# ARO-Net: Learning Implicit Fields from Anchored Radial Observations

Yizhi Wang[1,2*], Zeyu Huang[1*], Ariel Shamir[3], Hui Huang[1], Hao (Richard) Zhang[2], Ruizhen Hu[1†]

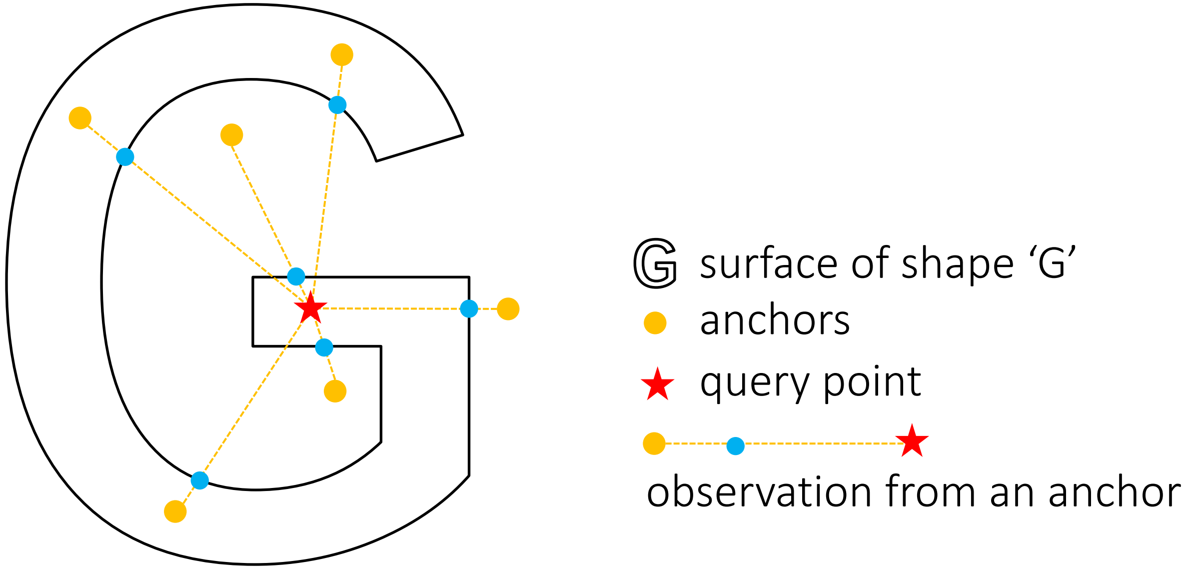
[1]Shenzhen University [2]Simon Fraser University [3]Reichman University

(* equal contribution, † corresponding author)

Code: https://github.com/yizhiwang96/ARO-Net

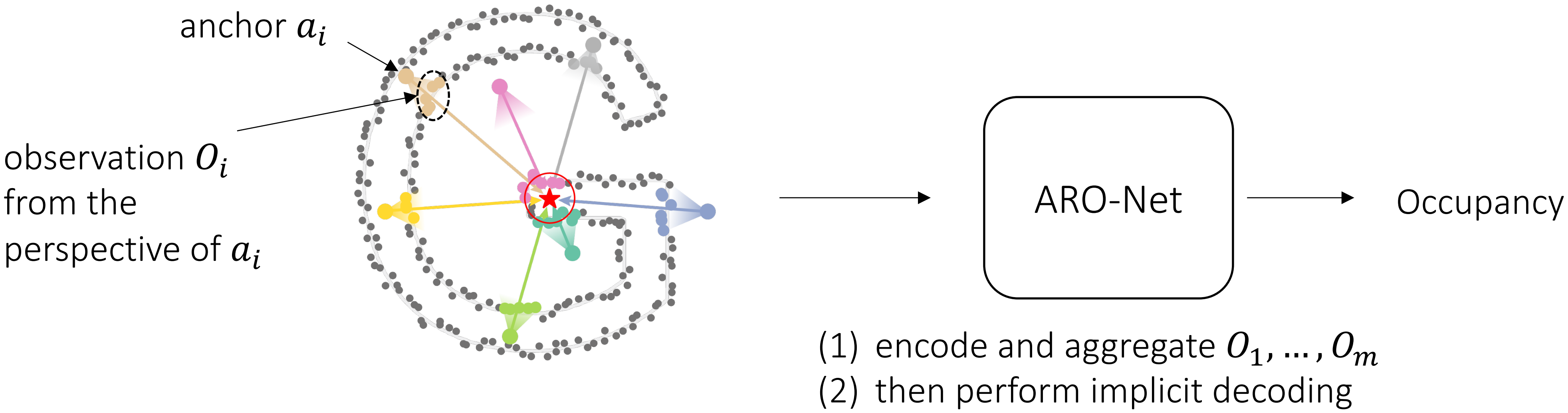
# Anchored Radial Observations (ARO)

- A novel shape encoding for learning implicit field representation of shapes that is category-agnostic and generalizable
- The core idea is to reason about shapes through partial observations from a set of viewpoints, called *anchors*



# Anchored Radial Observations (ARO)

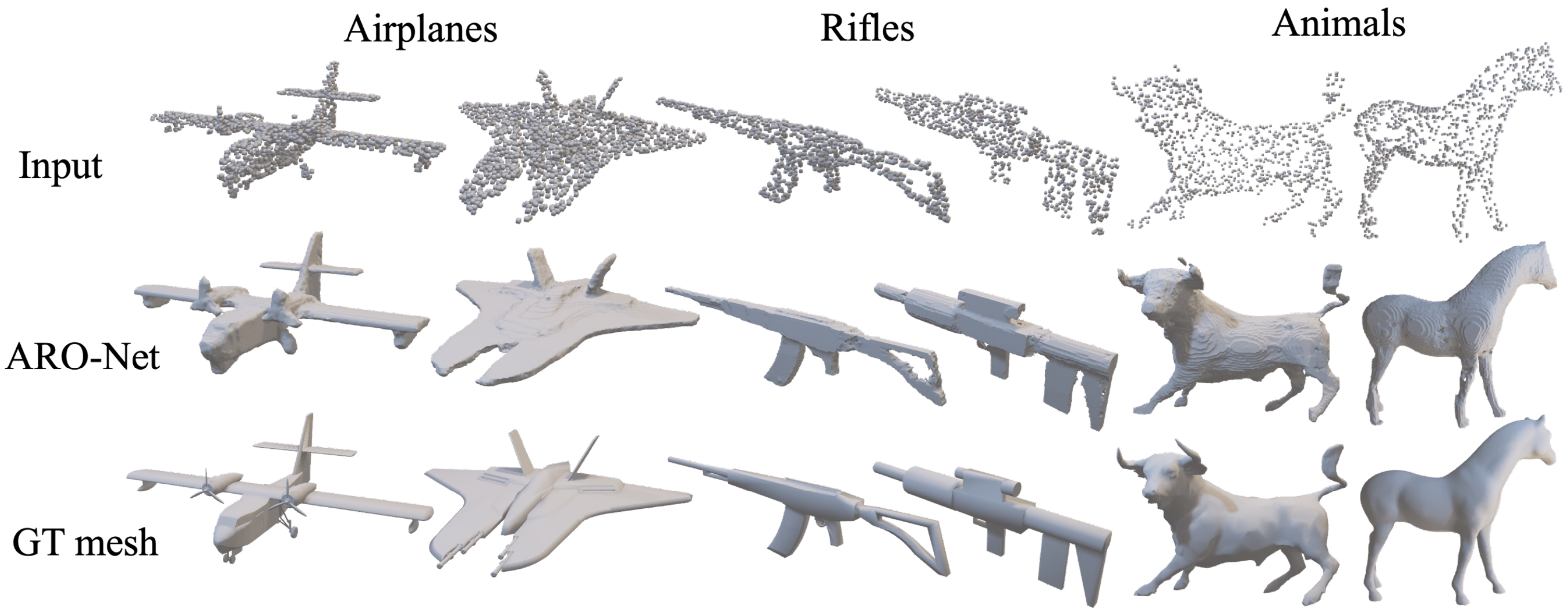
- We design ARO-Net, to predict the occupancy value of a query point in space from input point cloud




Input point cloud of ‘G’

# Anchored Radial Observations (ARO)

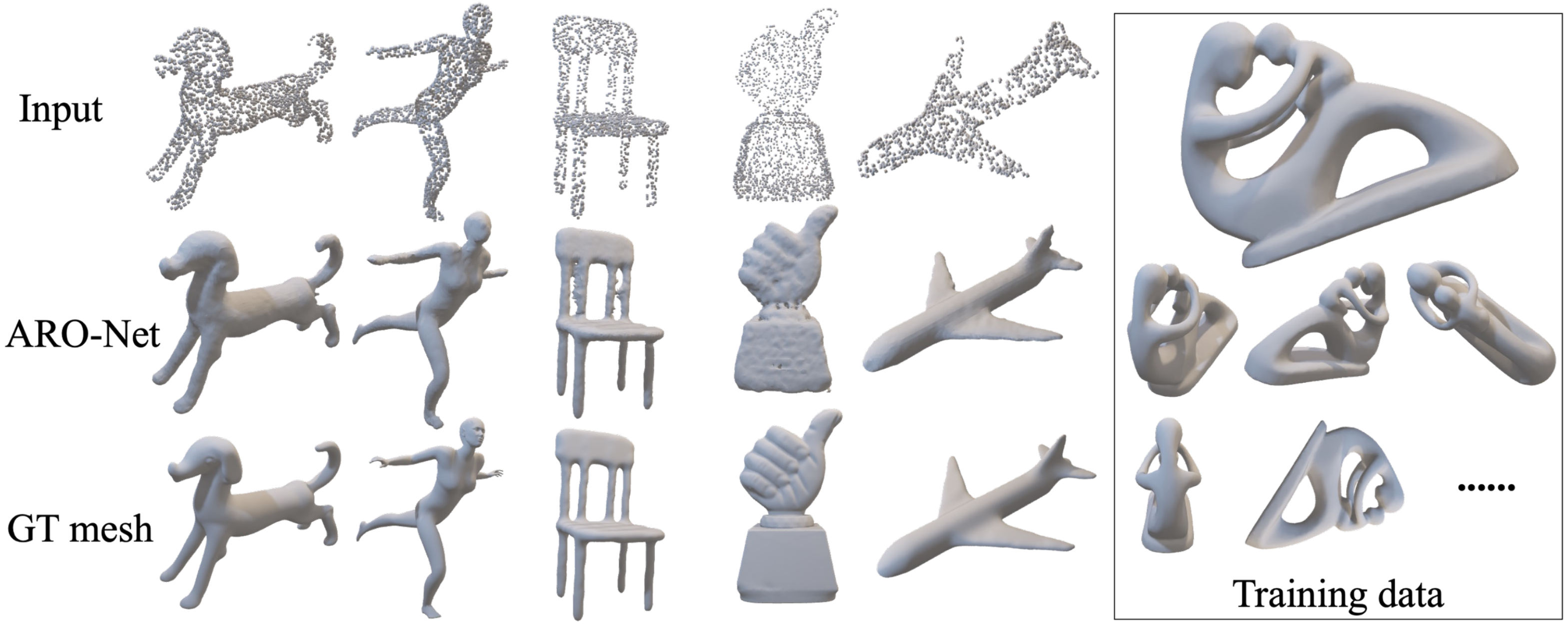
- ARO-Net is category-agnostic and generalizable amid significant shape variations




3D Reconstruction from sparse point clouds by ARO-Net *trained on 4K chairs*

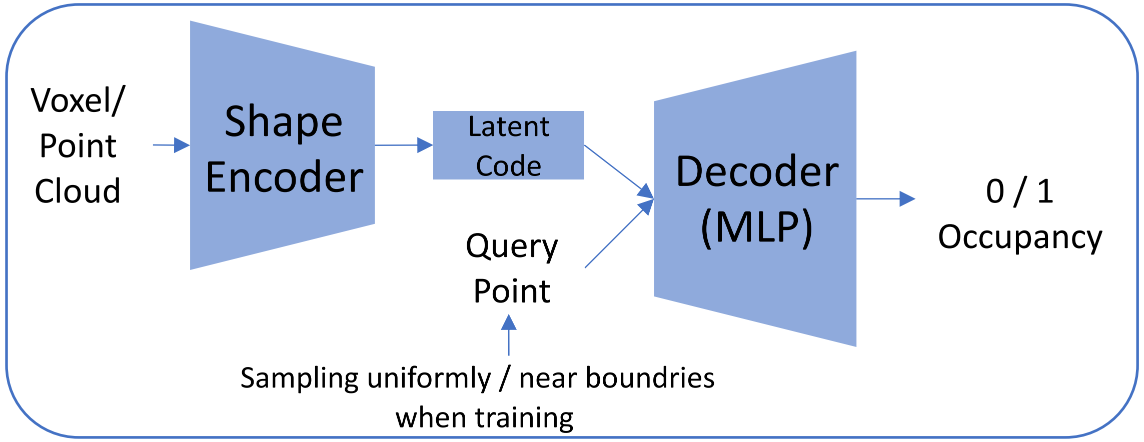
# Anchored Radial Observations (ARO)

- ARO-Net is category-agnostic and generalizable amid significant shape variations

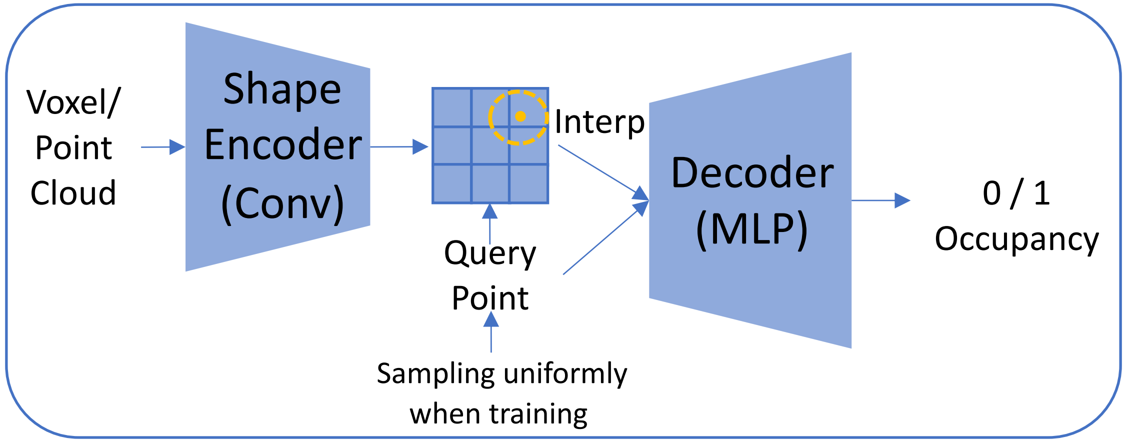



ARO-Net trained on *only the Fertility model* with rotation and scaling

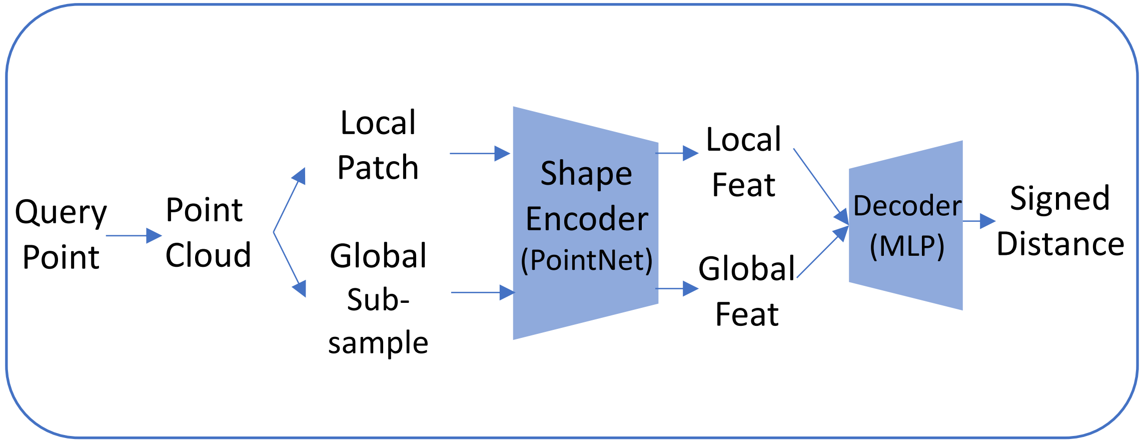
# Related work



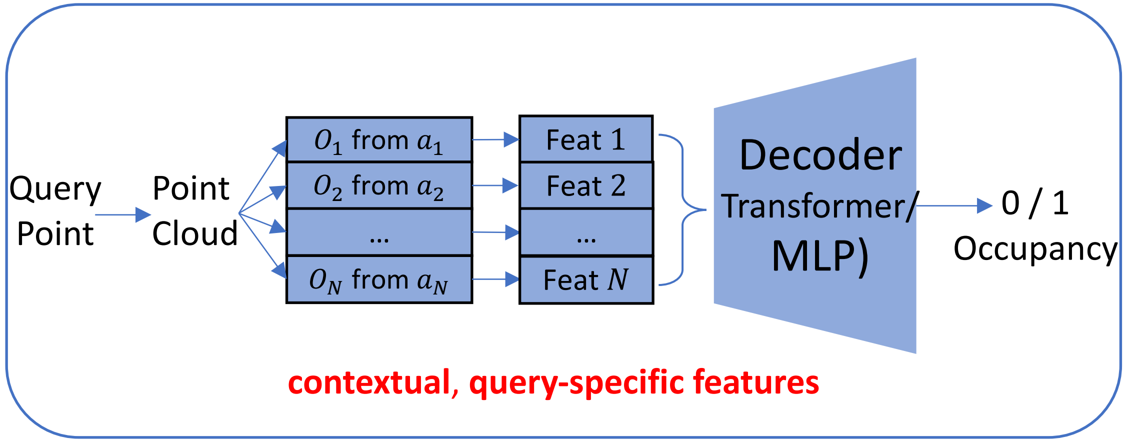

IM-Net / Occ Net (CVPR 19)




ConvONet (ECCV 20)




Points2Surf (ECCV 20)

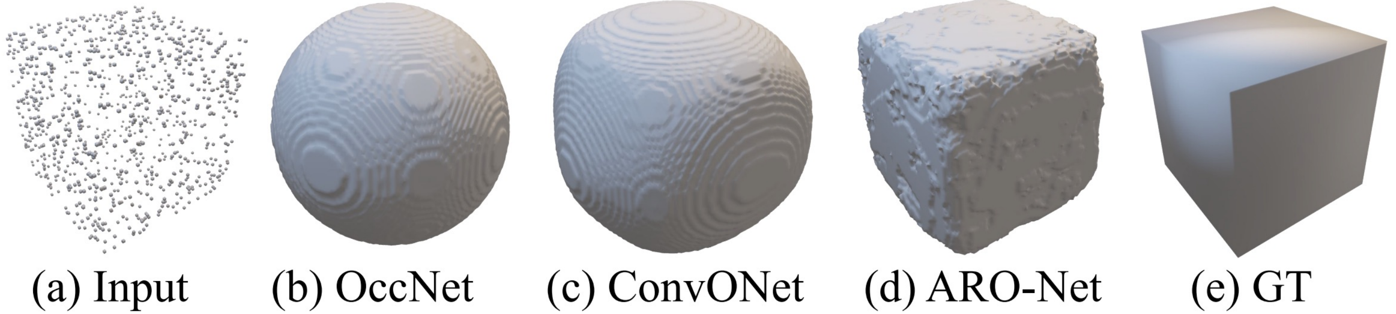



**ARO-Net**

(anchors: $a_1, a_2, \ldots, a_N$, $O_i$: radial observation from $a_i$)

# Related work

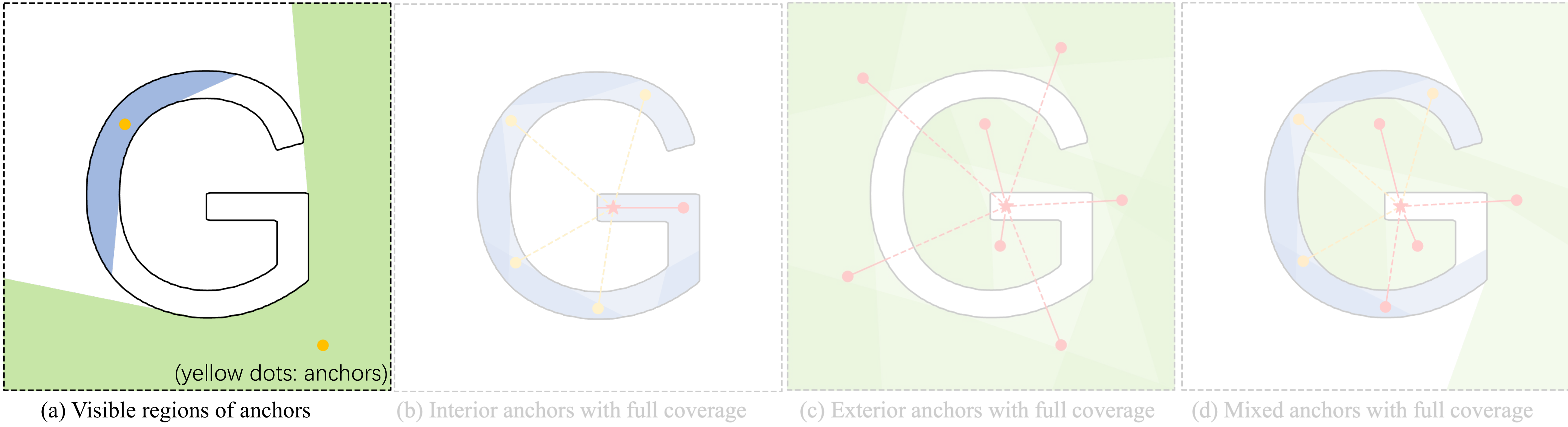
- Differently from prior neural implicit models, that use **global** or **local grid** shape feature, our shape encoder operates on **contextual**, **query-specific** features




3D reconstruction from a sparse point cloud of a **cube** (a), with training on a single **sphere**

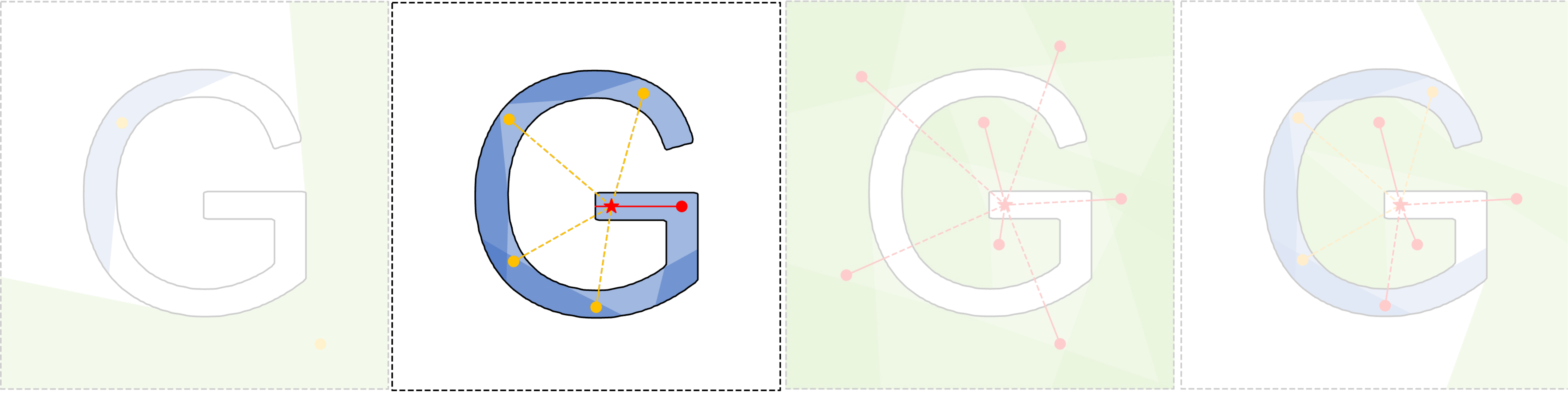
# Method

# Anchor visibility and point occupancy




The visible regions of different anchors,
colored in blue for interior anchors and green for exterior anchors

# Anchor visibility and point occupancy



(a) Visible regions of anchors

(b) Interior anchors with full coverage

(c) Exterior anchors with full coverage

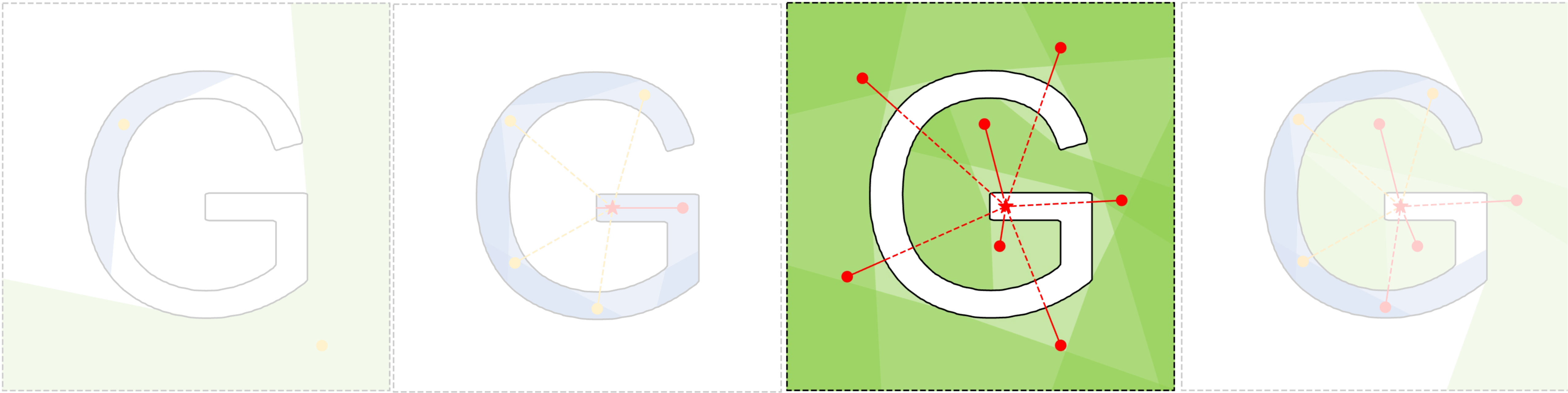
(d) Mixed anchors with full coverage

A query point (red star) is inside the shape

$\Leftrightarrow$

it is covered by at least one radial observation (the red anchor)

# Anchor visibility and point occupancy



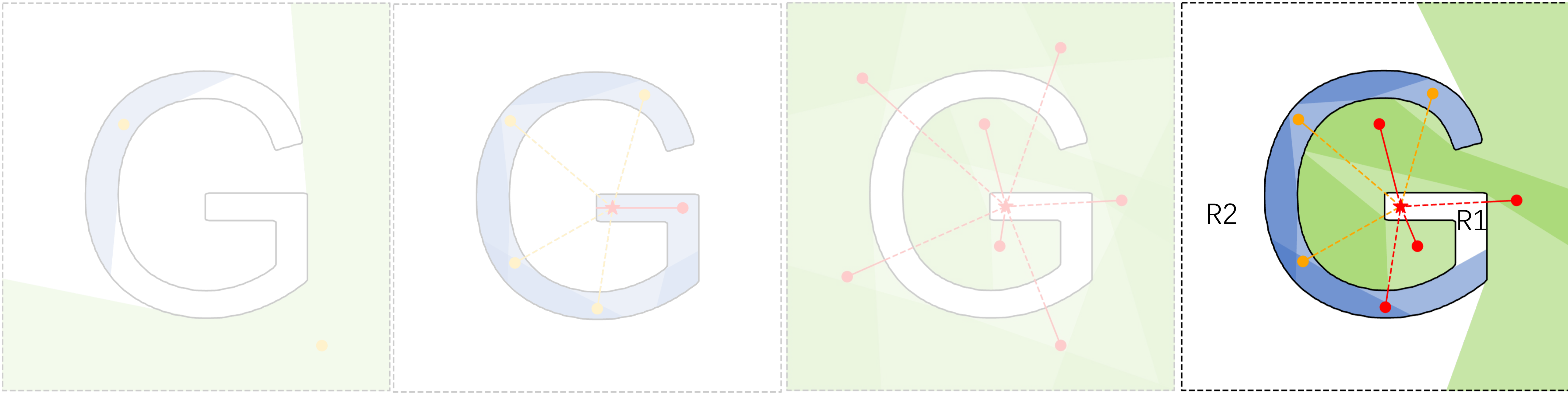
(a) Visible regions of anchors (b) Interior anchors with full coverage (c) Exterior anchors with full coverage (d) Mixed anchors with full coverage

A query point (red star) is inside the shape

⇔

it is not covered by any of the visible regions

# Anchor visibility and point occupancy




(a) Visible regions of anchors (b) Interior anchors with full coverage (c) Exterior anchors with full coverage (d) Mixed anchors with full coverage

If the query point is

- in the **blue** or **green** regions: the situations are the same as **(b)** or **(c)**
- in **R1:** R1 is bounded by the visible regions of the **red** anchors, the situation turns into **(c)**
- in **R2:** R2 is bounded by some anchors and the virtual bounding box, the situation turns into **(b)**

# Anchor visibility and point occupancy

If we have a set of anchors that together observe the entire surface of the shape:

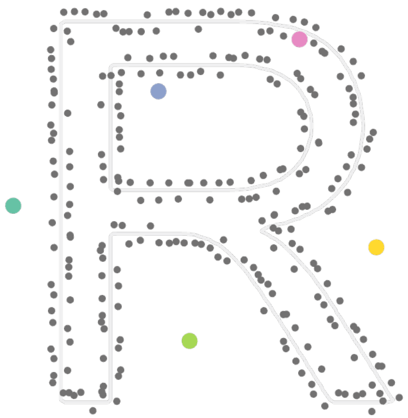
The inside/outside judgment of a spatial point relative to the shape ⇔ The point's relationship to the radial observations of the anchors

# Anchor visibility and point occupancy

What if the shape is only **partially** observed by the anchors
& we don’t know whether each anchor is inside or outside the shape?

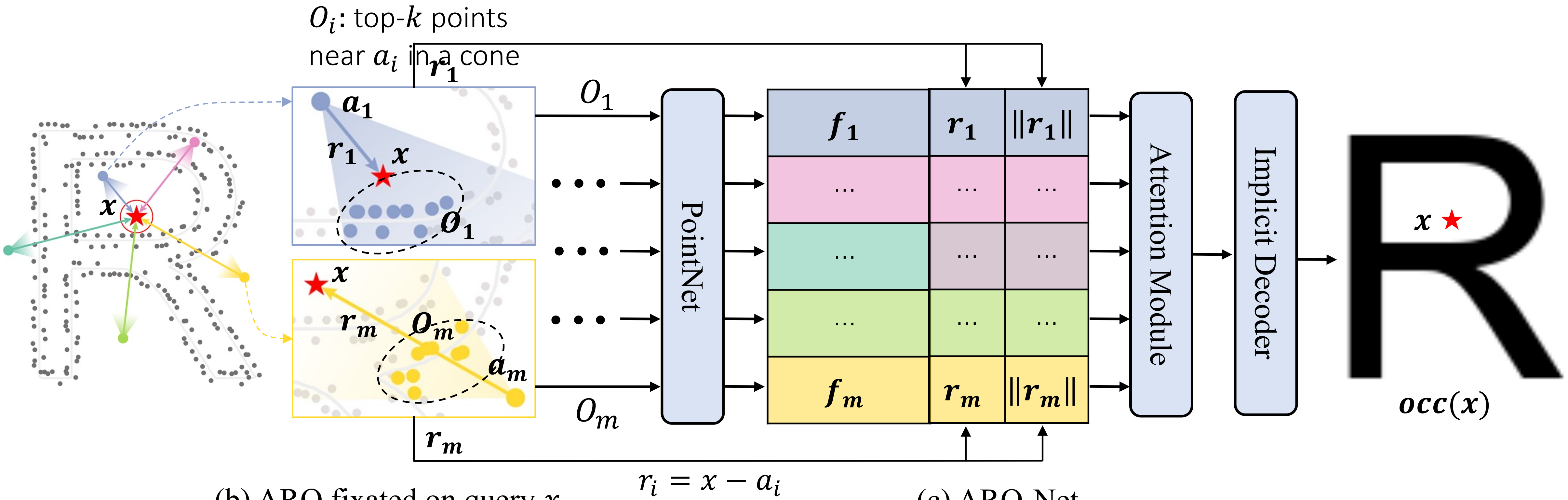
The inside/outside judgment of a spatial point relative to the shape $\xleftarrow{\text{learn}}$ ARO-Net $\xleftarrow{\text{Input}}$ The point’s coordinates & the radial observations of the anchors

# ARO Network



(a) Anchors and input point cloud

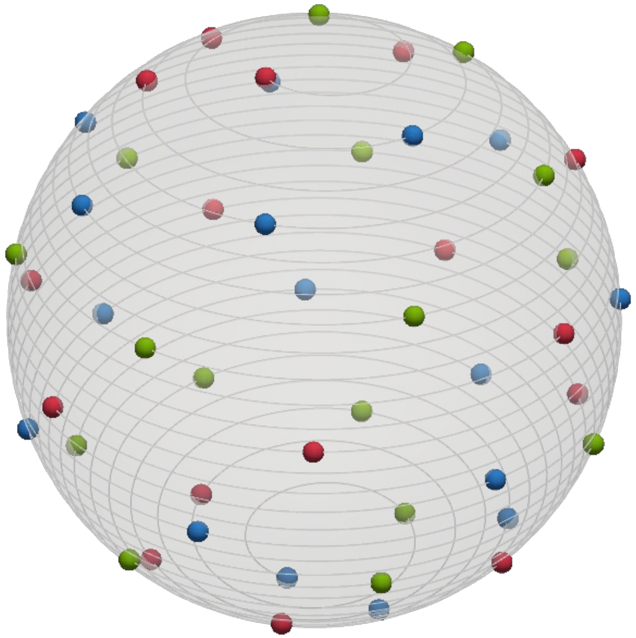
# ARO Network



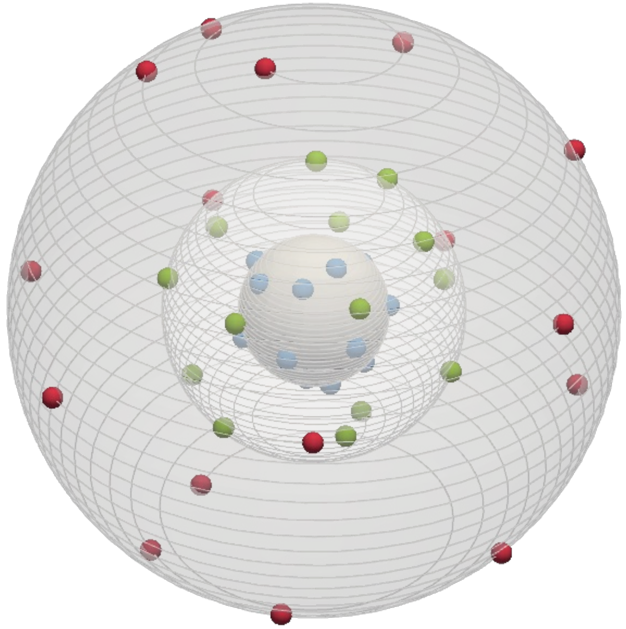

(b) ARO fixated on query $x$

(c) ARO-Net

# Anchor Placement



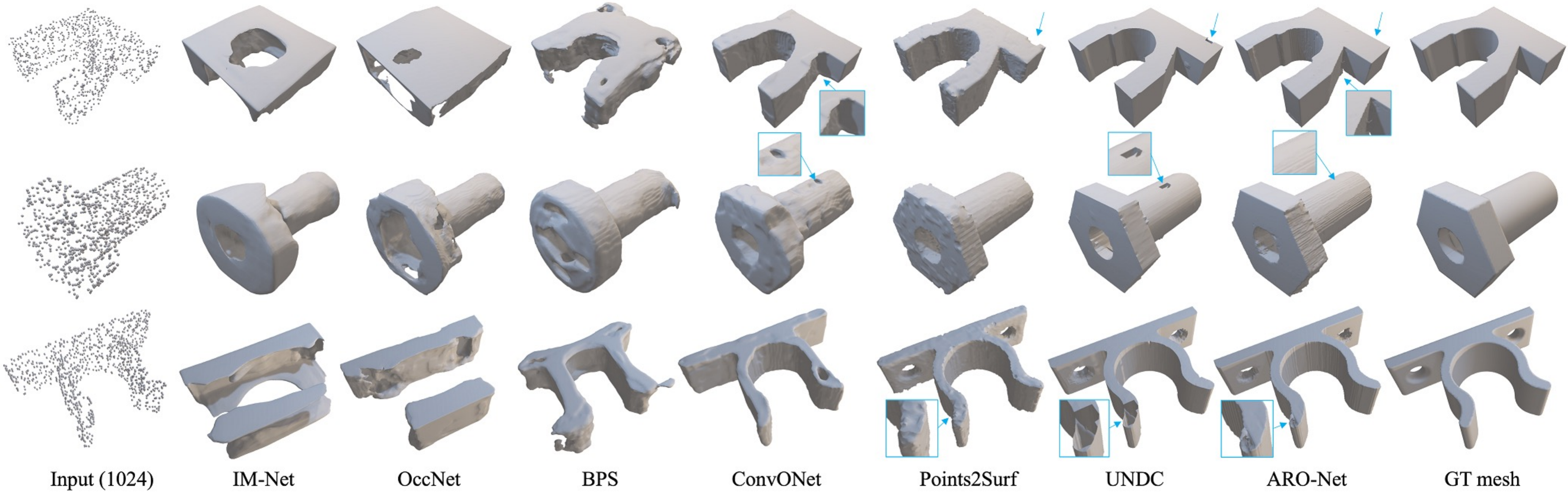
(1) uniformly sample $m$ (48 by default) points on the surface of the unit sphere with radius $r = 1/2$ using Fibonacci Sampling



(2) for the points with index $i\%3 = 1$, move them to the sphere with radius $r = 1/4$; for the points with index $i\%3 = 2$, further move them to the sphere with radius $r = 1/8$

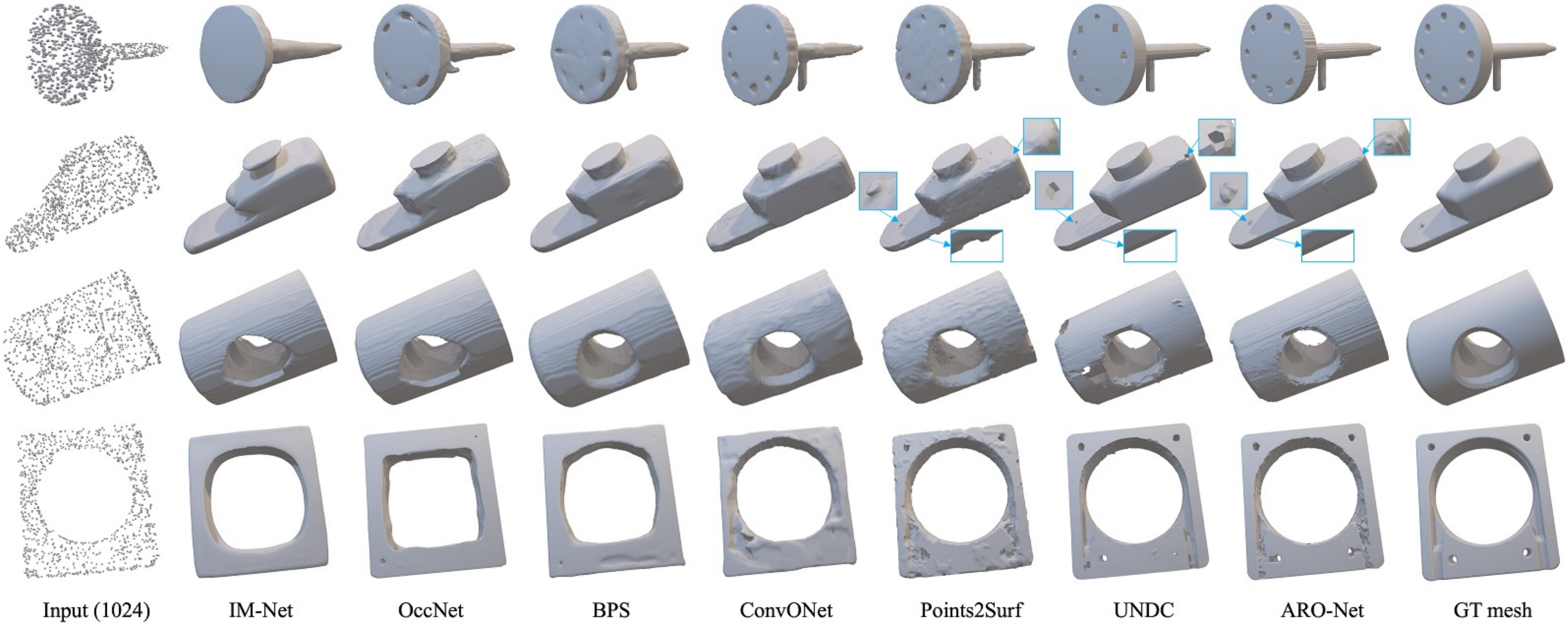
# Experiments

# Visual Comparisons - ABC

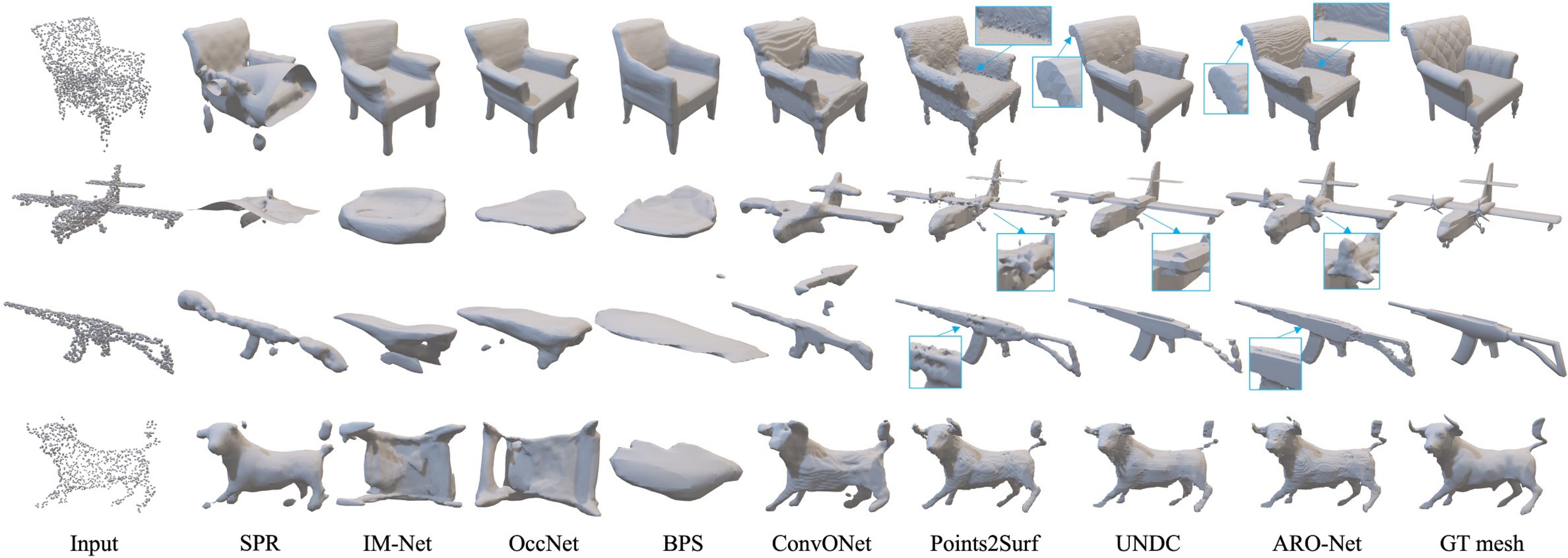



ARO-Net is able to reconstruct both **global structure** and **local detail** well, including **sharp features** as UNDC, and its results exhibit the least amount of artifacts in terms of holes and noise

# Visual Comparisons - ABC

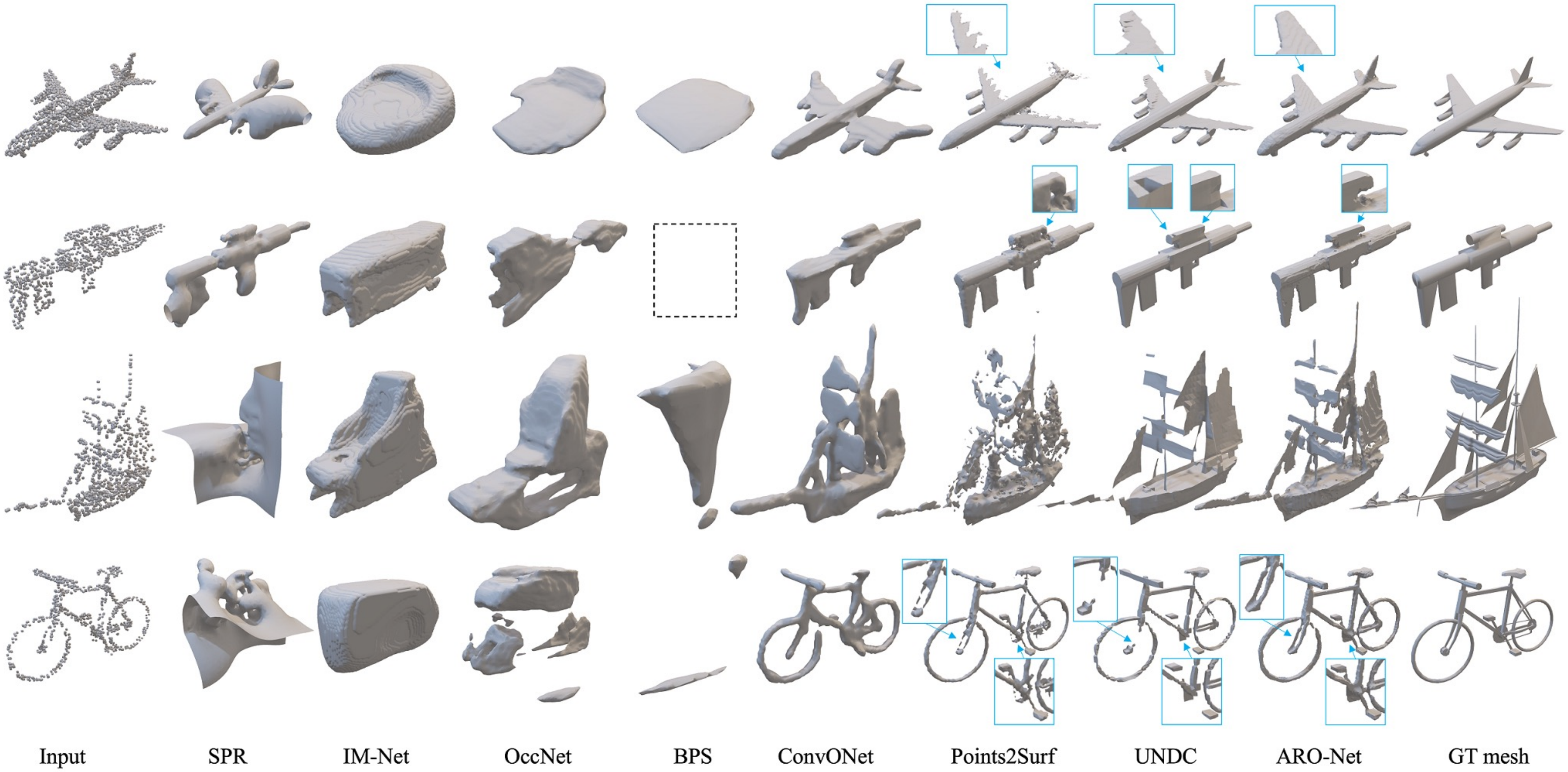


# Visual Comparisons – ShapeNet etc.




All methods (except SPR) were trained on **chairs** in ShapeNet and tested on chairs, airplanes, and rifles in ShapeNet V1 and animals in PSB.

# Visual Comparisons – ShapeNet etc.



# Quantitative Comparisons

| Method | LFD↓ | HD↓ | CD↓ | EMD↓ | IOU↑ |
|---|---|---|---|---|---|
| IM-Net [8] | 3.27 | 11.96 | 57.00 | 11.40 | 3.63 |
| OccNet [23] | 2.49 | 11.95 | 54.68 | 11.60 | 3.51 |
| BPS [26] | 2.64 | 5.03 | 11.72 | 2.66 | 6.56 |
| ConvONet [25] | 2.69 | 3.87 | 8.43 | 1.71 | 6.78 |
| Points2Surf [11] | 1.64 | <u>2.75</u> | 5.69 | 1.25 | <u>8.36</u> |
| UNDC [7] | **1.25** | 2.78 | **4.90** | <u>1.17</u> | 8.21 |
| ARO-Net | <u>1.35</u> | **2.25** | <u>5.46</u> | **1.12** | **8.79** |

Trained on ABC, tested on ABC

| Method | LFD↓ | HD↓ | CD↓ | EMD↓ |
|---|---|---|---|---|
| IM-Net [8] | 3.18 | 9.04 | 13.84 | 15.40 |
| OccNet [23] | 2.59 | 7.77 | 12.36 | 13.47 |
| BPS [26] | 4.31 | 12.28 | 20.51 | 23.30 |
| ConvONet [25] | <u>2.12</u> | 6.22 | 11.20 | 12.30 |
| Points2Surf [11] | 2.51 | 6.46 | 8.60 | <u>7.34</u> |
| UNDC [7] | 2.19 | <u>5.60</u> | **6.06** | **6.80** |
| ARO-Net | **1.92** | **5.33** | <u>7.14</u> | 7.70 |

Trained on chairs, tested on chairs

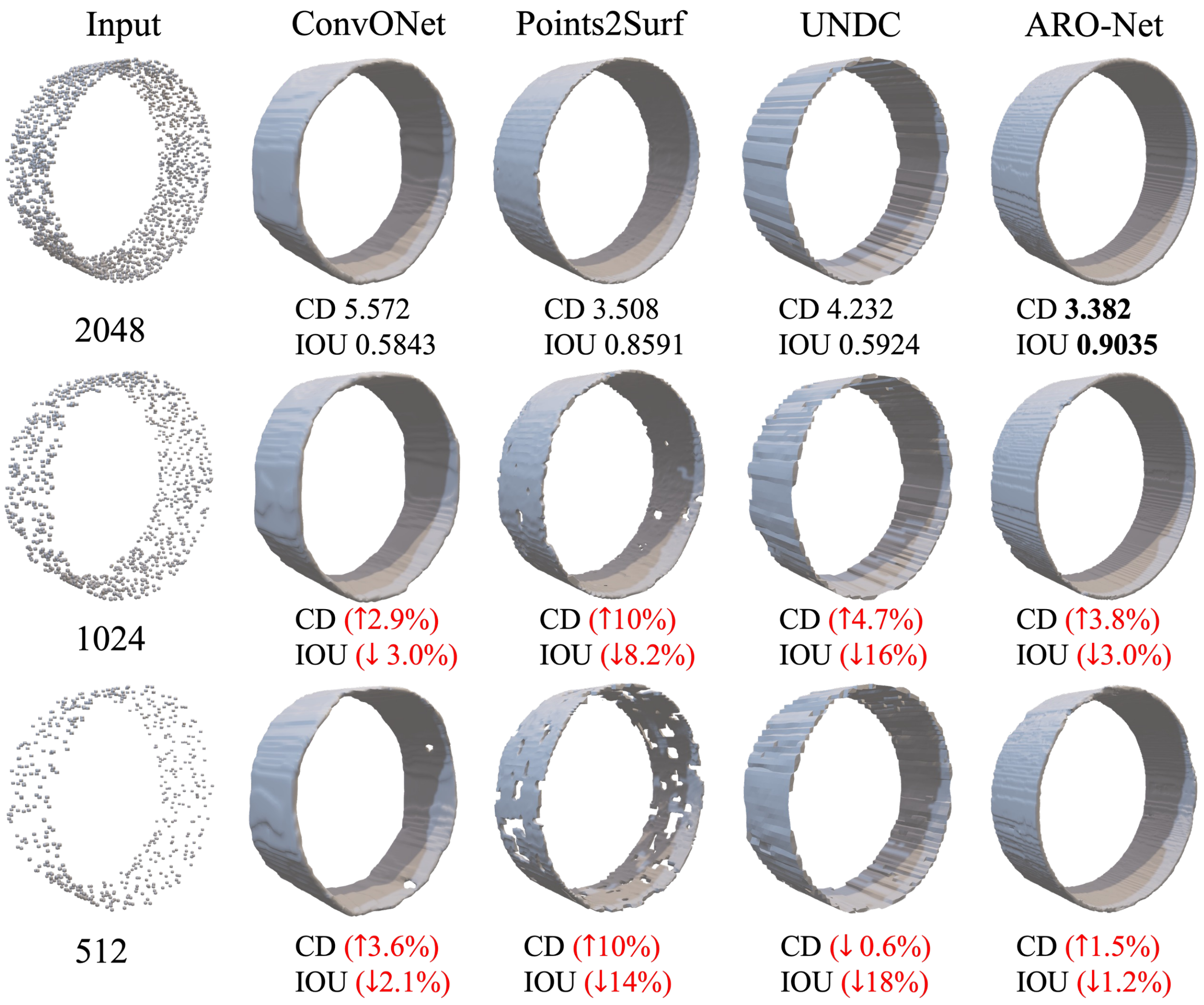
| Method | LFD↓ | HD↓ | CD↓ | EMD↓ |
|---|---|---|---|---|
| IM-Net [8] | 14.30 | 19.50 | 44.84 | 41.38 |
| OccNet [23] | 12.31 | 18.86 | 39.75 | 38.57 |
| BPS [26] | 13.73 | 19.48 | 38.10 | 36.61 |
| ConvONet [25] | 5.69 | 15.98 | 13.50 | 10.75 |
| Points2Surf [11] | 5.48 | 4.70 | 4.86 | 5.76 |
| UNDC [7] | <u>3.62</u> | <u>4.40</u> | **3.98** | **3.98** |
| ARO-Net | **3.56** | **4.32** | <u>4.61</u> | <u>4.73</u> |

Trained on chairs, tested on airplanes

LFD: light filed distance
HD: hausdoff distance
CD: chamfer distance
EMD: earth mover's distance
IOU: occupancy IOU

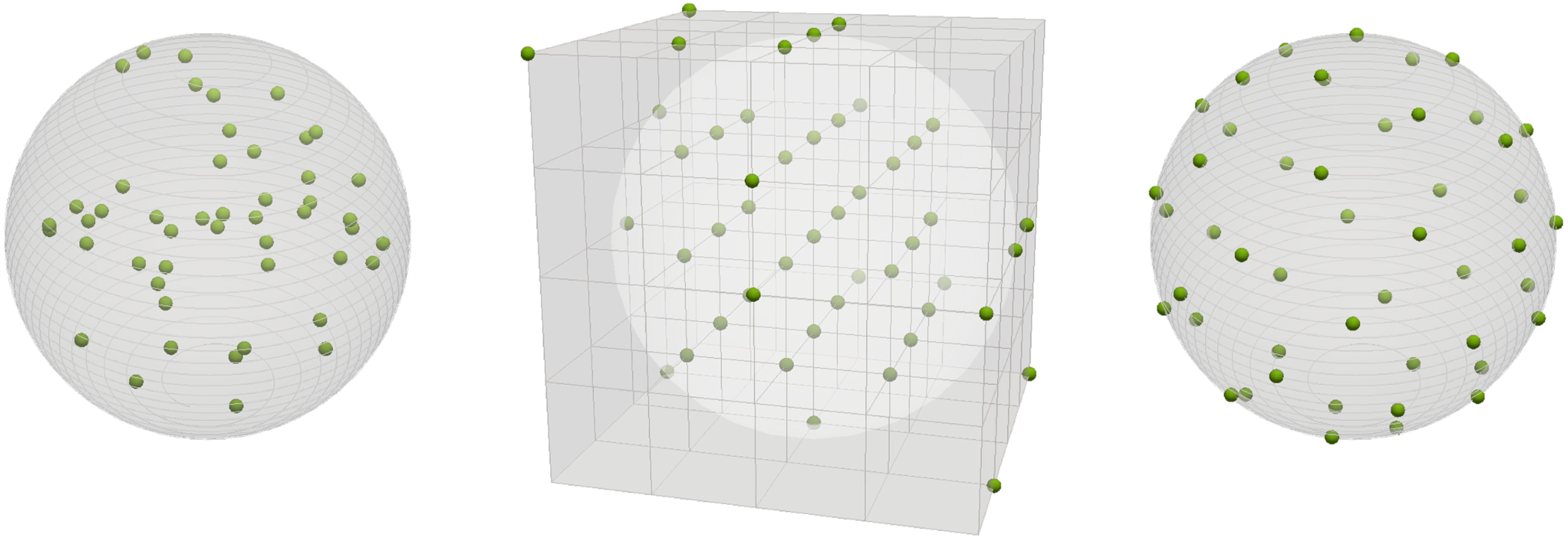
- ARO-Net exhibits a clear advantage in
  - IOU, EMD and HD on ABC
  - LFD and HD on ShapeNet
- For the remaining CD metric, ARO-Net is still a close runner-up

# Robustness Against Sparsity




- Once trained with 2,048-point input, ARO-Net can produce quality results with sparser inputs ***without re-training***.
- The reconstruction quality of both UNDC and Points2Surf degrades dramatically when the point num decreases from 2,048 to 512

# Different Anchor Placement Strategies

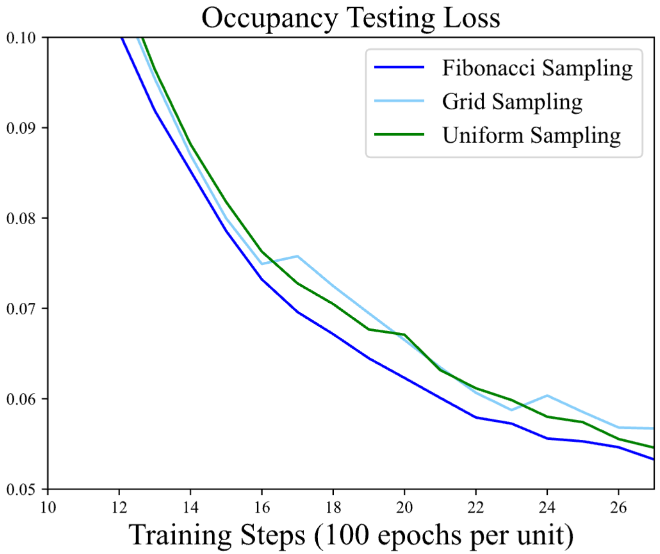


Uniform Sampling Gird Sampling Fibonacci Sampling

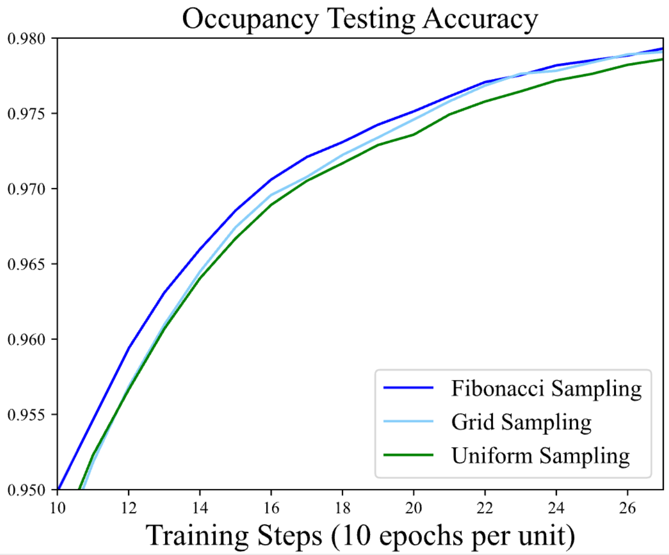
In uniform sampling, we randomly sample $m = 48$ points in a unit sphere as anchors

In grid sampling, we first choose all grid points $\{(x, y, z) | \, x, y, z \in \{-0.5, -0.25, 0, 0.25, 0.5\}\}$ inside the unit sphere and then randomly select from remaining grid points to make a total count of 48

# Different Anchor Placement Strategies




Grid sampling makes training relatively unstable

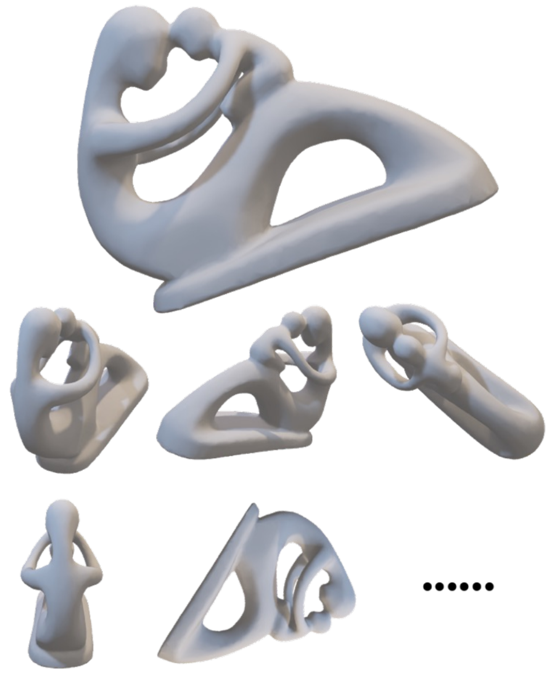



Fibonacci sampling makes training the smoothest

| Setting | LFD↓ | HD↓ | CD↓ | EMD↓ | IOU↑ |
|---|---|---|---|---|---|
| Uniform | 1.52 | 2.32 | 6.29 | 1.18 | 8.67 |
| Gird | 1.42 | 2.20 | 5.63 | 1.11 | 8.76 |
| Fibonacci | 1.35 | 2.25 | 5.46 | 1.12 | 8.79 |

Fibonacci sampling produces the overall best performance

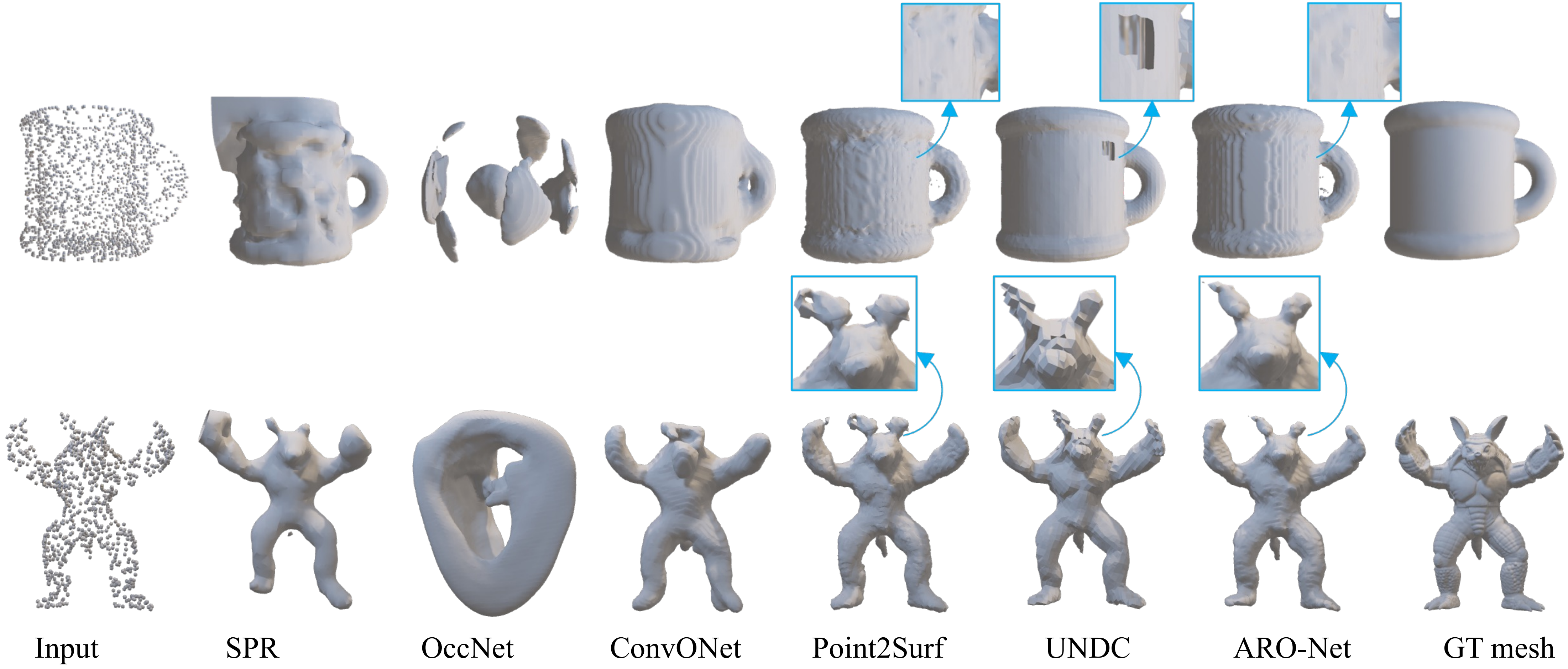
# Visual Comparisons – One Shape Training




Training data

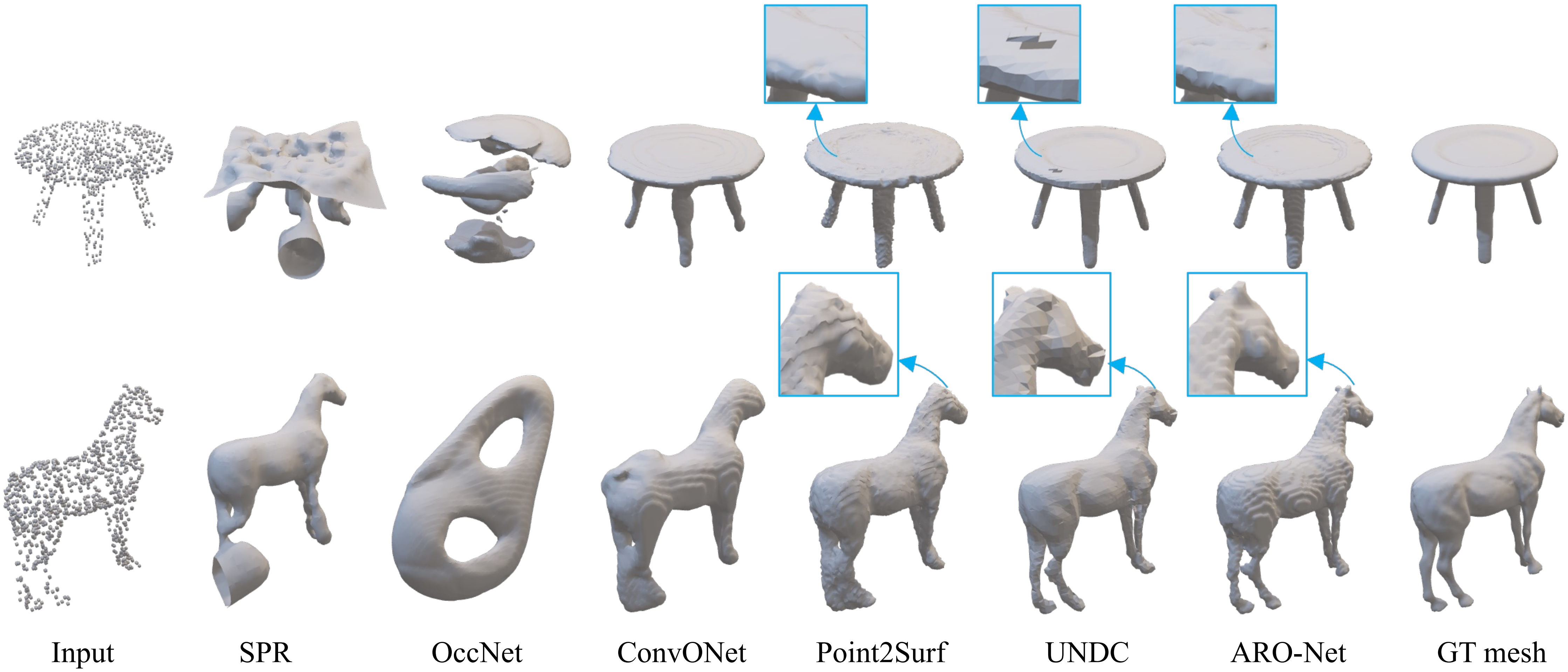
- One shape: the Fertility
- Data augmentations
  - Translations
  - Scaling
  - Rotations
  - …

# Visual Comparisons – One Shape Training




ARO-Net can reconstruct the best **details** from sparse point cloud

# Visual Comparisons – One Shape Training




ARO-Net can reconstruct the best **details** from sparse point cloud

# Ablation Studies

| Setting | LFD↓ | HD↓ | CD↓ | EMD↓ | IOU↑ |
| --- | --- | --- | --- | --- | --- |
| $m = 24$ | 1.48 | 2.33 | 5.94 | 1.28 | 8.66 |
| $m = 96$ | 1.30 | 2.16 | 5.31 | 1.03 | 8.92 |
| MLP | 1.44 | 2.30 | 5.90 | 1.14 | 8.75 |
| Default | 1.35 | 2.25 | 5.46 | 1.12 | 8.79 |

**Replacing Transformers with MLP**: the superior results by ARO-Net are predominantly owing to ARO rather than the decoder architecture
$\boldsymbol{m = 24, 96}$: a trade-off between reconstruction performance and computational cost

# Limitations

- ARO-Net runs slow because
  - Performing ***finding top-$k$ points in a cone*** operation for $\boldsymbol{m}$ times
  - Network parameters/computation cost grow with $\boldsymbol{m}$ linearly
- ARO-Net's performance depends on
  - The number of the anchors
  - Placement of the anchors

Thanks!