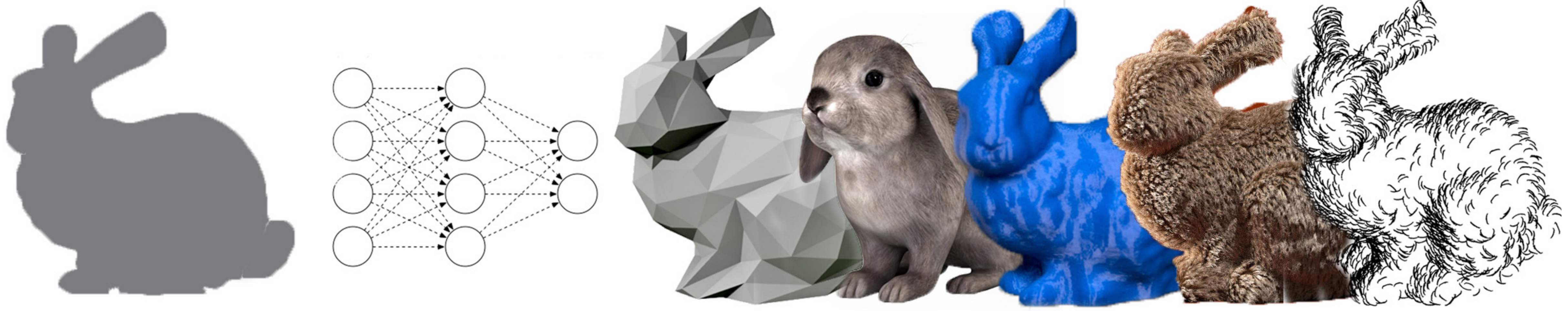


3D Domain (extrinsic)

Niloy Mitra — UCL

Iasonas Kokkinos — UCL

Federico Monti — USI Lugano

Emanuele Rodolà — La Sapienza

Michael Bronstein — Imperial College, USI Lugano

Or Litany — Stanford University, Facebook

Leonidas Guibas — Stanford University

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

Recap

Recap

- Supervised versus unsupervised; training/validation/test data

Recap

- Supervised versus unsupervised; training/validation/test data
- MLP, NN, CNN
    - backpropagation, SGD, momentum
    - overfitting, dropout, bottleneck, …
    - dilated convolutions
    - FCN, UNet, ….

Recap

- Supervised versus unsupervised; training/validation/test data
- MLP, NN, CNN
  - backpropagation, SGD, momentum
  - overfitting, dropout, bottleneck, …
  - dilated convolutions
  - FCN, UNet, ….
- AutoEncoded, VariationalAutoEncoder

Recap

- Supervised versus unsupervised; training/validation/test data
- MLP, NN, CNN
  - backpropagation, SGD, momentum
  - overfitting, dropout, bottleneck, …
  - dilated convolutions
  - FCN, UNet, ….
- AutoEncoded, VariationalAutoEncoder
- Generative Adversarial Network
  - Cycle consistency
  - Latent GANs
  - Progressive GANs, Disentanglement, etc.

Timetable

			Niloy	Federico	Iasonas	Emanuele
Theory/Basics	Introduction	9:00	X	X	X	X
	Machine Learning Basics	~ 9:05	X			
	Neural Network Basics	~ 9:35		X		
	Alternatives to Direct Supervision (GANs)	~11:00			X	
State of the Art	Image Domain	~11:45			X	
	3D Domains (extrinsic)	~13:30	X			
	3D Domains (intrinsic)	~ 14:15				X
	Physics and Animation	~ 16:00	X			
	Discussion	~ 16:45	X	X	X	X

Sessions: A. 9:00-10:30 (**coffee**) B. 11:00-12:30 [**LUNCH**] C. 13:30-15:00 (**coffee**) D. 15:30-17:00

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

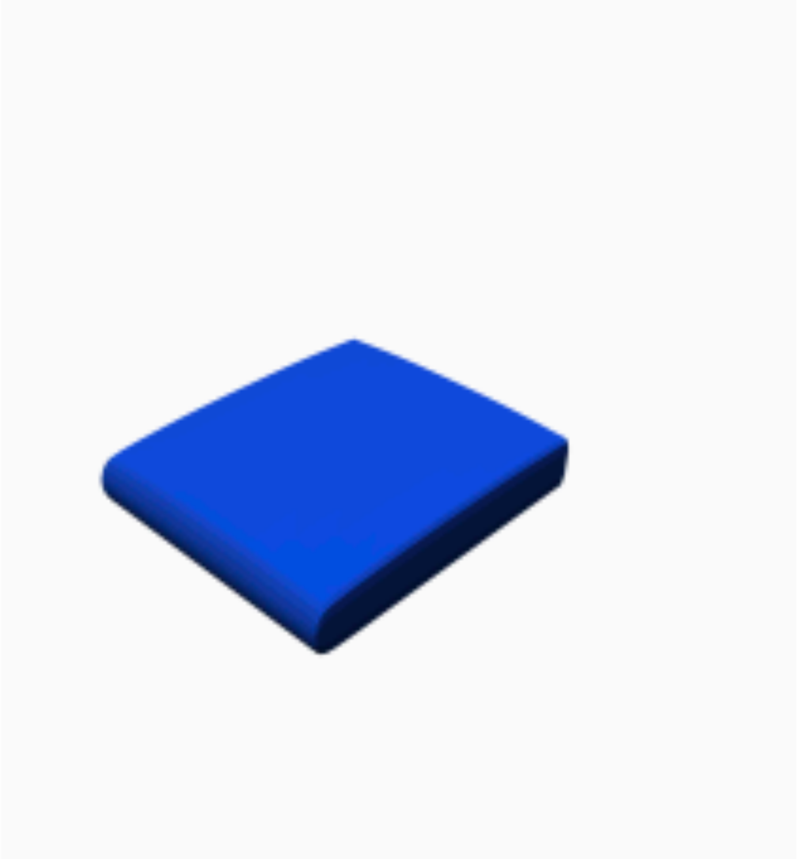
- 3D modeling, retrieval, classification for AR and VR
- Joint multi-modal understanding
- Semantic 3D reconstruction

Motivating Applications

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

Application #1: 3D Modeling

- Modeling by example, revisited

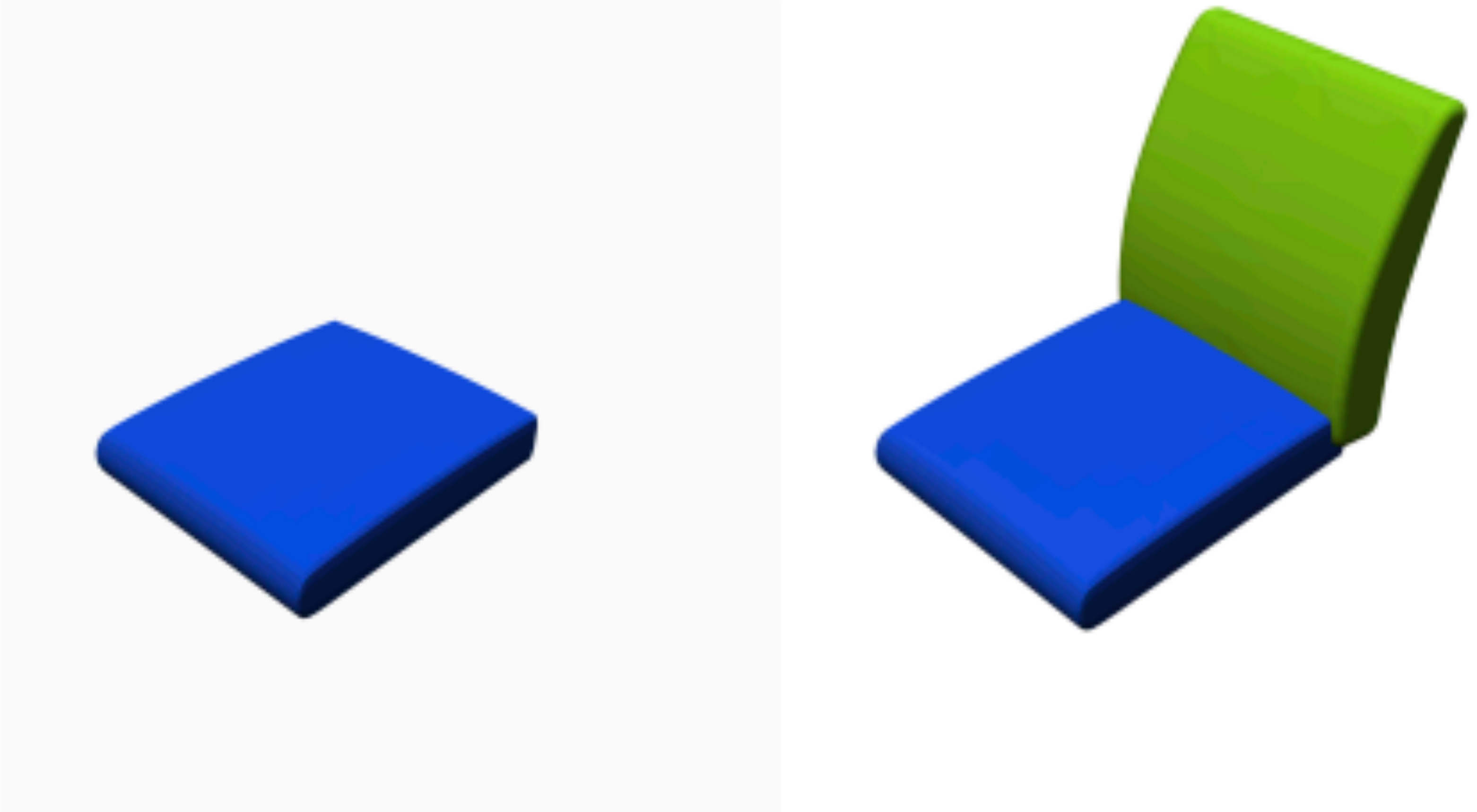


Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

[Sung et al. 2017]

Application #1: 3D Modeling

- Modeling by example, revisited



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]

Application #1: 3D Modeling

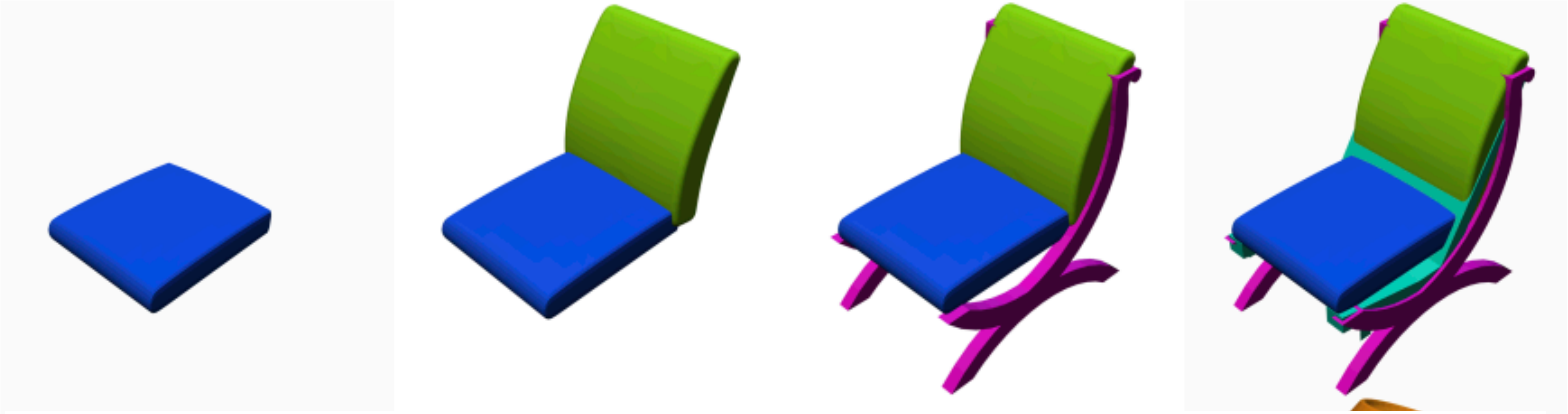
- Modeling by example, revisited

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]

Application #1: 3D Modeling

- Modeling by example, revisited

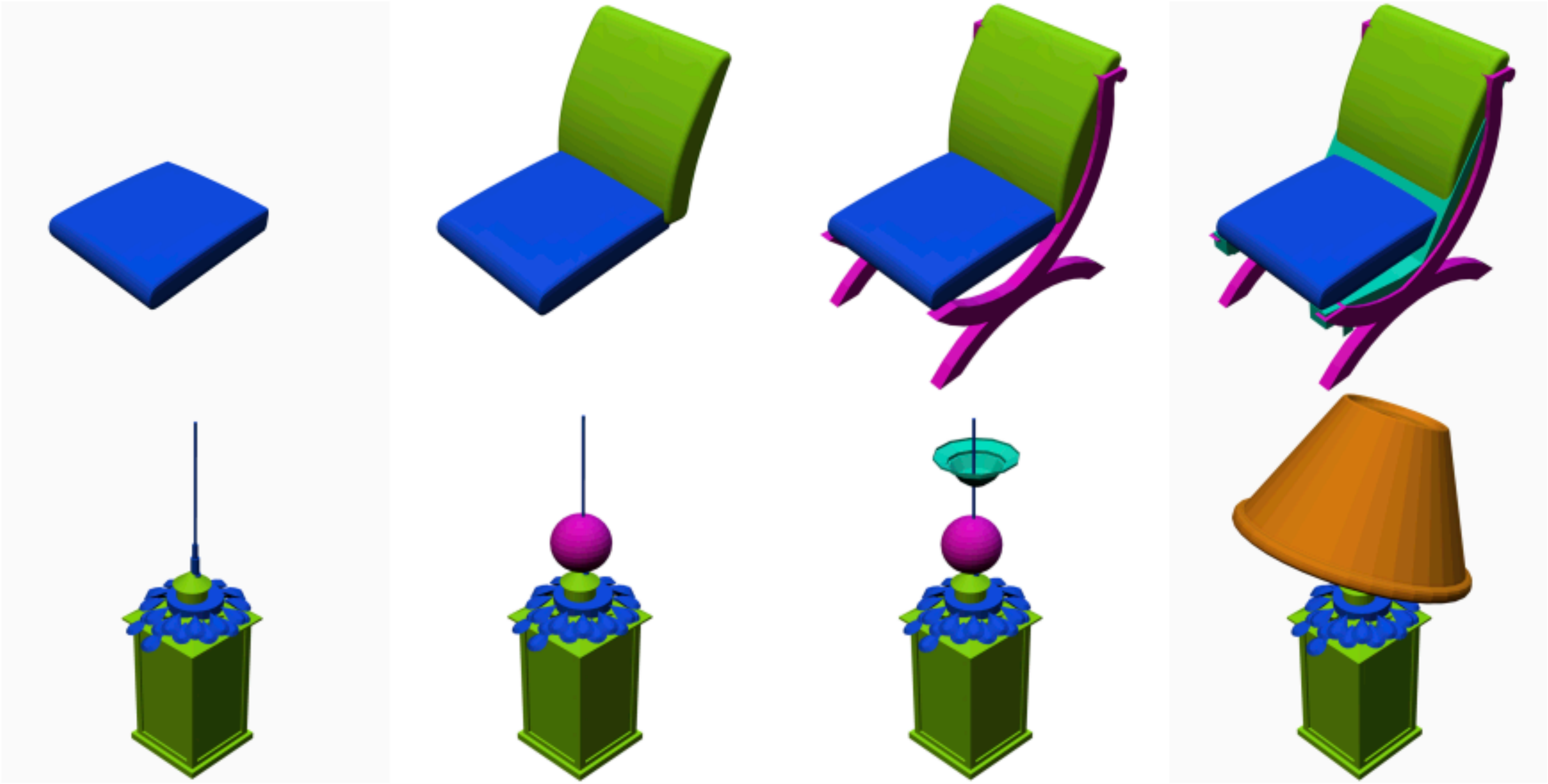


Deep neural network predicts the **next best part** to add and its **position** to enable non-expert users to create novel shapes.

[Sung et al. 2017]

Application #1: 3D Modeling

- Modeling by example, revisited



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]

Application #2: Image Understanding

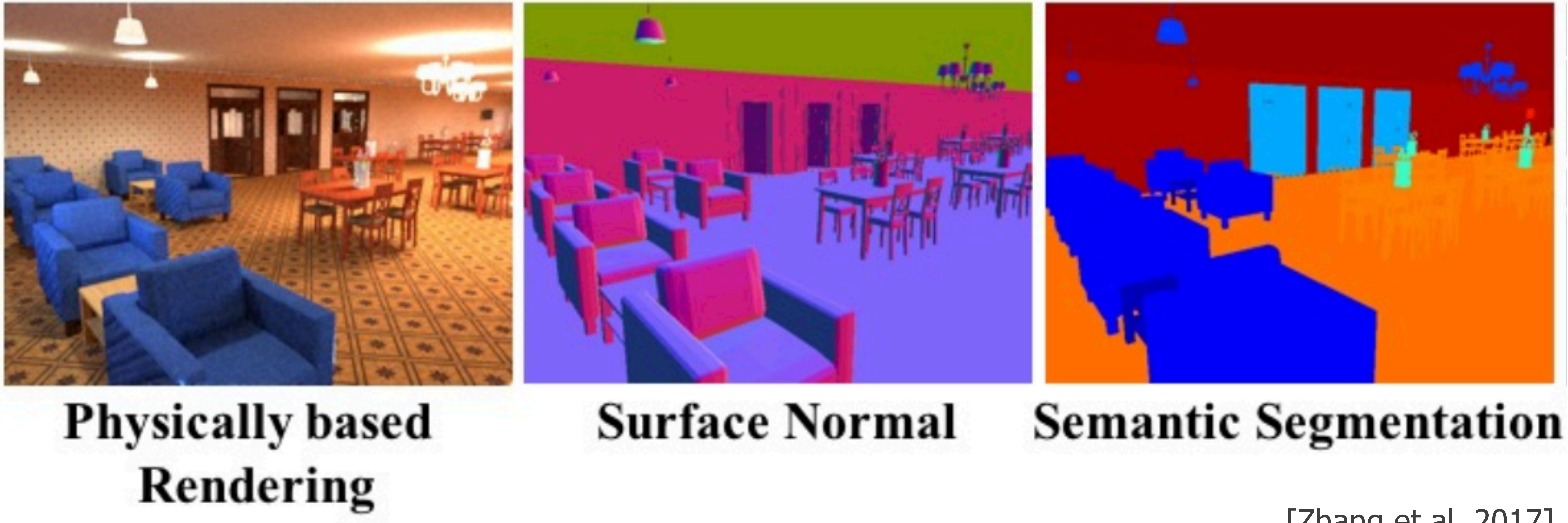
understanding 3D shapes can benefit image understanding

Physically based Rendering

[Zhang et al. 2017]

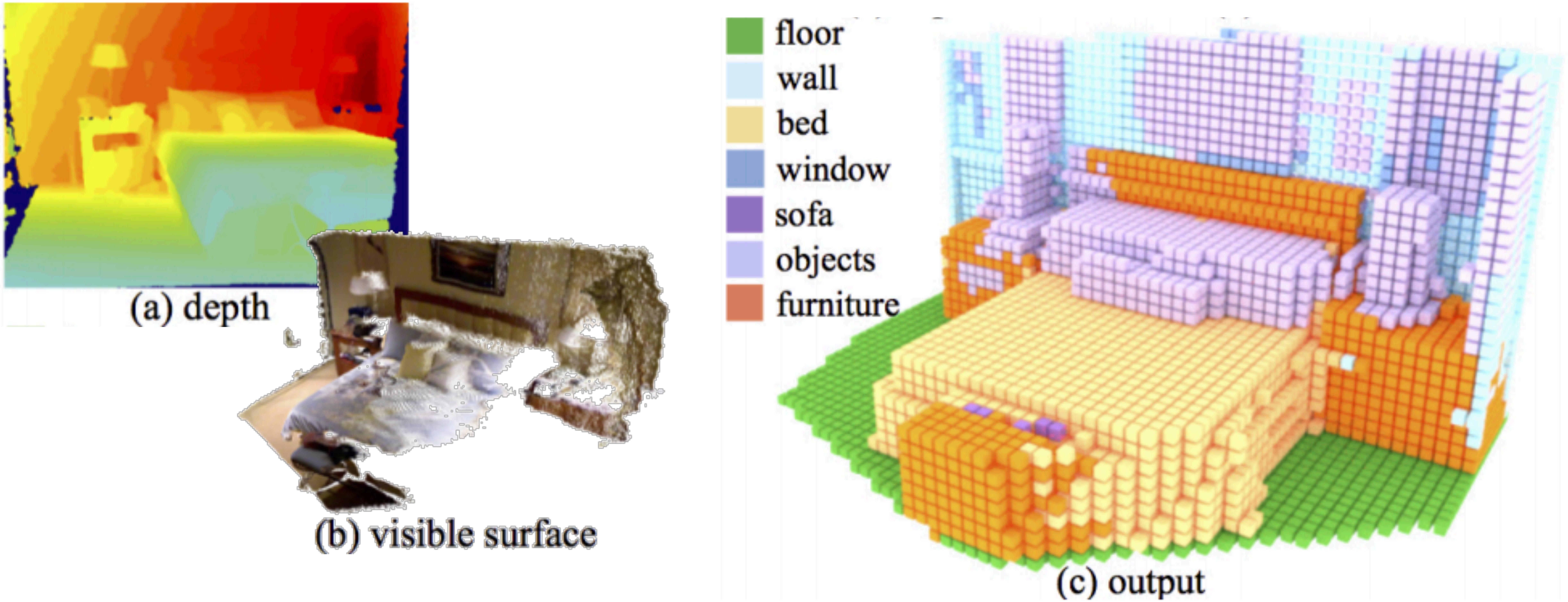
Application #2: Image Understanding

understanding 3D shapes can benefit image understanding



[Zhang et al. 2017]

Application #3: Semantic Scene Understanding



[Song et al. 2017]

Application #4: 3D Asset Creation

Application #4: 3D Asset Creation

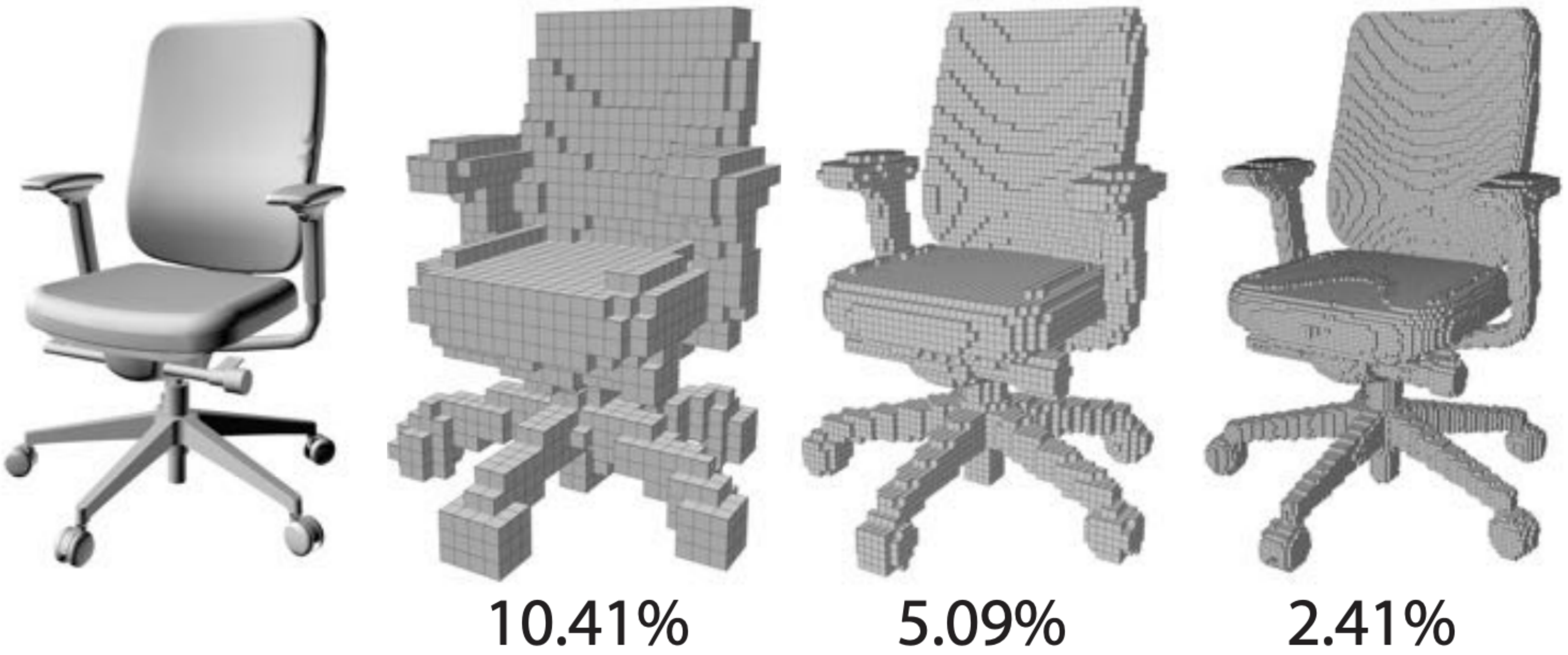
What’s Different in 3D?

What's Different in 3D?

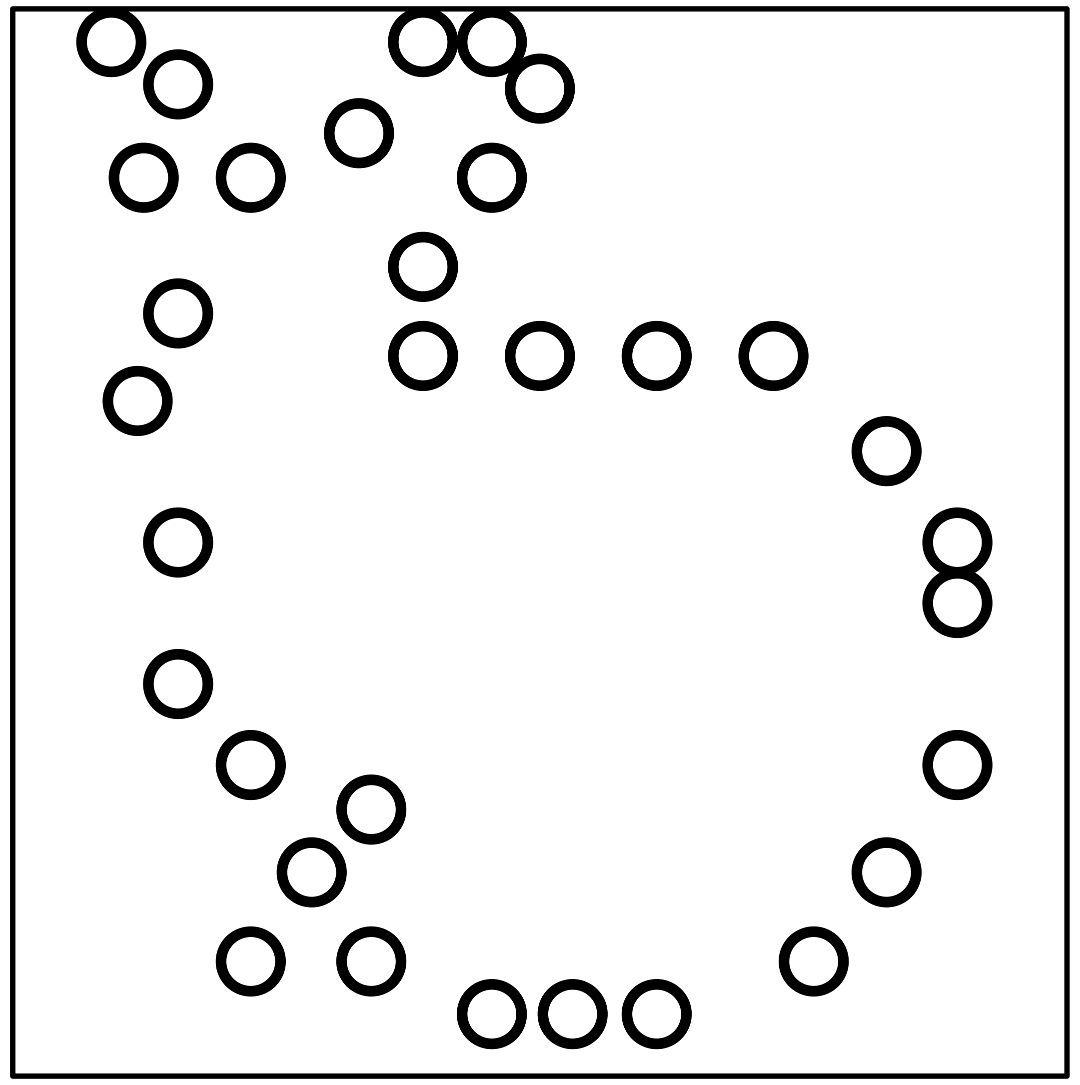
- Number of Voxels grows as $O(n^3)$ versus occupied **surface** $O(n^2)$

What's Different in 3D?

- Number of Voxels grows as $O(n^3)$ versus occupied **surface** $O(n^2)$

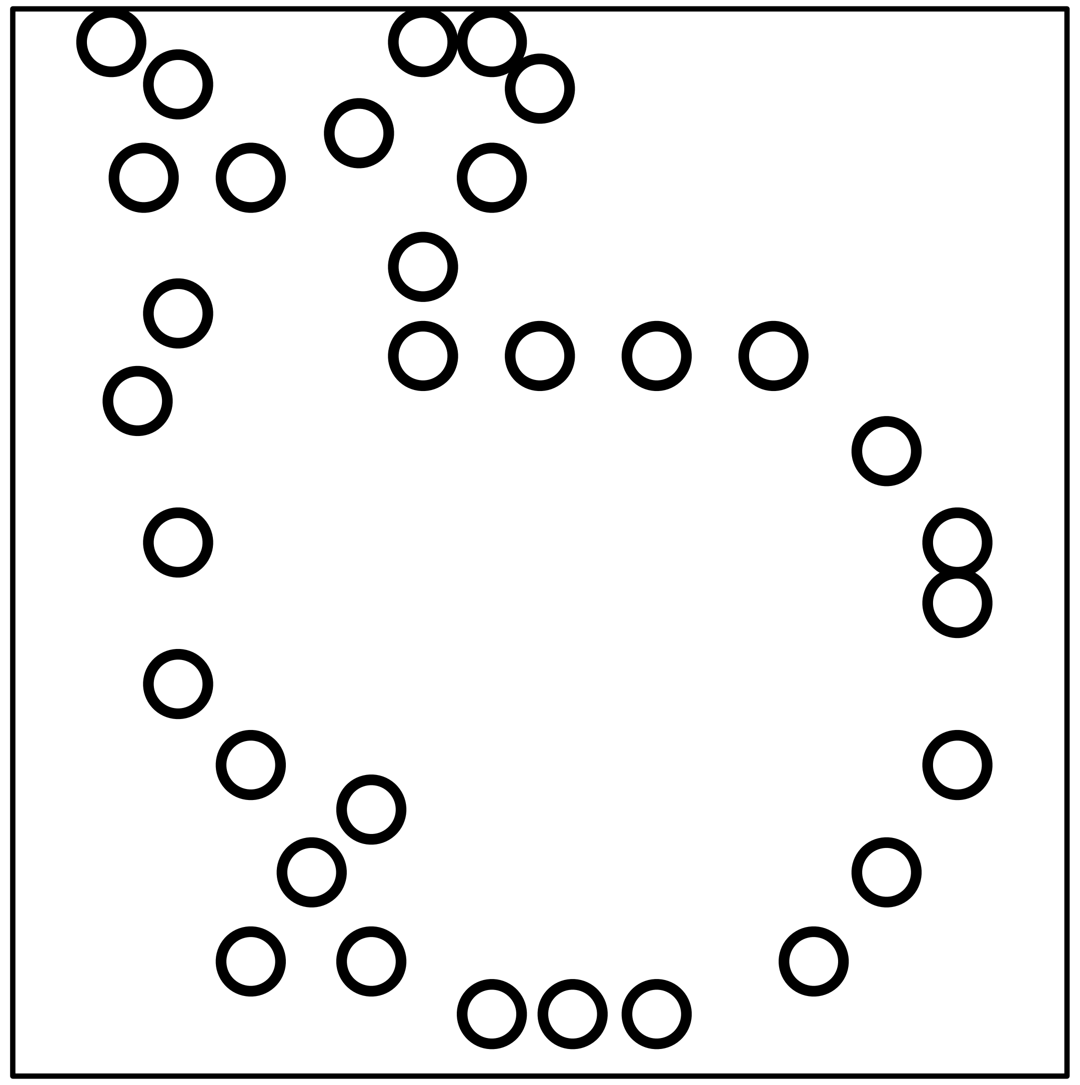


Data Representation .. Many Possibilities!

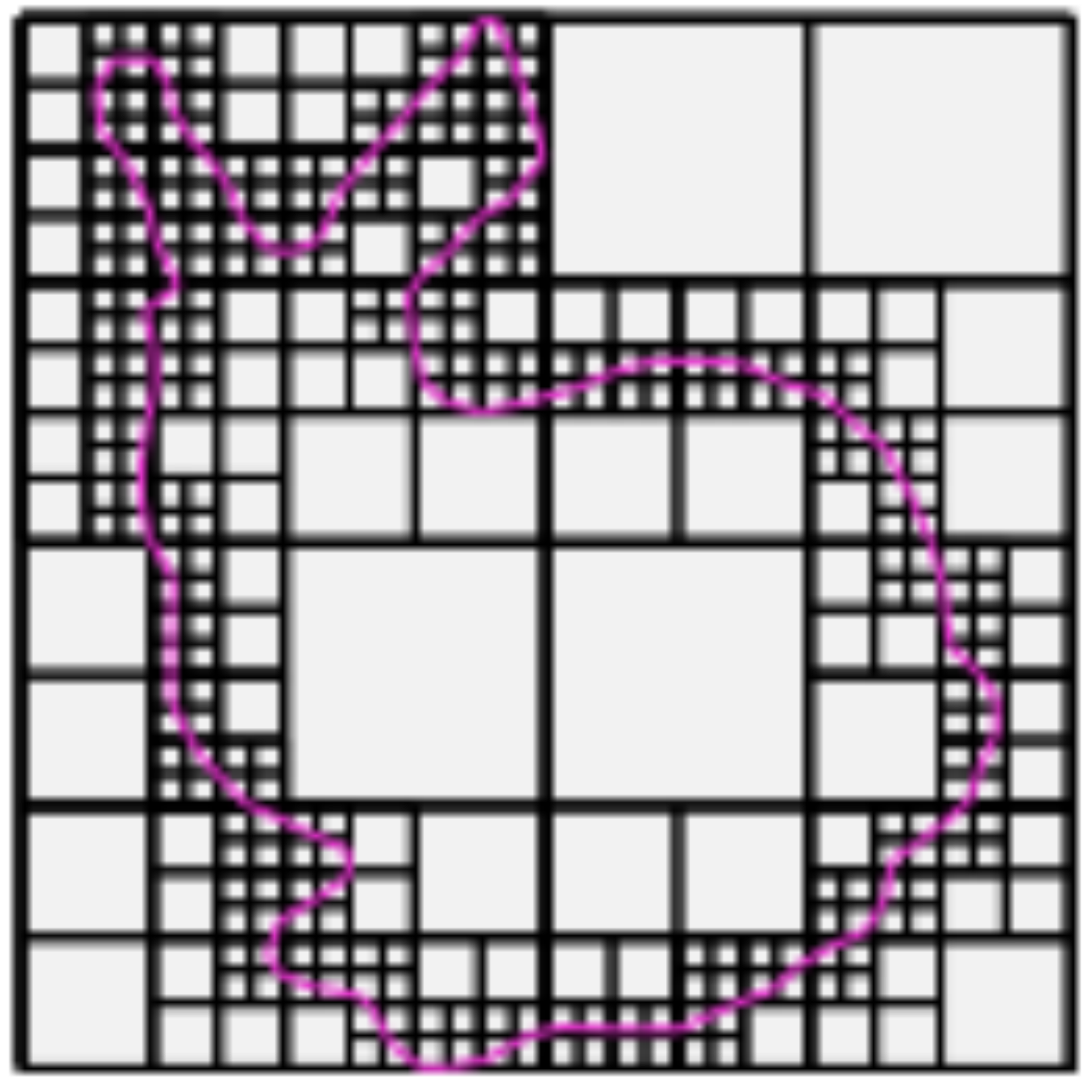


points

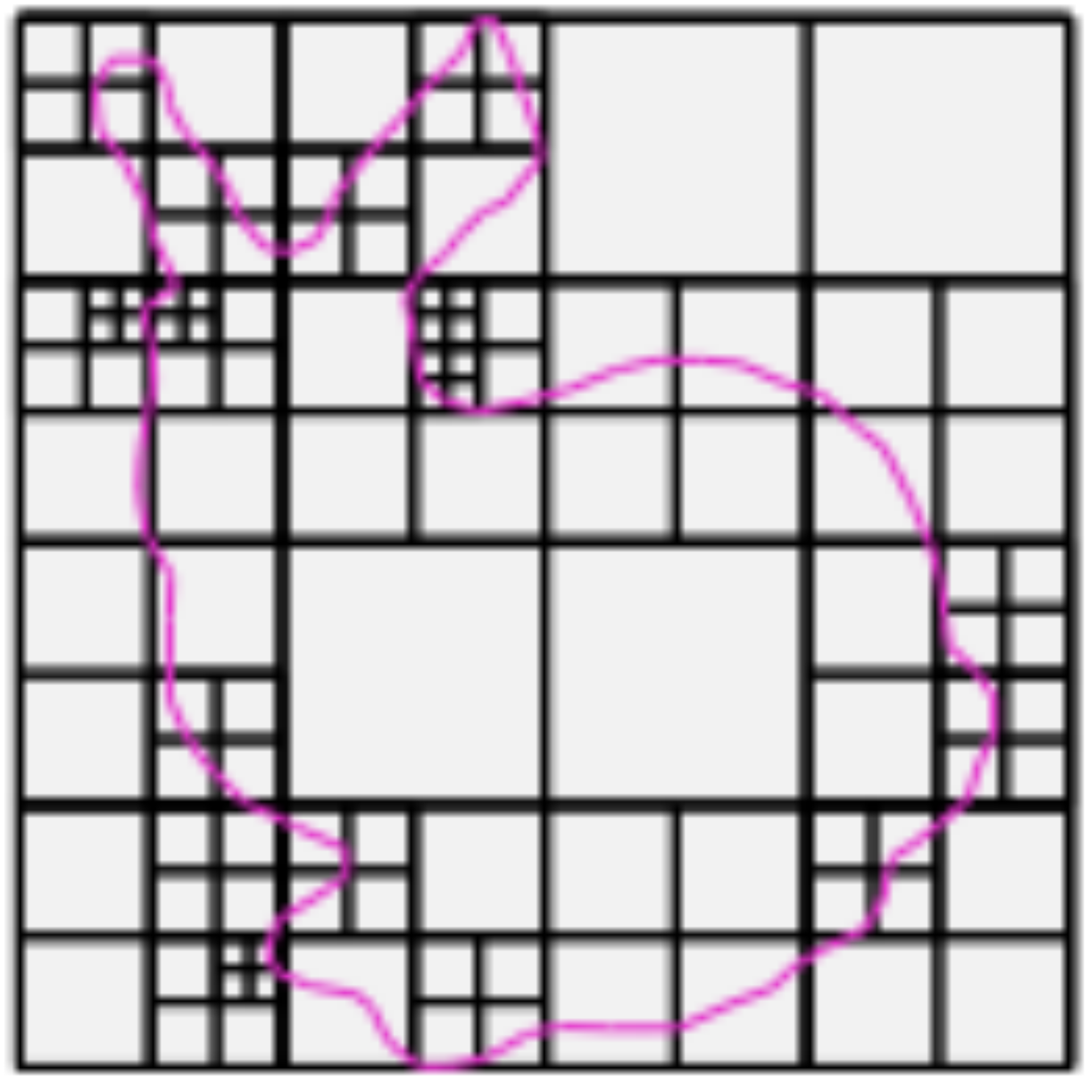
Data Representation .. Many Possibilities!



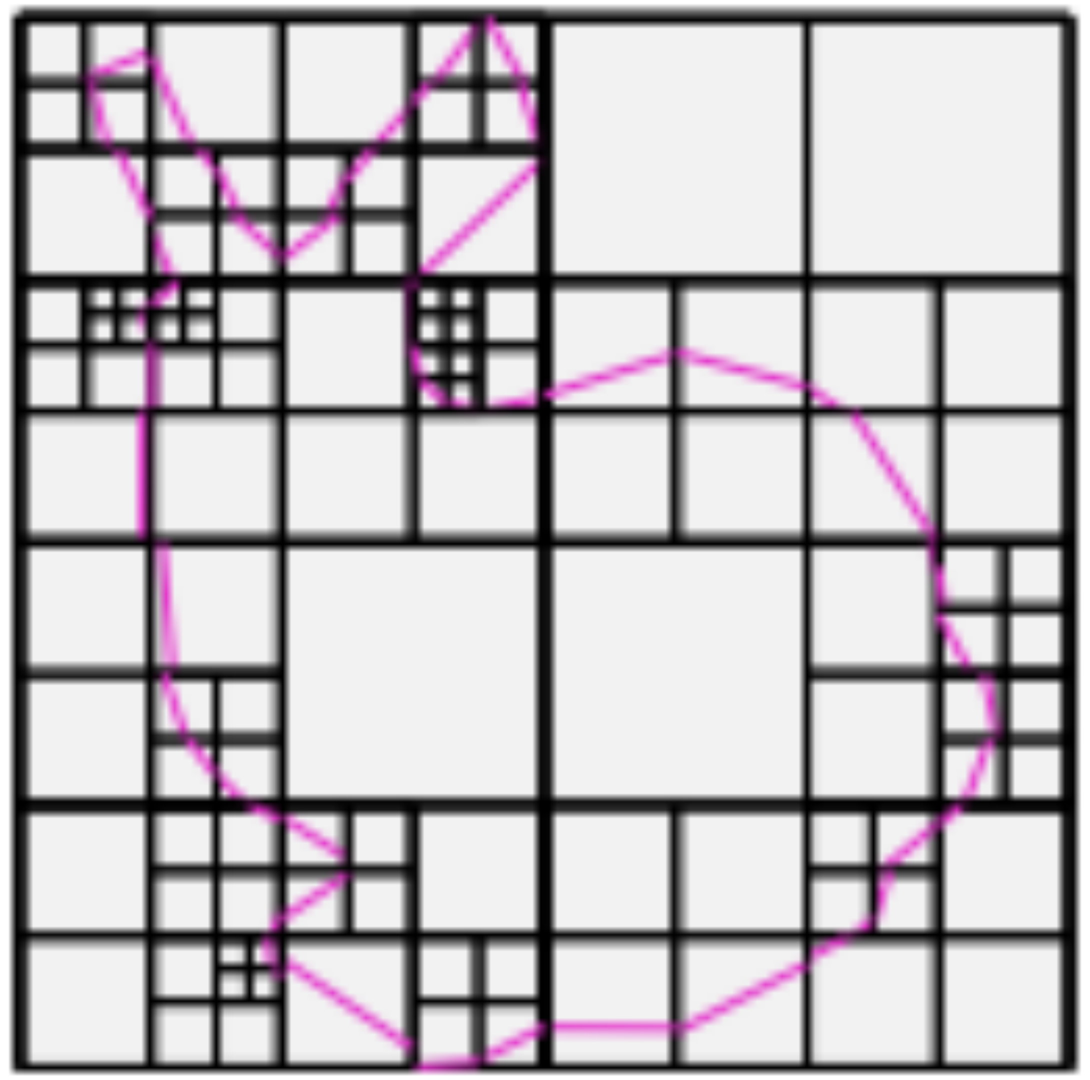
points



voxels



cells



patches

Challenges

Challenges

1. Representation

Challenges

1. Representation

2. **Neighborhood** information
   - who are the neighbouring elements
   - how are the elements ordered

Challenges

1. Representation

2. **Neighborhood** information
   - who are the neighbouring elements
   - how are the elements ordered

3. **Extrinsic** versus **intrinsic** representation

Challenges

1. Representation

2. **Neighborhood** information
   - who are the neighbouring elements
   - how are the elements ordered

3. **Extrinsic** versus **intrinsic** representation

4. Simplicity versus memory/runtime tradeoff

Representation for 3D

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

Representation for 3D

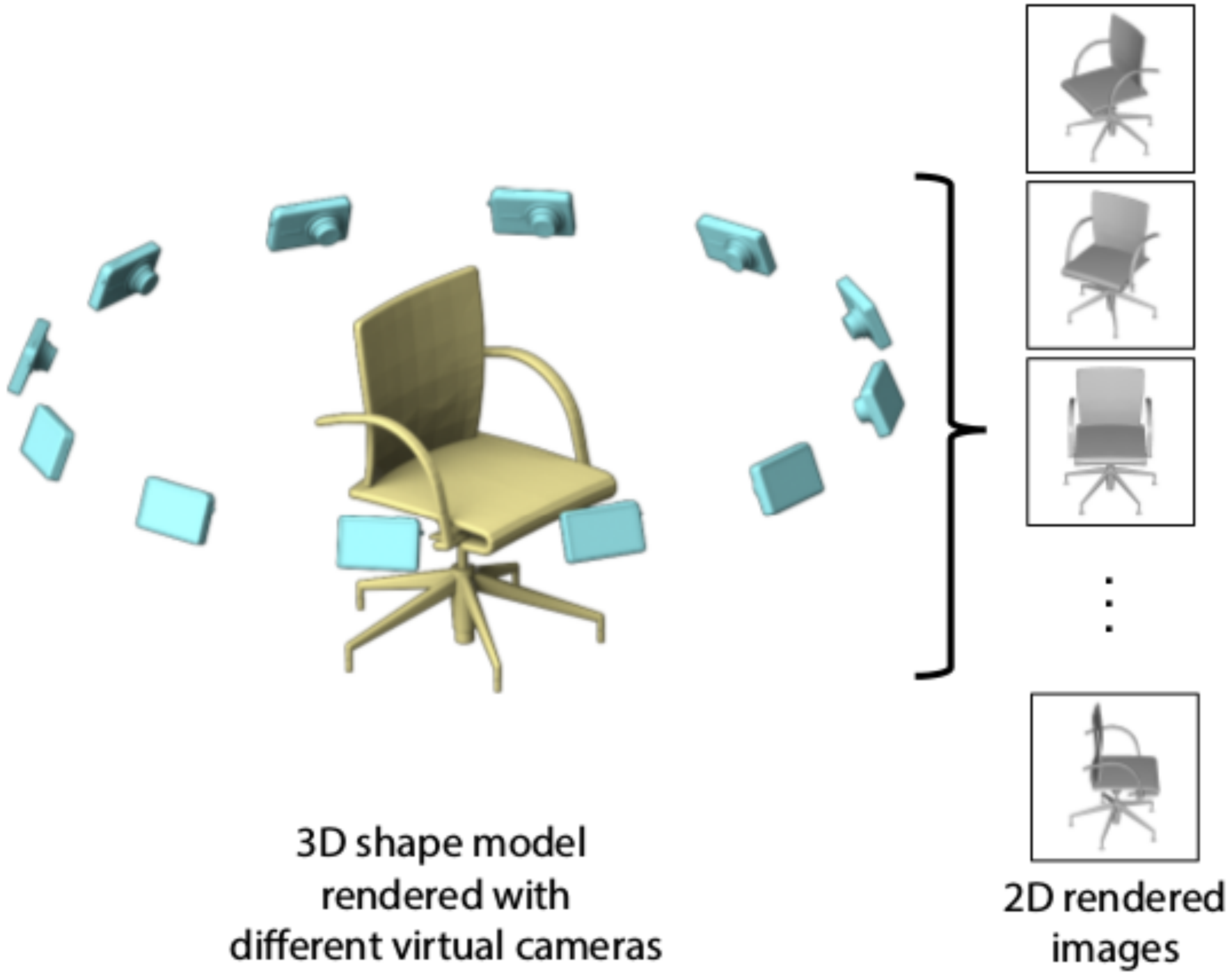
- **Image-based**
- Volumetric
- Surface-based
- Point-based

Representation for 3D: Multi-view CNN

3D shape model
rendered with
different virtual cameras

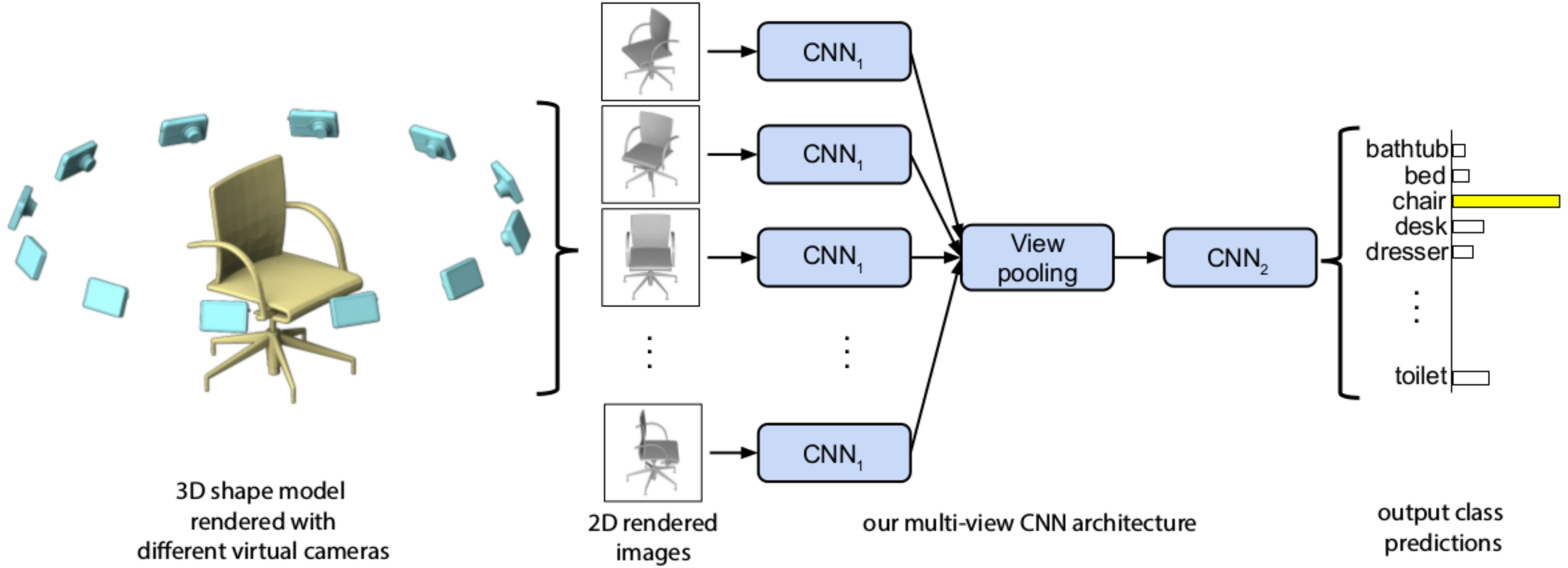
[Kalogerakis et al. 2015]

Representation for 3D: Multi-view CNN



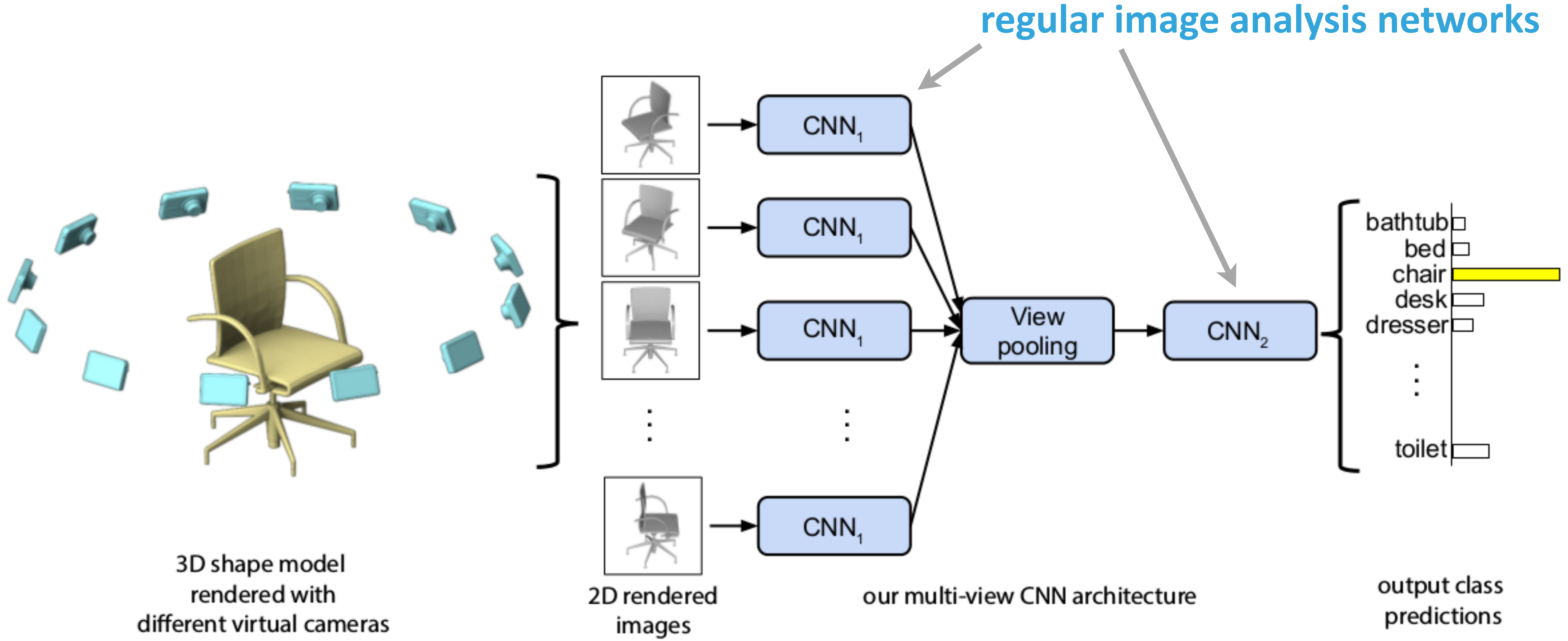
[Kalogerakis et al. 2015]

Representation for 3D: Multi-view CNN



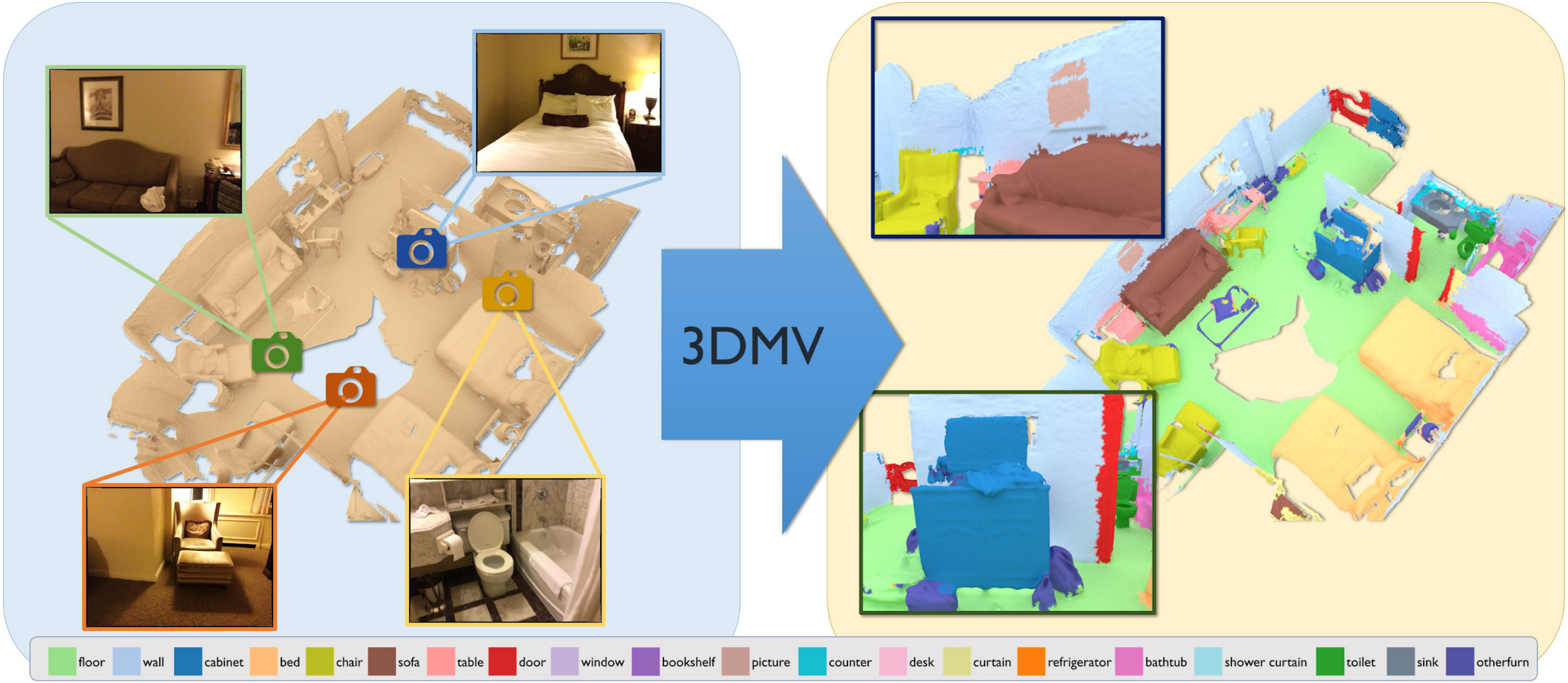
[Kalogerakis et al. 2015]

Representation for 3D: Multi-view CNN

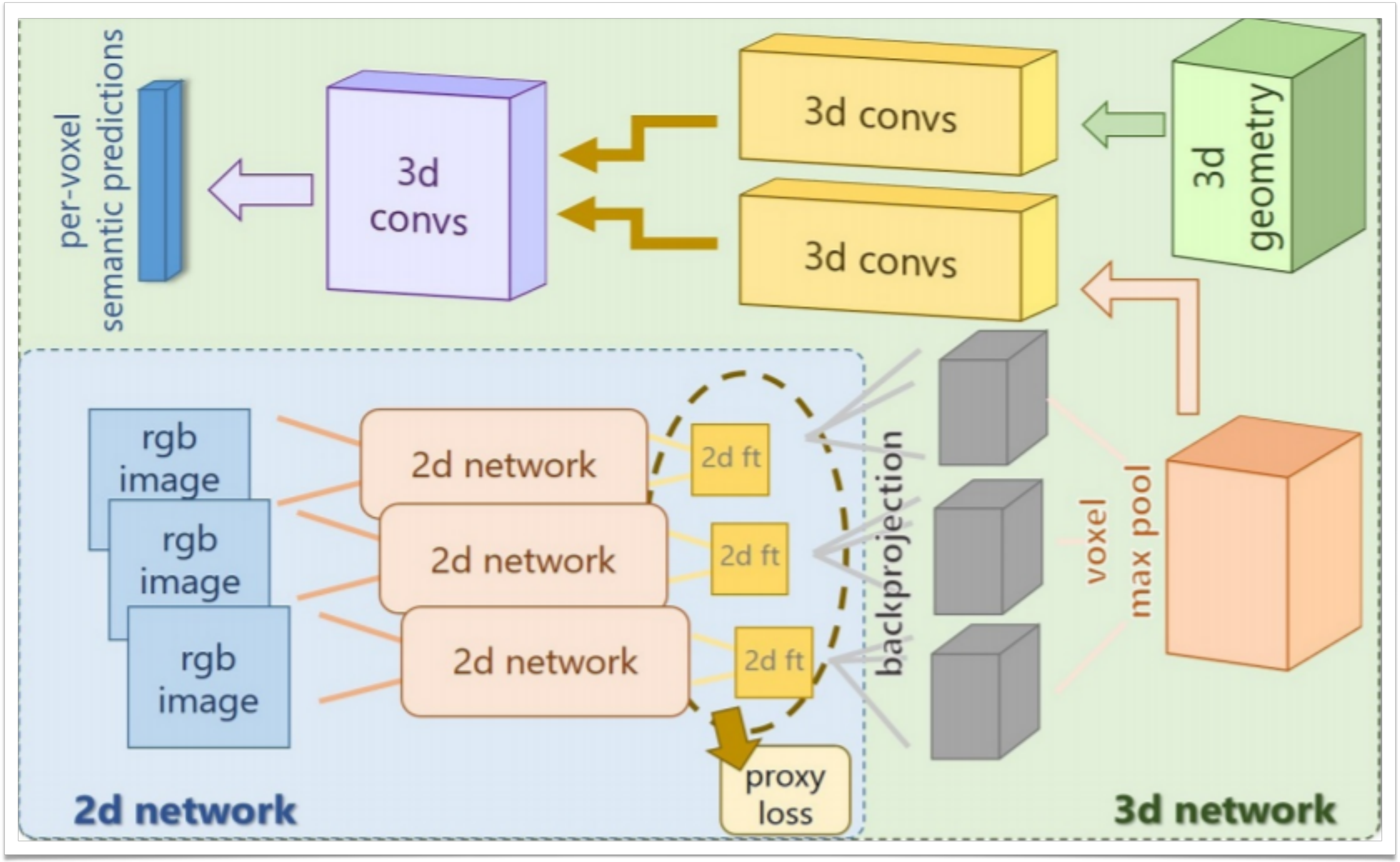


[Kalogerakis et al. 2015]

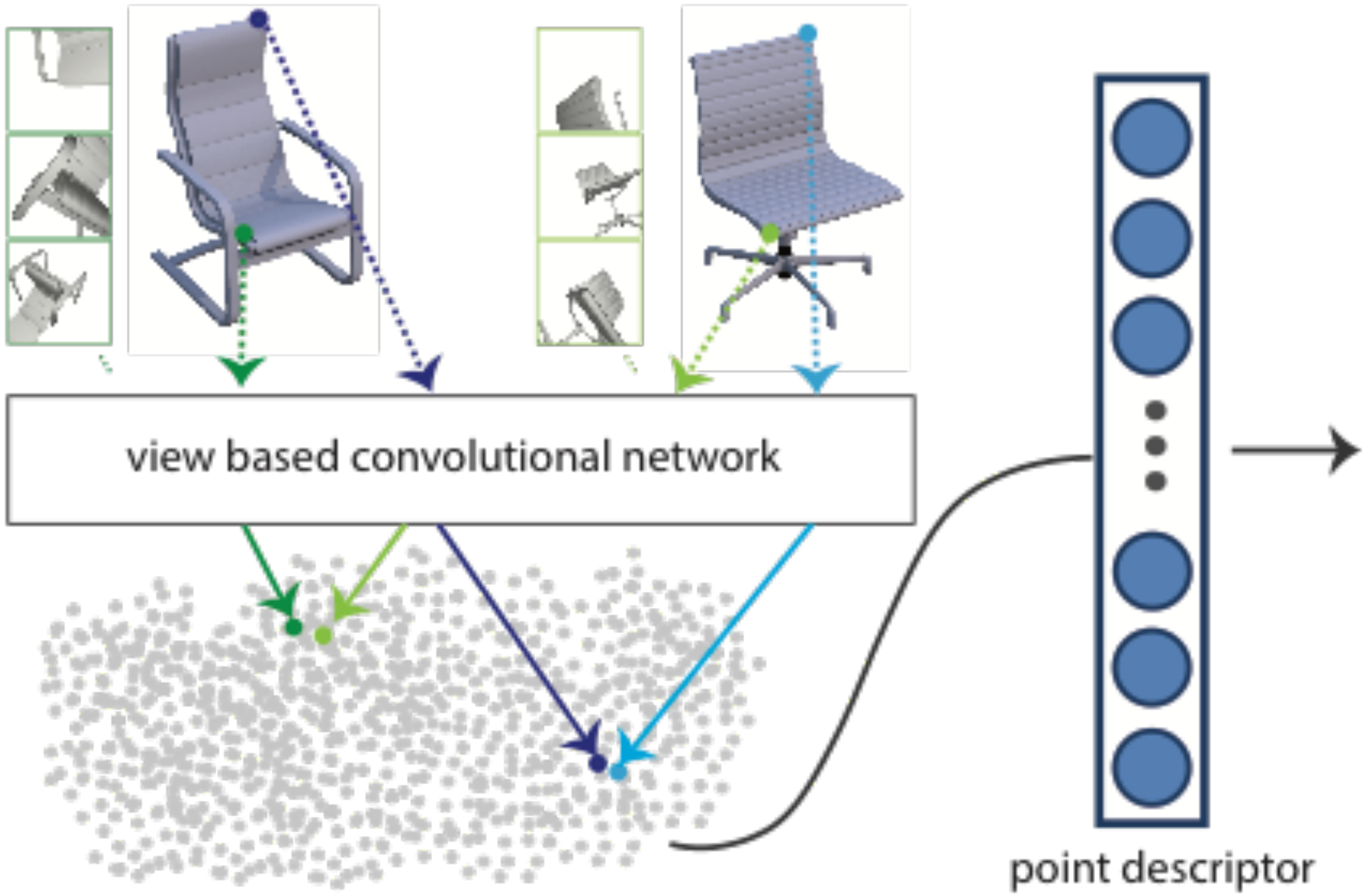
3DMV: Joint 3D-Multi-View Prediction for 3D Semantic Scene Segmentation



Integrating View Information

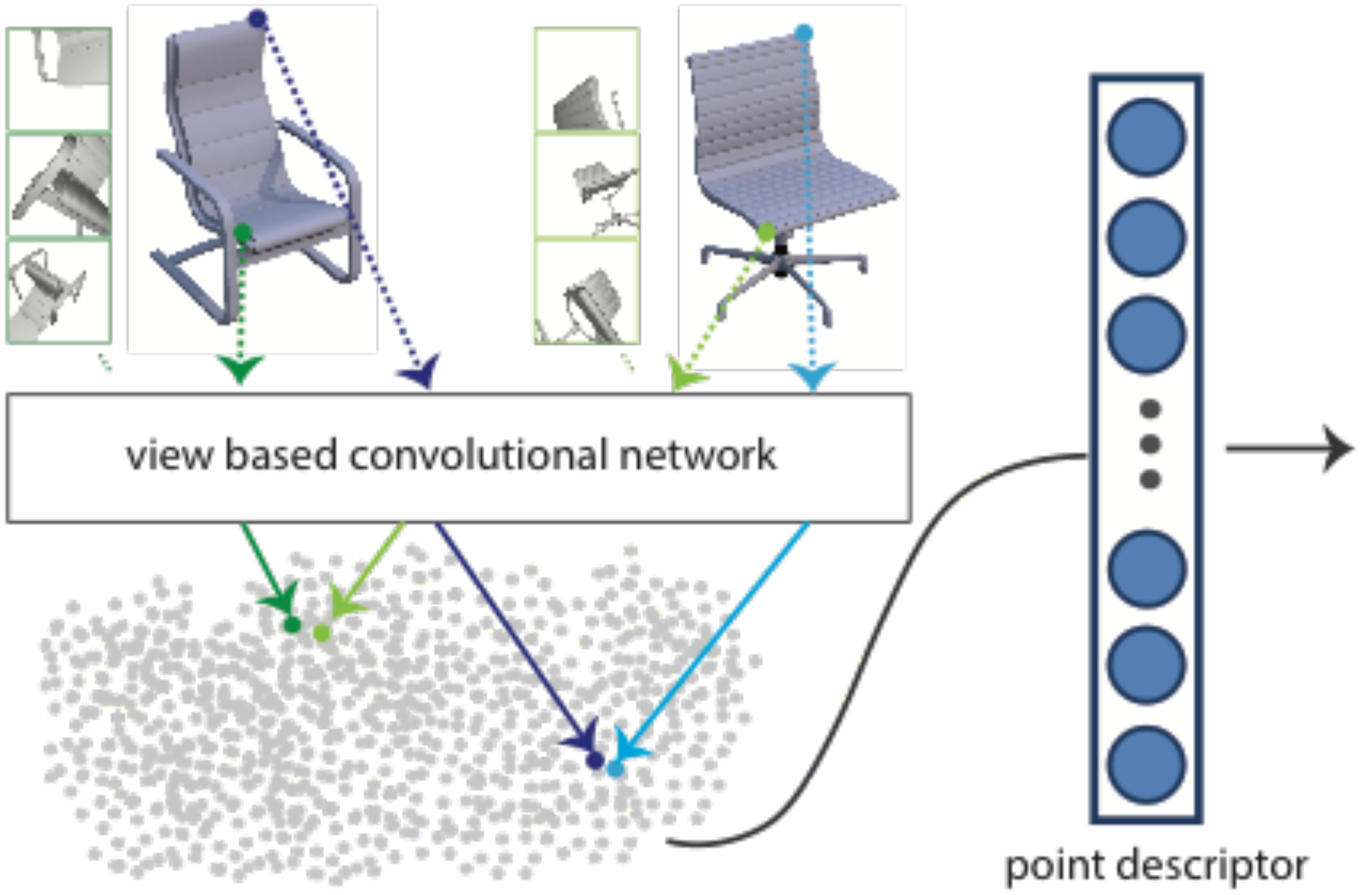


Representation for 3D: Local Multi-view CNN



[Huang et al. 2018]

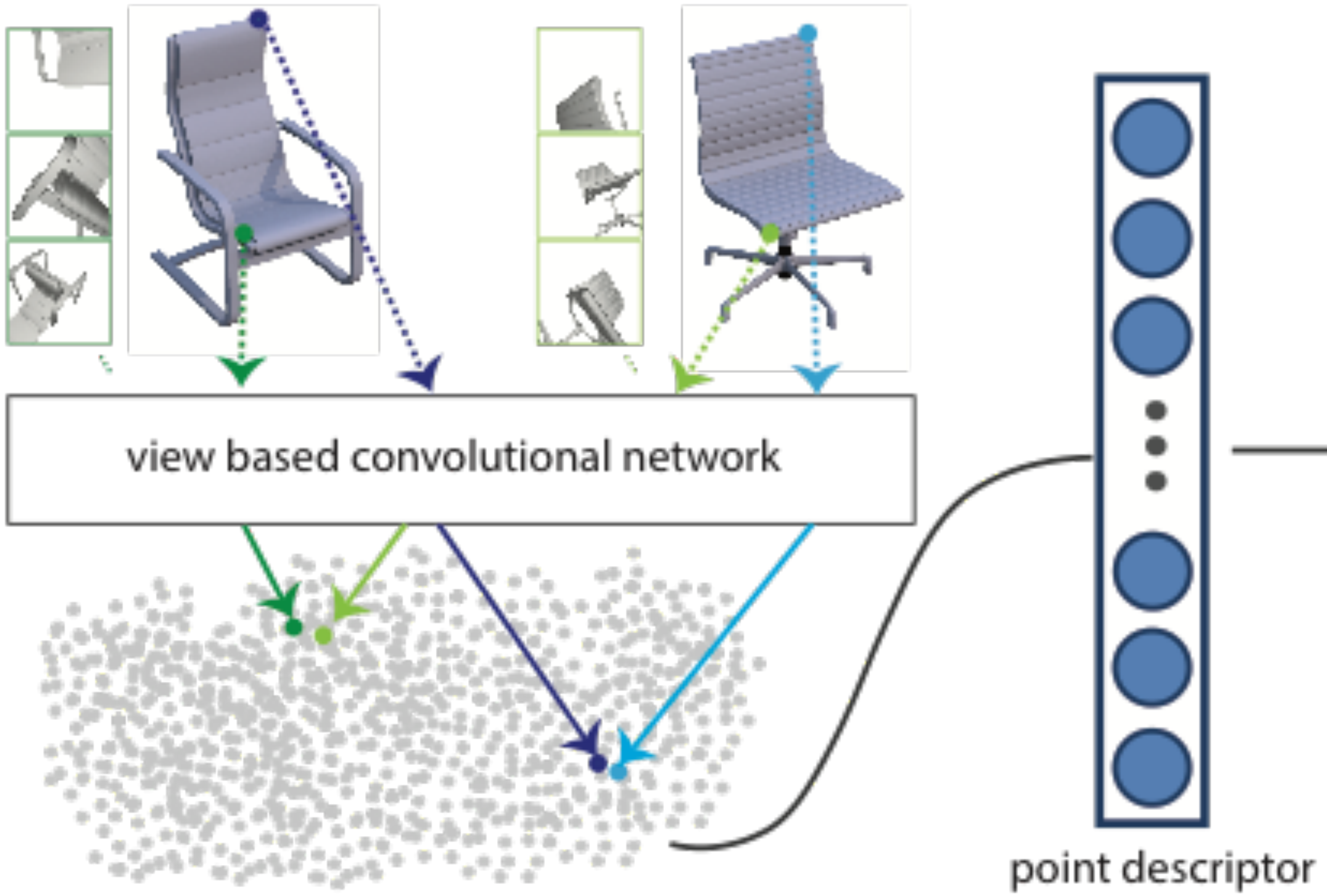
Representation for 3D: Local Multi-view CNN



Segmentation
Correspondence
Feature matching
Predicting semantic functions

[Huang et al. 2018]

Representation for 3D: Local Multi-view CNN

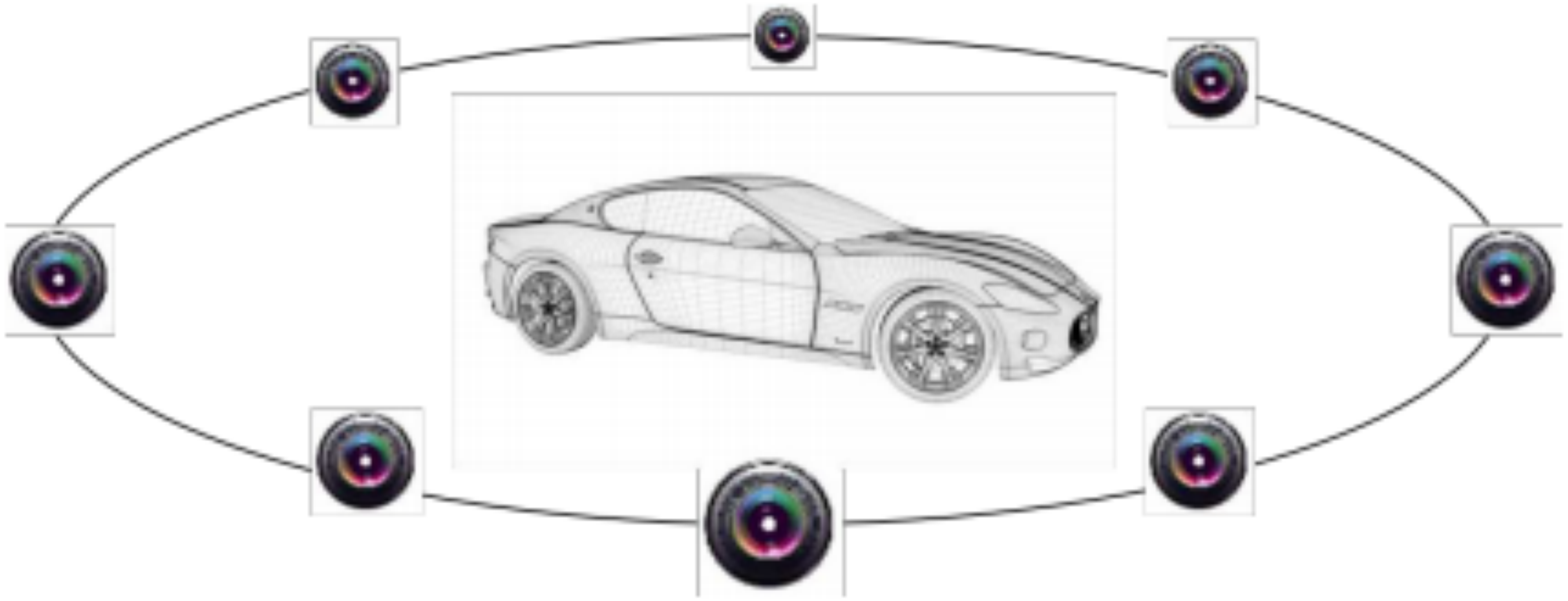


Segmentation
Correspondence
Feature matching
Predicting semantic functions

localized renderings for point-wise features

[Huang et al. 2018]

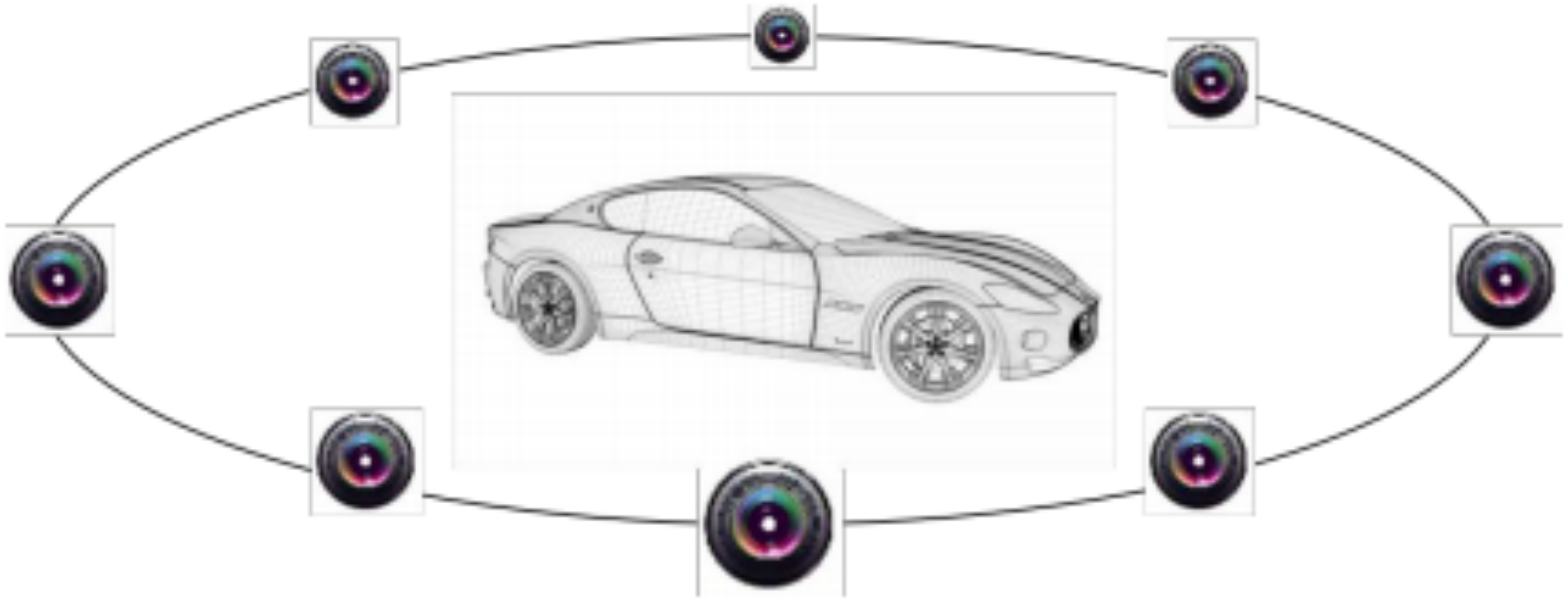
Tangent Convolutions



loses information due to occlusion

[Tatarchenko et al. 2018]

Tangent Convolutions

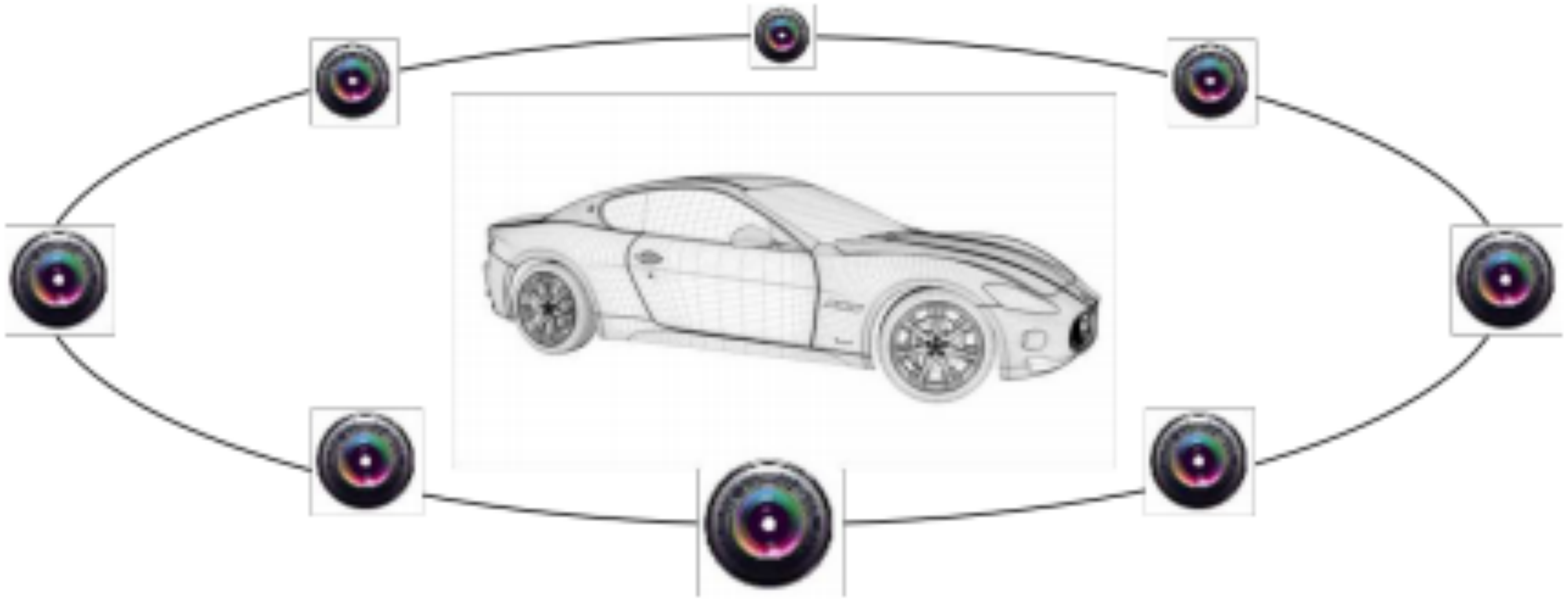


loses information due to occlusion

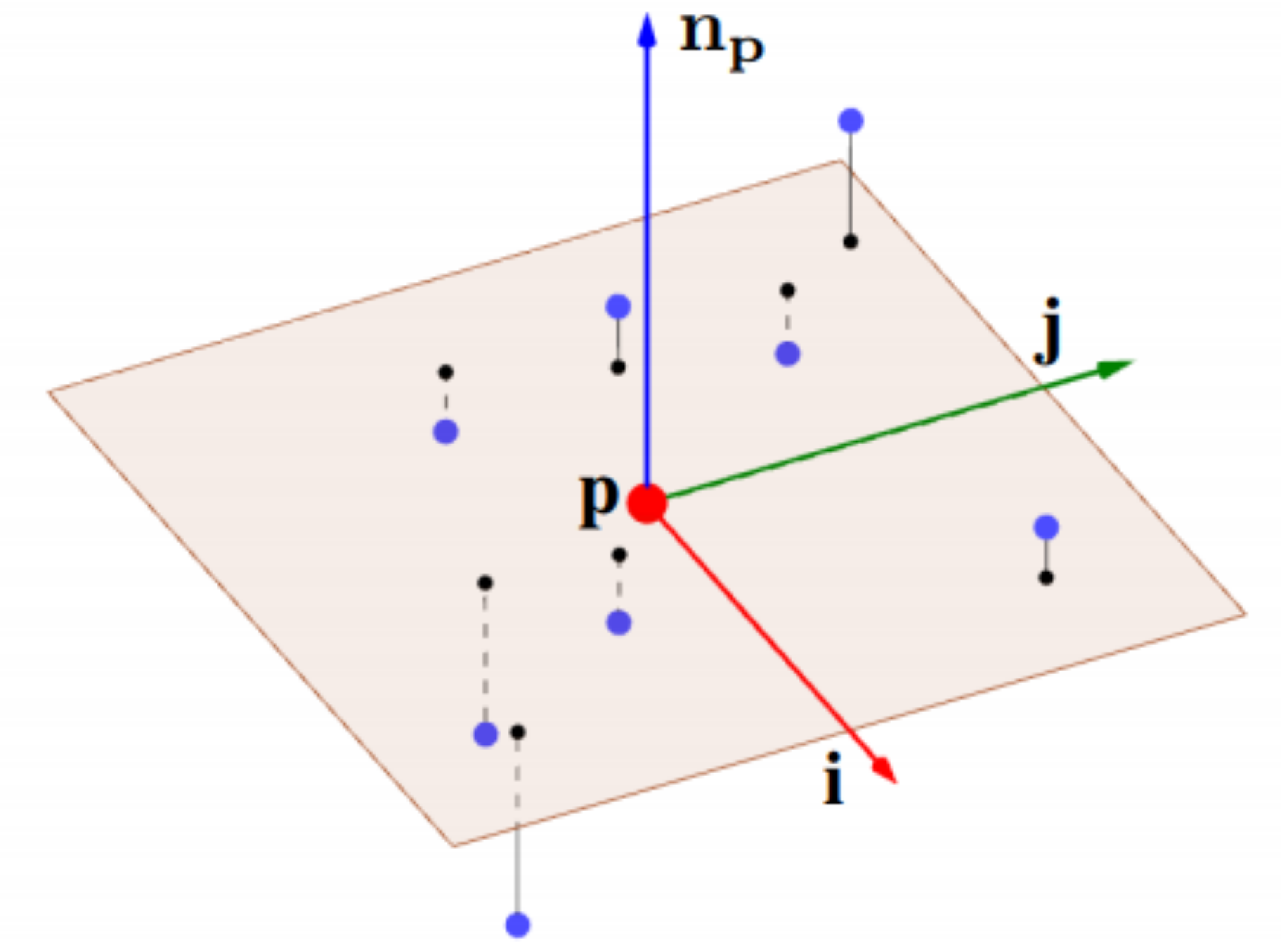
project to local patches
(contrast with PCPNet construction)

[Tatarchenko et al. 2018]

Tangent Convolutions



loses information due to occlusion



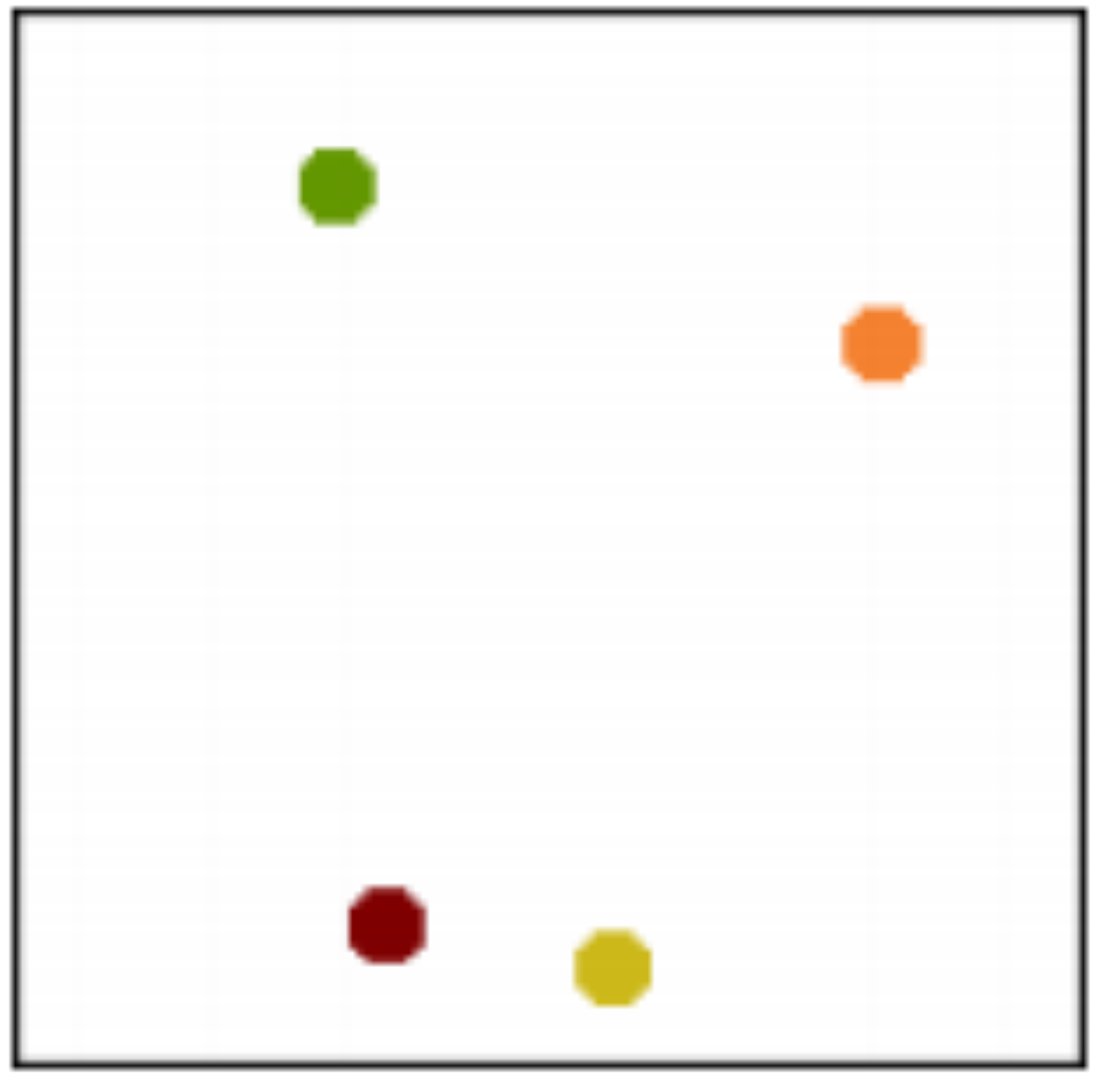
project to local patches
(contrast with PCPNet construction)

[Tatarchenko et al. 2018]

Dealing with Sparse Points

Signal Interpolation

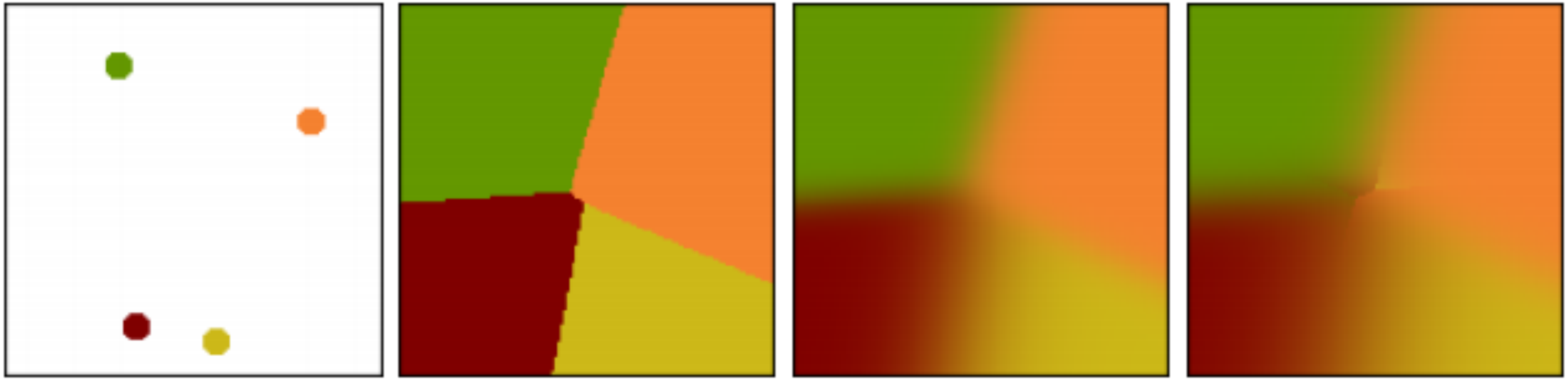
- Use nearest neighbor or Gaussian mixture based methods for interpolation.
- Now the signal is more dense



Dealing with Sparse Points

Signal Interpolation

- Use nearest neighbor or Gaussian mixture based methods for interpolation.

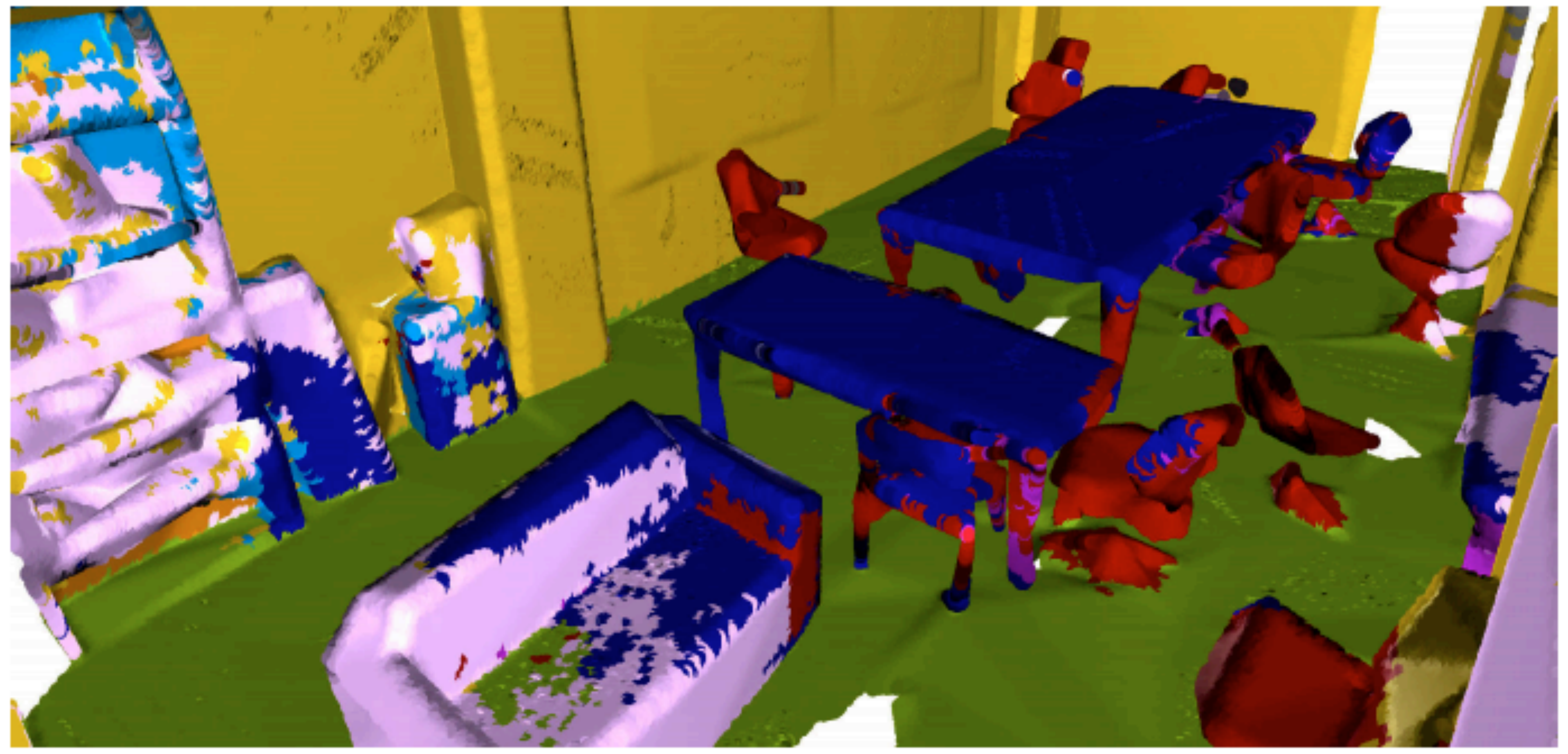


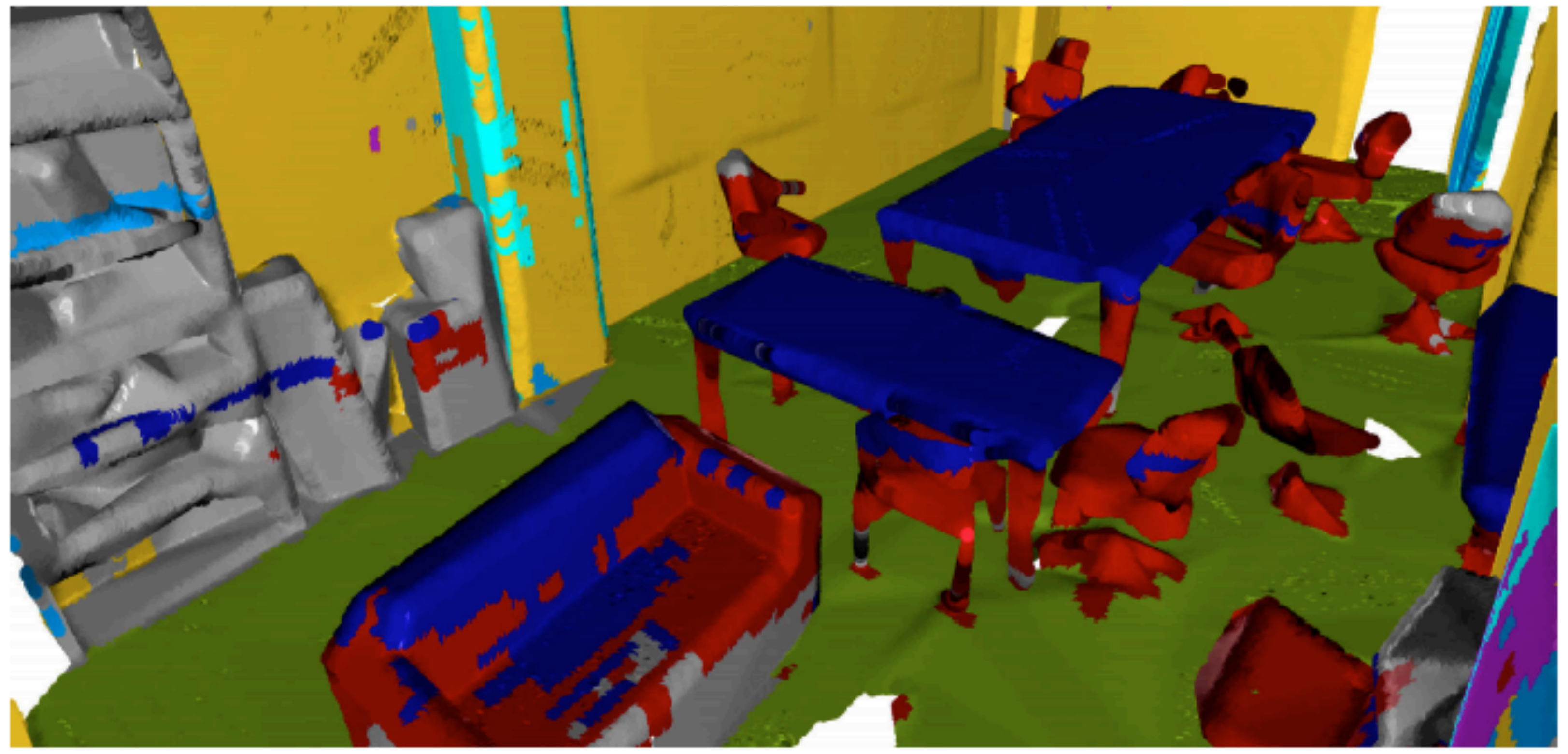
(a) (b) (c) (d)

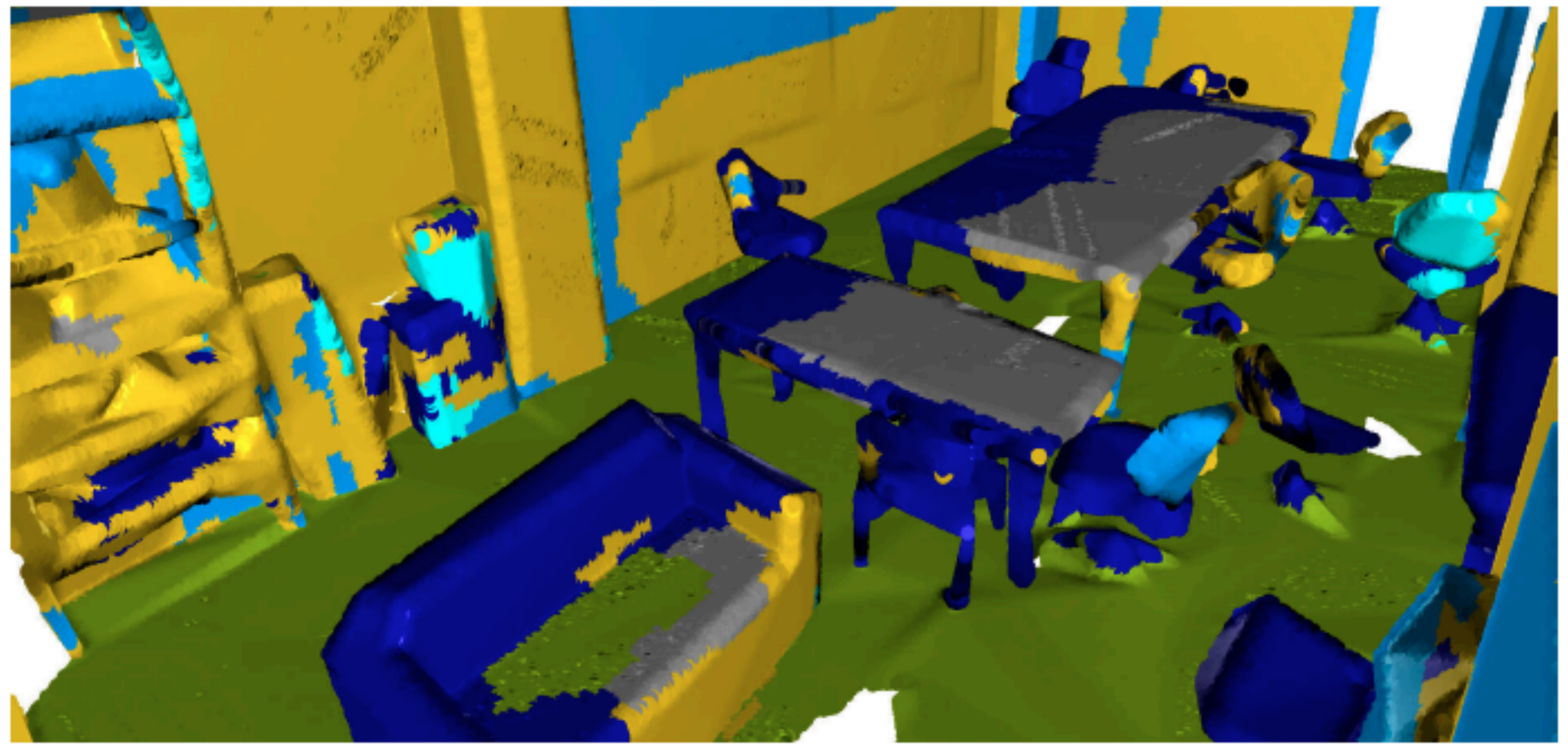
Improved Performance

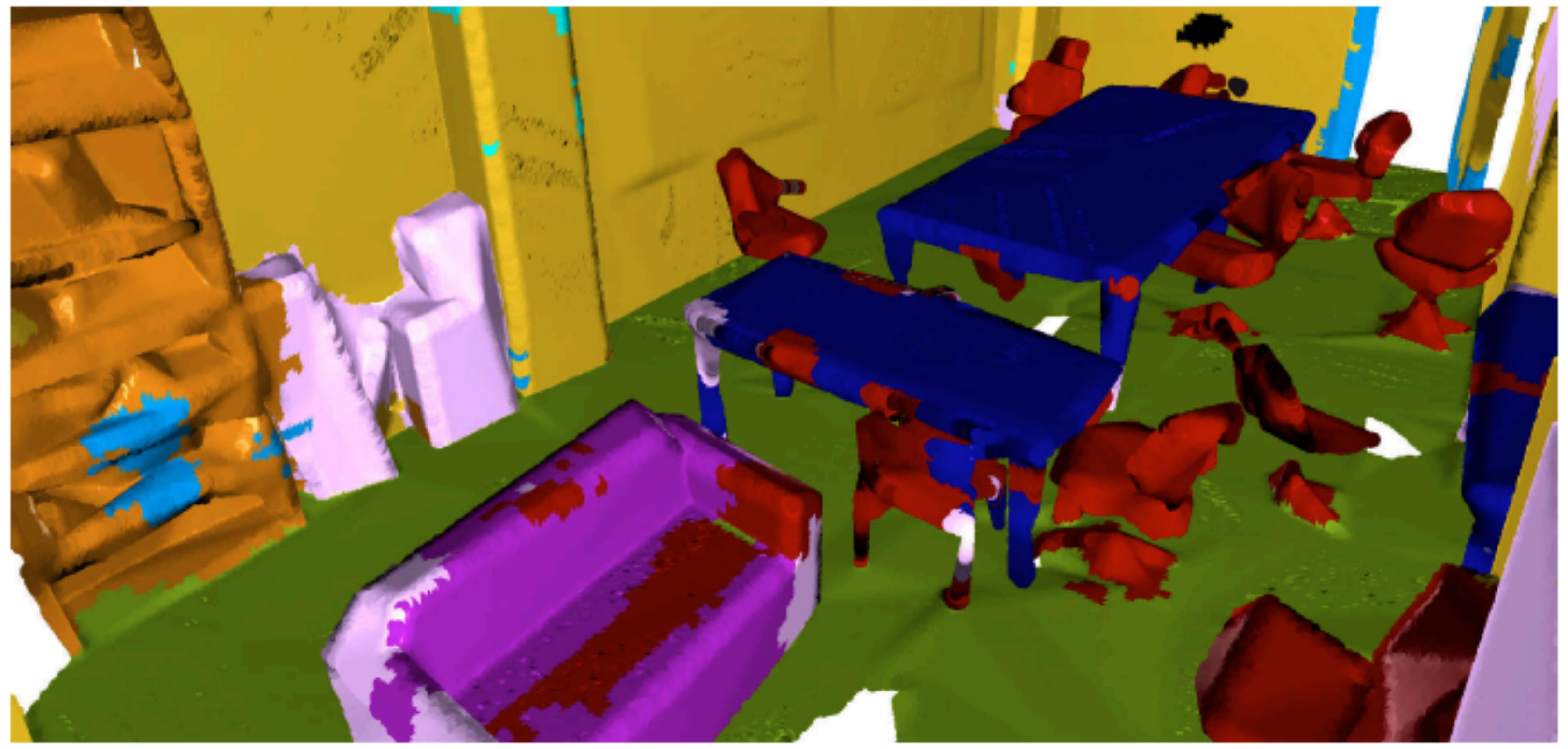
Tangent Convolutions

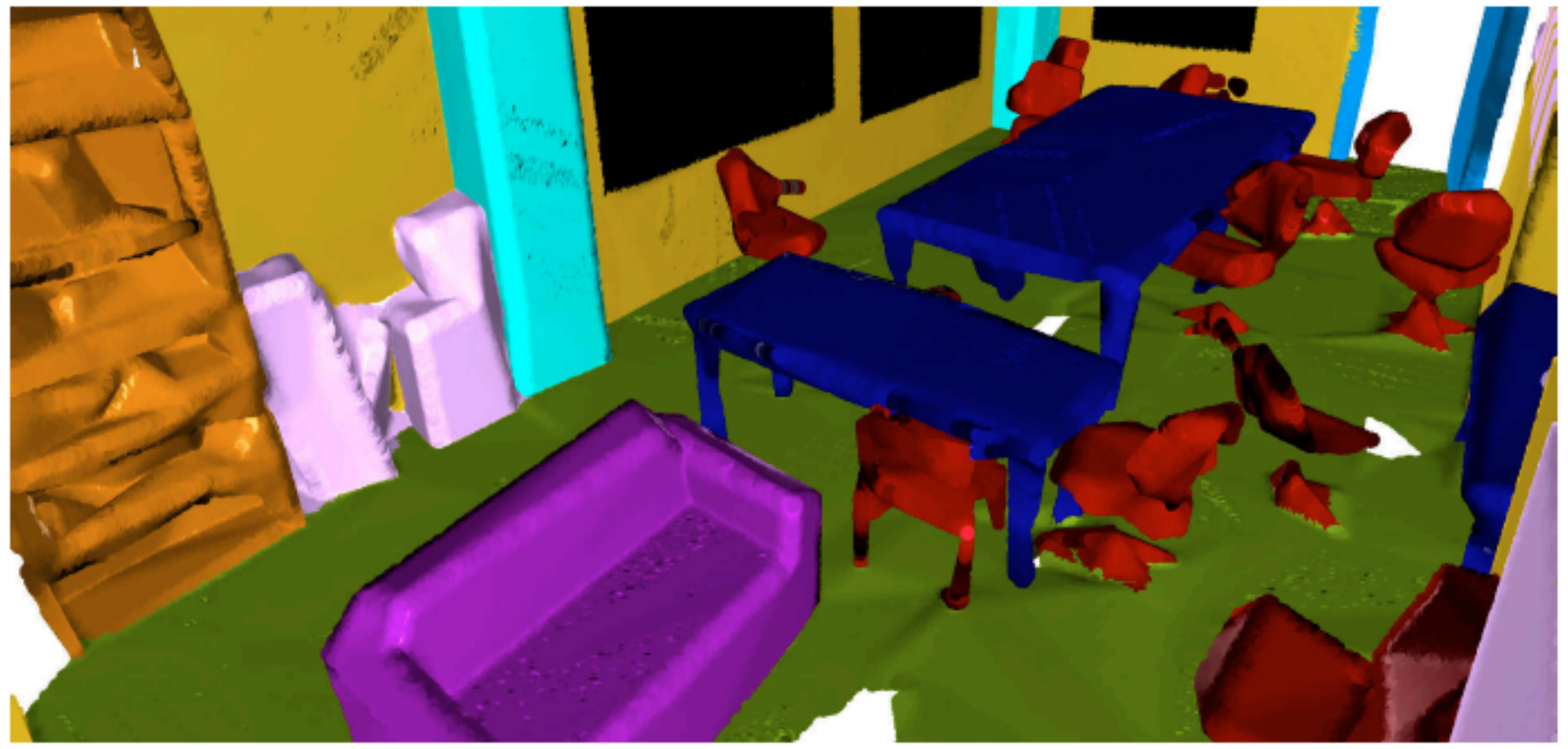

Color


PointNet [39]


ScanNet [10]


OctNet [43]


Ours (DHNRGB)


Ground truth

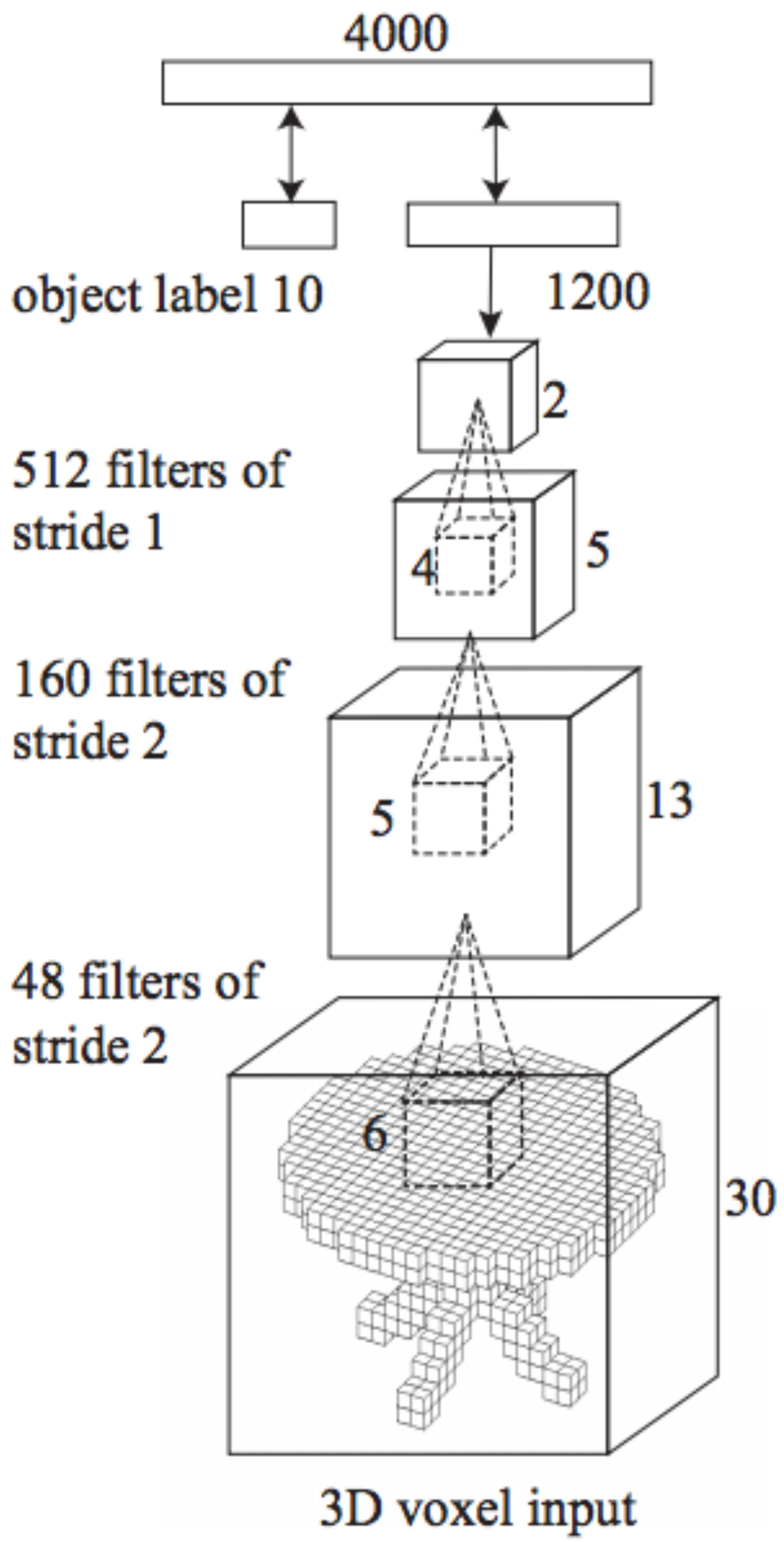
Representation for 3D

- **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**
- Surface-based
- Point-based

3D CNNs : Direct Approach

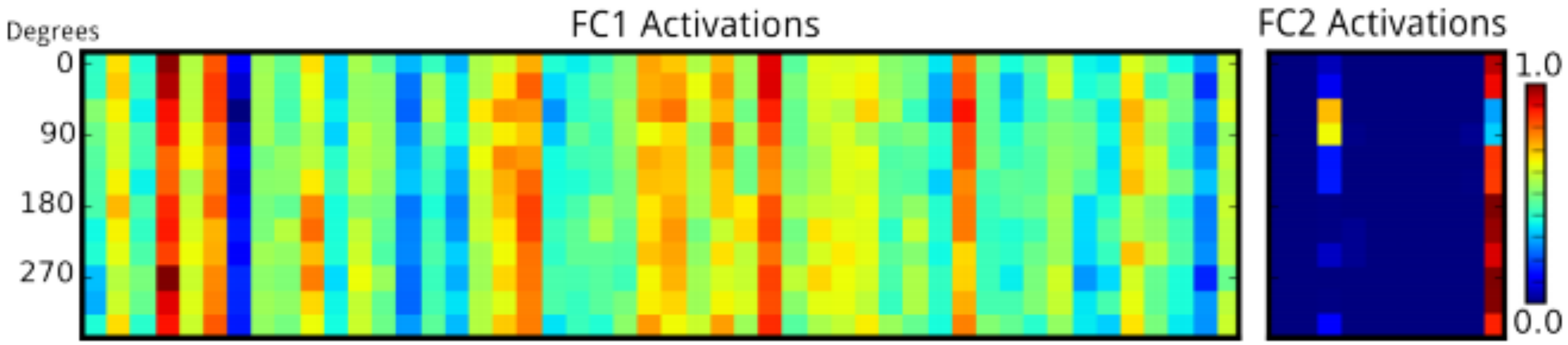


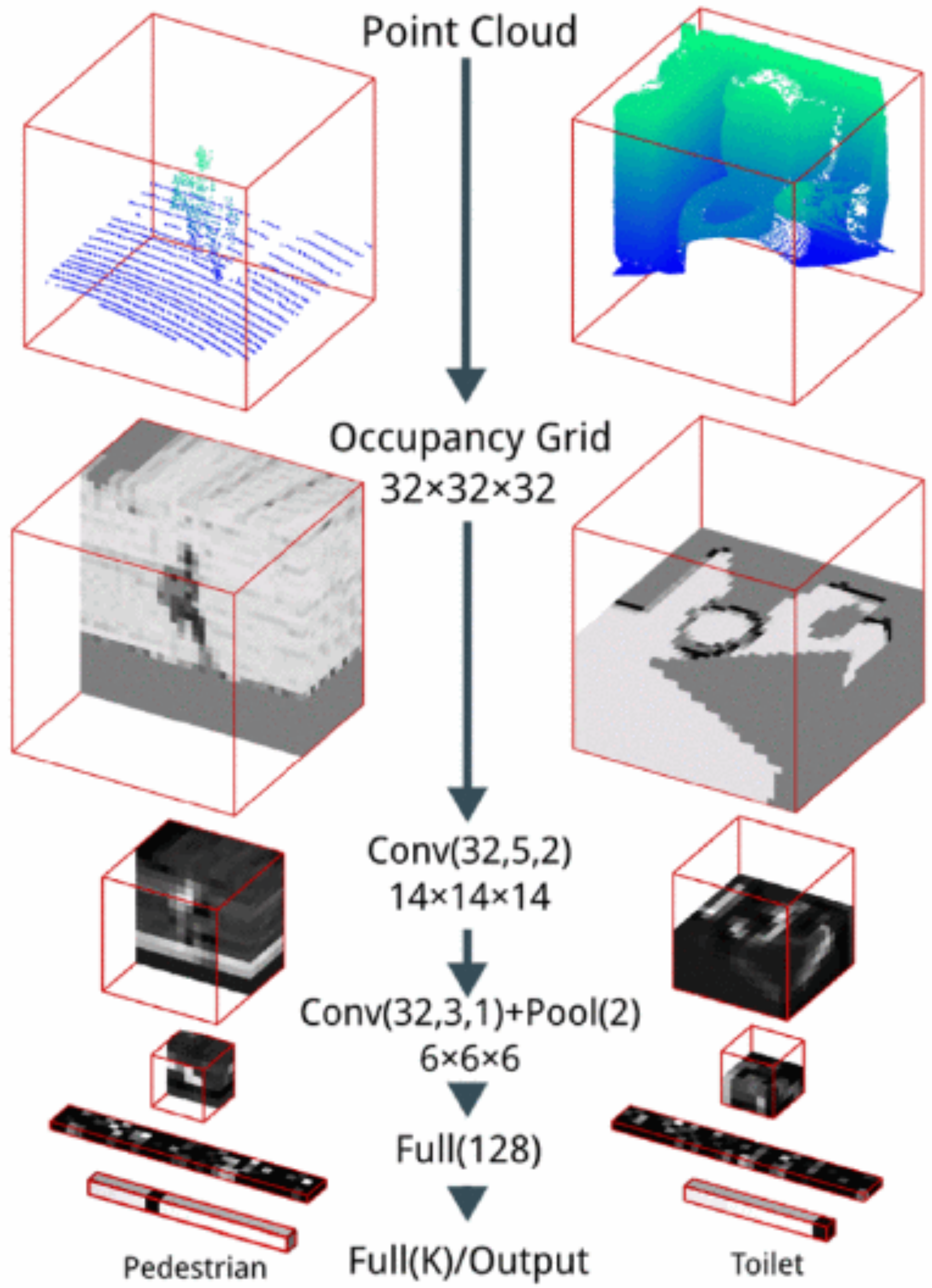
[Xiao et al. 2014]

VoxNet [Maturana et al. 15]

- Binary occupancy, density grid, etc.

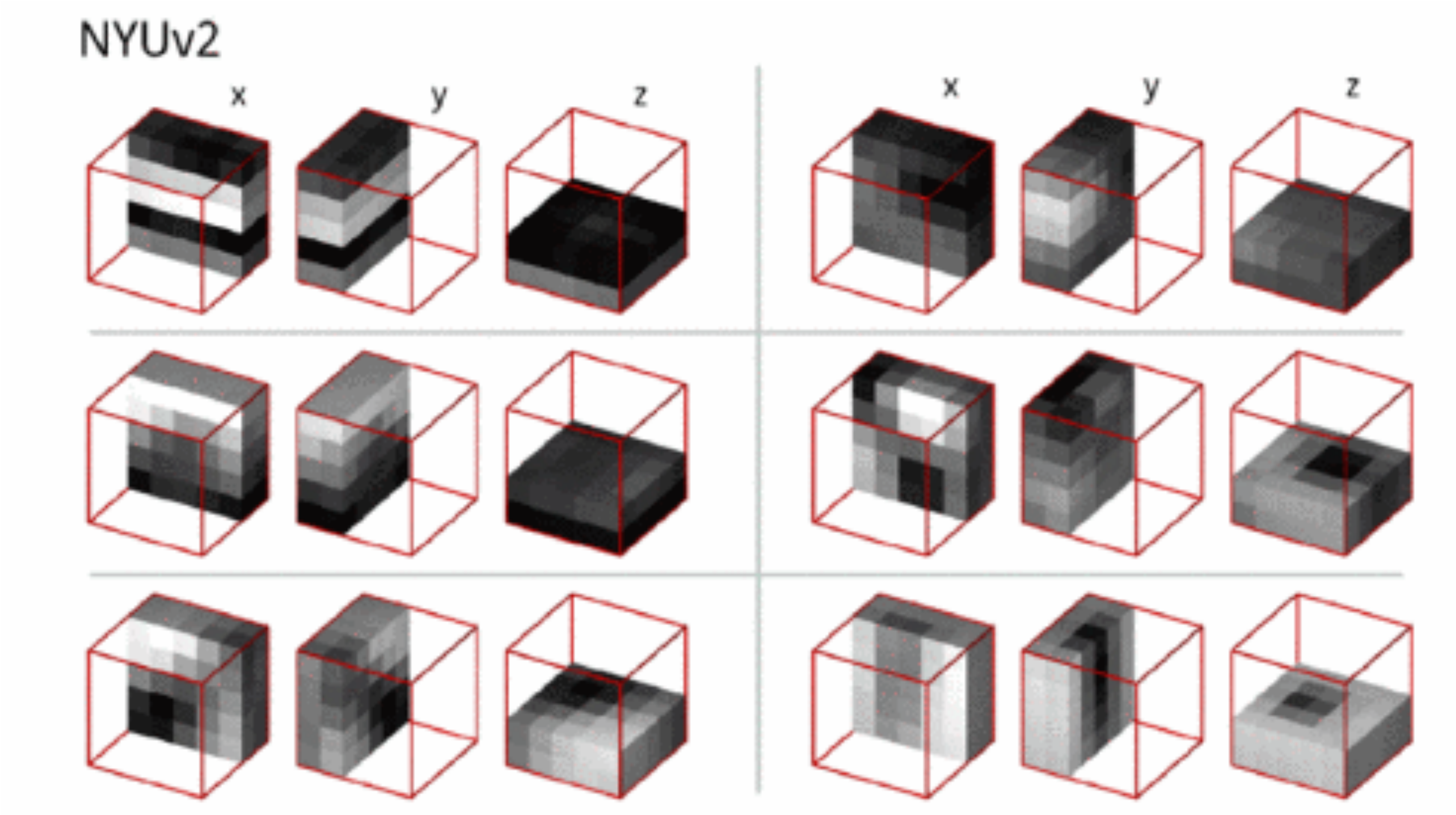
rotational invariance





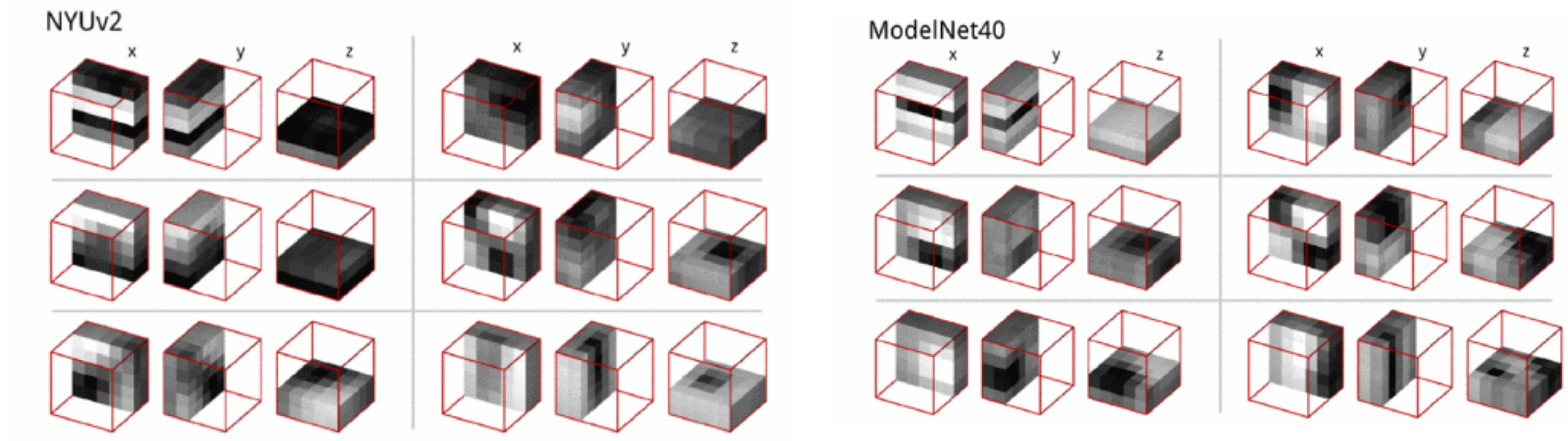
Visualization of First Level Filters

VISUALISATION OF FIRST LAYER FILTERS

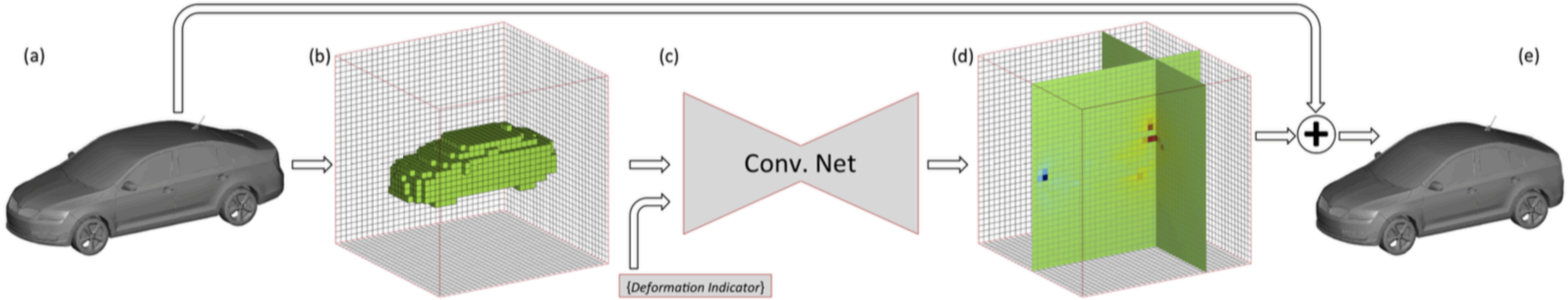


Visualization of First Level Filters

VISUALISATION OF FIRST LAYER FILTERS

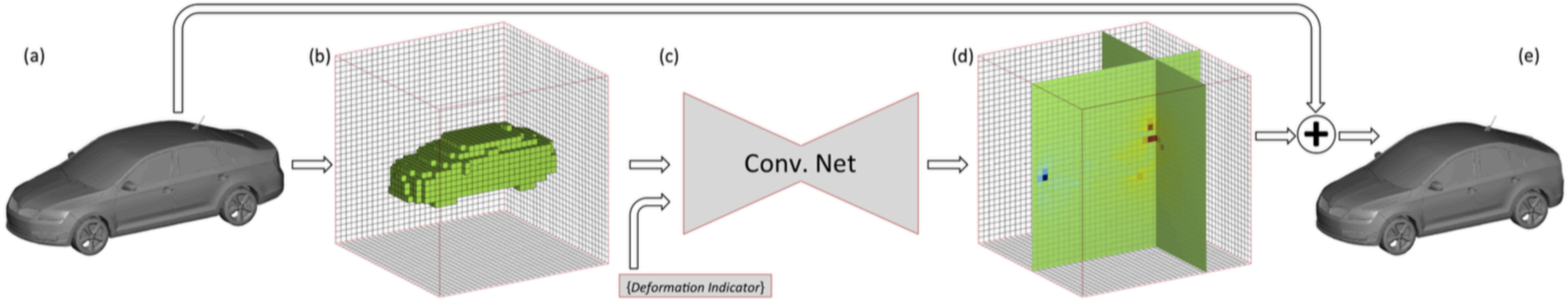


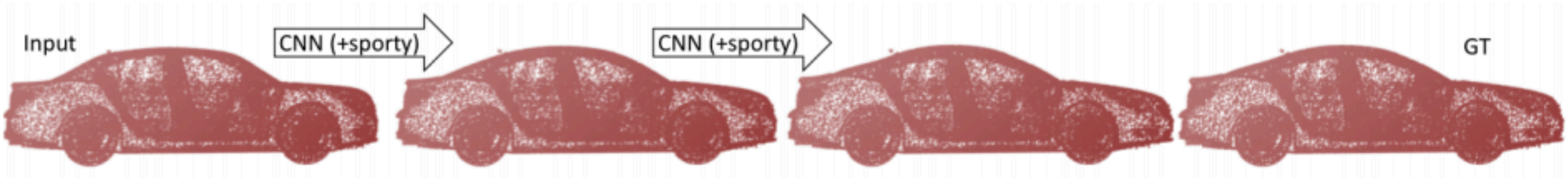
Representation for 3D: Volumetric Deformation



[Yumer and Mitra, ECCV, 2016]

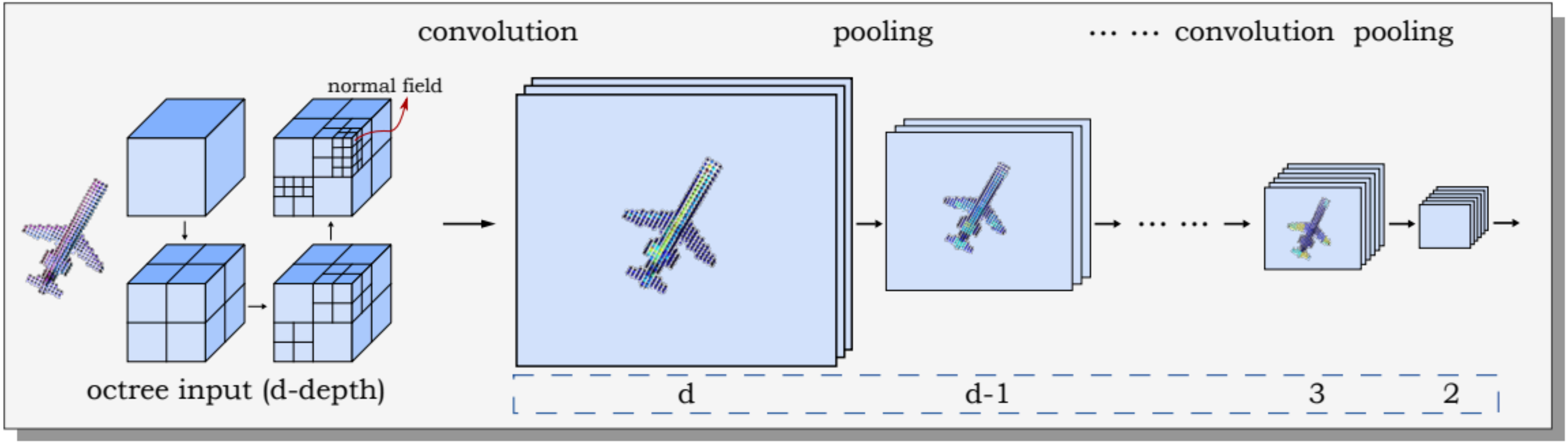
Representation for 3D: Volumetric Deformation





[Yumer and Mitra, ECCV, 2016]

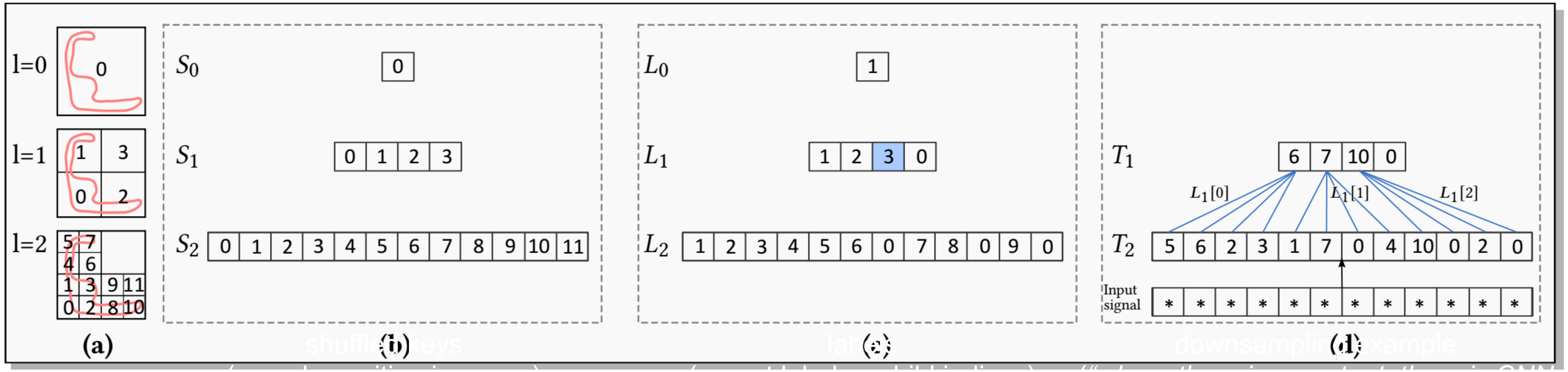
Efficient Volumetric Datastructures



[Wang et al. 2017]

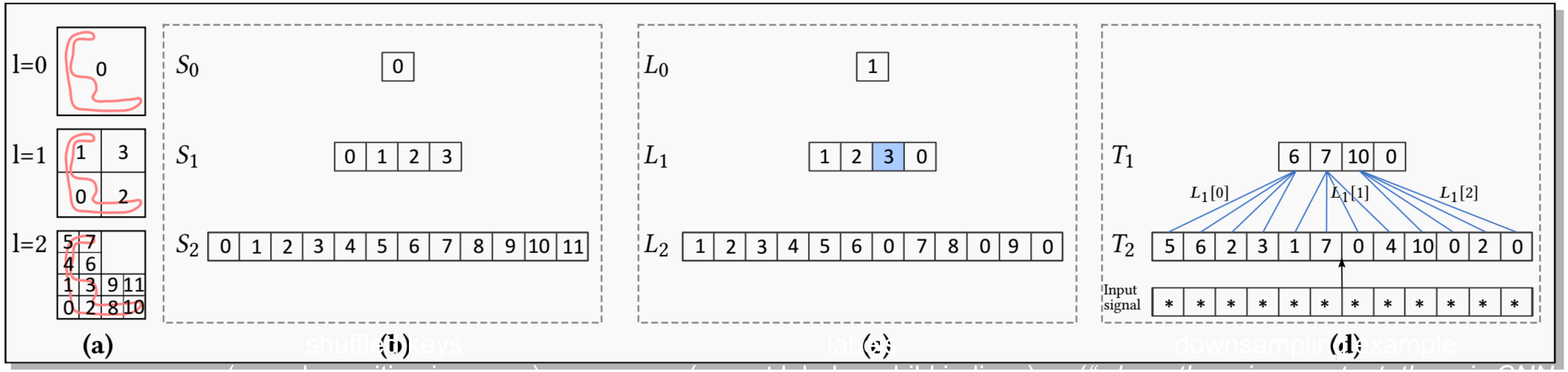
Data Structure and CNN Operations

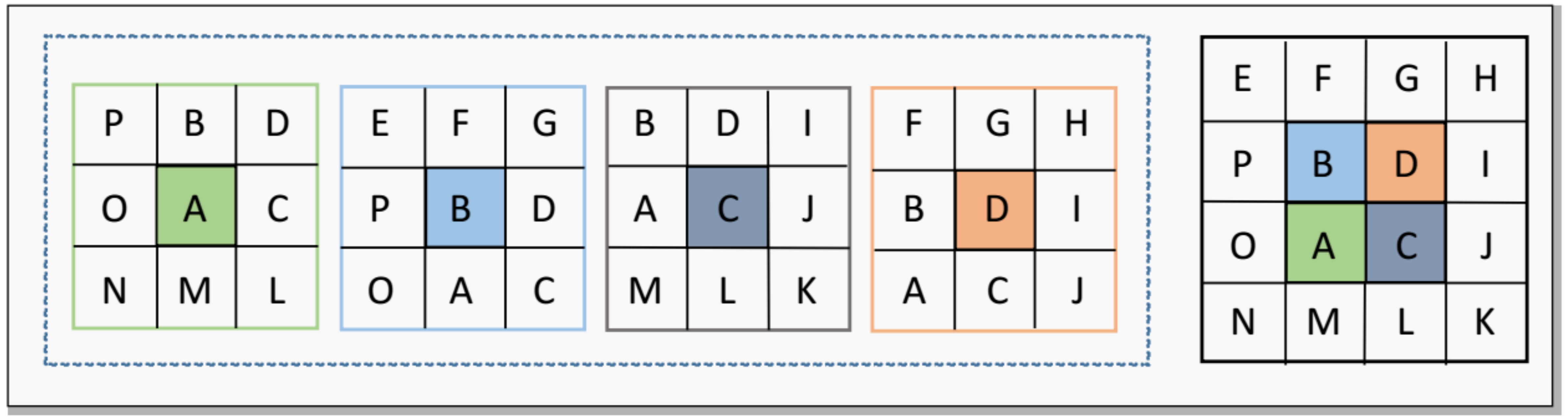
O-CNN



Data Structure and CNN Operations

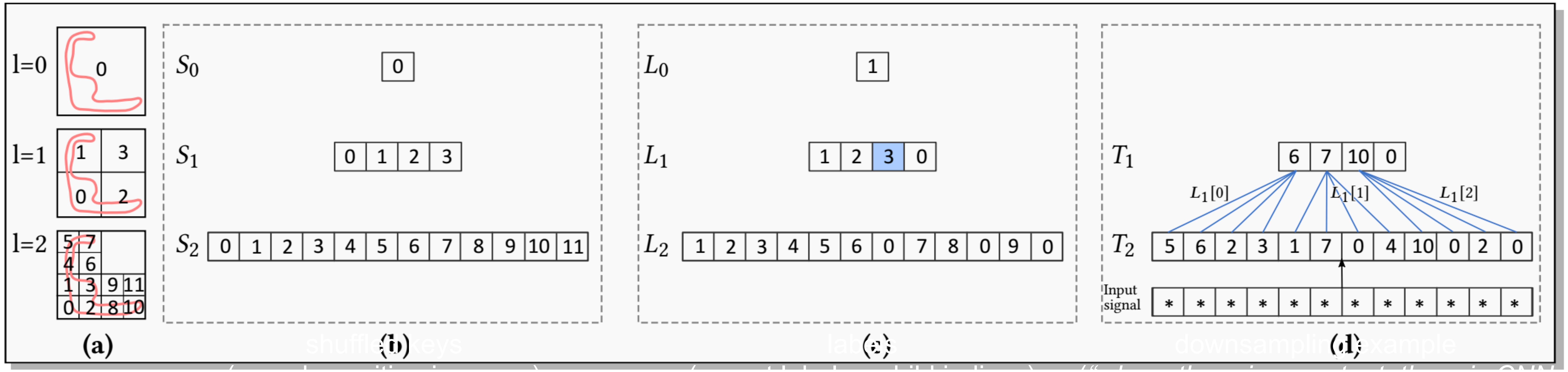
O-CNN

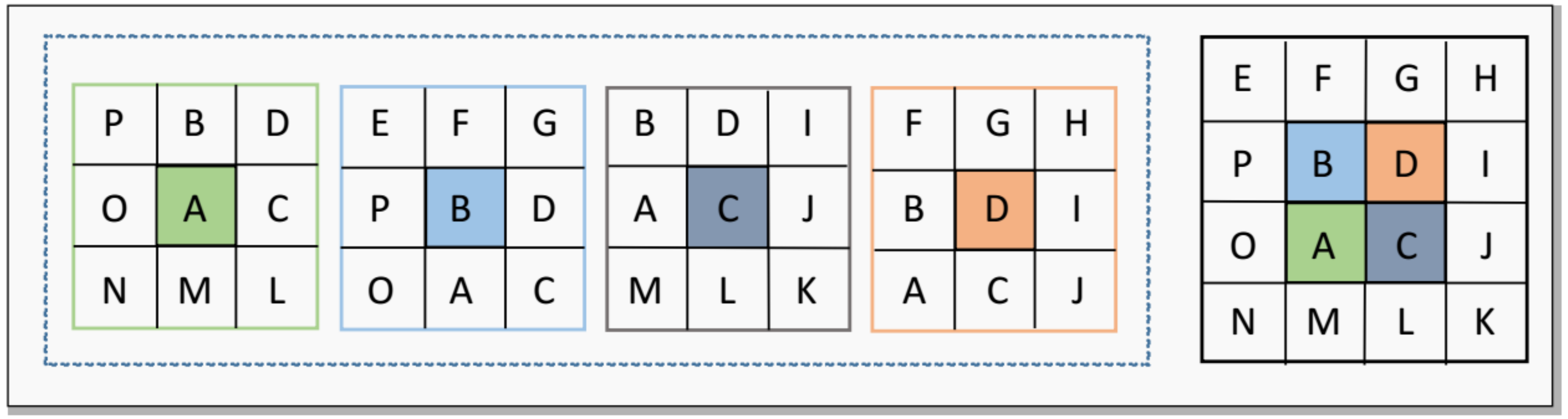


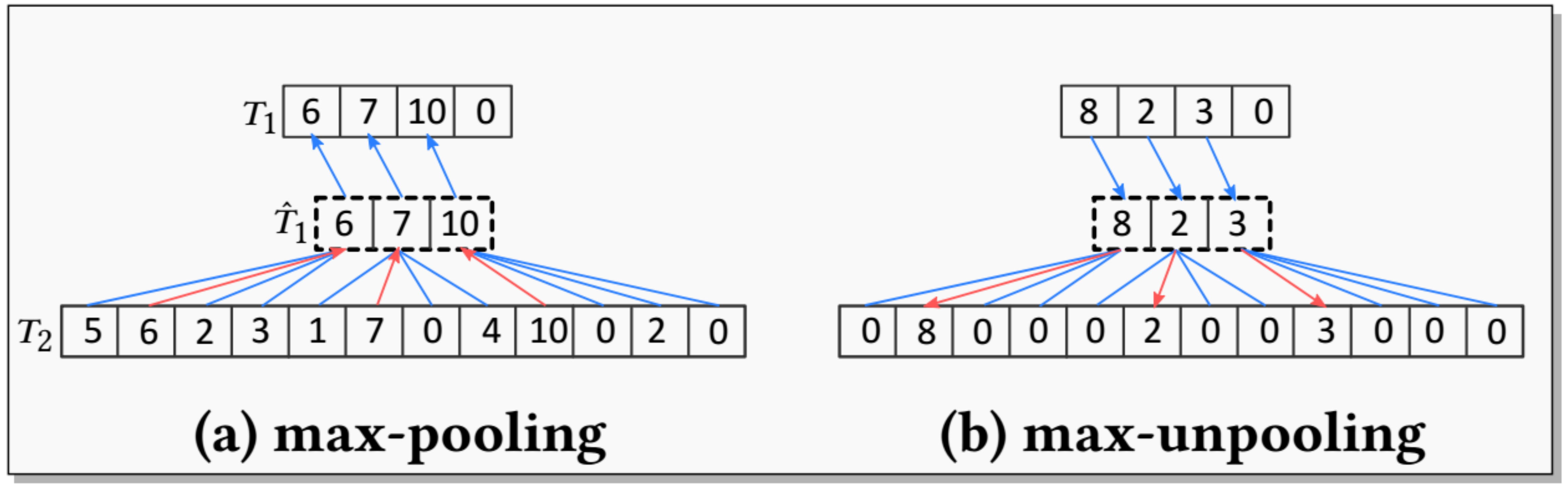


Data Structure and CNN Operations

O-CNN

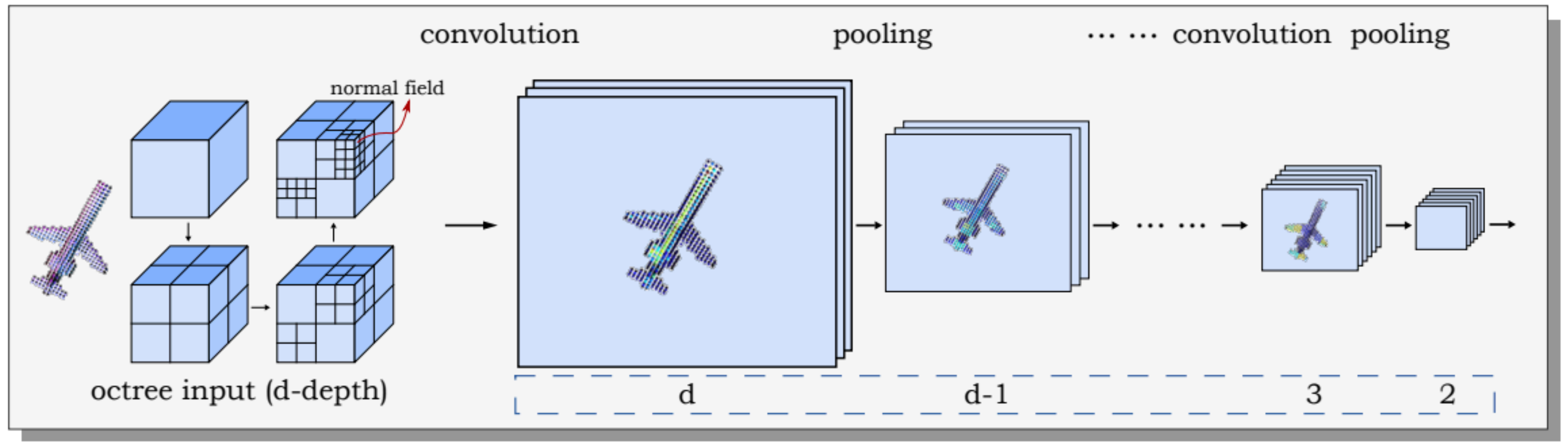






Efficient Volumetric Datastructures

Encoder



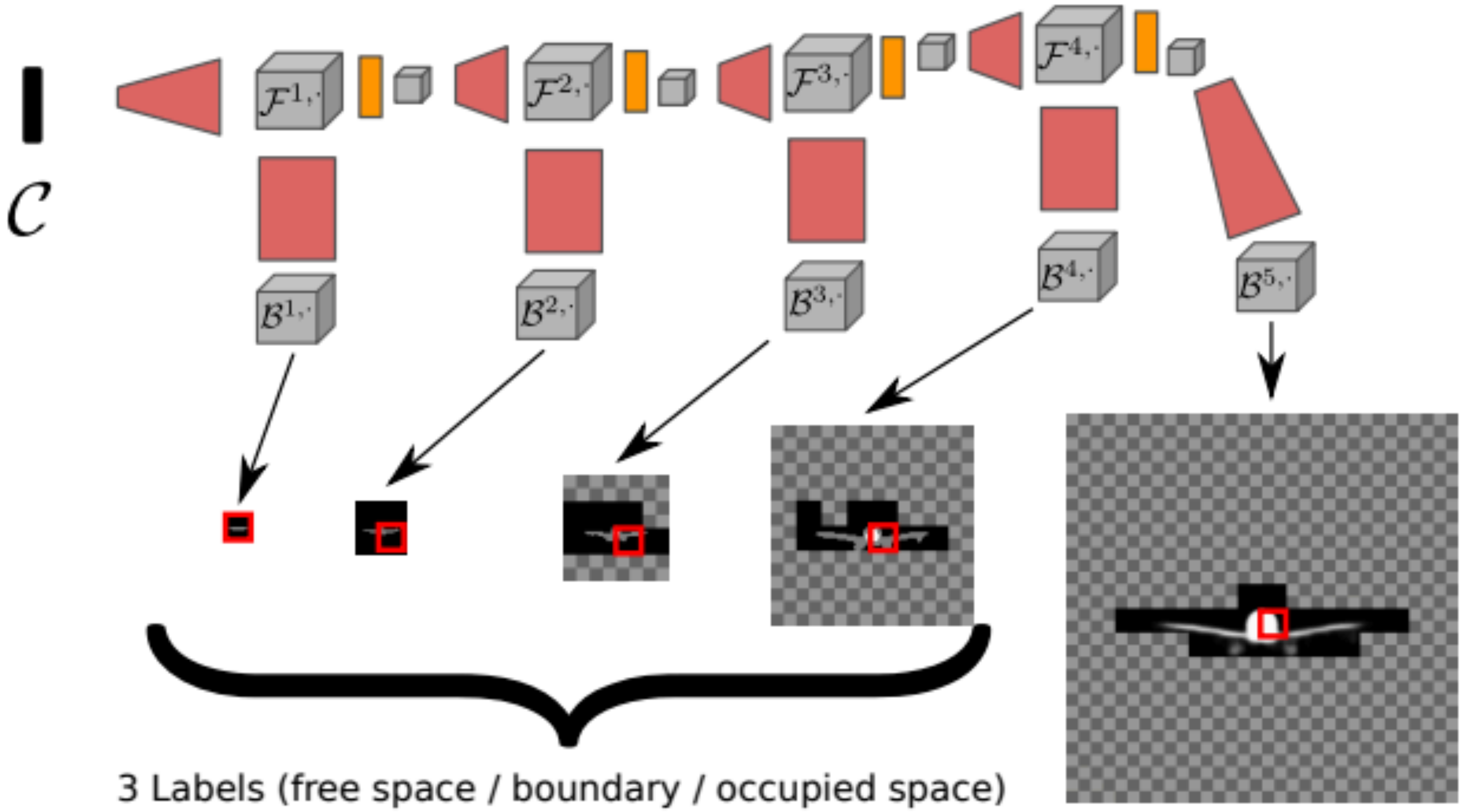
Wang et al. 2017

only generate non-empty voxels

Decoder/generator

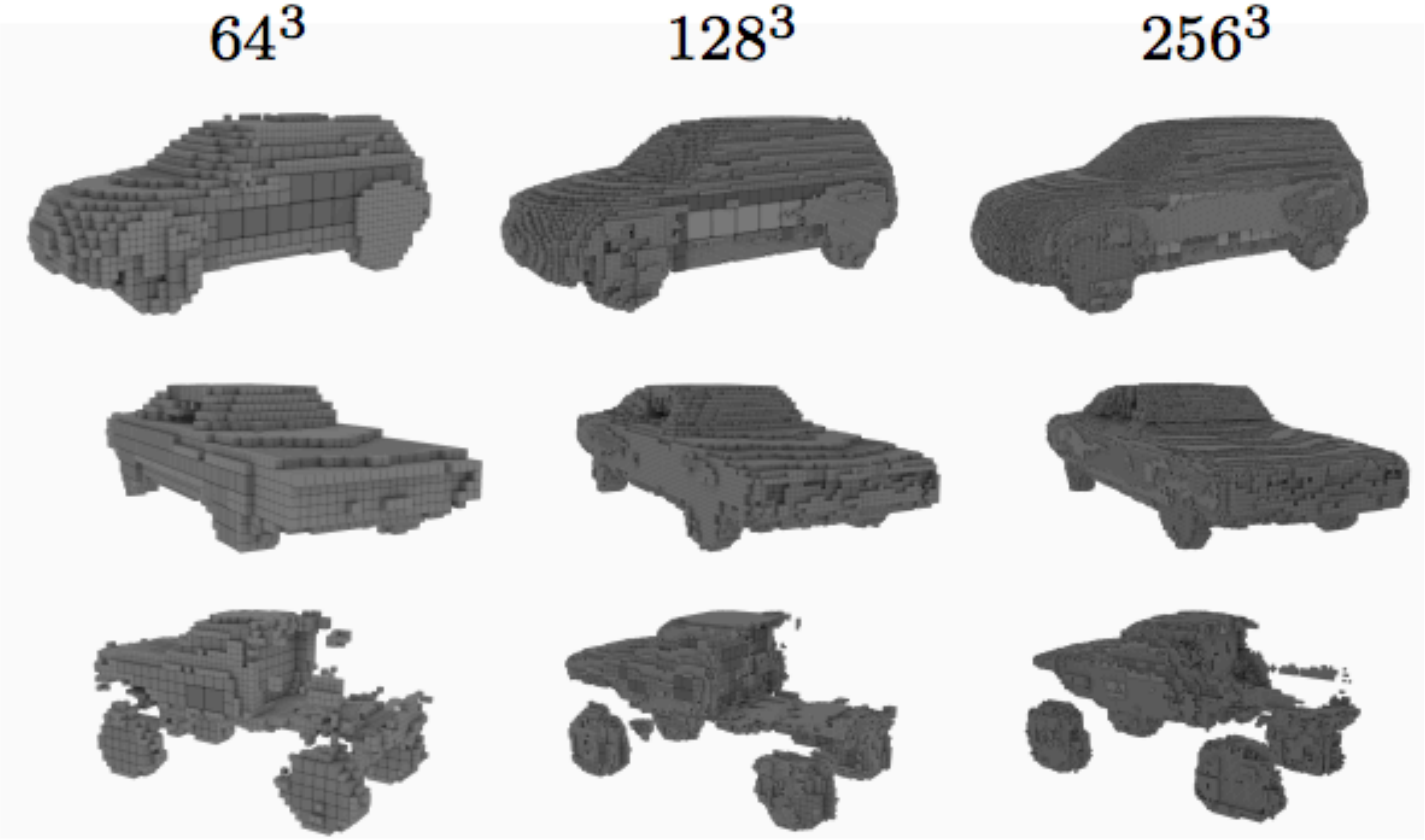
Volumetric (Up-) Convolutions

Cropping



[Hane et al. 2018]

Efficient Volumetric Datastructures



[Hane et al. 2018]

Lower Memory Footprint

O-CNI

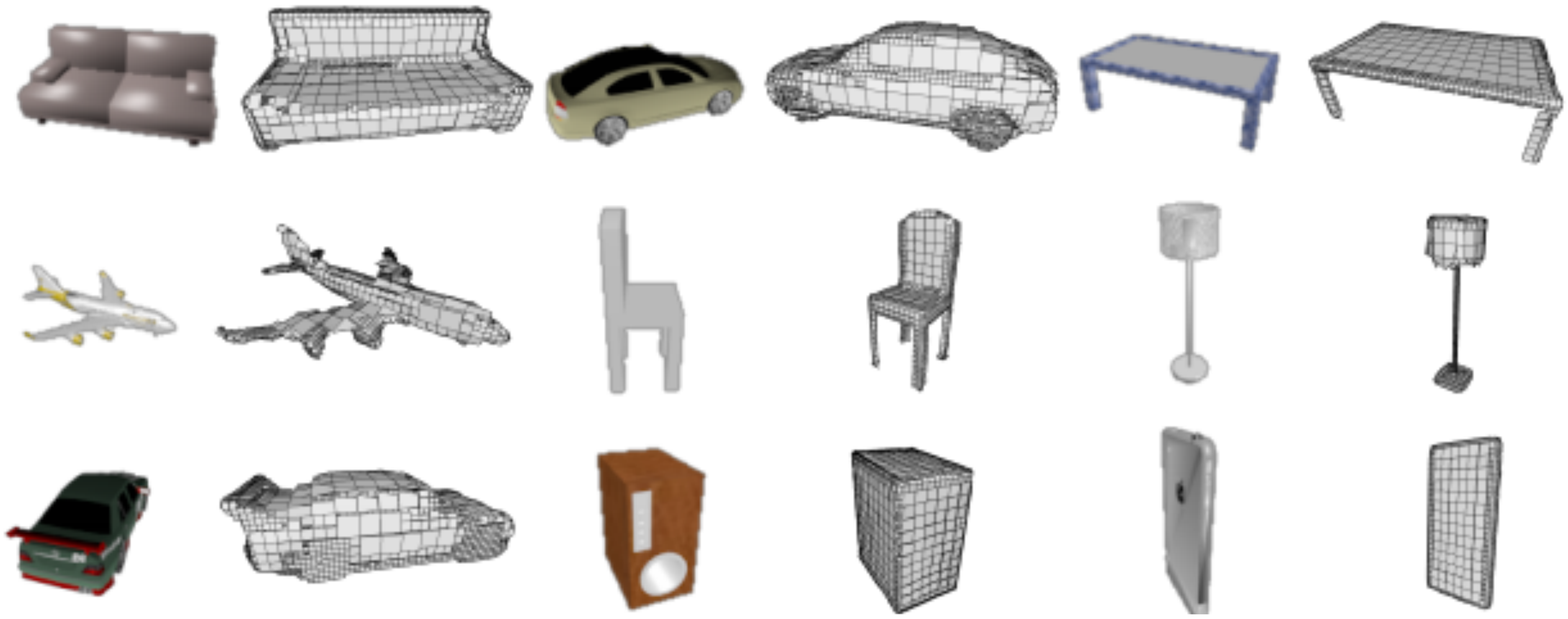
Method	16^3	32^3	64^3	128^3	256^3
O-CNN	0.32GB	0.58GB	1.1GB	2.7GB	6.4GB
full voxel+binary	0.23GB	0.71GB	3.7GB	Out of memory	Out of memory
full voxel+normal	0.27GB	1.20GB	4.3GB	Out of memory	Out of memory

Table 3. Comparisons on GPU-memory consumption. The batch size is 32.

Method	16^3	32^3	64^3	128^3	256^3
O-CNN	17ms	33ms	90ms	327ms	1265ms
full voxel+binary	59ms	425ms	1648ms	-	-
full voxel+normal	75ms	510ms	4654ms	-	-

Table 4. Timings of one backward and forward operation in milliseconds. The batch size is 32.

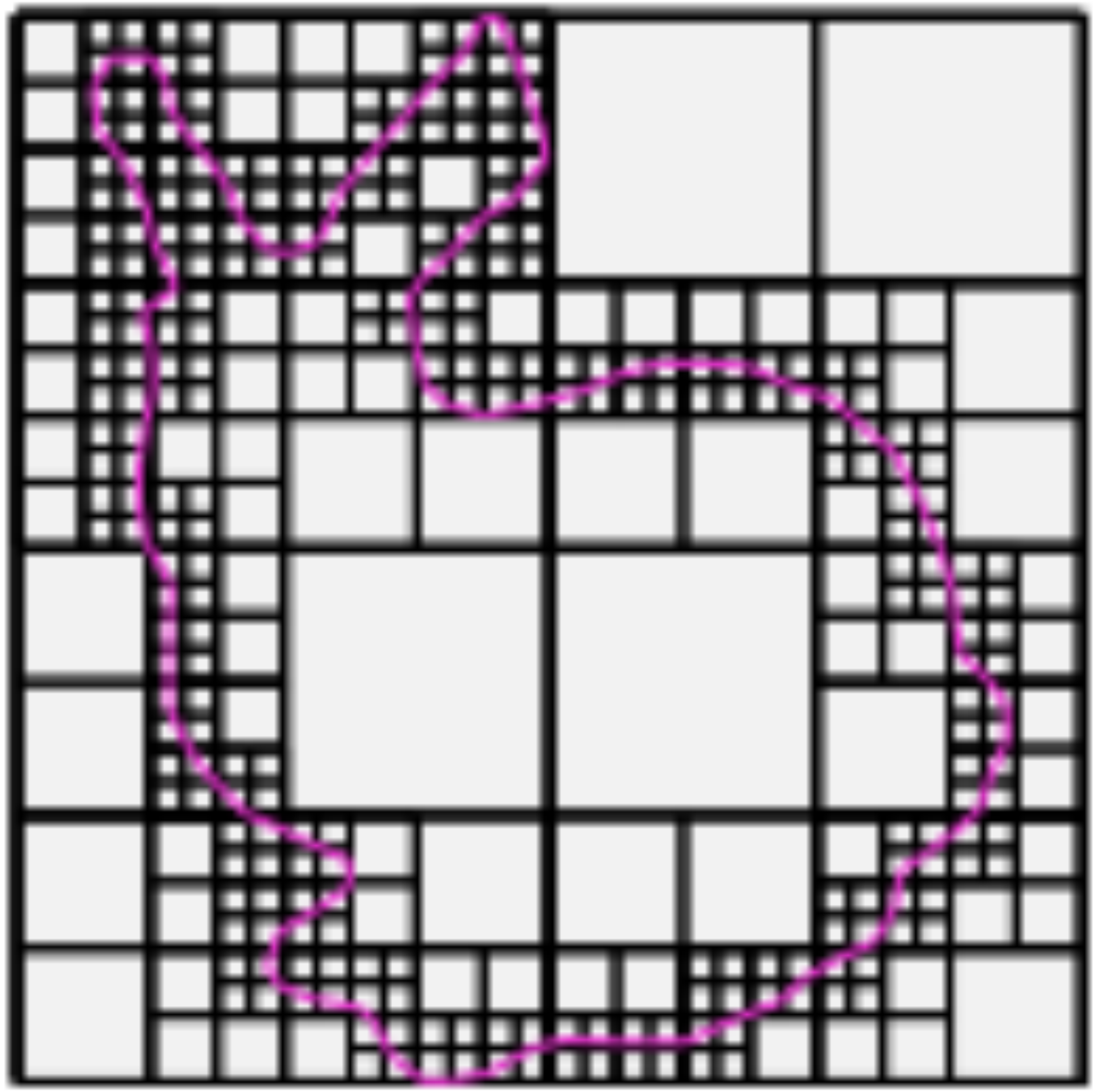
Adaptive O-CNN



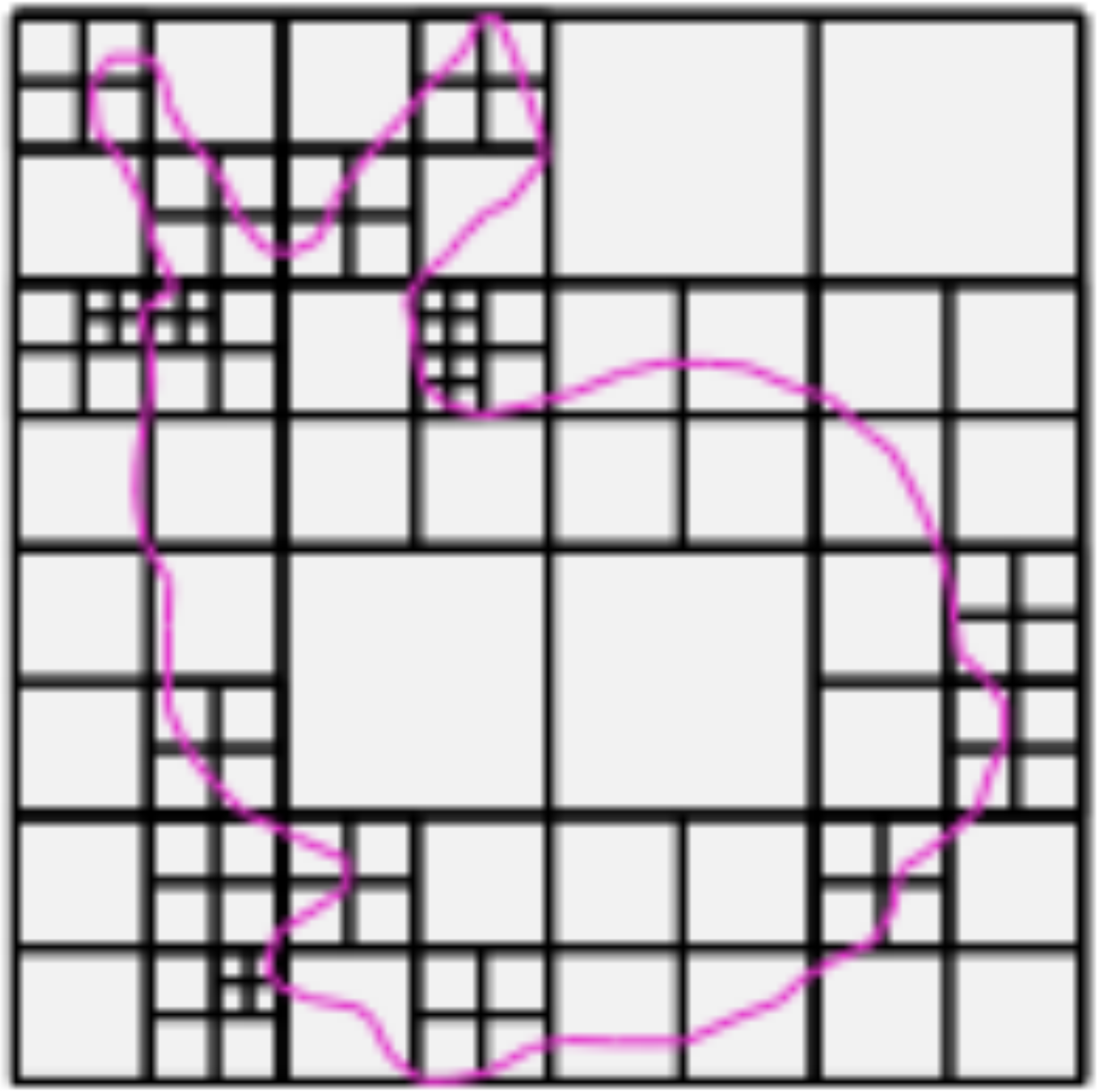
[Wang et al. 2018]

image to planar patch-based shapes

First-order Patches



OCNN



Adaptive OCNN

Motivation: Creating a Layered Representation

Input image

2 minutes

Single channel edit

[Innamorati, Ritschel, Weyrich, Mitra, EGSR, 2017]

Motivation: Creating a Layered Representation

Input image

2 minutes

Single channel edit

[Innamorati, Ritschel, Weyrich, Mitra, EGSR, 2017]

Motivation: Creating a Layered Representation

Input image

2 minutes

Single channel edit

30 seconds

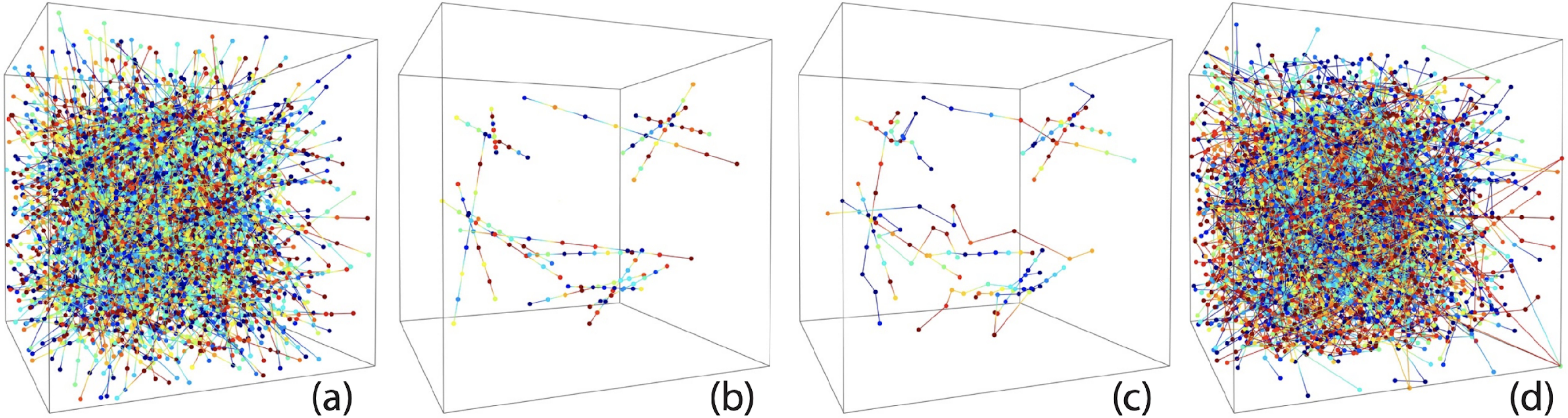
Our edit

[Innamorati, Ritschel, Weyrich, Mitra, EGSR, 2017]

Field Probing Neural Networks for 3D Data

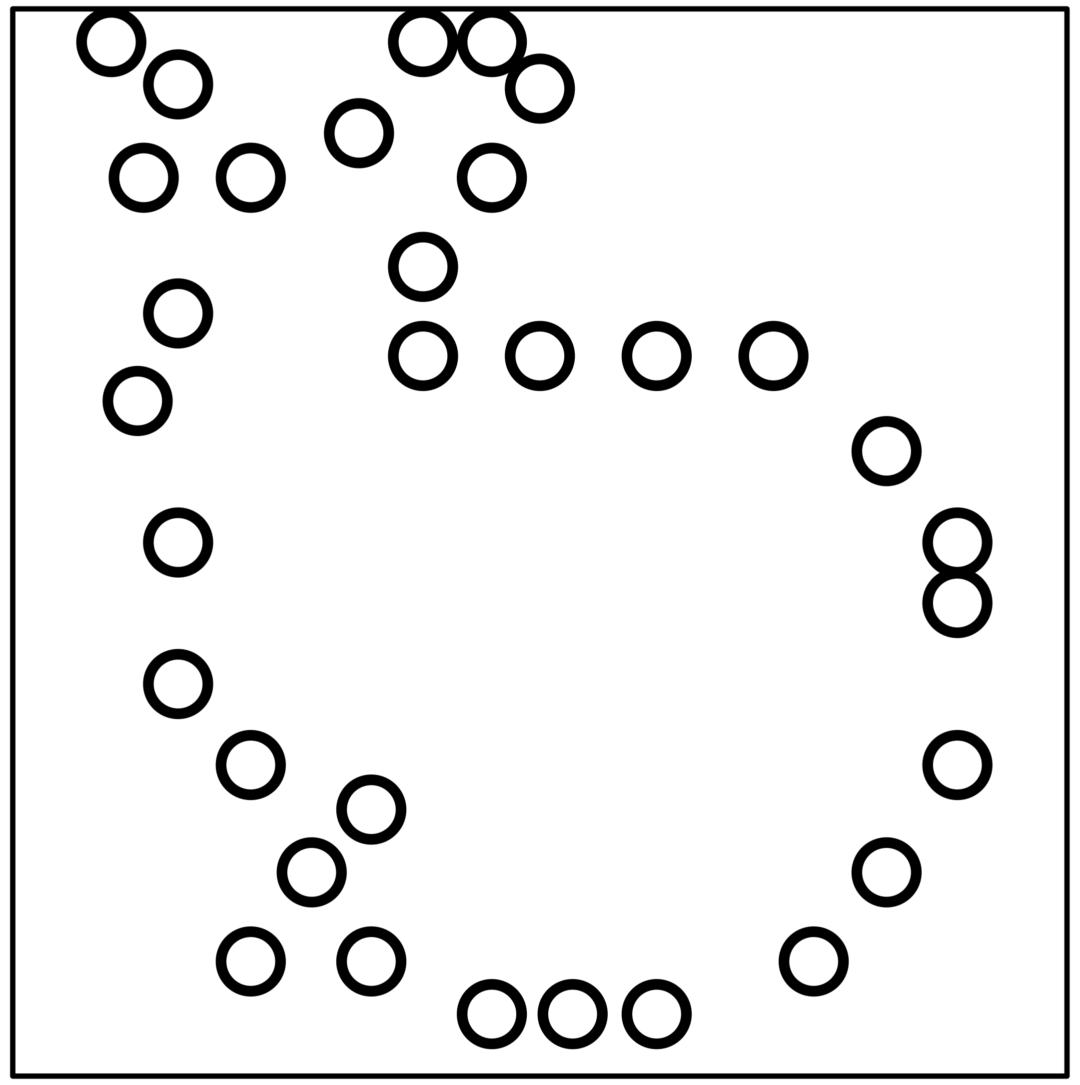
[Li et al. 2016]

Field Probing Neural Networks for 3D Data

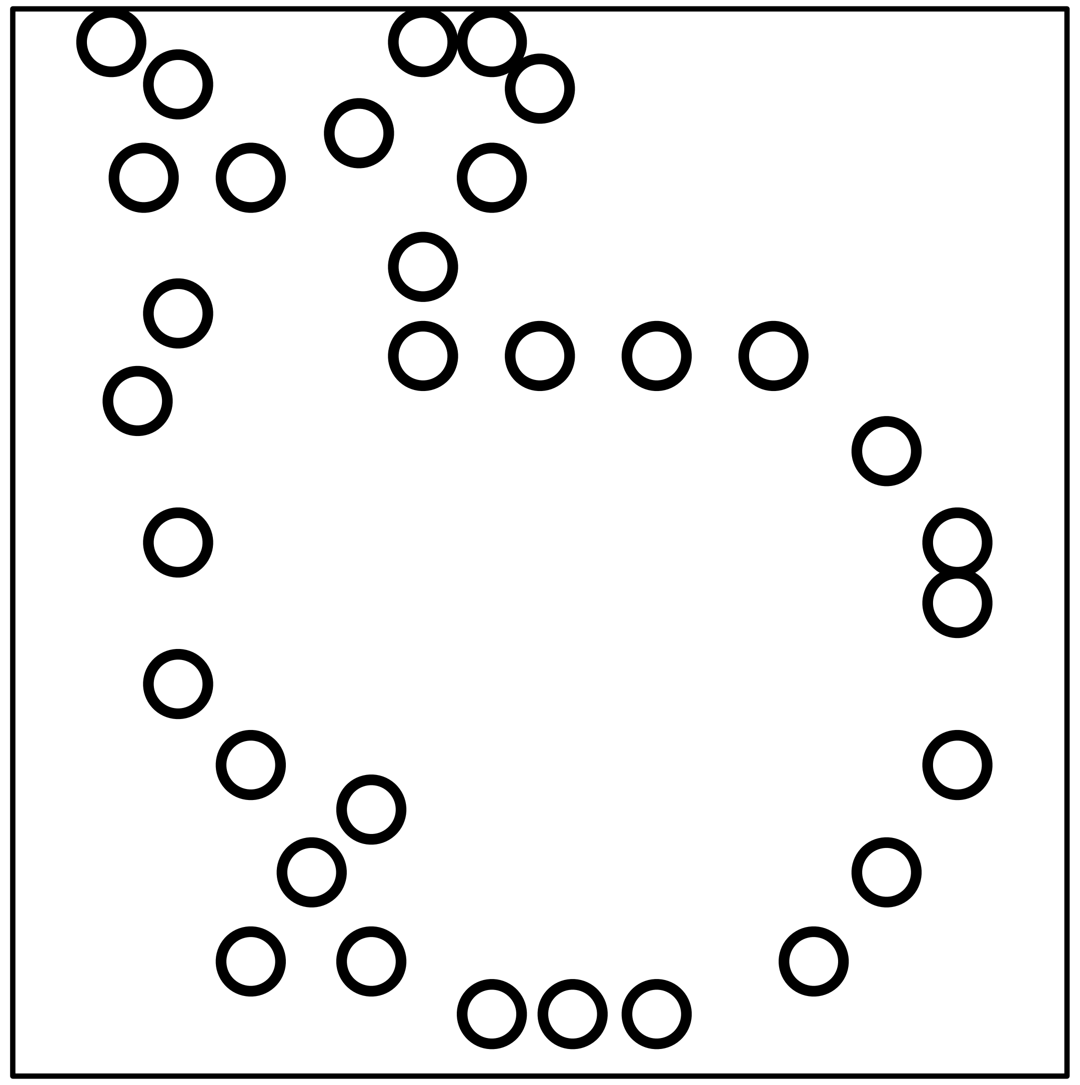


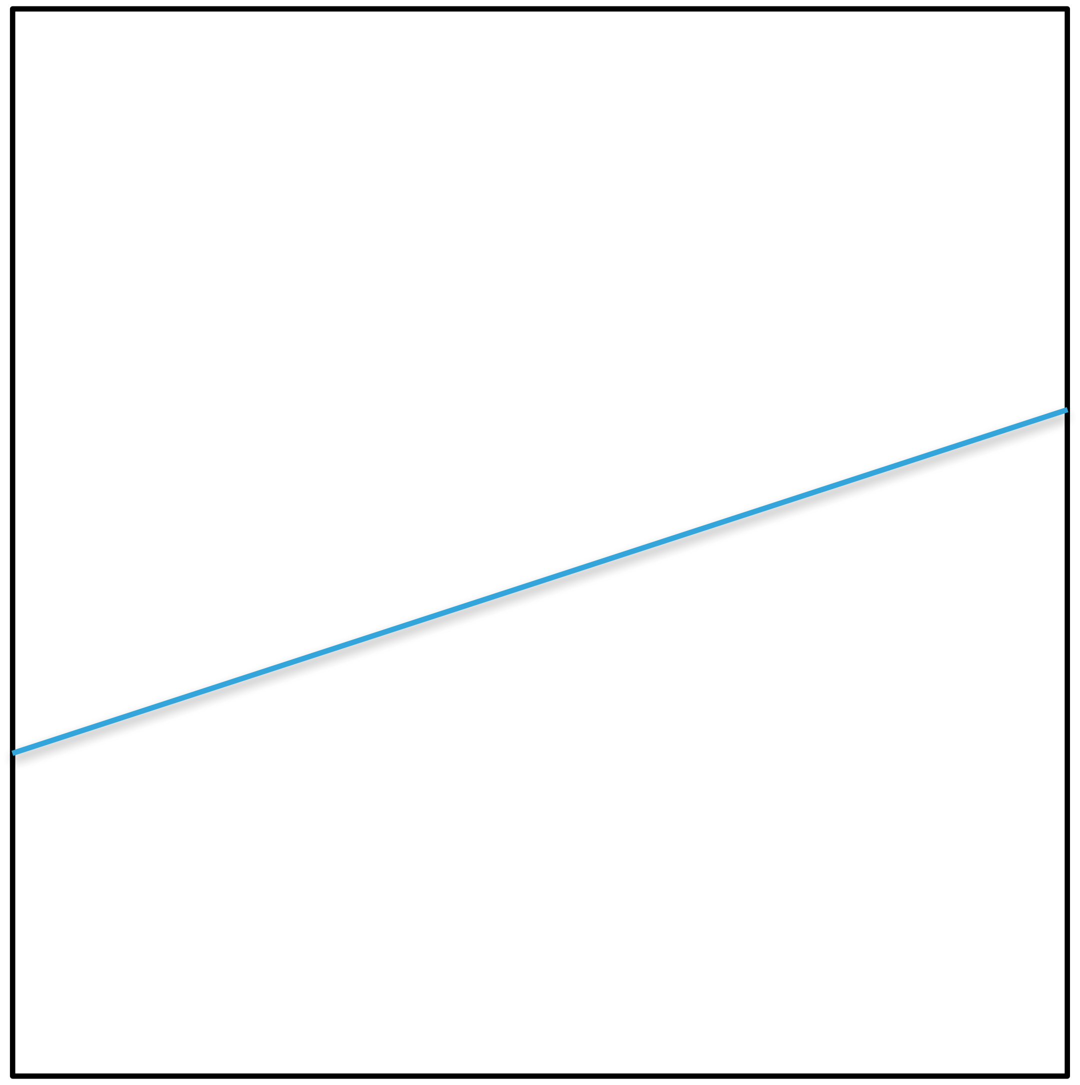
[Li et al. 2016]

Spatial Probes

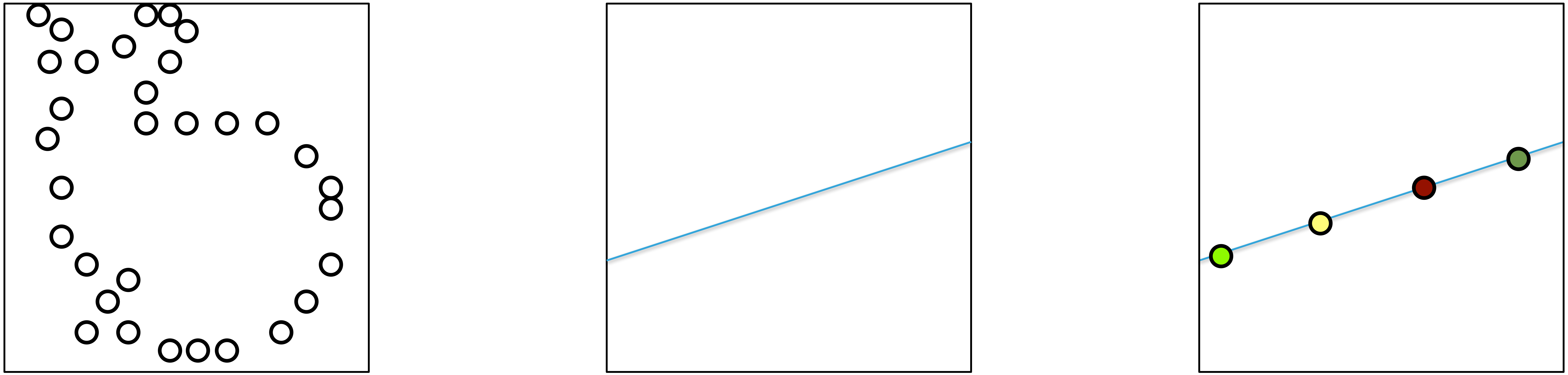


Spatial Probes





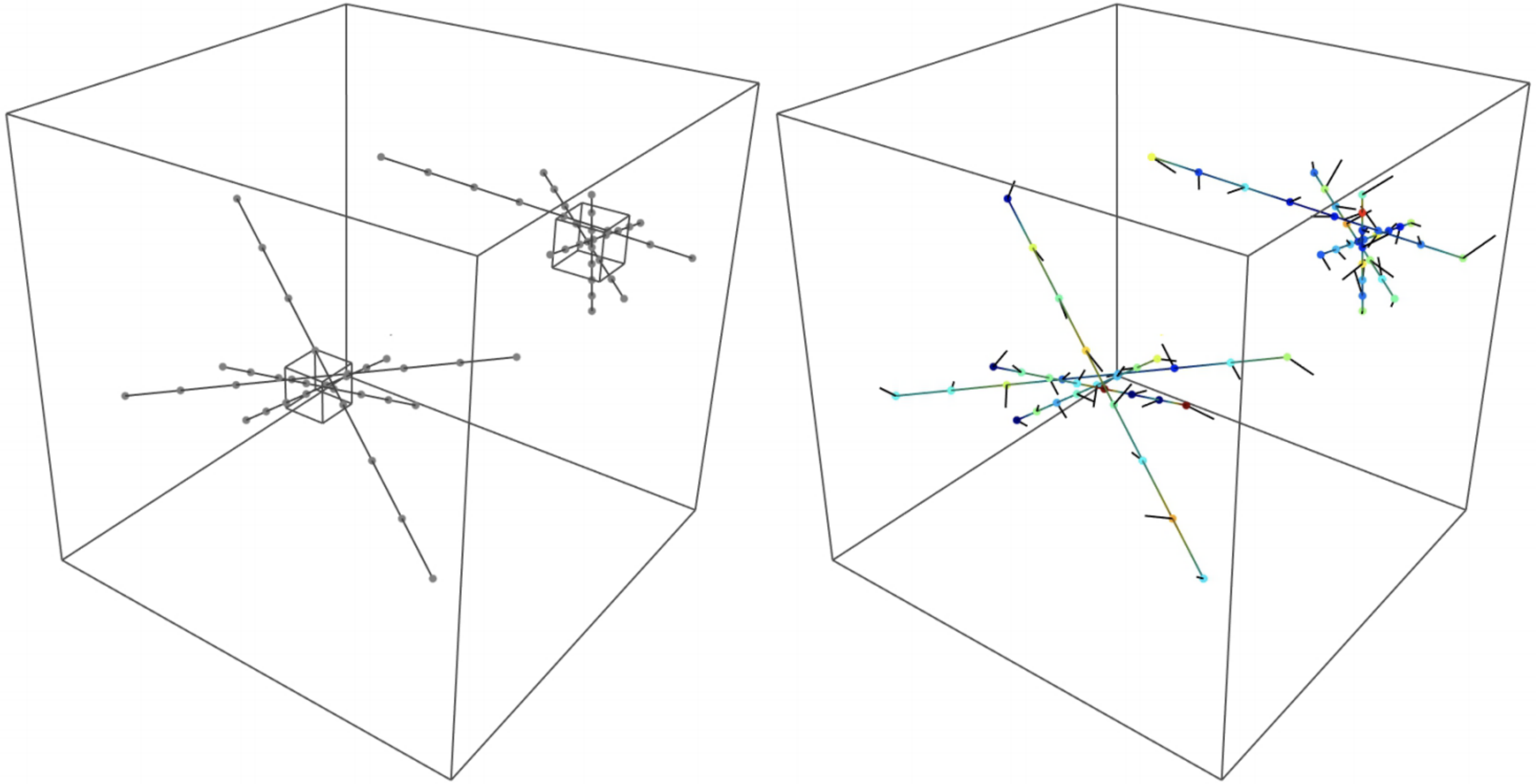
Spatial Probes



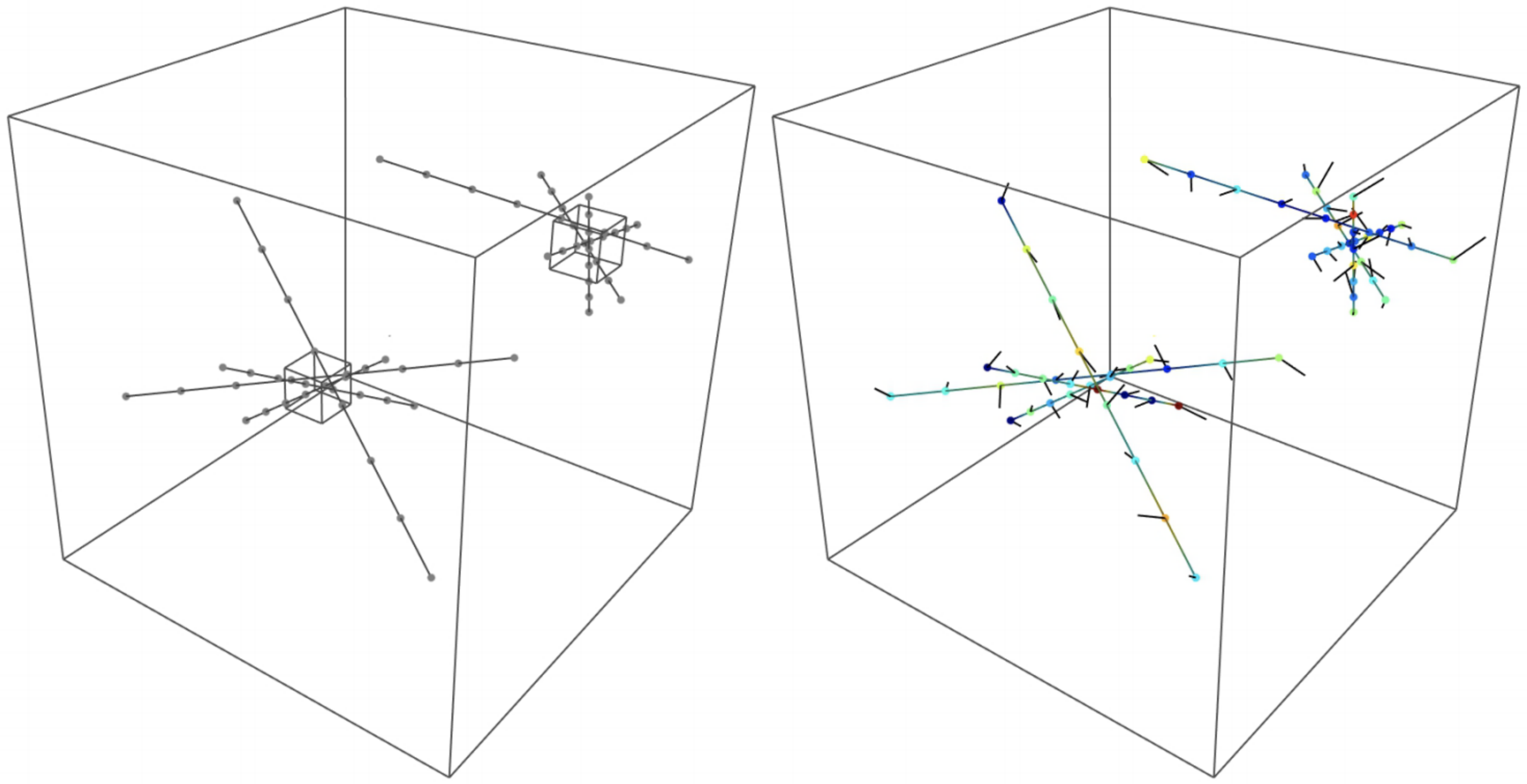
Method Details

Details

Method Details

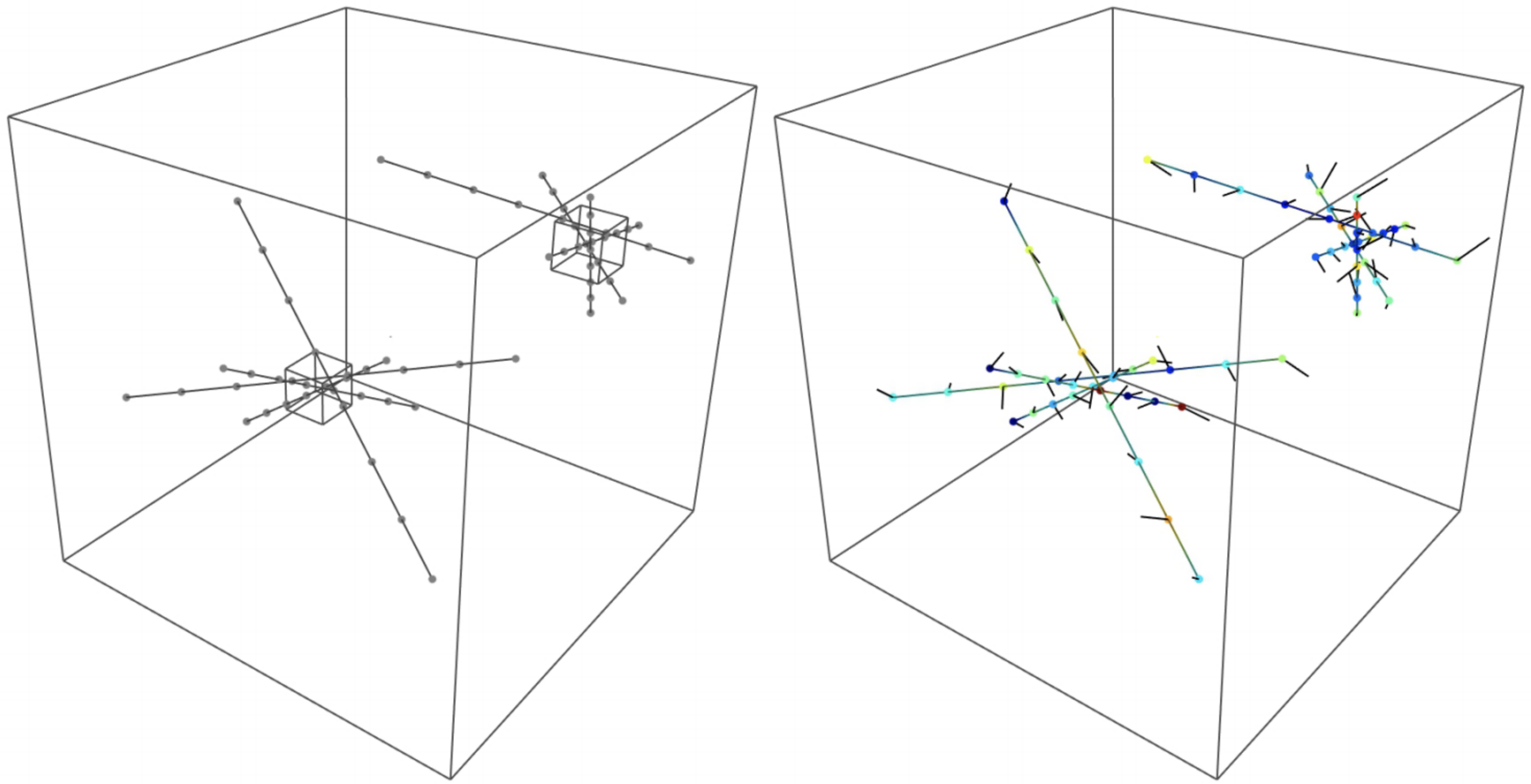


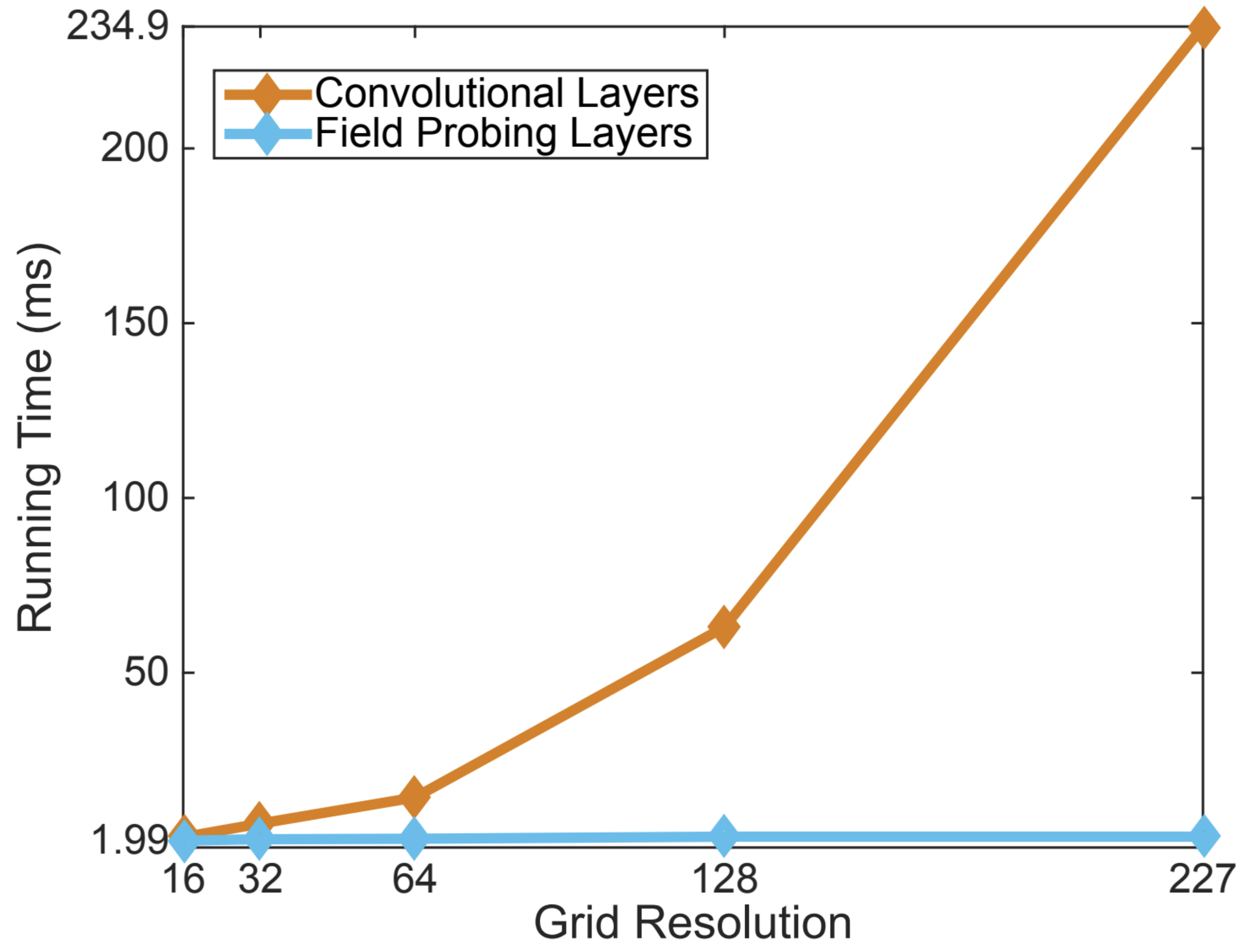
Method Details



$$v_c = v(\{p_{c,i,j}\}, \{w_{c,i,j}\}) = \sum_{\substack{i=1,\ldots,N \\ j=1,\ldots,T}} p_{c,i,j} \times w_{c,i,j}$$

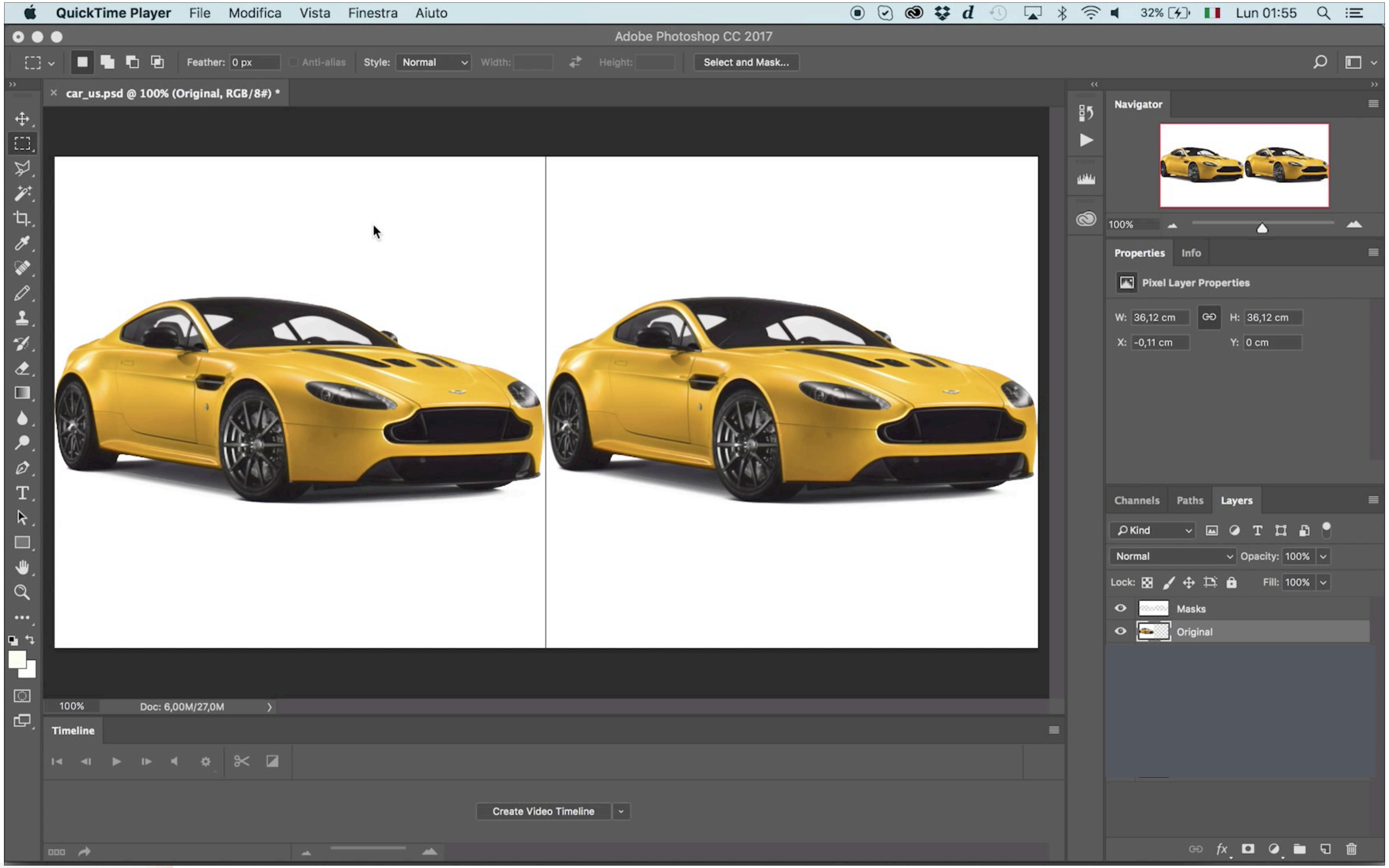
Method Details



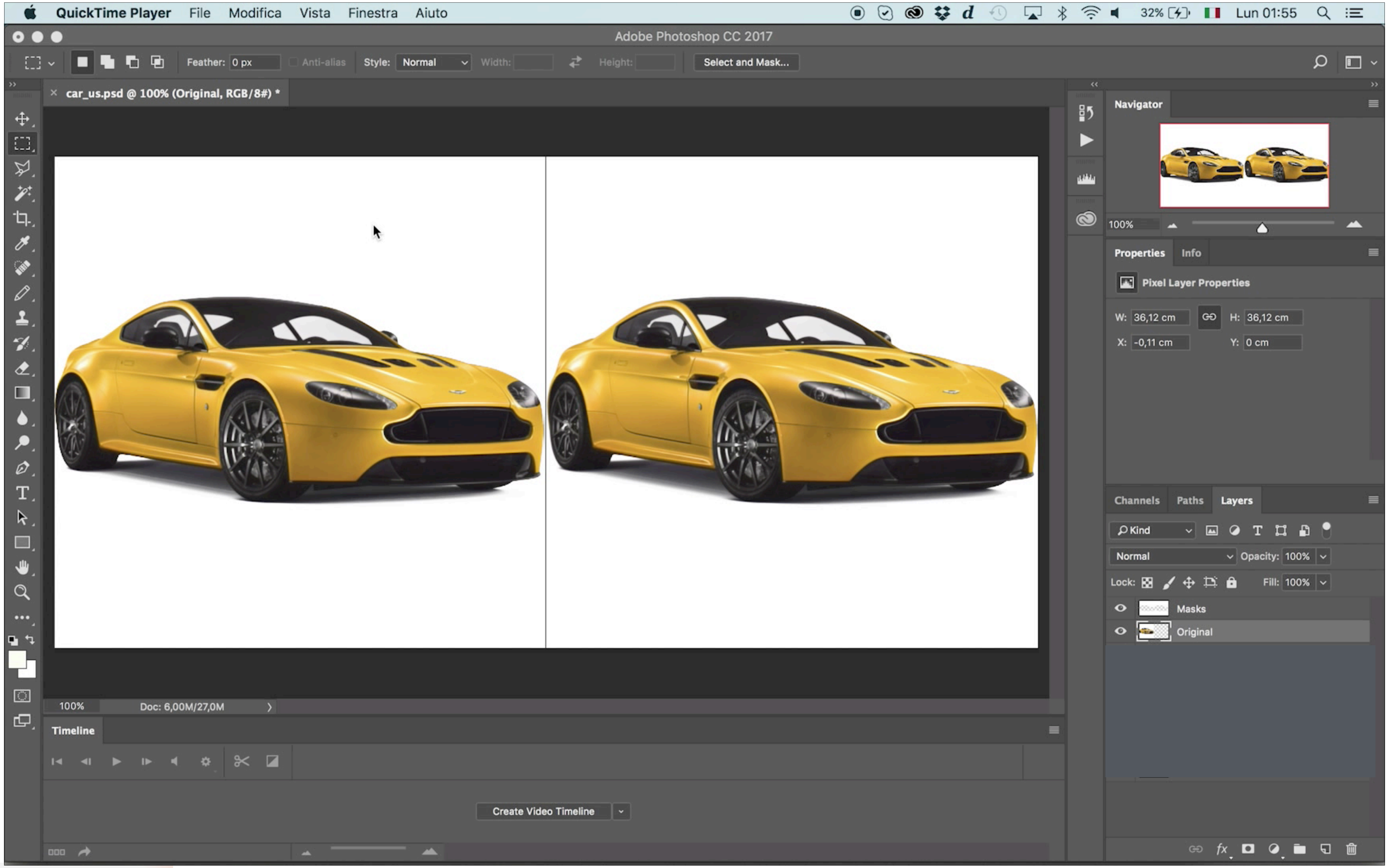


$$v_c = v(\{p_{c,i,j}\}, \{w_{c,i,j}\}) = \sum_{\substack{i=1,\ldots,N \\ j=1,\ldots,T}} p_{c,i,j} \times w_{c,i,j}$$

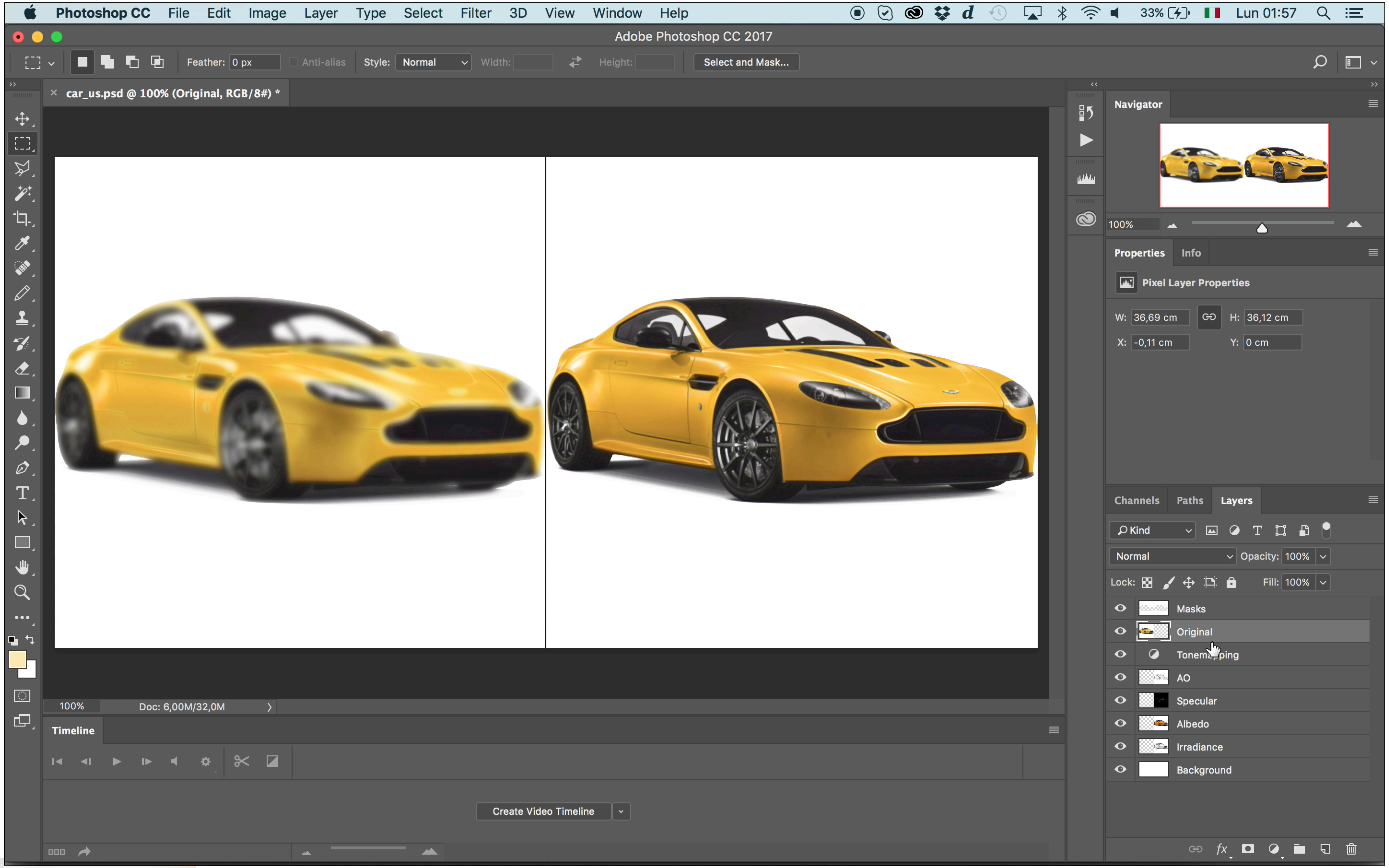
On Real Images



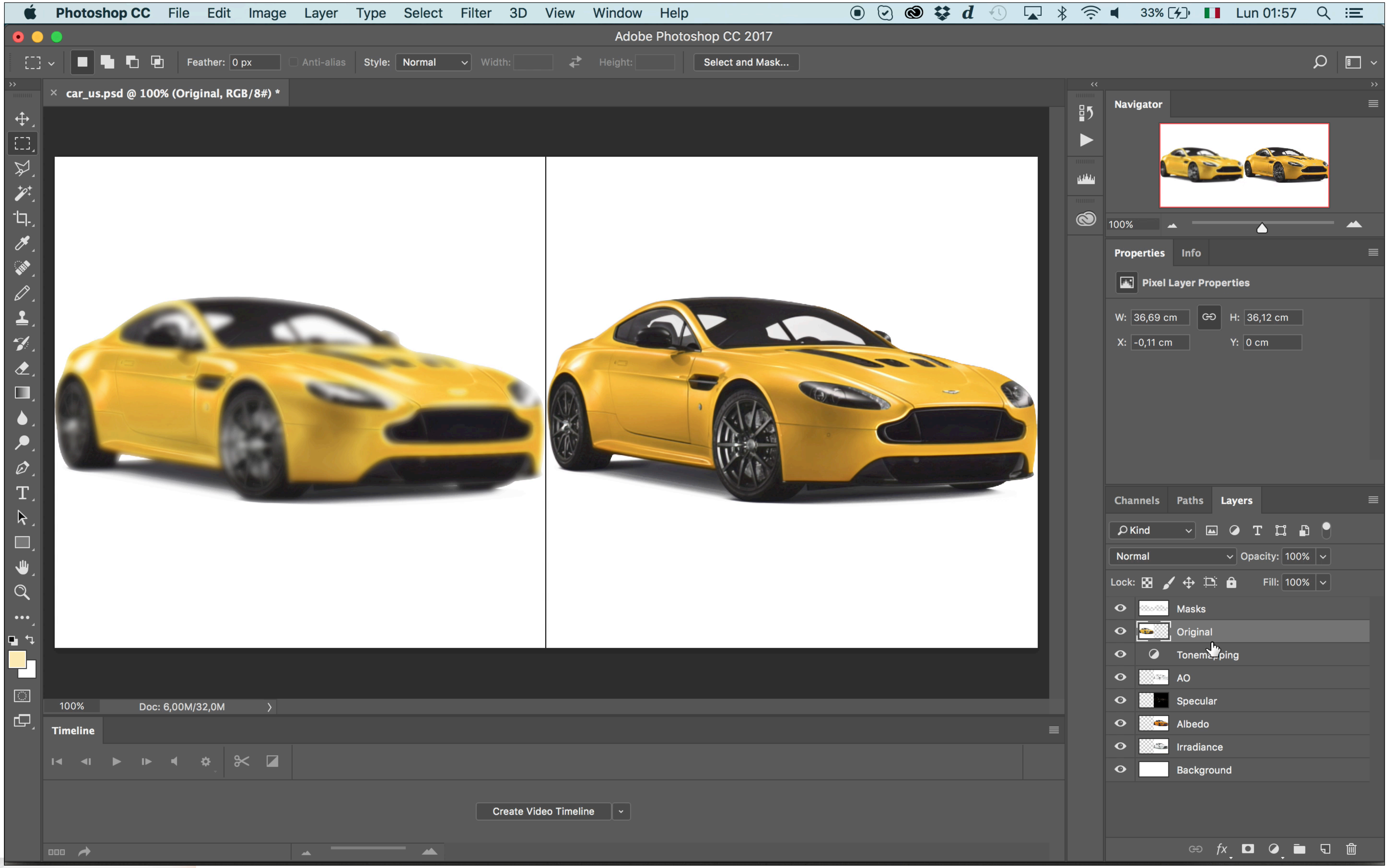
On Real Images



With Inferred Layered Representation



With Inferred Layered Representation



When We Do Not Have 3D Training Data

[Henzler, Mitra, Ritschel, Arxiv, 2019]

When We Do Not Have 3D Training Data

- Look back at image formation model (rendering equation)

[Henzler, Mitra, Ritschel, Arxiv, 2019]

When We Do Not Have 3D Training Data

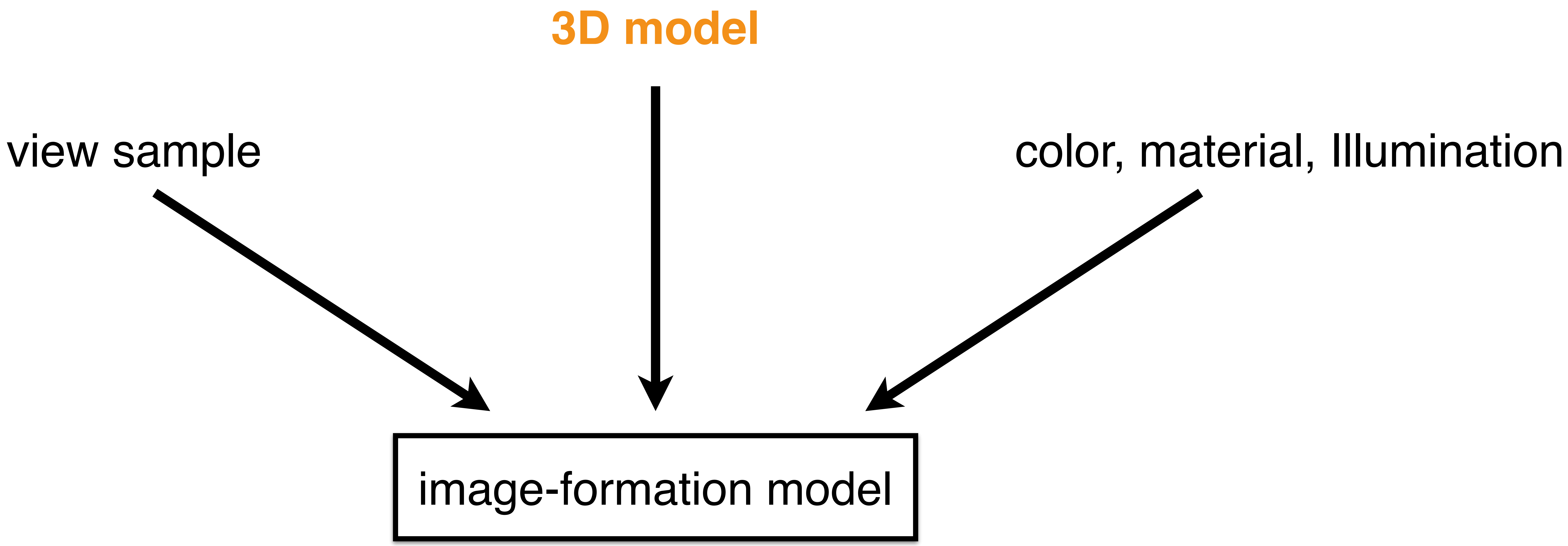
- Look back at image formation model (rendering equation)

3D model

[Henzler, Mitra, Ritschel, Arxiv, 2019]

When We Do Not Have 3D Training Data

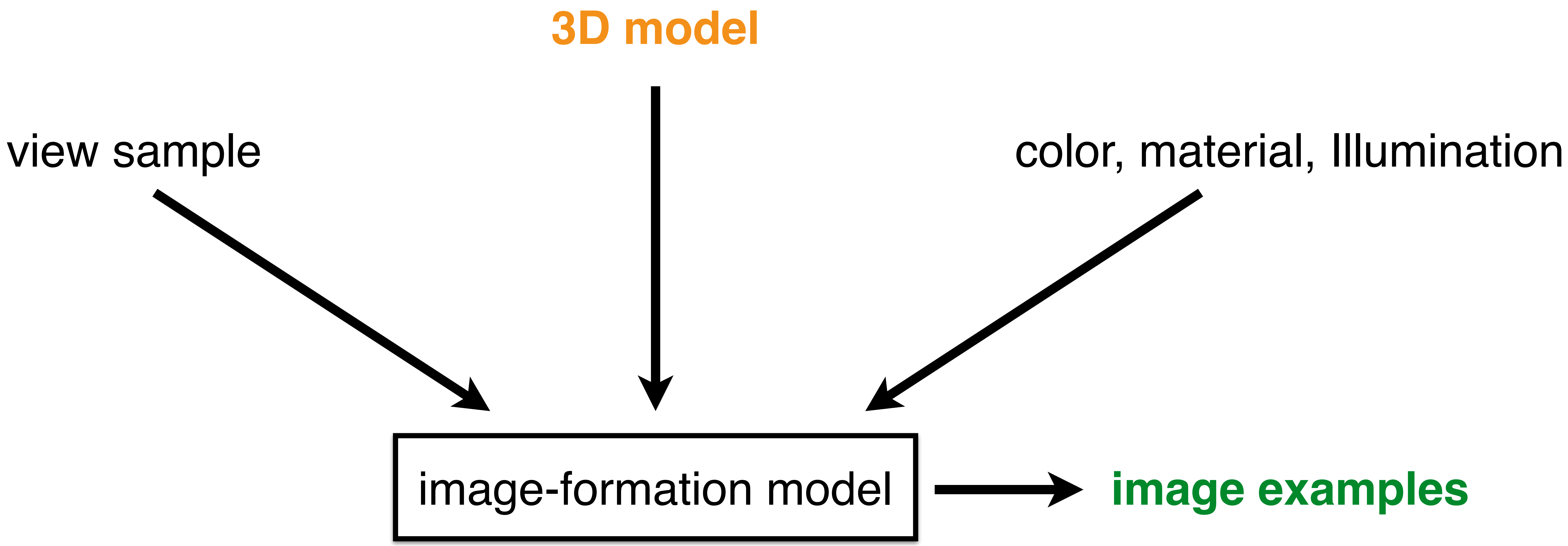
- Look back at image formation model (rendering equation)



[Henzler, Mitra, Ritschel, Arxiv, 2019]

When We Do Not Have 3D Training Data

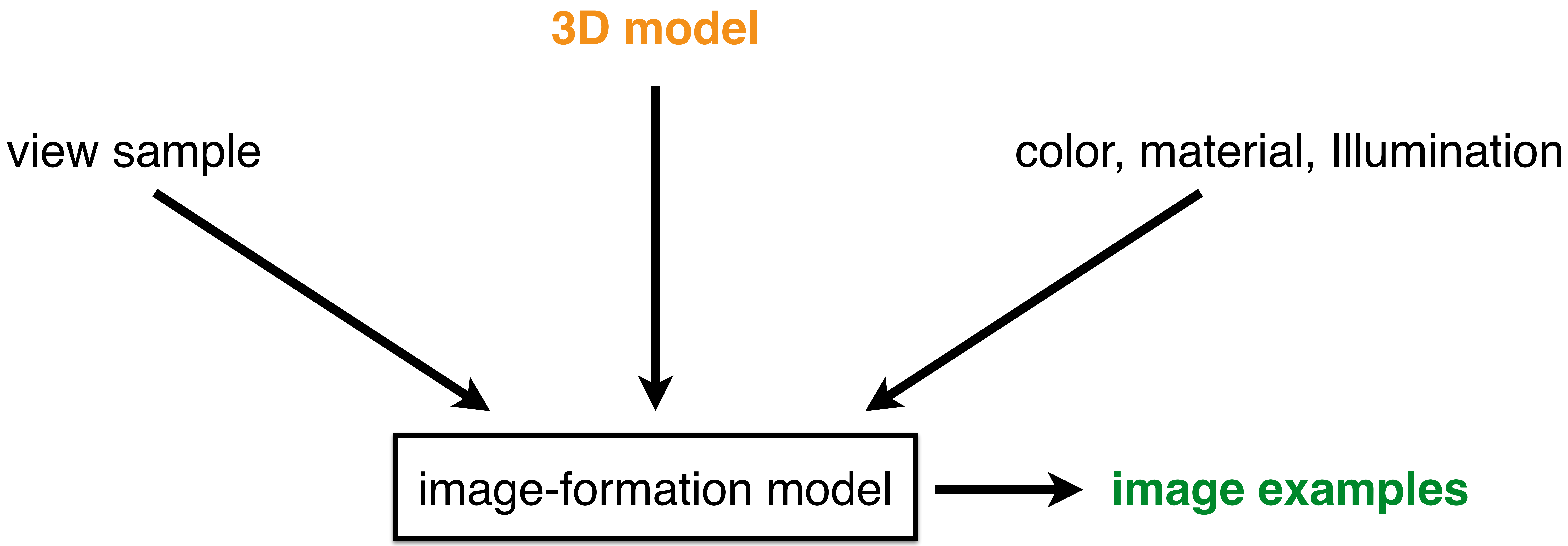
- Look back at image formation model (rendering equation)



[Henzler, Mitra, Ritschel, Arxiv, 2019]

When We Do Not Have 3D Training Data

- Look back at image formation model (rendering equation)



- Image formation, view transformation are **known** functions/transformations

[Henzler, Mitra, Ritschel, Arxiv, 2019]

Image-formation Models

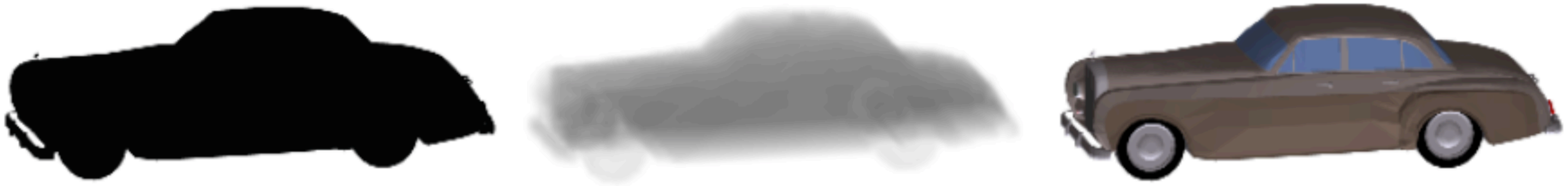
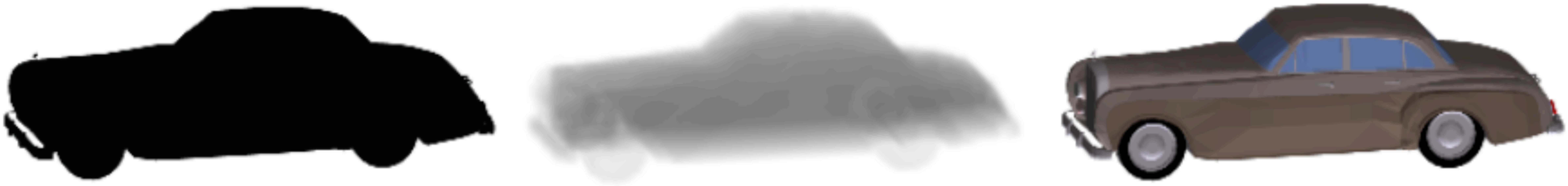


Image-formation Models



visual hull

$$\rho_{\mathrm{VH}}(\mathbf{v}) = 1 - e^{\sum_i (-N v_i)}/N$$

Image-formation Models

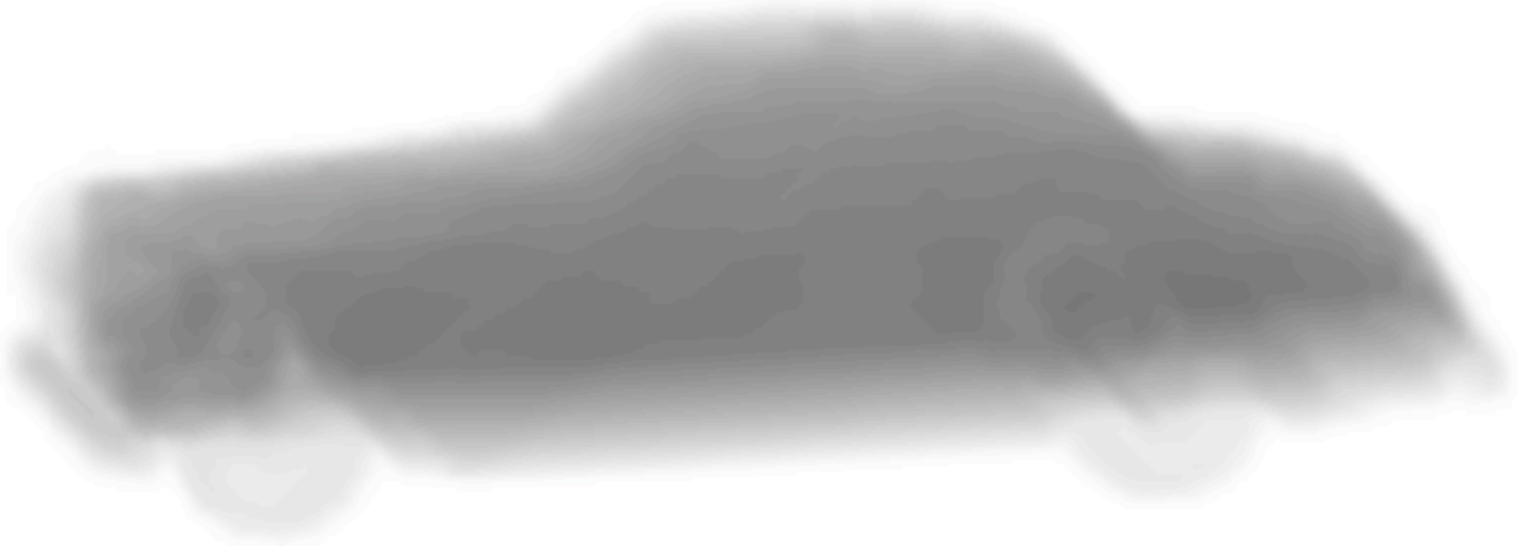
visual hull

absorption only

$$\rho_{\mathrm{VH}}(\mathbf{v}) = 1 - e^{\sum_i (-N v_i)}/N$$

$$\rho_{\mathrm{AO}}(\mathbf{v}) = 1 - \prod_i (1 - v_i)$$

Image-formation Models



visual hull

absorption only

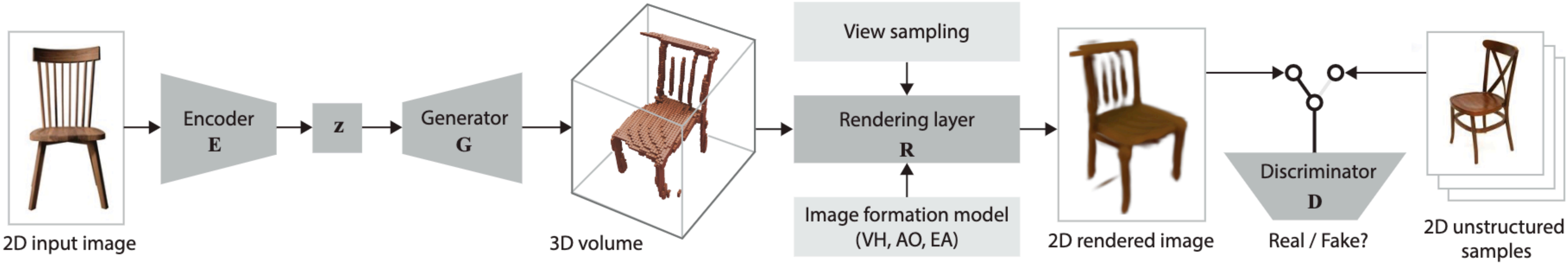
emission-absorption

$$\rho_{\mathrm{VH}}(\mathbf{v}) = 1 - e^{\sum_i (-N v_i)}/N$$

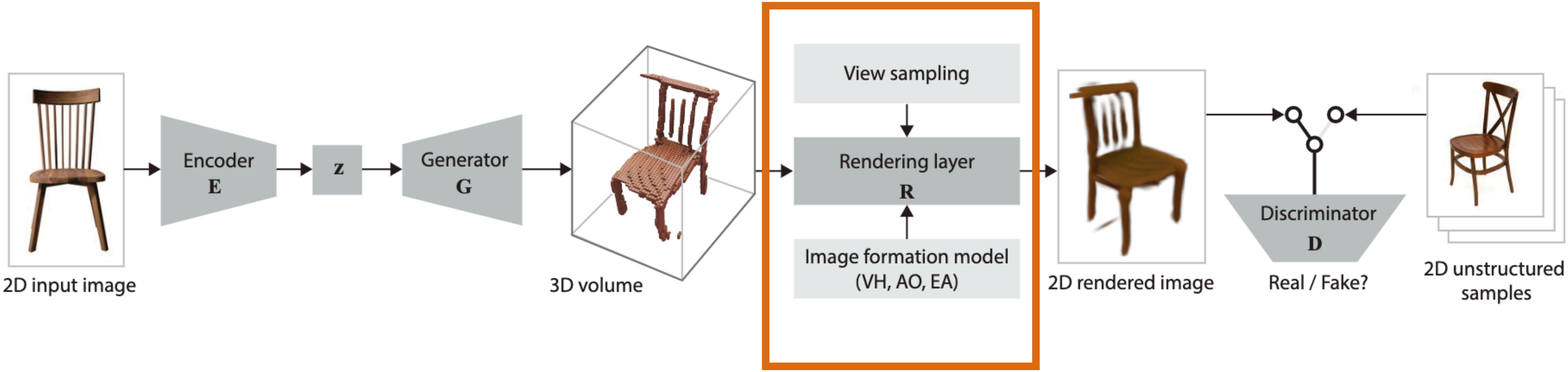
$$\rho_{\mathrm{AO}}(\mathbf{v}) = 1 - \prod_i (1 - v_i)$$

$$\rho_{\mathrm{EA}}(\mathbf{v}) = \sum_{i=1}^{n_z} \underbrace{\left(1 - \prod_{j=1}^{i} (1 - v_{\mathrm{a},j})\right)}_{\text{Transmission } t_i \text{ to voxel } i} v_{\mathrm{e},i}$$

PlatonicGAN

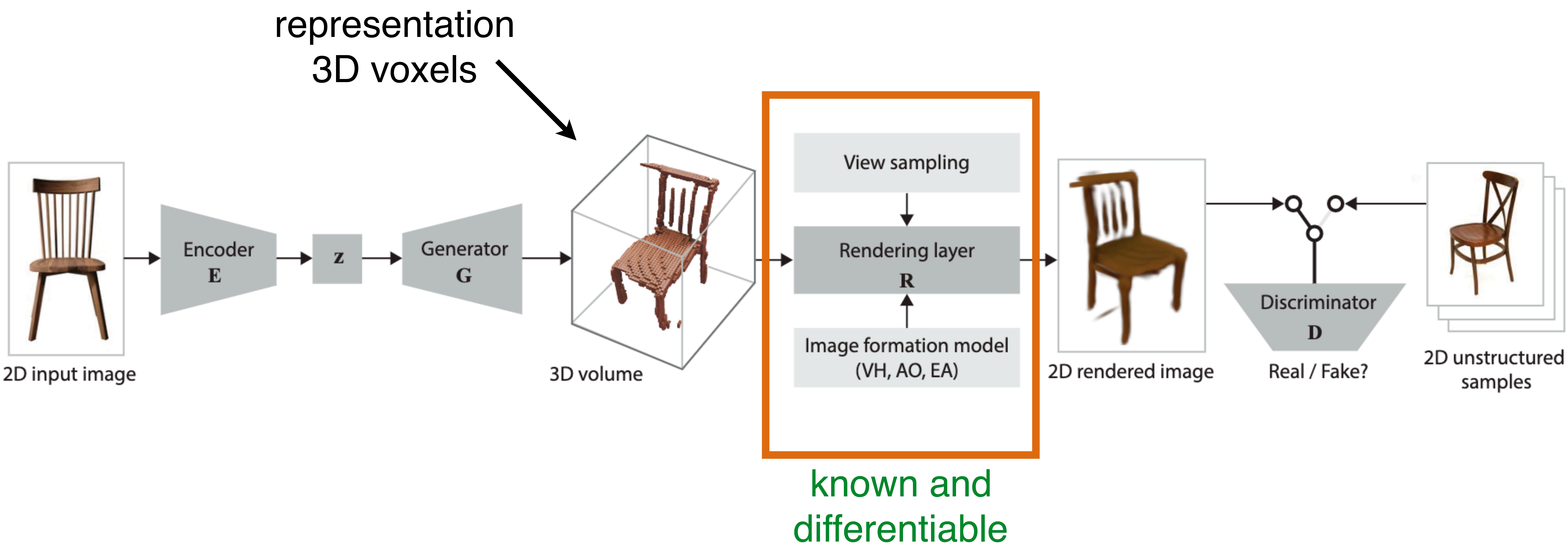


PlatonicGAN



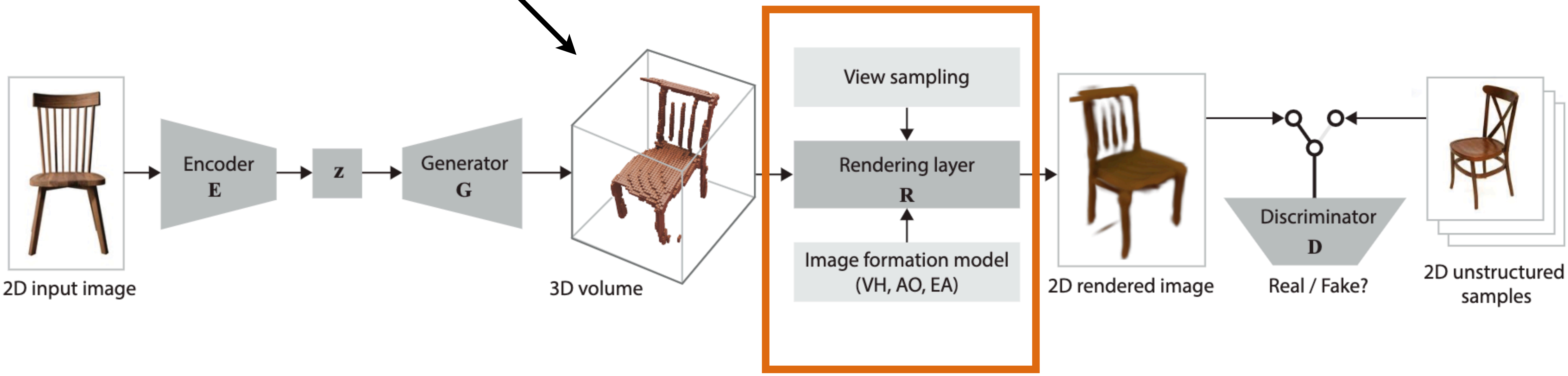
known and
differentiable

PlatonicGAN



PlatonicGAN

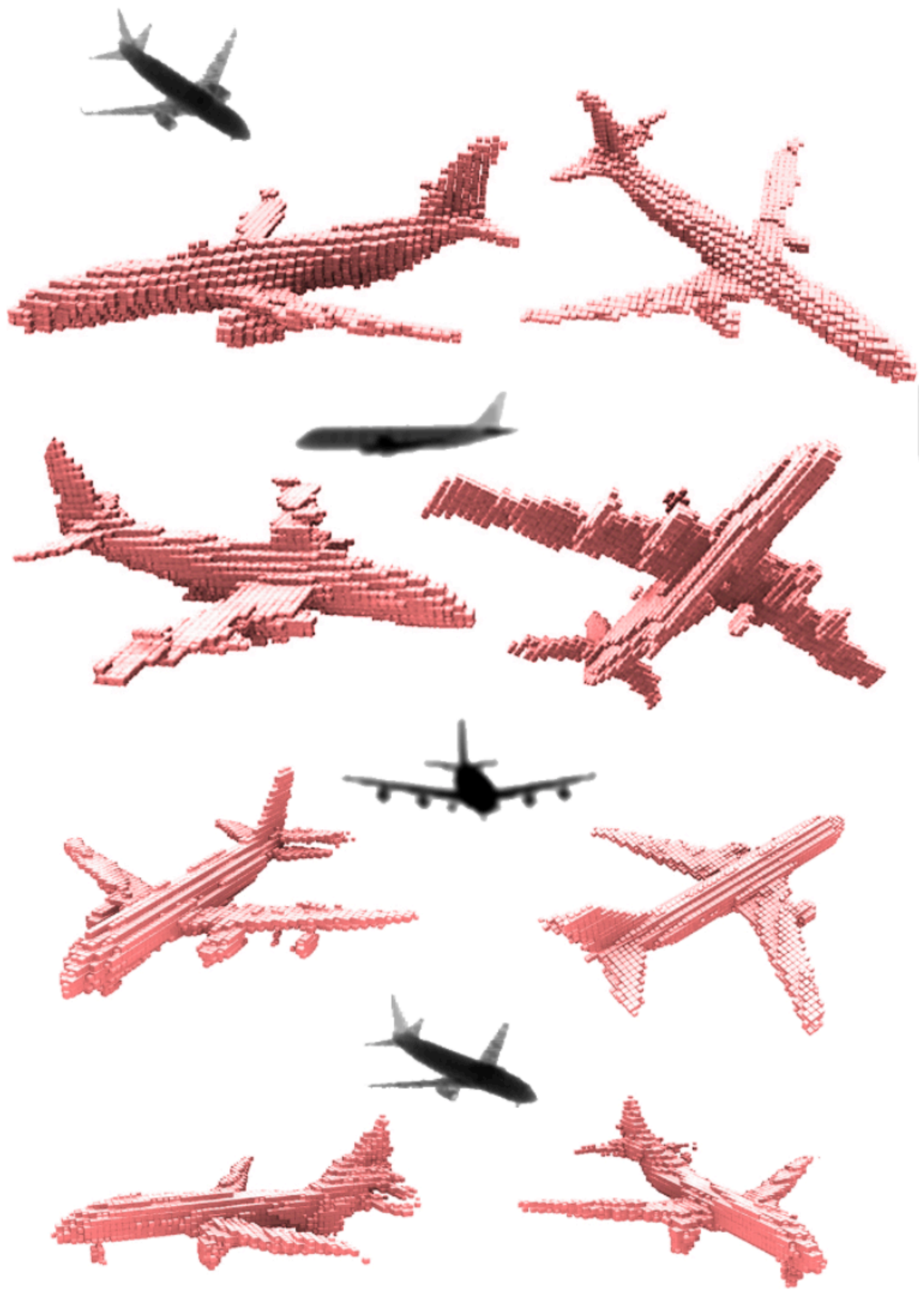
representation
3D voxels



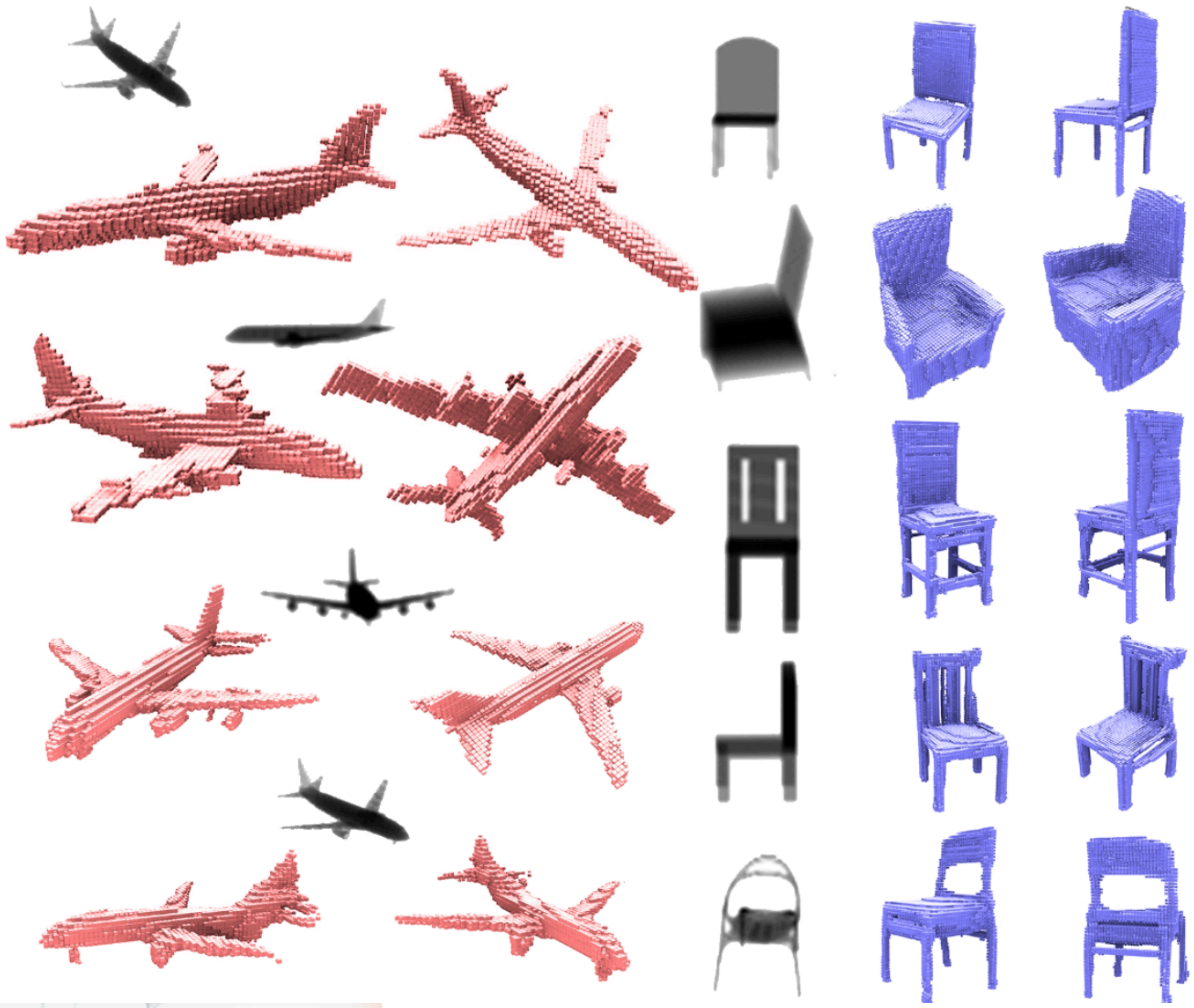
known and differentiable

$$\min_{\Psi} \max_{\Theta,\Phi} \; c_{\text{Disc}}(\Psi) + c_{\text{Gen}}(\Theta, \Phi) + \lambda c_{\text{Rec}}(\Theta, \Phi)$$

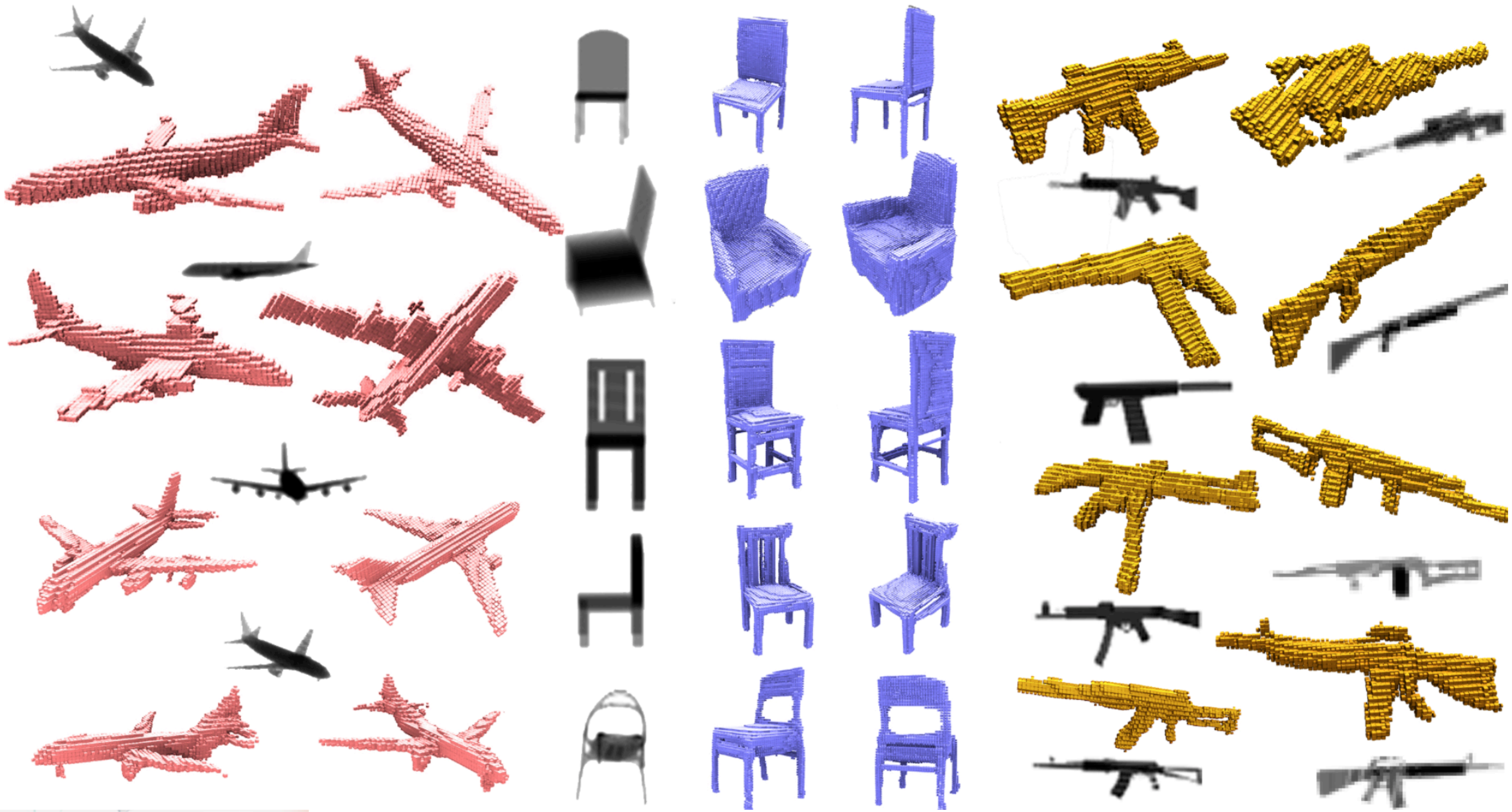
Results: Without any access to 3D Data for Training



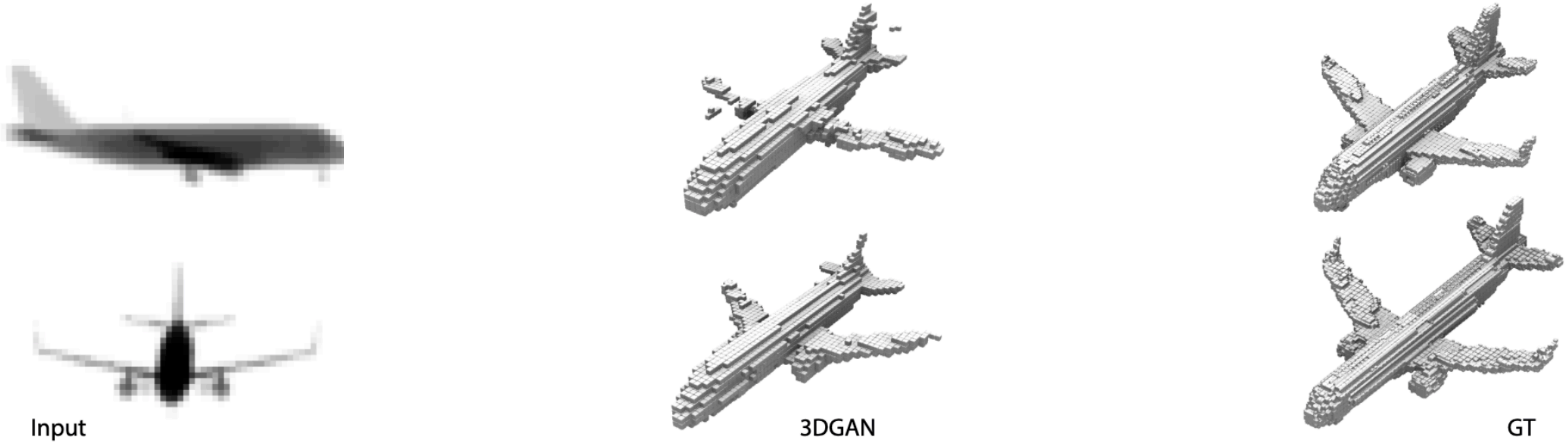
Results: Without any access to 3D Data for Training



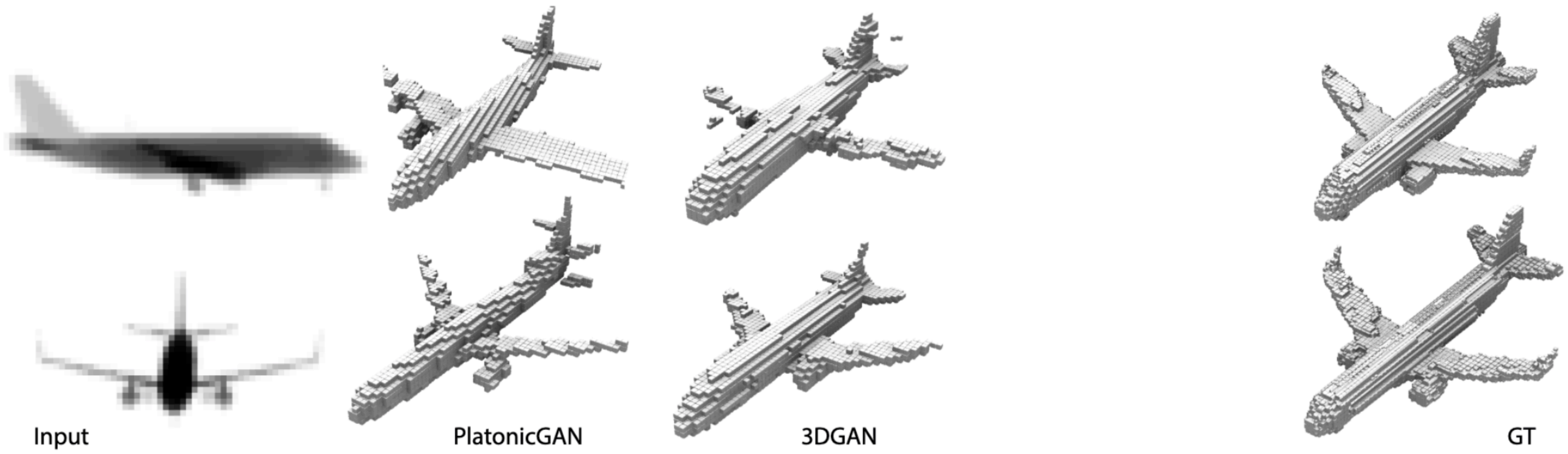
Results: Without any access to 3D Data for Training



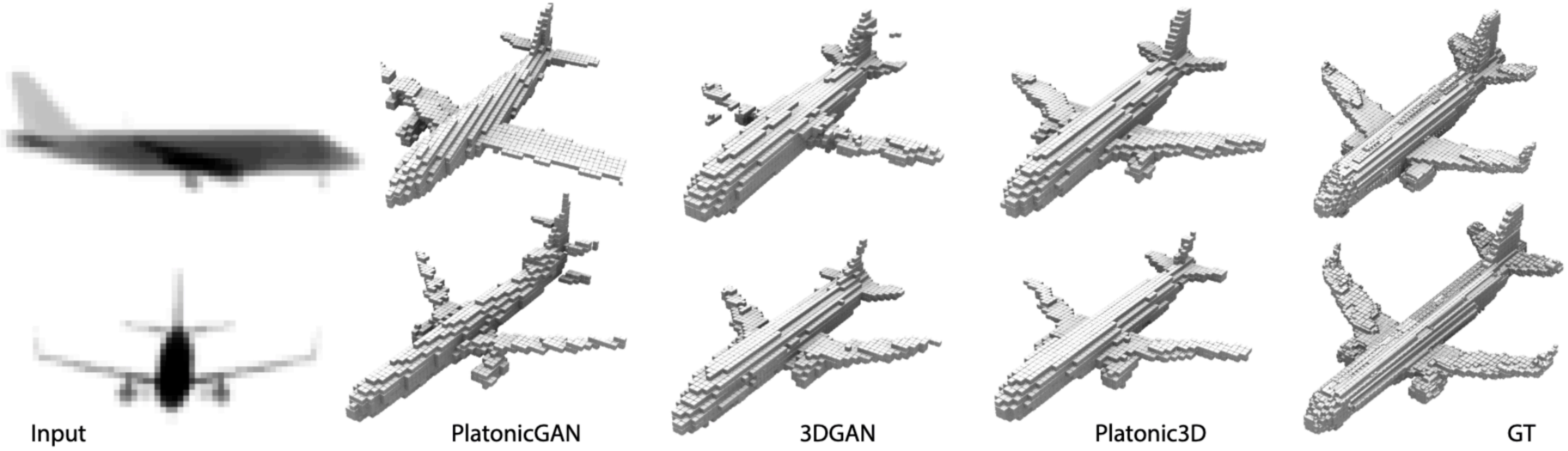
Effect of Access to 3D Information



Effect of Access to 3D Information



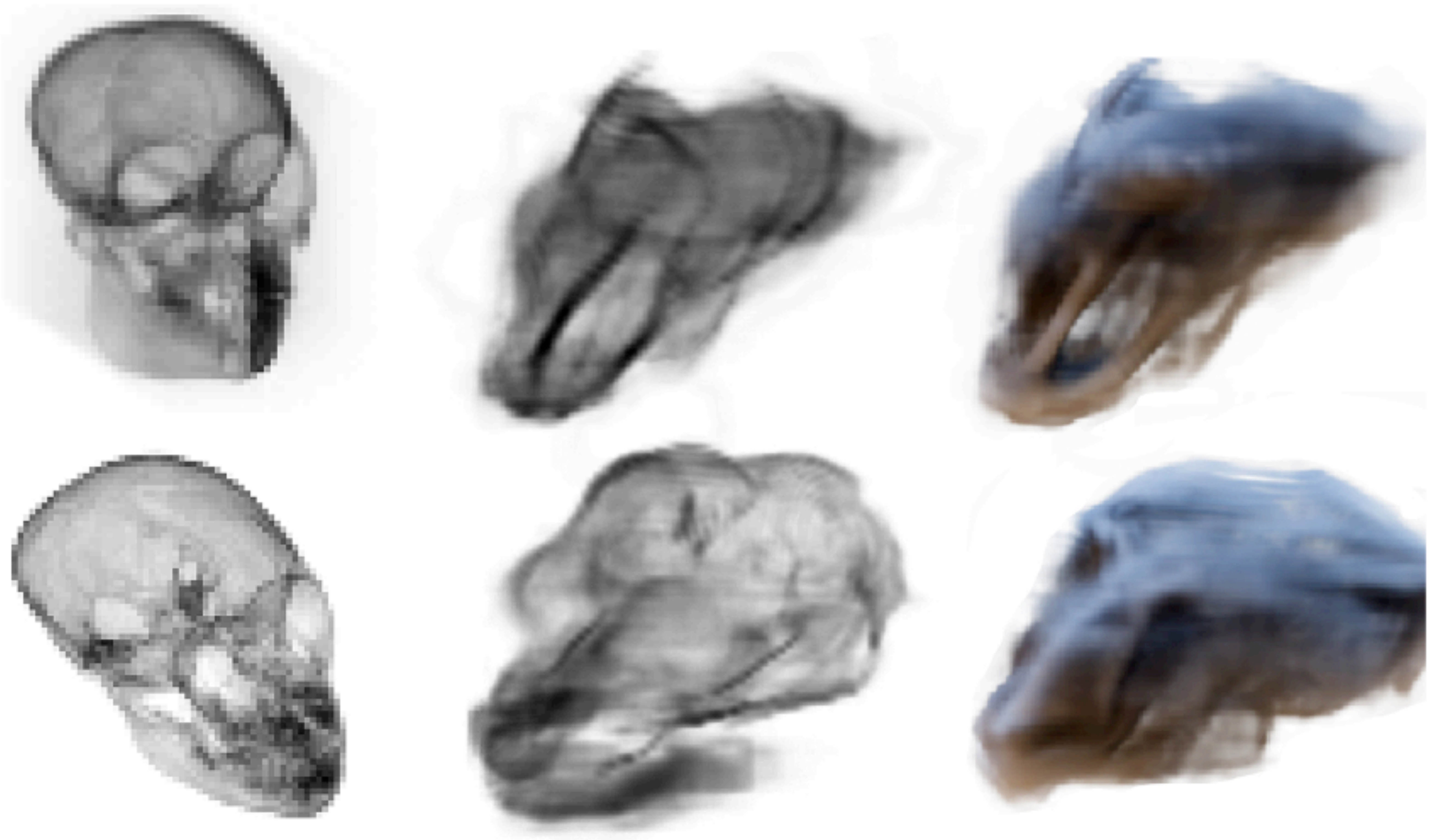
Effect of Access to 3D Information



In Absence of 3D Training Data

In Absence of 3D Training Data

Novel view from Xray data



a) 2D input x-ray b) X-ray c) Volume render

Representation for 3D

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

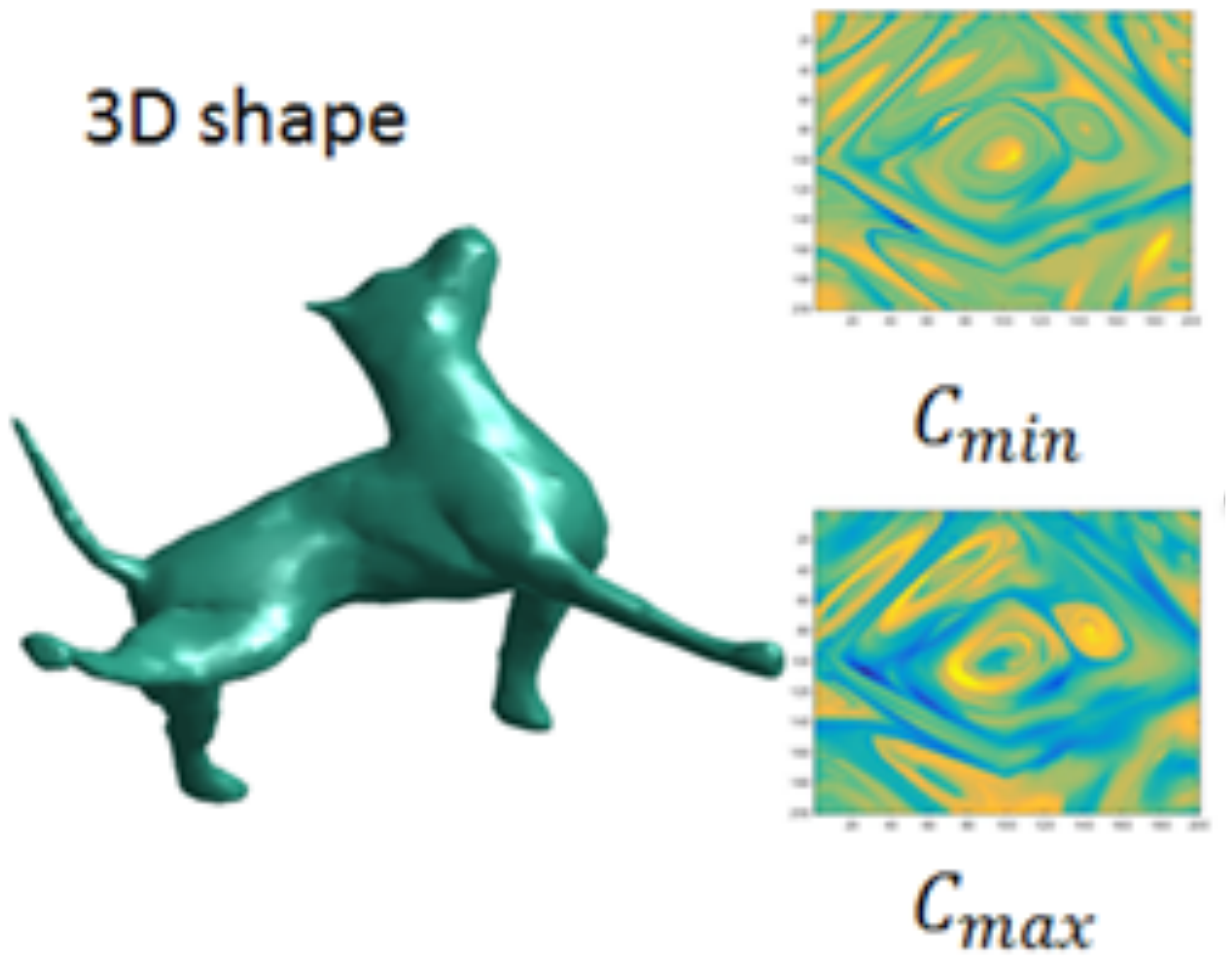
Representation for 3D

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

Local/Global Parameterizations

Local/Global Parameterizations

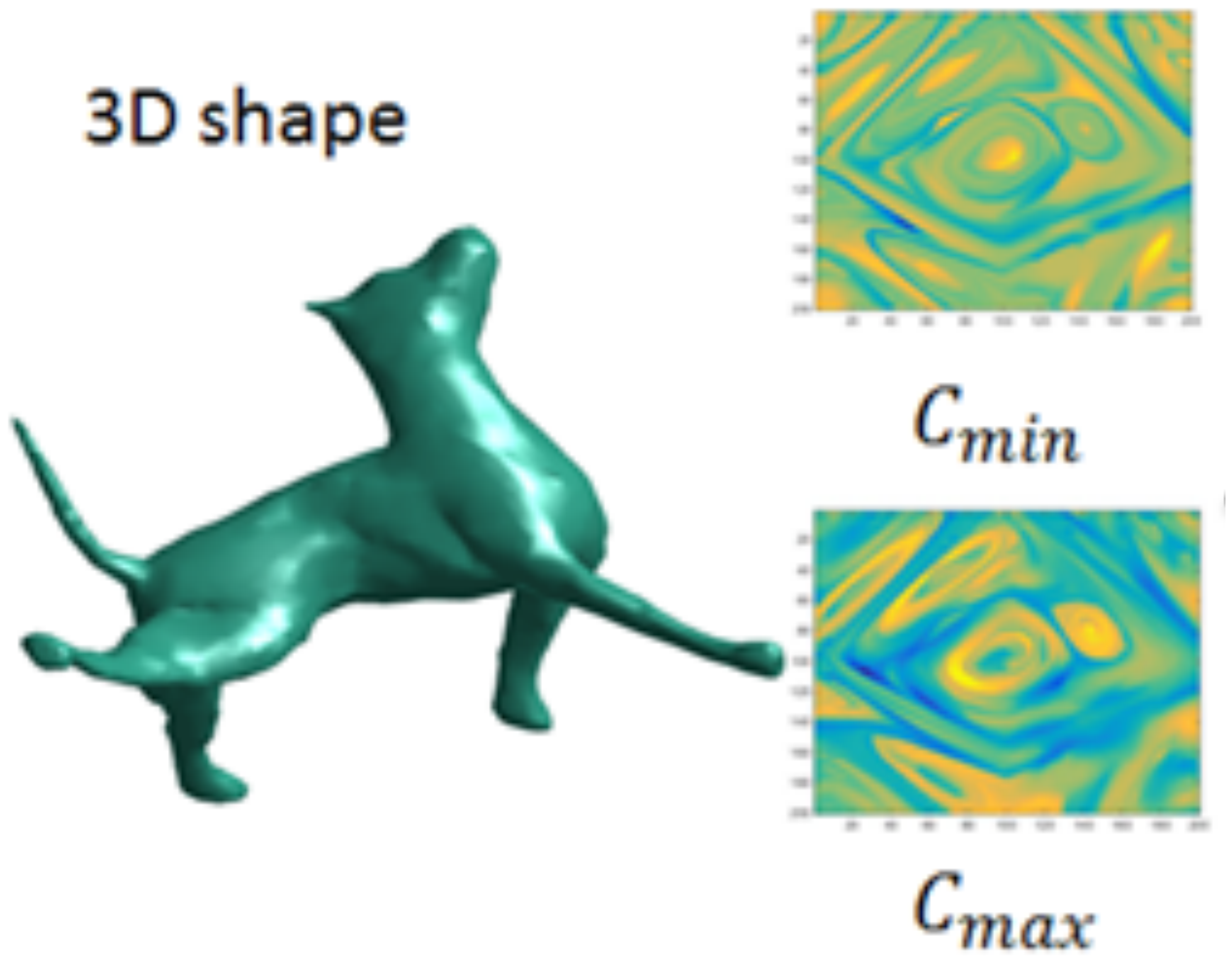
Geometry Image



[Sinha et al. 2016]

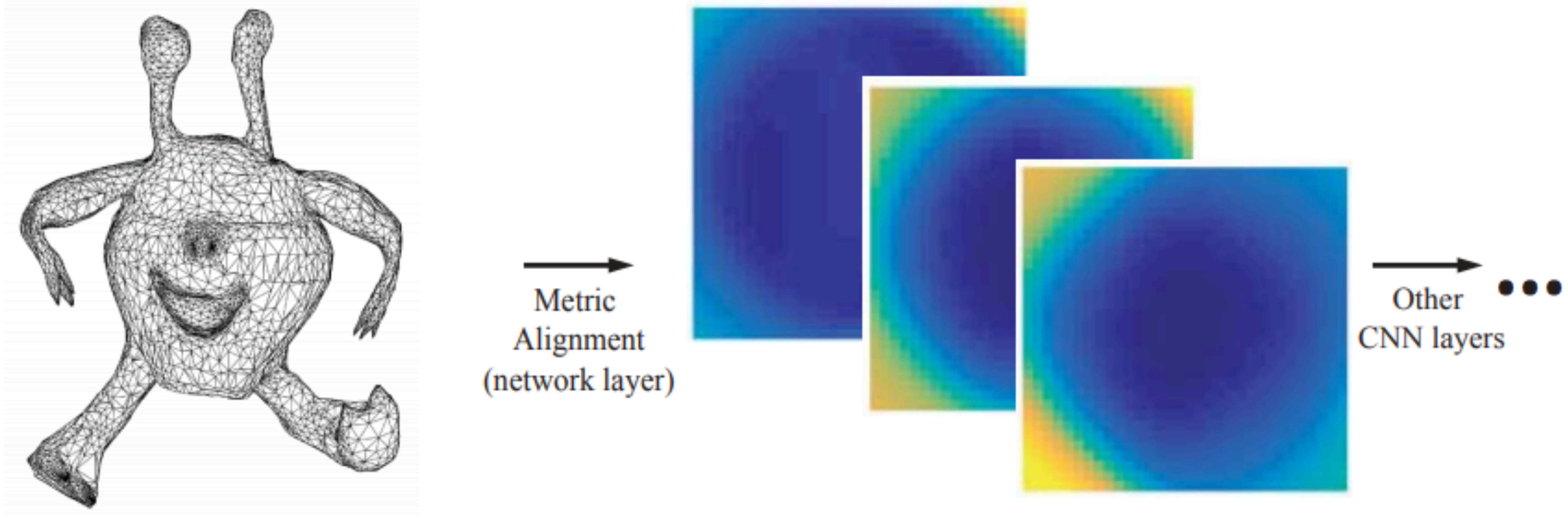
Local/Global Parameterizations

Geometry Image



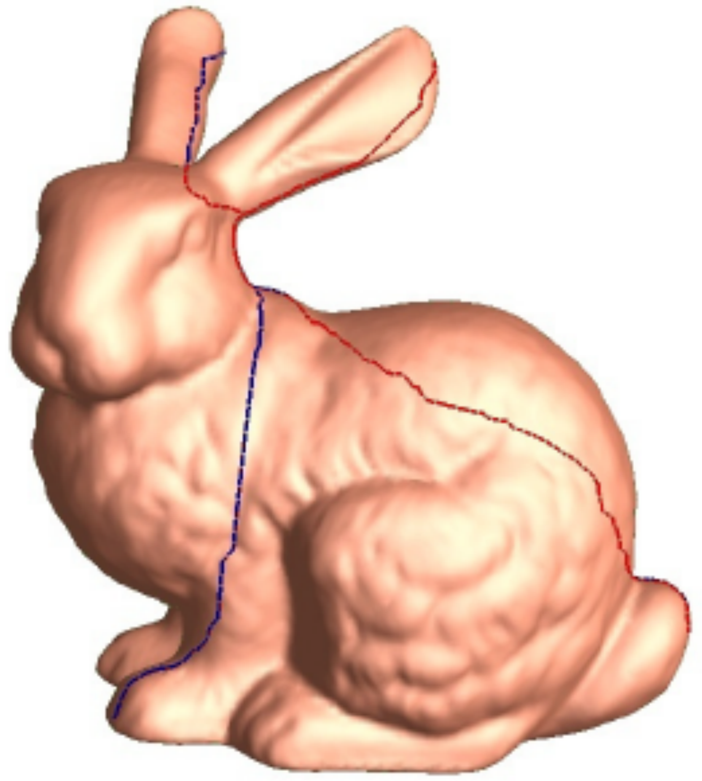
[Sinha et al. 2016]

Metric Alignment (GWCNN)

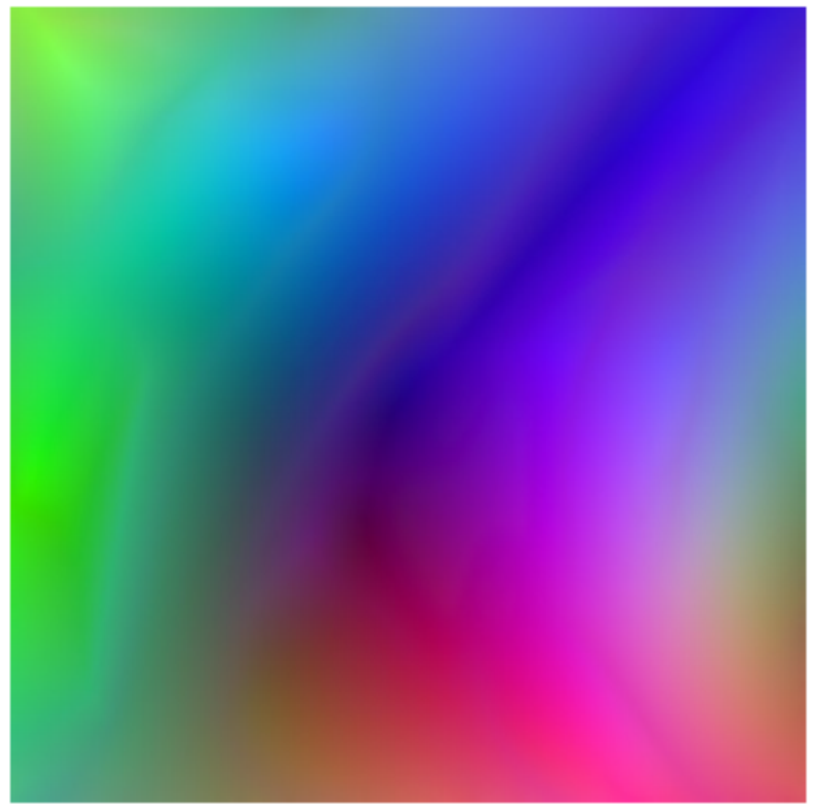


[Ezuz et al. 2017]

Shape Surfaces using Geometry Images

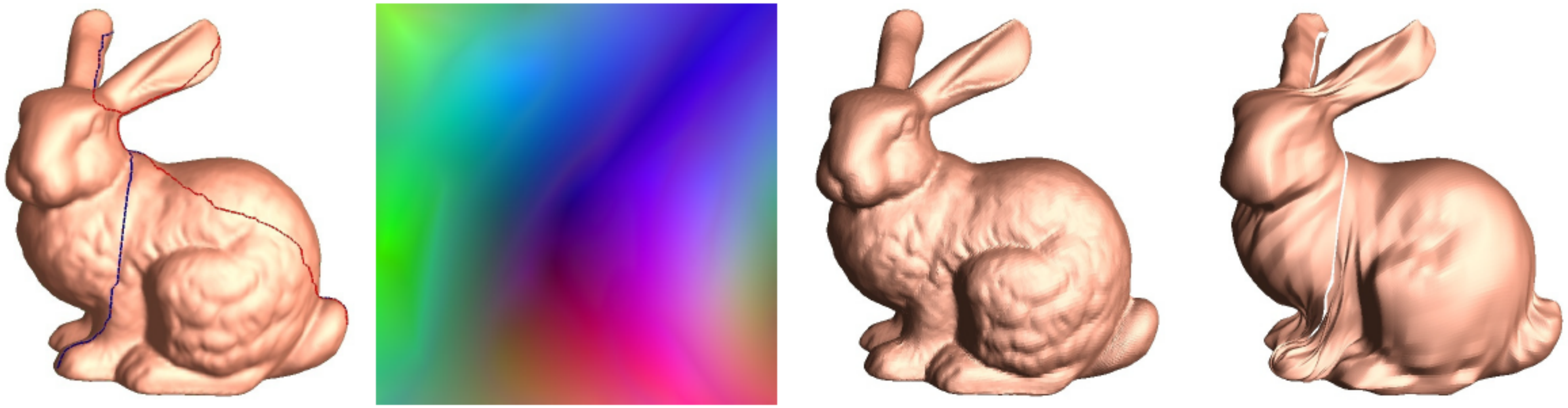


(a) Original mesh with cut
70K faces; genus 0



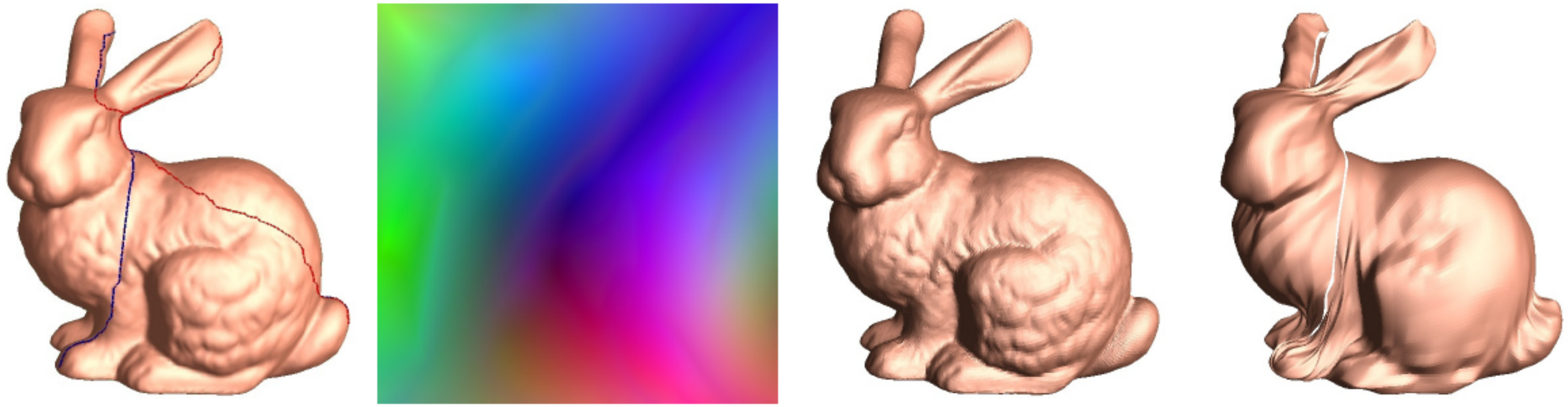
(b) Geometry image 257×257
(b∗) Compr. to 1.5KB (not shown)

Shape Surfaces using Geometry Images

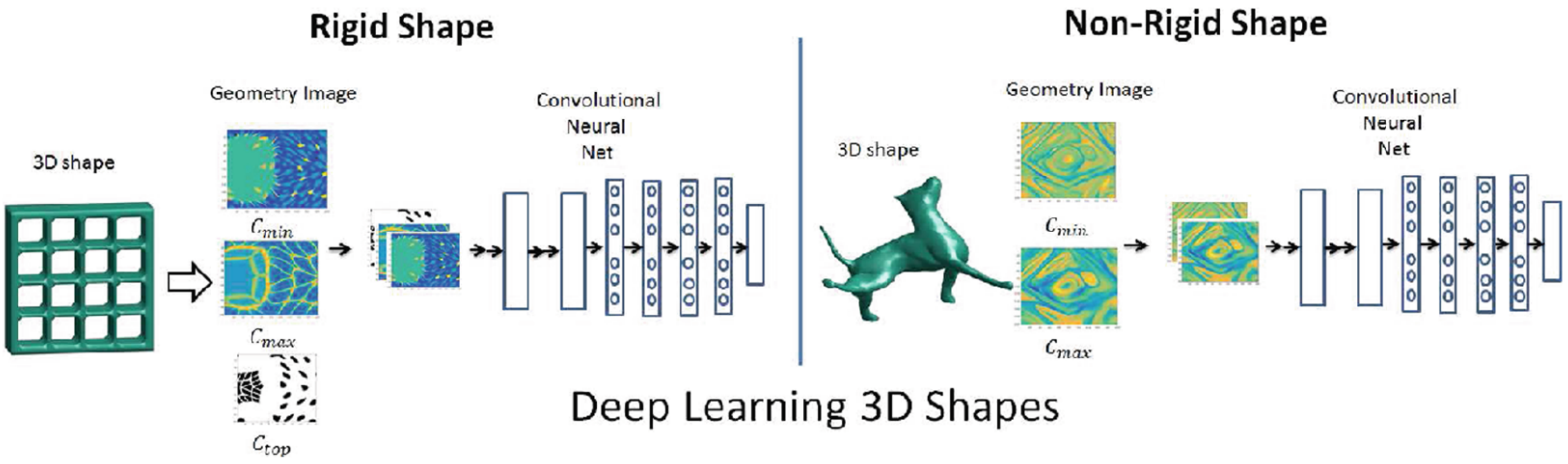


(a) Original mesh with cut 70K faces; genus 0 (b) Geometry image 257×257 (b∗) Compr. to 1.5KB (not shown) (c) Geometry reconstructed entirely from b (d) Geometry reconstructed entirely from b∗

Shape Surfaces using Geometry Images

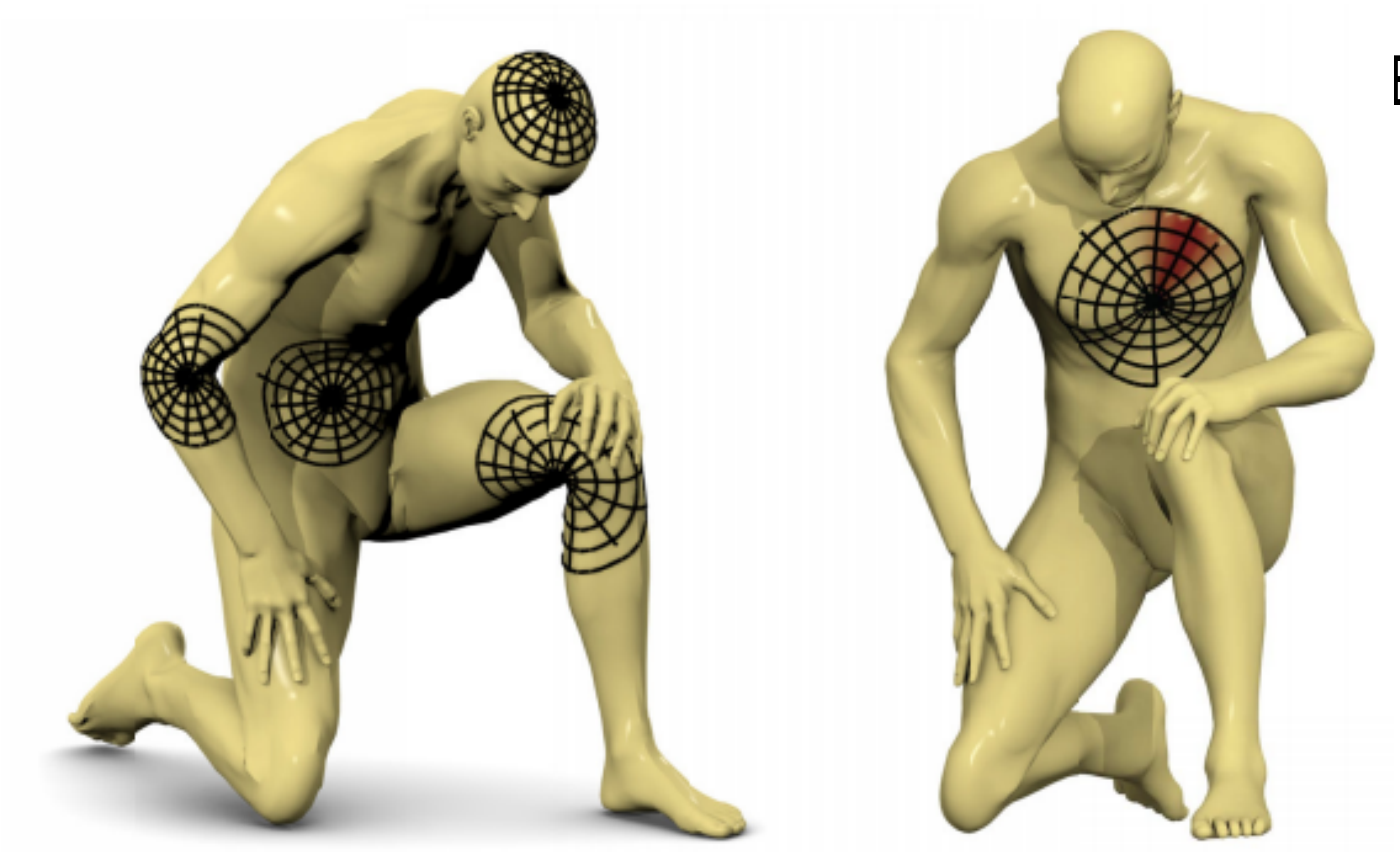


(a) Original mesh with cut 70K faces; genus 0
(b) Geometry image 257×257 (b∗) Compr. to 1.5KB (not shown)
(c) Geometry reconstructed entirely from b
(d) Geometry reconstructed entirely from b∗



Deep Learning 3D Shapes

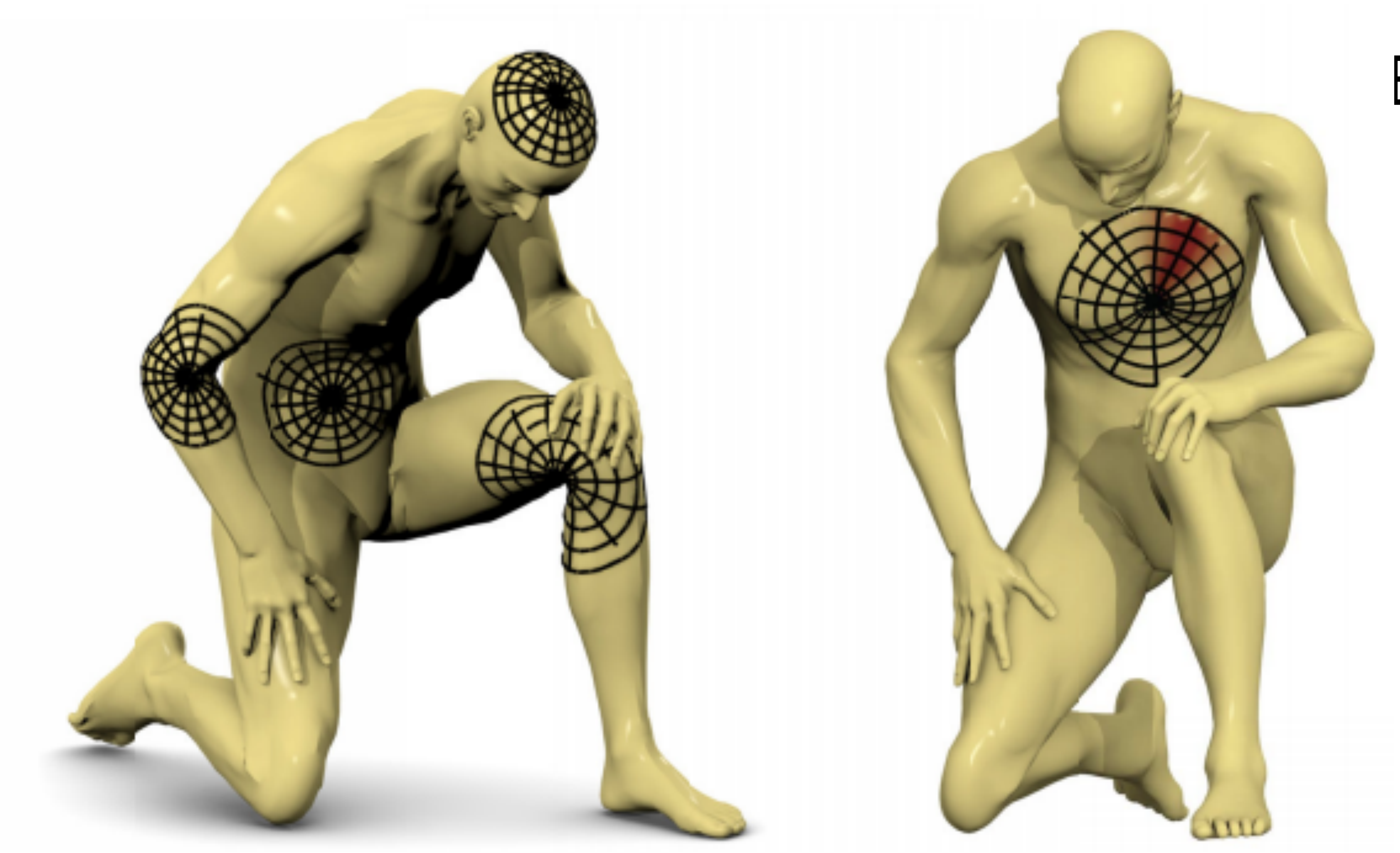
Using Geodesic Patches: GCNN



ETWORKS ON RIEMANNIAN
D VERSION]

[Masci et al. 2015]

Using Geodesic Patches: GCNN

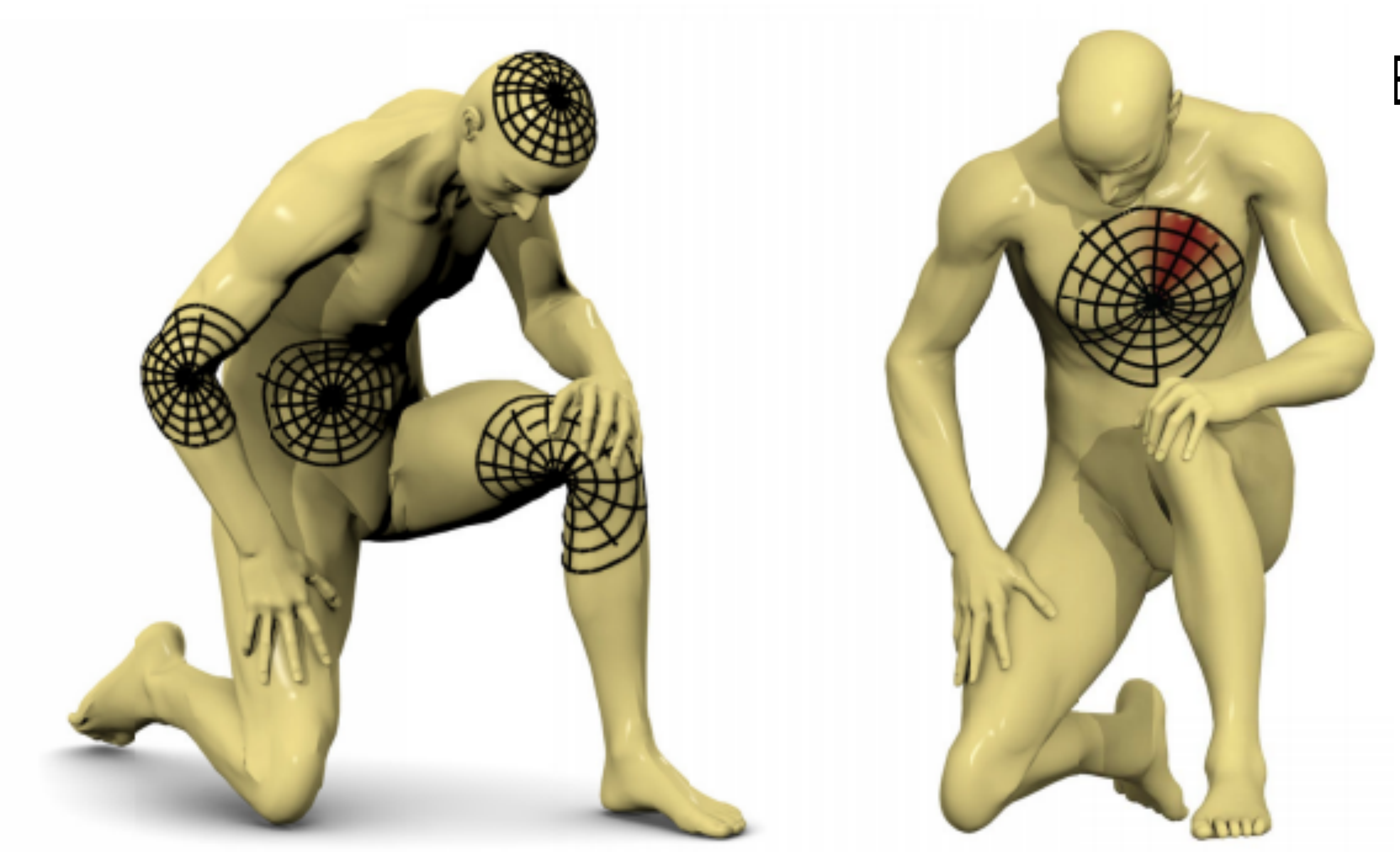


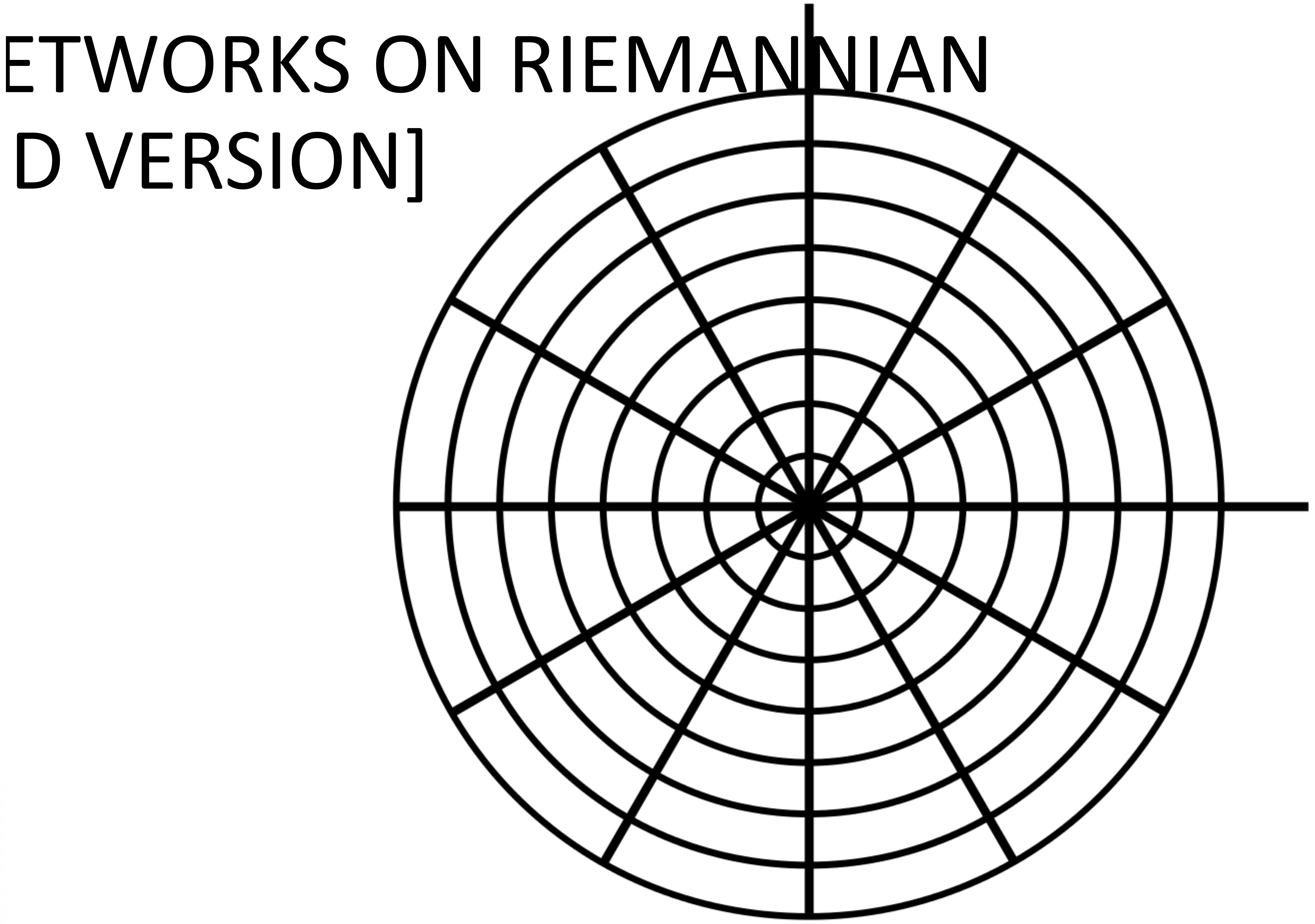
ETWORKS ON RIEMANNIAN
D VERSION]

$$(f \star a)(x) := \sum_{\theta,r} a(\theta + \Delta\theta, r)(D(x)f)(r,\theta)$$

[Masci et al. 2015]

Using Geodesic Patches: GCNN

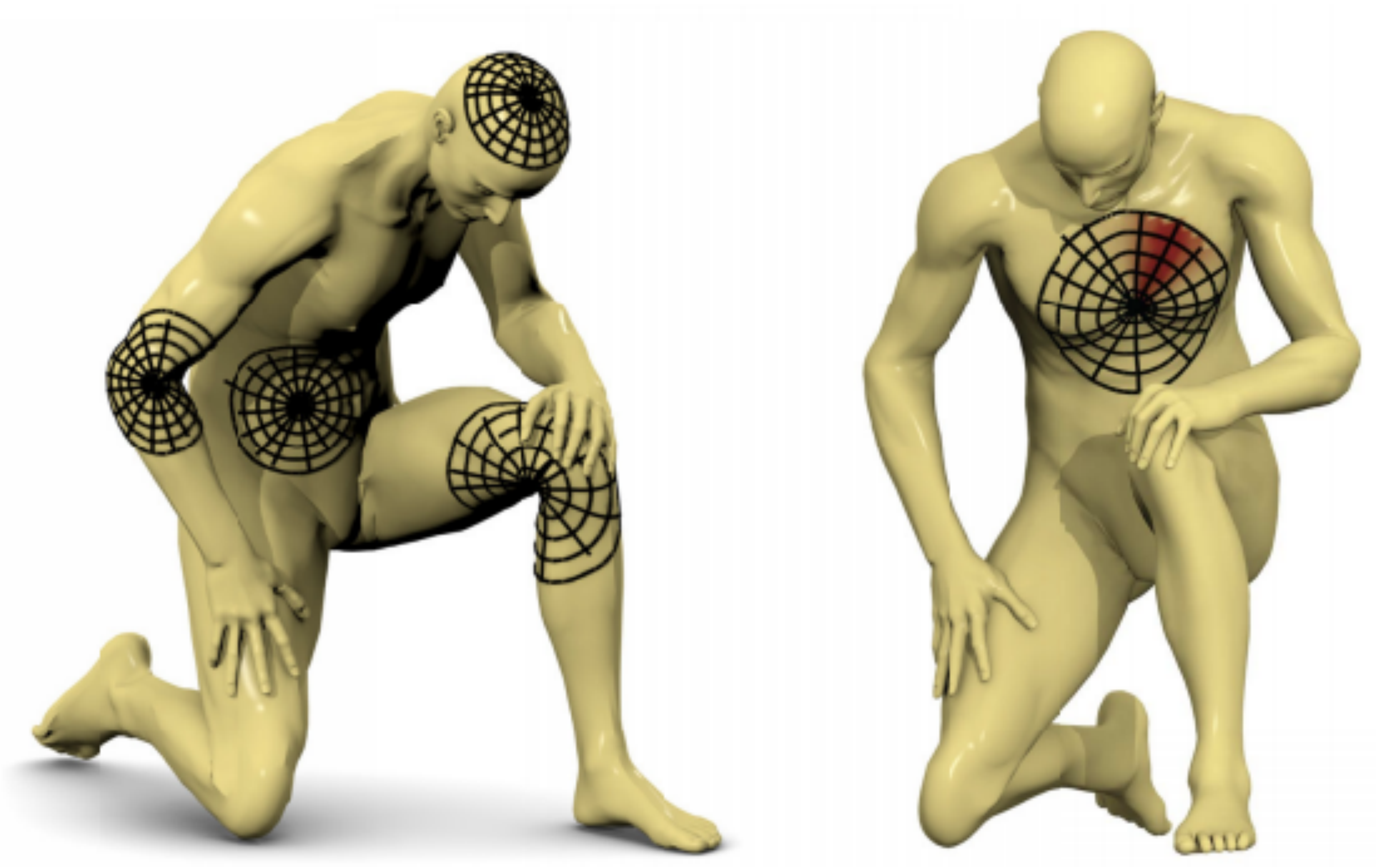




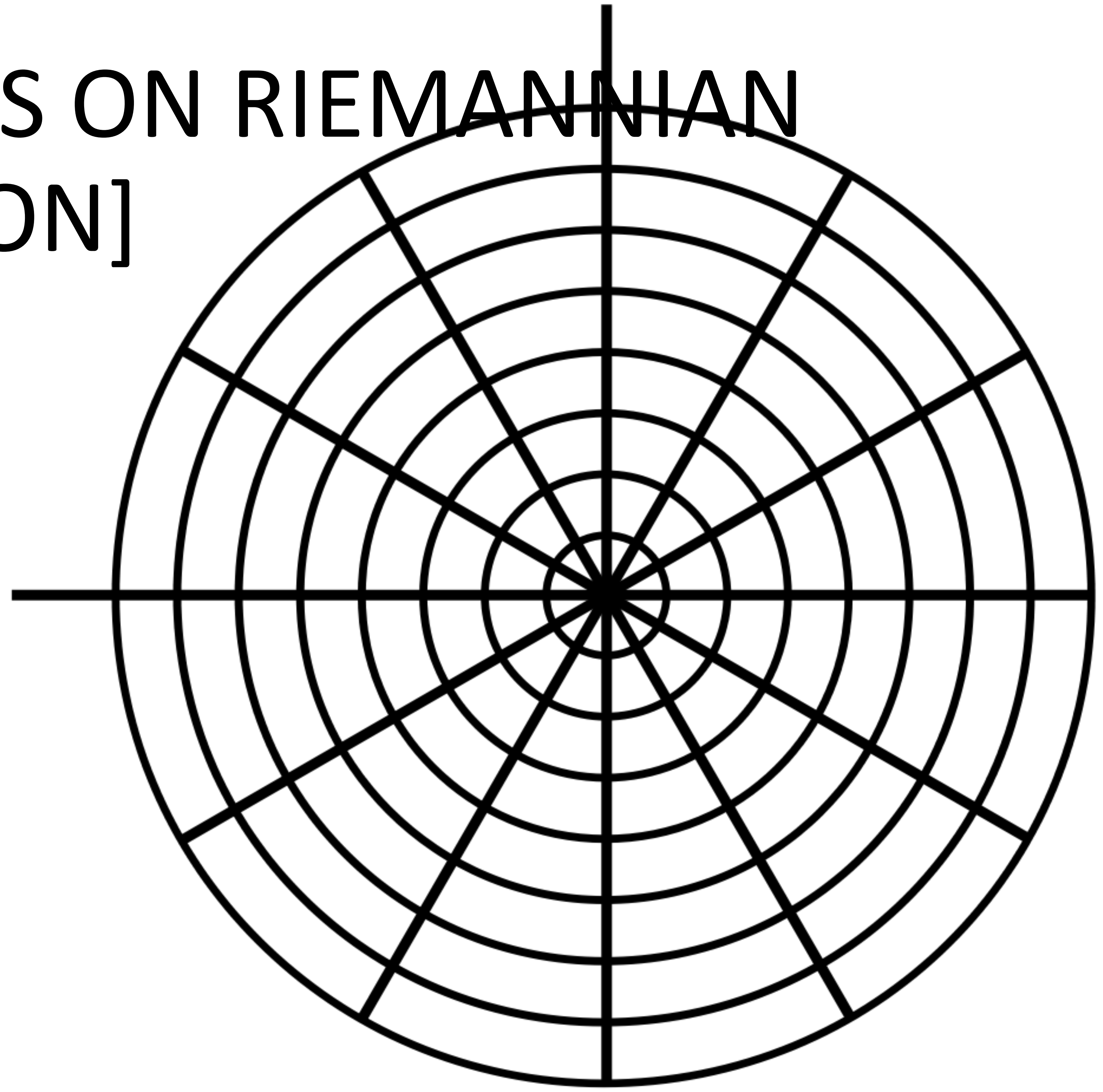
$$(f \star a)(x) := \sum_{\theta, r} a(\theta + \Delta\theta, r)(D(x)f)(r, \theta)$$

[Masci et al. 2015]

Using Geodesic Patches: GCNN



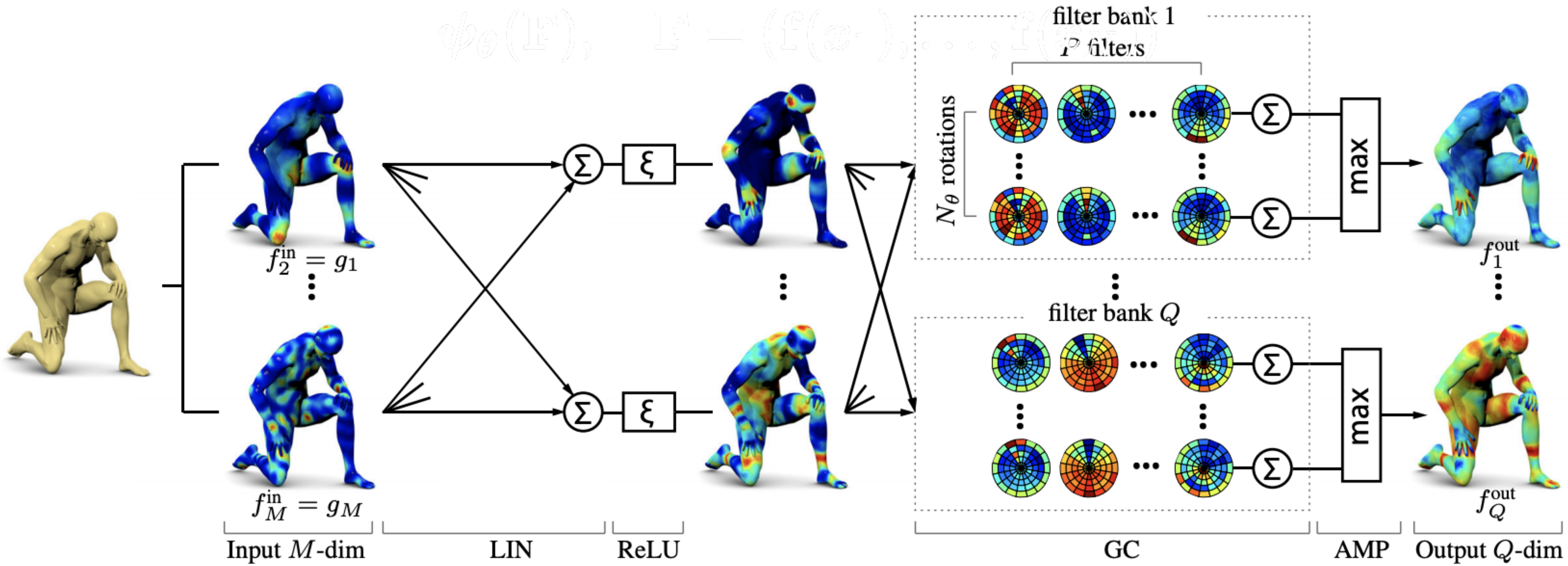
ETWORKS ON RIEMANNIAN
D VERSION]



$$(f \star a)(x) := \sum_{\theta, r} a(\theta + \Delta\theta, r)(D(x)f)(r, \theta)$$

[Masci et al. 2015]

GCNN Architecture



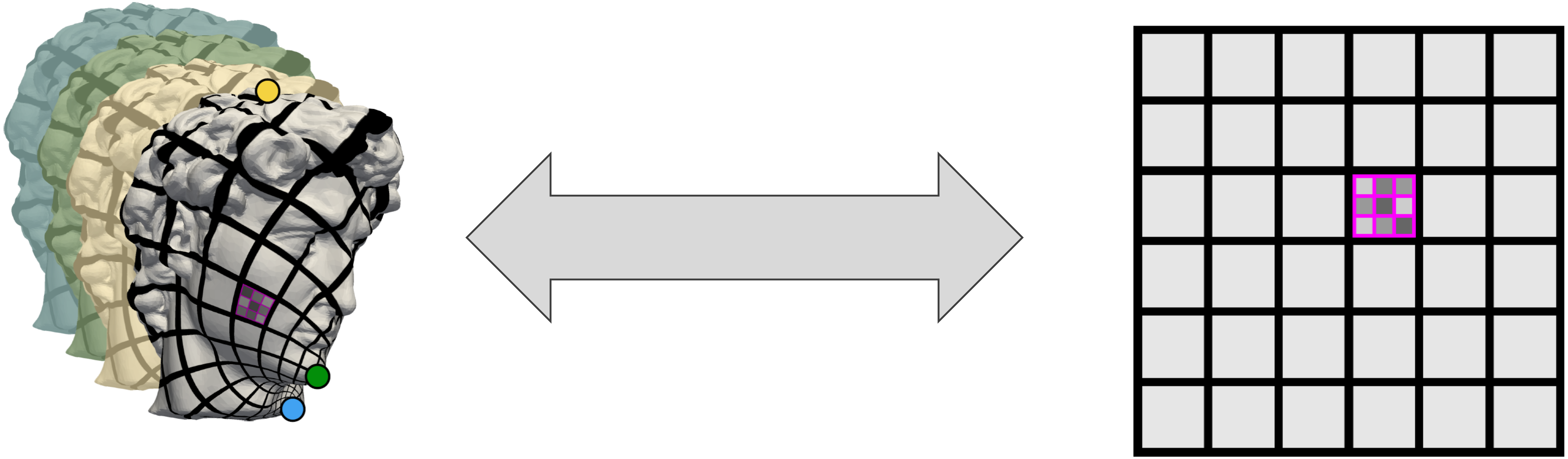
Parameterization for Surface Analysis

map 3D surface to 2D domain

[Maron et al. 2017]

Parameterization for Surface Analysis

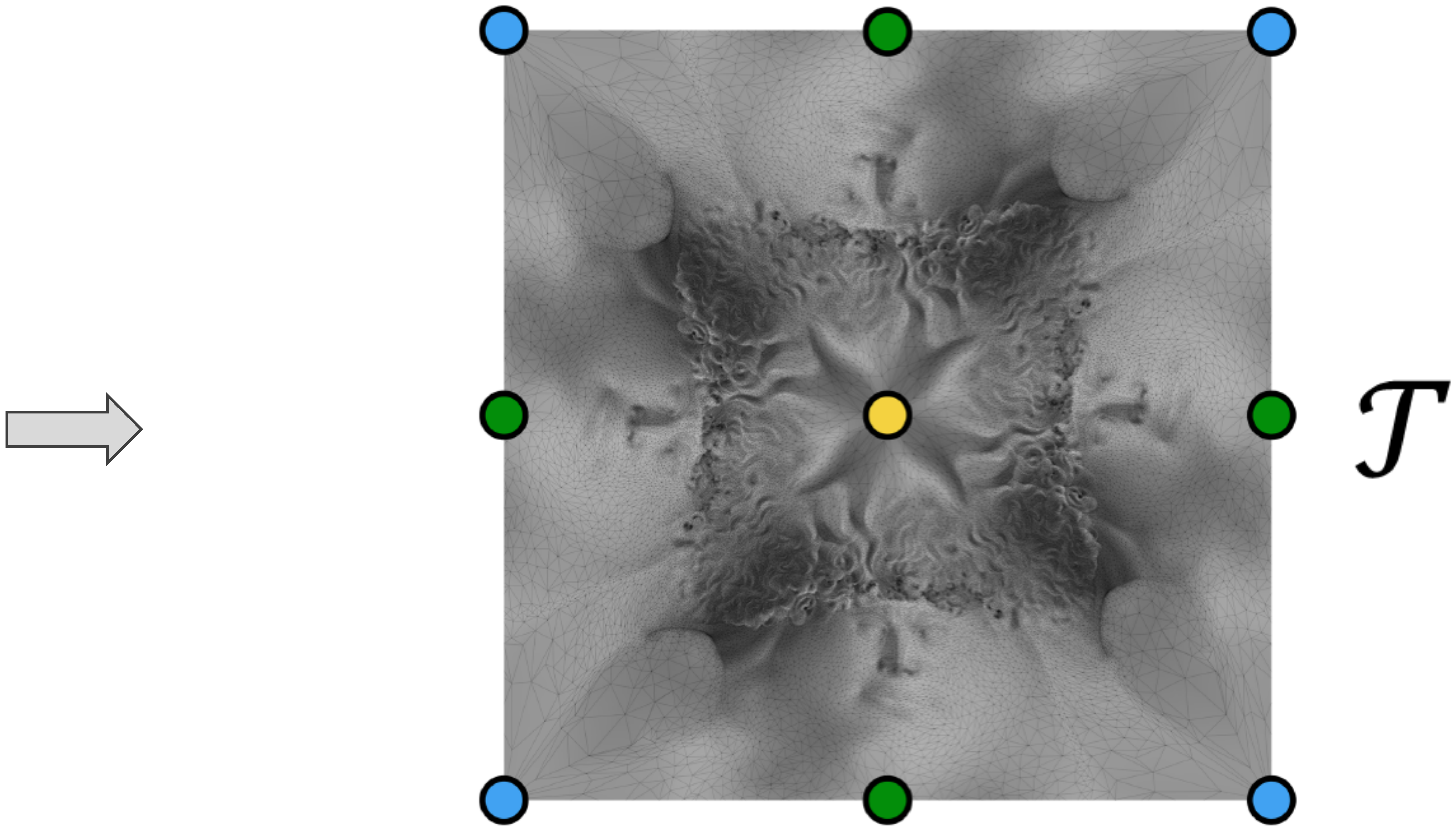
map 3D surface to 2D domain



[Maron et al. 2017]

Parameterization for Surface Analysis

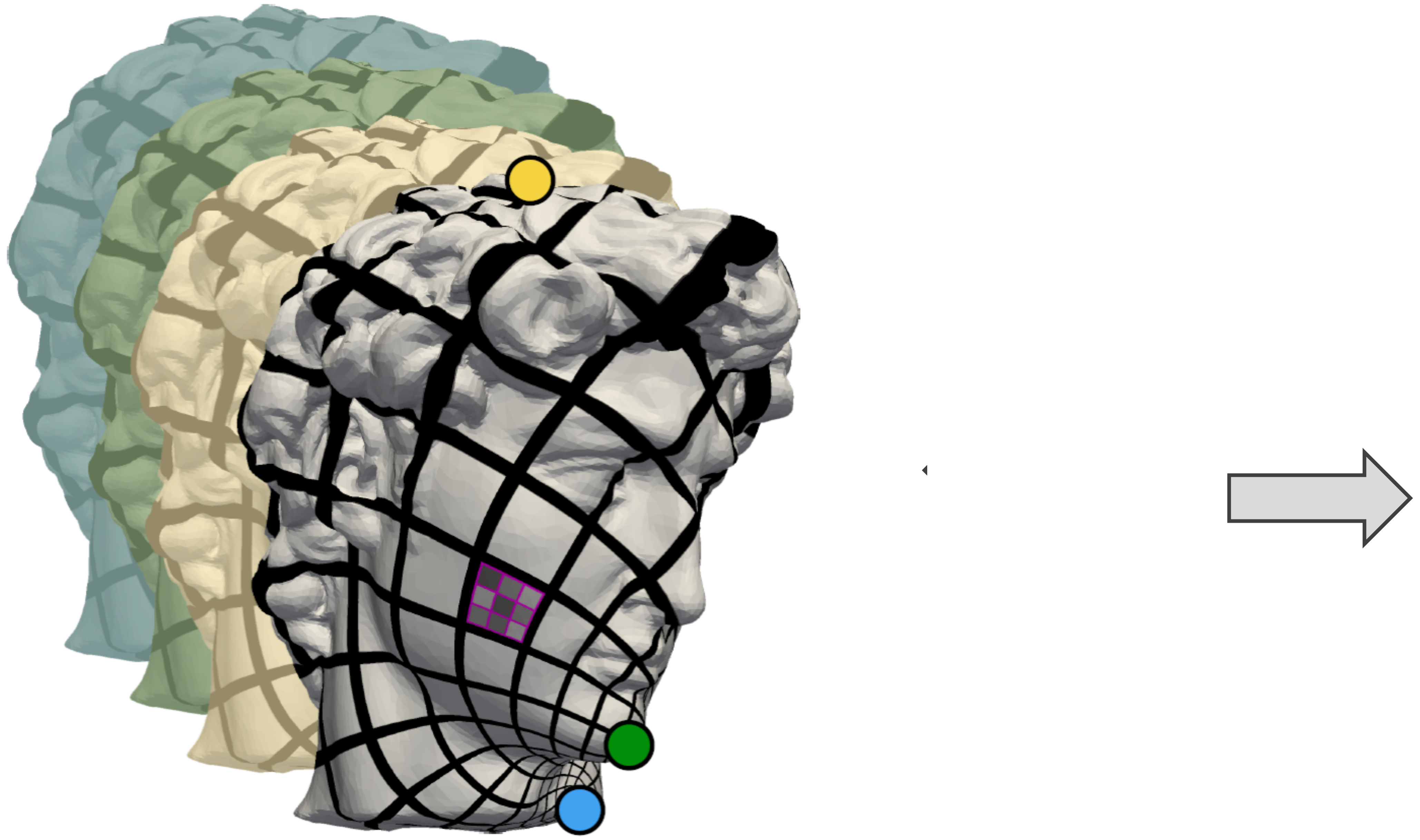
map 3D surface to 2D domain

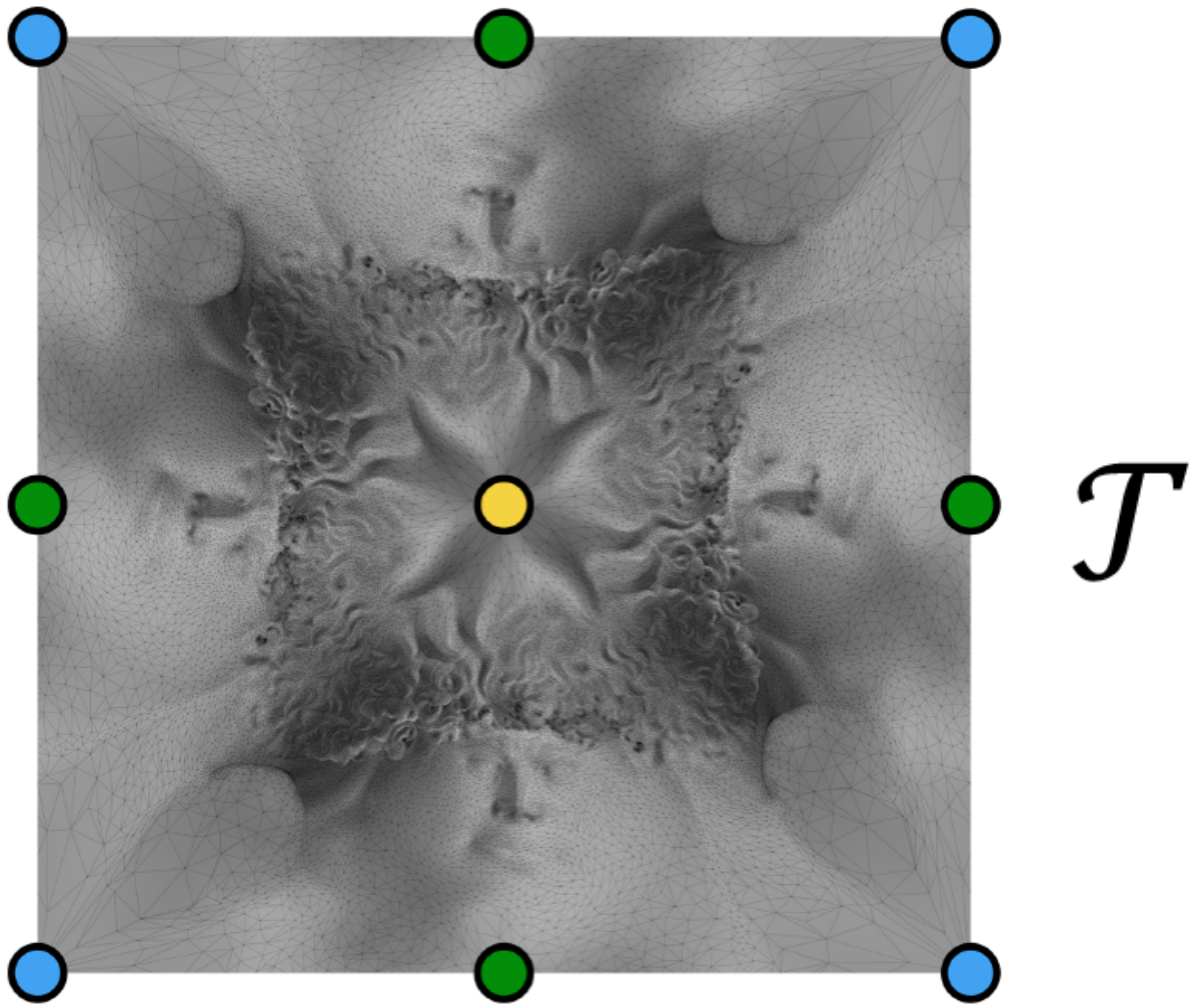


[Maron et al. 2017]

Parameterization for Surface Analysis

map 3D surface to 2D domain

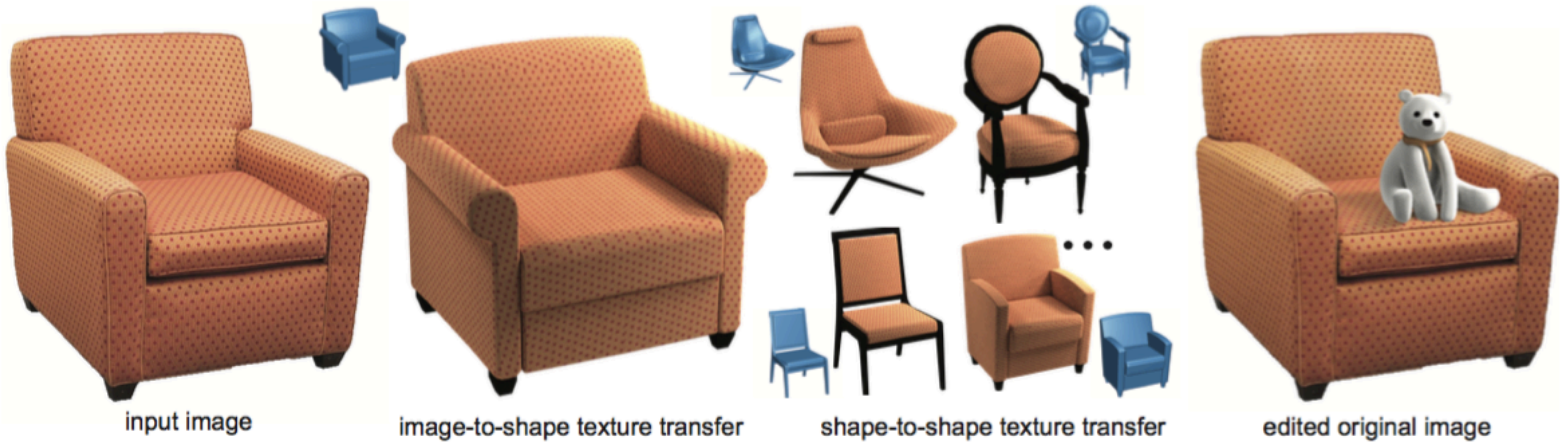




[Maron et al. 2017]

Texture Transfer (Parameterization + Alignment)

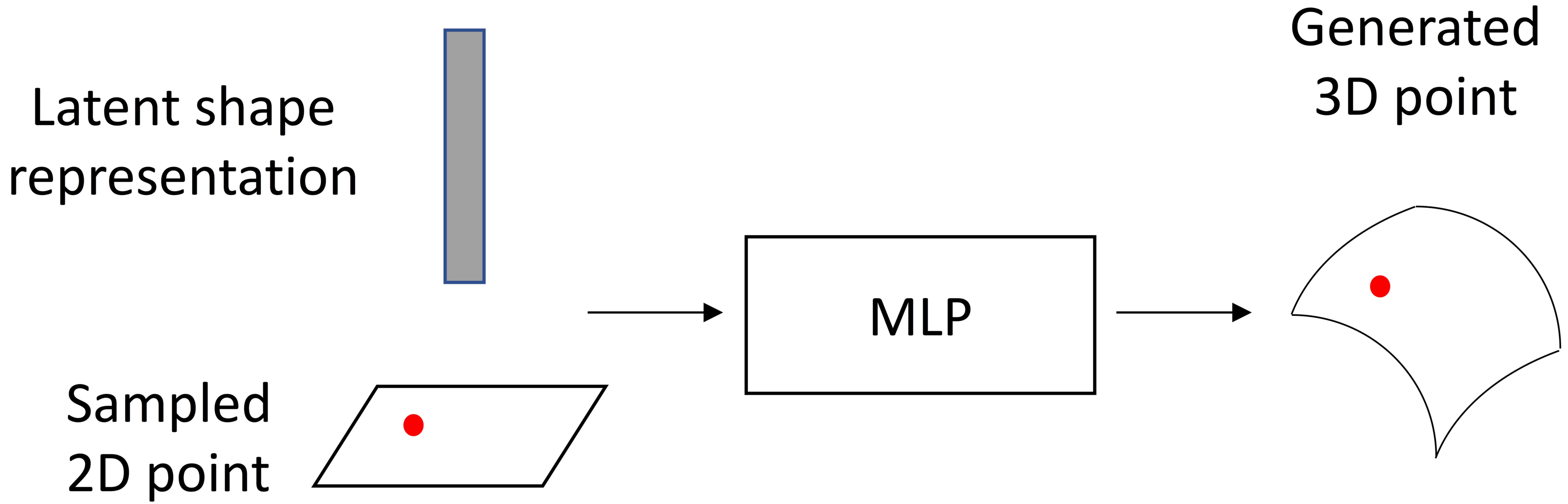
- Condition decoded points on 2D patches



[Wang, Su, Huang, Huang, Guibas, Mitra, Siggraph Asia 2016]

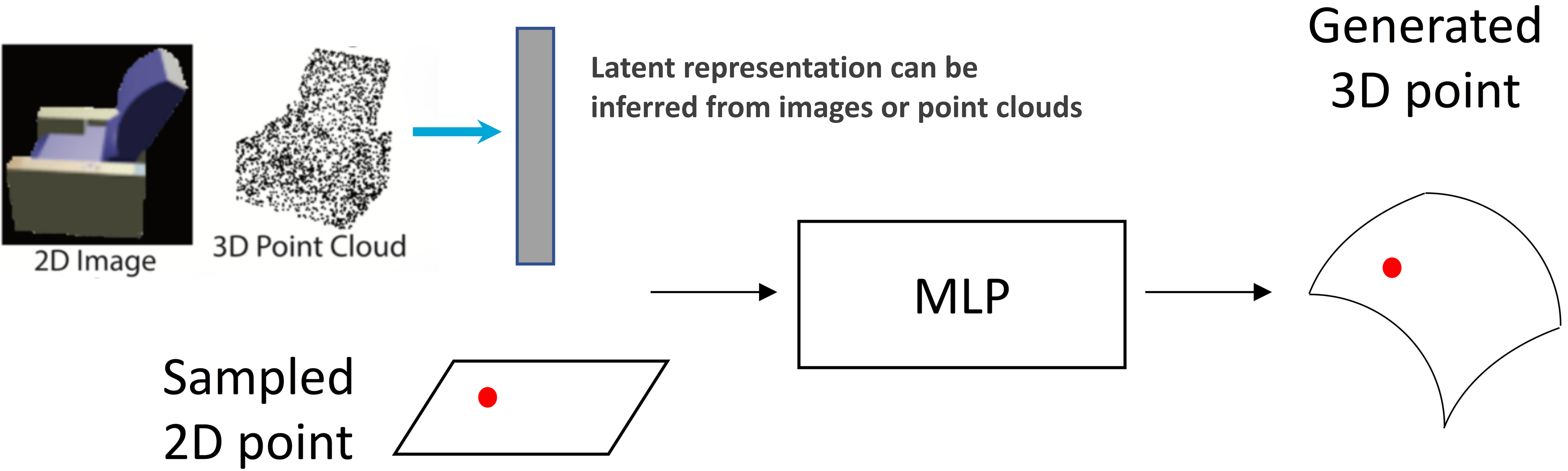
AtlasNet for Surface Generation

condition decoded points on 2D patches



[Groueix et al. 2018]

AtlasNet for Surface Generation



[Groueix et al. 2018]

AtlasNet for Surface Generation

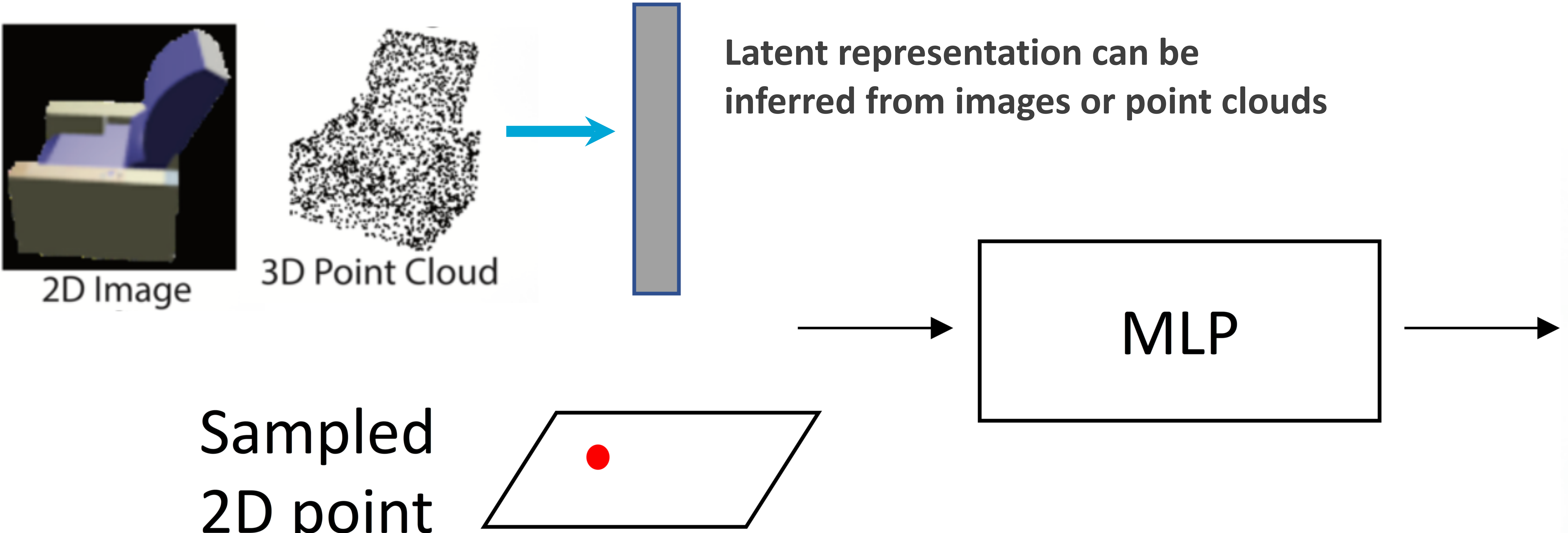
- Condition decoded points on 2D patches

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

AtlasNet for Surface Generation

- Condition decoded points on 2D patches

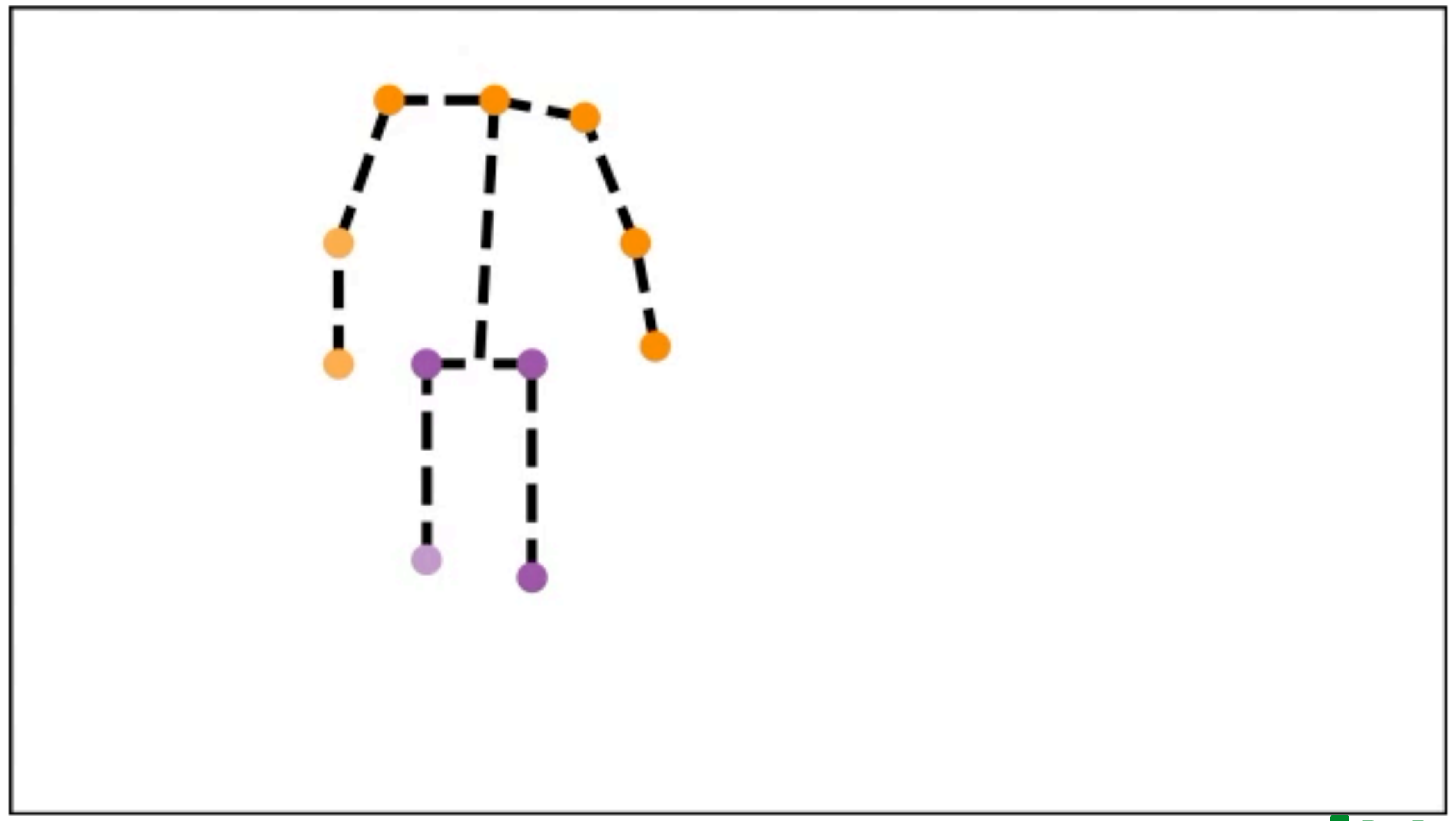
texture coordinates come for free!!



3. Data-driven Human-Object Interaction Capture

input video:

initial skeletal estimate:

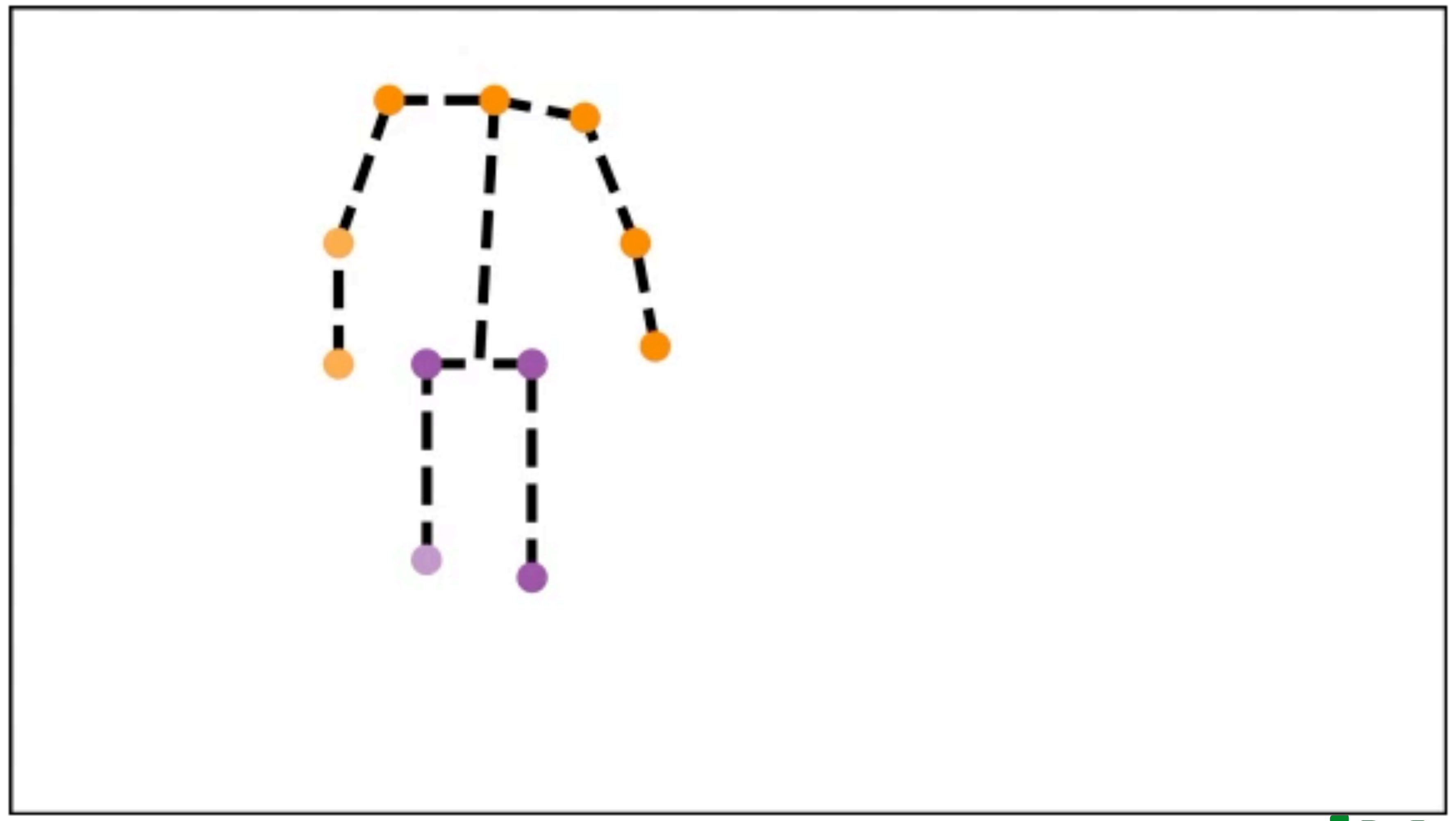


[Monszpart, Guerrero, Ceylan, Yumer, Mitra, Siggraph, 2019]

3. Data-driven Human-Object Interaction Capture

input video:

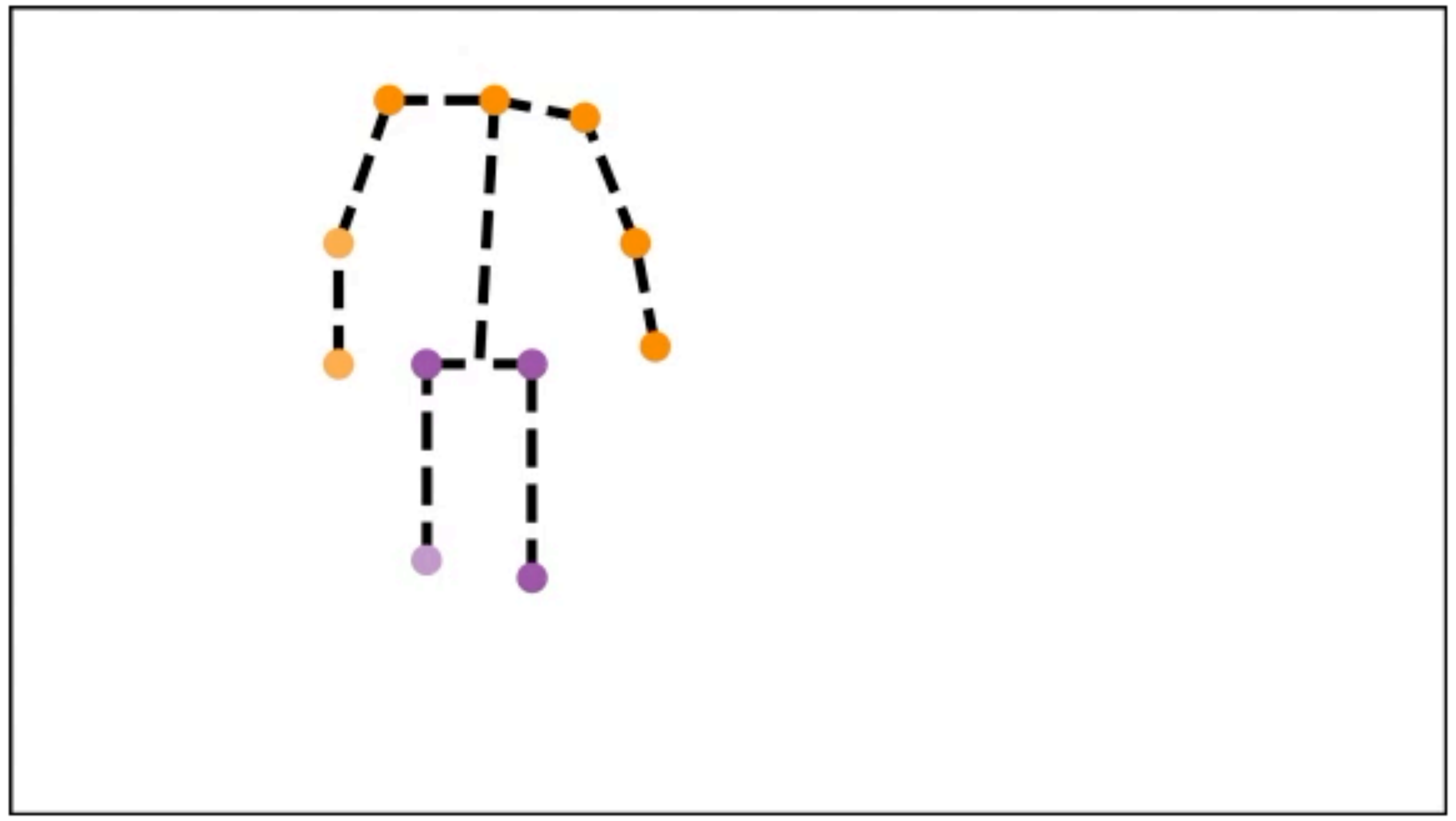
initial skeletal estimate:



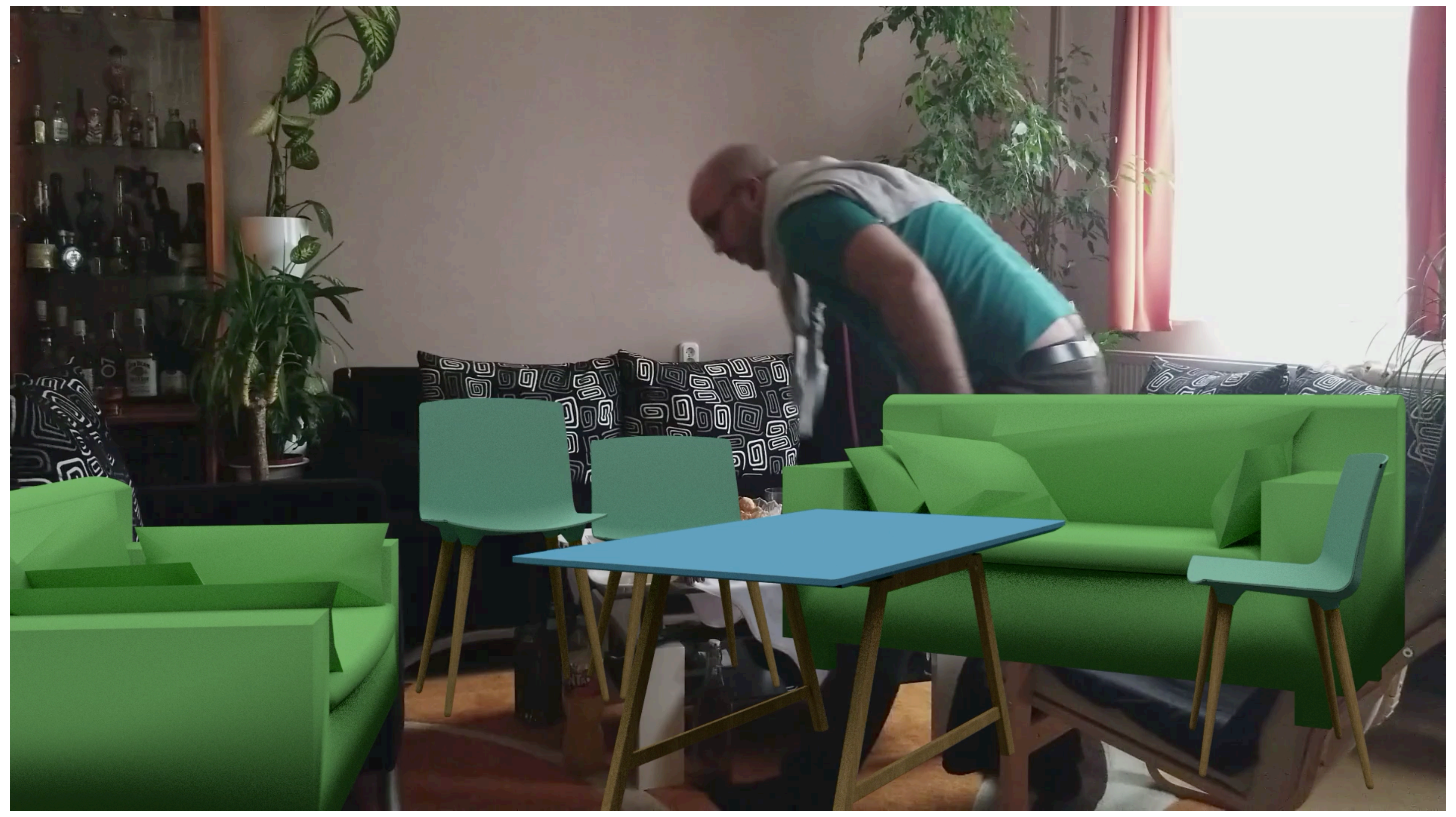
[Monszpart, Guerrero, Ceylan, Yumer, Mitra, Siggraph, 2019]

3. Data-driven Human-Object Interaction Capture

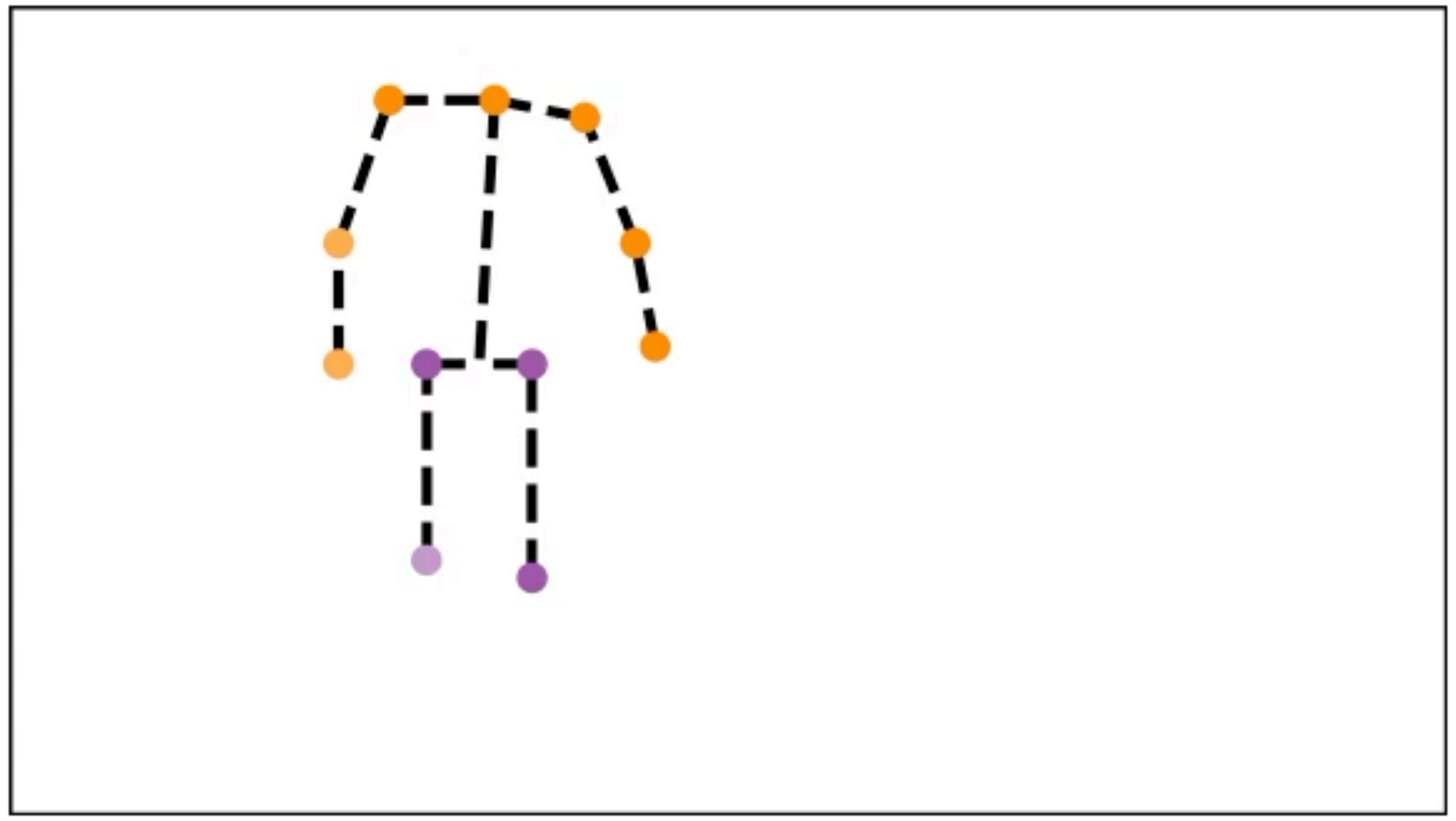
initial skeletal estimate:



3. Data-driven Human-Object Interaction Capture



initial skeletal estimate:



Representation for 3D

- Image-based
- Volumetric
- Surface-based
  - **PROS**: parameterize + image networks (instrinsic representation)
  - **CONS**: suffers from parameterisation artefacts (local versus global distortion), requires good quality mesh
- Point-based

Representation for 3D

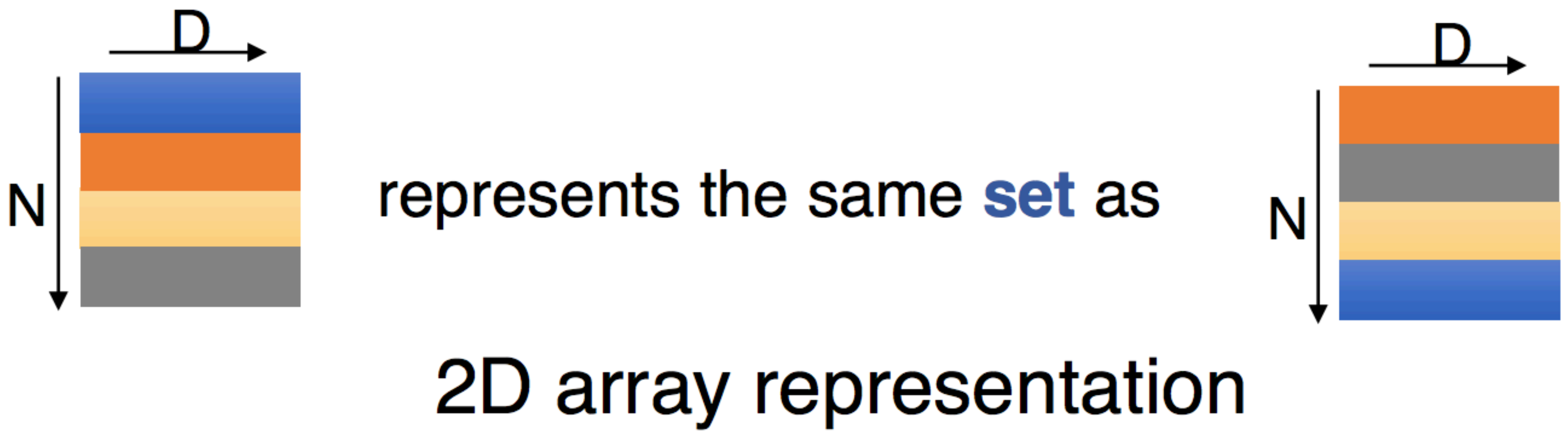
- Image-based
- Volumetric
- Surface-based
- **Point-based**

Representation for 3D: Point-based

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

In Original Representation

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



[Qi 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

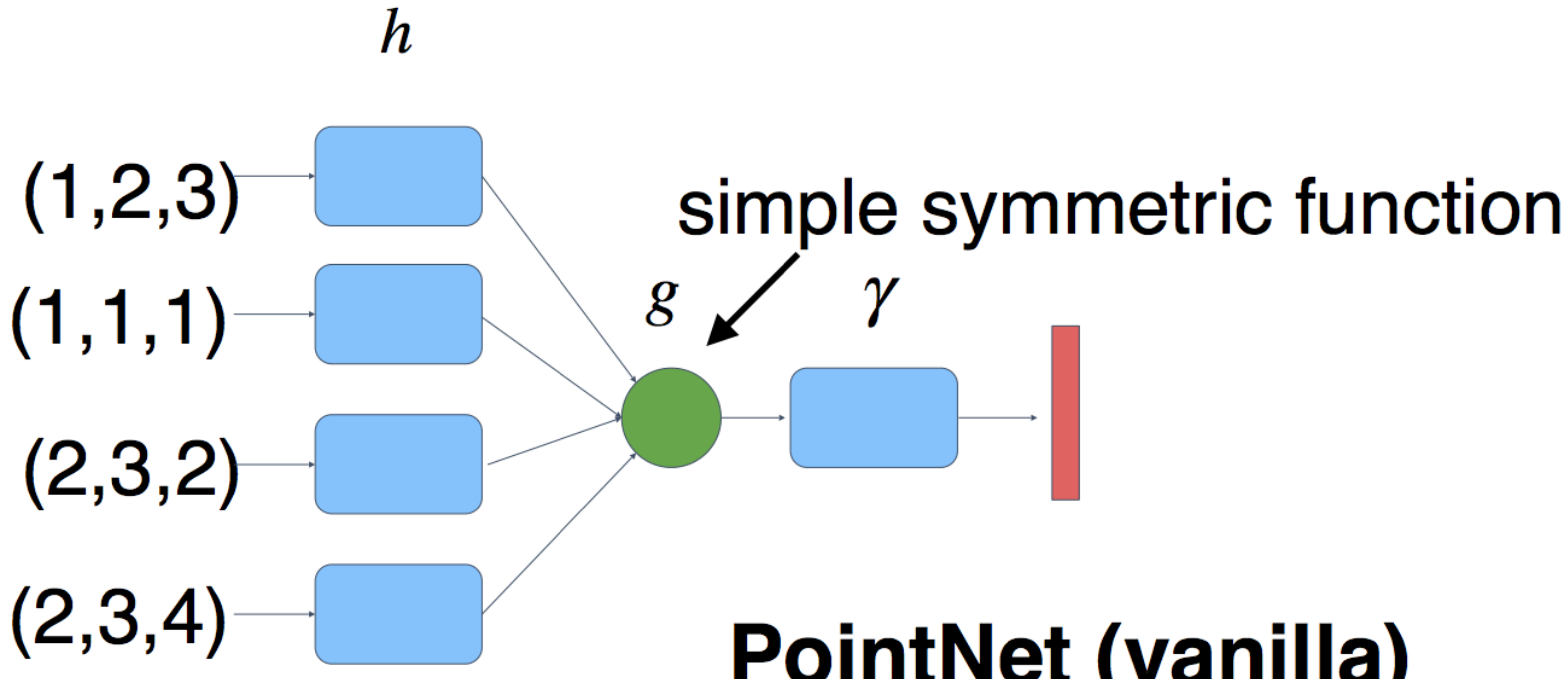
[Qi et al. 2017]

PointNet for Point Cloud Analysis

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

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

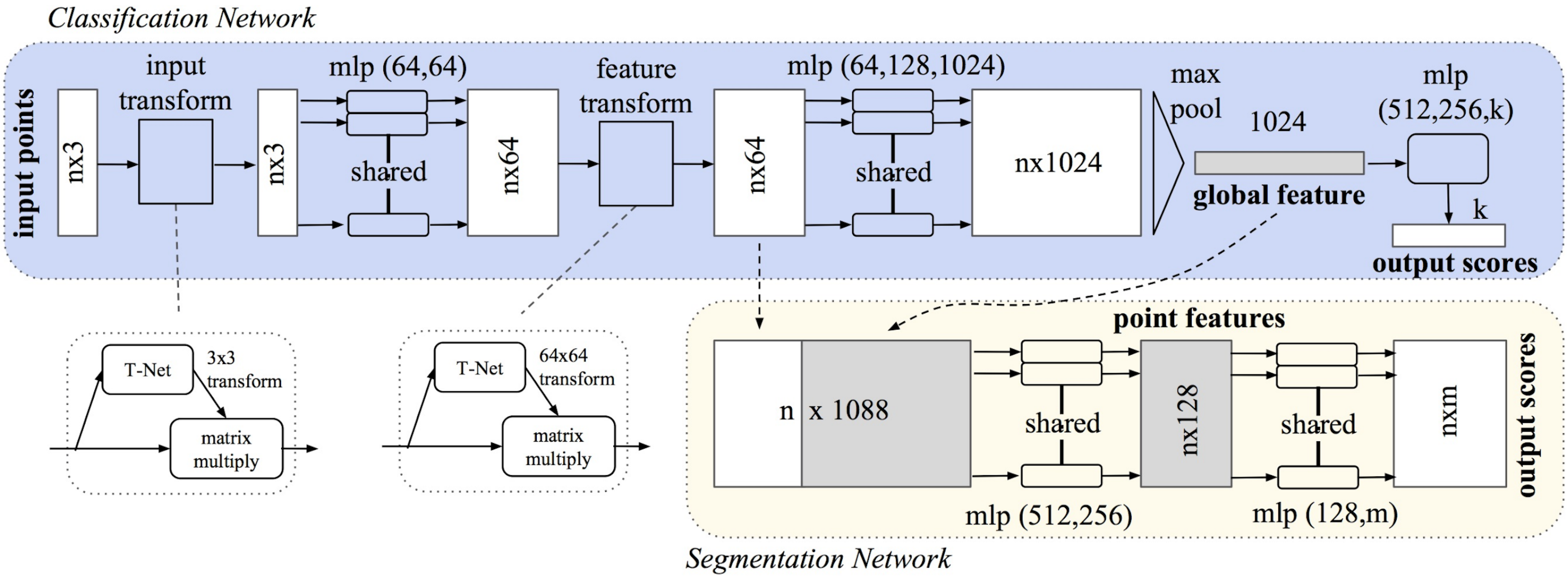
$$f(x_1, x_2, \ldots, x_n) = \gamma \circ g(h(x_1), \ldots, h(x_n))$$



PointNet (vanilla)

[Qi et al. 2017]

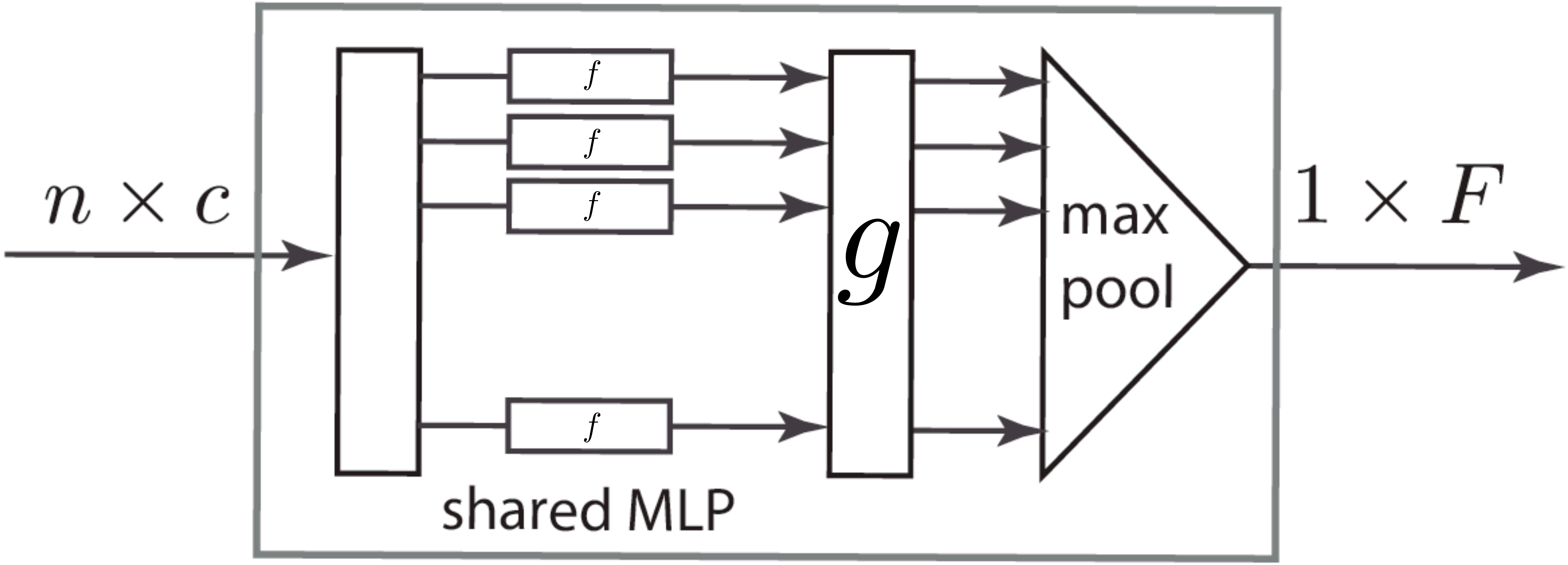
PointNet Architecture



[Qi et al. 2017]

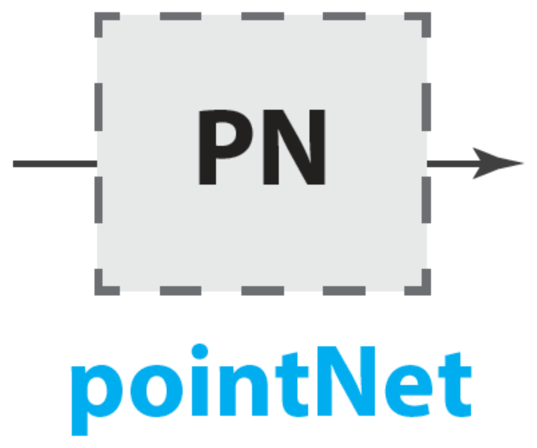
PointNet Revisited

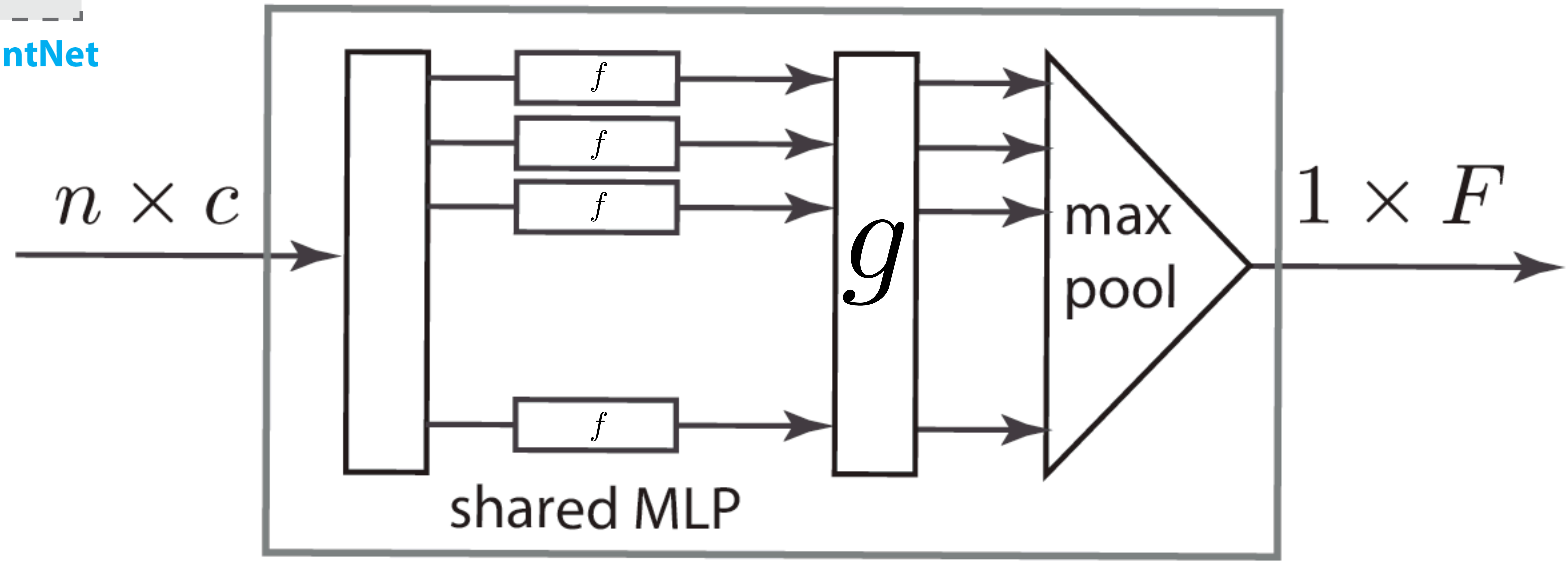
$$\sum_{p_j \in \mathbb{P}} g(f(p_j))$$



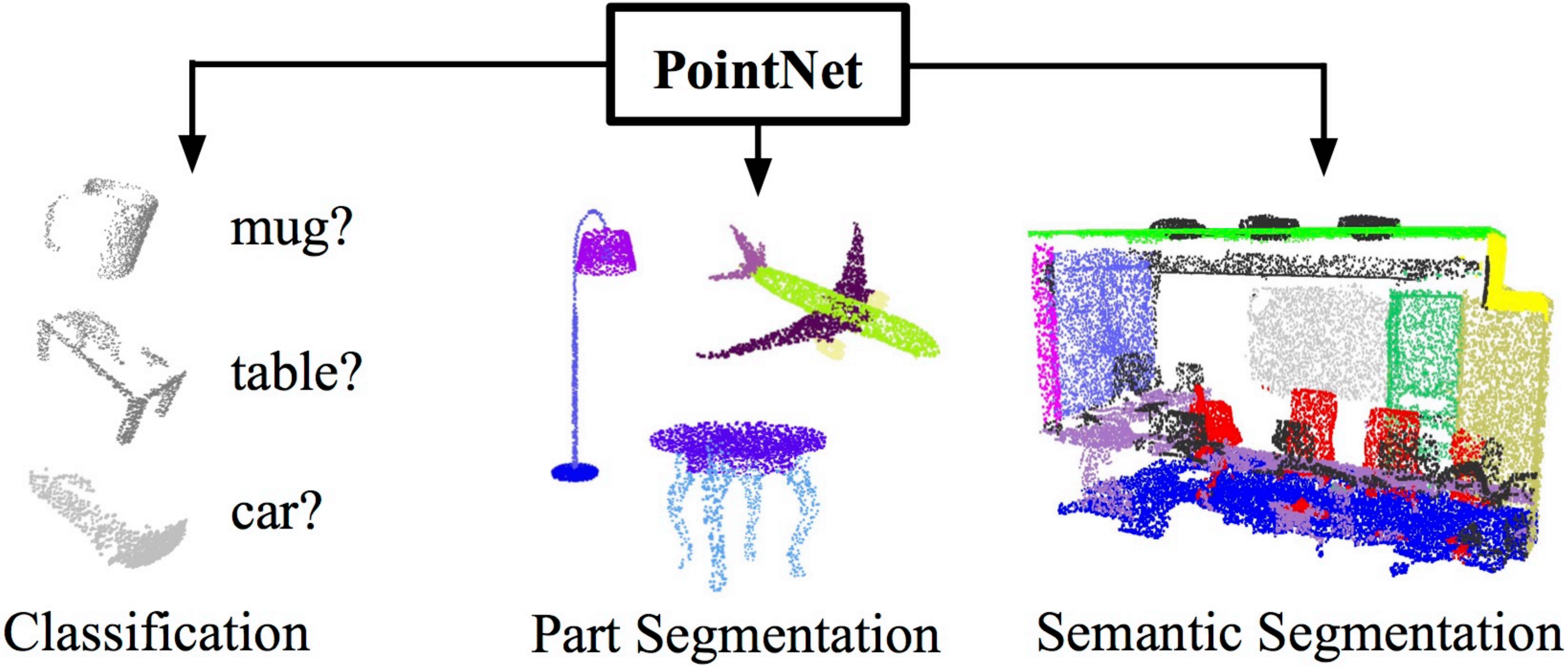
PointNet Revisited

$$\sum_{p_j \in \mathbb{P}} g(f(p_j))$$

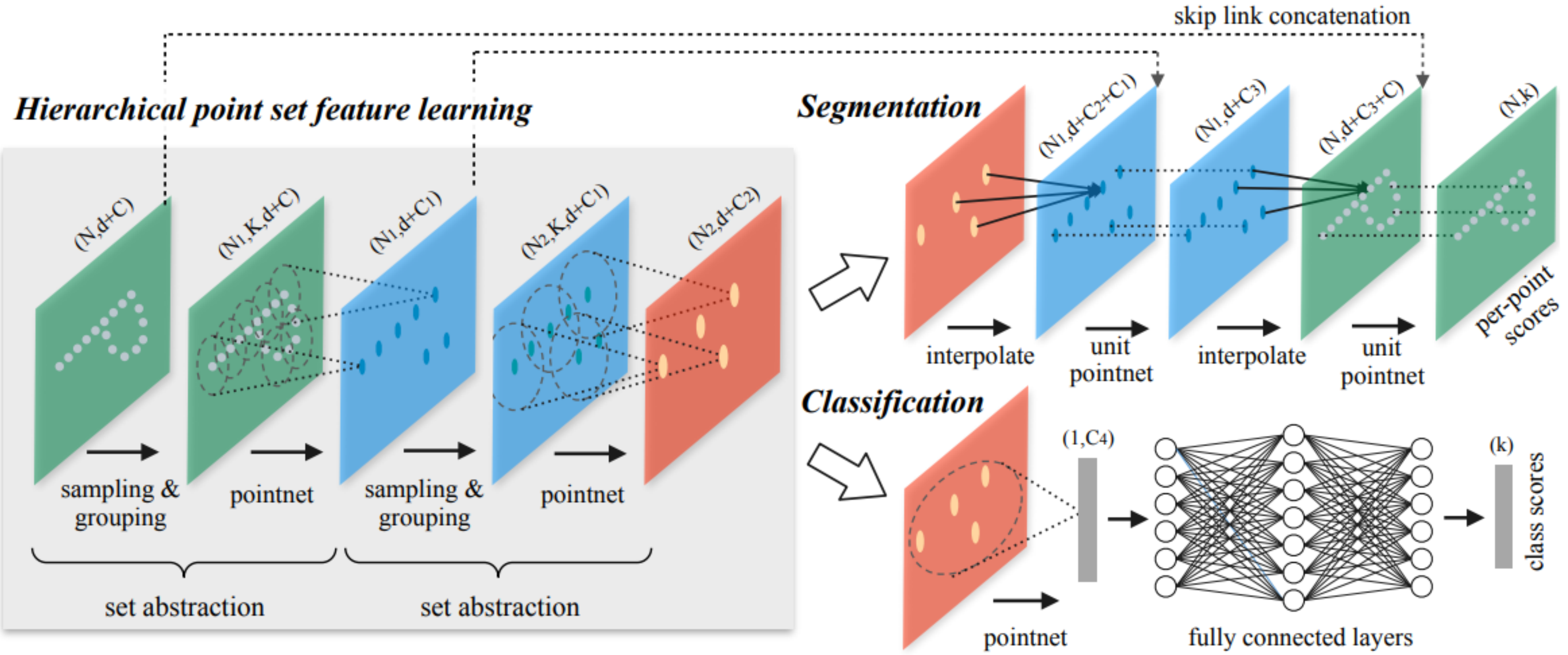




PointNet for Point Cloud Analysis

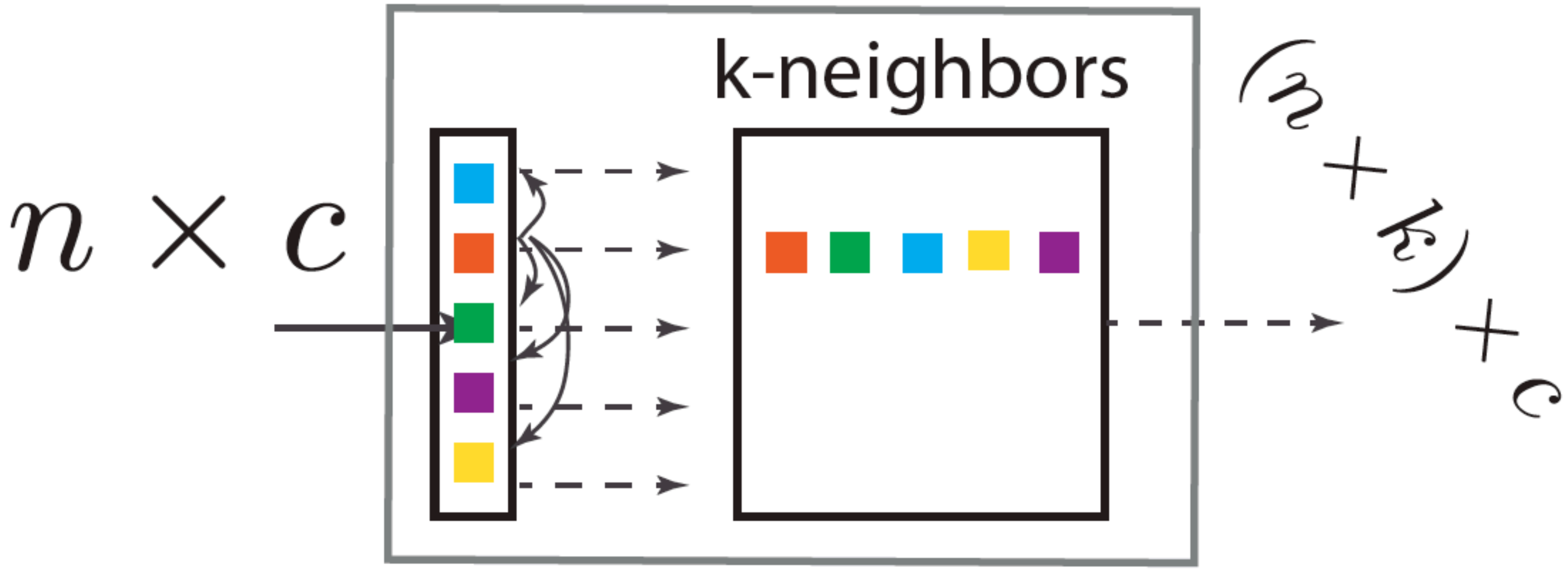


PointNet++

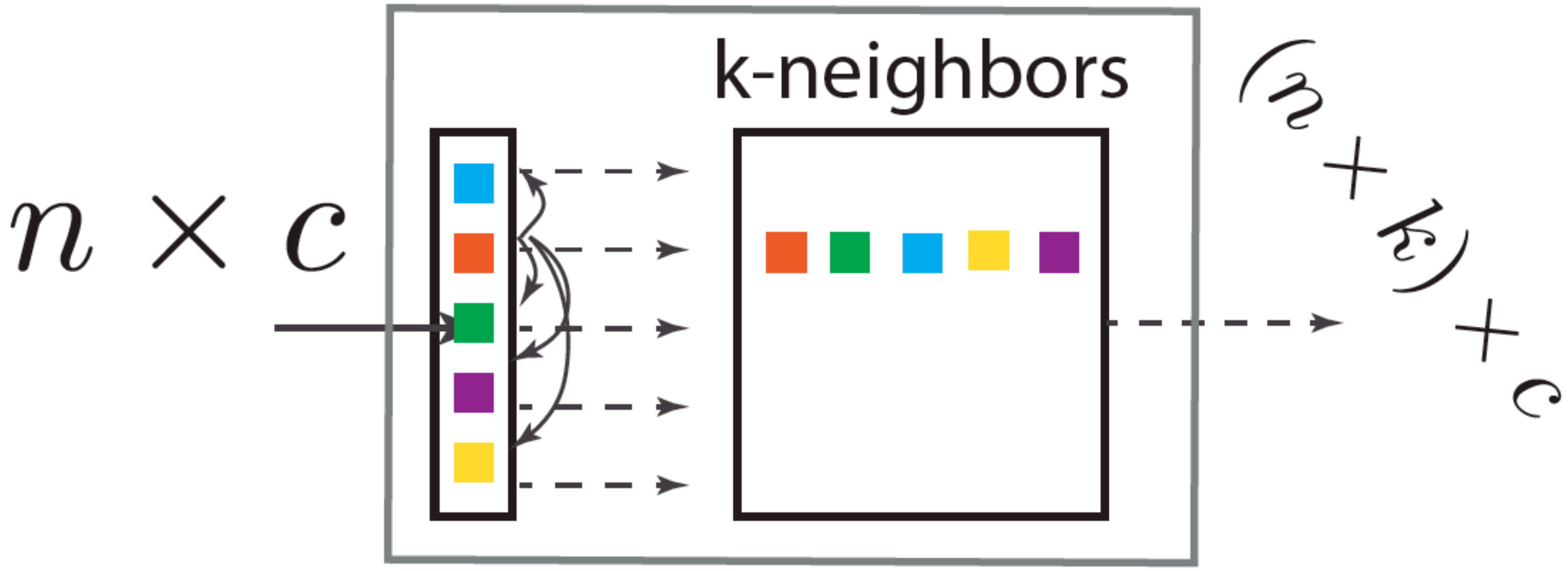


[Qi et al. 2017]

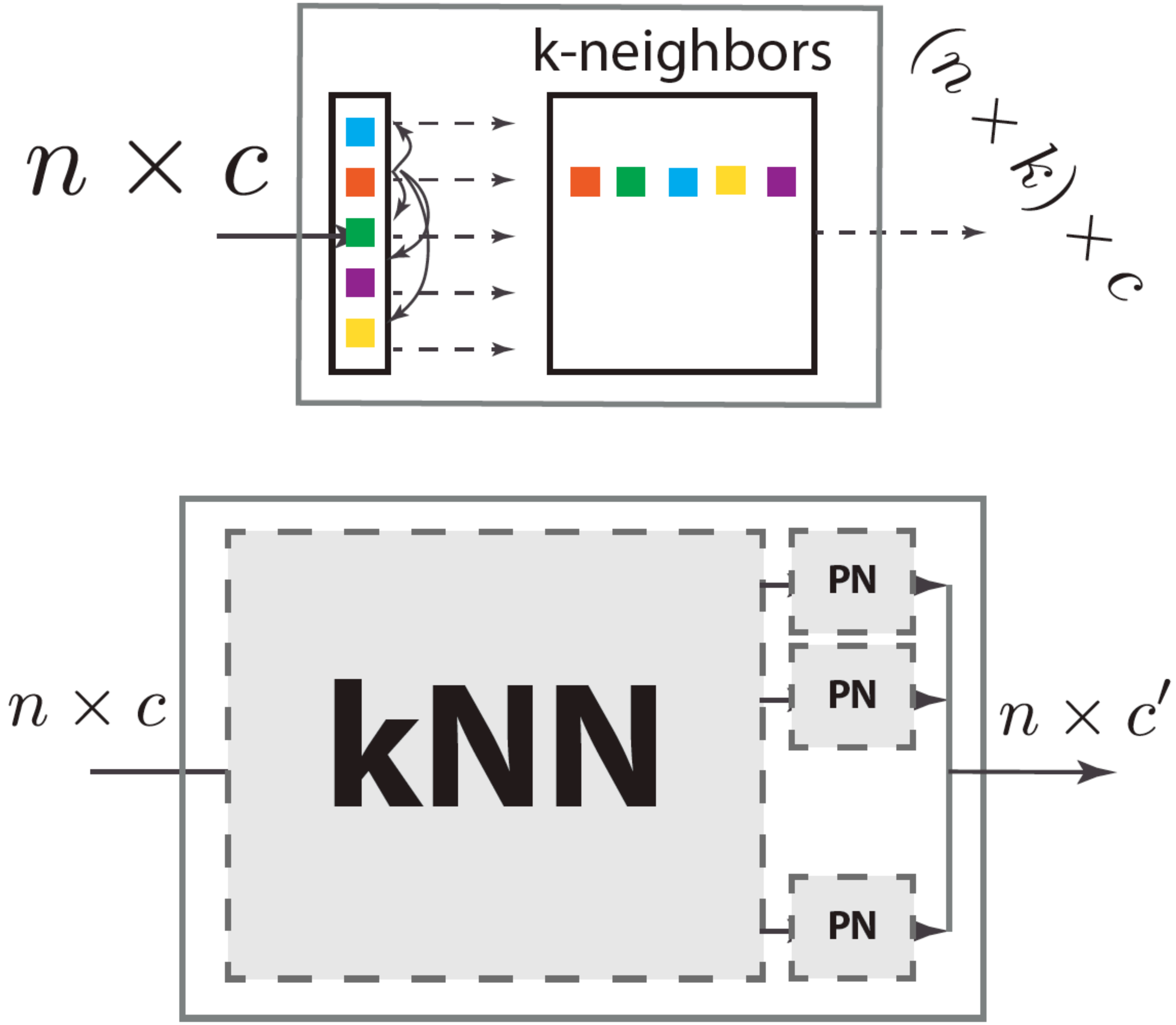
PointNet++ Revisited



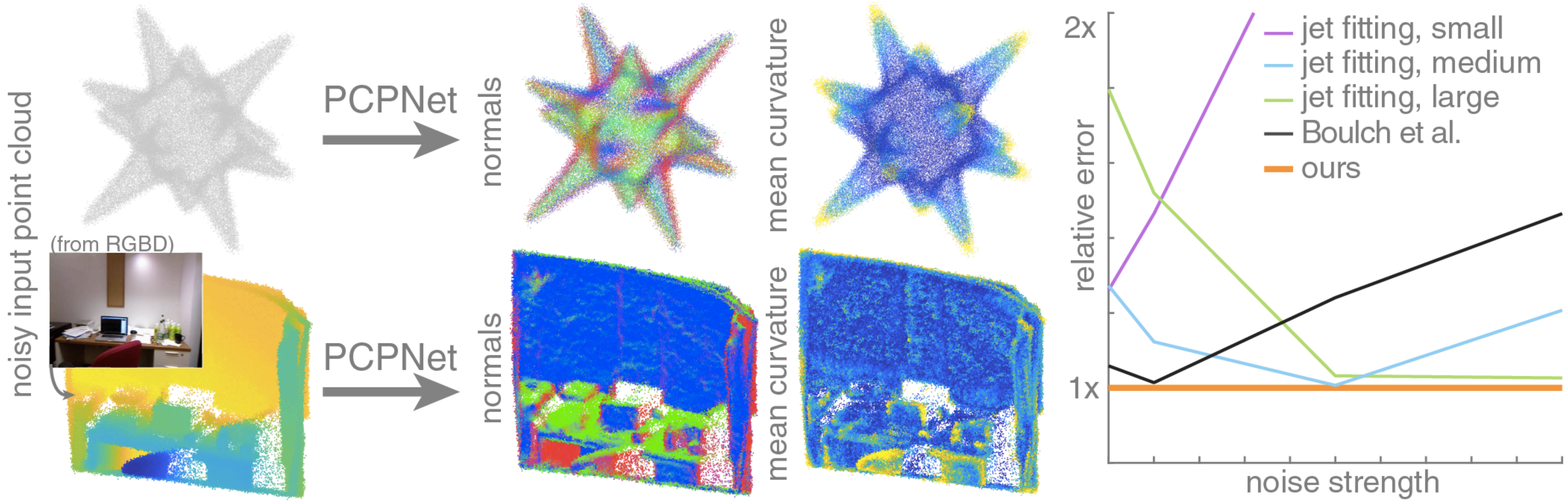
PointNet++ Revisited



PointNet++ Revisited

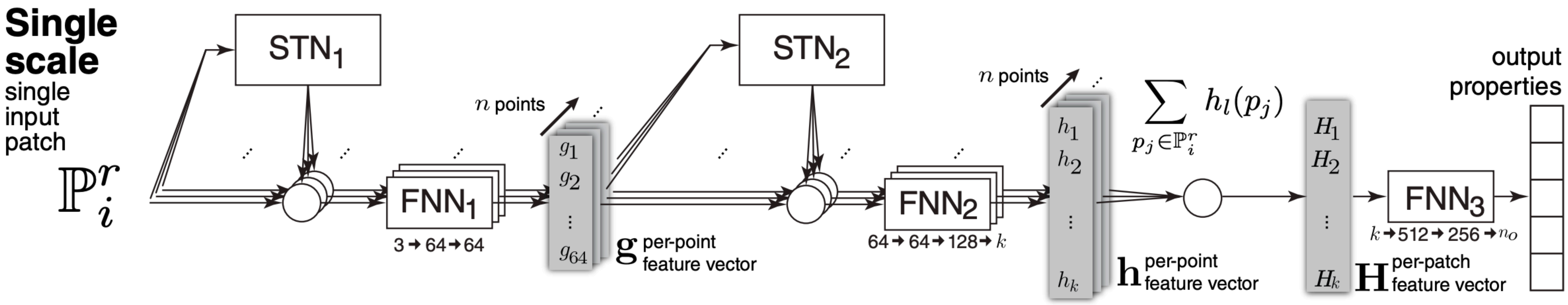


PCPNet for Local Point Cloud Analysis

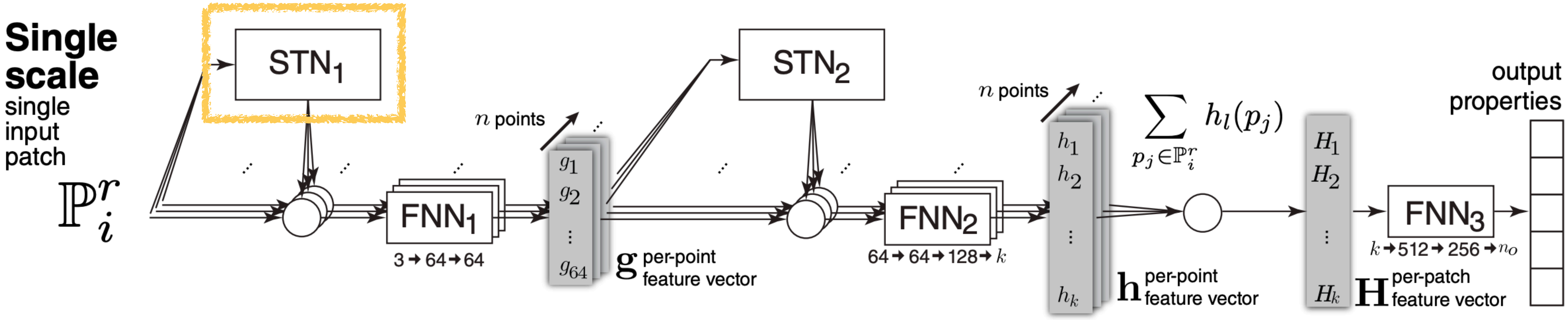


[Guerrero, Kleiman, Ovsjanikov, Mitra, EG, 2018]

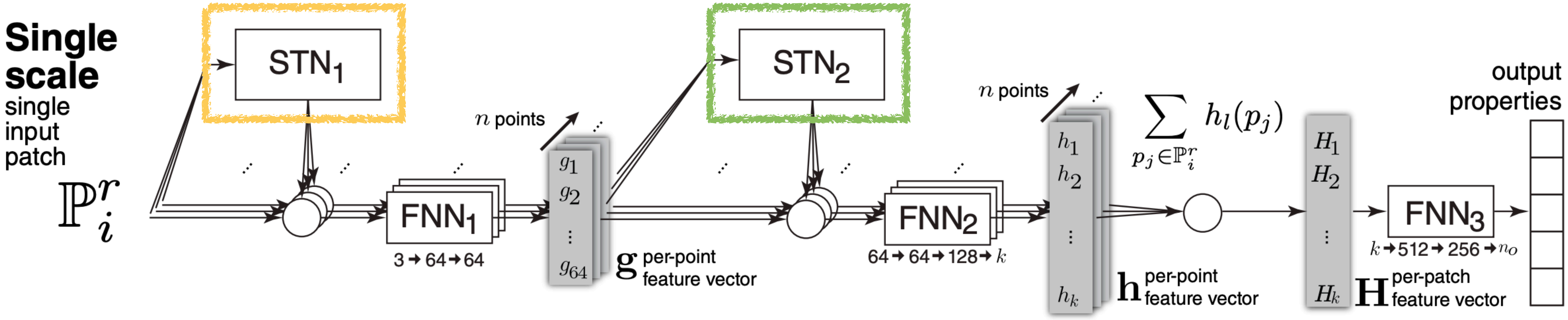
PCPNet Architecture



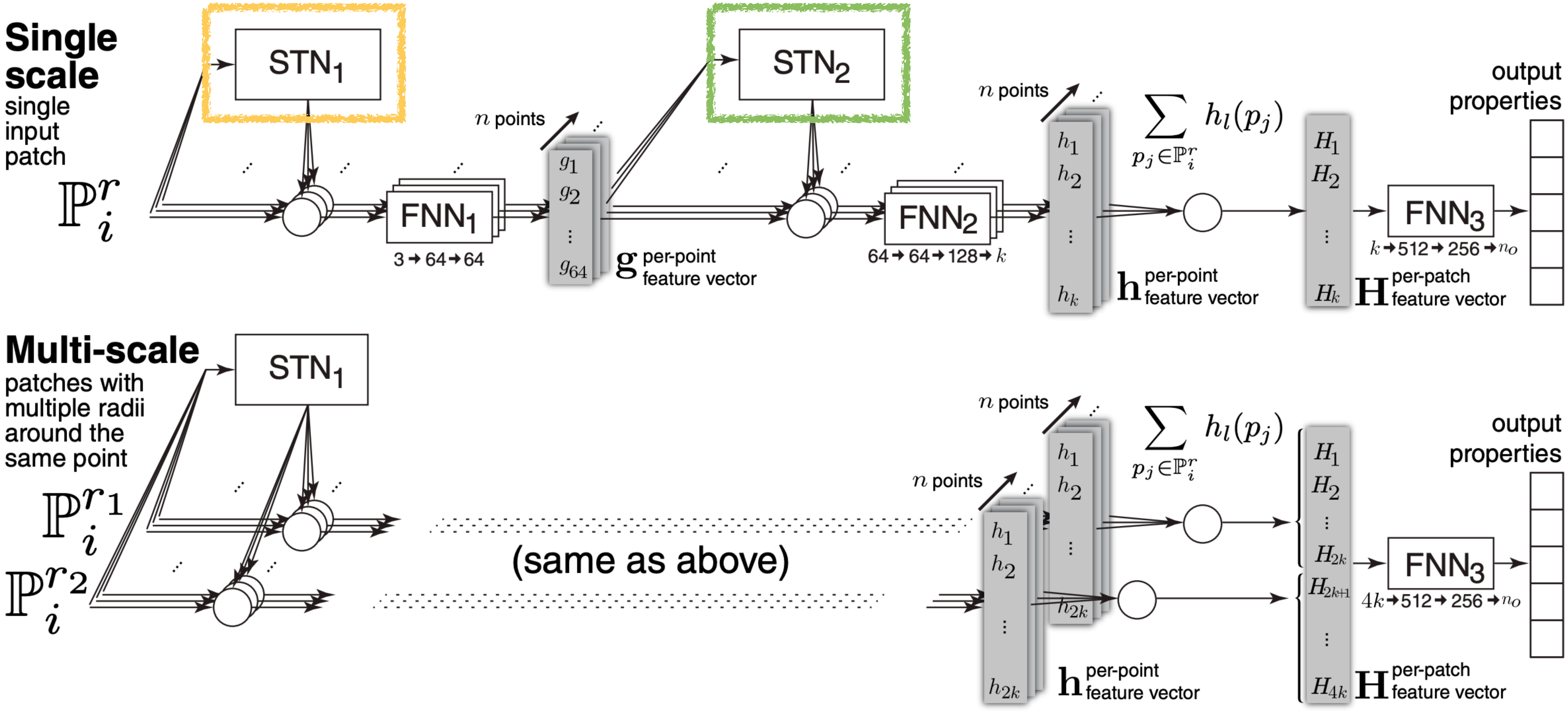
PCPNet Architecture



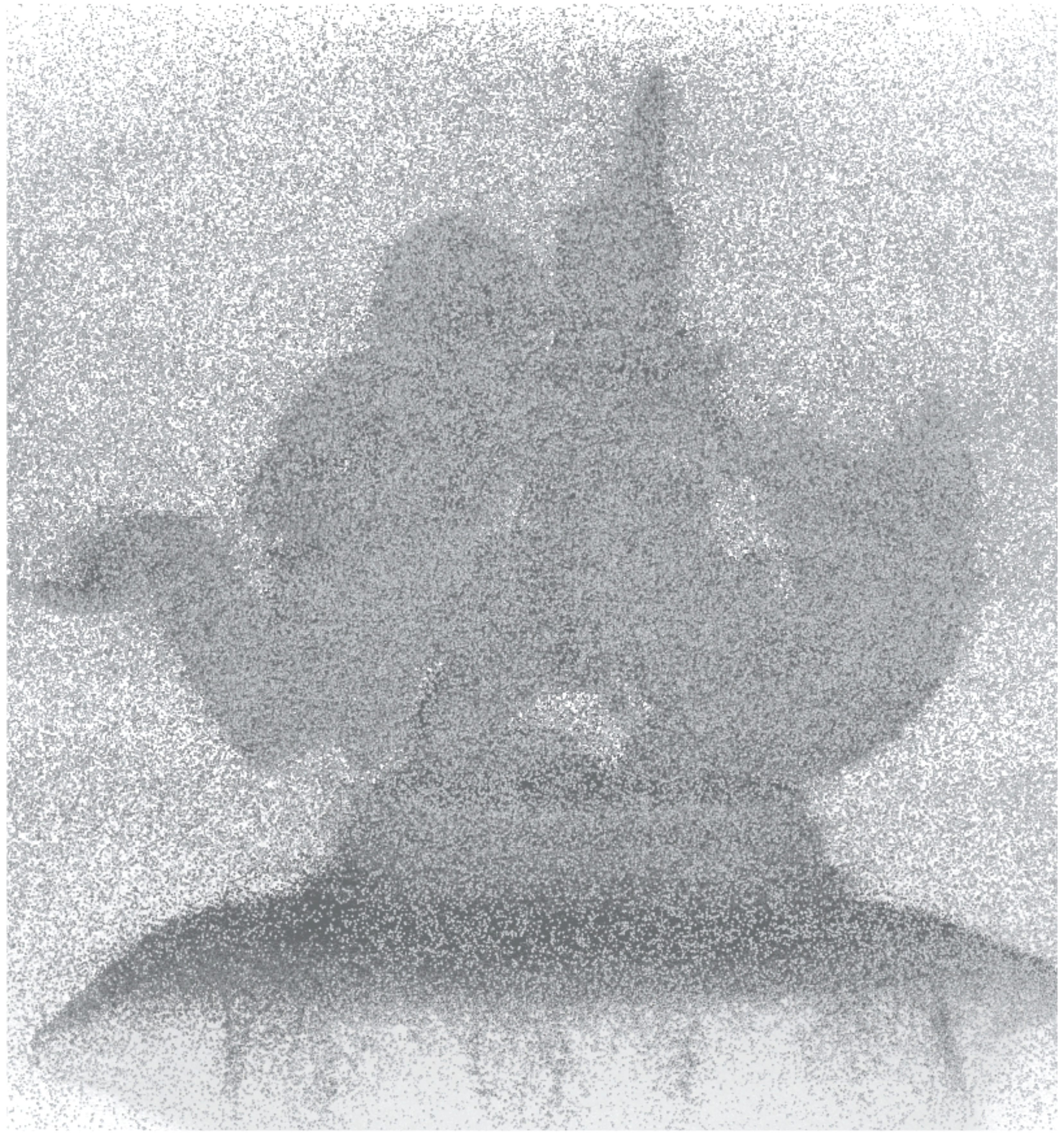
PCPNet Architecture



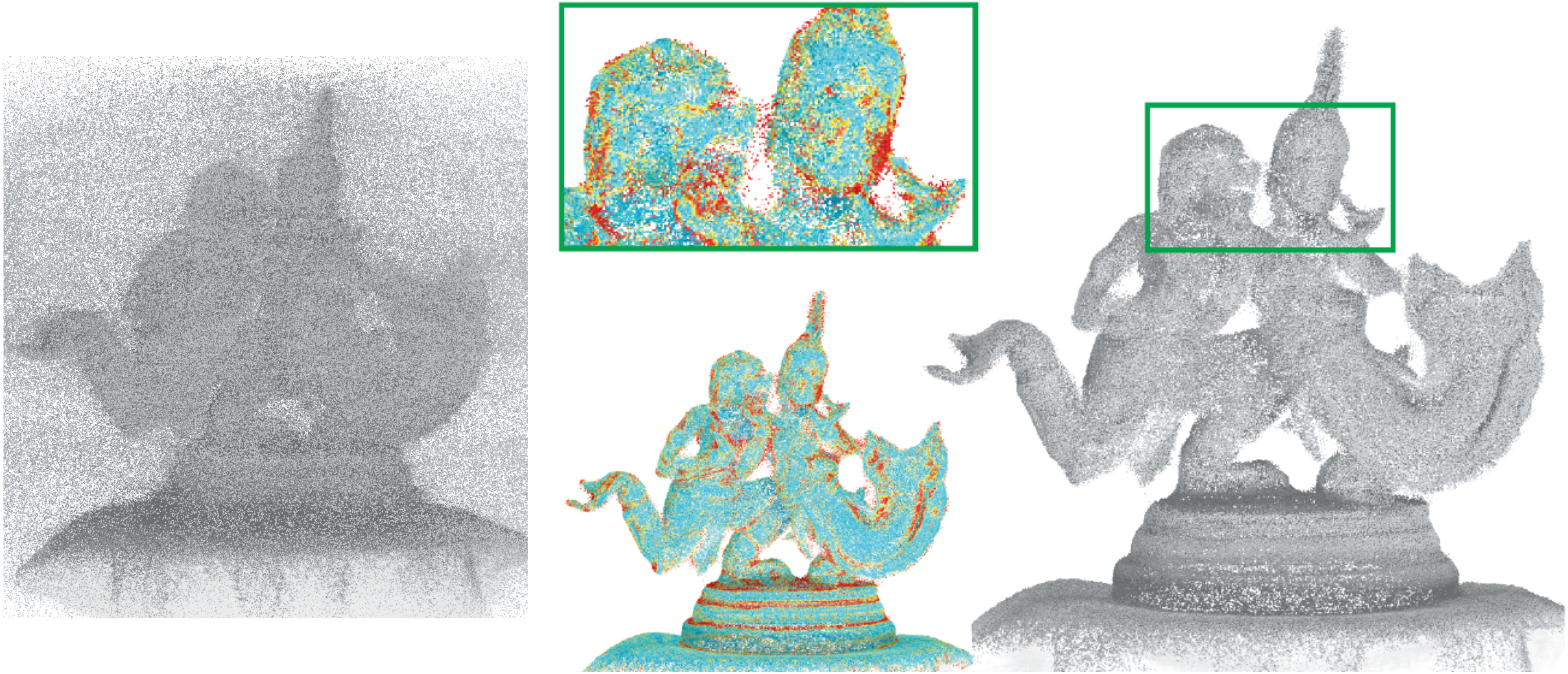
PCPNet Architecture



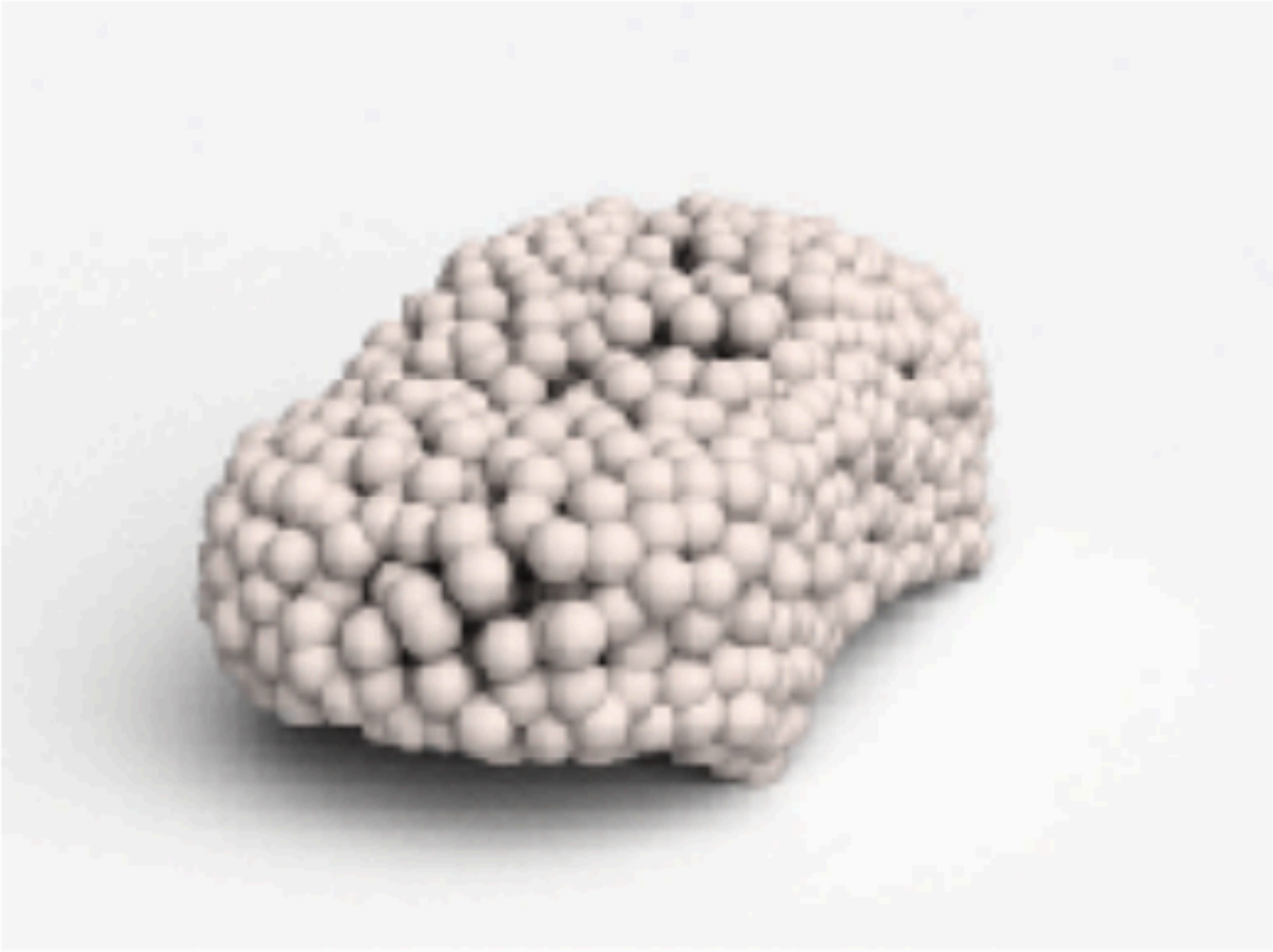
Denoising Results



Denoising Results

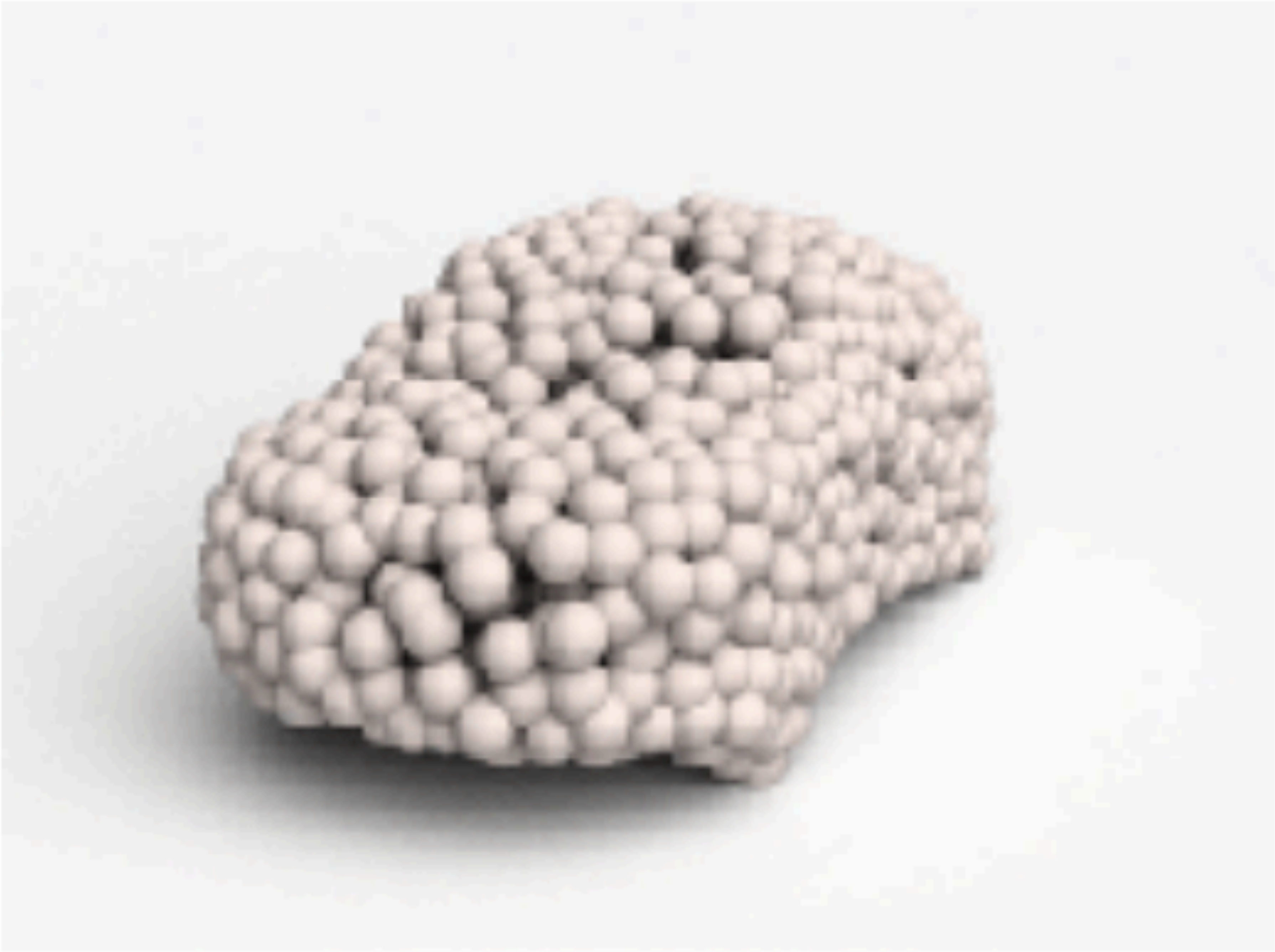


PointNet for Point Cloud Synthesis



[Su et al. 2017]

PointNet for Point Cloud Synthesis



Earth Mover Distance as loss function

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

[Su et al. 2017]

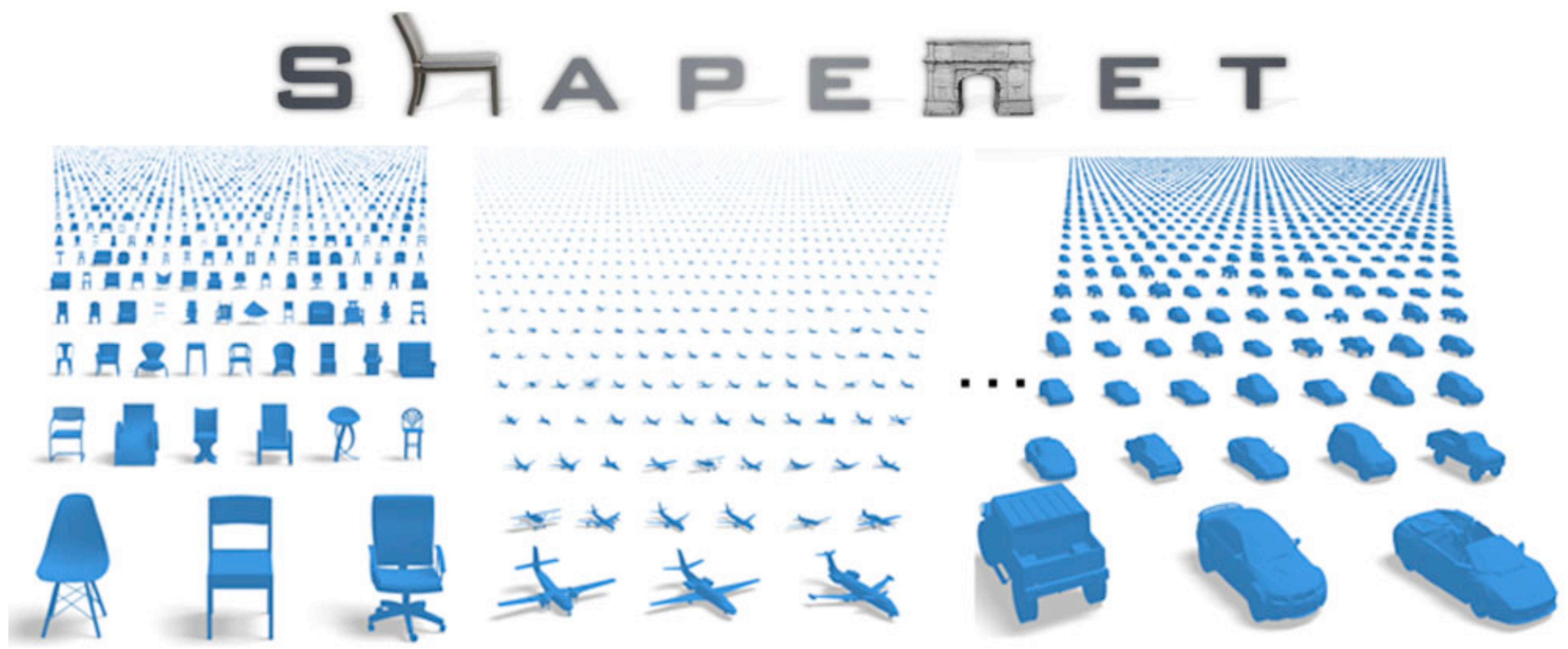
Shape Completion as Unpaired Data Translation

[Chen, Chen, Mitra, Arxiv, 2019]

scanned models
e.g., ScanNet,
MatterPort, NYU, etc.

Shape Completion as Unpaired Data Translation

[Chen, Chen, Mitra, Arxiv, 2019]

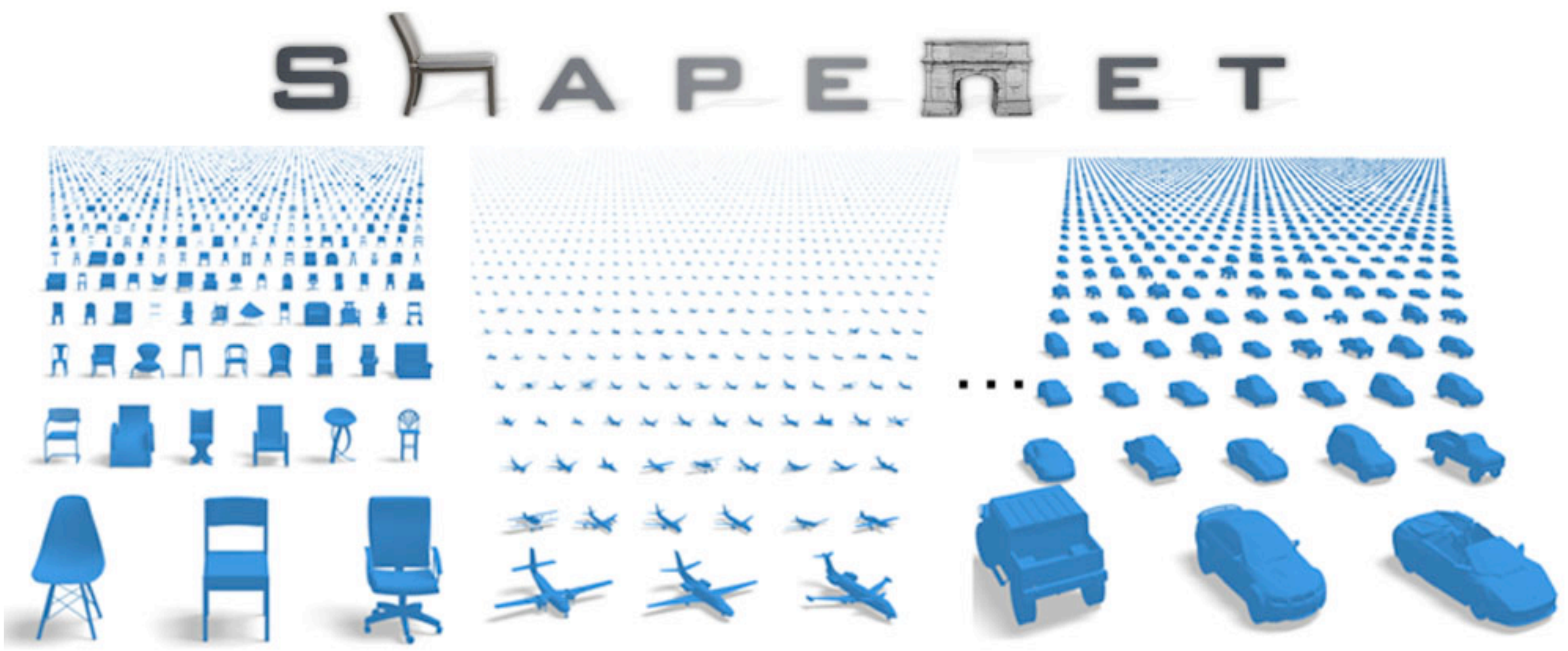


scanned models
e.g., ScanNet,
MatterPort, NYU, etc.

3D synthetic models
e.g., ShapeNet,
SunCG, Ikea, etc.

Shape Completion as Unpaired Data Translation

[Chen, Chen, Mitra, Arxiv, 2019]

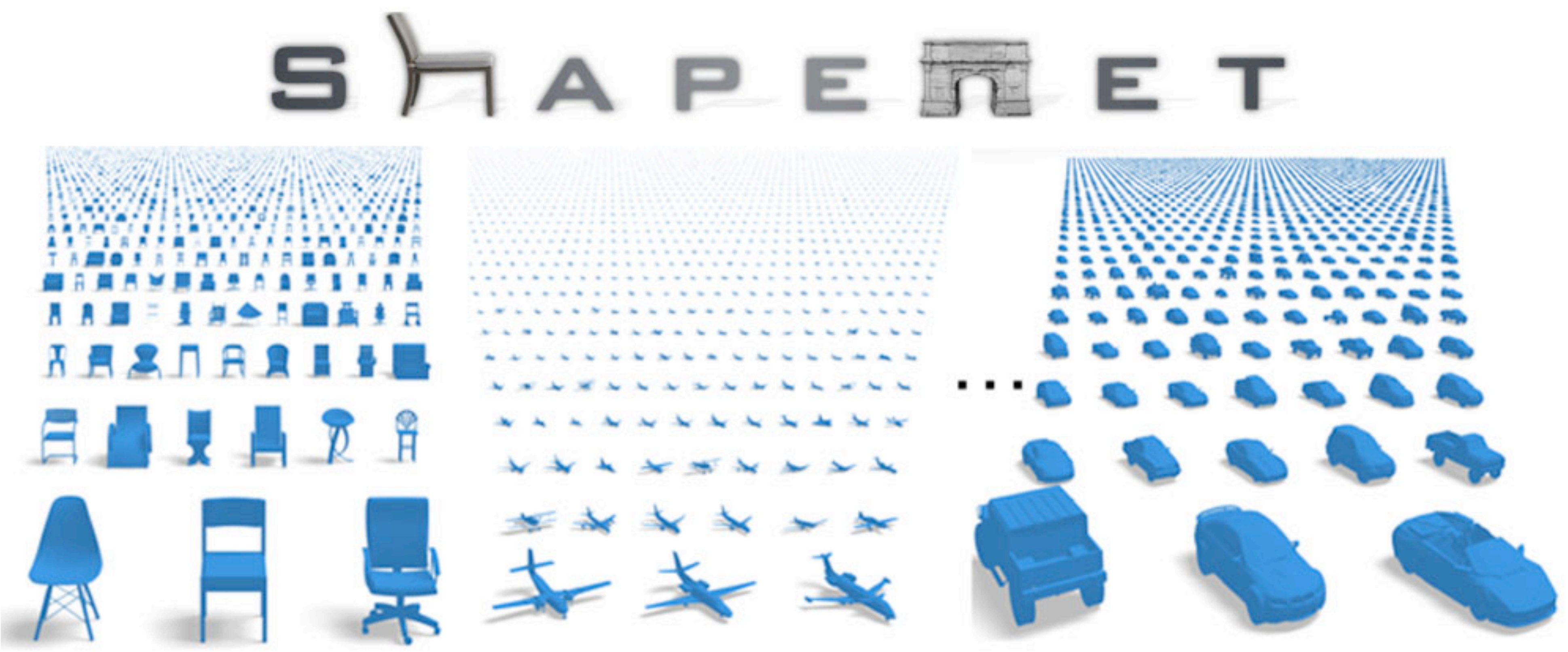


scanned models
e.g., ScanNet,
MatterPort, NYU, etc.

3D synthetic models
e.g., ShapeNet,
SunCG, Ikea, etc.

Shape Completion as Unpaired Data Translation

[Chen, Chen, Mitra, Arxiv, 2019]

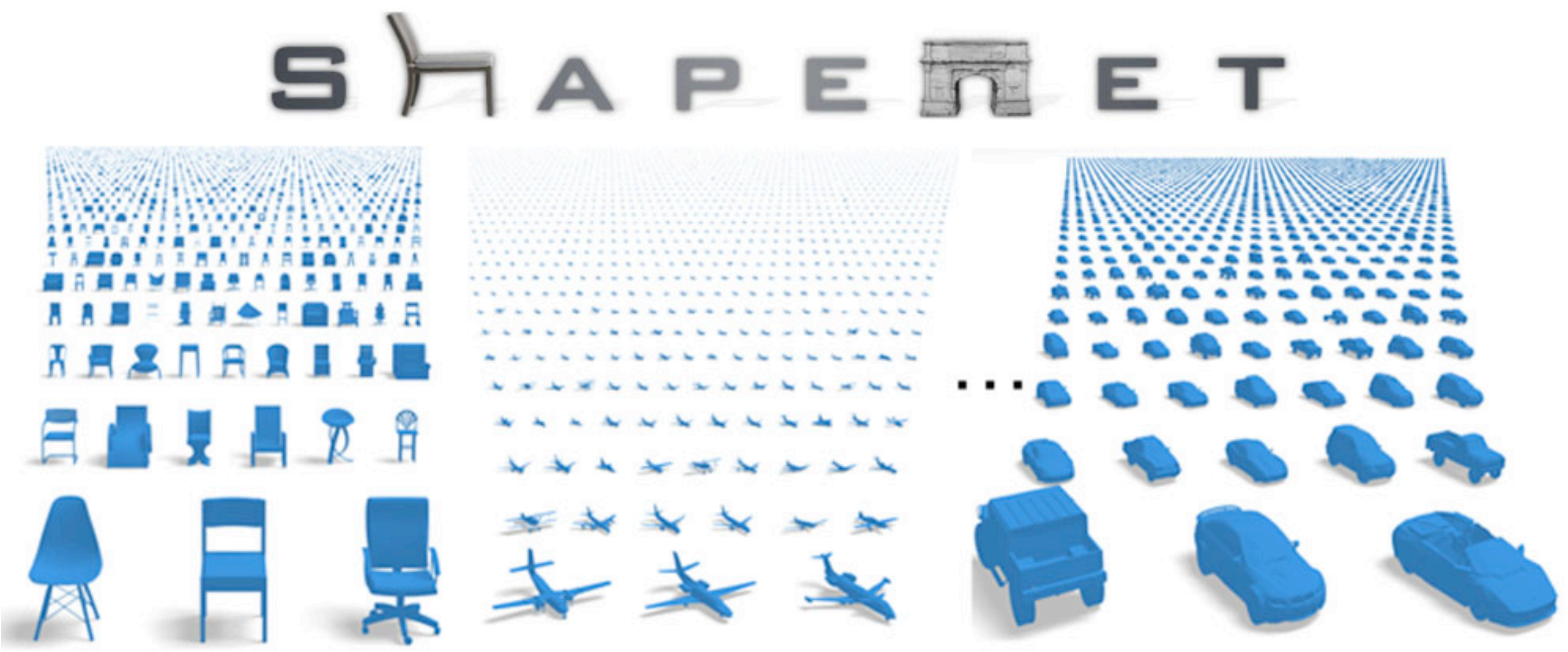


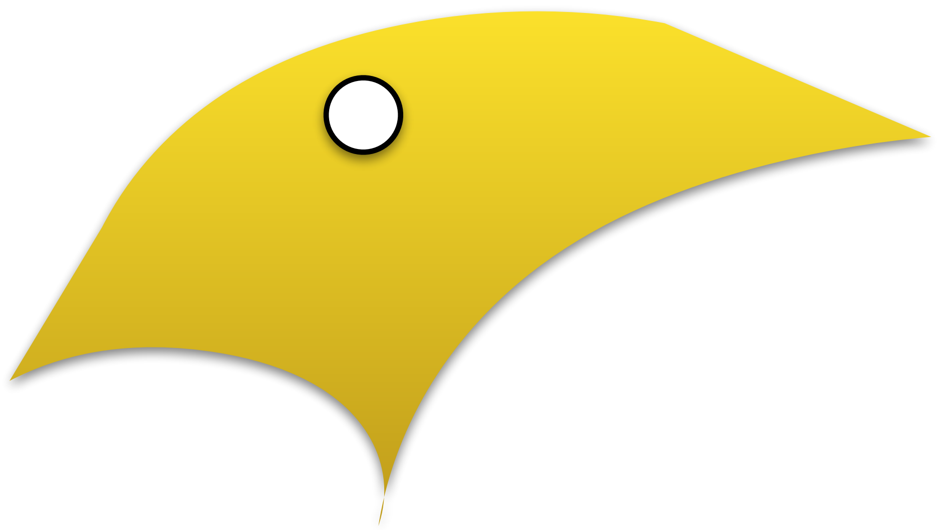
scanned models
e.g., ScanNet,
MatterPort, NYU, etc.

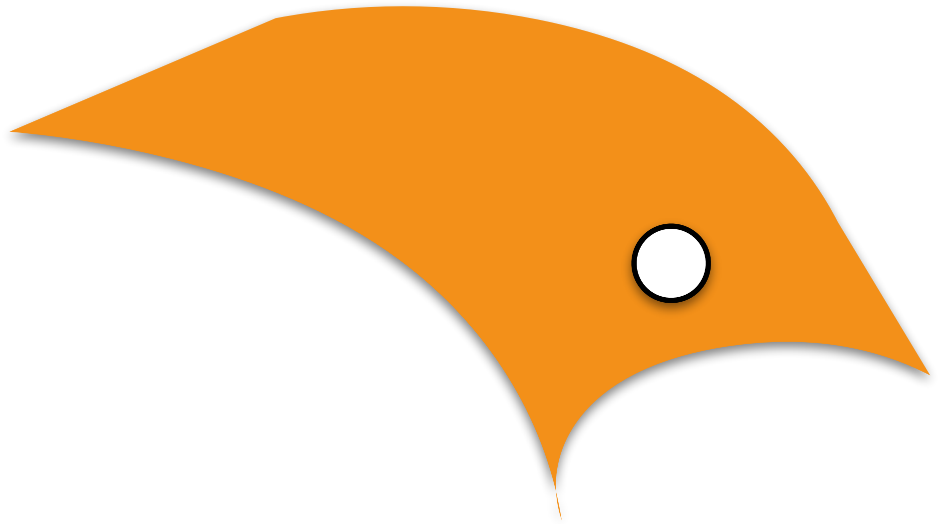
3D synthetic models
e.g., ShapeNet,
SunCG, Ikea, etc.

Shape Completion as Unpaired Data Translation

[Chen, Chen, Mitra, Arxiv, 2019]





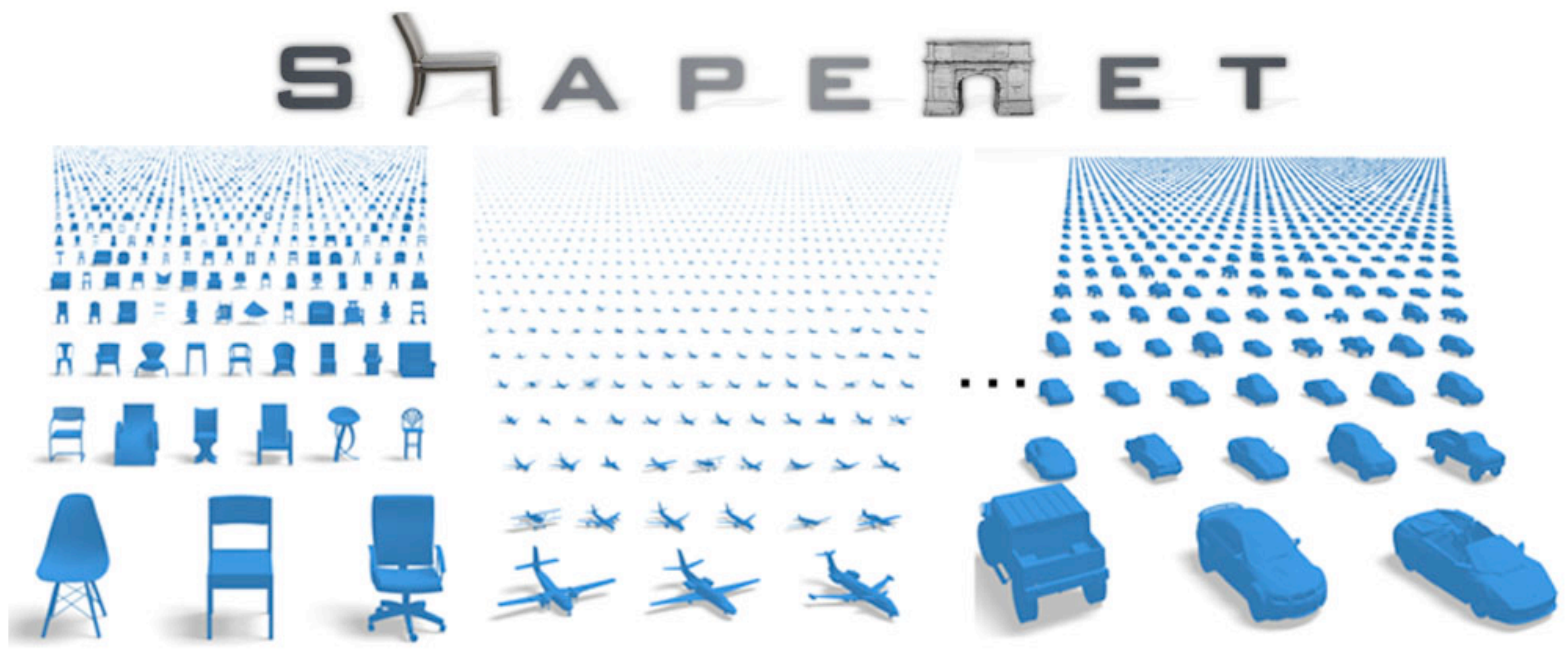


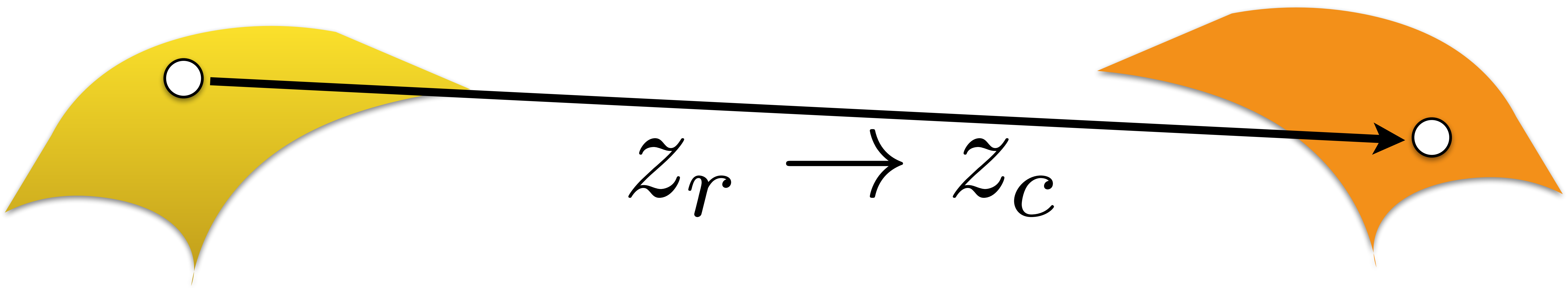
scanned models
e.g., ScanNet,
MatterPort, NYU, etc.

3D synthetic models
e.g., ShapeNet,
SunCG, Ikea, etc.

Shape Completion as Unpaired Data Translation

[Chen, Chen, Mitra, Arxiv, 2019]

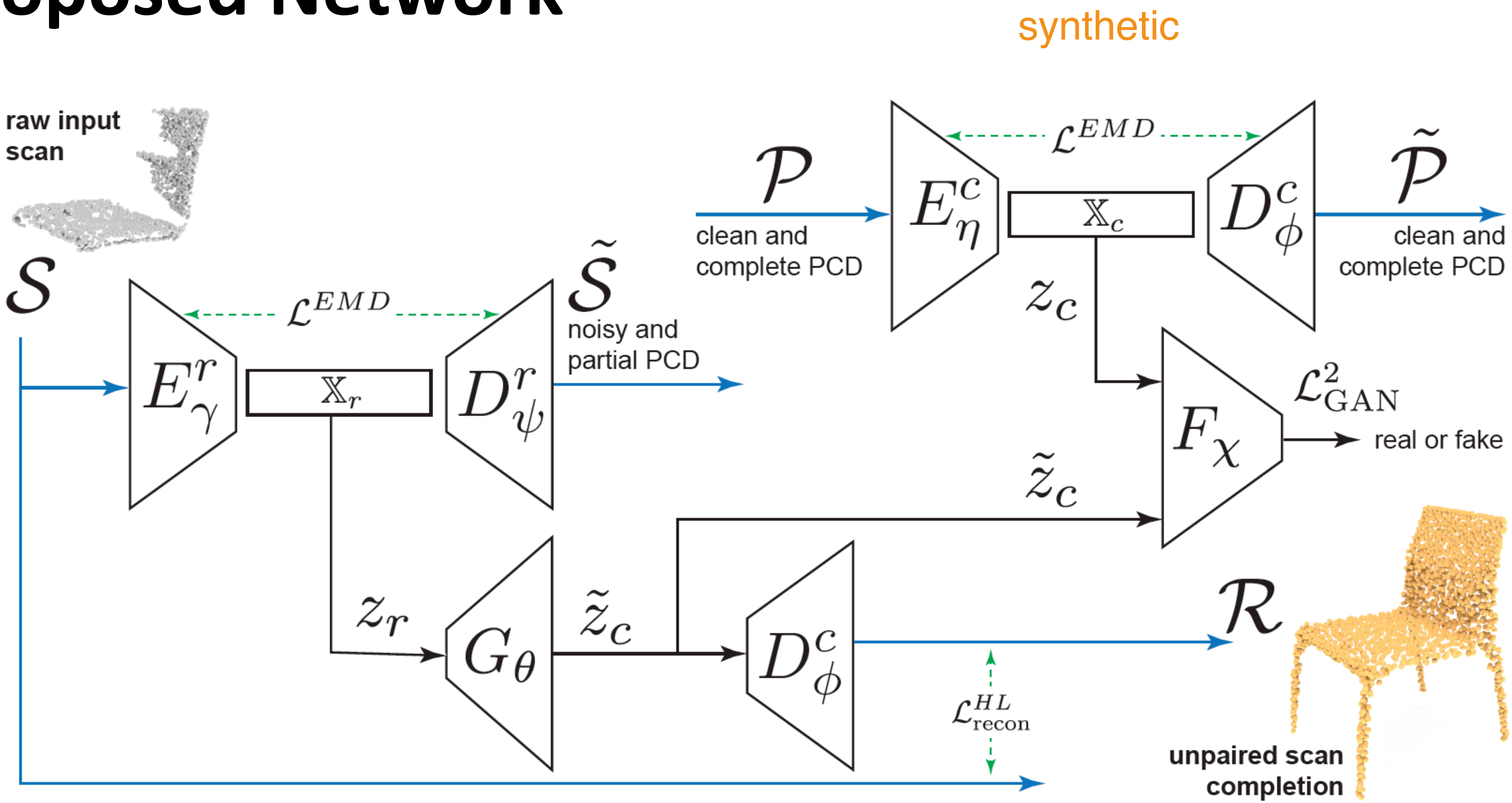




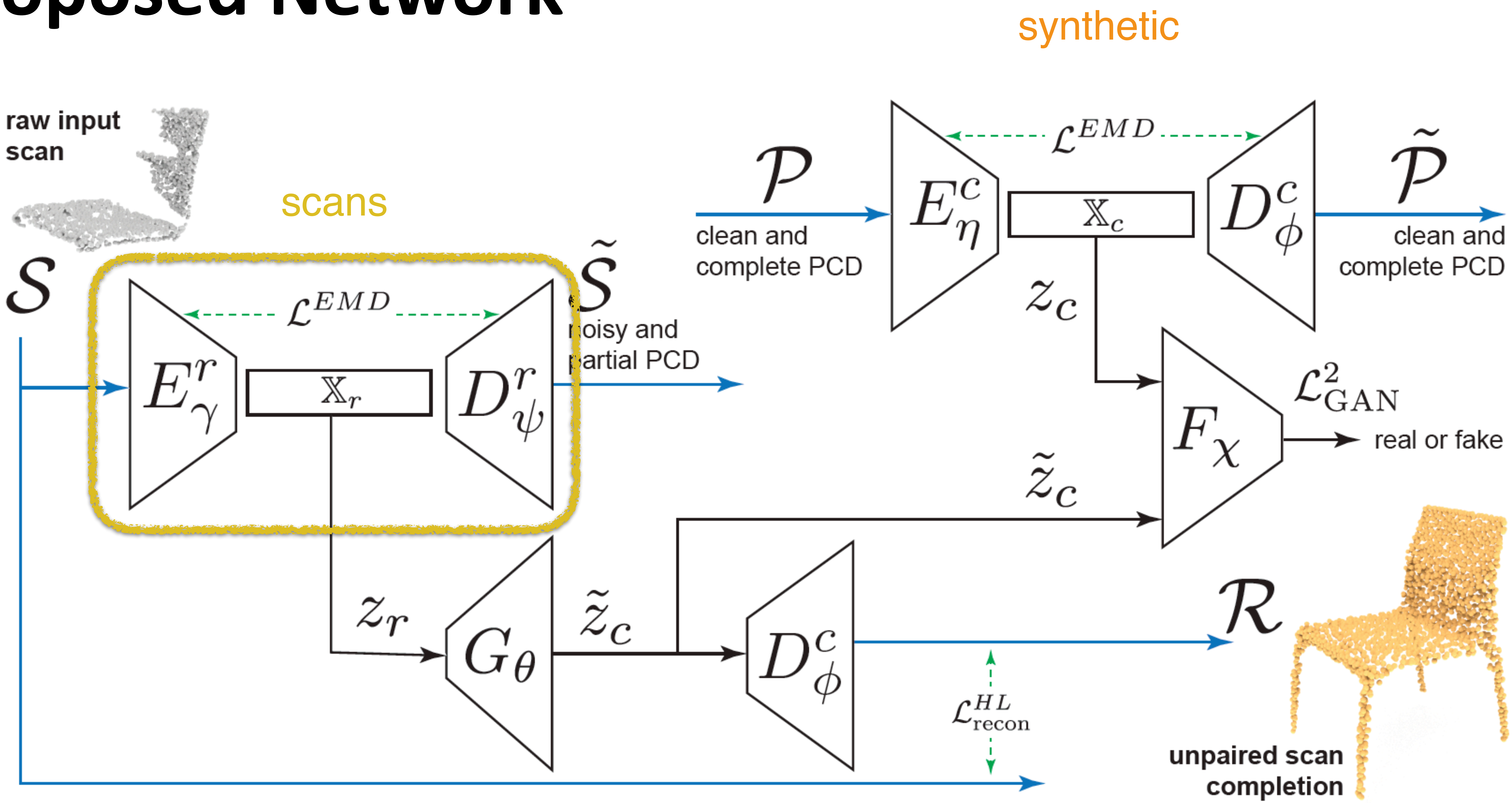
scanned models
e.g., ScanNet,
MatterPort, NYU, etc.

3D synthetic models
e.g., ShapeNet,
SunCG, Ikea, etc.

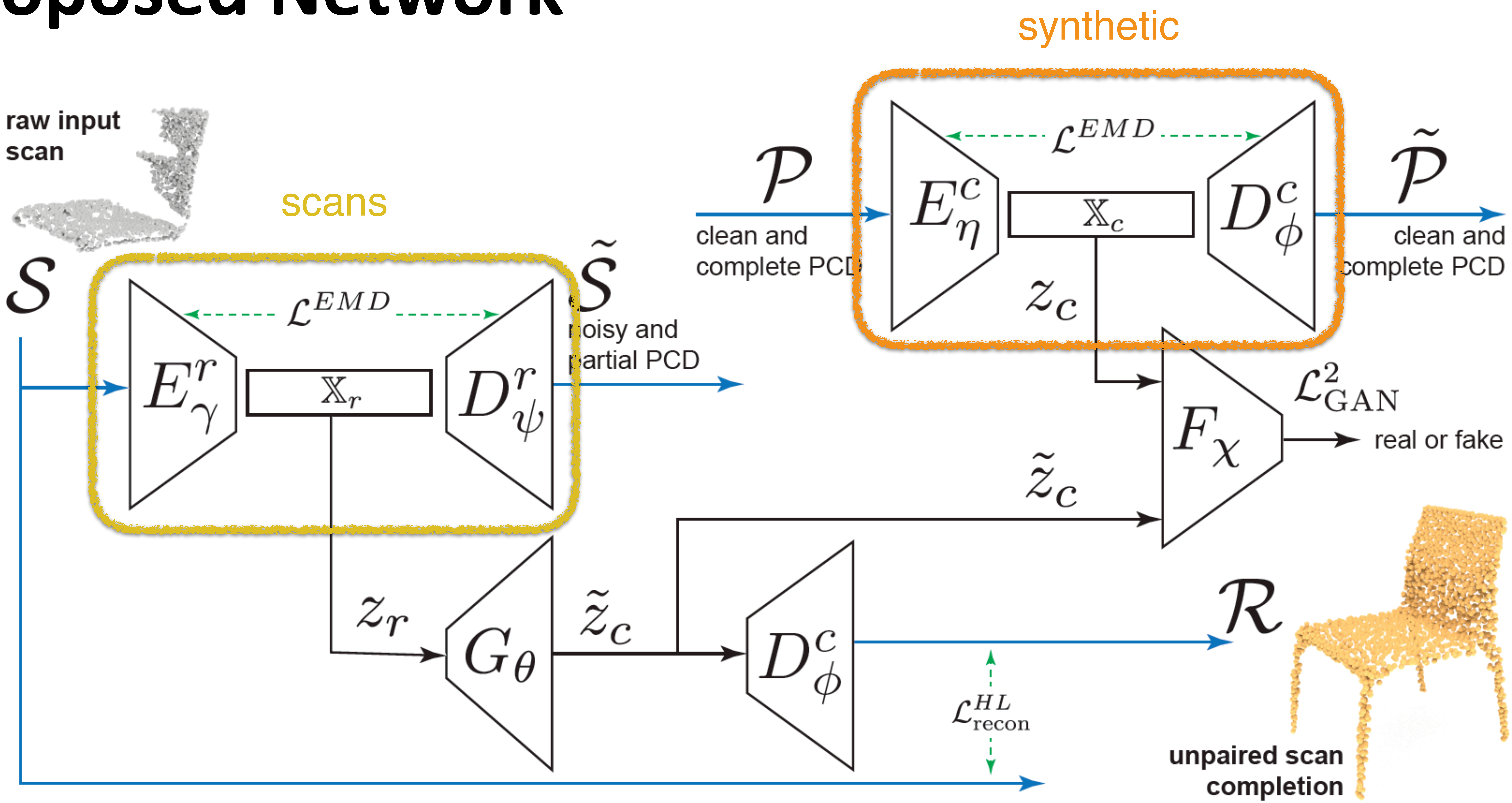
Proposed Network



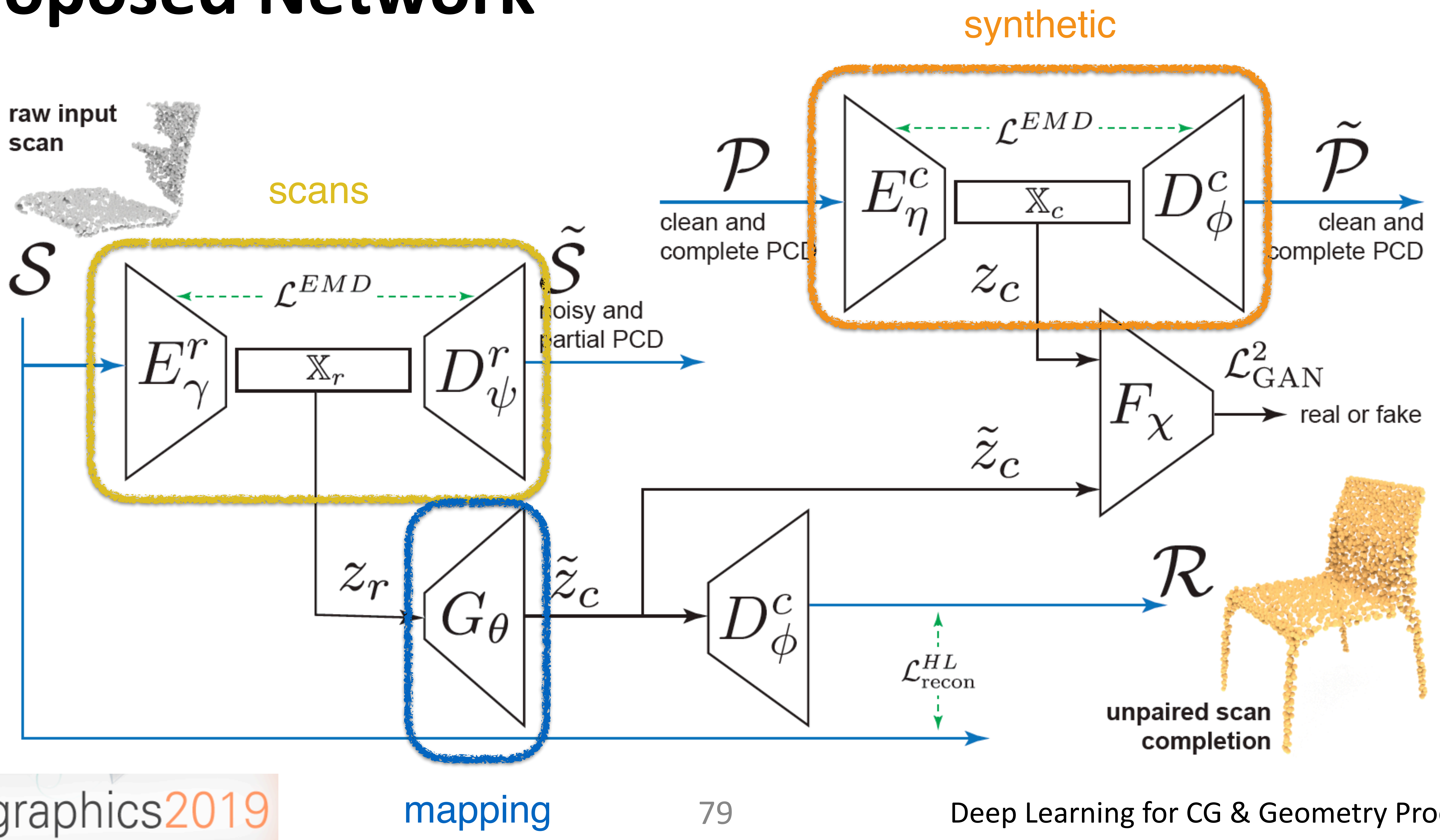
Proposed Network



Proposed Network



Proposed Network



Loss Terms

$$\mathcal{L}_F(\chi) \;= \mathbb{E}_{x\sim p_{\text{clean-complete}}}\left[F_\chi\left(E^c_\eta(x)\right) - 1\right]^2 + \mathbb{E}_{y\sim p_{\text{noisy-partial}}}\left[F_\chi\left(G_\theta(E^r_\gamma(y))\right)\right]^2$$

$$\mathcal{L}_G(\theta) \;= \alpha\mathbb{E}_{y\sim p_{\text{noisy-partial}}}\left[F_\chi\left(G_\theta(E^r_\gamma(y))\right) - 1\right]^2 + \beta\mathcal{L}^{\text{HL}}_{\text{recon}}(\mathcal{S}, D^c_\psi(\tilde{z}_c))$$

Loss Terms

$$\mathcal{L}_F(\chi) \; = \mathbb{E}_{x\sim p_{\text{clean-complete}}}\left[F_\chi\left(E^c_\eta(x)\right) - 1\right]^2 + \mathbb{E}_{y\sim p_{\text{noisy-partial}}}\left[F_\chi\left(G_\theta(E^r_\gamma(y))\right)\right]^2$$

$$\mathcal{L}_G(\theta) \; = \alpha\mathbb{E}_{y\sim p_{\text{noisy-partial}}}\left[F_\chi\left(G_\theta(E^r_\gamma(y))\right) - 1\right]^2 + \beta\boxed{\mathcal{L}^{\text{HL}}_{\text{recon}}(\mathcal{S}, D^c_\psi(\tilde{z}_c))}$$

reconstruction

Loss Terms

$$\mathcal{L}_F(\chi) = \mathbb{E}_{x\sim p_{\text{clean-complete}}}\left[F_\chi\left(E^c_\eta(x)\right) - 1\right]^2 + \mathbb{E}_{y\sim p_{\text{noisy-partial}}}\left[F_\chi\left(G_\theta(E^r_\gamma(y))\right)\right]^2$$

$$\mathcal{L}_G(\theta) = \alpha\mathbb{E}_{y\sim p_{\text{noisy-partial}}}\left[F_\chi\left(\boxed{G_\theta}(E^r_\gamma(y))\right) - 1\right]^2 + \beta\boxed{\mathcal{L}^{\text{HL}}_{\text{recon}}(\mathcal{S}, D^c_\psi(\tilde{z}_c))}$$

mapper

reconstruction

Without and With Reconstruction Loss

input point set

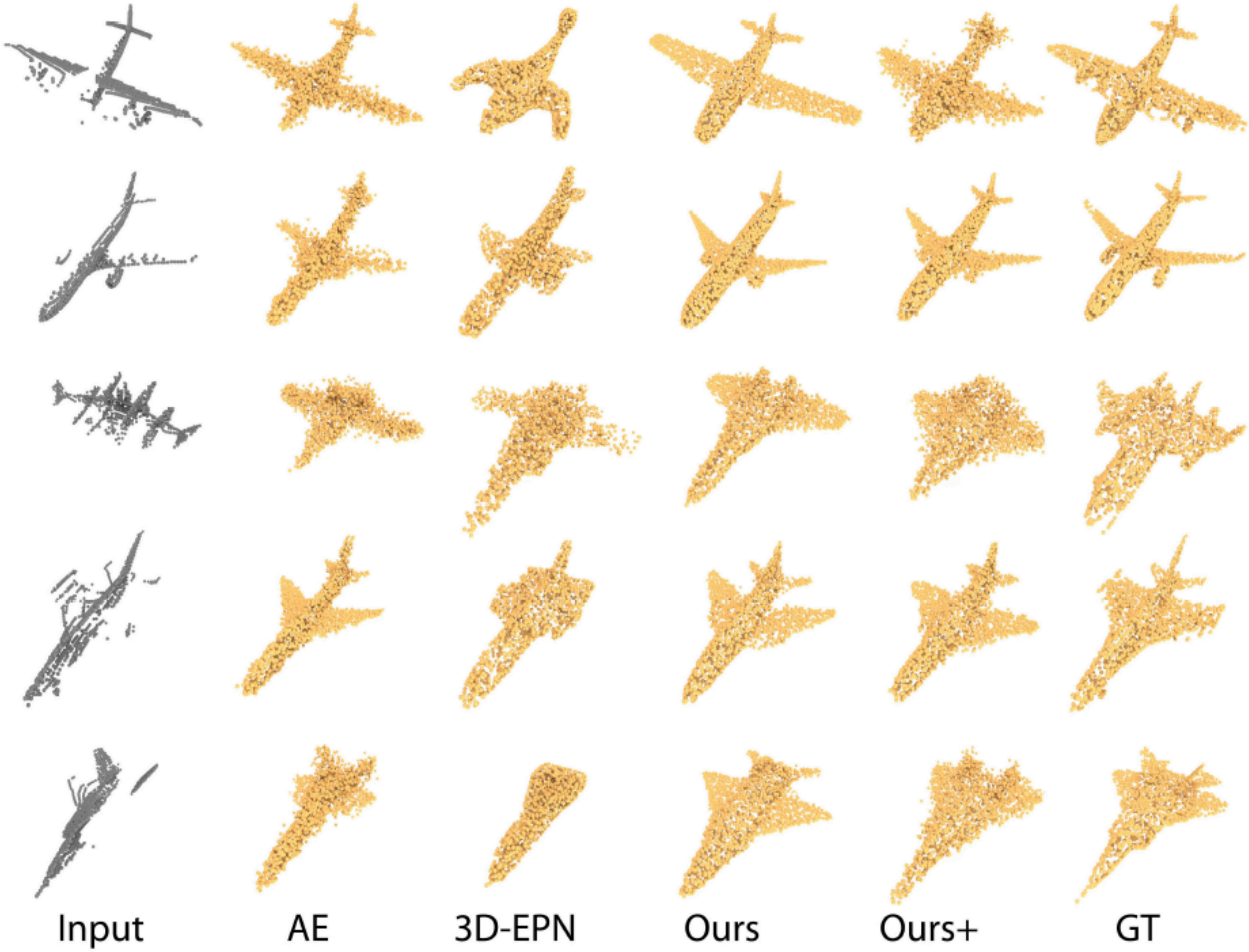
Without and With Reconstruction Loss

input point set

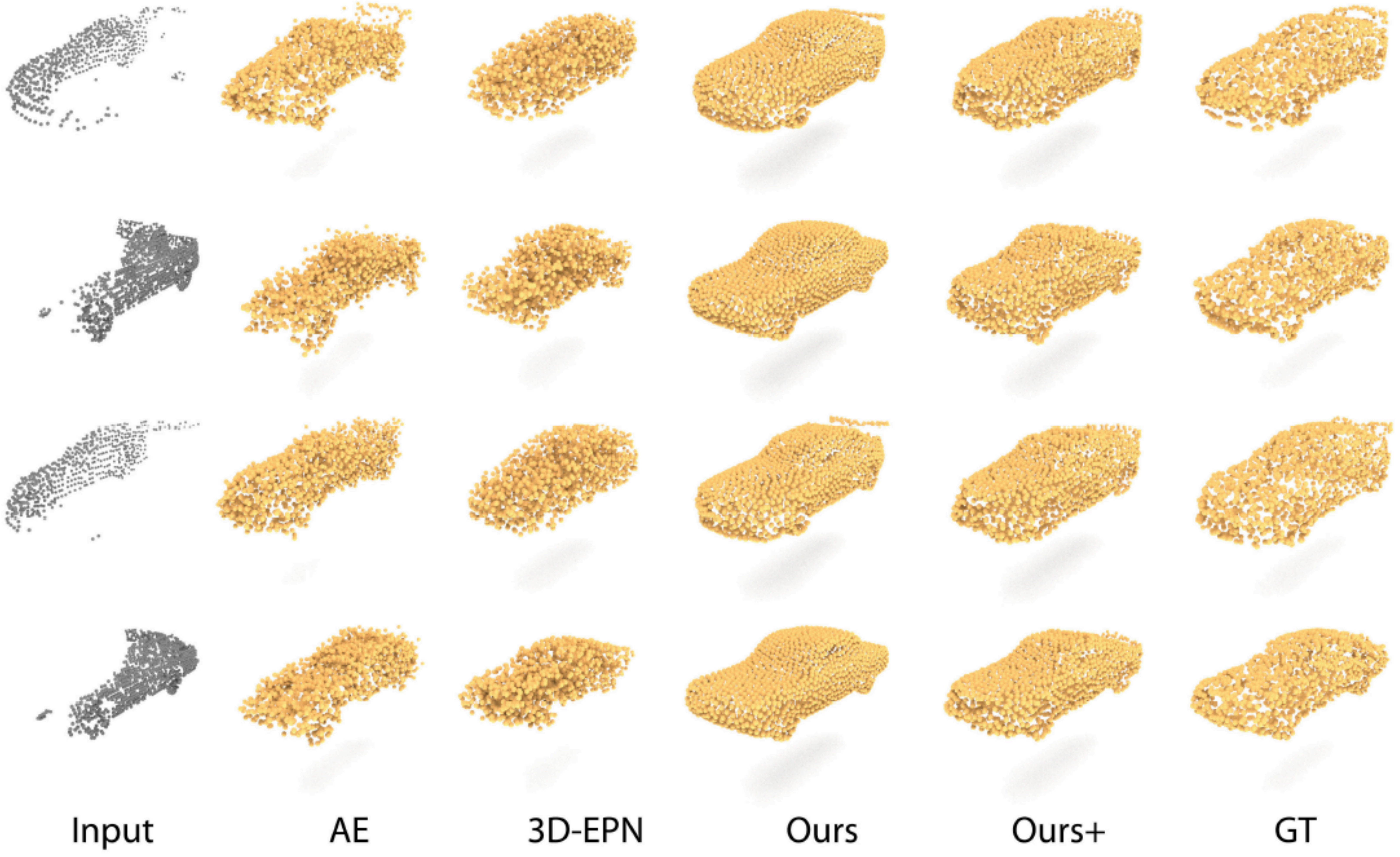
completion without HL

Without and With Reconstruction Loss

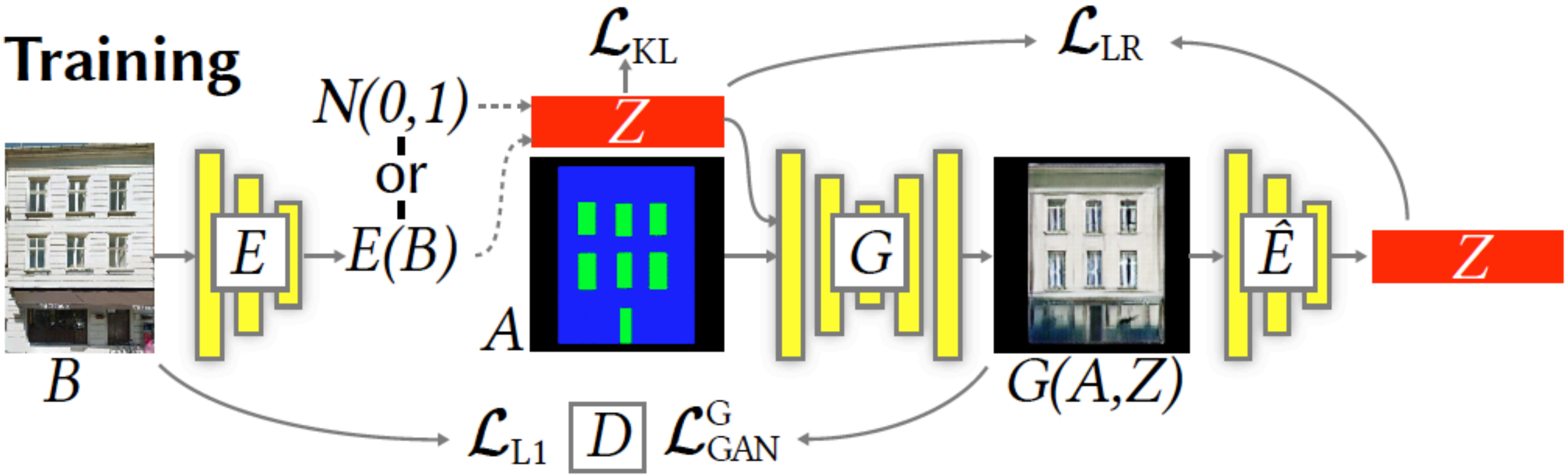
Evaluation on Synthetic Data



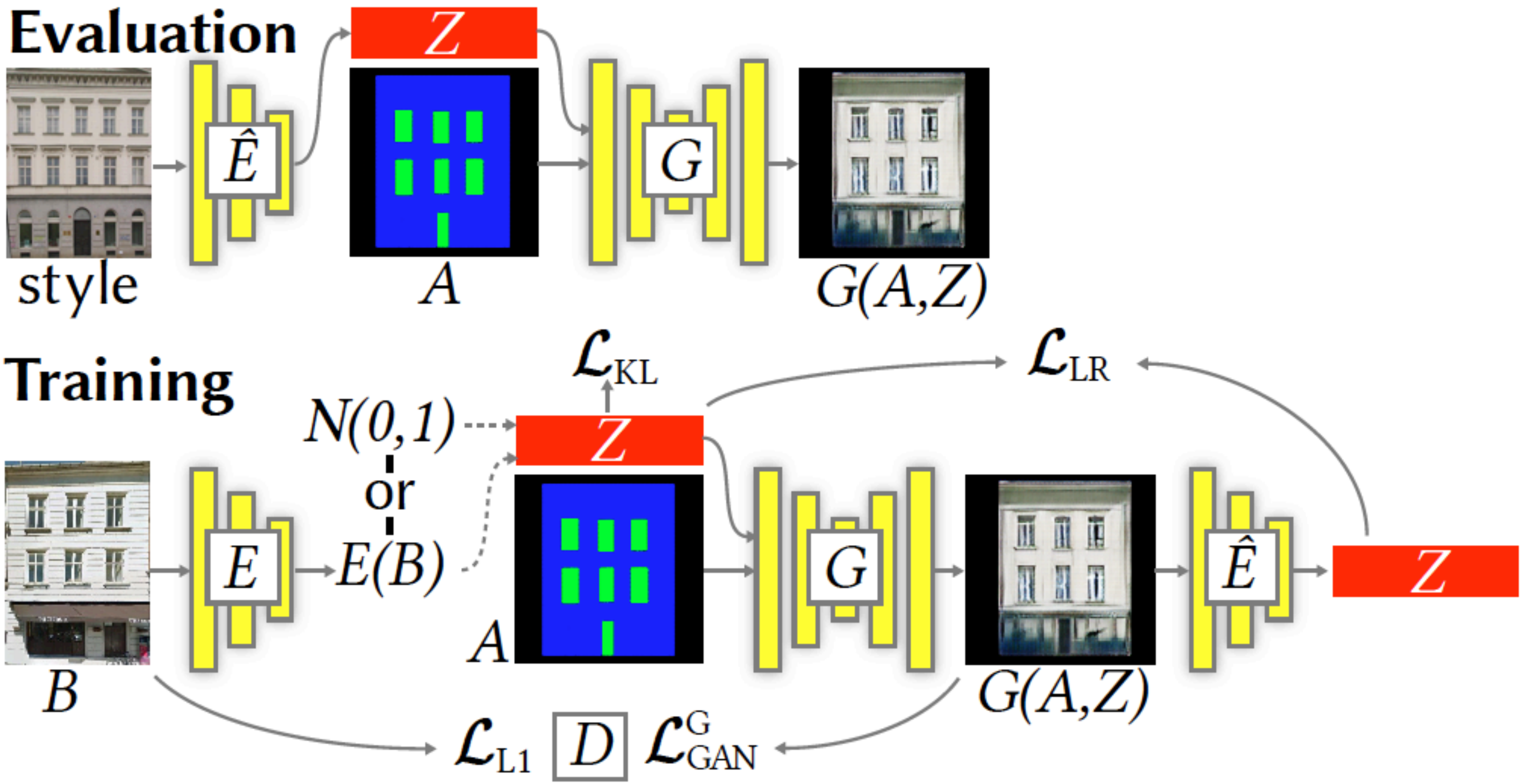
Evaluation on Synthetic Data



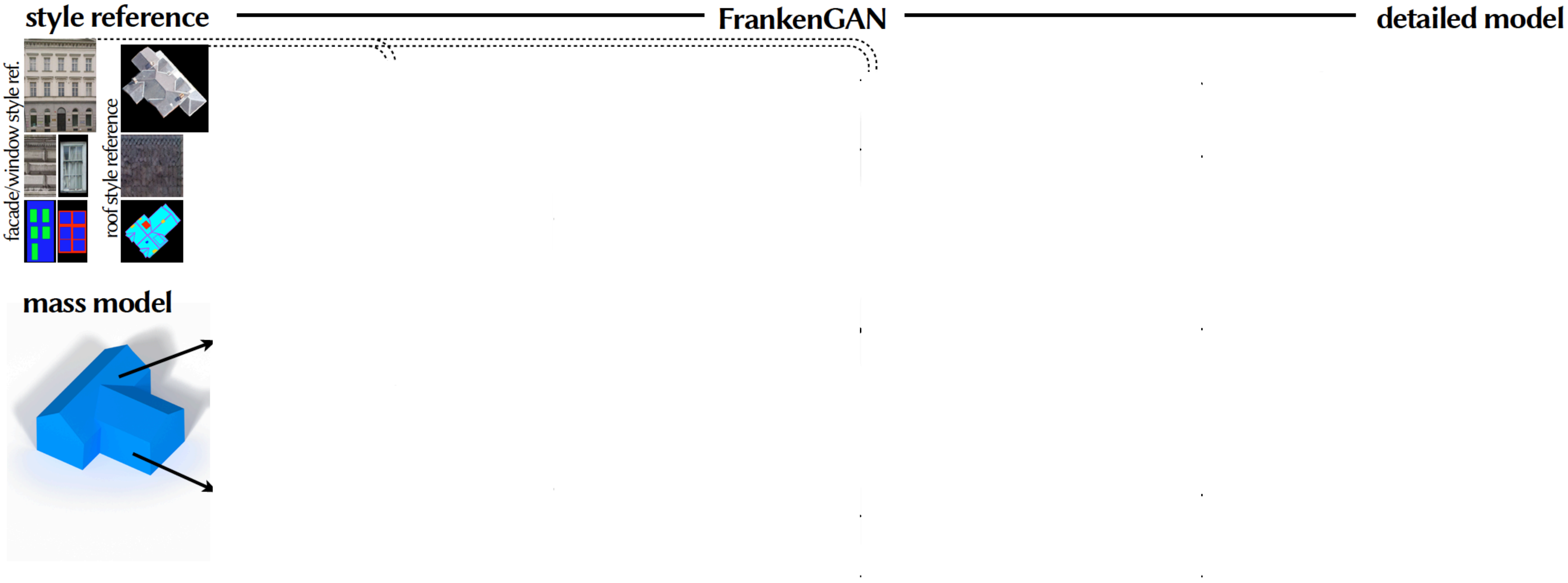
Building GAN Block: Consistent Latent Space



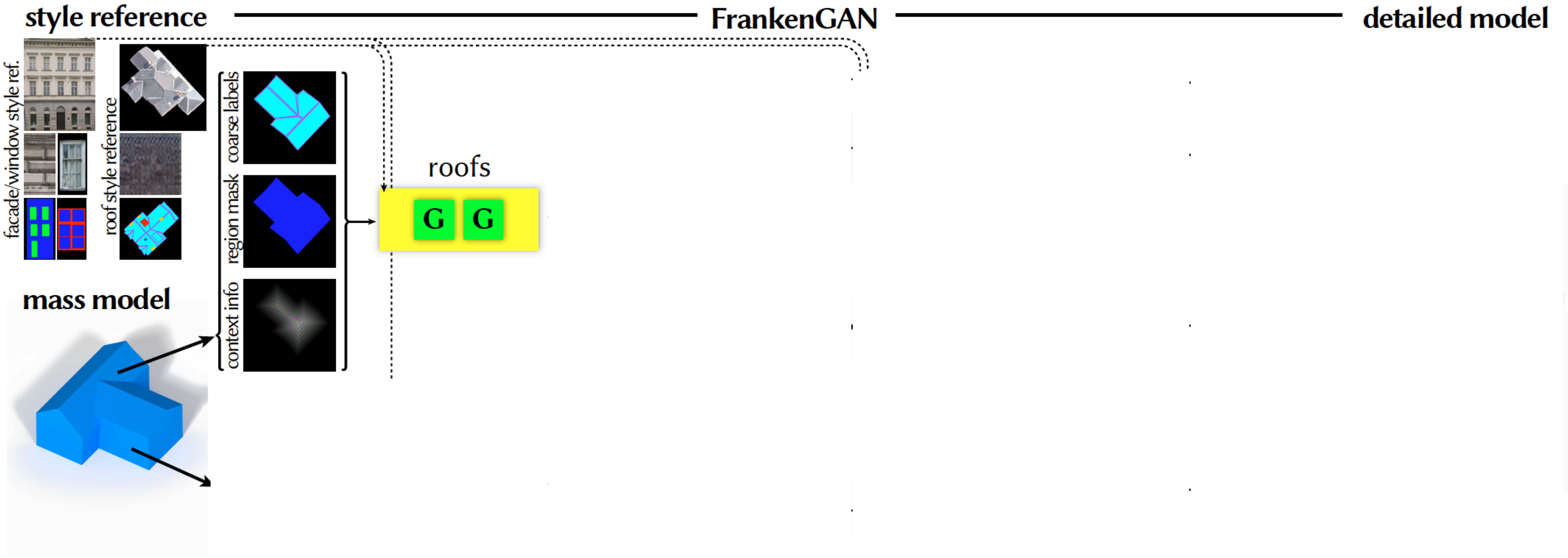
Building GAN Block: Consistent Latent Space



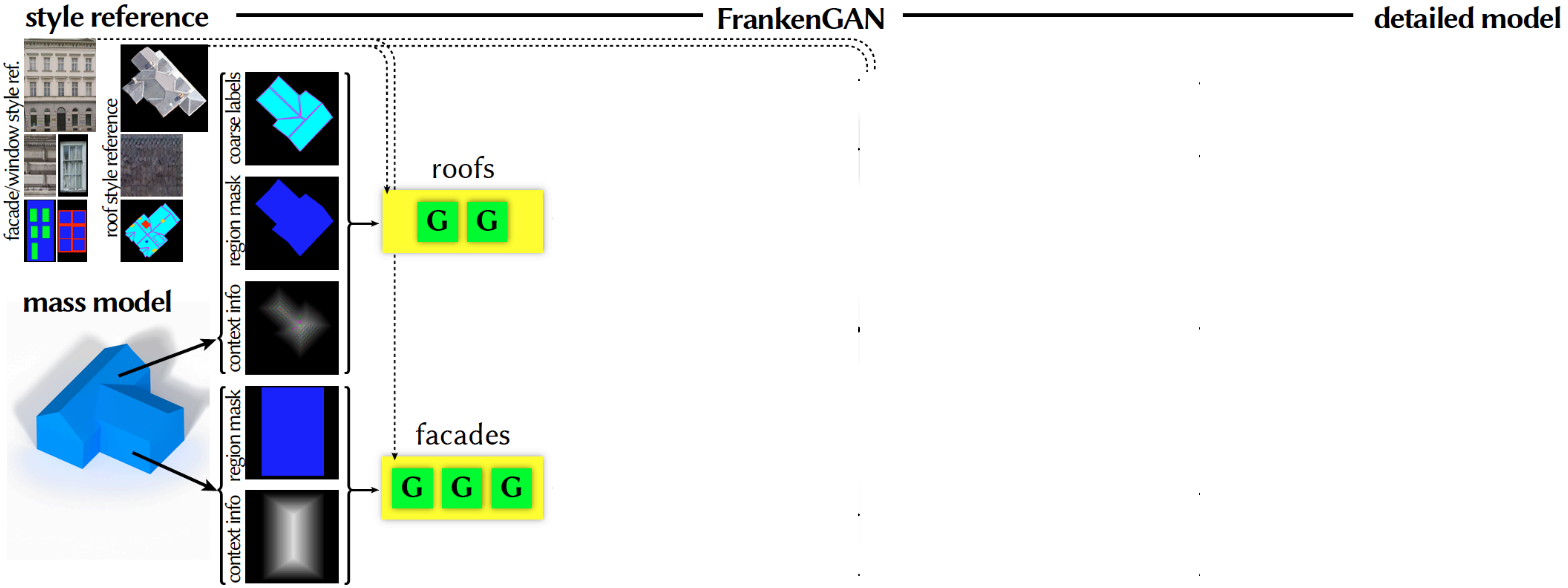
FrankenGAN Architecture



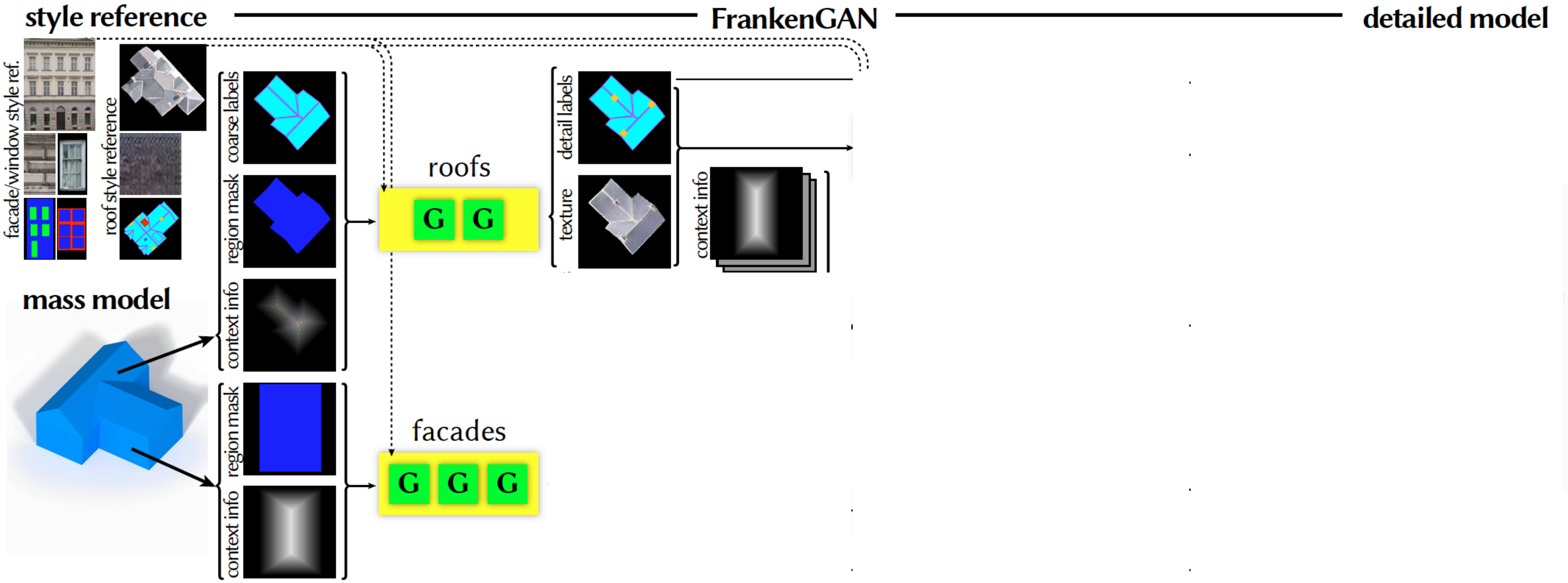
FrankenGAN Architecture



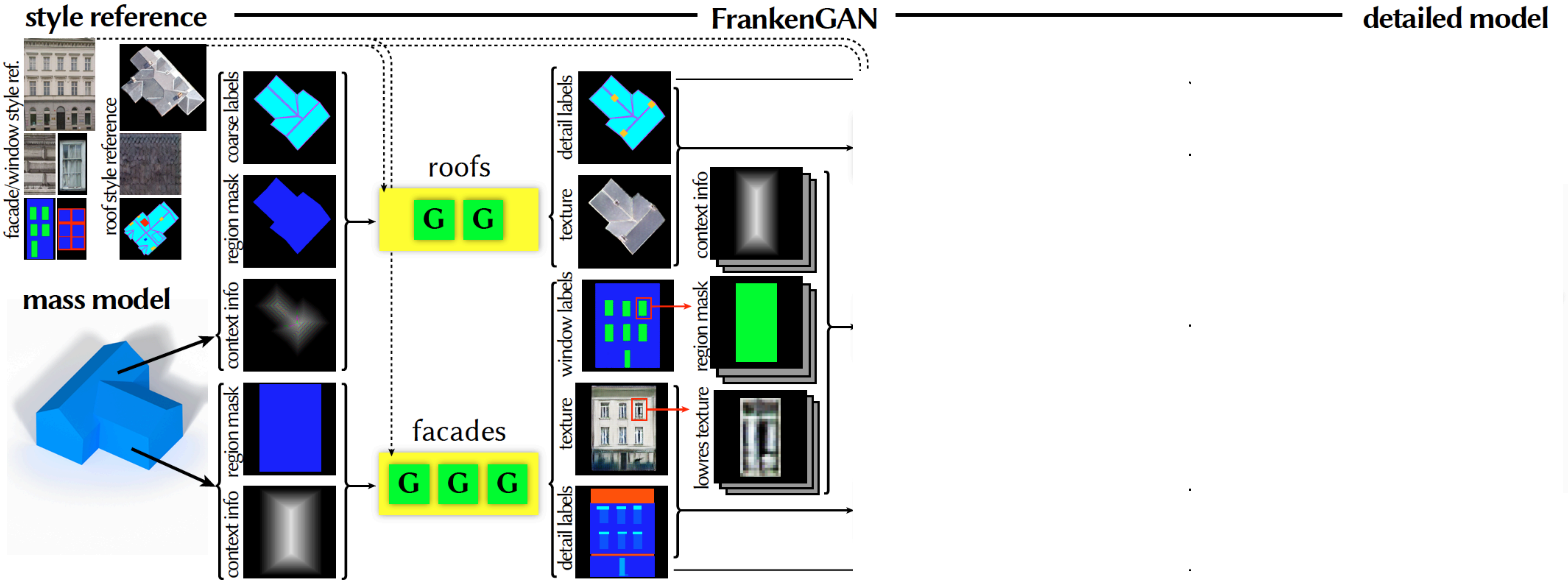
FrankenGAN Architecture



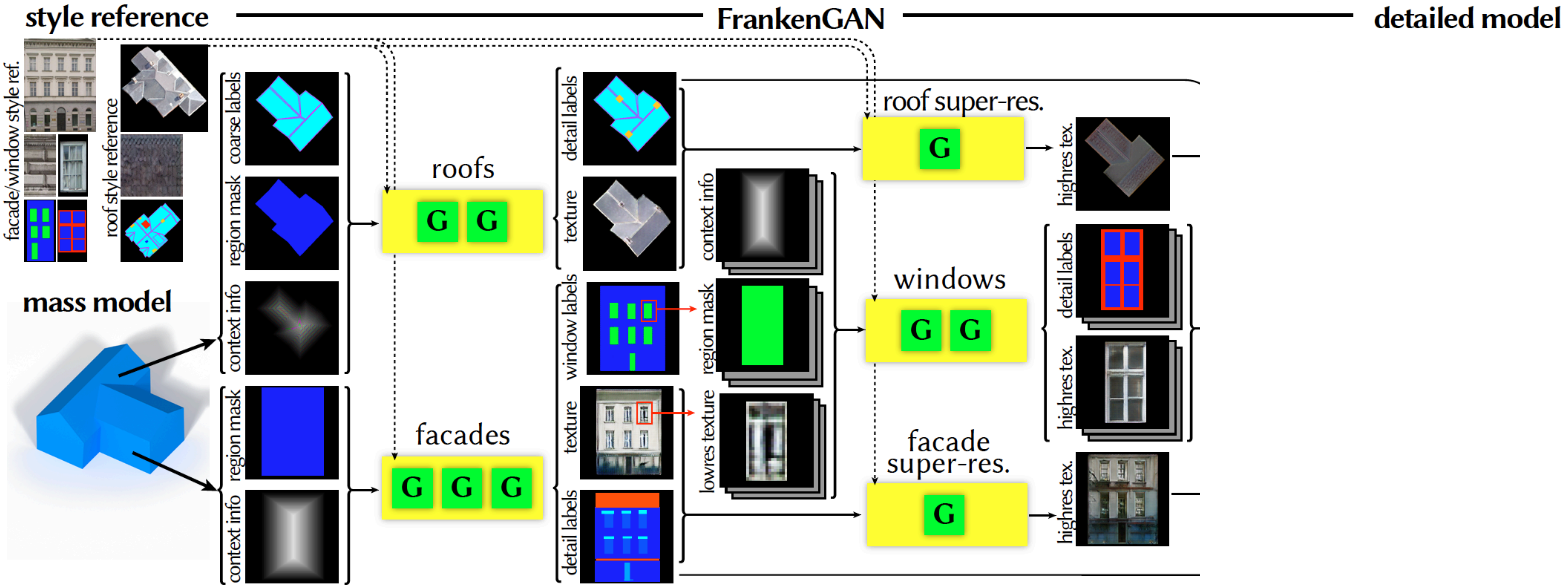
FrankenGAN Architecture



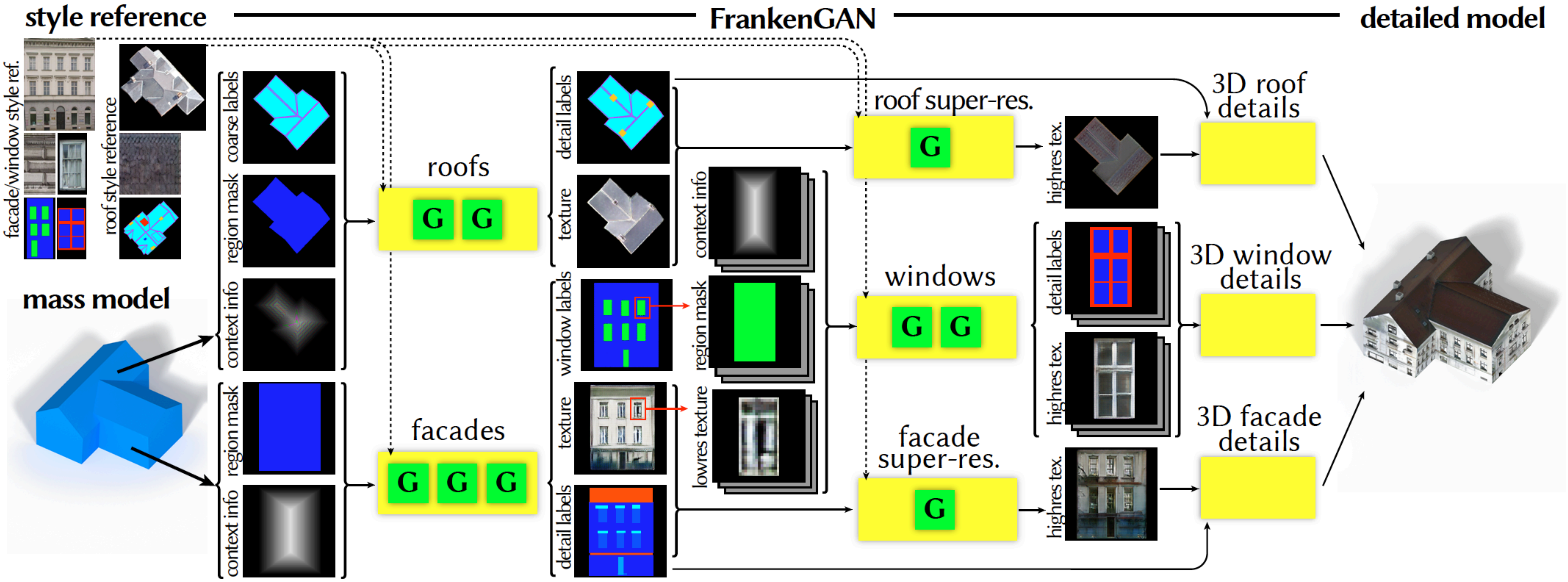
FrankenGAN Architecture



FrankenGAN Architecture



FrankenGAN Architecture

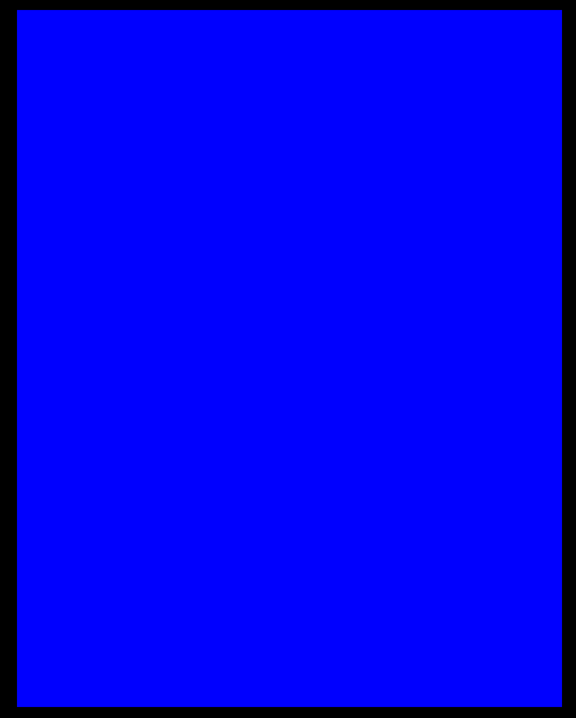



chordatlas blankland: joint distriubtion editor
file
name my dist
+ - < >
p 0.5
Facade Label on
fixZ: building
Facade Gree... on
fixZ: building
Facade Super on
fixZ: building
Facade Tex on
fixZ: building
Panes Label on
fixZ: building
Panes Tex on
fixZ: building
Roof Tex on
fixZ: building
Roof Super on
fixZ: building
modes:
p:
σ:
ok
10541.png
10772.png
12345.png
12855.png
14110.png
tools:
world: -48.4829, 47.2488

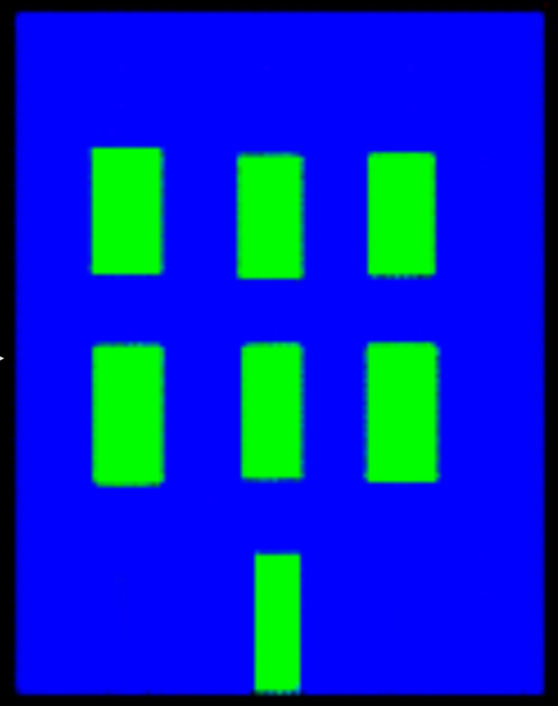

10541.png
10772.png
12345.png
12855.png
14110.png
chordatlas blankland: joint distriubtion editor
file
name my dist
+
-
<
>
p
0.5
Facade Label
on
fixZ: building
Facade Gree...
on
fixZ: building
Facade Super
on
fixZ: building
Facade Tex
on
fixZ: building
Panes Label
on
fixZ: building
Panes Tex
on
fixZ: building
Roof Tex
on
fixZ: building
Roof Super
on
fixZ: building
modes:
+
p:
x
σ:
p:
x
σ:
ok
tools:
world: -48.4829, 47.2488
EPSG:27700: 528681.9034, 18

FrankenGAN: ‘Procedural’ Steps

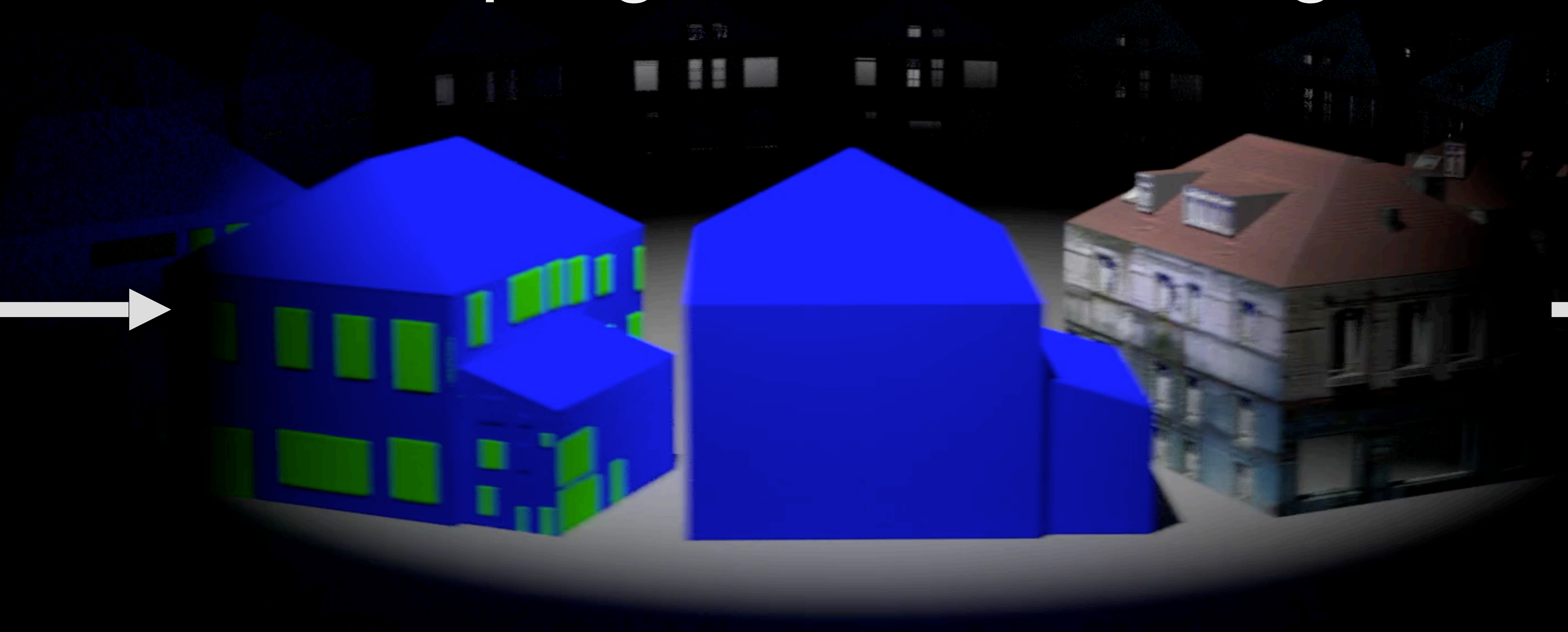
input

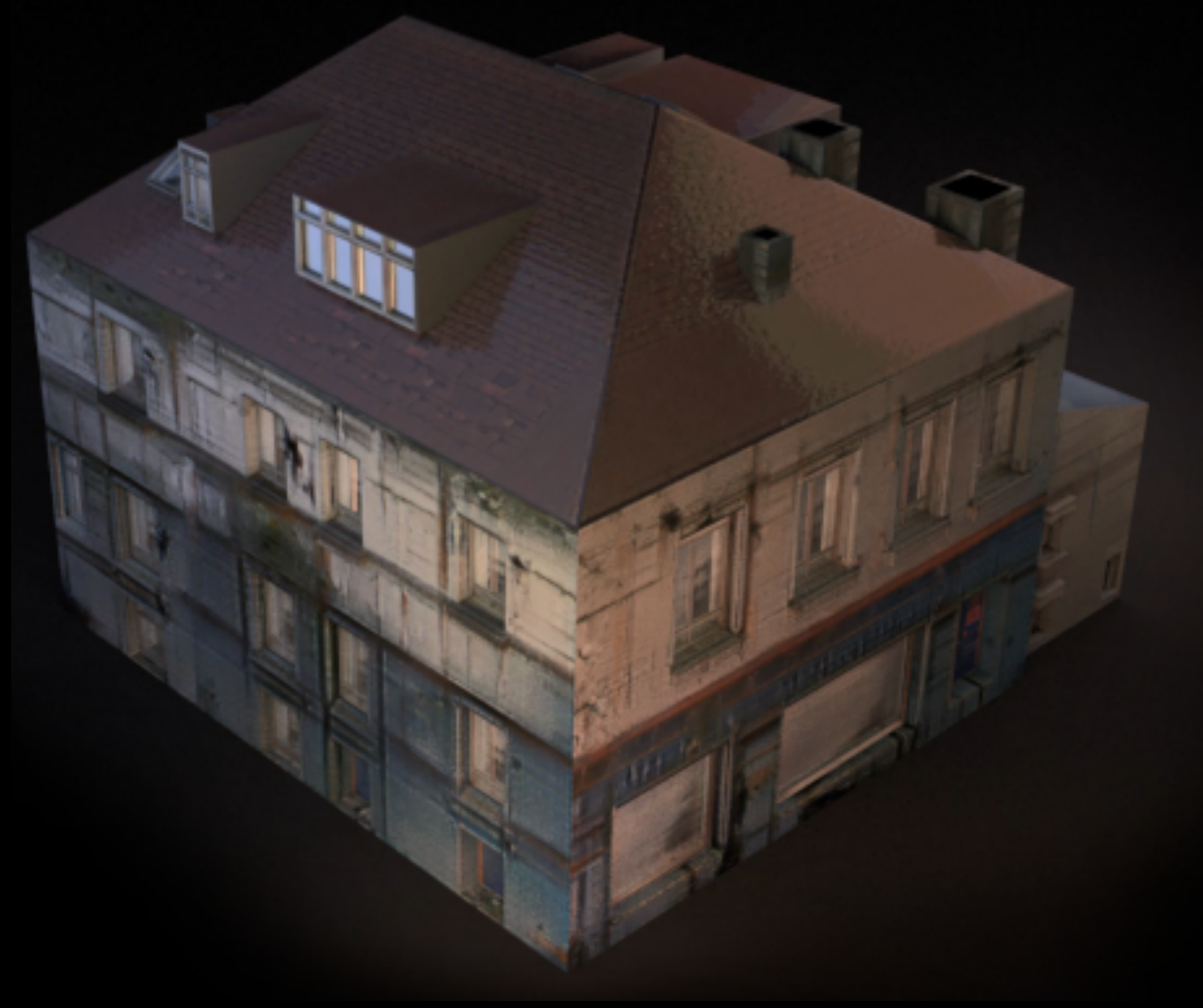


1st step:
windor
layout



… progressive modeling …

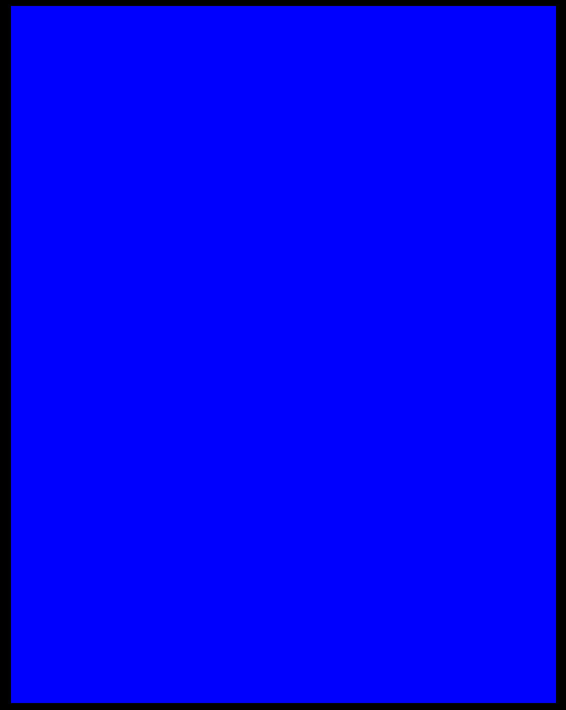




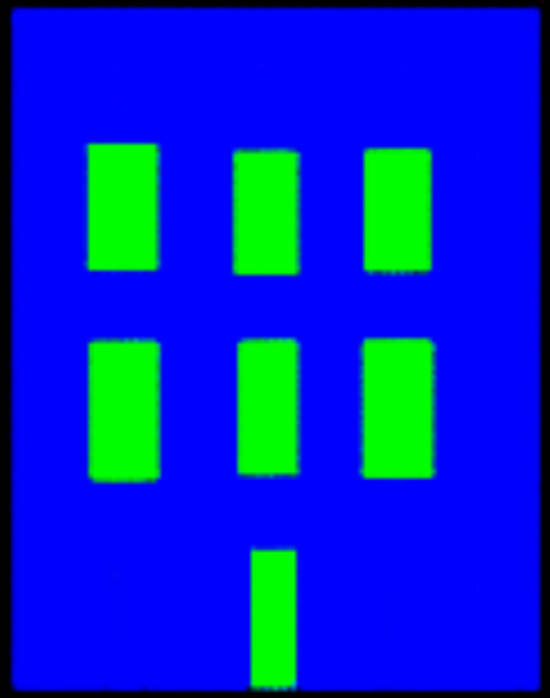
FrankenGAN: ‘Procedural’ Steps

FrankenGAN: ‘Procedural’ Steps

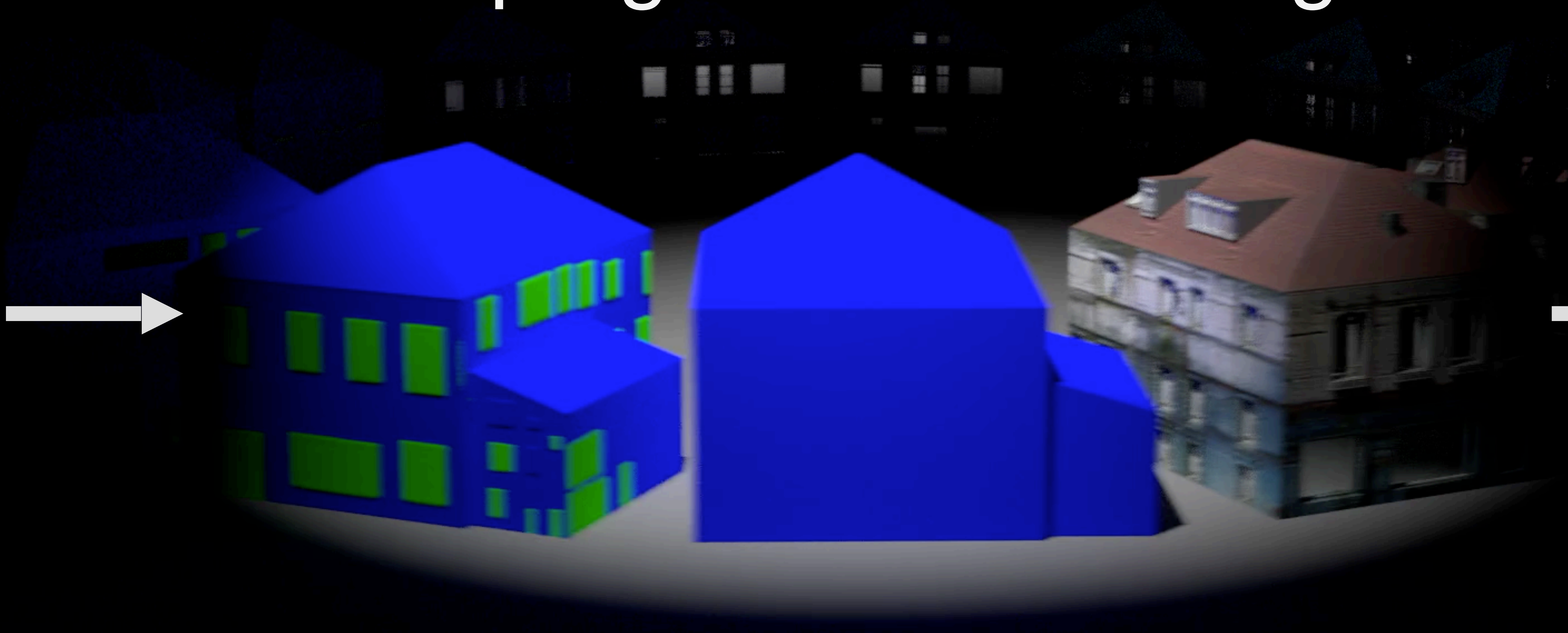
input



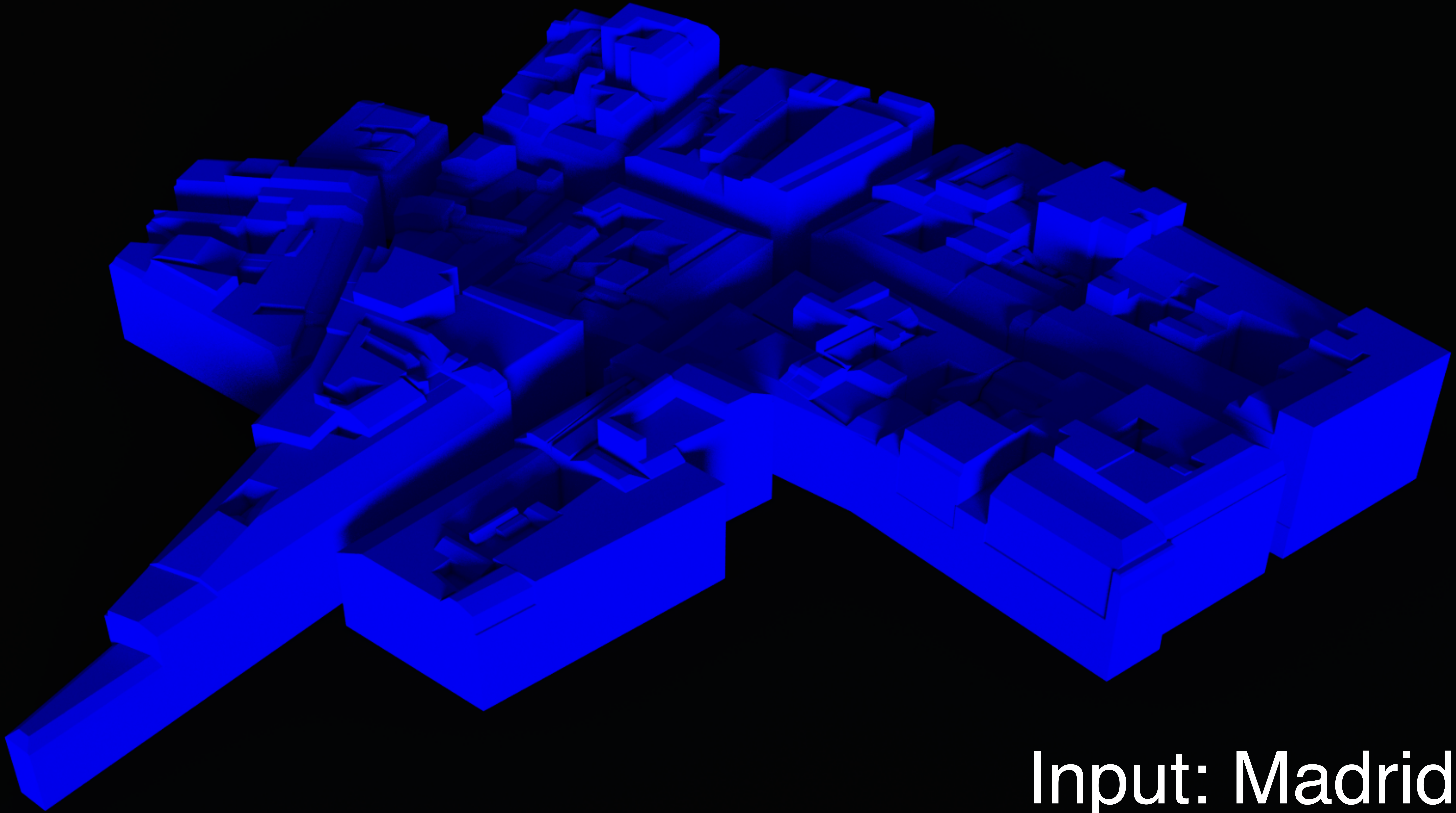
1st step:
windor
layout



… progressive modeling …



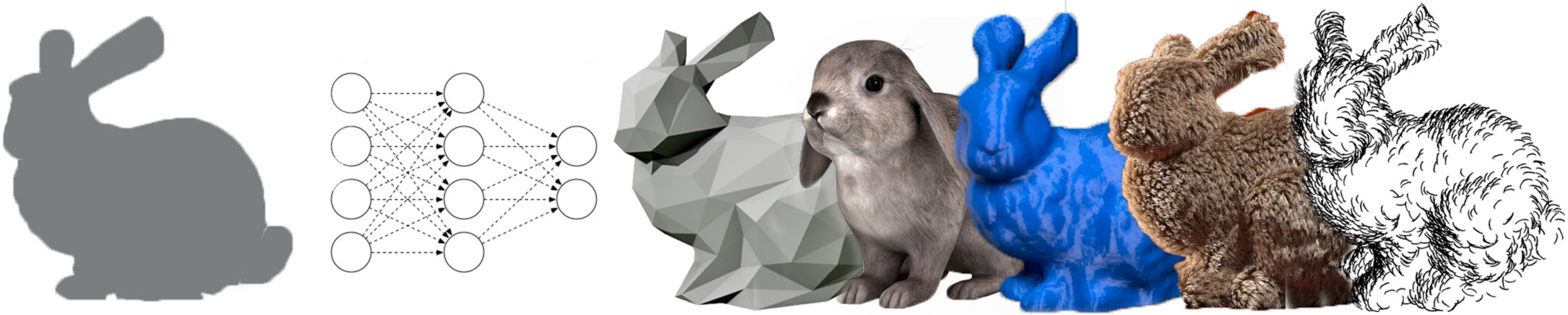
- No mode collapse
- Style and scale control
- Regularisation
- Manual edits
- High-res



Input: Madrid

Output: Madrid

Course Information (slides/code/comments)



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

Scan me