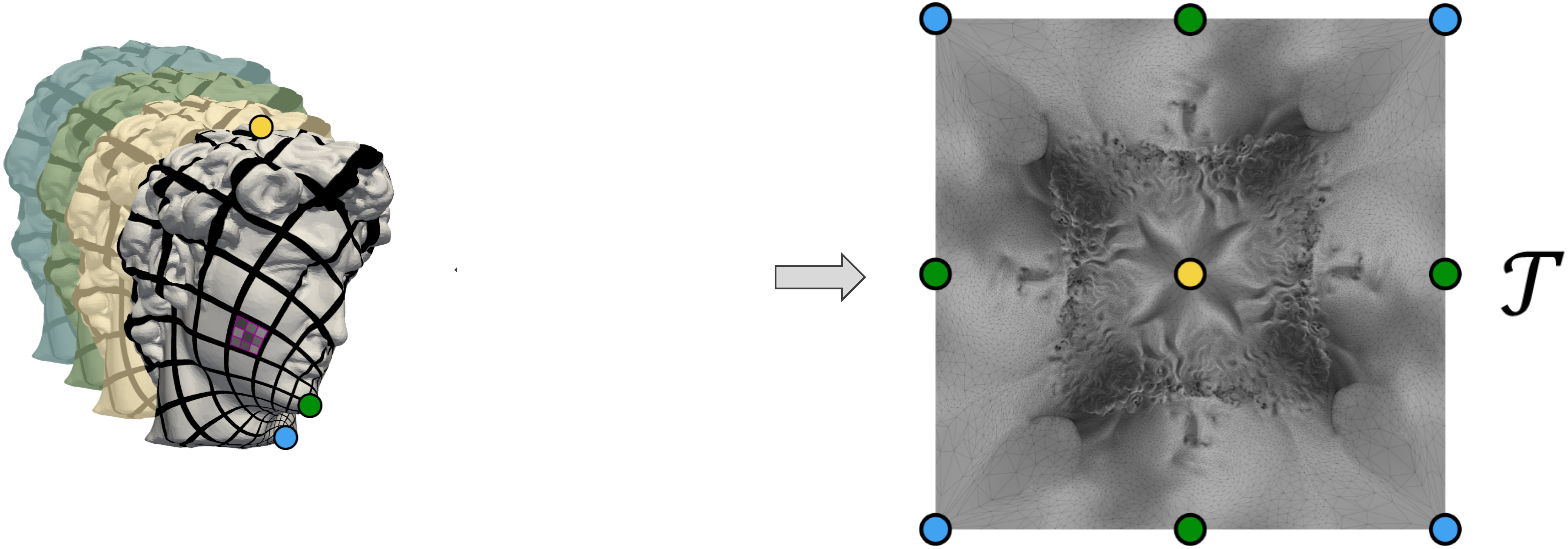
map 3D surface to 2D domain



[Maron et al. 2017]

Parameterization for Surface Analysis

map 3D surface to 2D domain



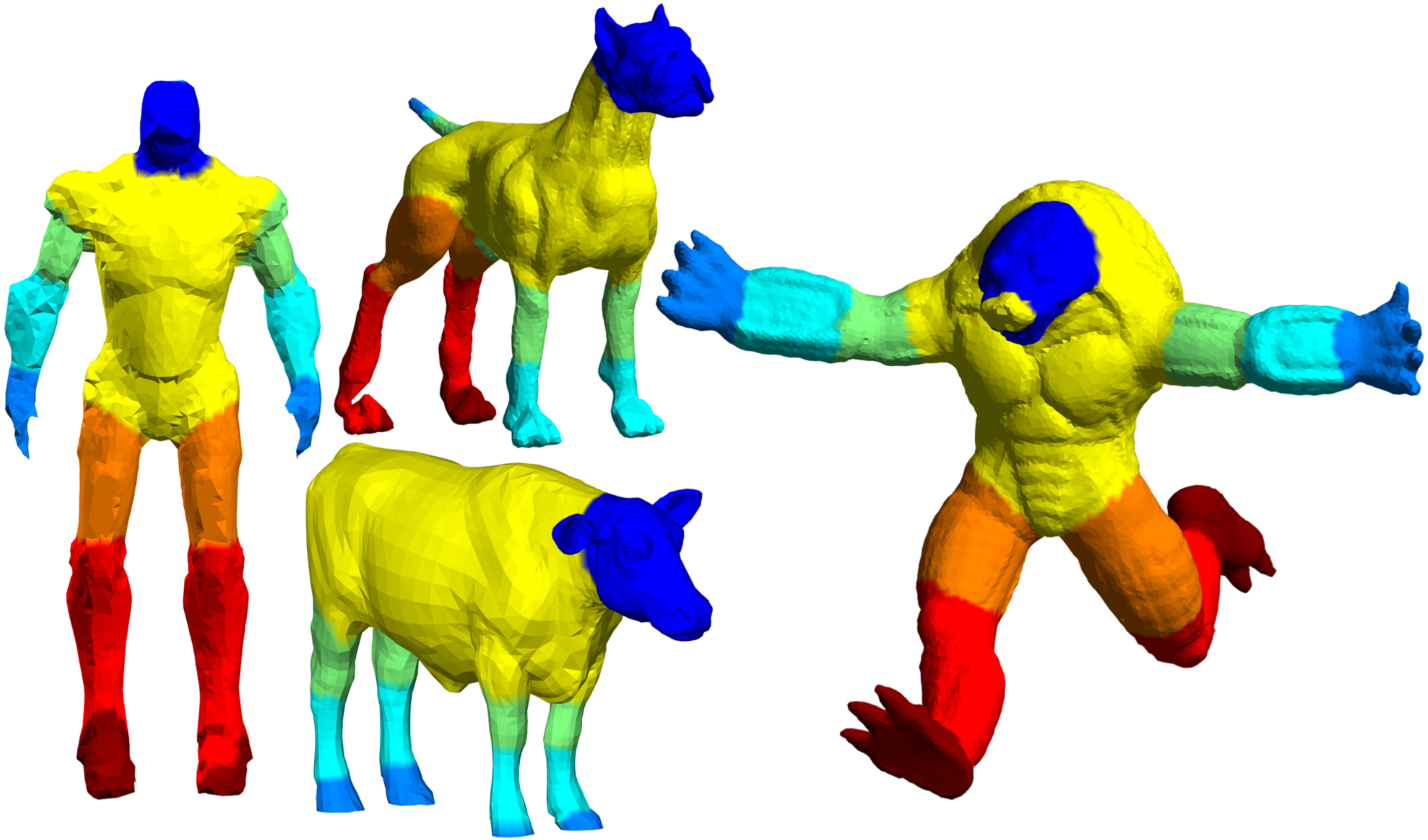
[Maron et al. 2017]

Parameterization for Surface Analysis

- Map 3D surface to 2D domain
  - One such mapping: flat torus (seamless => translation-invariant)
  - Many mappings exists: sample a few and average result
  - Which functions to map?
    XYZ, normals, curvature, …

[Maron et al. 2017]

Parameterization for Surface Analysis

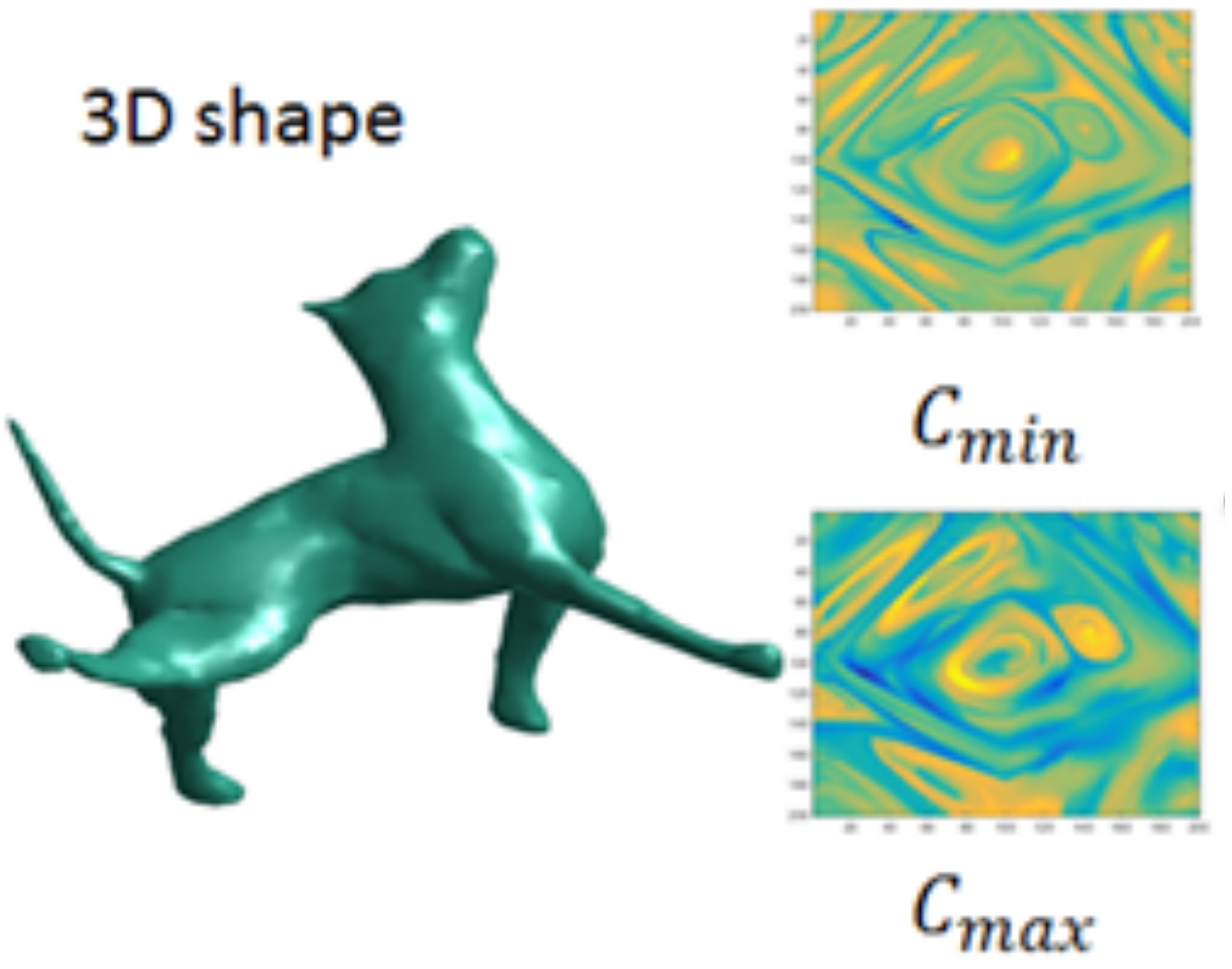


[Maron et al. 2017]

Other Parameterizations

Other Parameterizations

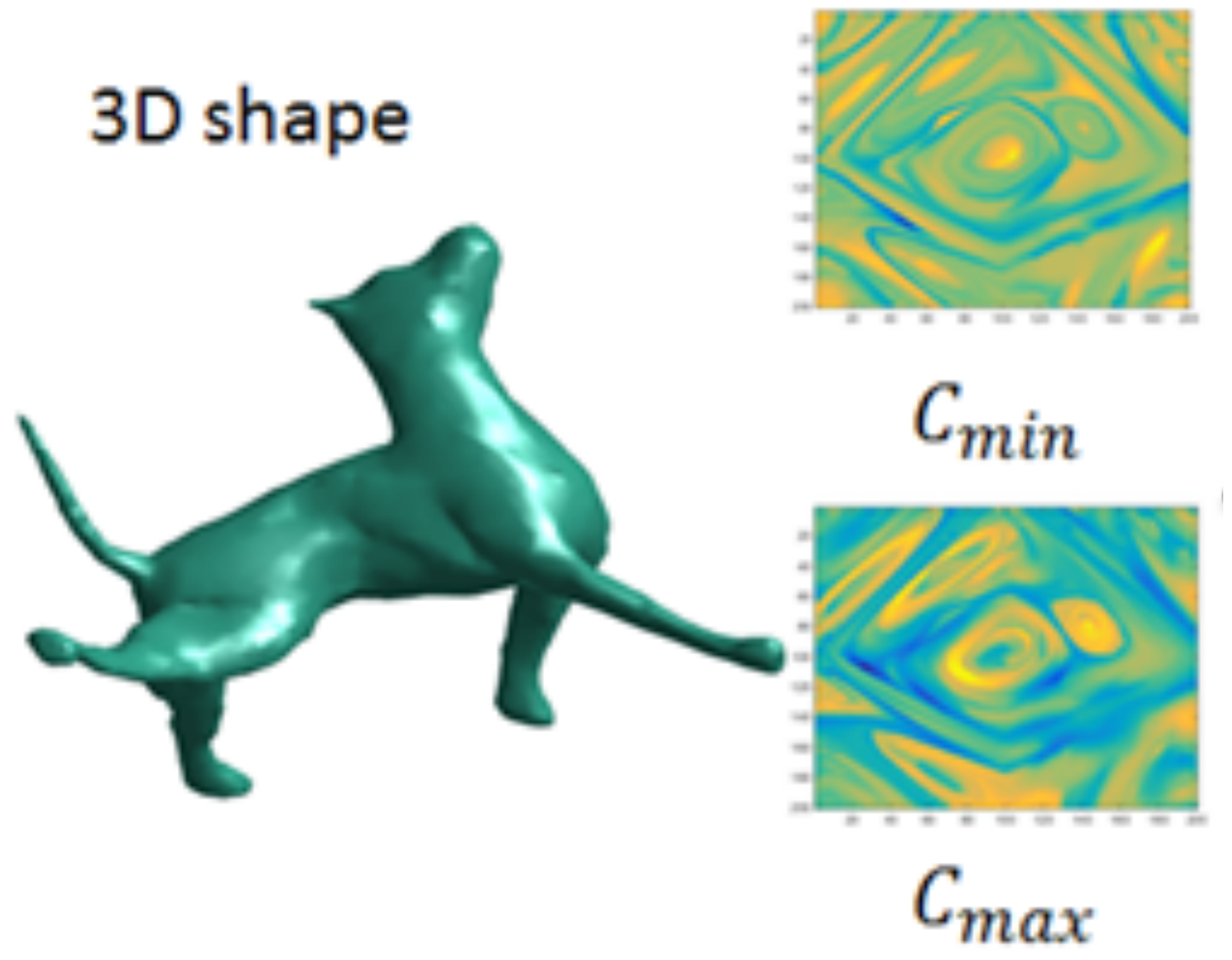
Geometry Image



[Sinha et al. 2017]

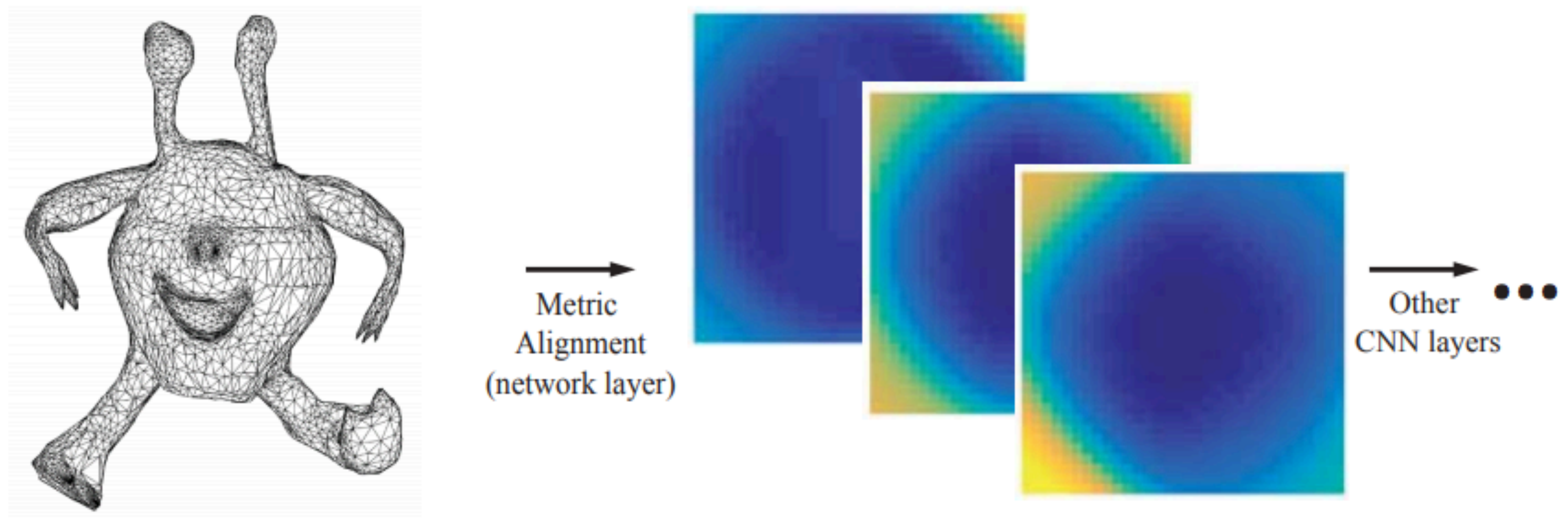
Other Parameterizations

Geometry Image



[Sinha et al. 2017]

Metric Alignment



[Ezuz et al. 2017]

Other Parameterizations

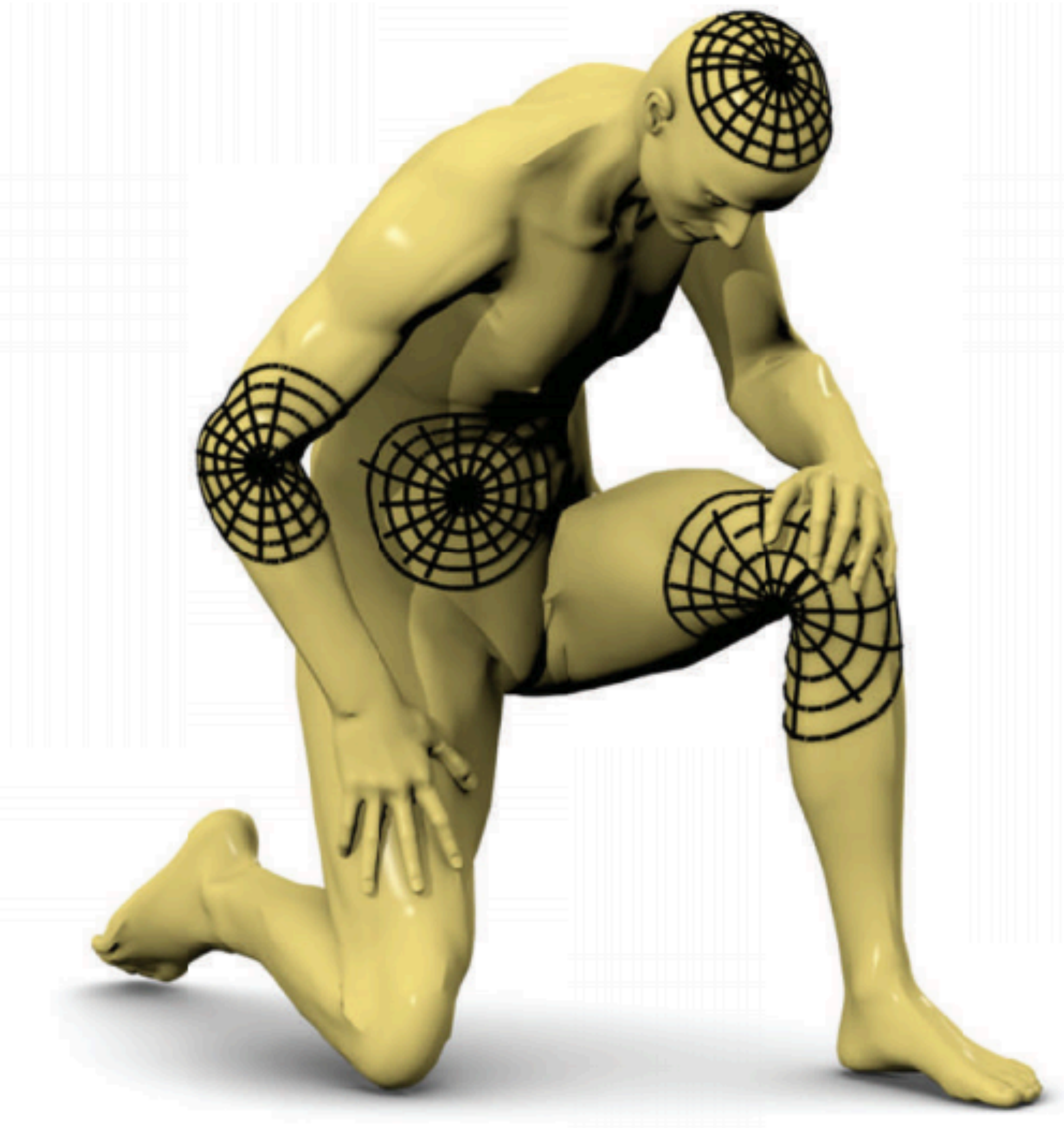
Other Parameterizations

geodesic discs

Spatial domain

Other Parameterizations

geodesic discs

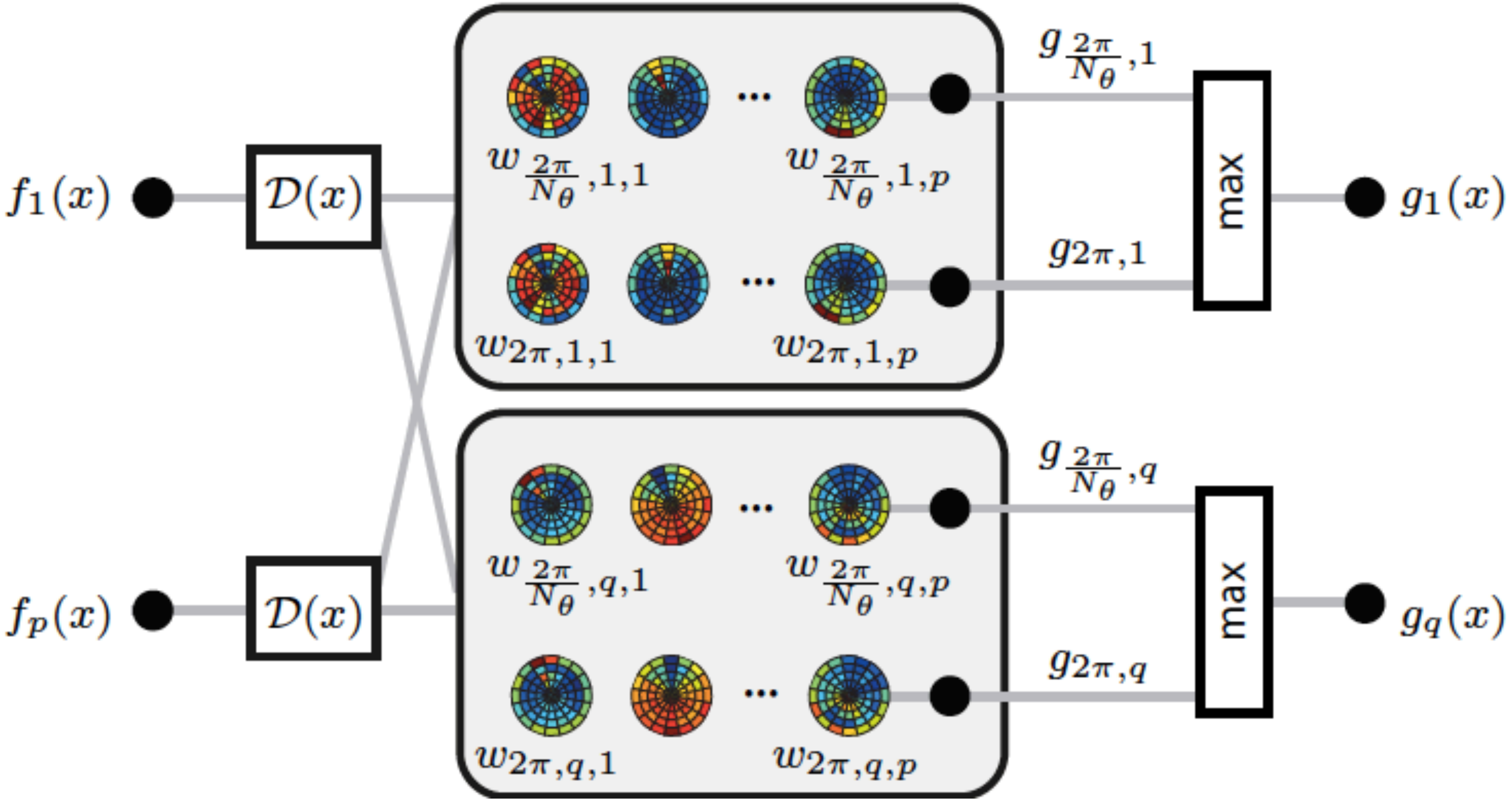


Spatial domain

parameterize in spectral domain

Spectral domain

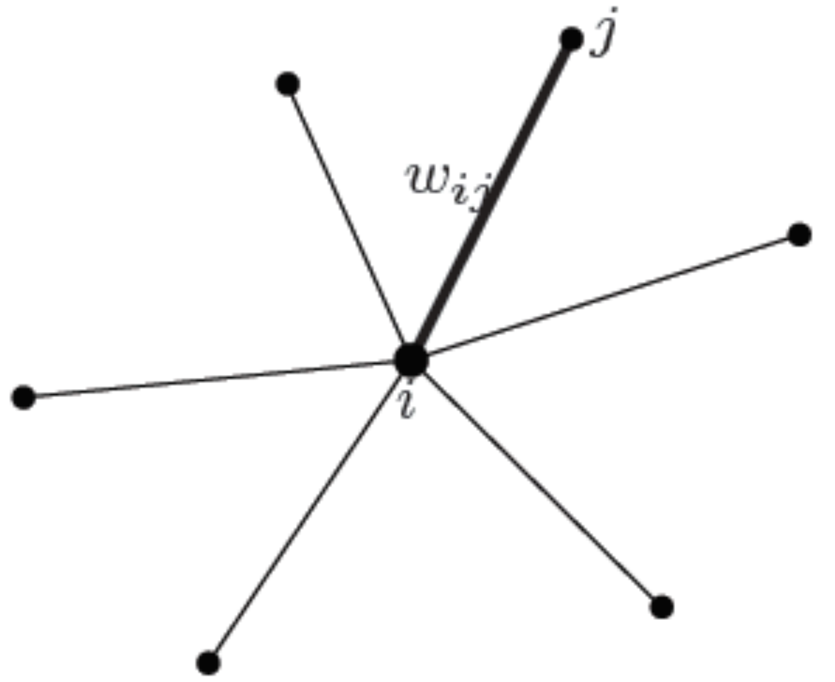
Other Parameterizations



[Masci et al. 2015]

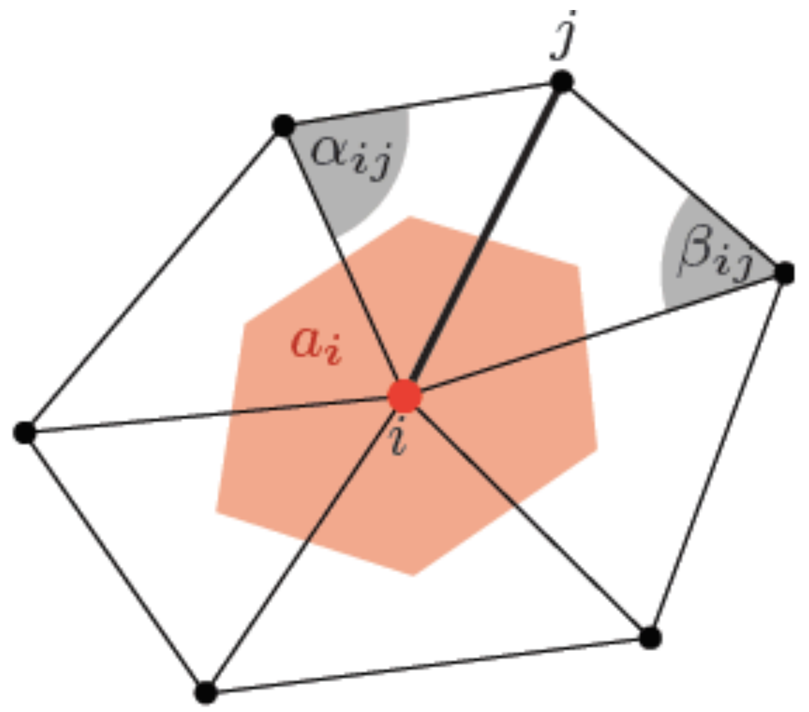
Discrete Laplacian

(slide credit: Michael Bronstein)



Undirected graph $(\mathcal{V}, \mathcal{E})$

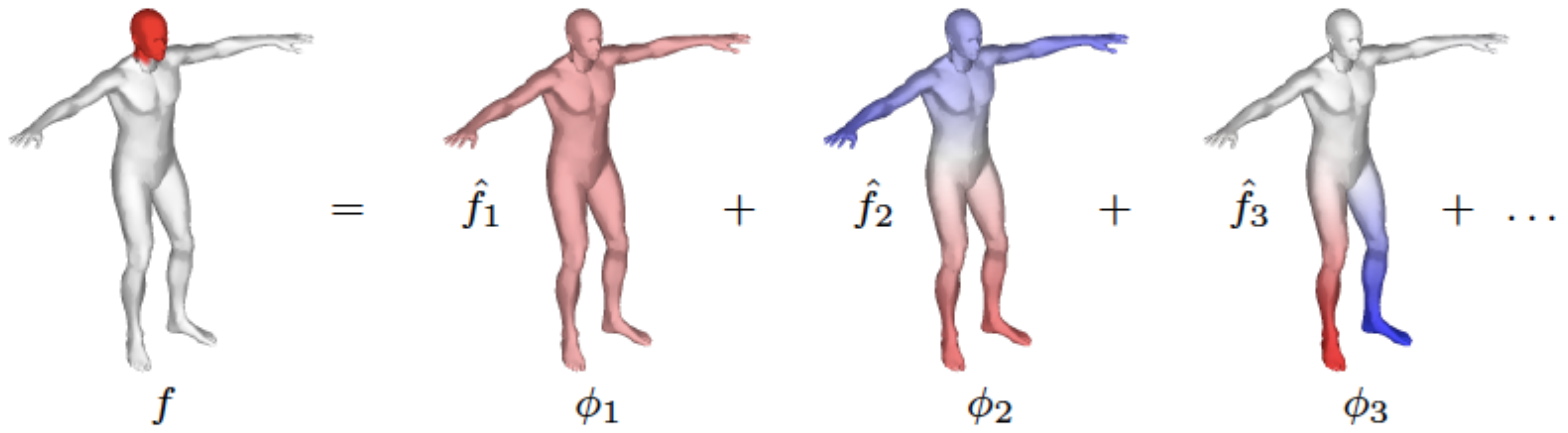
$$(\Delta f)_i \approx \sum_{(i,j)\in\mathcal{E}} w_{ij}(f_i - f_j)$$



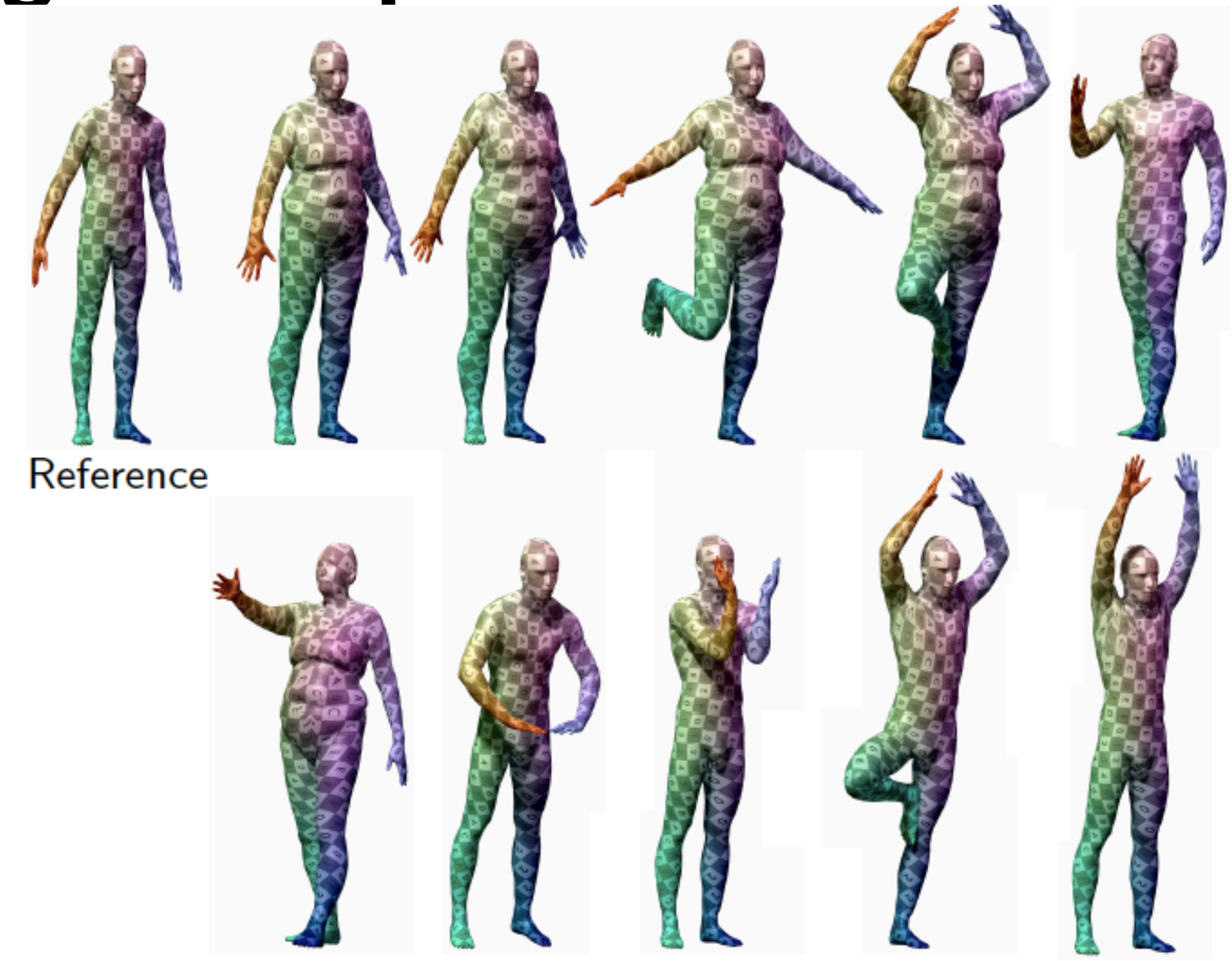
Triangular mesh $(\mathcal{V}, \mathcal{E}, \mathcal{F})$

$$(\Delta f)_i \approx \frac{1}{a_i} \sum_{(i,j)\in\mathcal{E}} \frac{\cot\alpha_{ij} + \cot\beta_{ij}}{2}(f_i - f_j)$$

a_i = local area element



Transferring Correspondence

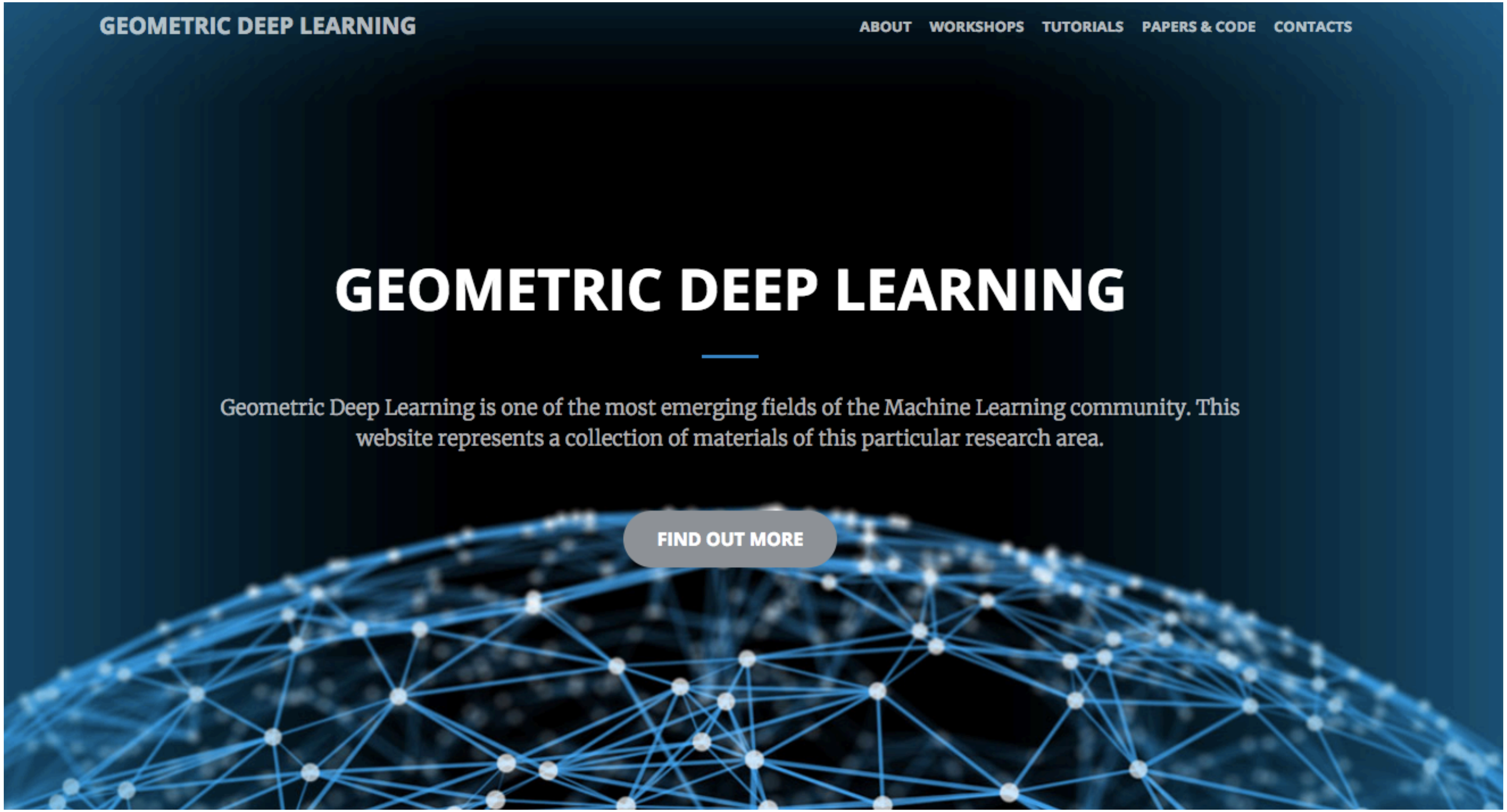


Texture transferred from reference to query shapes

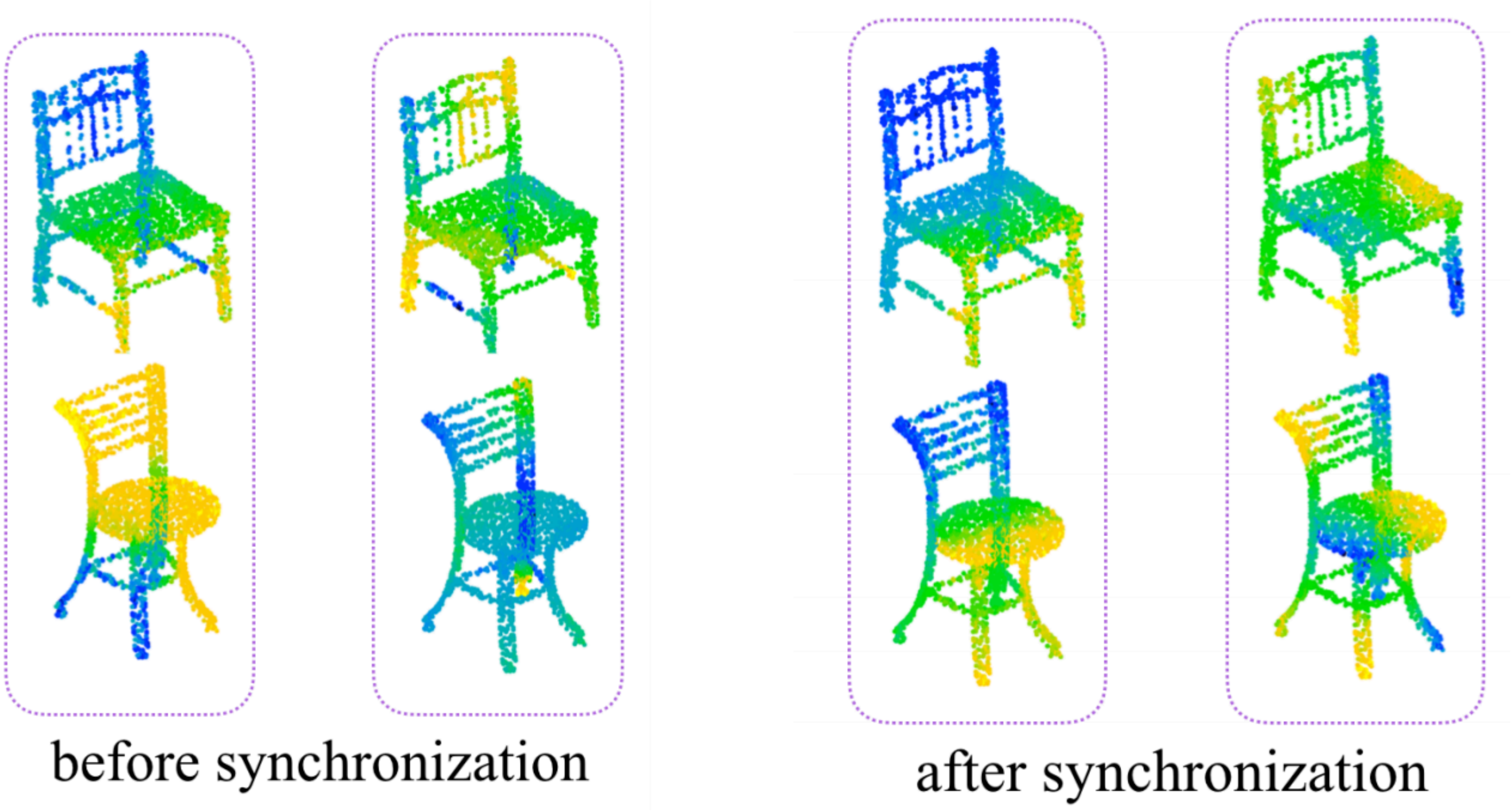
[Monti et al. 2016]

Spectral Methods

(slide credit: Michael Bronstein)



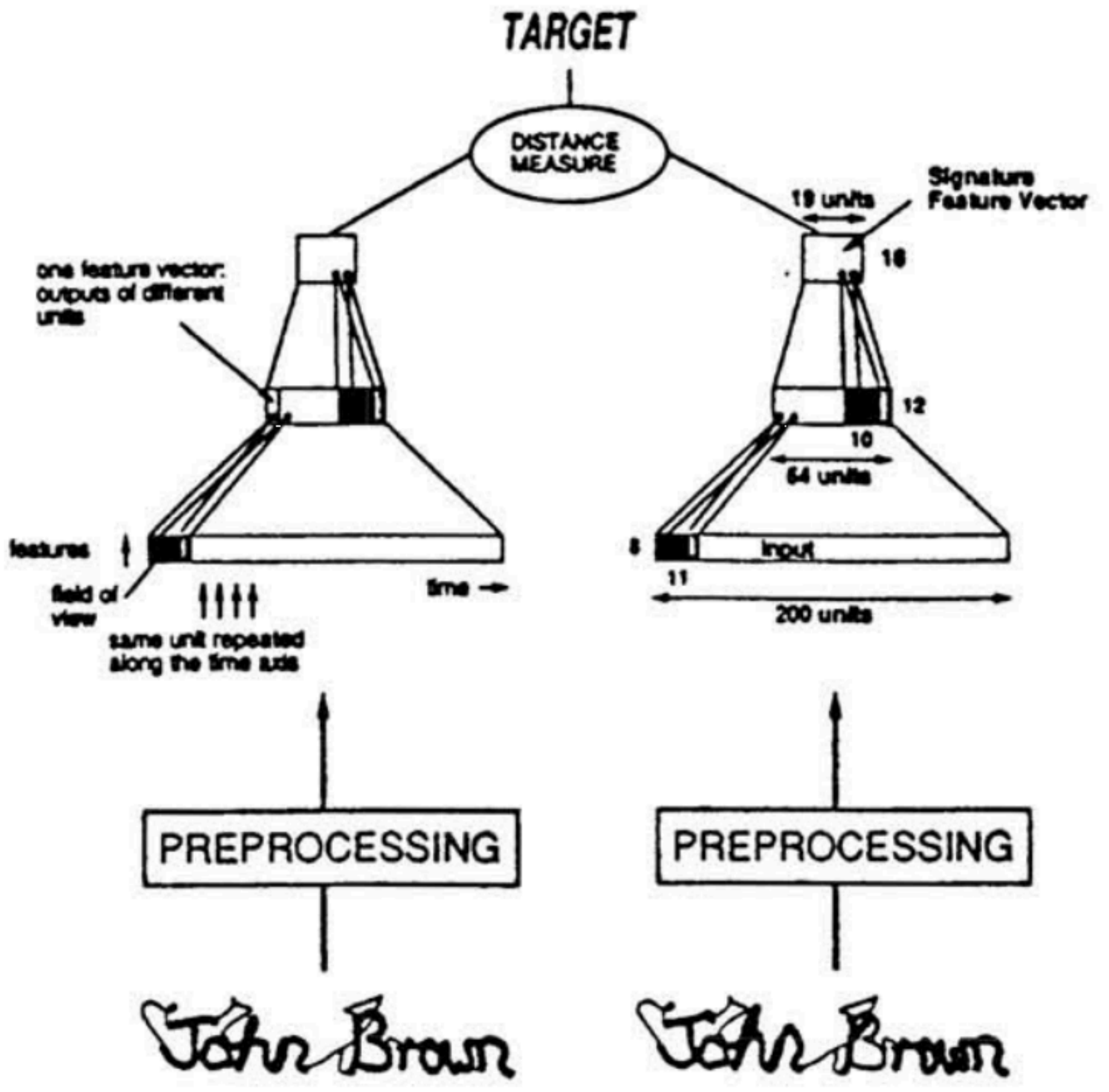
SyncSpecCNN



[Yi et al. 2017]

Siamese Networks

- Used to estimate quantities that depend on **pairs of representations**
- **Two networks** that merge at the end and that share parameters
- Triplet networks are used in practice as well



t-SNE (think MDS)

- Technique for dimensionality reduction of high-dimensional data
- focuses on **preserving local structure** (keeps similar points close together), *at the expense of global structure*
- Works well for visualizations

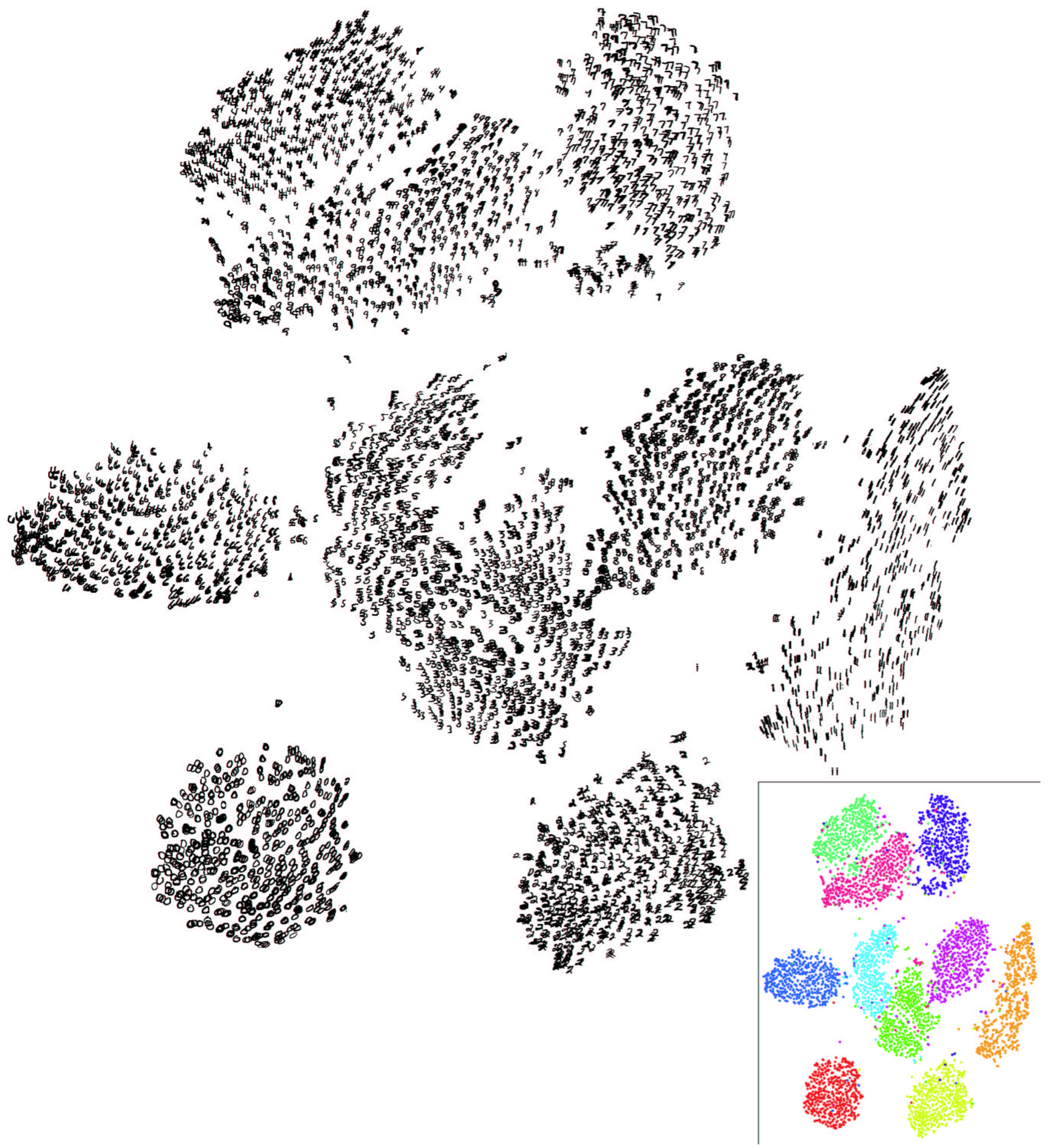
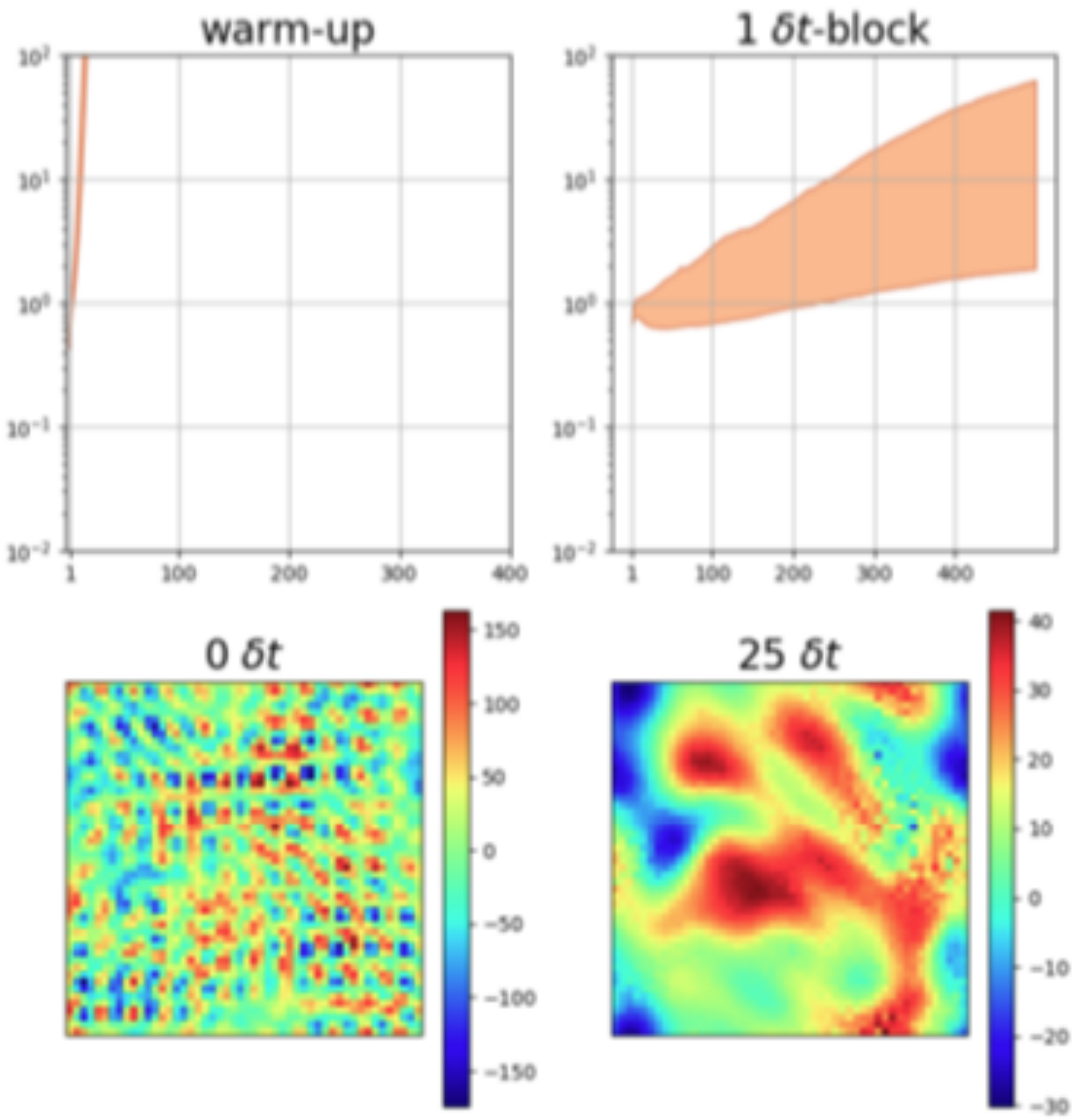


Image Credit: *Visualizing Data using* t-SNE, Maaten and Hinton

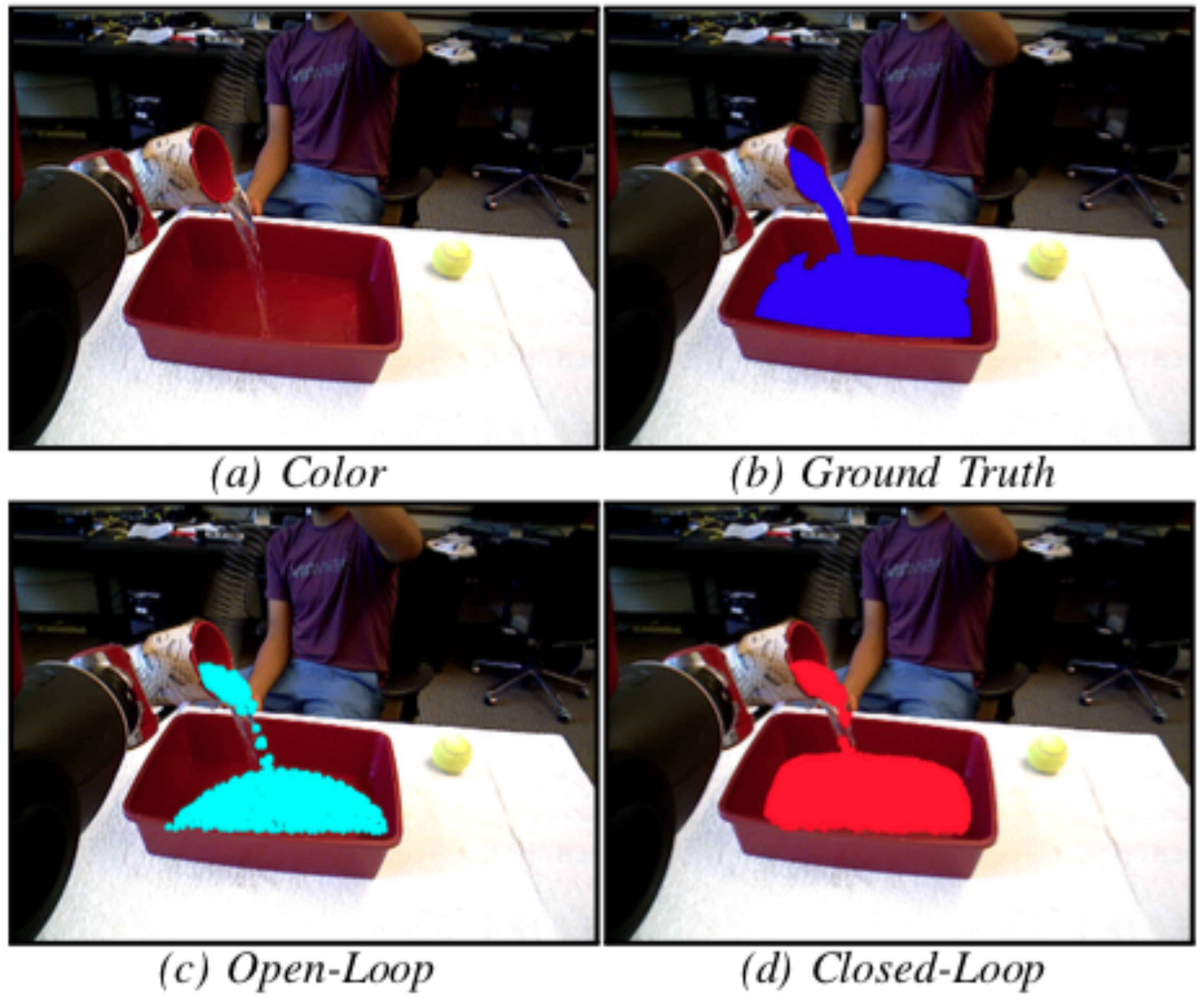
Deep Learning *for Fluids*

(slide credit: Nils Thuerey)

Tompson et. al 2017



Long et. al 2017



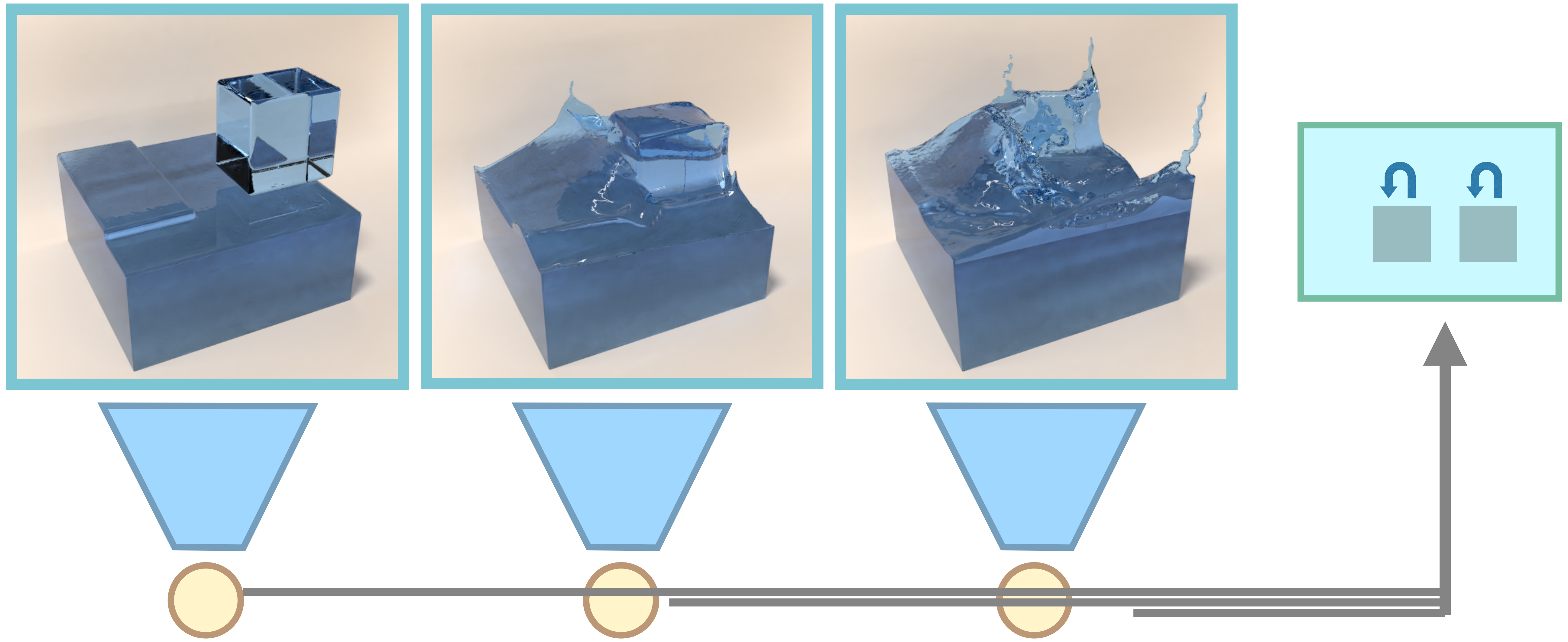
Schenck et. al 2017

High Resolution Simulation of Liquids (slide credit: Nils Thuerey)

High Resolution Simulation of Liquids (slide credit: Nils Thuerey)

Latent-space encoding

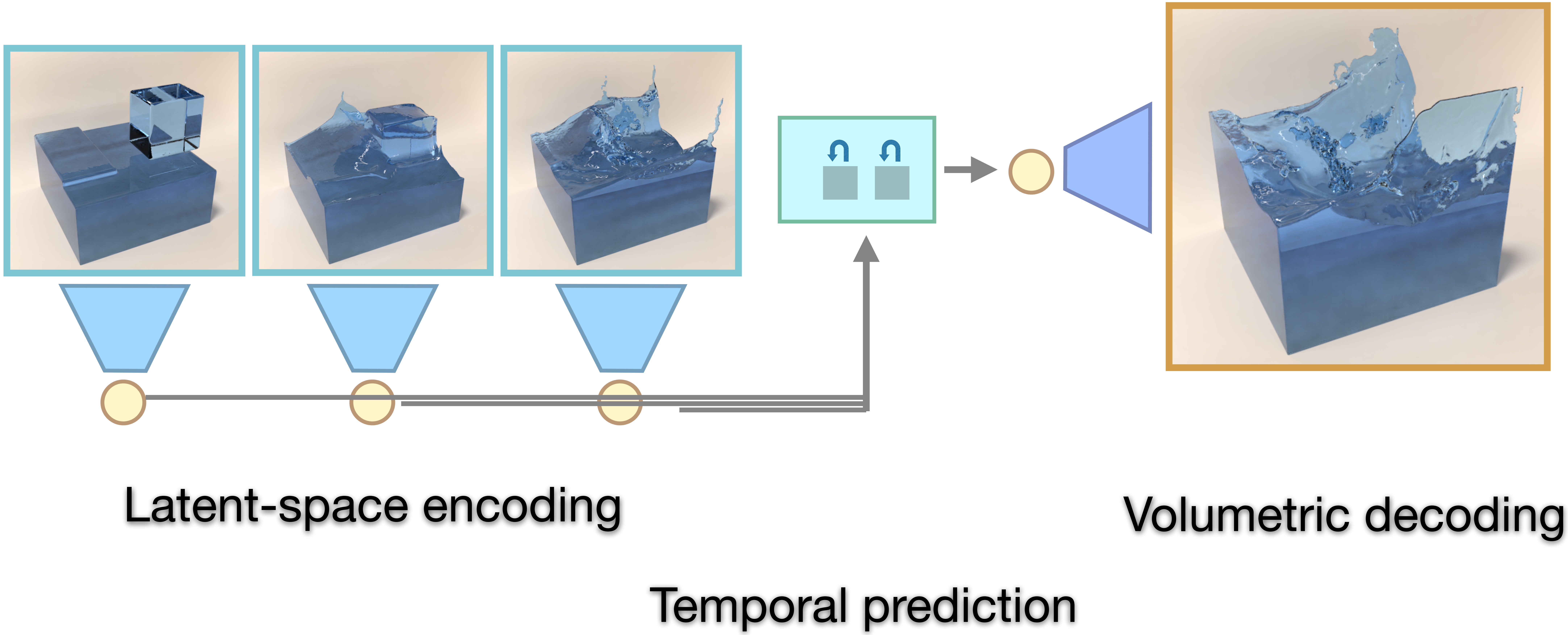
High Resolution Simulation of Liquids (slide credit: Nils Thuerey)



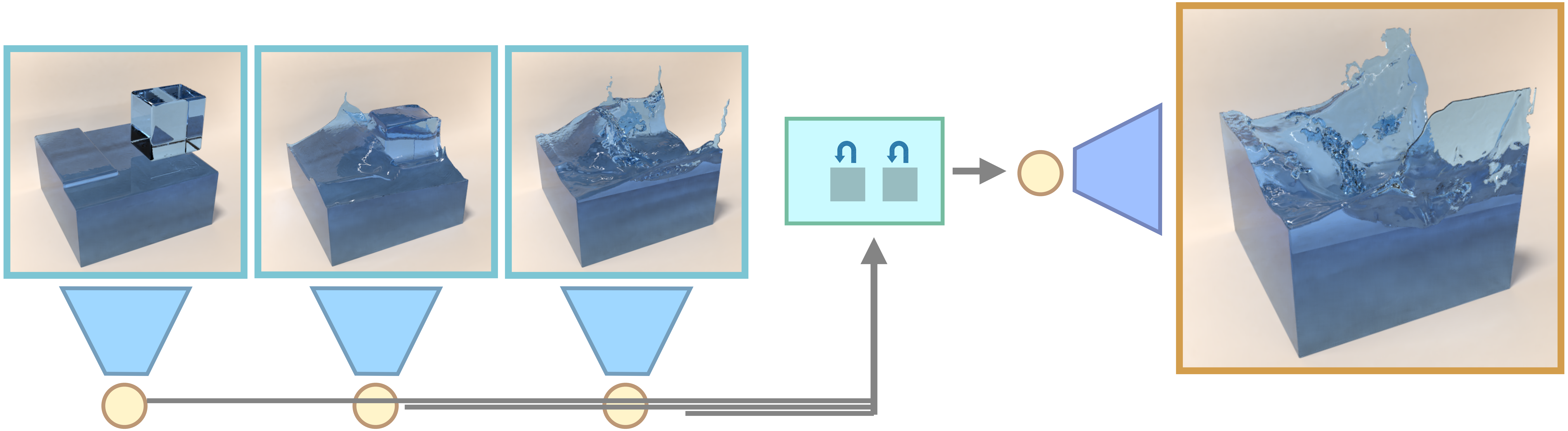
Latent-space encoding

Temporal prediction

High Resolution Simulation of Liquids (slide credit: Nils Thuerey)



High Resolution Simulation of Liquids (slide credit: Nils Thuerey)



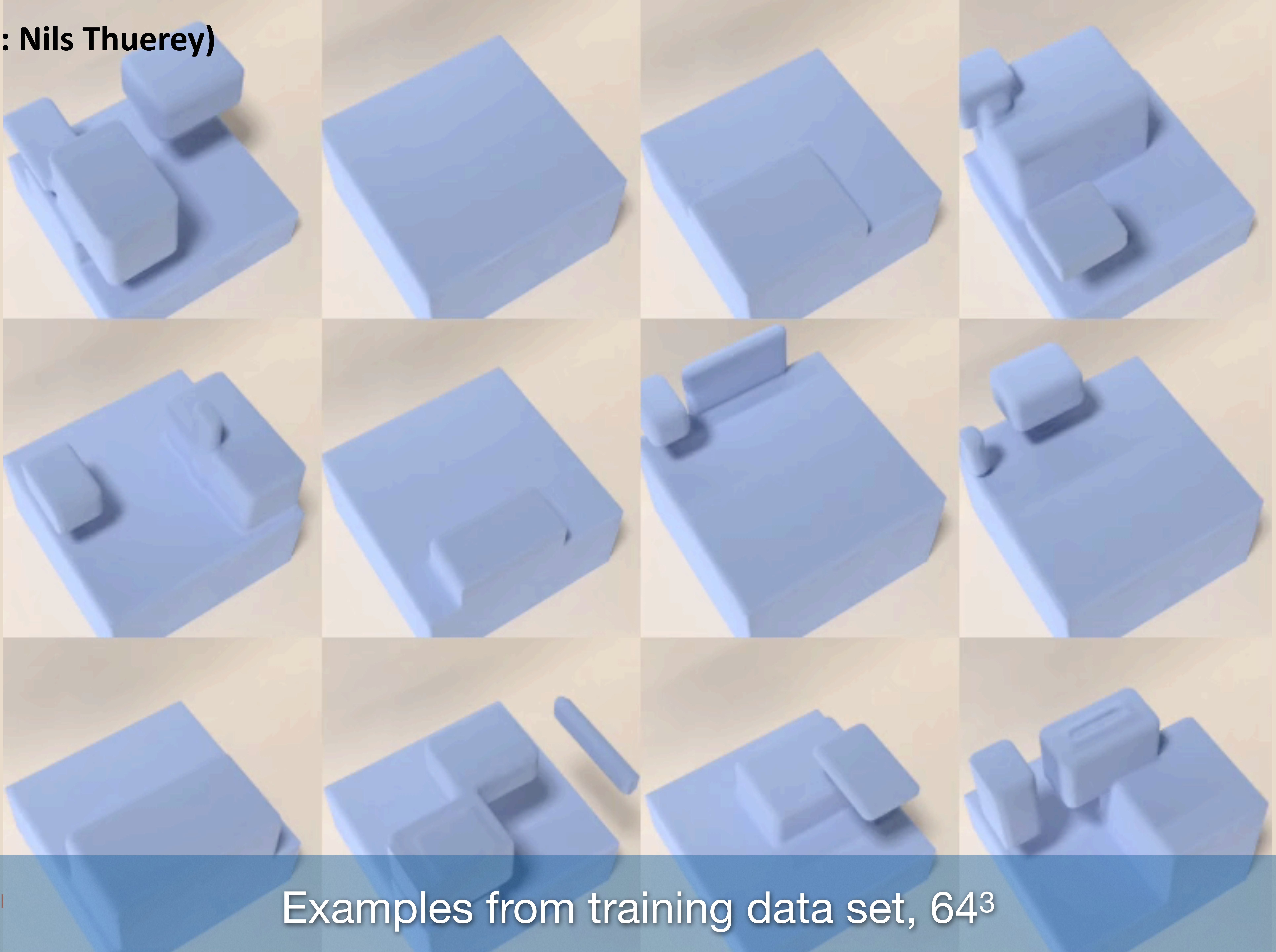
Latent-space encoding

Volumetric decoding

Temporal prediction

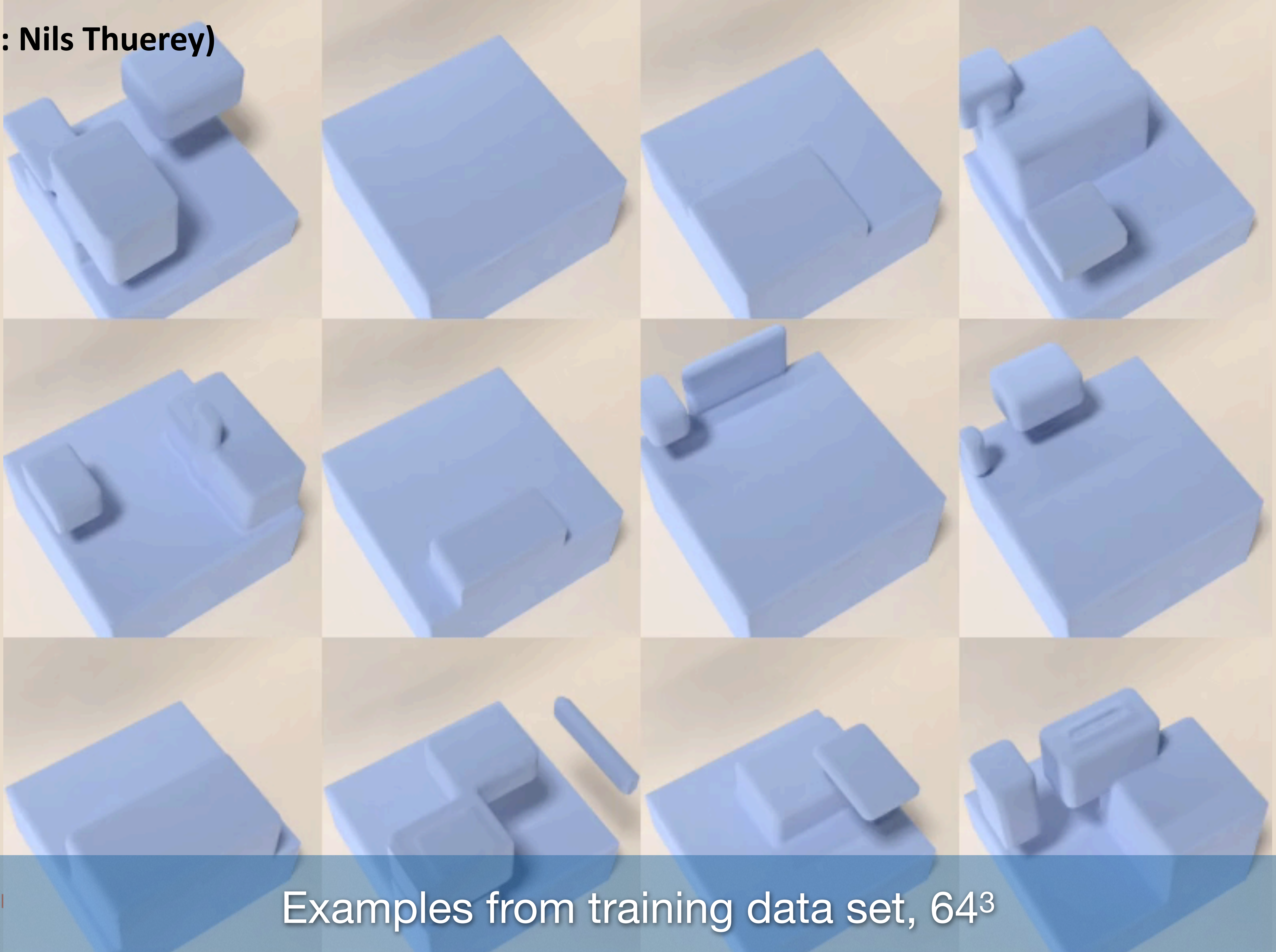
[Latent-space Physics: Towards Learning the Temporal Evolution of Fluid Flow, arXiv 2018]

(slide credit: Nils Thuerey)



Examples from training data set, 64^3

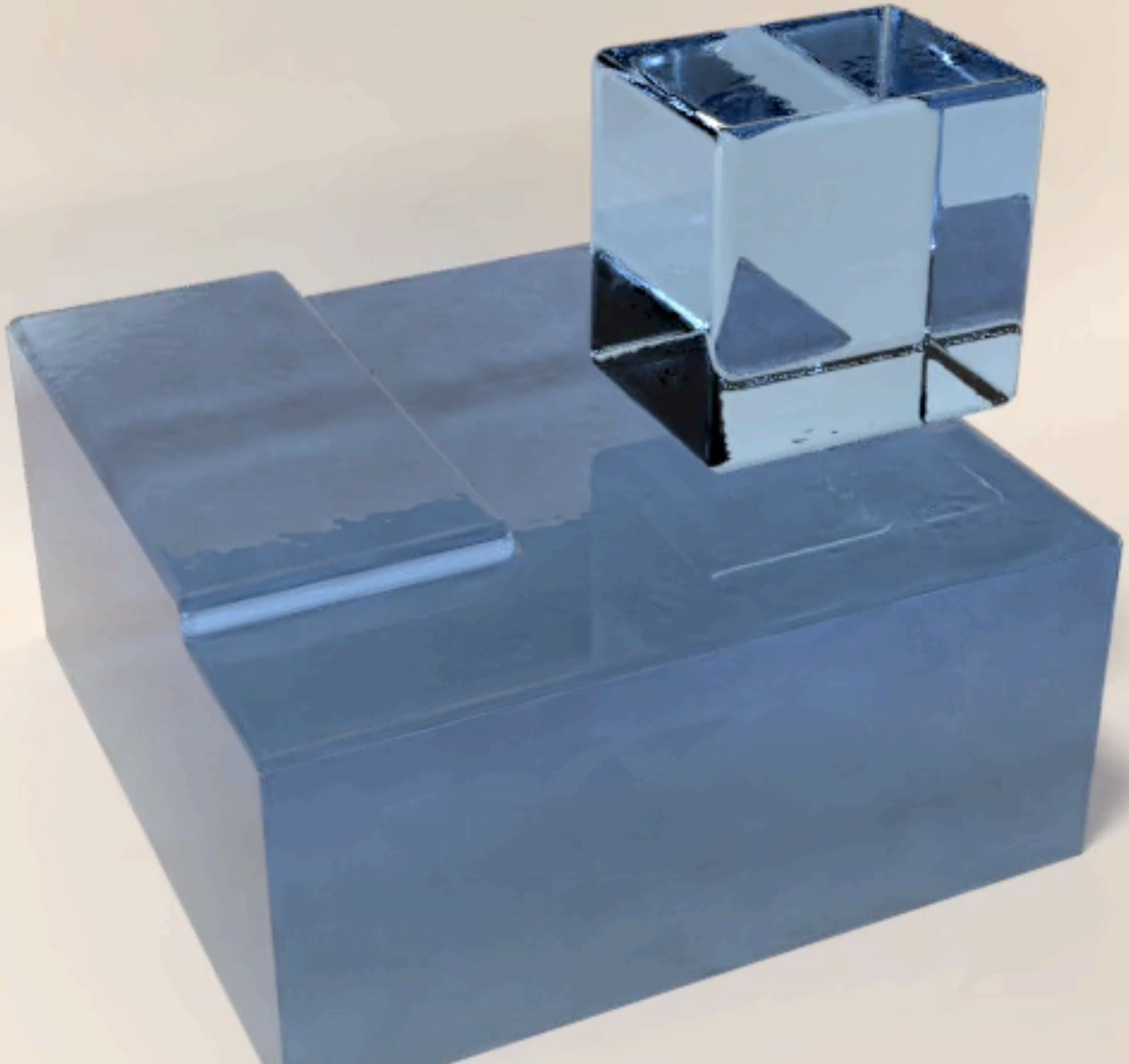
(slide credit: Nils Thuerey)



Examples from training data set, 64^3

(slide credit: Nils Thuerey)

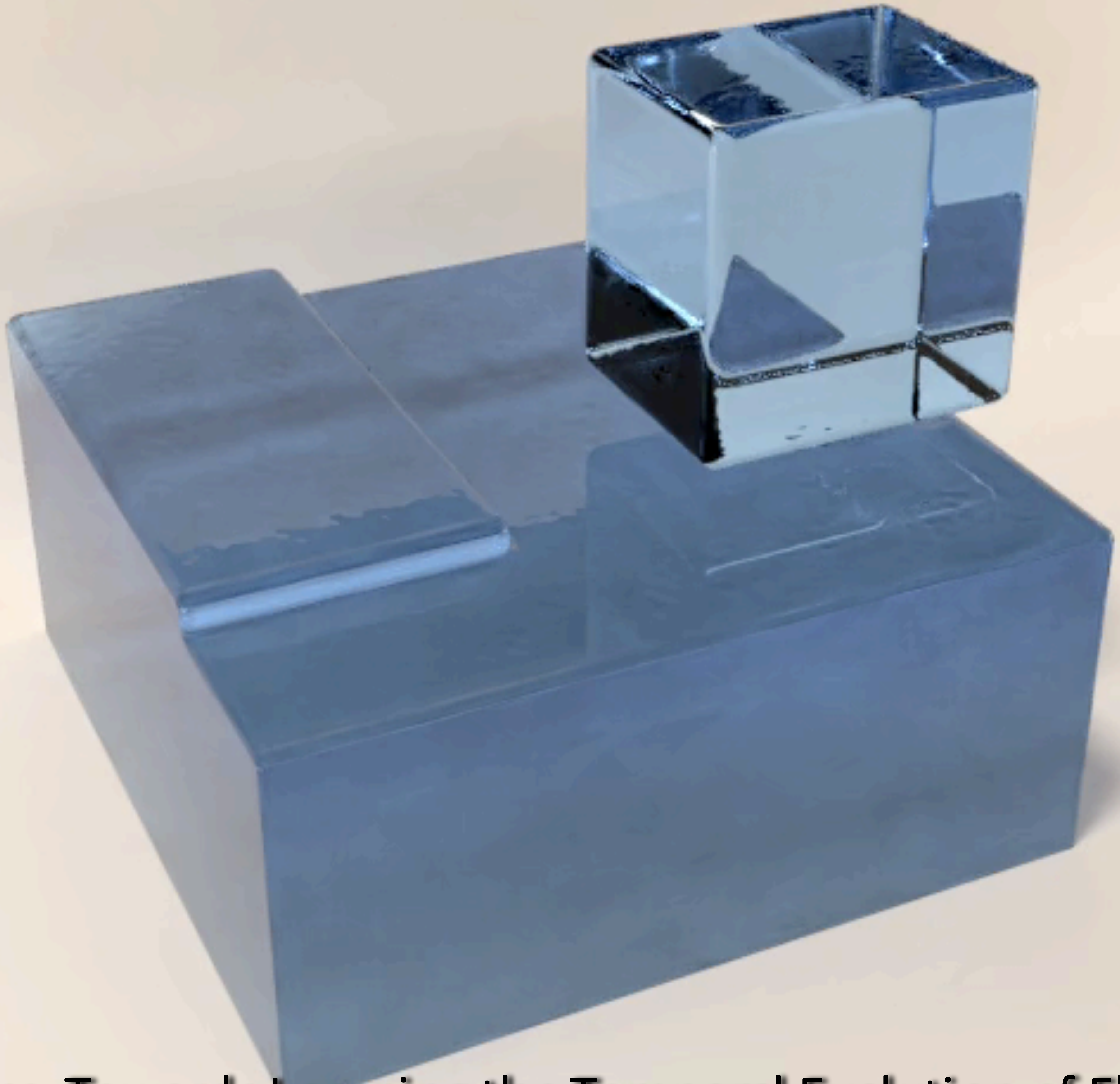
Further Examples, 128^3



[Latent-space Physics: Towards Learning the Temporal Evolution of Fluid Flow, arXiv 2018]

(slide credit: Nils Thuerey)

Further Examples, 128^3



[Latent-space Physics: Towards Learning the Temporal Evolution of Fluid Flow, arXiv 2018]

Example input (slide credit: Nils Thuerey) Example target (4x)

[tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow , SIGGRAPH 2018]

Example input

(slide credit: Nils Thuerey)

Example target (4x)

[tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow , SIGGRAPH 2018]

Example input

(slide credit: Nils Thuerey)

Example target (4x)

[tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow , SIGGRAPH 2018]

Example input

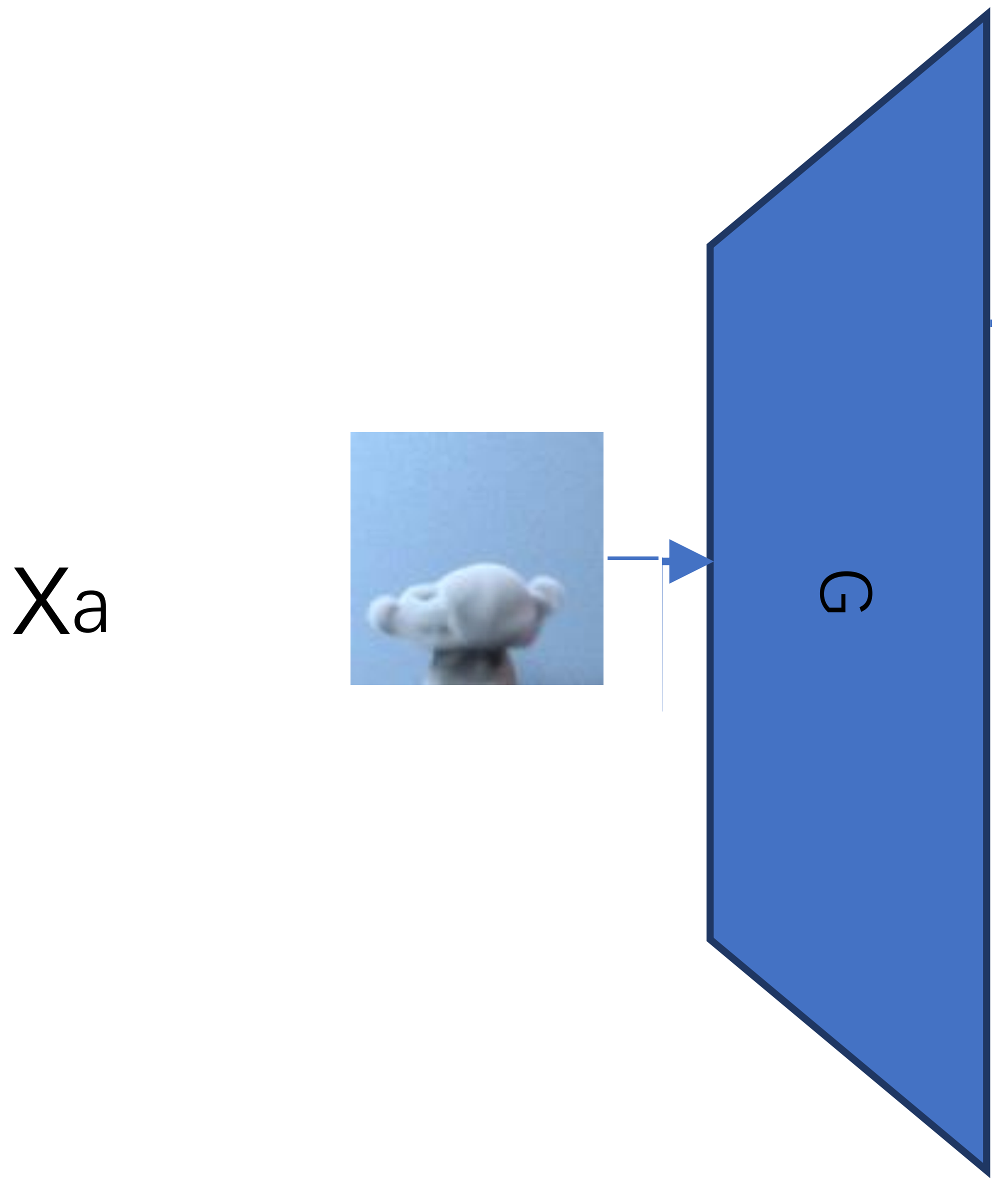
(slide credit: Nils Thuerey)

Example target (4x)

[tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow , SIGGRAPH 2018]

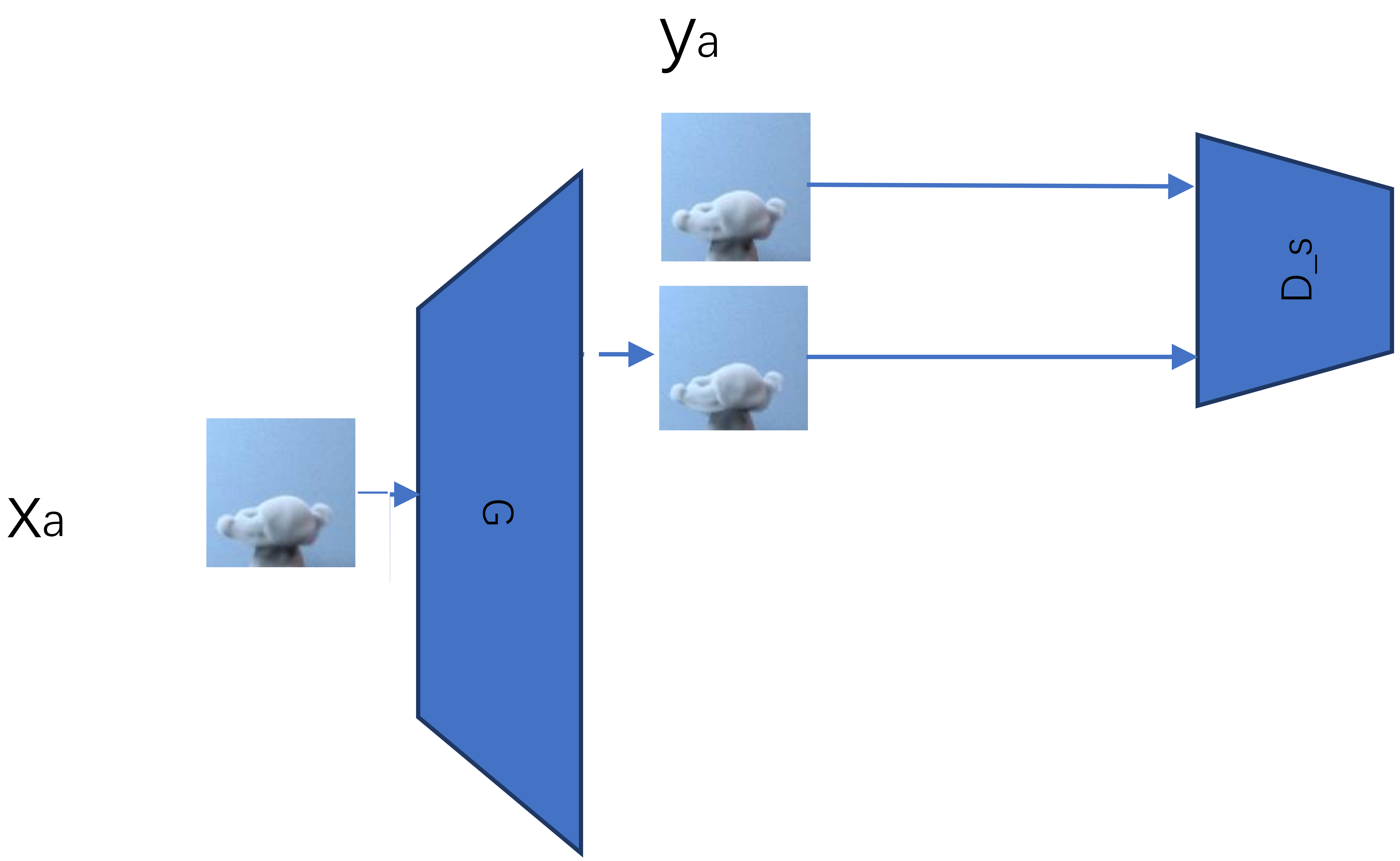
Architecture Overview

(slide credit: Nils Thuerey)



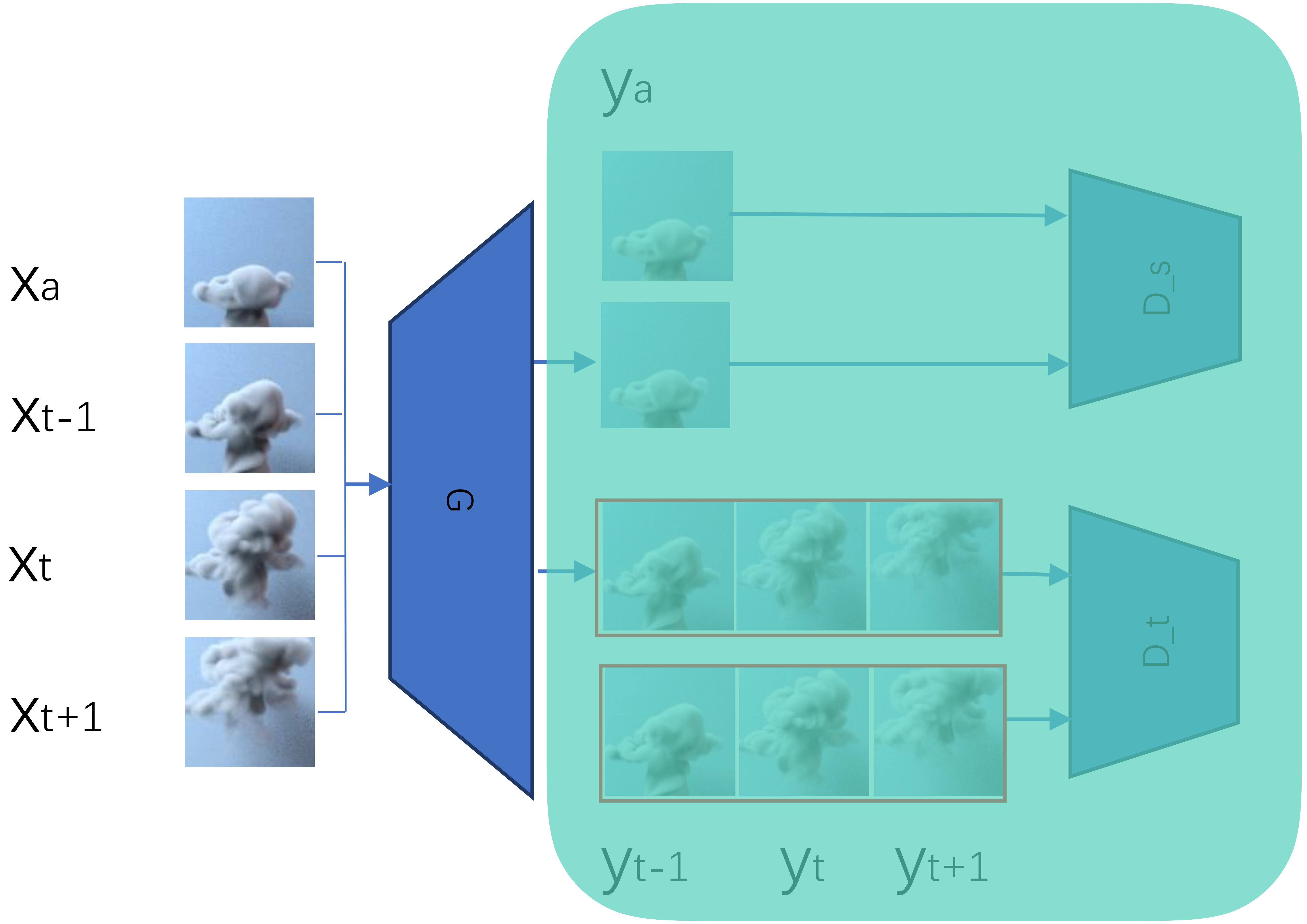
Architecture Overview

(slide credit: Nils Thuerey)



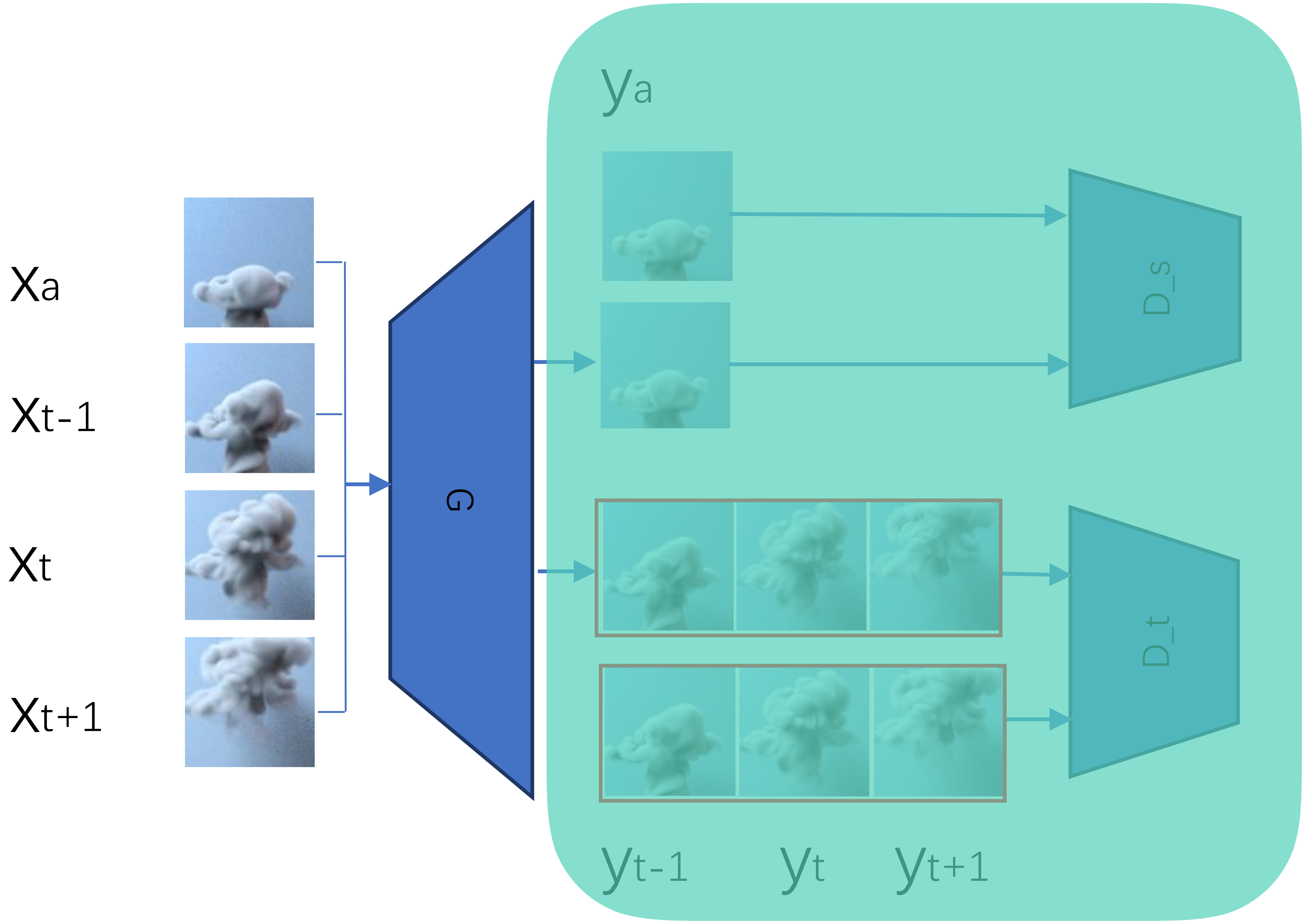
Architecture Overview

(slide credit: Nils Thuerey)



Architecture Overview

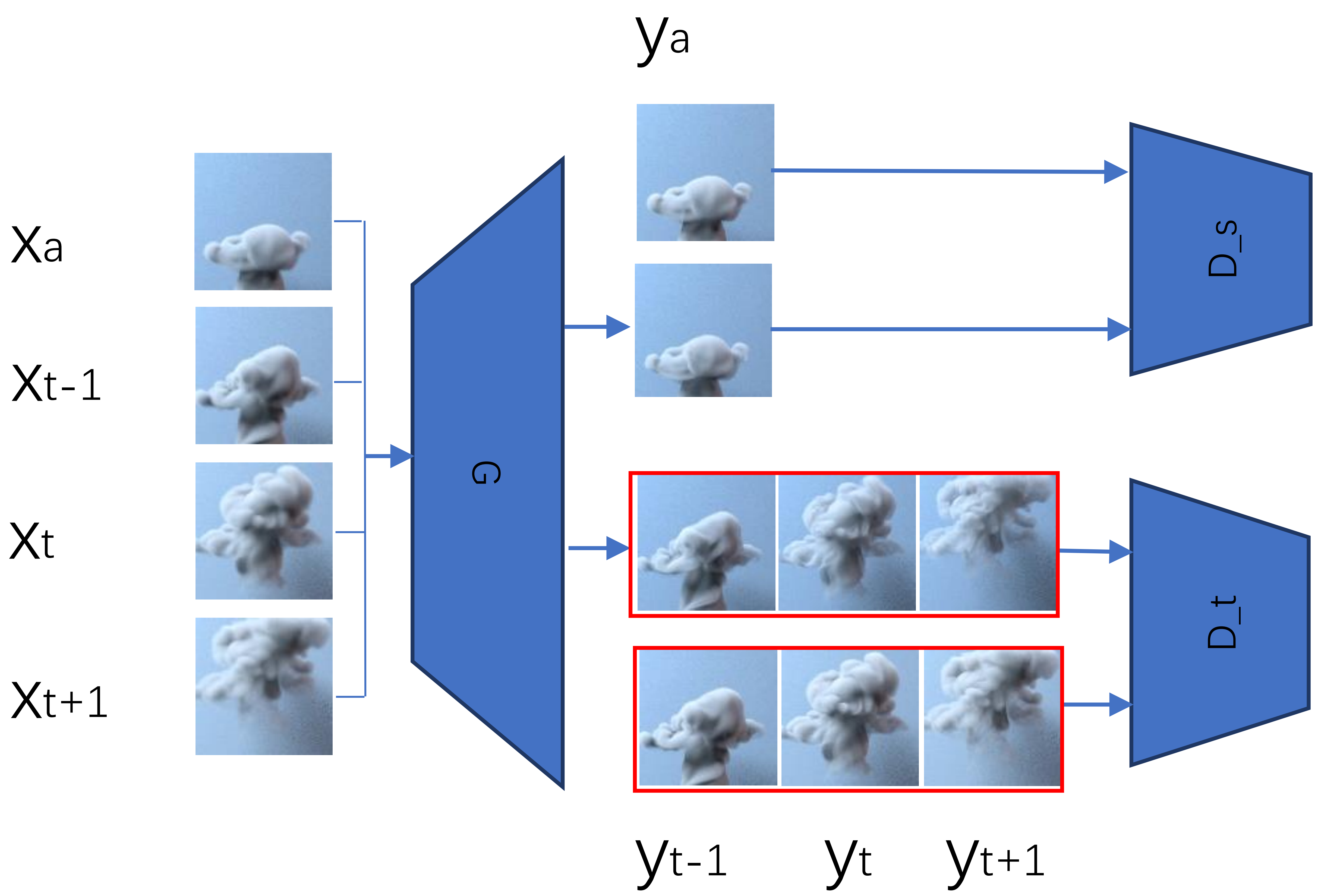
(slide credit: Nils Thuerey)



"Loss" for generator

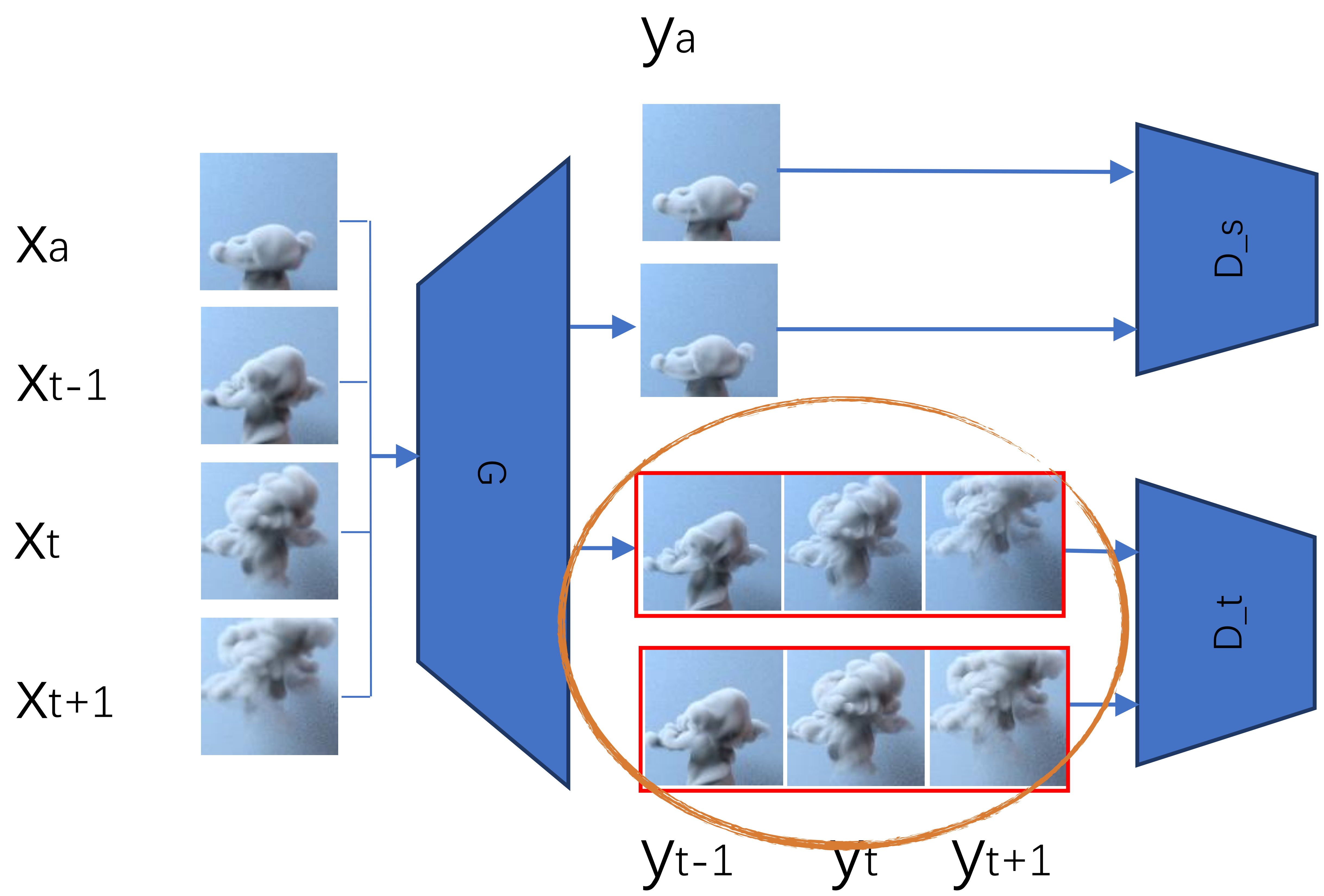
Architecture Overview

(slide credit: Nils Thuerey)



Architecture Overview

(slide credit: Nils Thuerey)

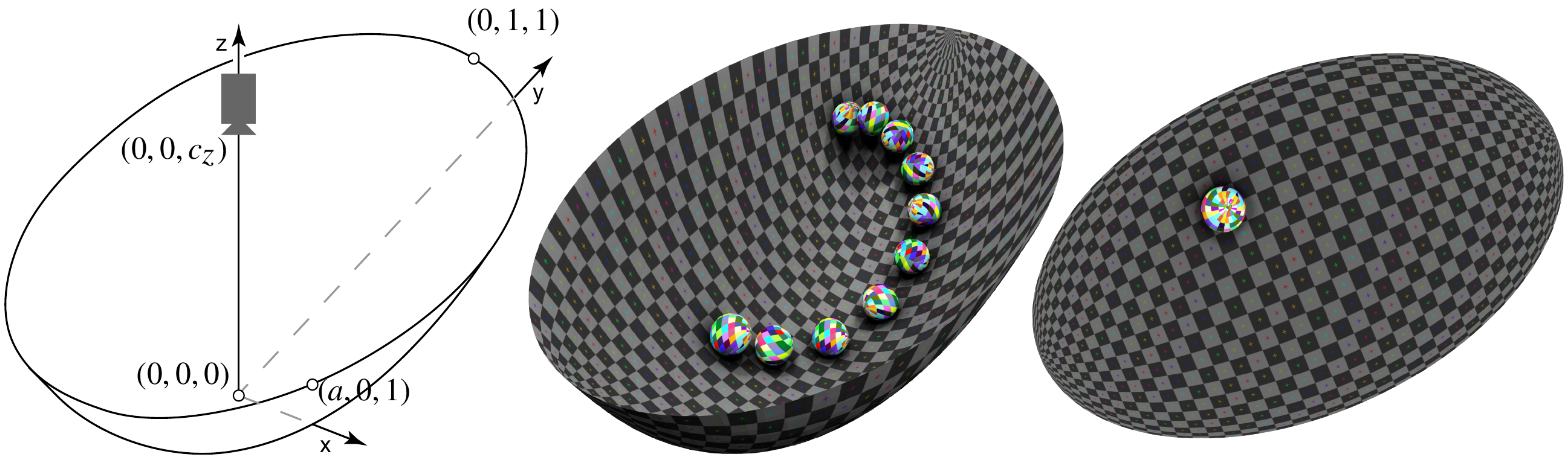


Advection encoded in loss for G


Input
tempoGAN
Target

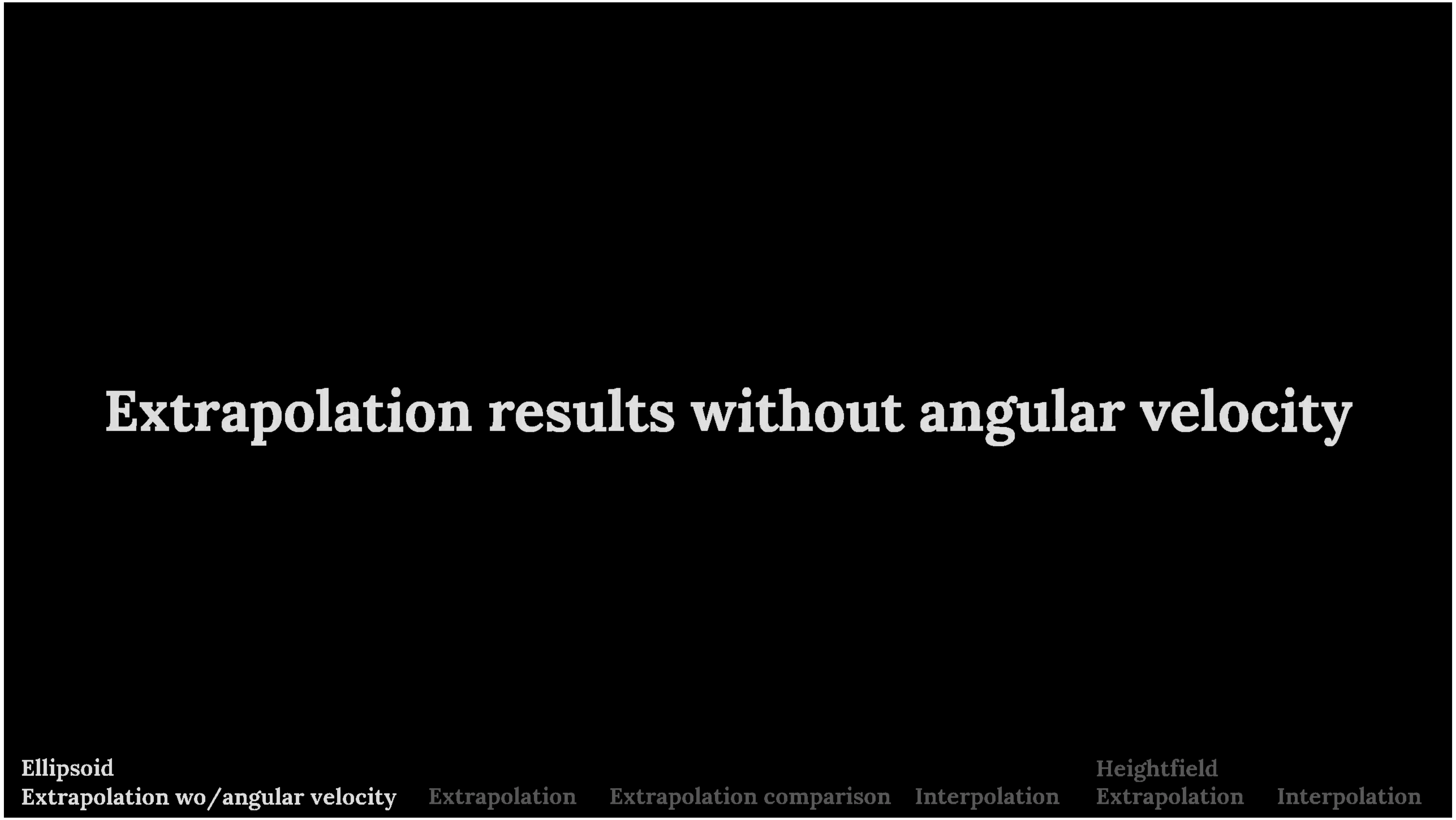

Input
tempoGAN
Target

Learning Rolling Motion

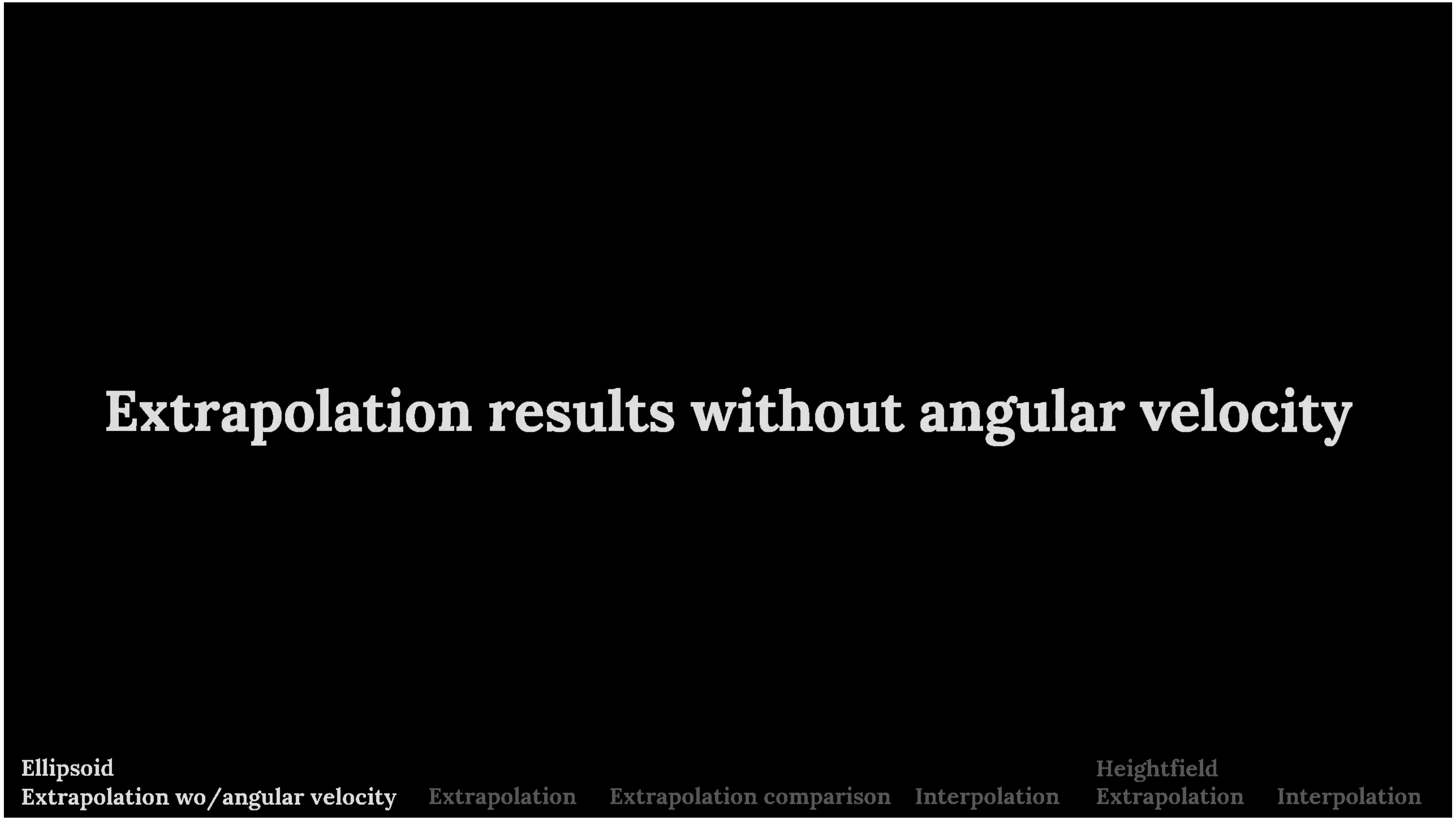


(a) (b) (c)

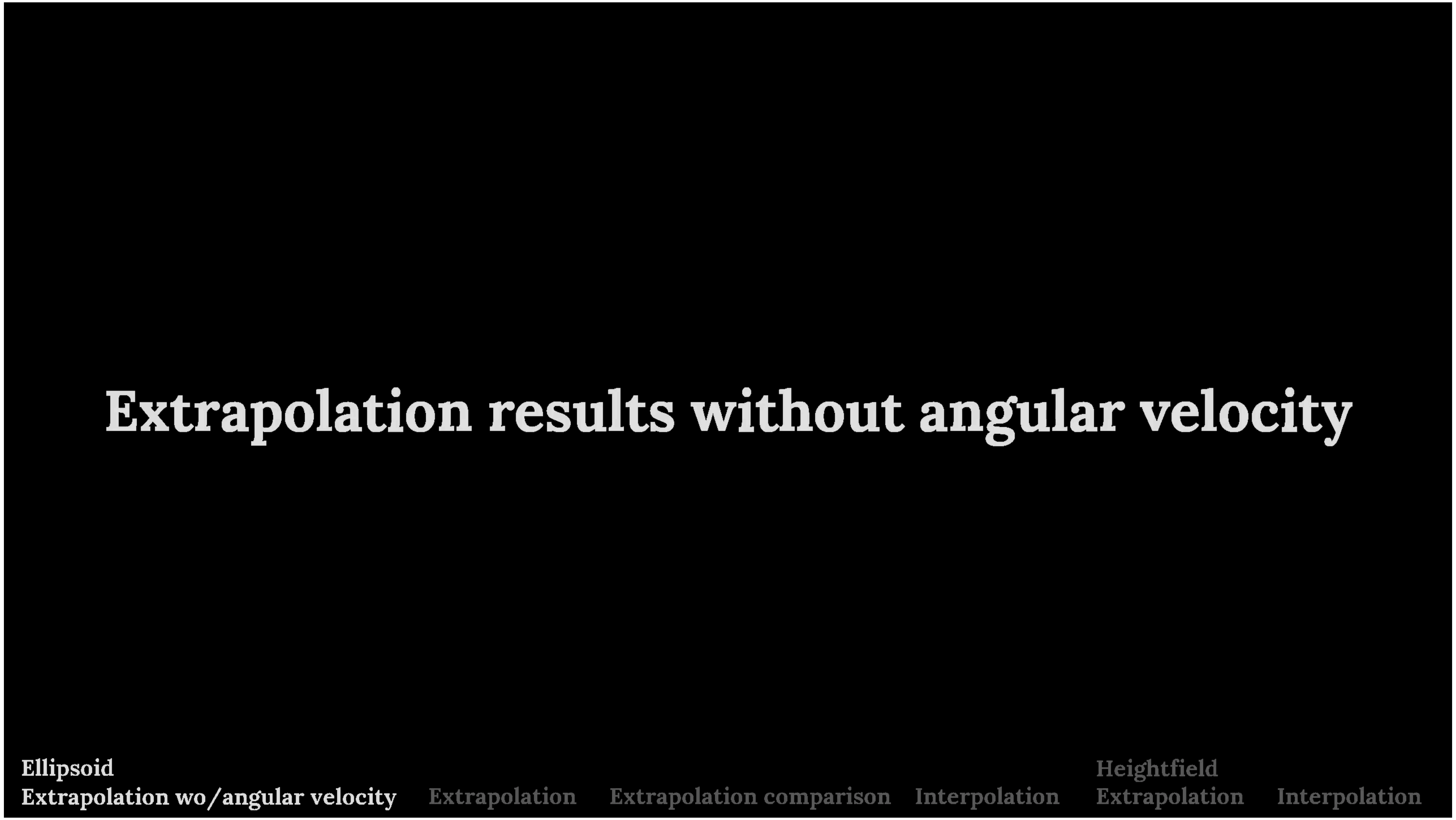
Learning Rolling Motion



Learning Rolling Motion



Learning Rolling Motion



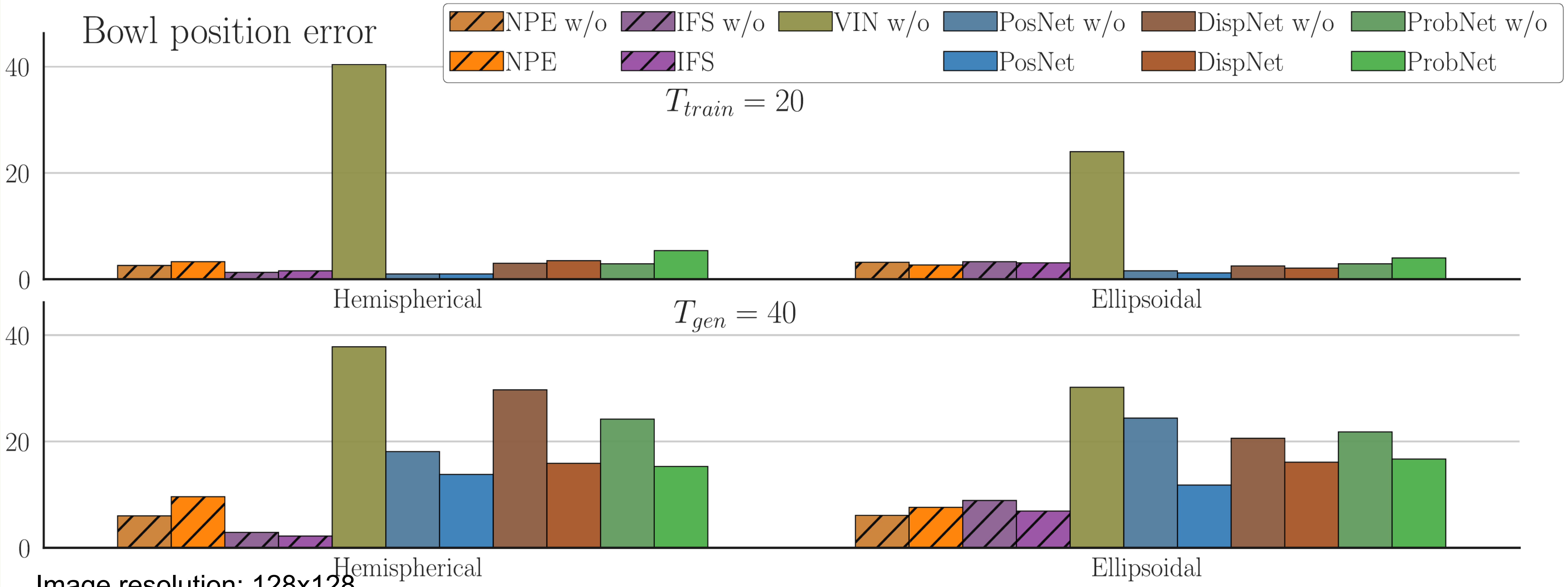


Image resolution: 128x128

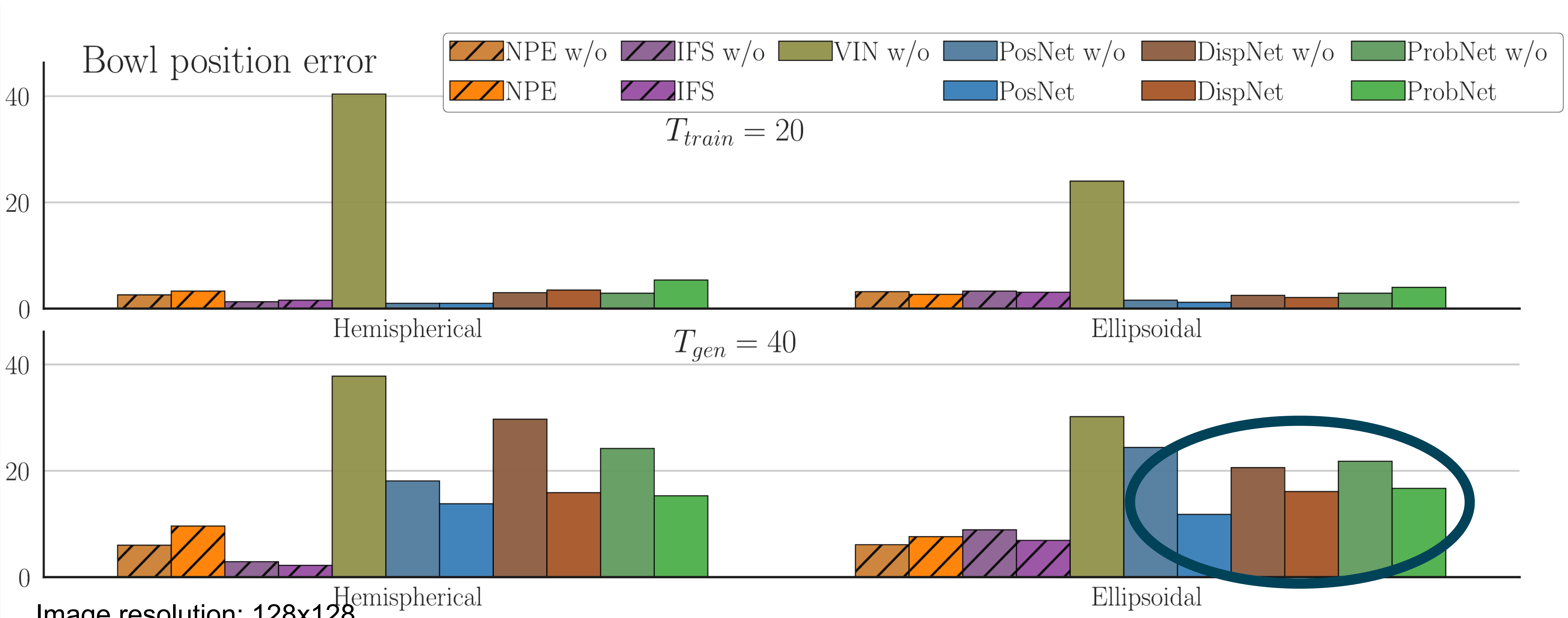


Image resolution: 128x128

Nicholas Watters, Andrea Tacchetti, Theophane Weber, Razvan Pascanu, Peter Battaglia, Daniel Zoran (DeepMind): **Visual Interaction Networks**, NIPS 2017

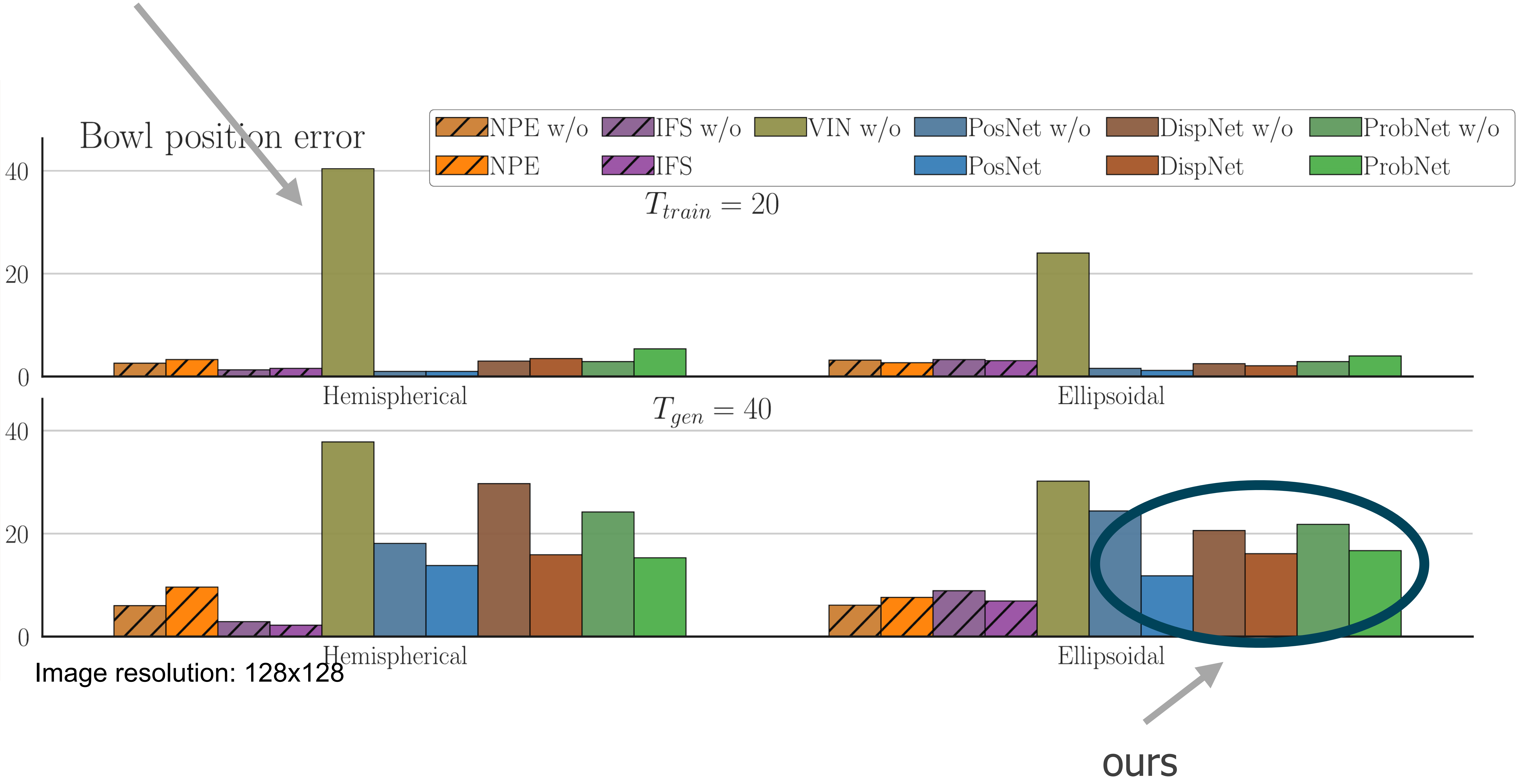
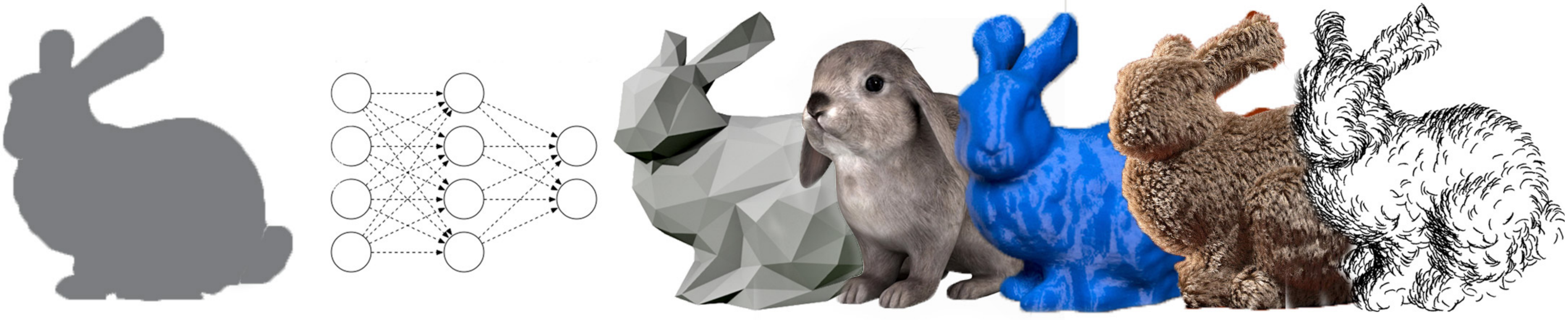


Image resolution: 128x128

Course Information (slides/code/comments)



http://geometry.cs.ucl.ac.uk/dl4g/