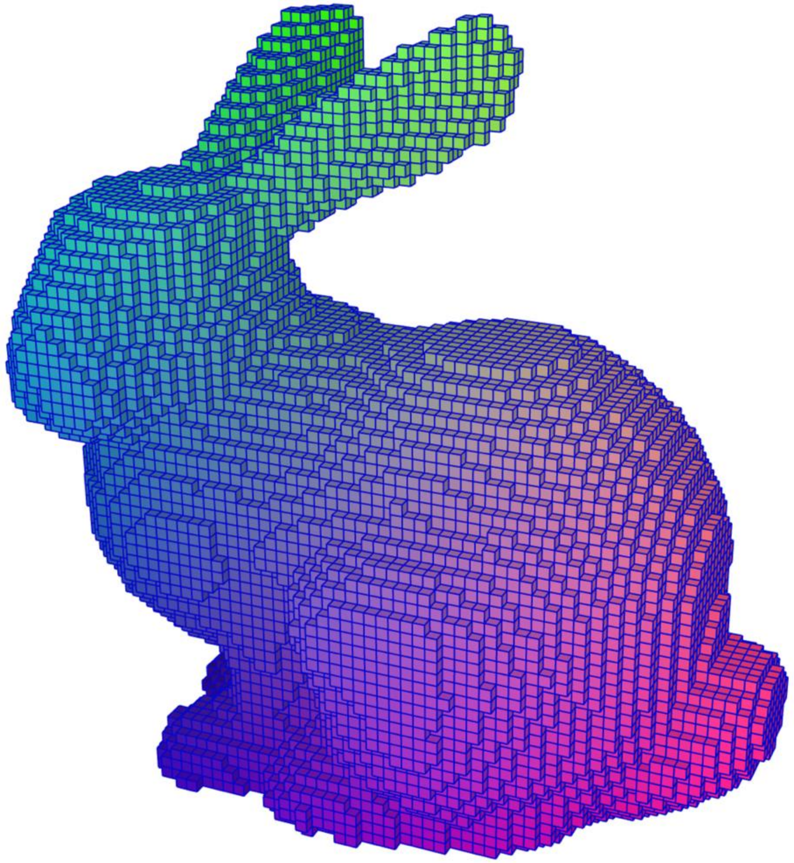
Deep 3D Vision

Xiaolong Wang

This Class – 3D representations

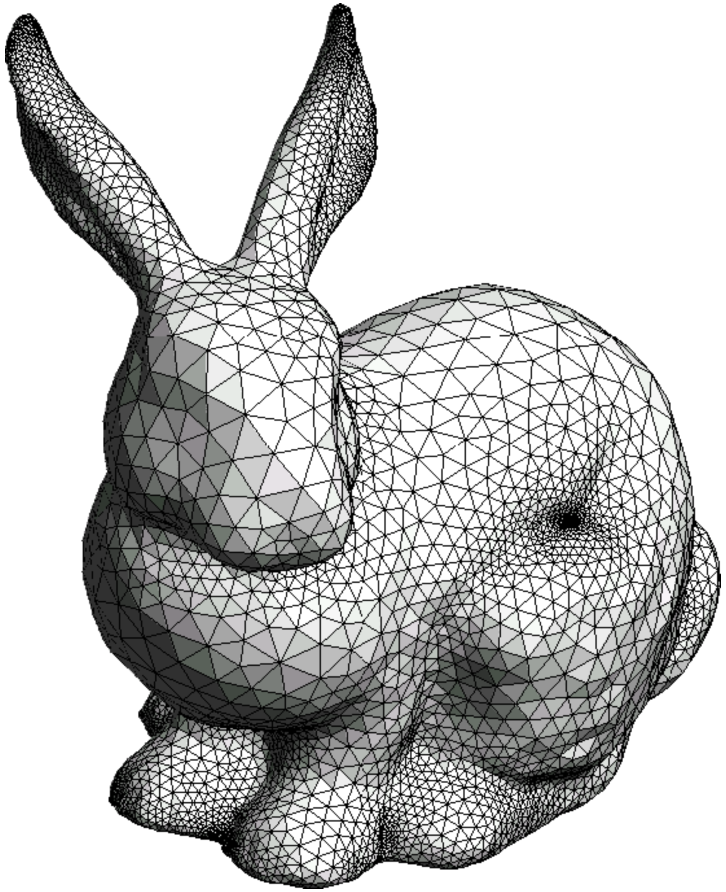
- Volumetric representation
- Mesh
- Point cloud
- Implicit functions

How to represent objects in 3D?



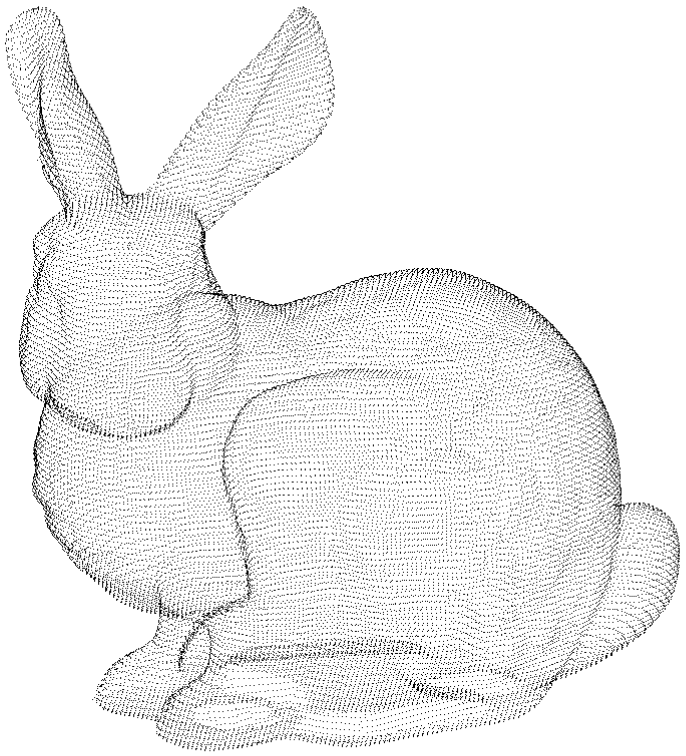
- Volumetric
- Mesh
- Point cloud
- Implicit function

How to represent objects in 3D?



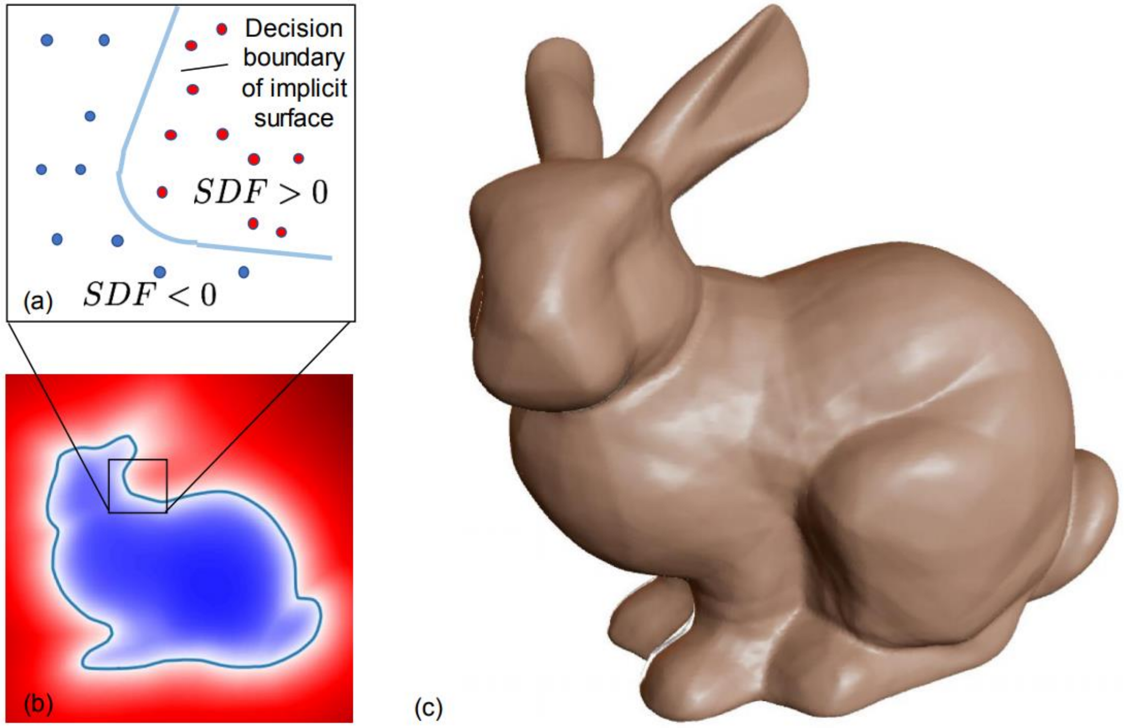
- Volumetric
- Mesh
- Point cloud
- Implicit function

How to represent objects in 3D?



- Volumetric
- Mesh
- Point cloud
- Implicit function

How to represent objects in 3D?



- Volumetric
- Mesh
- Point cloud
- Implicit function

1.Voxel: Discretizing into grids

Voxels

Pros:

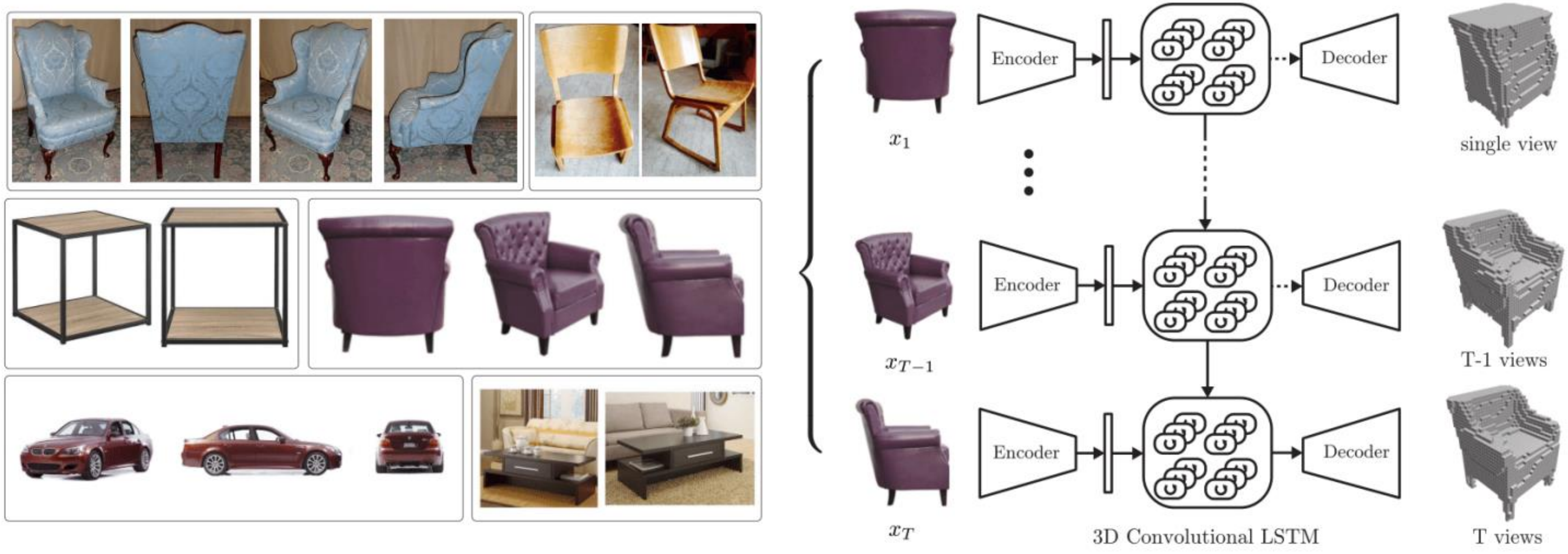
- *Easy to process with Neural Networks*

Cons:

- *Cubic memory*
- *Limited resolution*

3D-R2N2

Given single/multi-view images, output 3D voxel occupancies

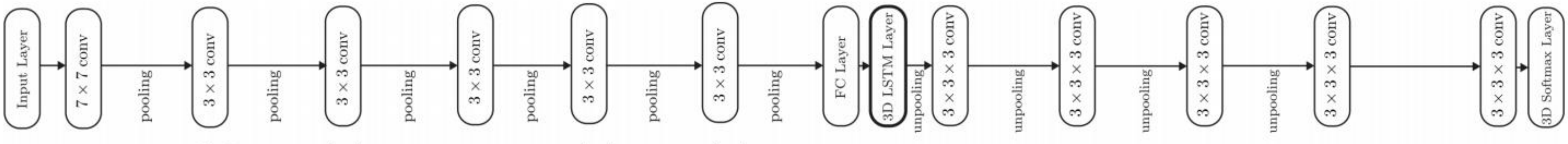


(a) Images of objects we wish to reconstruct (b) Overview of the network

3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction, ECCV, 2016.

3D-R2N2

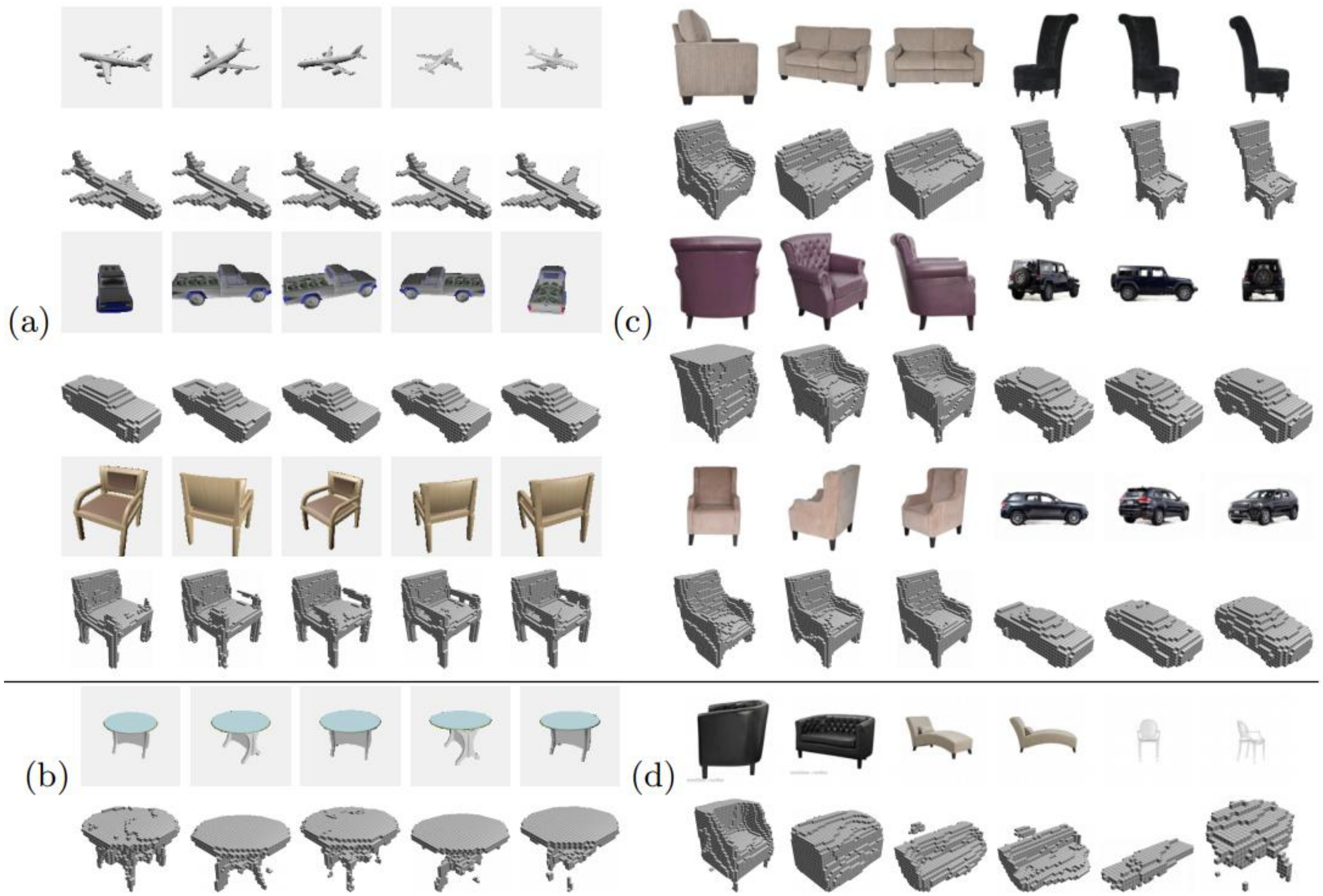
Overall architecture



3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction, ECCV, 2016.

3D-R2N2

Do not need to be trained per category.



3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction, ECCV, 2016.

2.Mesh: Discretizing into vertices

Mesh

Pros:

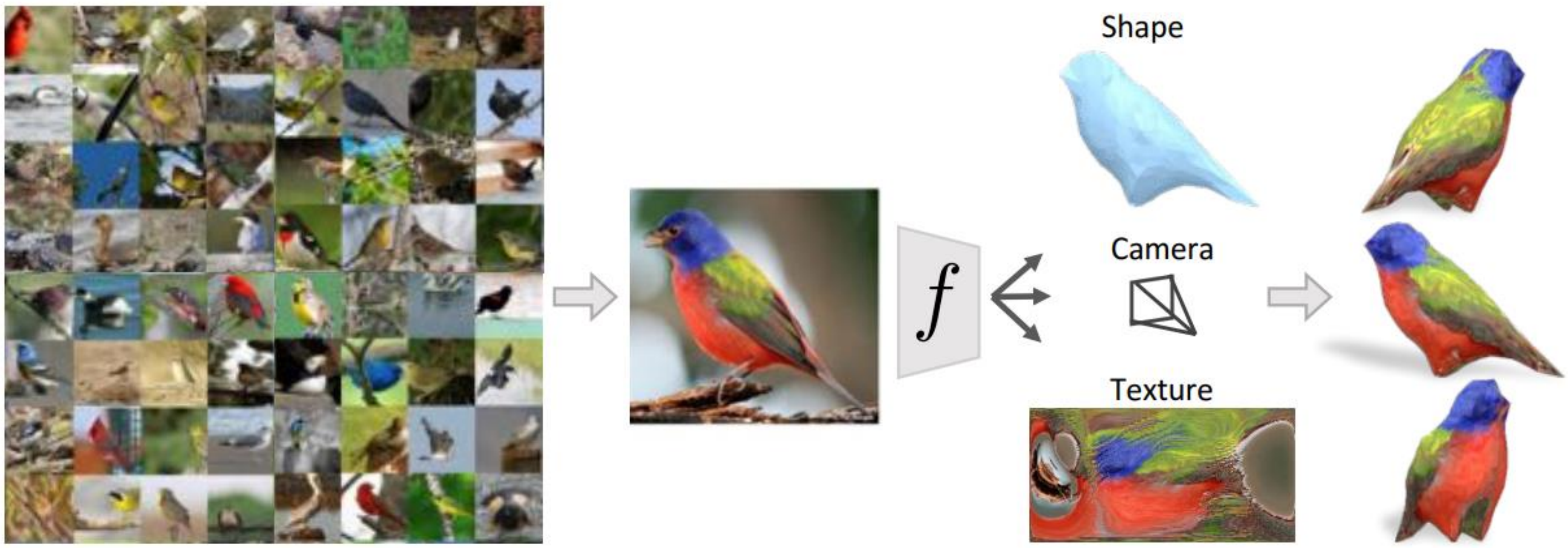
- *Easy to process with Neural Networks*

Cons:

- *Usually require class-specific template*
- *Limited by the number of vertices*

Learn mesh deformation

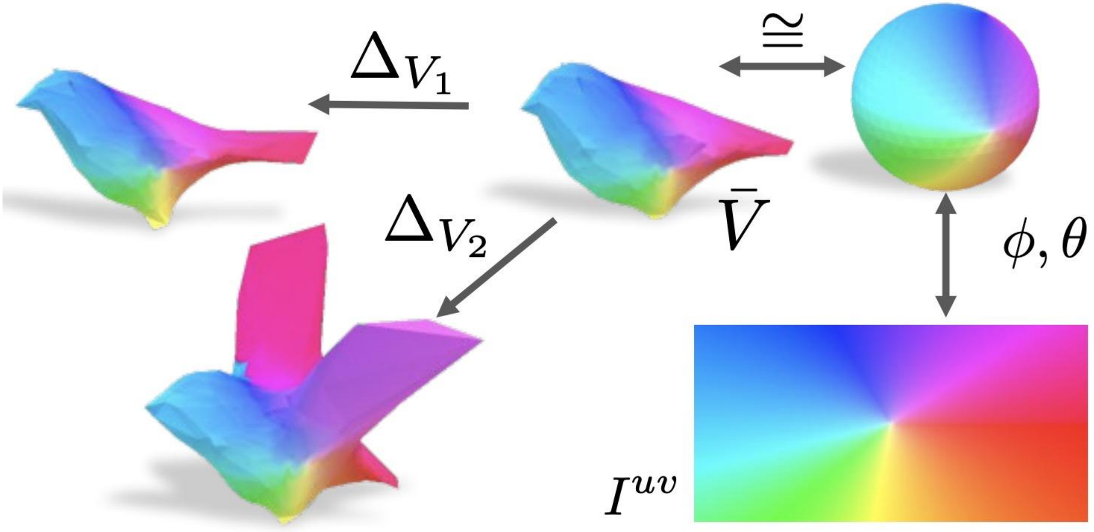
1. Given a collection of images from the same category
2. Jointly predict 3D mesh, texture and camera parameters

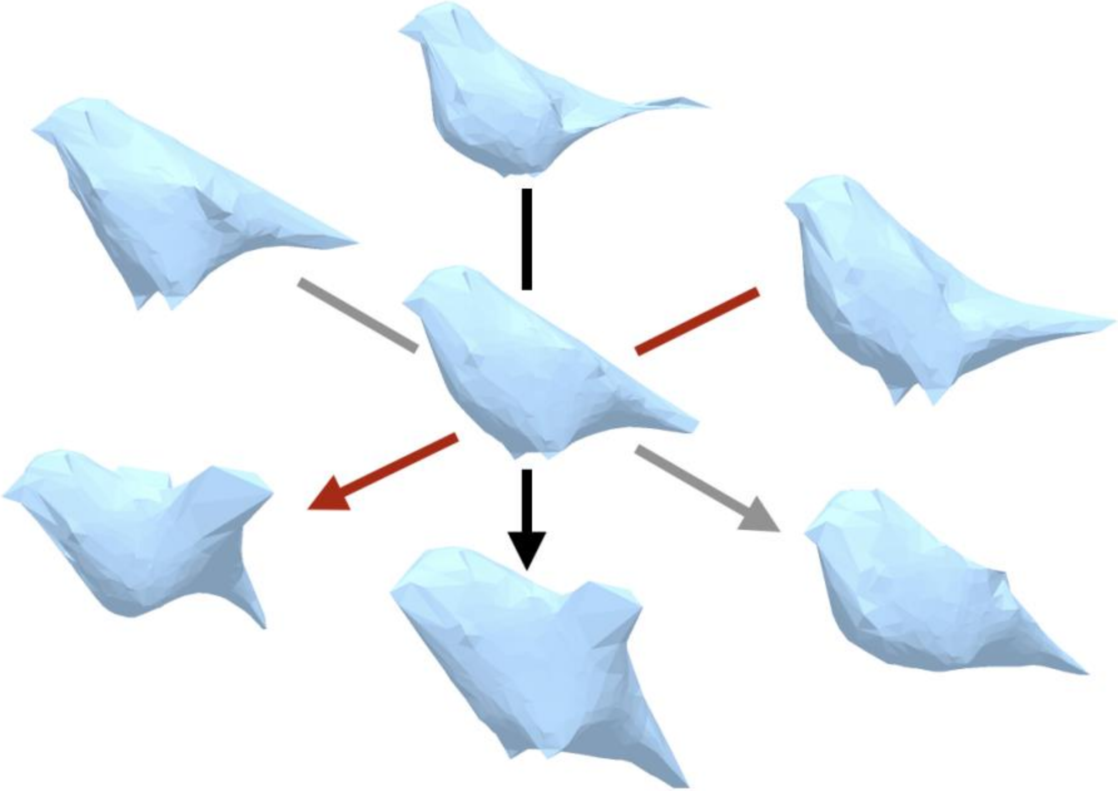


Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV, 2018.

Learn mesh deformation

Key 1: Learned deformation from mean shape

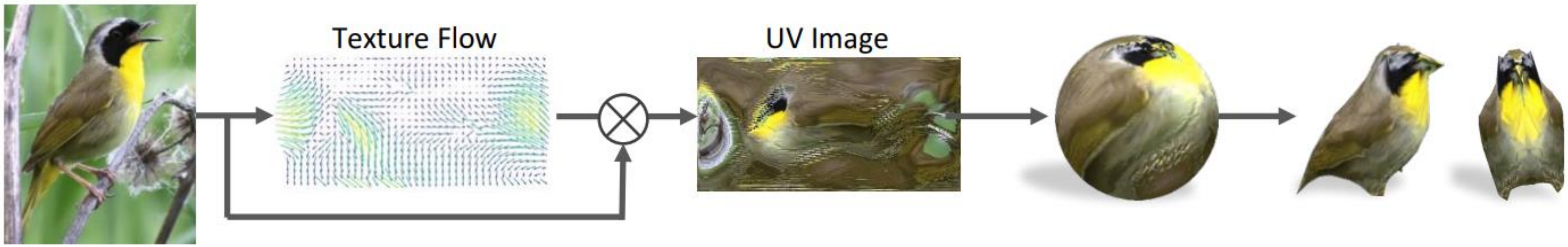


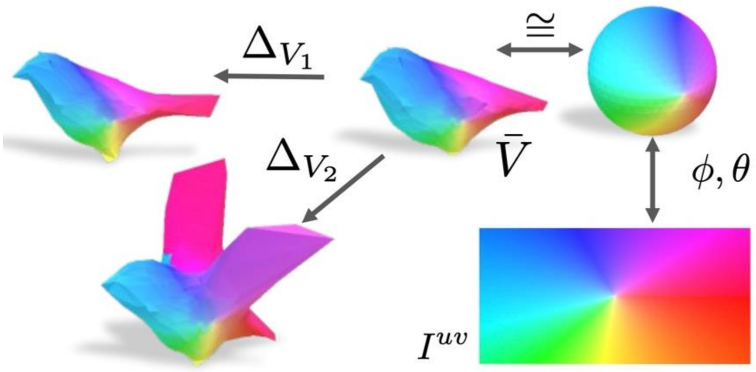


Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV, 2018.

Learn mesh deformation

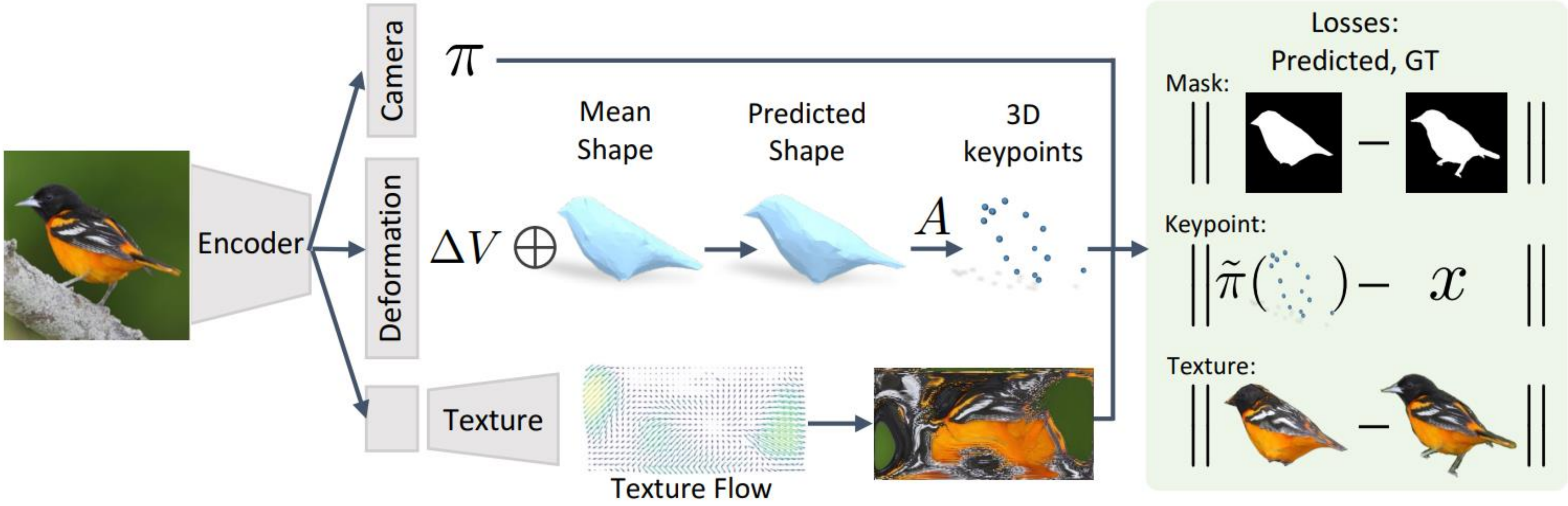
Key 2: Predict texture UV map





Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV, 2018.

Learn mesh deformation

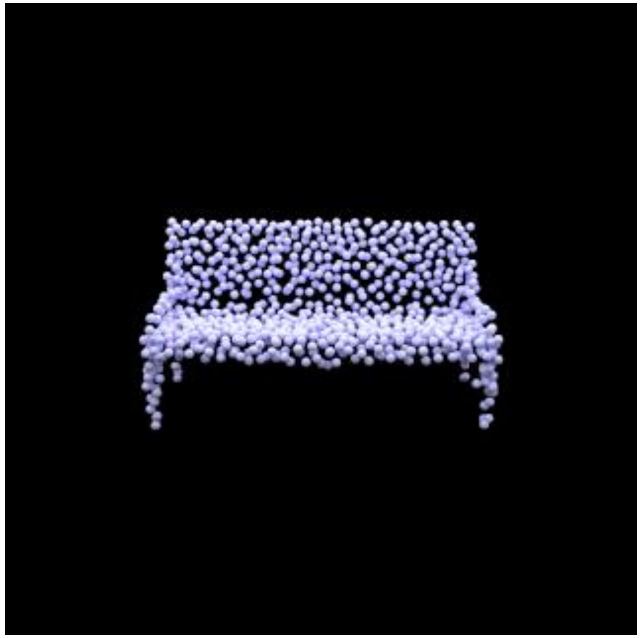


Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV, 2018.

Learn mesh deformation

Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV, 2018.

3.Point cloud: Discretizing into points



Point cloud

Pros:

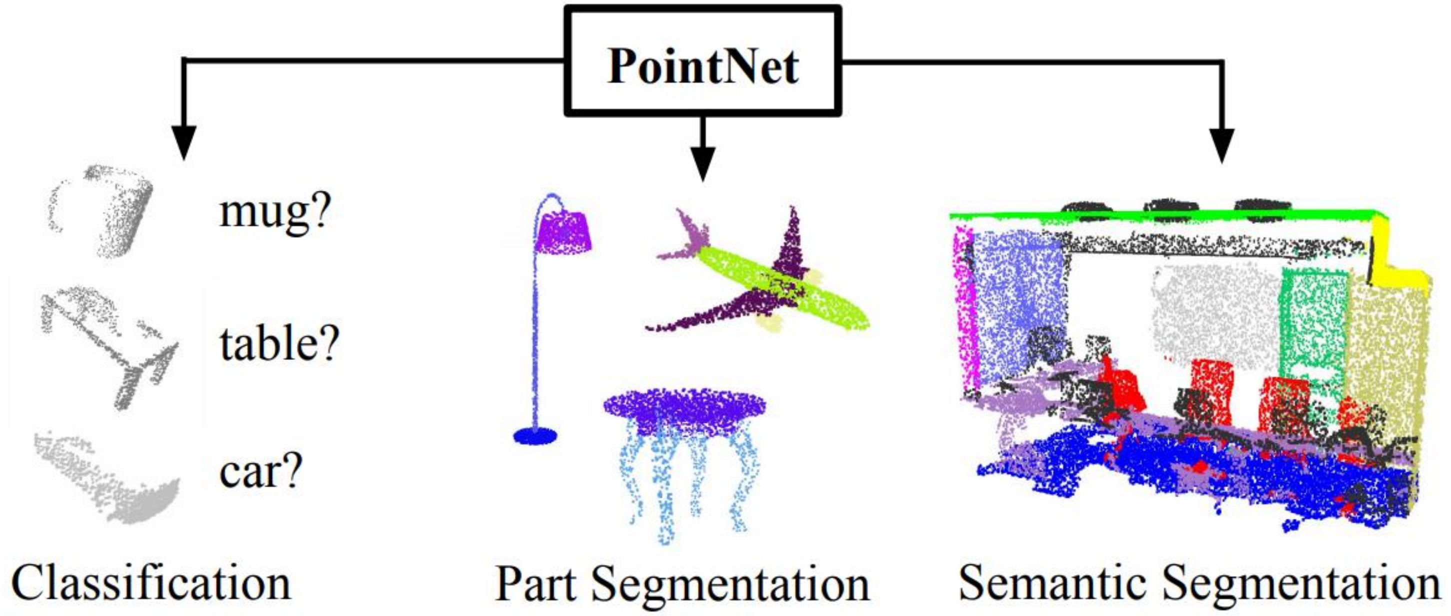
- *Flexible and memory efficient*

Cons:

- *Not model topology / connection*
- *Limited by the number of input points*

PointNet

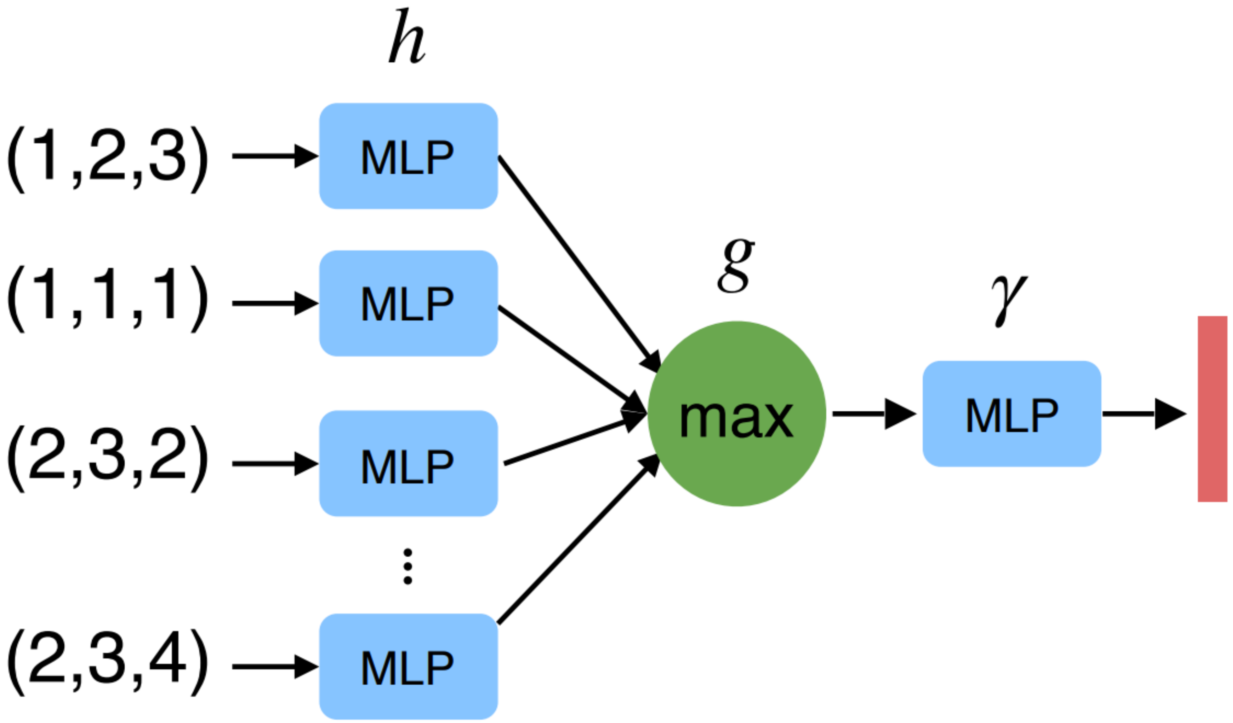
End-to-end learning given unordered and scattered point cloud, extracted features can be applied to multiple down-streaming tasks



PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR, 2017.

PointNet

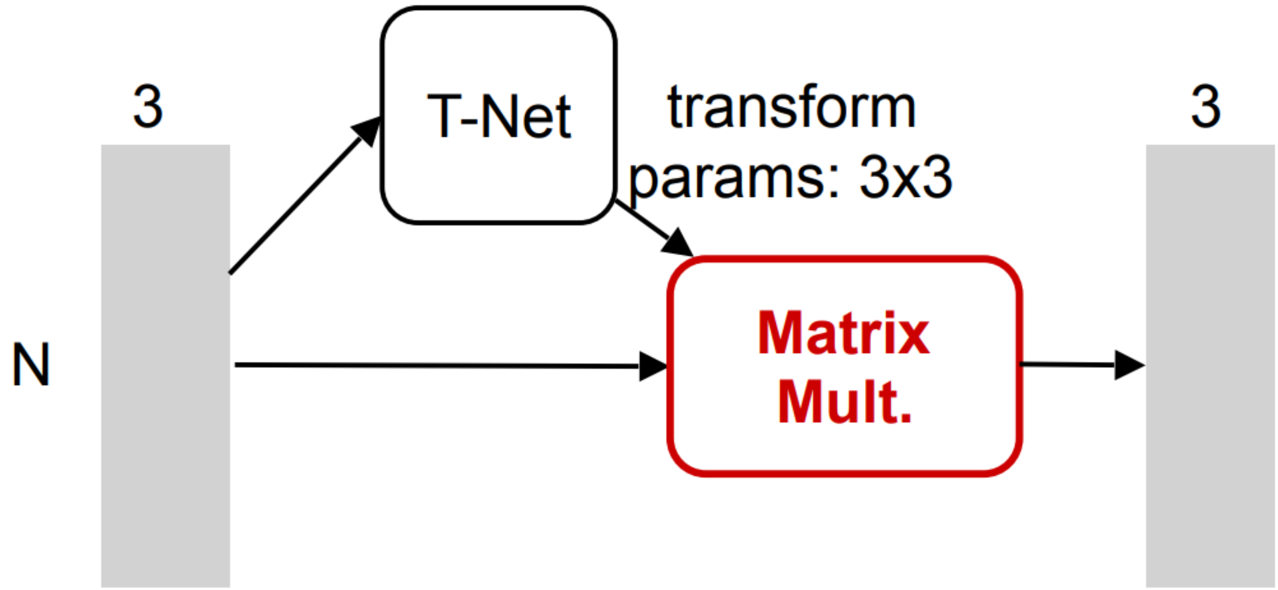
Key 1: the model should be invariant to input point permutations
Proposed solution: use symmetric functions, e.g., max pooling



Picture from http://stanford.edu/~rqi/pointnet/docs/cvpr17_pointnet_slides.pdf

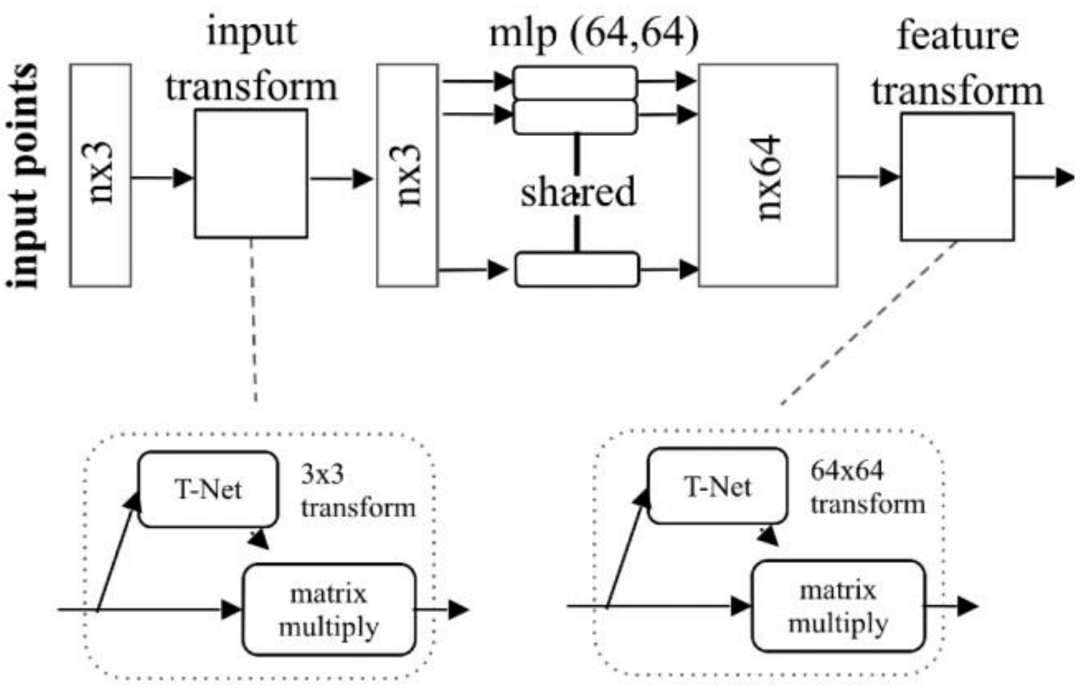
PointNet

Key 2: invariance under geometric transformations
Proposed solution: affine transformation



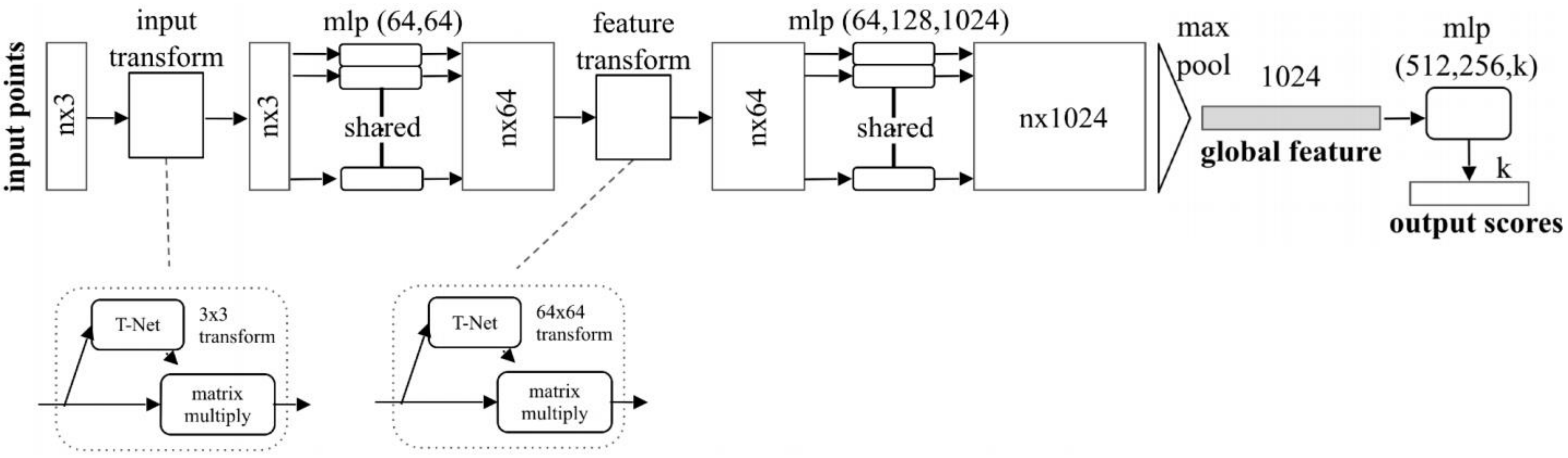
PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR, 2017.

PointNet



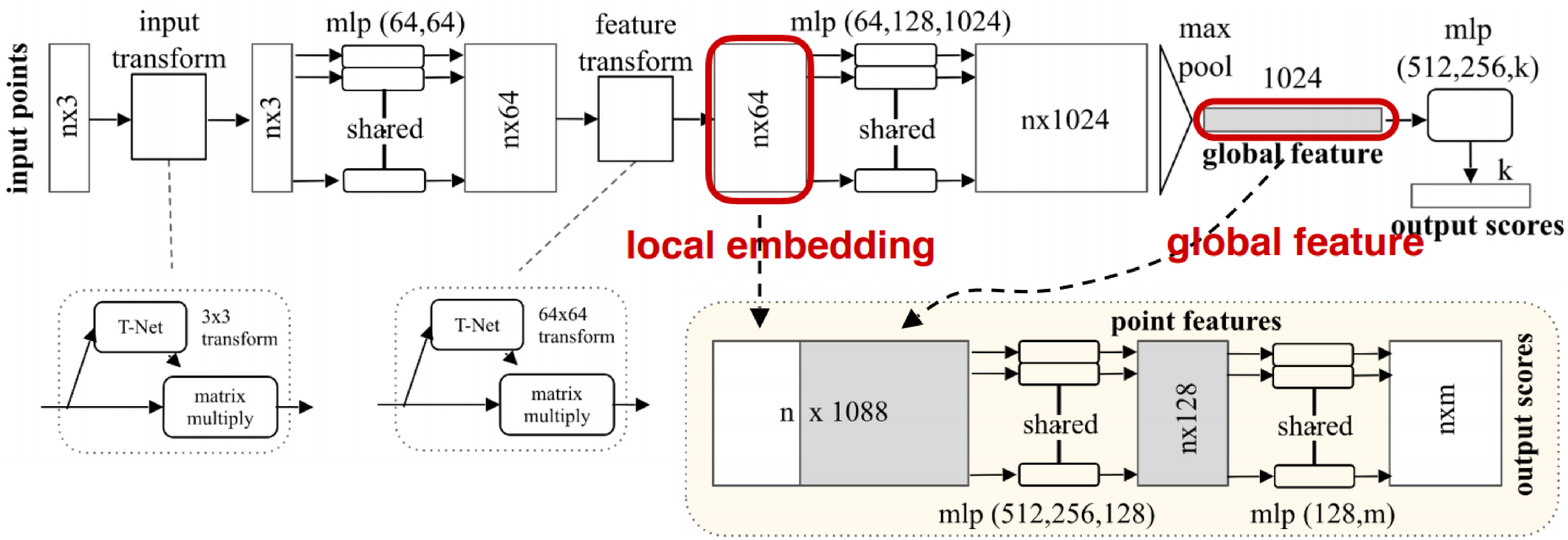
Picture from http://stanford.edu/~rqi/pointnet/docs/cvpr17_pointnet_slides.pdf

PointNet



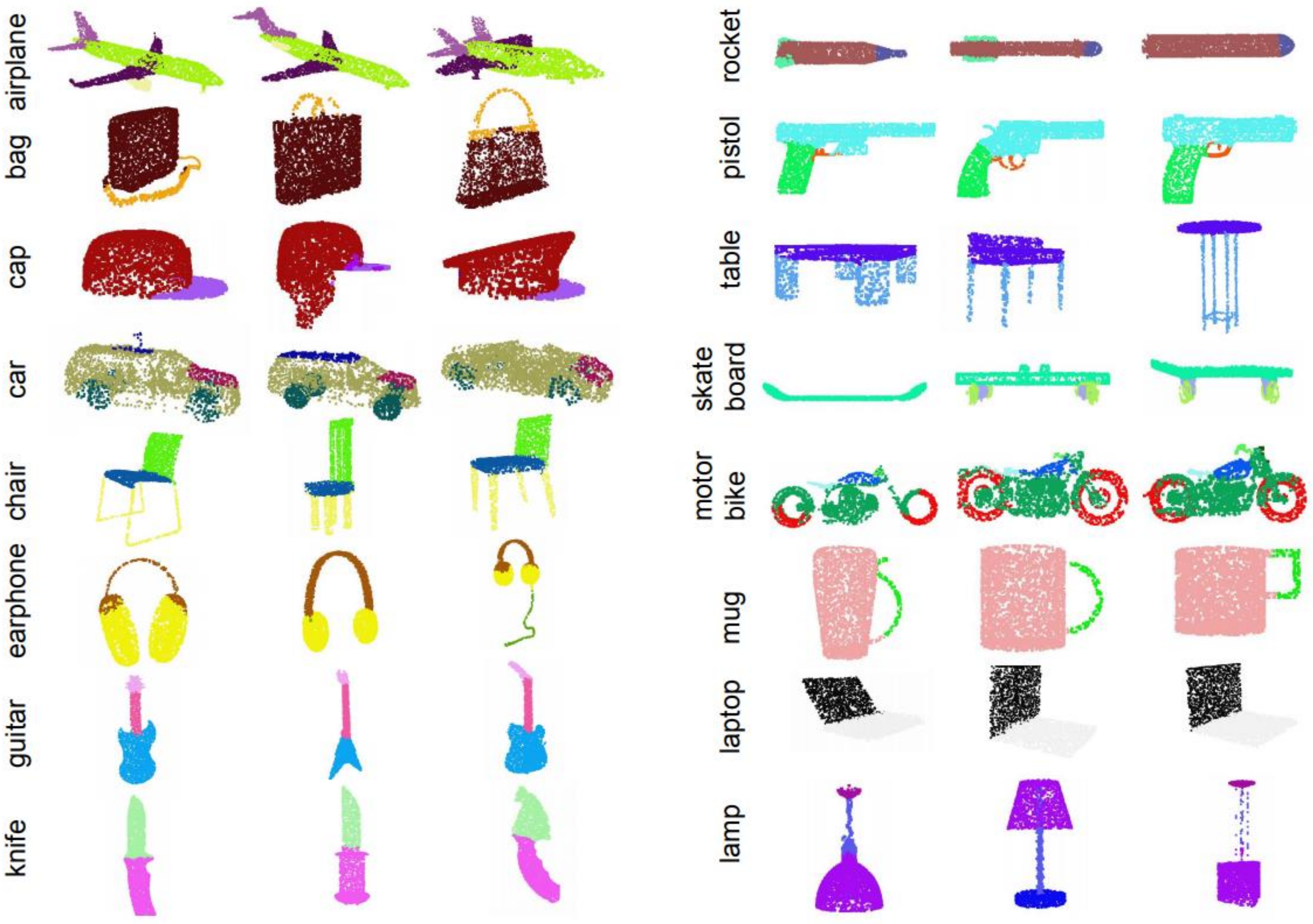
Picture from http://stanford.edu/~rqi/pointnet/docs/cvpr17_pointnet_slides.pdf

PointNet



Picture from http://stanford.edu/~rqi/pointnet/docs/cvpr17_pointnet_slides.pdf

PointNet



PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR, 2017.

4. Implicit function: continuous representation

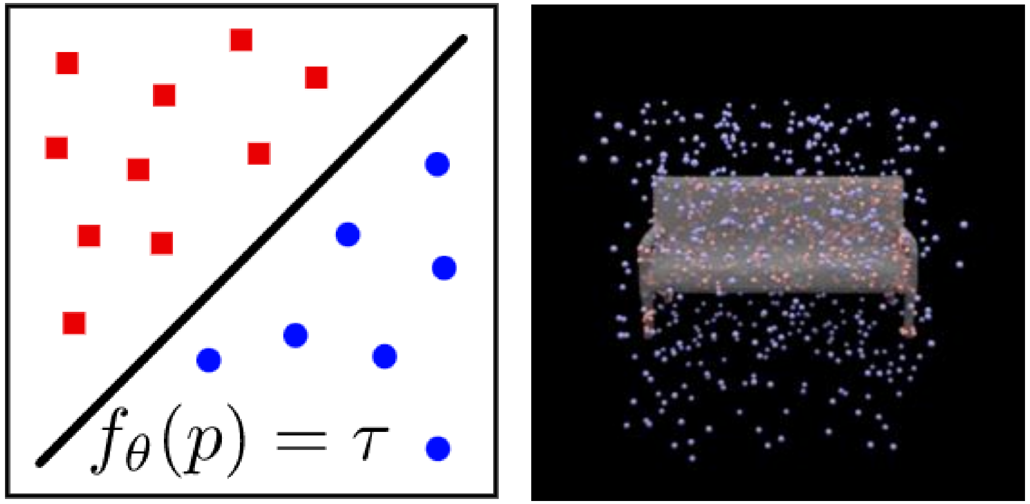
Pros:

- *No discretization*
- *Arbitrary resolution and memory efficient*
- *Arbitrary topology*
- *Not restricted to specific class*

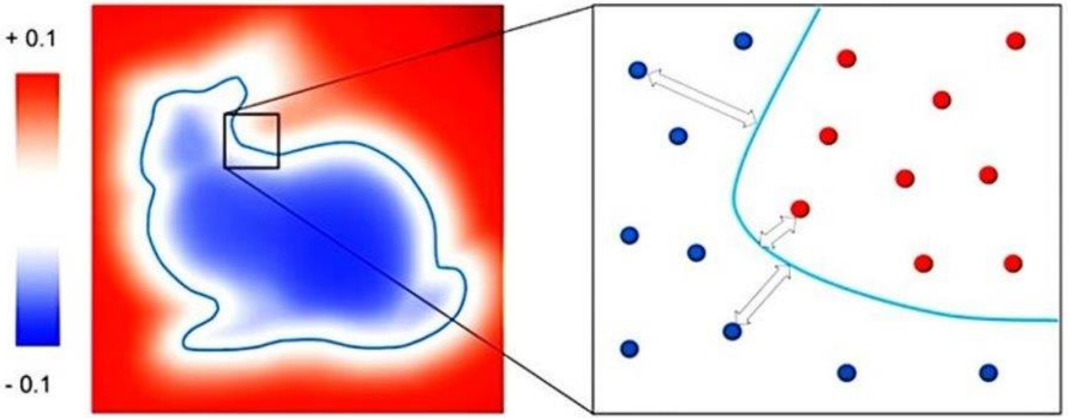
Cons:

- *Post-processing to extract meshes*

Occupancy VS Signed Distance



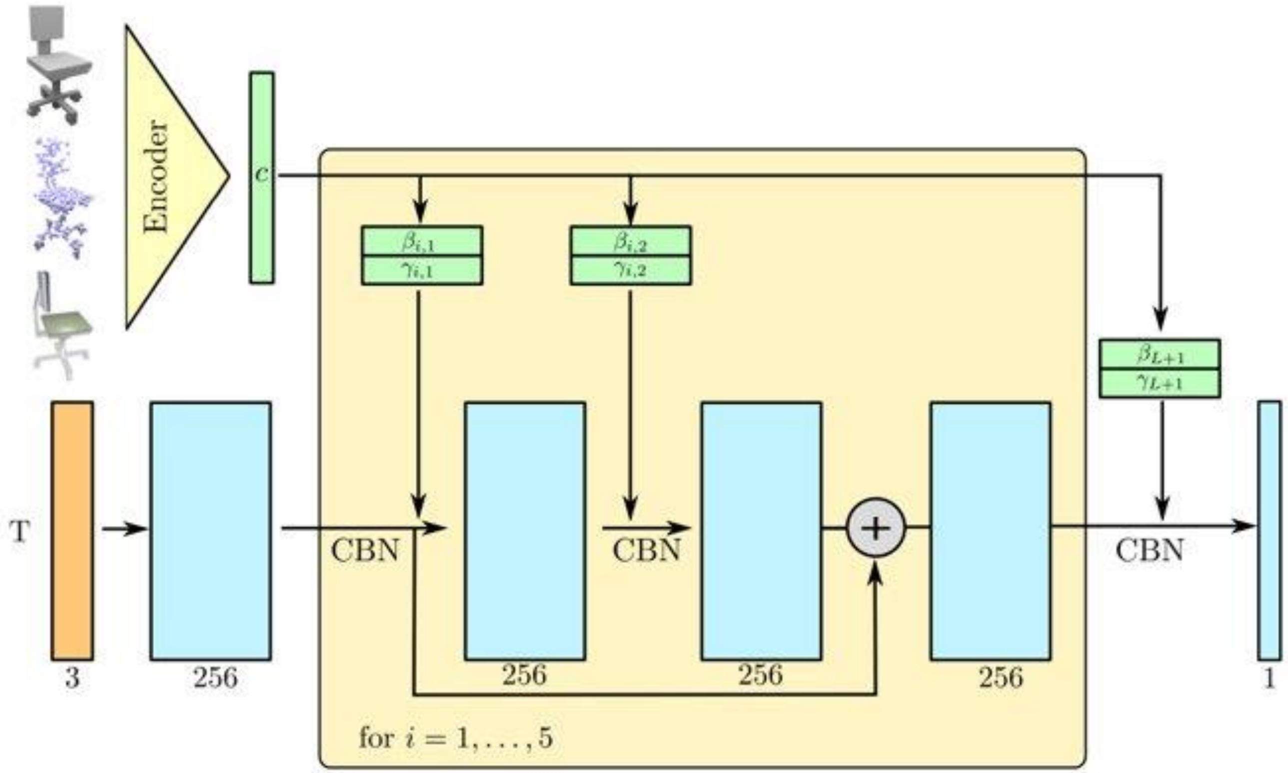
Occupancy Network



Deep Signed Distance Function

DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.
Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019.
Learning Implicit Fields for Generative Shape Modeling, CVPR, 2019

Occupancy Network



- *Input sampled 3D points and encoded features; output 0/1*

$$f_\theta : \mathbb{R}^3 \times \mathcal{X} \to [0, 1]$$

- *Fully connected layers + conditional batch normalization (CBN)*
- *Handling various input data*
  - *Image: Resnet*
  - *Point cloud: PointNet*
  - *Voxel: 3D CNN*

Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019.

Occupancy Network

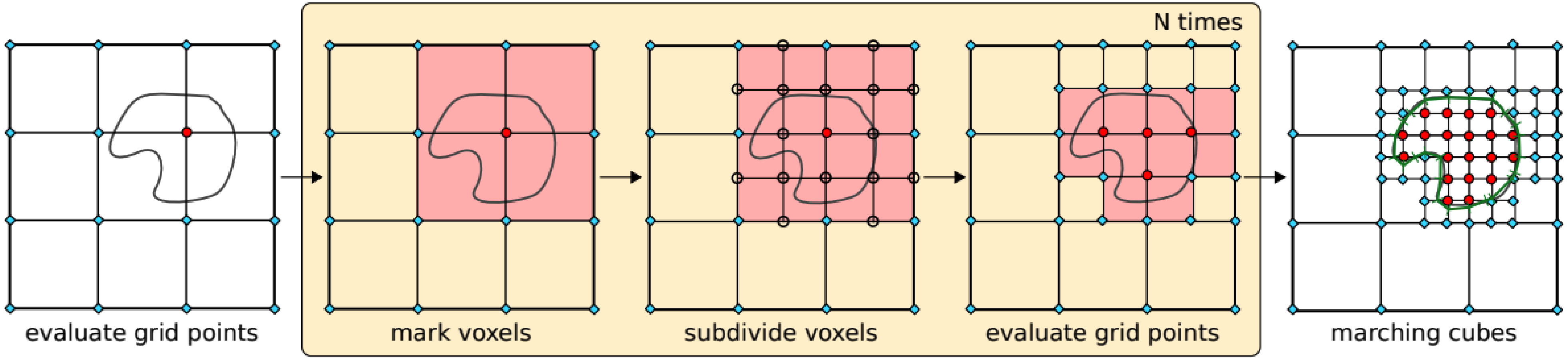
Occupancy Classification $\mathcal{L}_i(\theta)$

$$\mathcal{L}_i(\theta) = \sum_{j=1}^{K} \boldsymbol{BCE}\left(f_\theta\left(p_{ij}, x_i\right), o_{ij}\right)$$

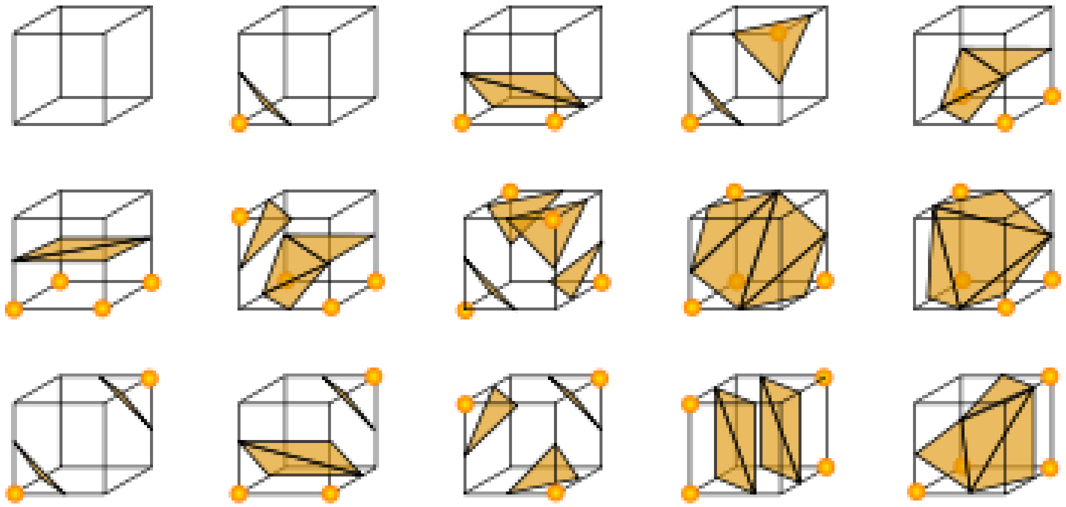
- p_{ij}: location of the j^{th} point in the i^{th} sample.
- x_i: the i^{th} sample the f_θ conditioned.
- o_{ij}: occupancy of the j^{th} point in the i^{th} sample.

Marching cubes to extract meshes

2D



3D

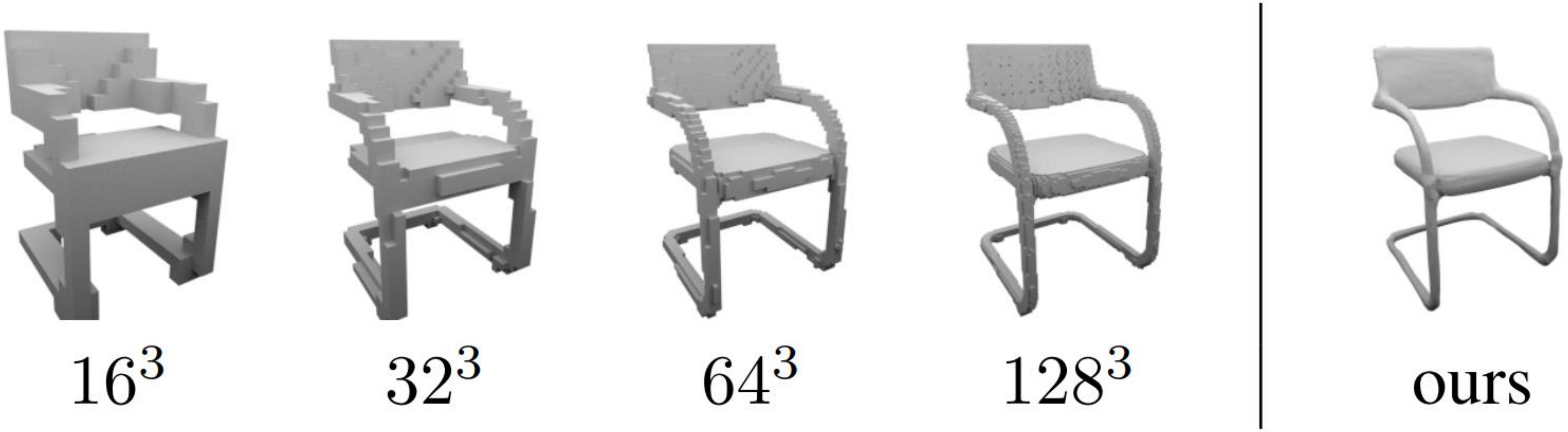


Highlighted points: outside points

Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019.

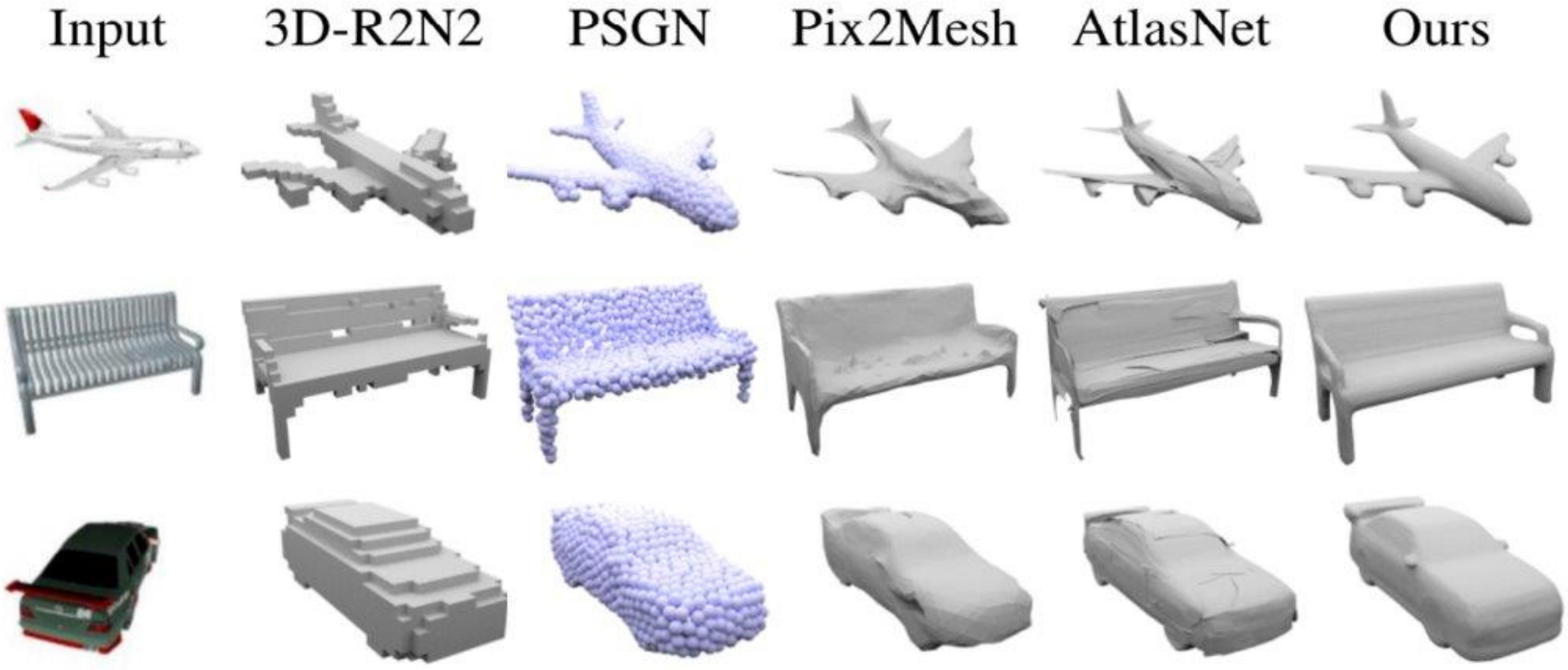
Occupancy Network

Discretize Vs continuous



Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019.

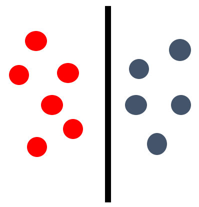
Occupancy Network



Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019.

DeepSDF

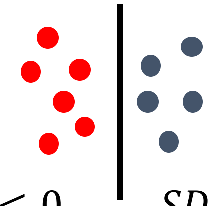
Occupancy



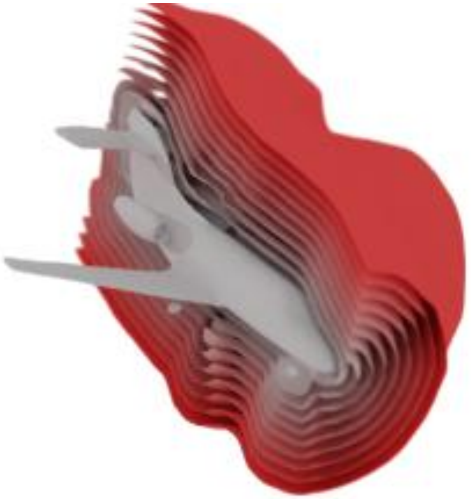
$SDF = 0$

0 level-set

Sign-distance-function

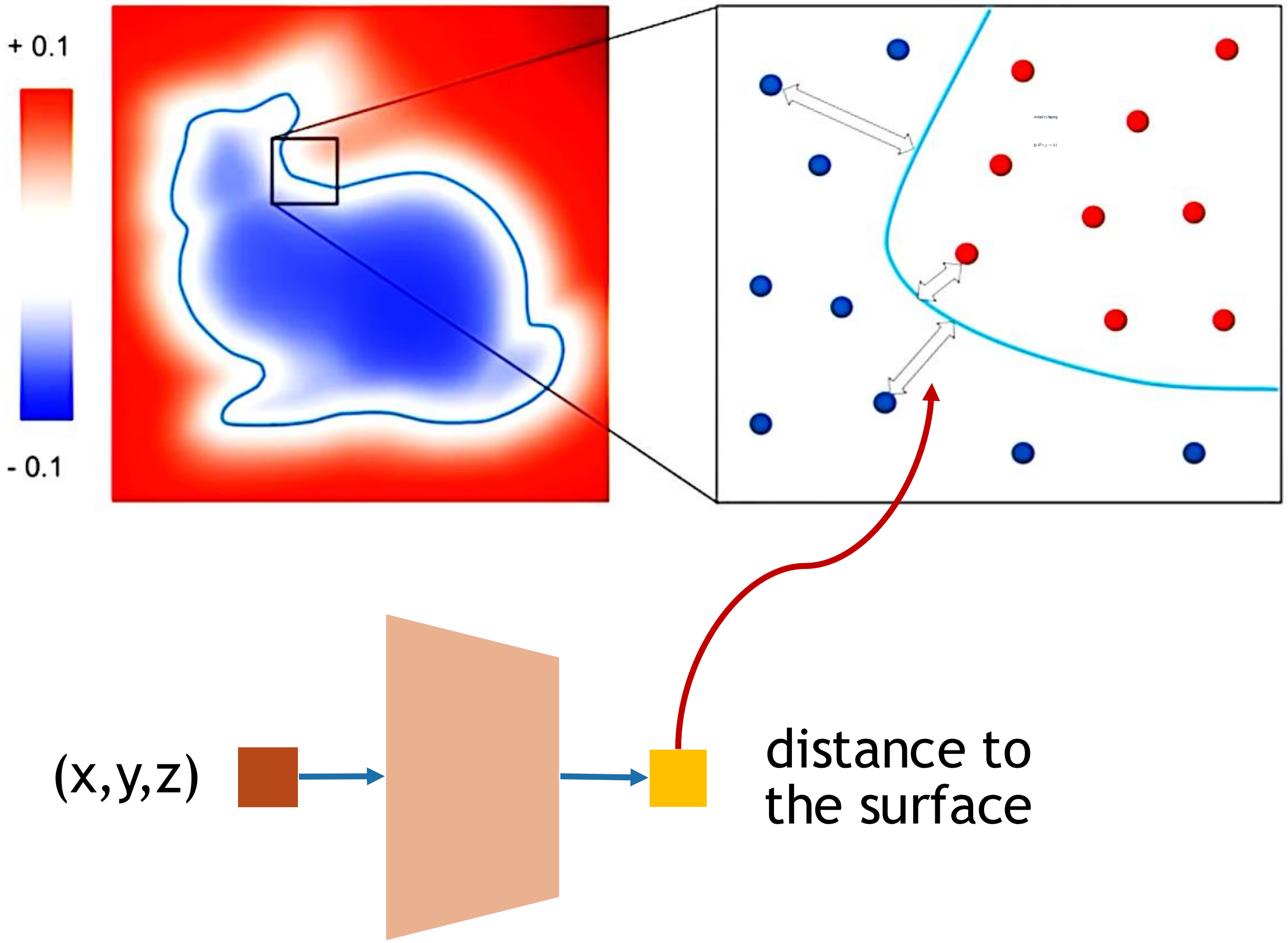


$SDF < 0$ $SDF > 0$



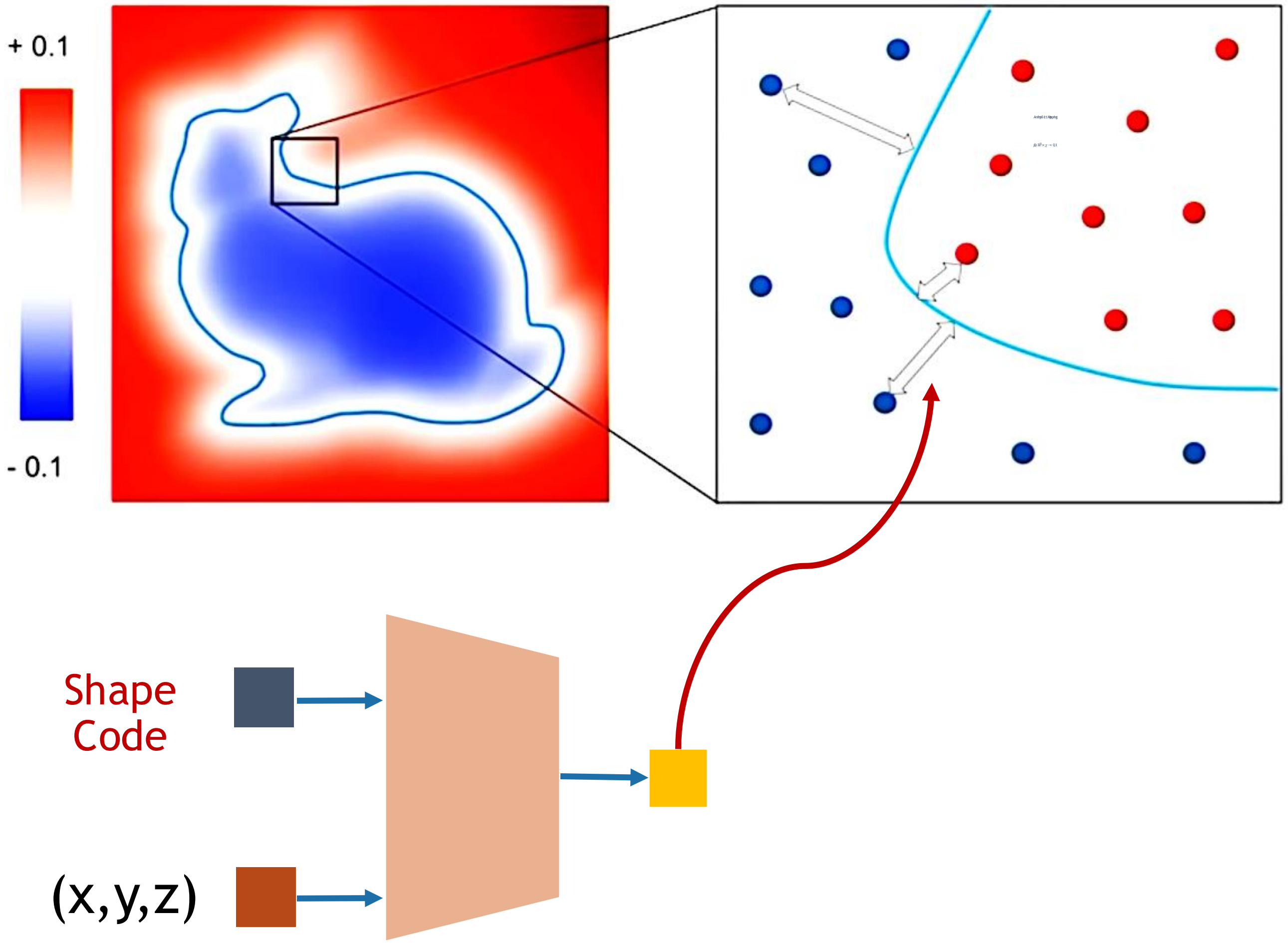
full level-set

DeepSDF



$$f_\theta(x) \approx SDF(x), \forall x \in \Omega$$

DeepSDF

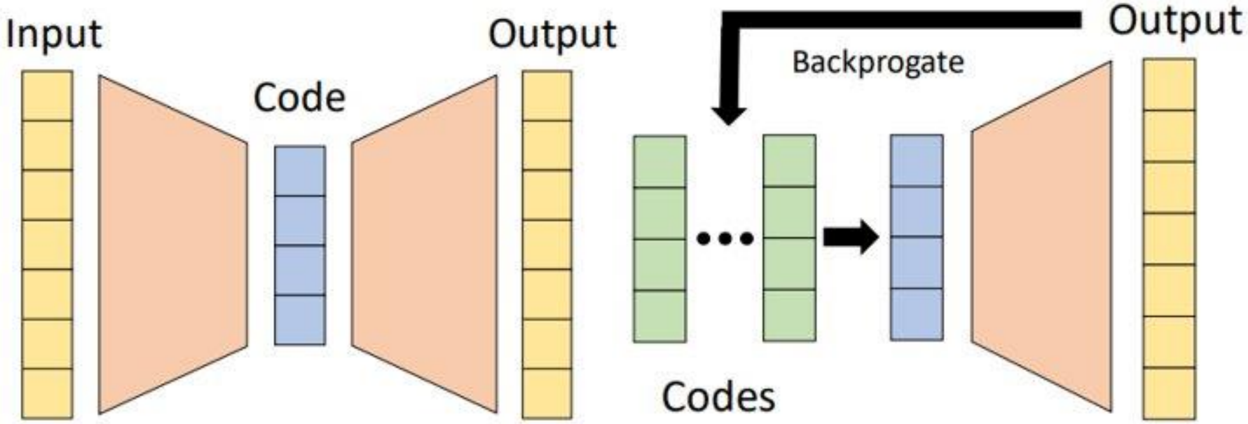


$$f_\theta(z_i, x) \approx SDF^i(x)$$

Function

- 8 fully-connected layers
- weight-norm

DeepSDF



(a) Auto-encoder **(b)** Auto-decoder

- *Auto-decoder architecture*

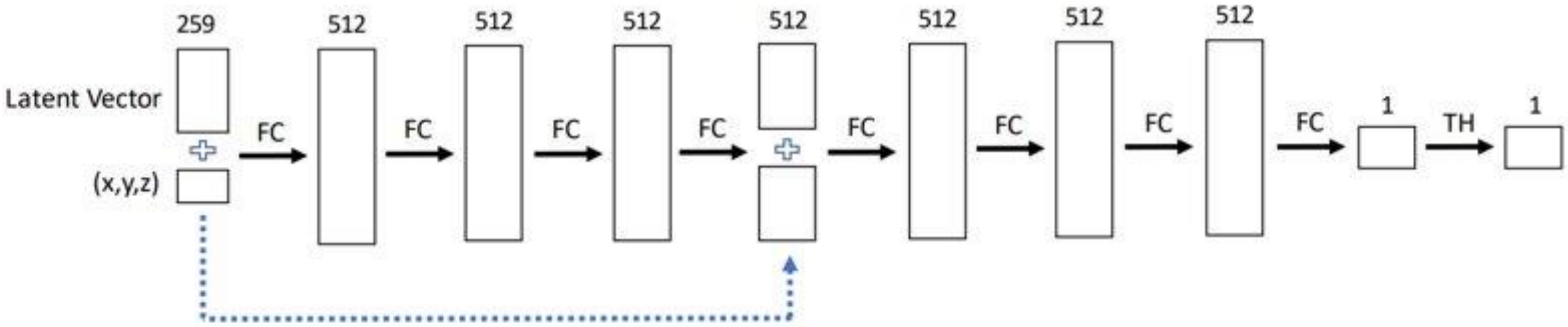
$$f_\theta(\boldsymbol{z}_i, \boldsymbol{x}) \approx SDF^i(\boldsymbol{x})$$

Shape code   Sampled point

- *Training:*
  - *Jointly learning of shape codes and model parameters*
  - *Shape codes are randomly initialized*

DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.

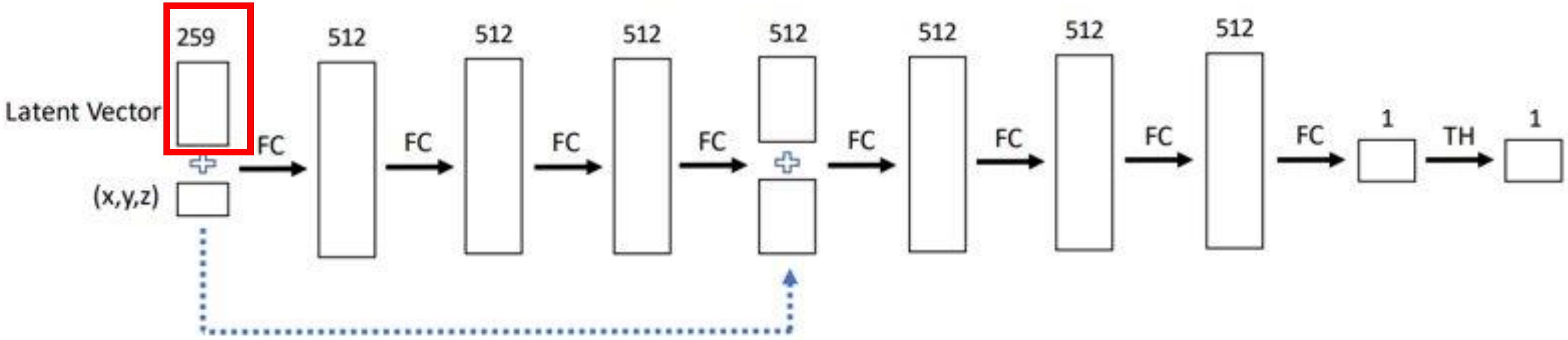
DeepSDF



- *Fully connected layers + Weight normalization*
- *Tanh() at the output*
- *Signed Distance clipping*

DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.

DeepSDF



- *Optimization-based inference*
  - *Fix model parameters*
  - *Infer the shape code of the obervation*

$$\underset{\theta,\{\boldsymbol{z}_i\}_{i=1}^{N}}{\arg\min} \sum_{i=1}^{N} \left(\sum_{j=1}^{K} \mathcal{L}(f_\theta(\boldsymbol{z}_i, \boldsymbol{x}_j), s_j) + \frac{1}{\sigma^2} ||\boldsymbol{z}_i||_2^2 \right)$$

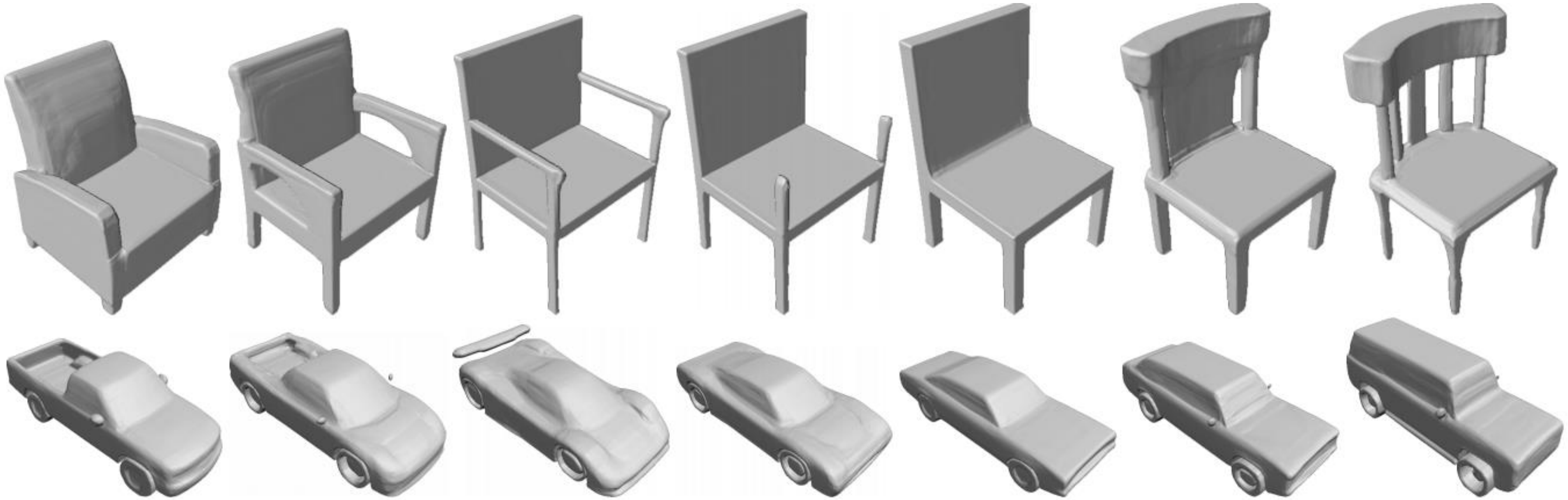
Signed Distance Loss

Shape code regularization

DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.

DeepSDF

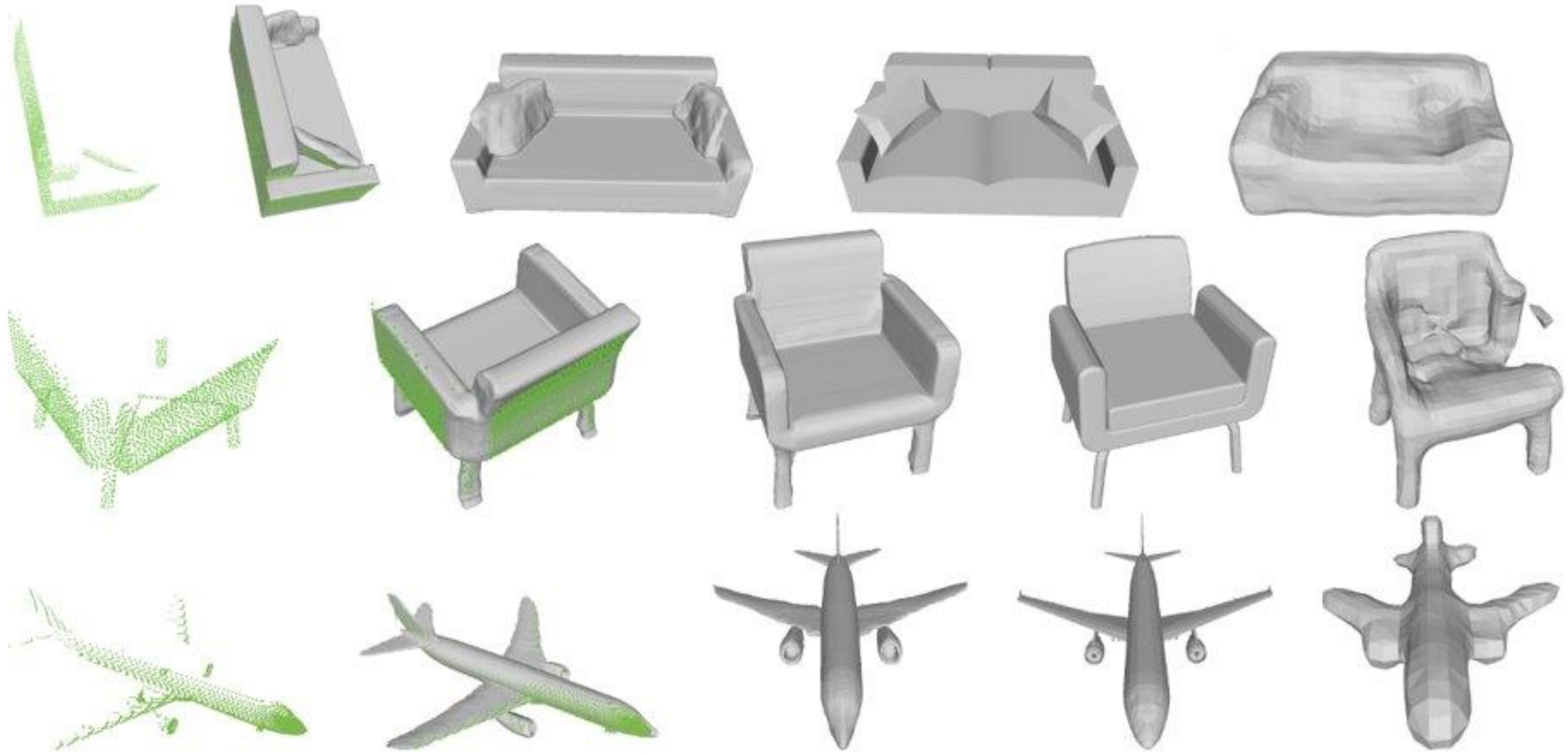
Generate shapes via shape code interpolation



DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.

DeepSDF

Partial point cloud completion



(a) Input Depth (b) Completion (ours) (c) Second View (ours) (d) Ground truth (e) 3D-EPN

DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.

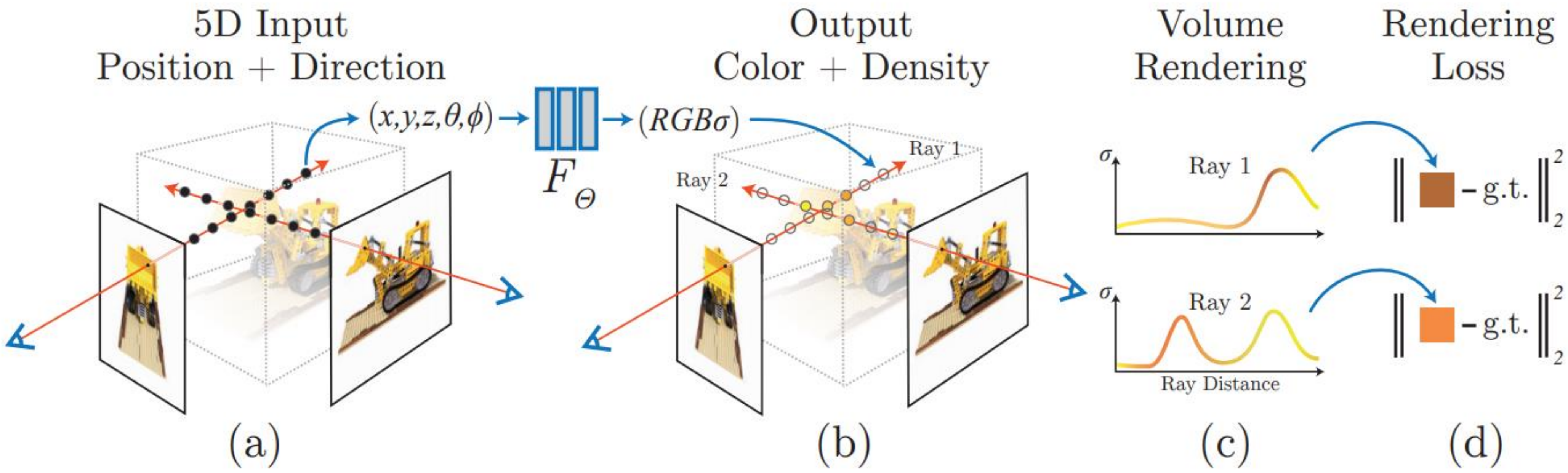
NeRF

Applied to image rendering

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020.

NeRF

1. Given sampled 3D points + camera directions, output RGB values and density
2. Image can be generated via volume rendering



Summary

- Volumetric representation
- Mesh
- Point cloud
- Implicit functions