# 3D Reconstruction based on Implicit Representation

Dong Du

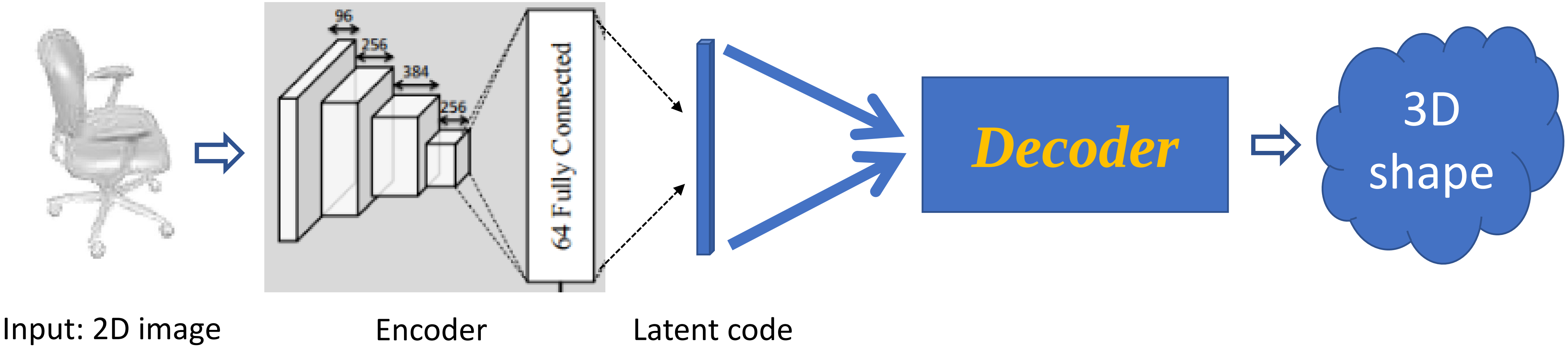
dongdu@mail.ustc.edu.cn

USTC; SRIBD

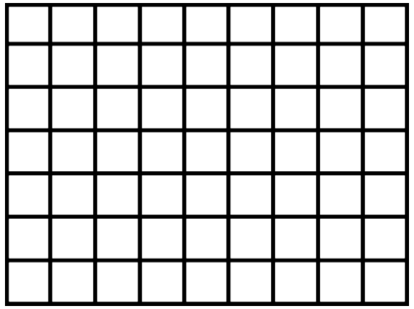
2019/08/22

# Task: 3D Reconstruction

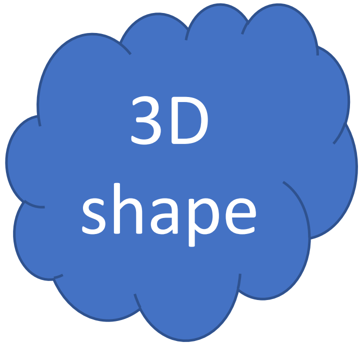
- Single-view object reconstruction



# 3D Representation



Pixel

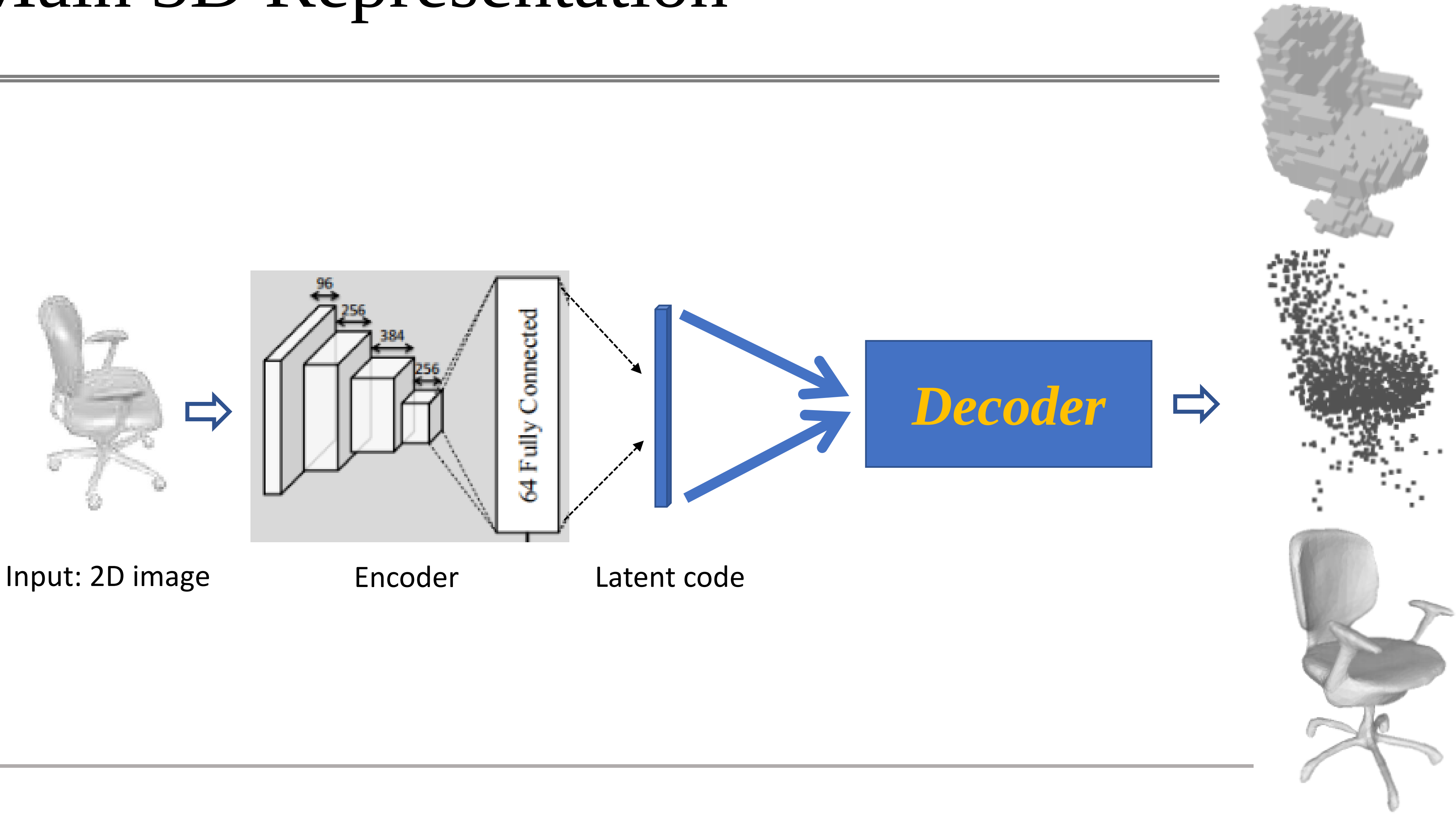



Volumetric grid(**Voxel**)
Multi-view depth/normal maps
Parametric model
Primitive-based CAD model
**Point cloud**
**Mesh**
....

2D: Uniform structure

3D: Non-uniform structure

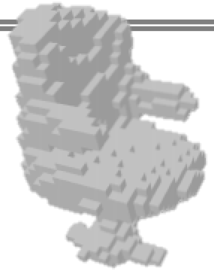
# Main 3D Representation



# Main 3D Representation

Explicit Representation

- Voxel (3D-CNN)
  - natural extension of 2D
  - high computational cost, low info efficiency, low-res output



- Point cloud (PointSetGeneration)
  - lightweight, high info efficiency
  - coarse, no surface or topological information, hard to extract a mesh

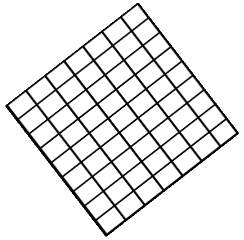
- Mesh (AtlasNet, Pixel2Mesh)
  - many good geometry properties for processing and rendering
  - irregular and hard to learning; overlapping and messy output(AtlasNet), limited topology(Pixel2Mesh)

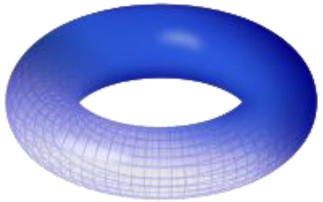
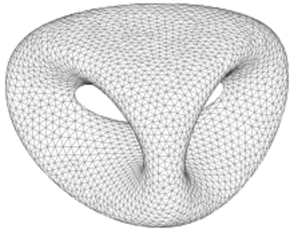
- Hybrid (Tang&Han et al.)

# Implicit Surface

- Plane: $x + 2y - 3z + 1 = 0.$
- Sphere: $x^2 + y^2 + z^2 - 4 = 0.$
- Torus: $(x^2 + y^2 + z^2 + R^2 - a^2)^2 - 4R^2(x^2 + y^2) = 0.$
- Genus 2 surface: $2y(y^2 - 3x^2)(1 - z^2) + (x^2 + y^2)^2 - (9z^2 - 1)(1 - z^2) = 0.$

Continuous surface
High resolution
Good geometry/property …

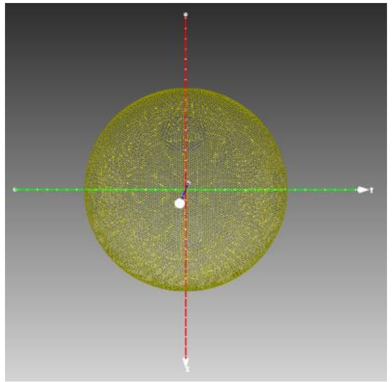


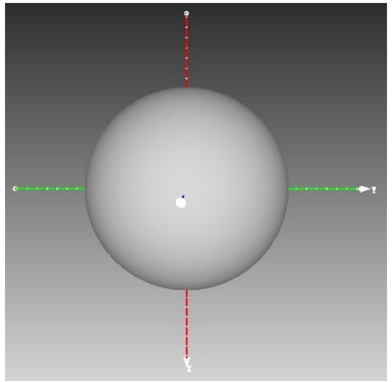
# Implicit Representation in Graphics

- Iso-surface

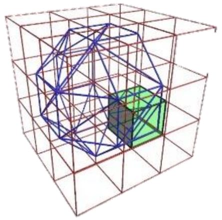
$$x^2 + y^2 + z^2 - 4 = 0.$$



- Level set
  - 0 level set (0 iso-surface)
  - volumetric representation (1 for inside, 0 for outside)



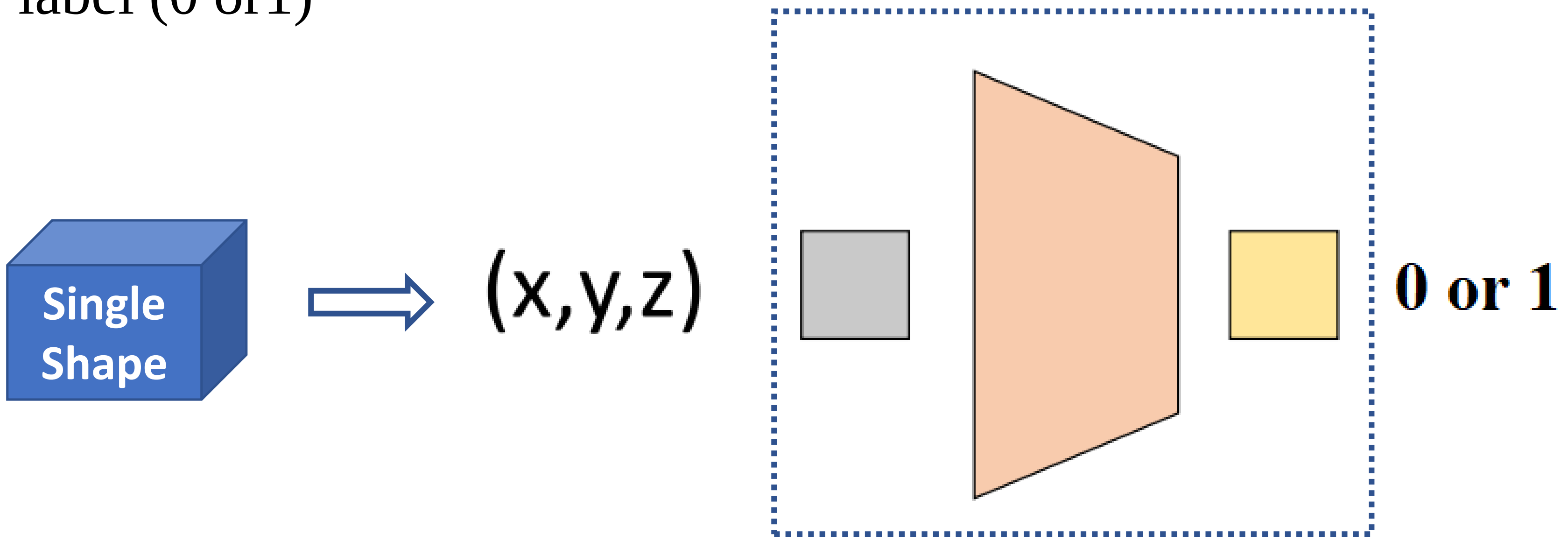
- Surface extraction method



  - Marching Cube algorithm
  - (Screened) Poisson Surface Reconstruction

# How to apply implicit representation to learning?

- Binary Classifier
  - Classify all points 0 or 1
  - Input: (x, y, z)
  - Output: label (0 or1)




Deep Network: e.g. MLP

# Paper list

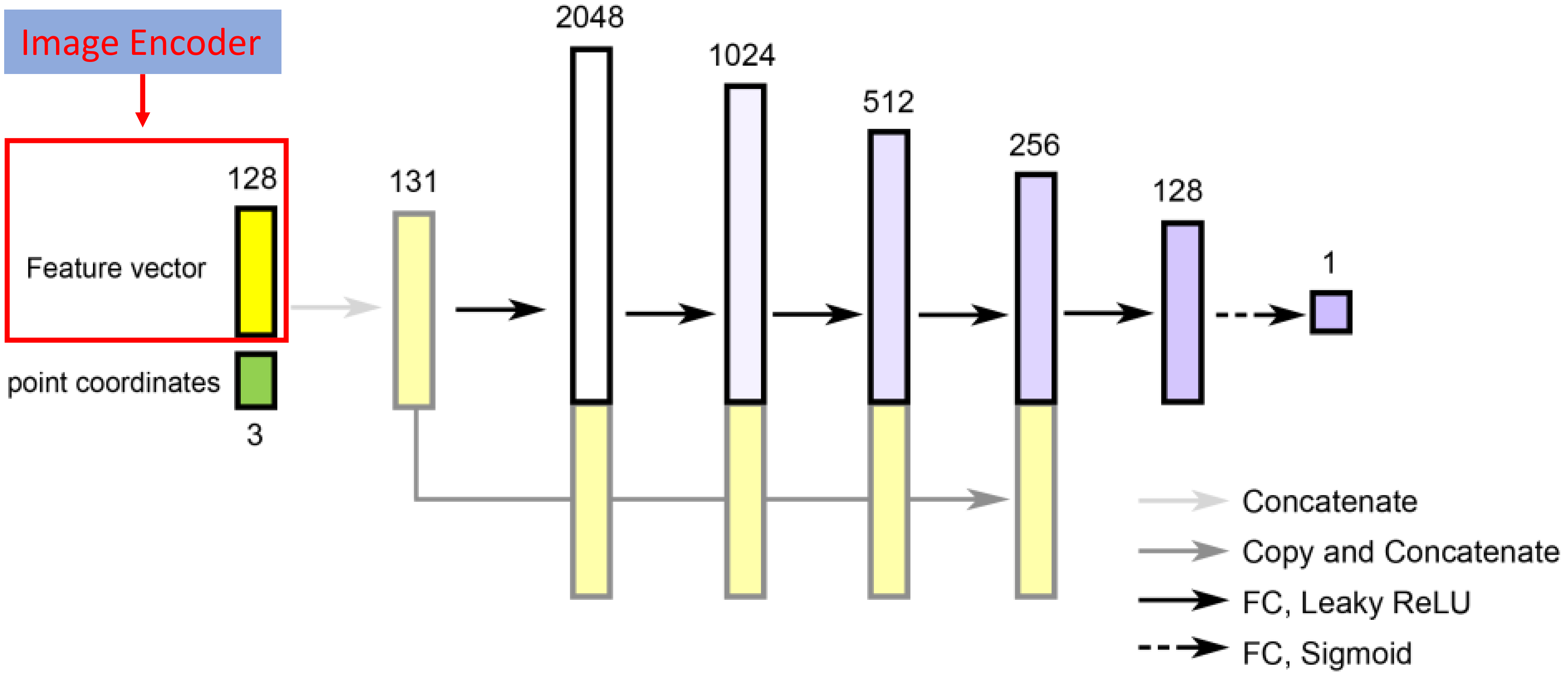
- **Learning Implicit Fields for Generative Shape Modeling**, Chen et al. CVPR 2019
- **DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation**, Park et al. CVPR oral 2019
- **Occupancy Networks: Learning 3D Reconstruction in Function Space**, Mescheder et al. CVPR oral 2019
- **DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction**, Wang et al. 2019
- **Deep Level Sets: Implicit Surface Representations for 3D Shape Inference**, Michalkiewicz et al. 2019

# Learning Implicit Fields for Generative Shape Modeling

Zhiqin Chen
Simon Fraser University
zhiqinc@sfu.ca

Hao Zhang
Simon Fraser University
haoz@sfu.ca

# Implicit Decoder



# Training

- Data preparation
  - image, point set, point label
- **Point sampling**
  - **uniform**: $16^3, 32^3, 64^3, 128^3$ volumetric points
  - **adaptive**: sample more points near shape surfaces, neglect most points far away
- **Progressive training** techniques
  - first train on $16^3$ resolution data, then increase resolution gradually
- Loss function

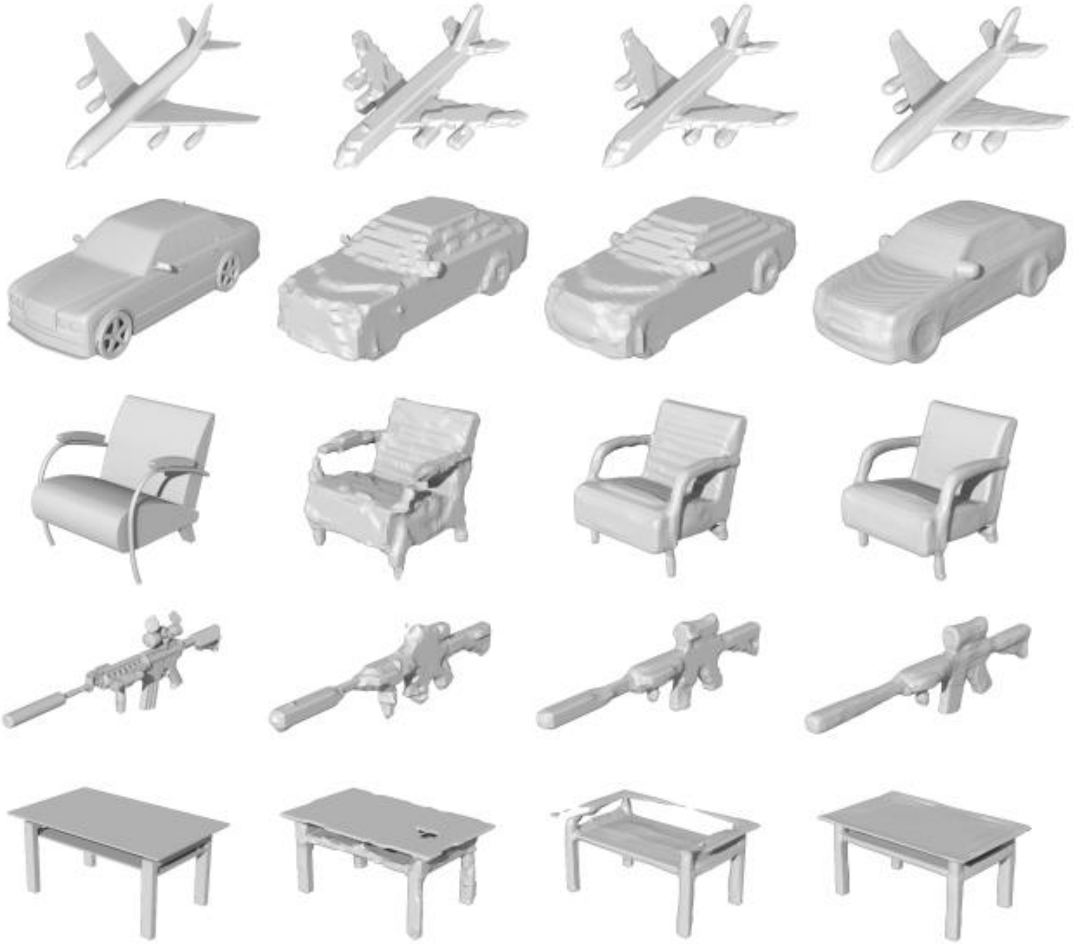
$$\mathcal{L}(\theta) = \frac{\sum_{p \in S} |f_\theta(p) - \mathcal{F}(p)|^2 \cdot w_p}{\sum_{p \in S} w_p}$$ , $w_p$ represents the inverse of sampling density near p

# Quality Metrics

- CD (Chamfer Distance)
- MSE (Mean Squared Error)
- IoU (Intersection over Union)
- **LFD (Light Field Descriptor)**
  - a set of rendered views of a 3D shape from various camera angles
  - compare the similarity of projected images

| | Plane | Car | Chair | Rifle | Table |
|---|---|---|---|---|---|
| CNN64-MSE | **1.47** | **4.37** | **7.76** | **1.62** | **5.80** |
| IM64-MSE | 2.14 | 4.99 | 11.43 | 1.91 | 10.67 |
| CNN64-IoU | **86.07** | **90.73** | **74.22** | **78.37** | **84.67** |
| IM64-IoU | 78.77 | 89.26 | 65.65 | 72.88 | 71.44 |
| CNN64-CD | **3.51** | 5.31 | **7.34** | **3.48** | **7.45** |
| IM64-CD | 4.22 | **5.28** | 8.96 | 3.78 | 12.05 |
| IM256-CD | 4.23 | 5.44 | 9.05 | 3.77 | 11.54 |
| CNN64-LFD | 3,375 | 1,323 | 2,555 | **3,515** | **1,824** |
| IM64-LFD | 3,371 | 1,190 | 2,515 | 3,714 | 2,370 |
| IM256-LFD | **3,236** | **1,147** | **2,453** | 3,602 | 2,201 |

# Visual Results



(a) Ground truth (b) CNN-AE (c)IM-AE64 (d)IM-AE256

# 3D Shape Interpolation Results

# Occupancy Networks: Learning 3D Reconstruction in Function Space

Lars Mescheder[1] Michael Oechsle[1,2] Michael Niemeyer[1] Sebastian Nowozin[3†] Andreas Geiger[1]
[1]Autonomous Vision Group, MPI for Intelligent Systems and University of Tübingen
[2]ETAS GmbH, Stuttgart
[3]Google AI Berlin
{firstname.lastname}@tue.mpg.de nowozin@gmail.com

Implicitly represent the 3D surface as the continuous decision boundary of a deep neural network classifier

# Difference

- Loss function

$$\mathcal{L}_{\mathcal{B}}^{\text{gen}}(\theta, \psi) = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \Big[ \sum_{j=1}^{K} \mathcal{L}(f_\theta(p_{ij}, z_i), o_{ij}) + \mathrm{KL}\left(q_\psi(z|(p_{ij}, o_{ij})_{j=1:K}) \,\|\, p_0(z)\right) \Big]$$

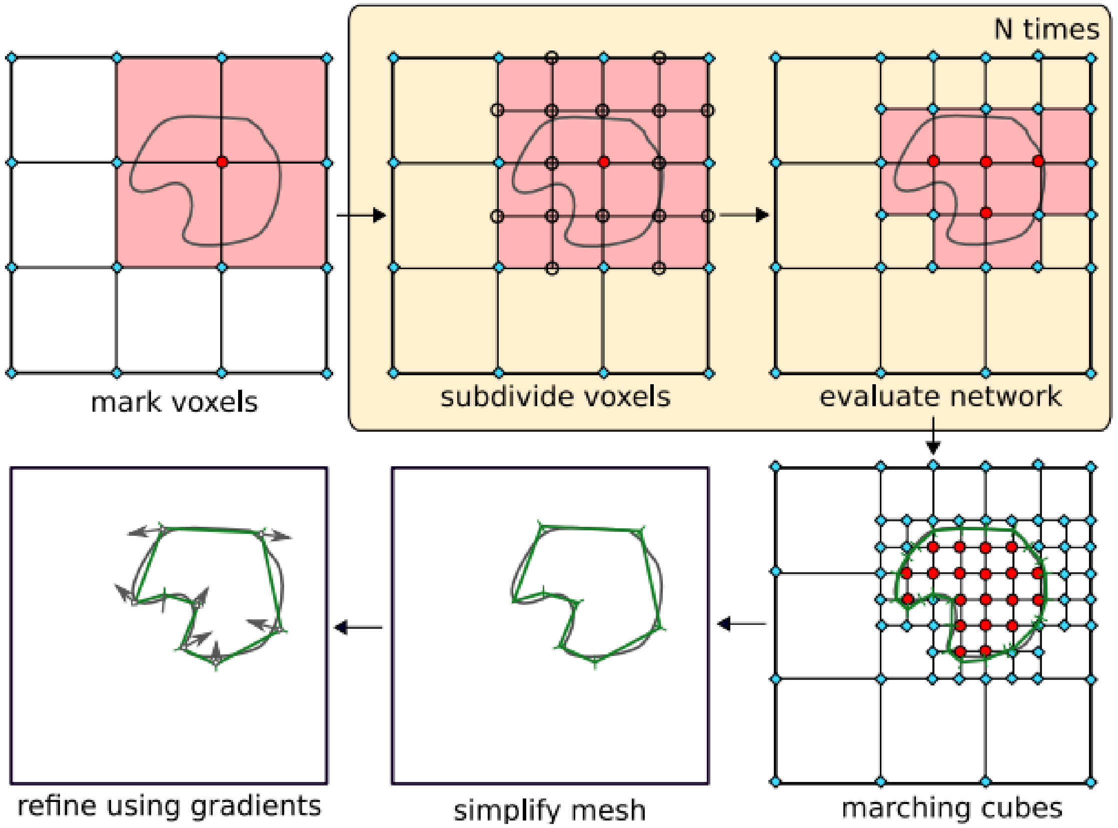
where $q_\psi(z|(p_{ij}, o_{ij})_{j=1:K})$
is Gaussian distribution.

- Inference: MISE (Multiresolution Iso-Surface Extraction)

  - Extract high-res meshes without densely evaluating all points

# MISE

- Extract high-res meshes without densely evaluating all points



# DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation

Jeong Joon Park[1,3†] Peter Florence [2,3†] Julian Straub[3] Richard Newcombe[3] Steven Lovegrove[3]

[1]University of Washington [2]Massachusetts Institute of Technology [3]Facebook Reality Labs

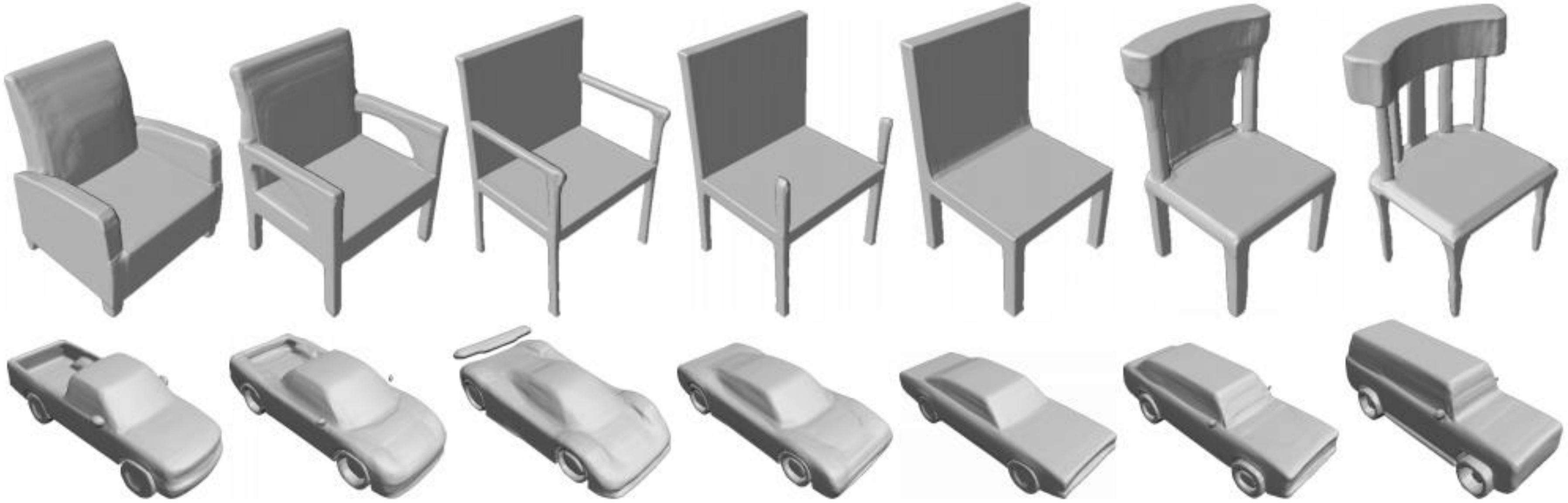
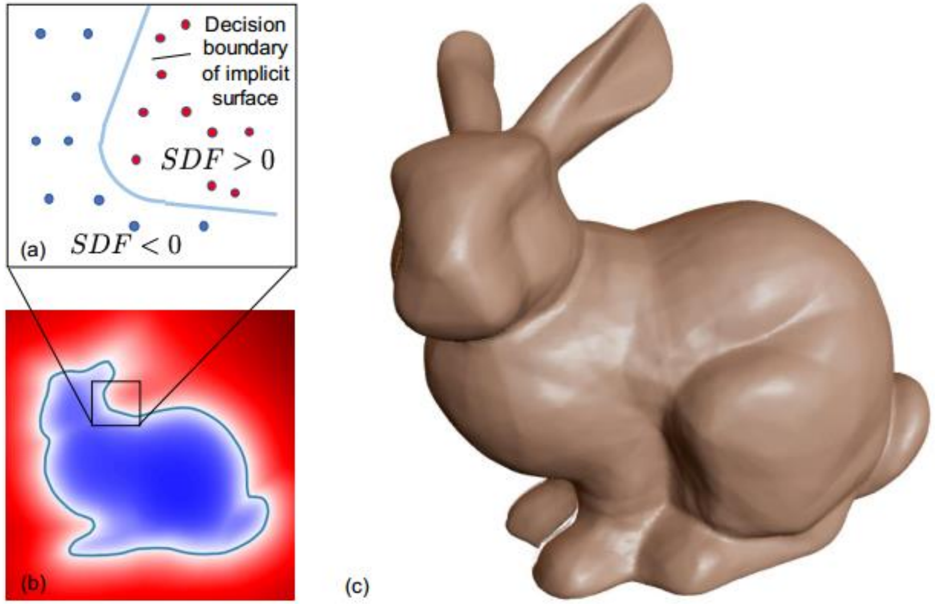


**Figure 1:** DeepSDF represents signed distance functions (SDFs) of shapes via latent code-conditioned feed-forward decoder networks. Above images are raycast renderings of DeepSDF interpolating between two shapes in the learned shape latent space. Best viewed digitally.

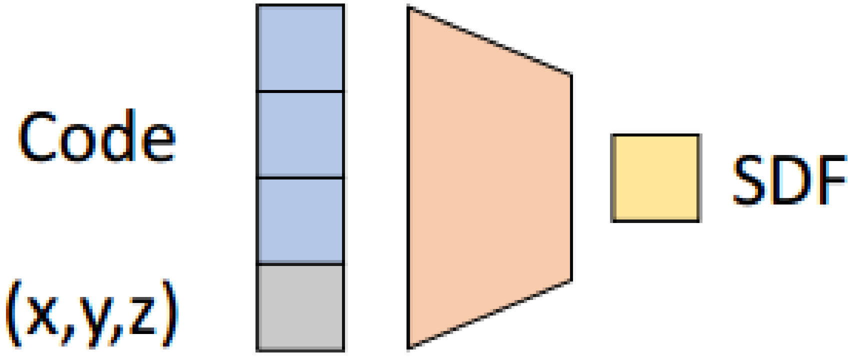
# SDF(Signed Distance Function)

- Represent shape's surface by a continuous volumetric field
  - **magnitude**: the distance of a point to the surface boundary
  - **sign**: inside (-) or outside (+) of the shape

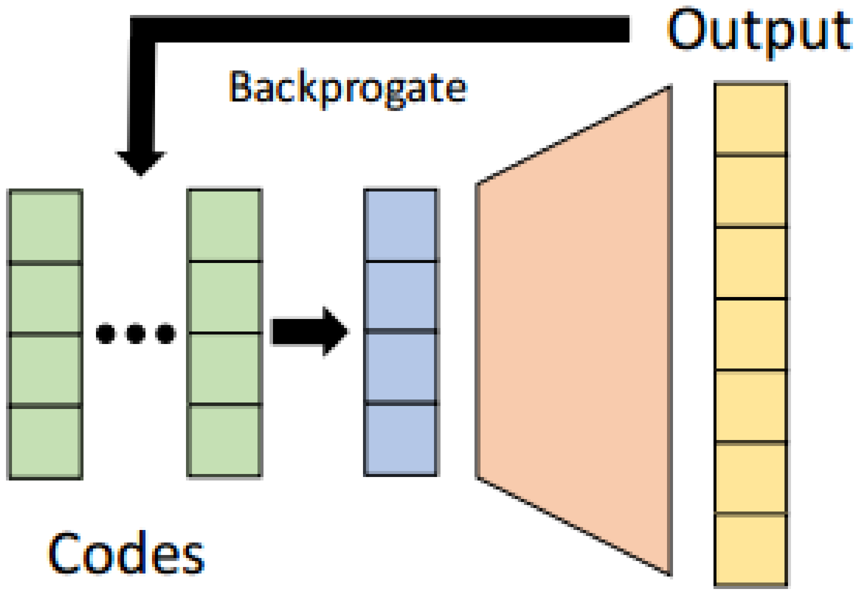


# Learning the Latent Space of Shapes

- Auto-decoder-based DeepSDF Formulation
  - latent vector $z$ is initialized randomly from $N(0, 0.01^2)$
  - training data: vector $z$ and shape $X\ pairs(set\ of\ points\ and\ SDFs)$




(a) Coded Shape DeepSDF




(b) Auto-decoder

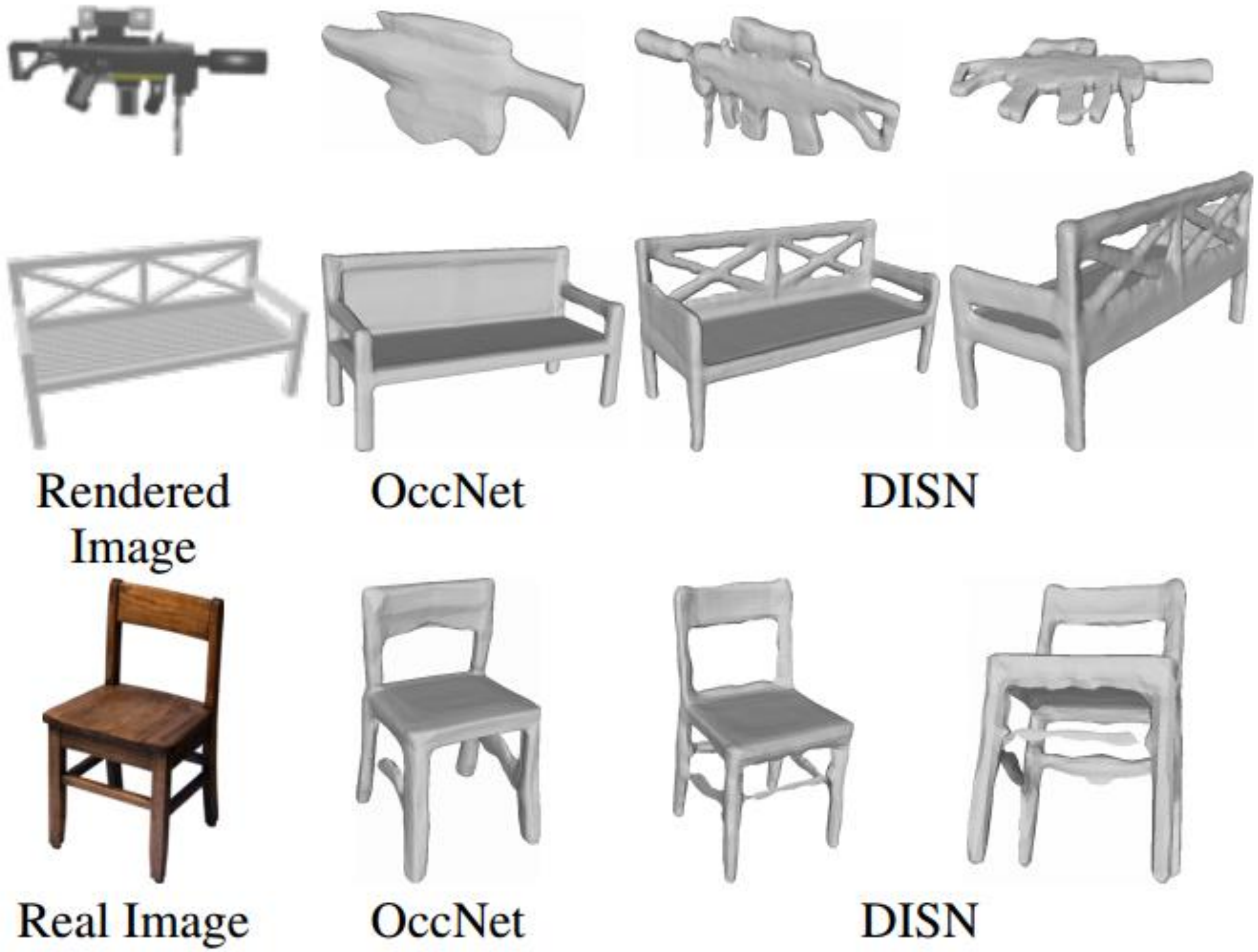
# DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction

**Weiyue Wang*,[1]** **Qiangeng Xu*,[1]** **Duygu Ceylan[2]** **Radomir Mech[2]** **Ulrich Neumann[1]**

[1]University of Southern California
Los Angeles, California
{weiyuewa,qiangenx,uneumann}@usc.edu

[2]Adobe
San Jose, California
{ceylan,rmech}@adobe.com

# Problem: ignore details (holes or thin structures)



# Key Idea: Combine global and local features

- DISN: two parts
  - camera pose estimation(for projection)
  - SDF prediction

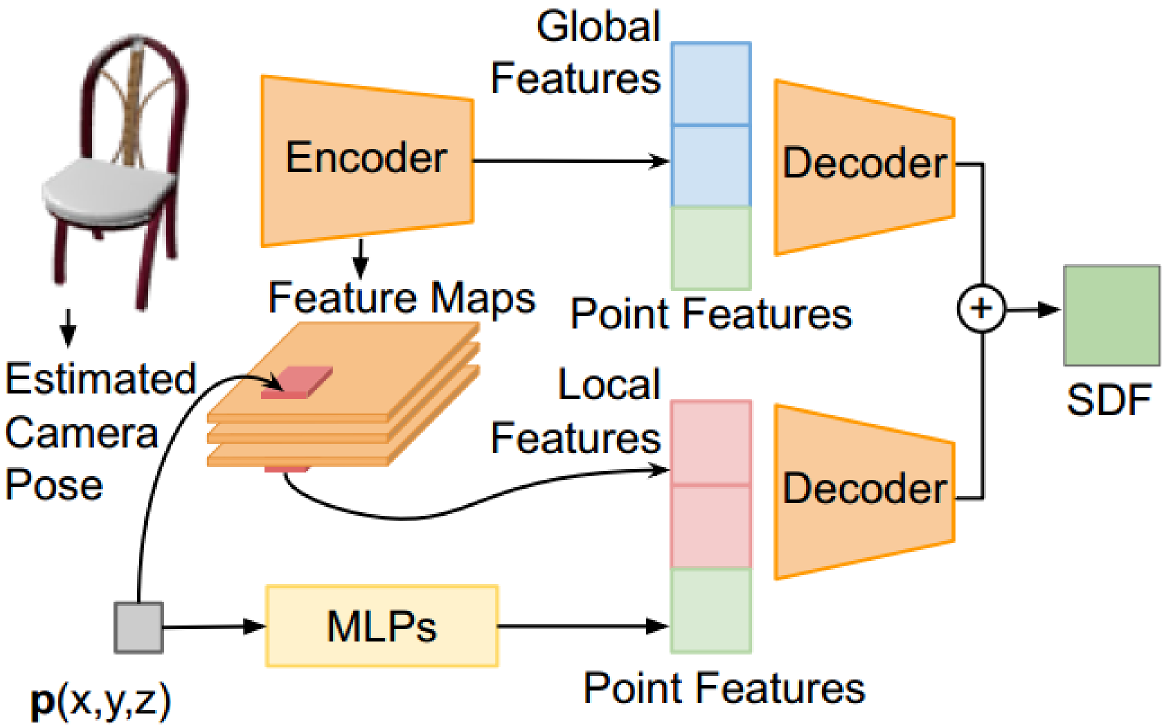
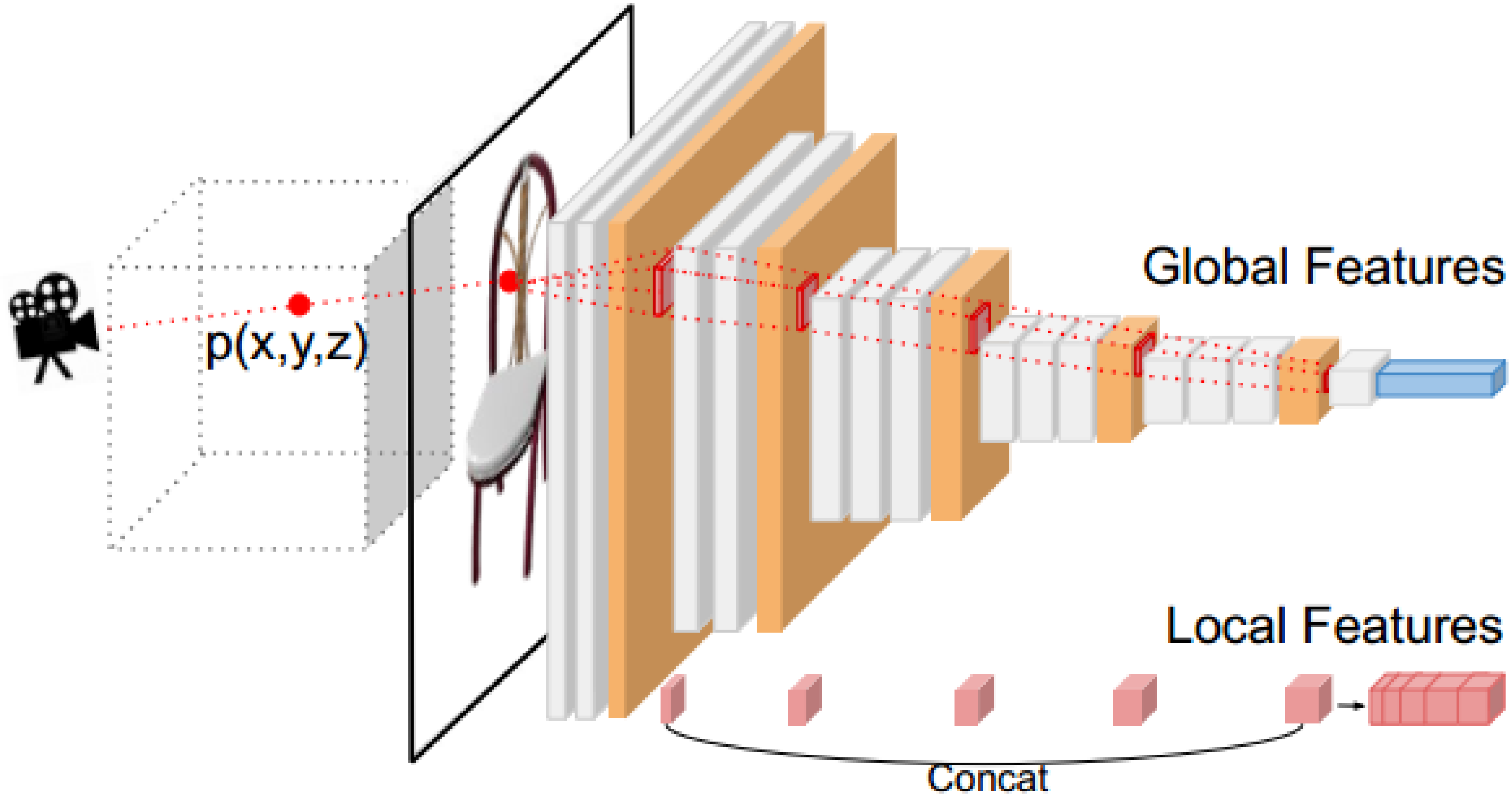


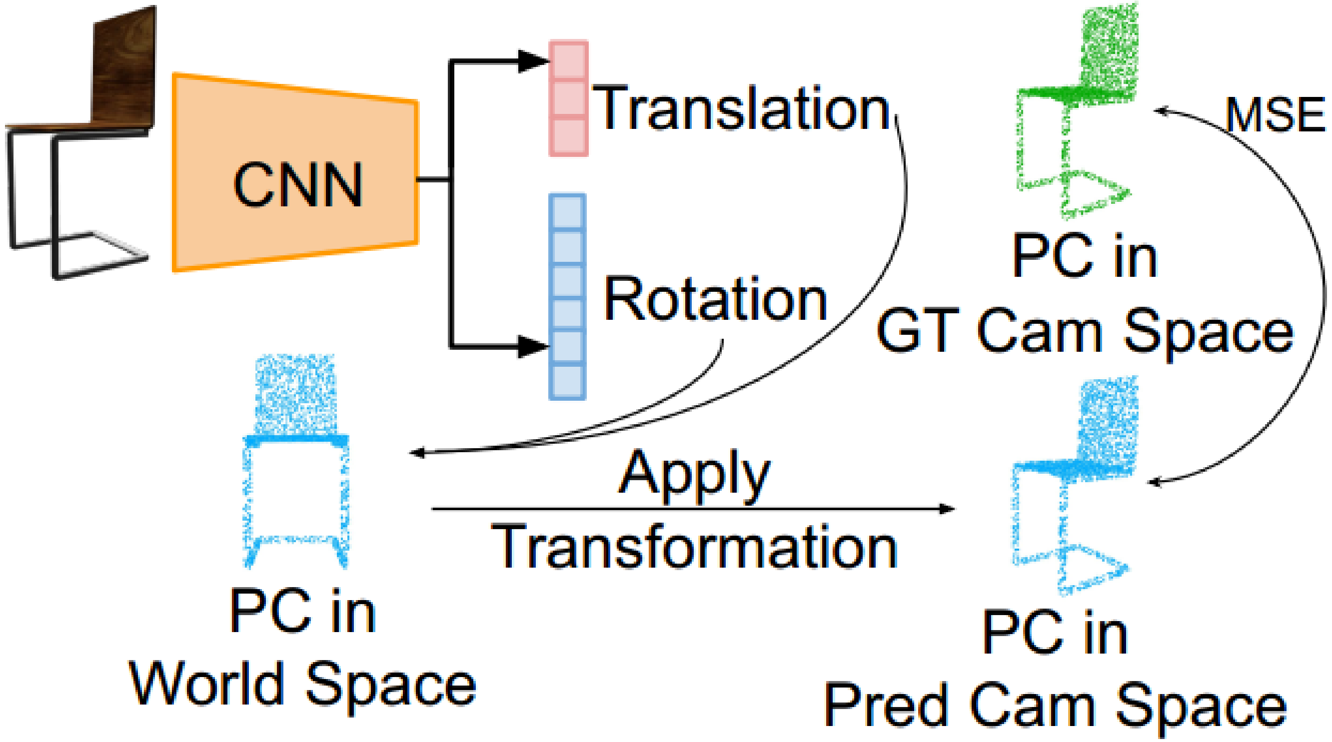

Figure 3: Given an image and a point **p**, we estimate the camera pose and project **p** onto the image plane. DISN uses the local features at the projected location, the global features, and the point features to predict the SDF of **p**. 'MLPs' denotes multi-layer perceptrons.

# Key Idea: Combine global and local features

- Local feature extraction

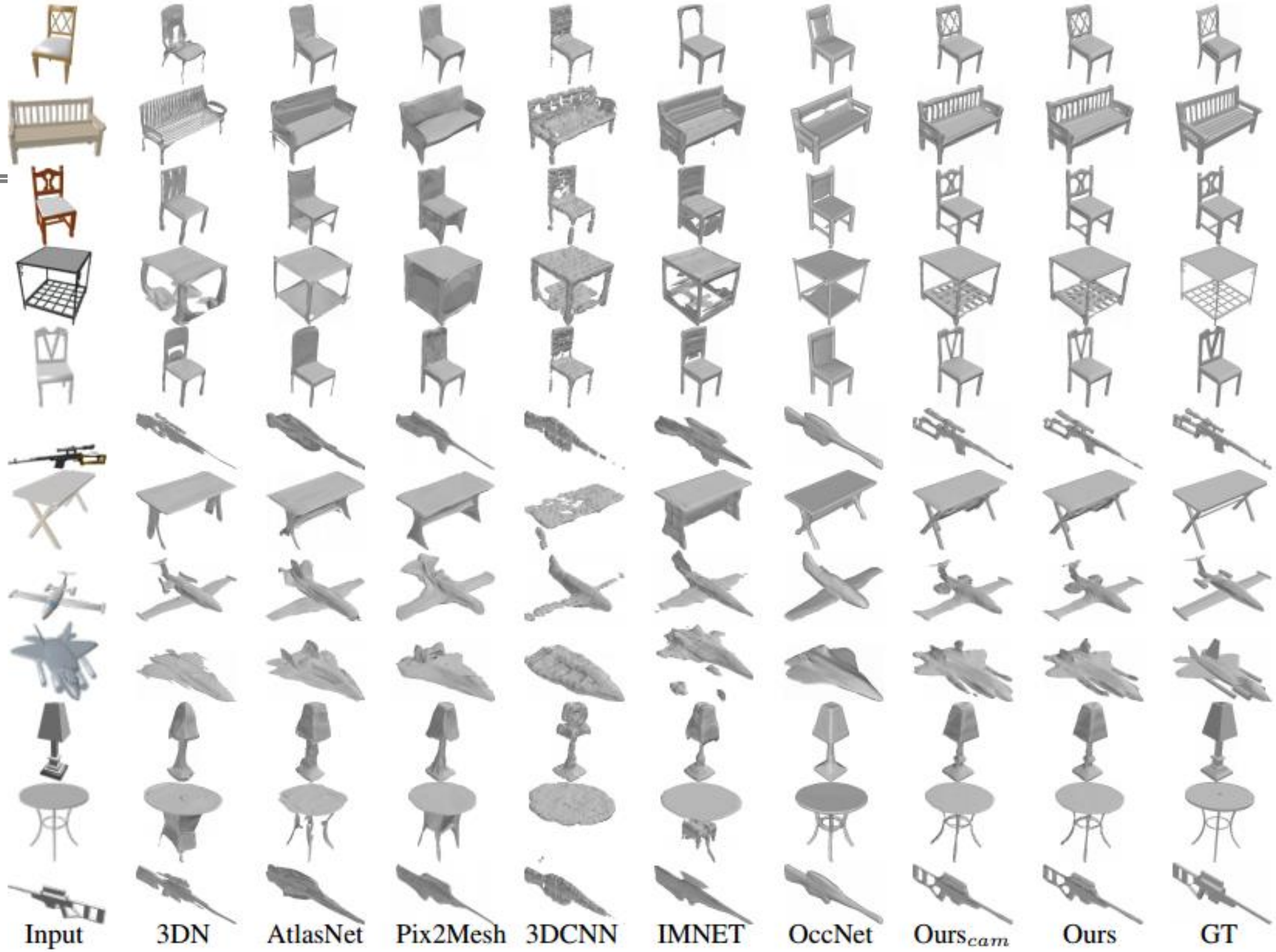


# Camera Pose Estimation Network




$$L_{cam} = \frac{\sum_{\mathbf{p}_w \in PC_w} ||\mathbf{p}_G - (\mathbf{R}\mathbf{p}_w + \mathbf{t}))||_2^2}{\sum_{\mathbf{p}_w \in PC_w} 1},$$

# Results


Input
3DN
AtlasNet
Pix2Mesh
3DCNN
IMNET
OccNet
Ours$_{cam}$
Ours
GT

# Deep Level Sets: Implicit Surface Representations for 3D Shape Inference

Mateusz Michalkiewicz[1], Jhony K. Pontes[1], Dominic Jack[1], Mahsa Baktashmotlagh[2], Anders Eriksson[2]

[1]School of Electrical Engineering and Computer Science, Queensland University of Technology
[2]School of Information Technology and Electrical Engineering, University of Queensland

# Minimal Oriented Surface Models

- (Screened) Poisson Surface Reconstruction (2006, 2013)

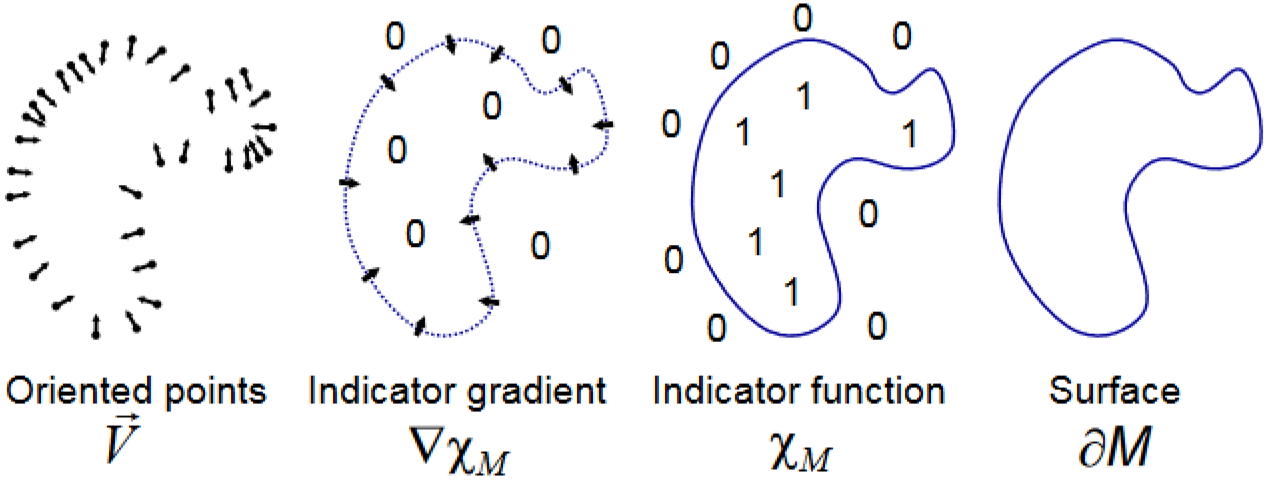



**Figure 1:** *Intuitive illustration of Poisson reconstruction in 2D.*

Solve the Poisson equation: $\Delta\tilde{\chi} = \nabla \cdot \vec{V}.$

# Minimal Oriented Surface Models

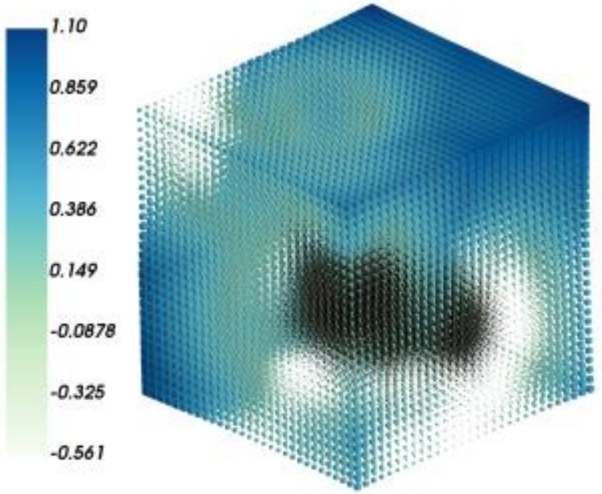
- (Screened) Poisson Surface Reconstruction (2006, 2013)

**Figure 2:** *Points from scans of the "Armadillo Man" model (left), our Poisson surface reconstruction (right), and a visualization of the indicator function (middle) along a plane through the 3D volume.*

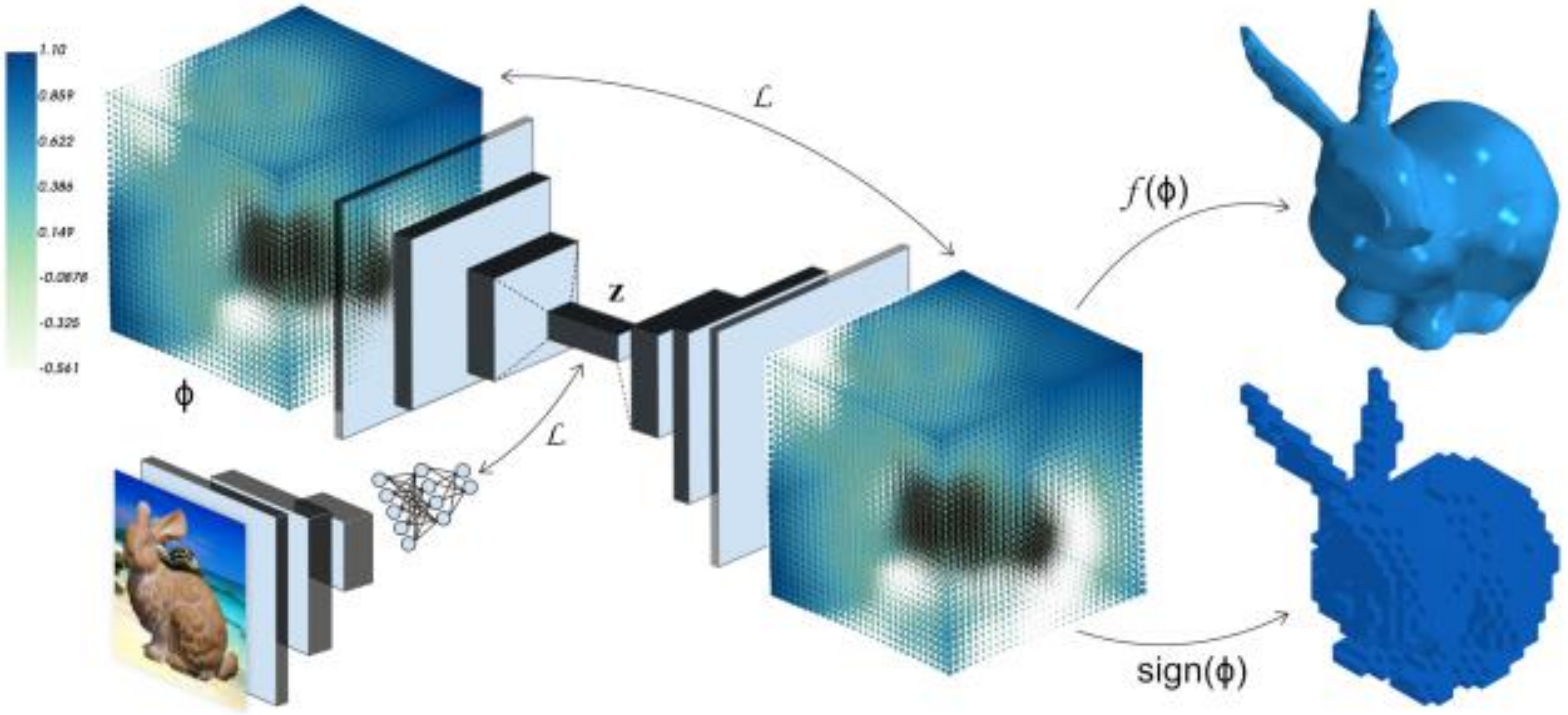
# Oriented Surface Learning

- Input: single image
- Output: points $\mathcal{X} = \{x_i\}_{i=1}^m$ with normals $\mathcal{N} = \{n_i\}_{i=1}^m$

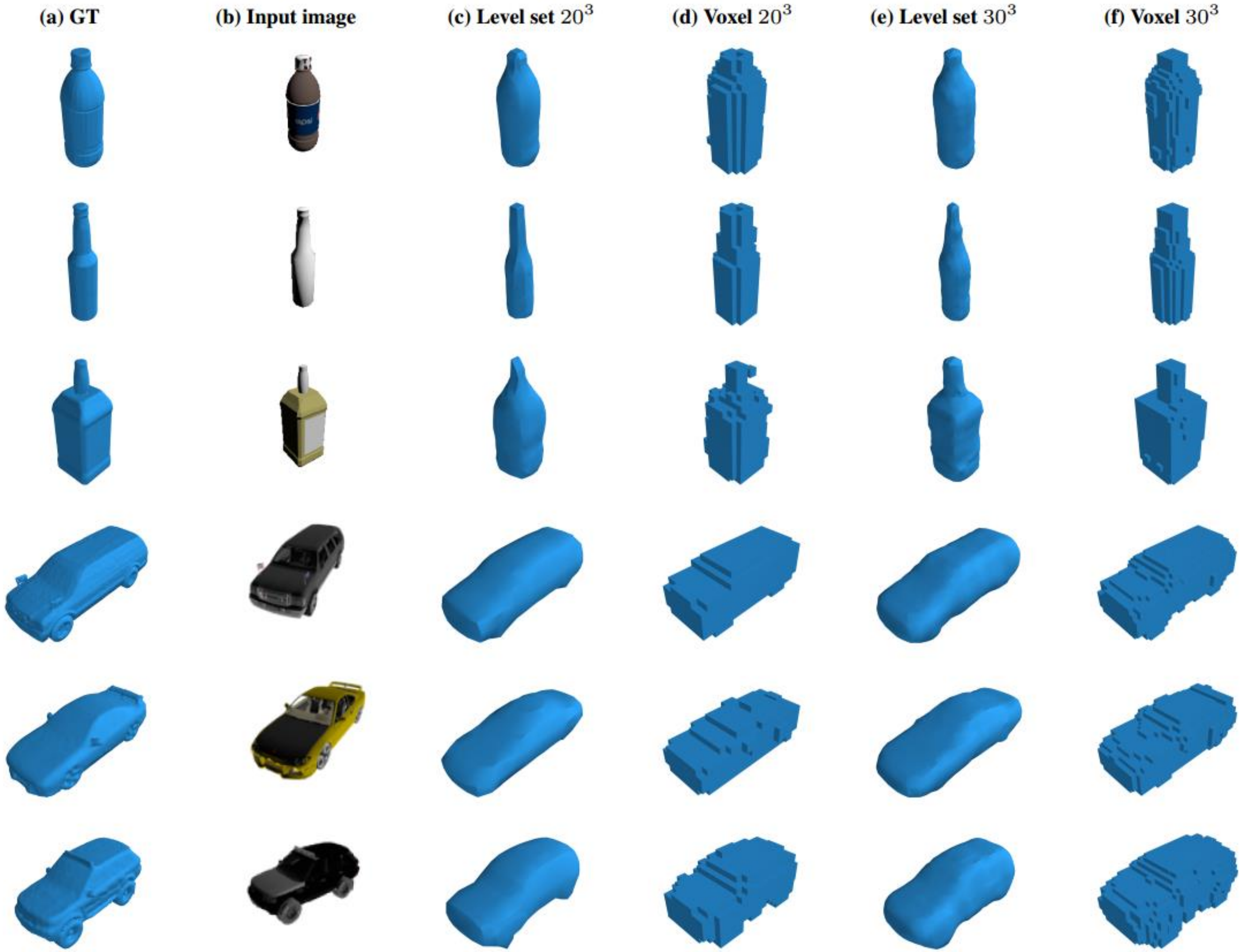
Level Set



# Network



# Results


(a) GT
(b) Input image
(c) Level set $20^3$
(d) Voxel $20^3$
(e) Level set $30^3$
(f) Voxel $30^3$

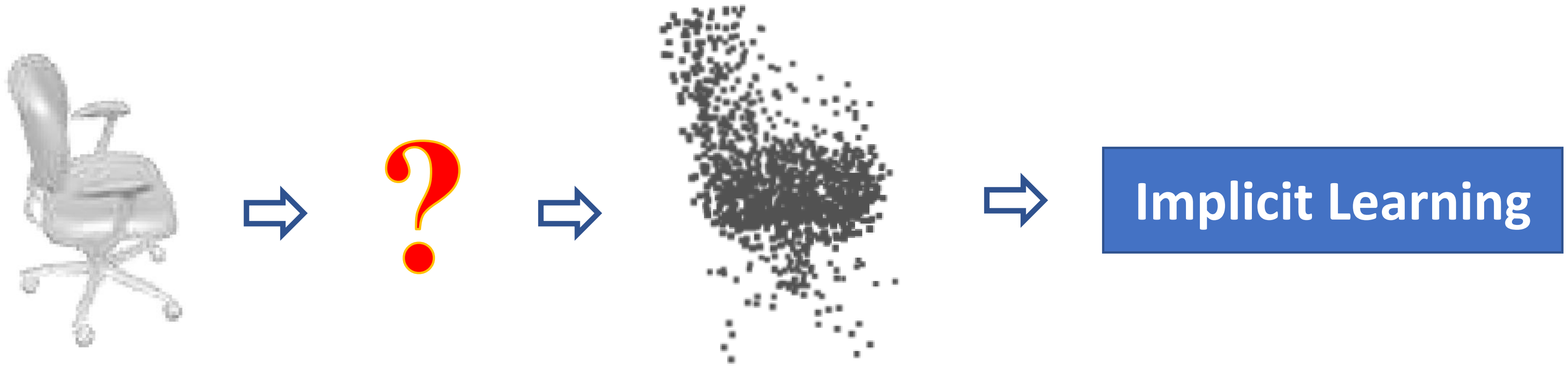
# Summary

- New shape representation learning for 3D reconstruction
- Lightweight network, but high-resolution output
- Easy to extract mesh with better properties
- …

# Limitation

- More training time
  - 15x CNN, 5x AtlasNet
- **Low inference efficiency**
  - OccNet: 3s per mesh
- Details depend on more points sampling (**sampling algorithm**)
  - uniform sampling results better than adaptive ones
- …

# Idea: sampling learning



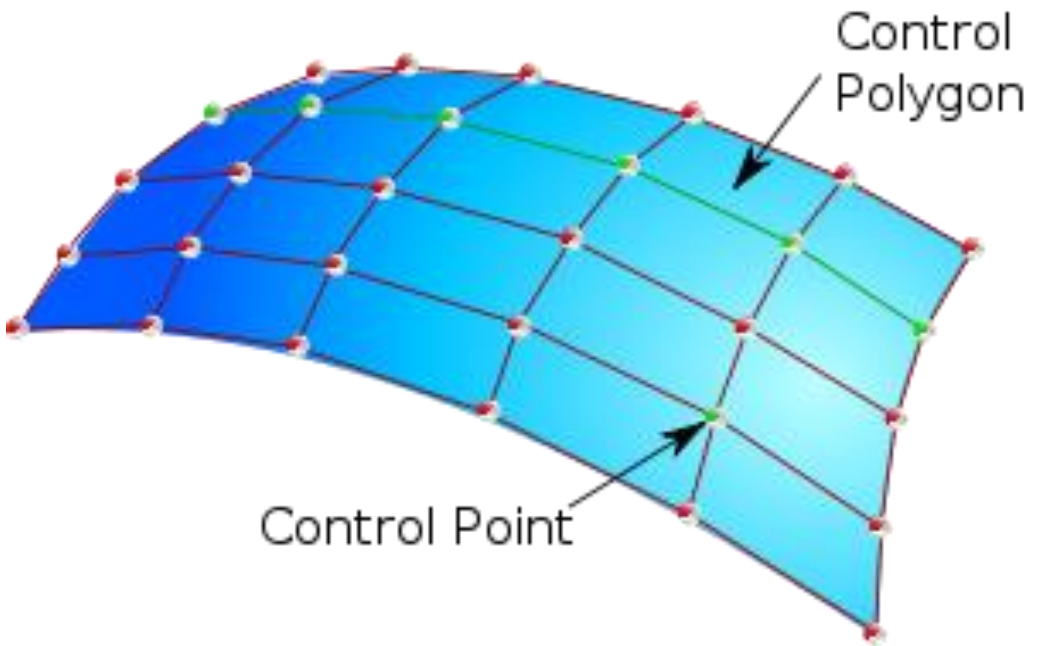

input image

inferred sampling points

# Idea: spline learning

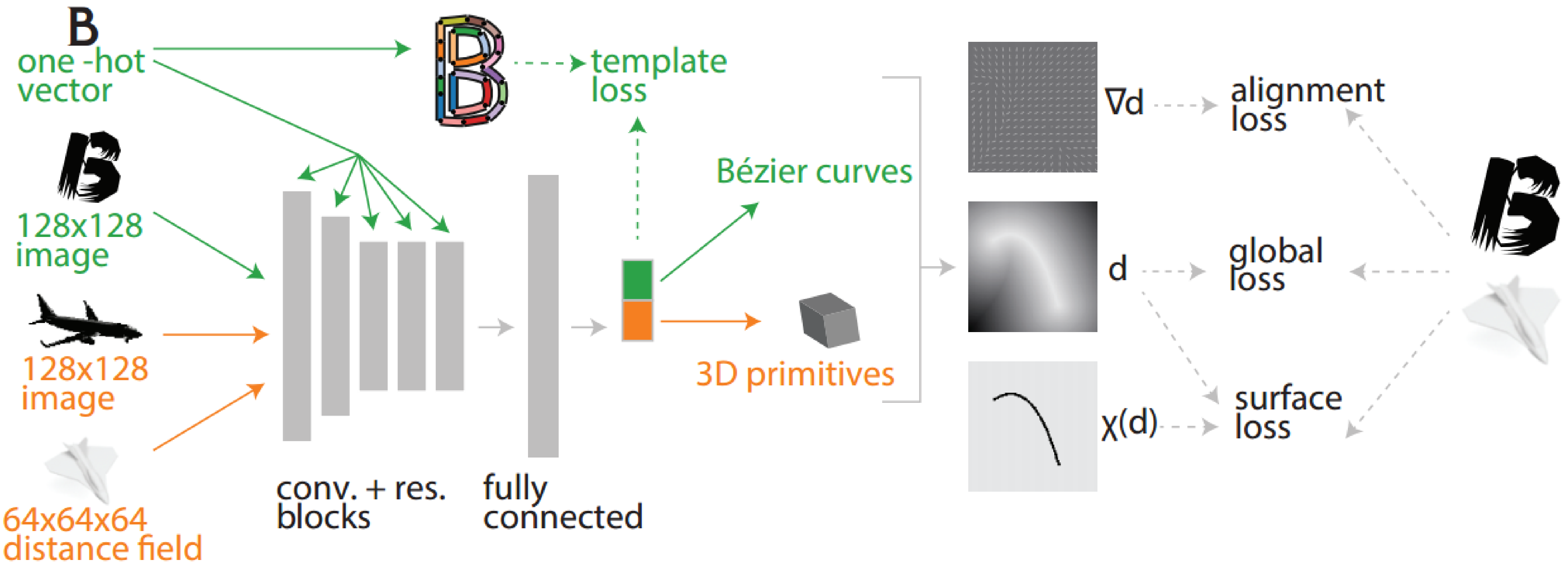
- Bezier surface
- B-Spline surface
- Non-Uniform Rational B-Spline (NURBS)

$$S(u,v) = \sum_{i=1}^{k} \sum_{j=1}^{l} R_{i,j}(u,v) P_{i,j}$$



# Idea: spline learning

- ***Deep Parametric Shape Predictions using Distance Fields***, *Smirnov et al. 2019*



# Idea: attributes learning

- Semantic labels for segmentation
- UV-coordinates/point color for texture generation
- …

# Thanks!