Data Compression Conference (DCC) 2022

High-fidelity 3D Model Compression based on Key Spheres

*Yuanzhan Li, Yuqi Liu, Yujie Lu, Siyu Zhang, Shen Cai**, *and Yanting Zhang*

Visual and Geometric Perception Lab, Donghua University, Shanghai, CHINA

*Corresponding author: hammer_cai@163.com

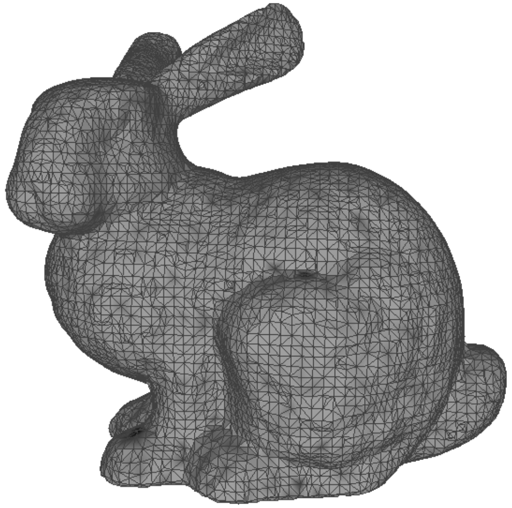
Idea in One Sentence

Using rough shape information to make shape fitting easier

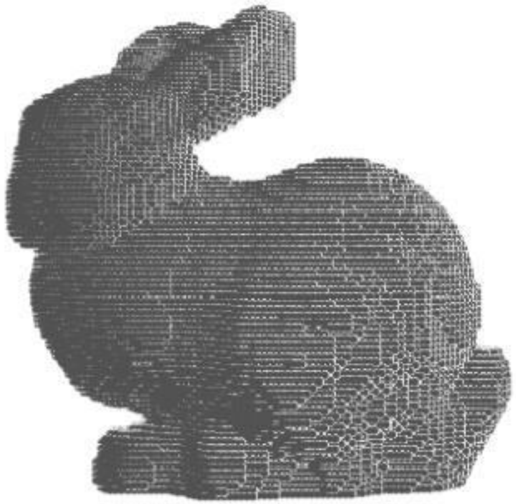
Background

Traditional 3D Object Representations

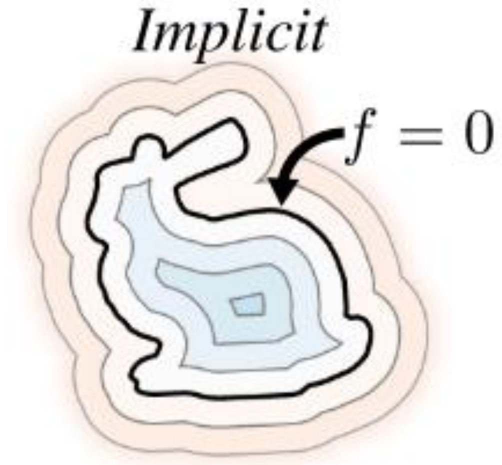
Point Cloud
(1024)



Mesh
(V:8171 F:16301)



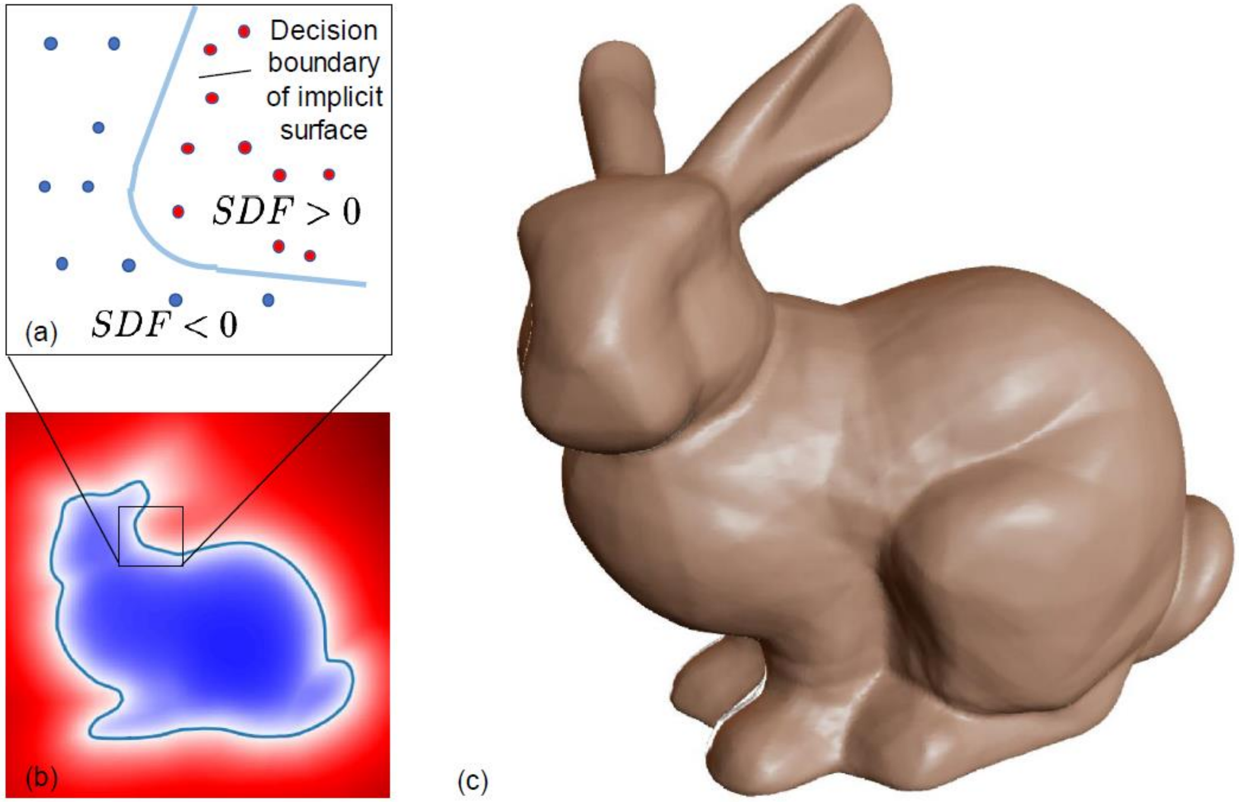
Voxel
(128^3)



Implicit SDF
(Continuous)

[1] Turk, Mark. The Stanford Bunny. www.cc.gatech.edu

Signed Distance Field (SDF)



[2] J. Park, et al, “DeepSDF: Learning continuous signed distance functions for shape representation,” in CVPR, 2019.

Traditional 3D Reconstructions

- PC of each frame (Laser) $\xrightarrow{\text{ICP}}$ stitched PCs [3]
- Images (Camera) $\xrightarrow{\text{triangulation}}$ sparse PC $\xrightarrow{\text{epipolar matching}}$ dense PC $\xrightarrow{\text{Delauney triangulation}}$ dense mesh [4]
- Depth images (Kinect) $\xrightarrow{\text{ICP\&fusion}}$ TSDF volume $\xrightarrow{\text{Marching Cube}}$ mesh [5]

[3] P. J. Besl and N. D. McKay, “A method for registration of 3-D shapes,” in TPAMI, 14(2), 239-256, 1992.

[4] Schonberger, J. L. , and J. M. Frahm, "Structure-from-Motion Revisited," in CVPR, 2016.

[5] R. A. Newcombe, et al., "KinectFusion: Real-time dense surface mapping and tracking," in ISMAR, 2011.

3D Reconstructions from Image(s) Utilizing Neural Networks

- Single view $\xrightarrow{\text{CNN}}$ mesh [6] / voxel or volume [7]
- Stereo views $\xrightarrow{\text{CNN}}$ matched image points $\xrightarrow{\text{triangulation}}$ PC (depth map) [8]
- Multiple views $\xrightarrow{\text{CNN}}$ cost volume $\xrightarrow{\text{prediction}}$ PC (depth map) [9]
- Multiple views $\xrightarrow{\text{MLP fitting}}$ density volume $\xrightarrow{\text{Marching Cube}}$ mesh [10][11] (NeRFs)

3D Reconstructions from Image(s) Utilizing Neural Networks

[6] N. Wang, et al, "Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images," in ECCV, 2018.

[7] L. Mescheder, et al, “Occupancy Networks: Learning 3D Reconstruction in Function Space,” in CVPR, 2019.

[8] R. Chabra, et al, “StereoDRNet: Dilated Residual StereoNet, ” in CVPR, 2019.

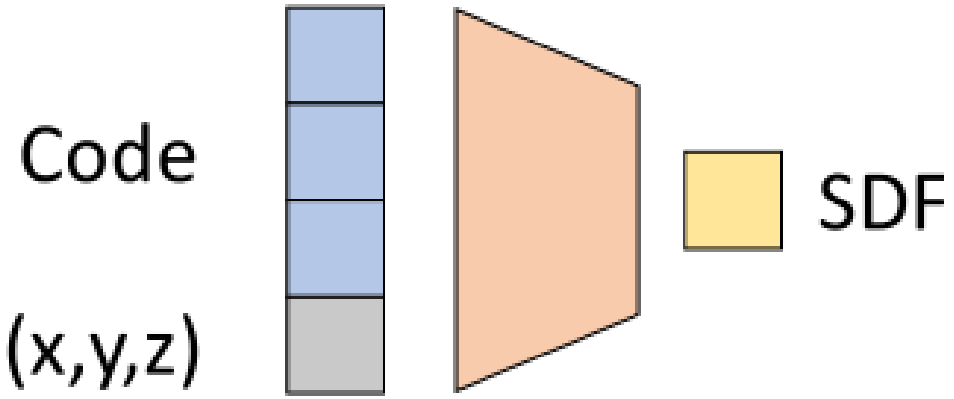
[9] Y. Yao, et al, “MVSNet: Depth Inference for Unstructured Multi-view Stereo,” in ECCV, 2018.

[10] B. Mildenhall, et al, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” in ECCV, 2020.

[11] A. Yu, et al, “Plenoxels: Radiance Fields without Neural Networks,” in ArXiv, 2021.

Neural Implicit Reconstructions from 3D Models ------ DeepSDF [2]

- SDFs of one category models →→→ SDF of a trained or new object
- The first work to introduce the auto-decoder method in 3D learning
- One code (latent) vector for each object

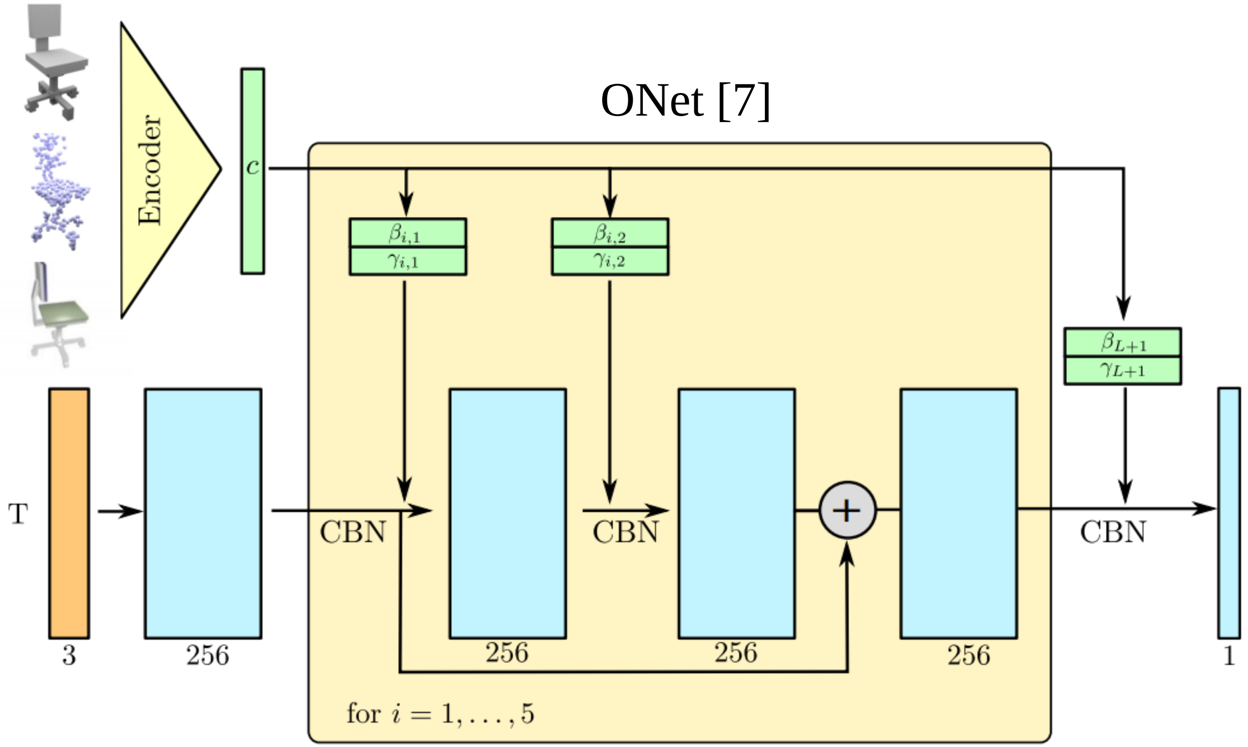


8*512 MLP

[2] J. Park, et al, "DeepSDF: Learning continuous signed distance functions for shape representation," in CVPR, 2019.
[12] V. Sitzmann, et al, " Implicit Neural Representations with Periodic Activation Functions," in NIPS, 2020.
[13] M. Tancik, et al, "Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains," in NIPS, 2020.

Neural Implicit Reconstructions from 3D Models ------ ONet [7] / IM-Net [14]

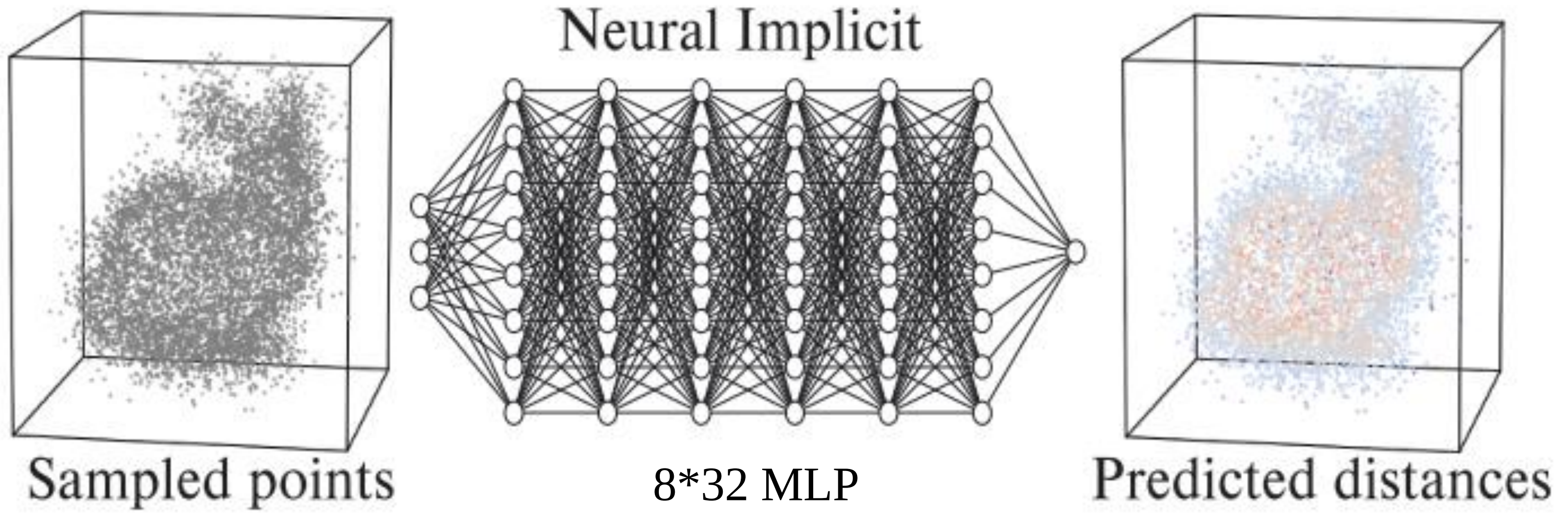
- Voxels of one category objects →→→ occupancy of one trained object
- The encoder-decoder architecture without latent vectors
- Represent the 3D surface as the continuous decision boundary of a deep neural network classifier



[7] L. Mescheder, et al, "Occupancy Networks: Learning 3D Reconstruction in Function Space," in CVPR, 2019.
[14] Z. Chen and H. Zhang, "Learning Implicit Fields for Generative Shape Modeling," in CVPR, 2019.

Neural Implicit Reconstructions from 3D Model ------ NI [15]

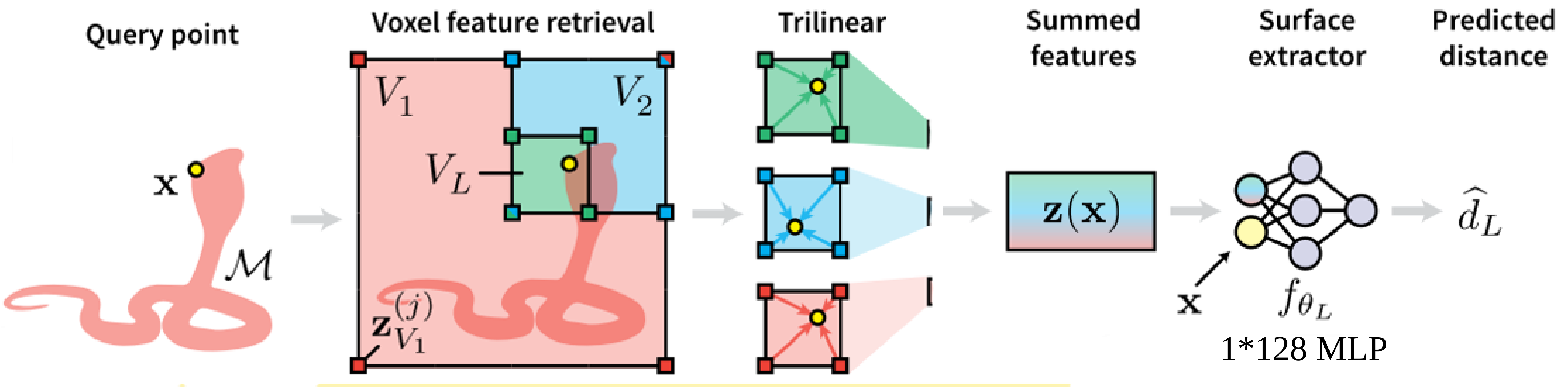
- SDF of one model →→→ SDF of this model
- The global overfitting scheme for each model (traditional compression way)
- Very small network (with 7553 paras.) and effective for simple objects



[15] T. Davies, et al, “Overfit Neural Networks as a Compact Shape Representatio”(2020.9), “On the Effectiveness of Weight-Encoded Neural Implicit 3D Shapes”(2021.1), in arXiv.

Neural Implicit Reconstructions from 3D Model ------ NGLOD [16]

- Follow the basic process of NI, but using local fitting instead of global fitting
- Learn latent vectors for vertices of octree and interpolate them to get point feature
- One-layer network and a lot of latent vectors (depending on different level in LOD)
- More effective for objects with moderate complexity, compared to NI

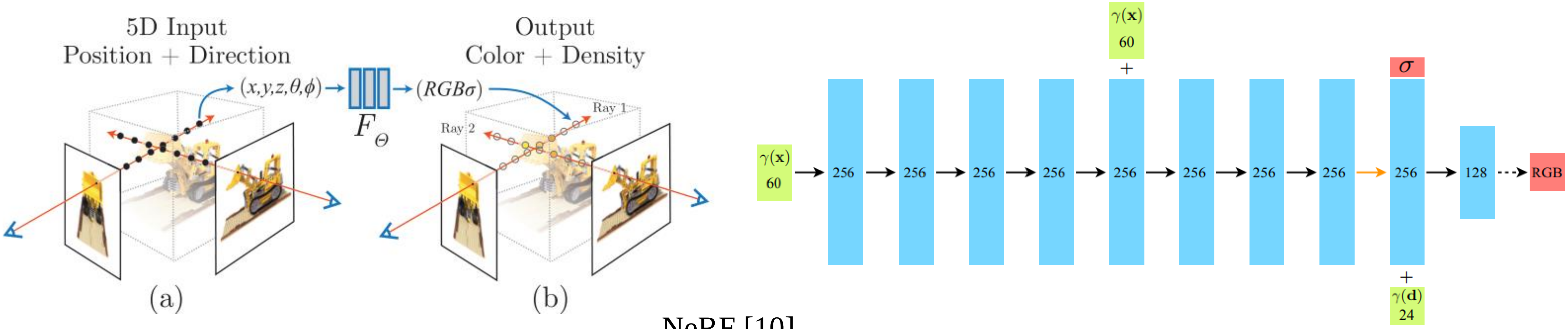


1*128 MLP

[16] T. Takikawa, et al, “Neural geometric level of detail: Real-time rendering with implicit 3d shapes,” in CVPR, 2021.

Other Overfitting Reconstructions for Image or Scene

- Image →→→ Image [17]
- Images →→→ Scene [10] [11] (NeRFs)
- Similarity: MLP networks, global/local fitting, and w./w.o. latent vectors



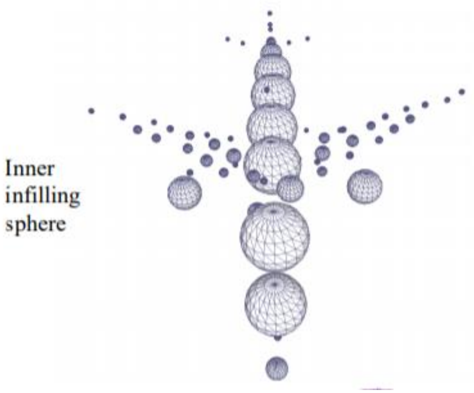
NeRF [10]

[17] Y. Strumpler, et al, "Implicit Neural Representations for Image Compression," in ArXiv, 2021.

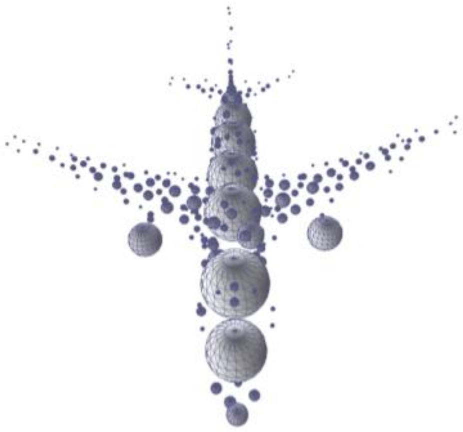
Motivation

Spheres Representation of 3D Models

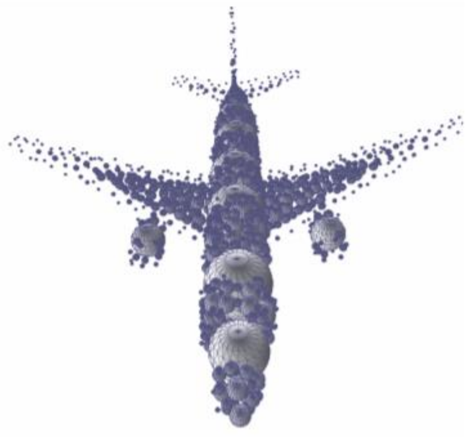
- Discrete but concise
- Extract spheres from SDF
- More effective than PC and voxel for 3D object classification



64 spheres



256 spheres



1024 spheres

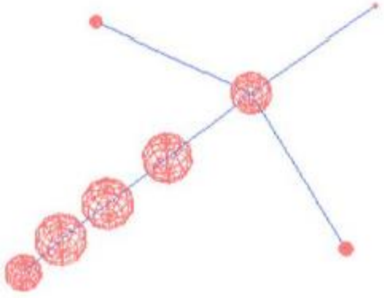
Table 1: Overall classification accuracy on ModelNet40

Method	Input	Acc(%)
MVCNN 12x	images	89.5
3D Shapenets	voxels	84.7
VoxNet	voxels	85.9
PointNet	points	89.2
PointNet++	points without normal	90.7
PointNet++	points with normal	91.9
Ours (1024 interior)	infilling spheres	90.2
Ours (1024 exterior)	infilling spheres	90.6
Ours (512 interior and 512 exterior)	infilling spheres	90.3
Ours (1024 interior on PointNet++)	infilling spheres	92.1

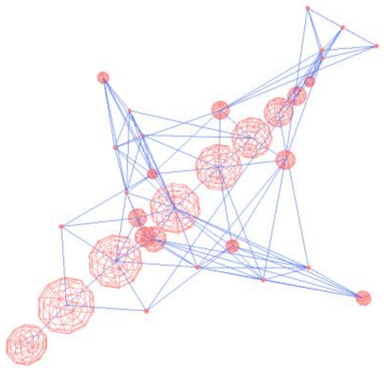
[18] Hui Cao, Haikuan Du, Siyu Zhang, Shen Cai*. “Inspherenet: a concise representation and classification method for 3d object,” in MultiMedia Modeling (MMM), 2020.

Sphere-Node Graph of 3D Models

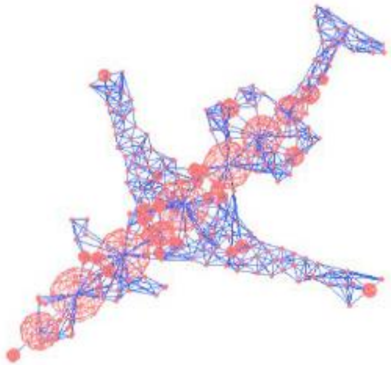
- Revise the method of extracting spheres
- Connect sphere nodes to form a graph (suitable for GNNs)
- More effective object classification under low resolution



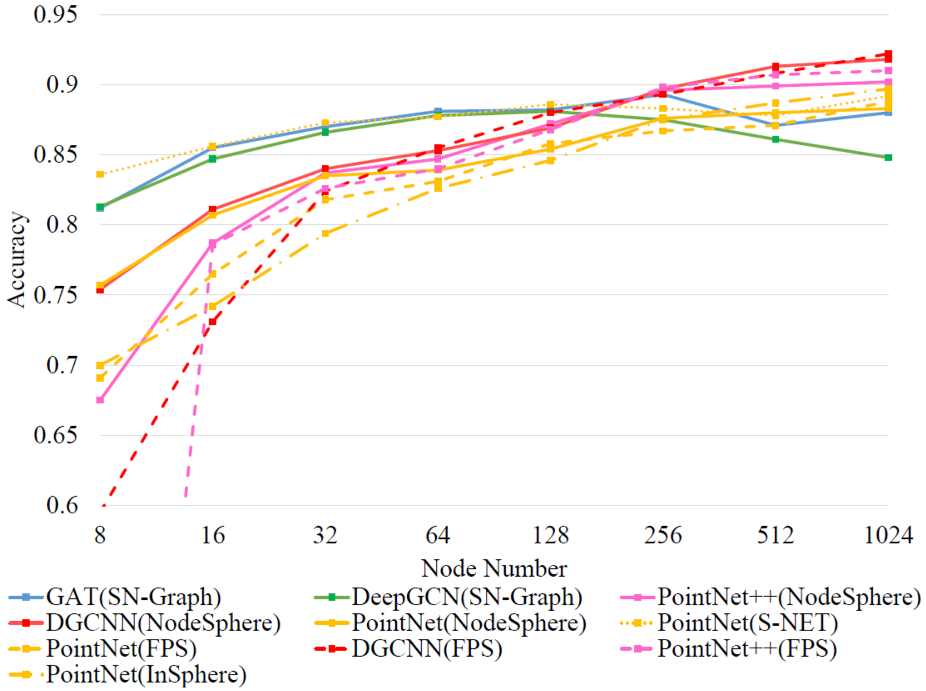
8 nodes



32 nodes



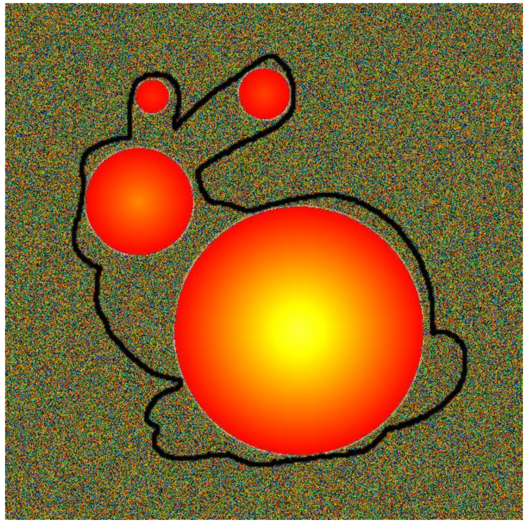
256 nodes



[19] Siyu Zhang, Hui Cao, Yuqi Liu, Shen Cai*, Yanting Zhang, Yuanzhan Li, and Xiaoyu Chi, "Sn-graph: A minimalist 3d object representation for classification," in ICME, 2021. (oral presentation)

Birth of This Idea

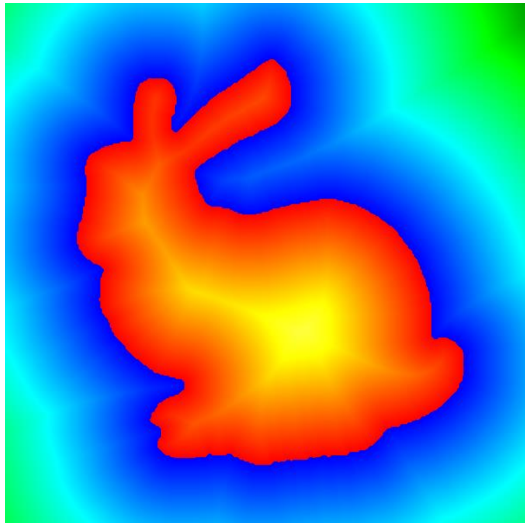
- Interior spheres or SN-Graph is visually attracted, so objects can be easily recognized.
- Except classification, what else can the spheres representation do? (2020.06)
- If rough shape information is obtained, SDF fitting will be easier. (2020.11)



Key Spheres SDF

Spheres indicate upper and lower bounds of local SDF for most regions of object.

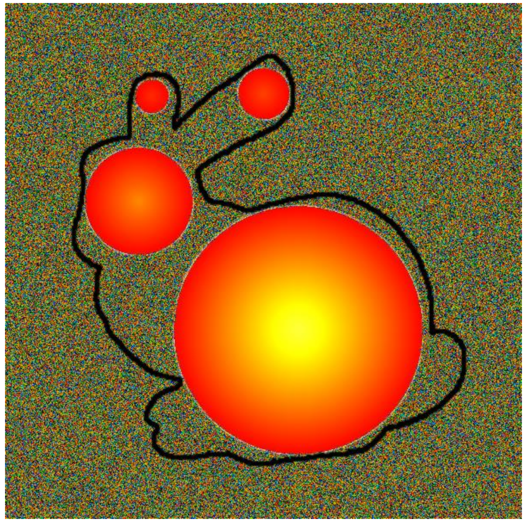
Overfitting



Ground Truth SDF

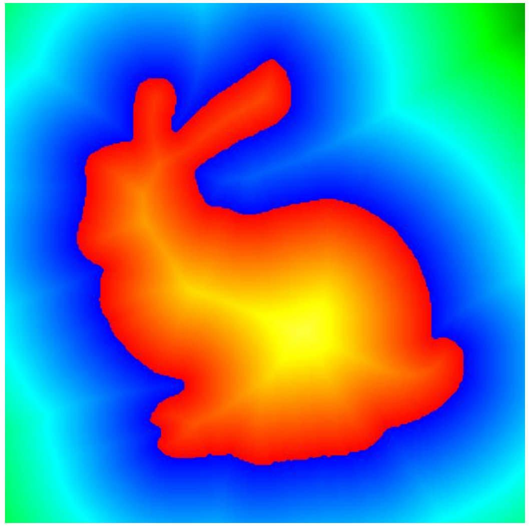
Detailed Problems

- How to embed the spheres information into neural network? We try two simple ways
- Global fitting or local fitting? Global fitting is enough for complex objects
- How much improvement in reconstruction accuracy can this fitting lead to? Significant



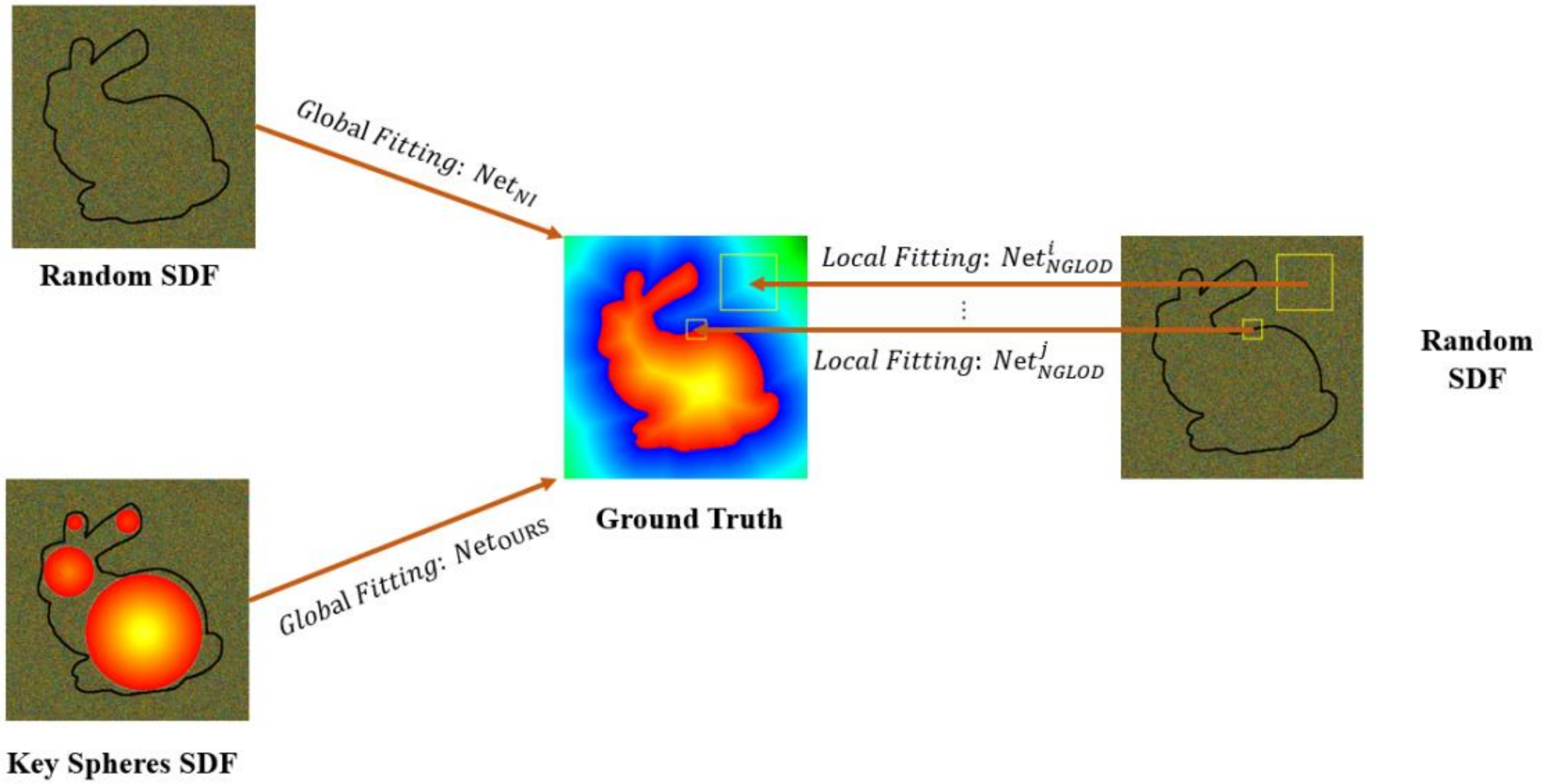
Key Spheres SDF

Overfitting



Ground Truth SDF

Fitting Comparison with NI and NGLOD



Key Spheres based 3D Model Compression

Basic Process

GT model

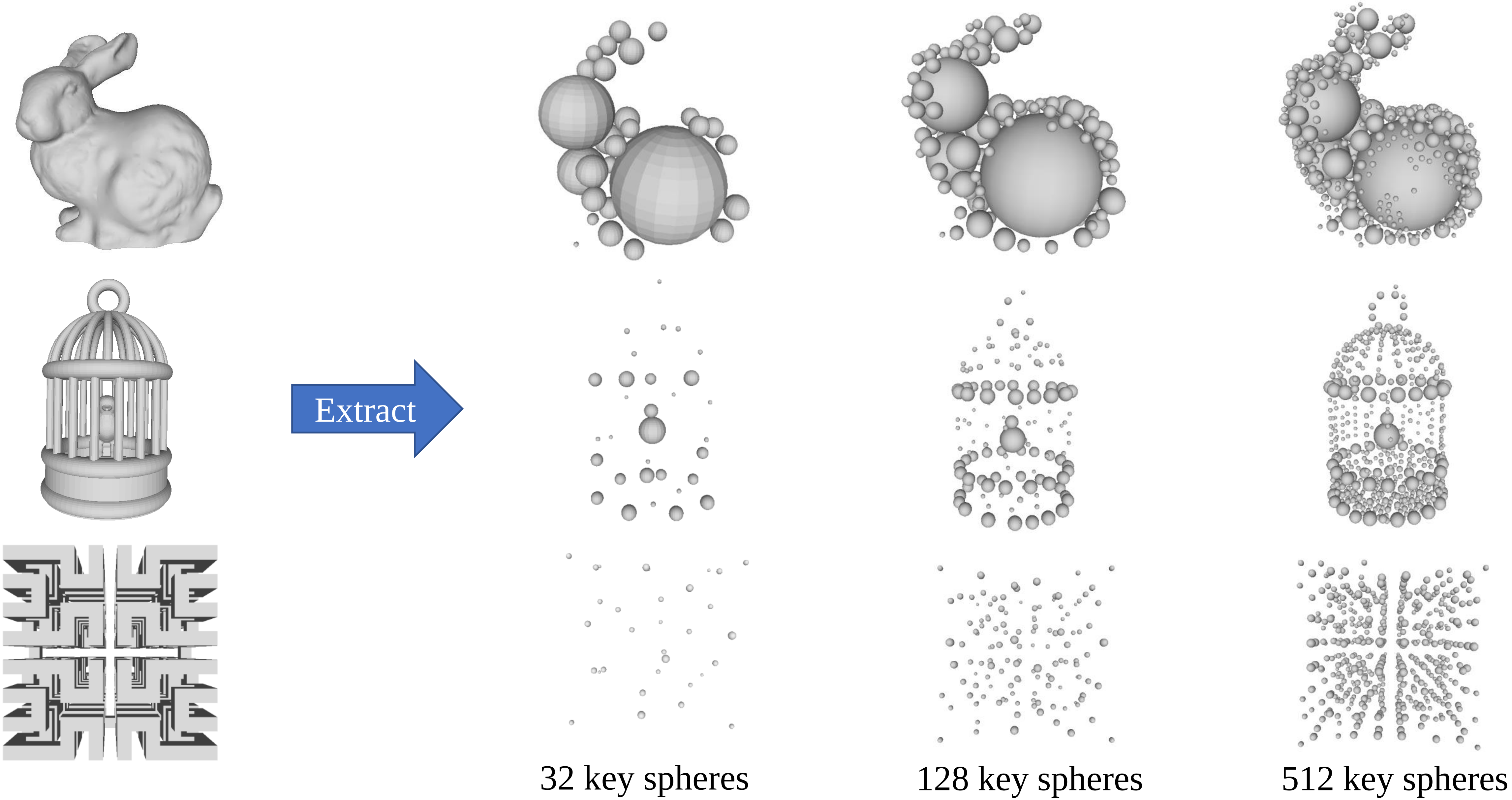
Neural Network

$\mathrm{SDF}(x)$

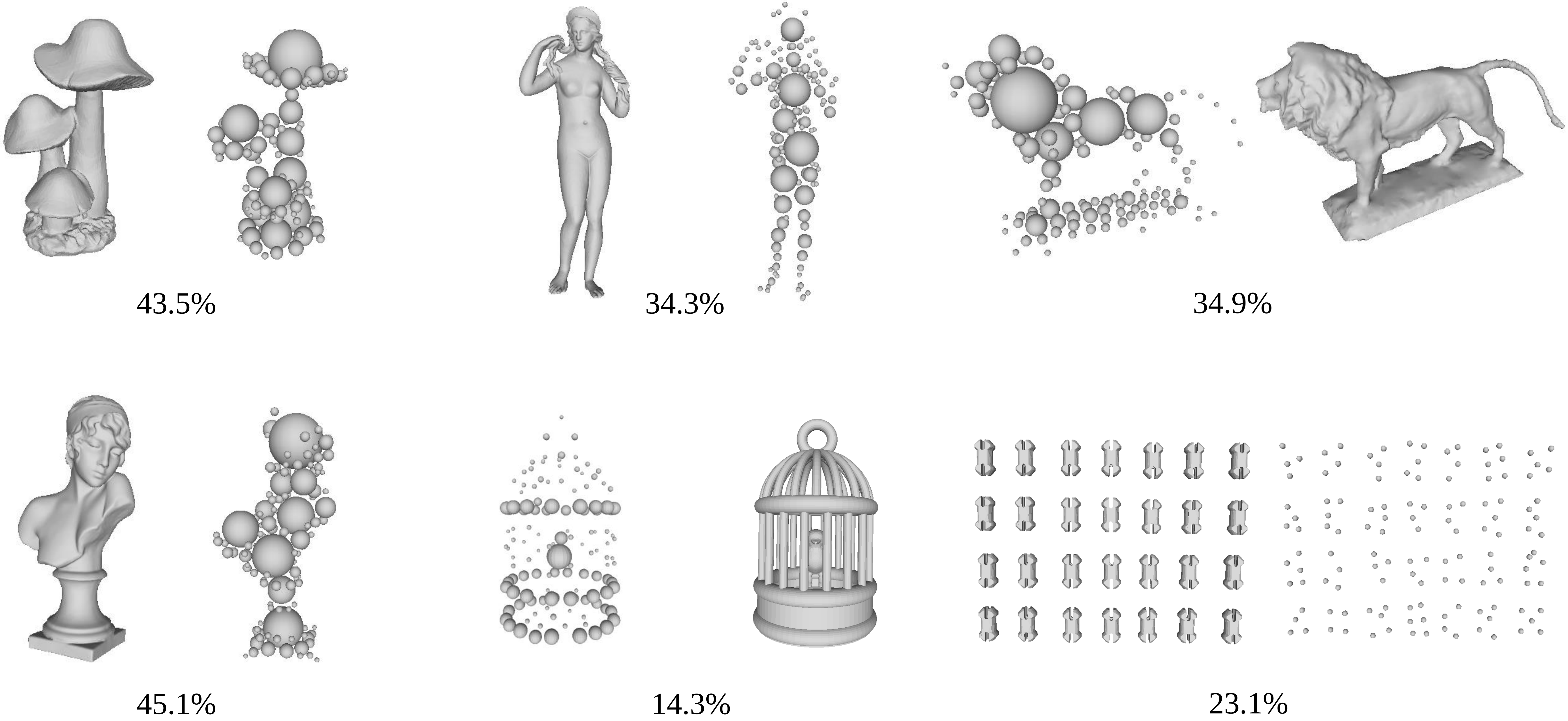
Marching Cube

reconstructed model

Key Spheres Extraction ------ Different Resolution



Key Spheres Extraction ------ Different Models

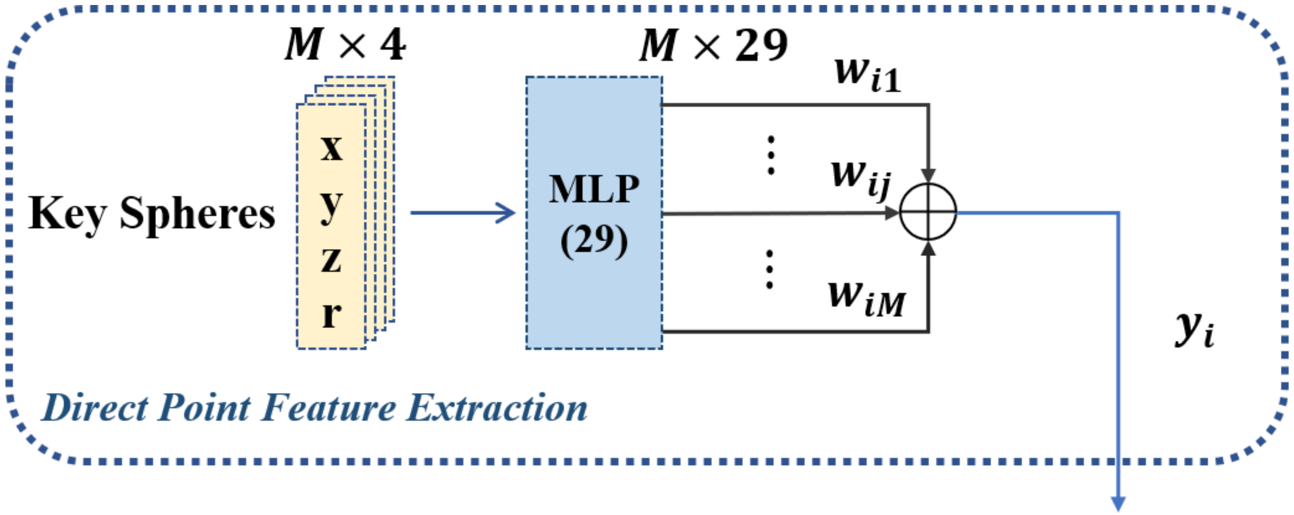


128 Key Spheres

Key Spheres Feature Extraction and Aggregation

- One layer 4*29 MLP promoting sphere feature to 29 dimensions
- Linear weighted aggregation

$$\mathbf{y}_i = \sum_{j=1}^{M} (w_{ij} * \mathbf{z}_j), \quad \text{with } w_{ij} = d_{ij} \Big/ \sum_{j=1}^{M} d_{ij}, \quad \text{where } d_{ij} = 2 * rad(\mathcal{S}_j) + dis(\mathbf{x}_i, \mathbf{c}_j).$$

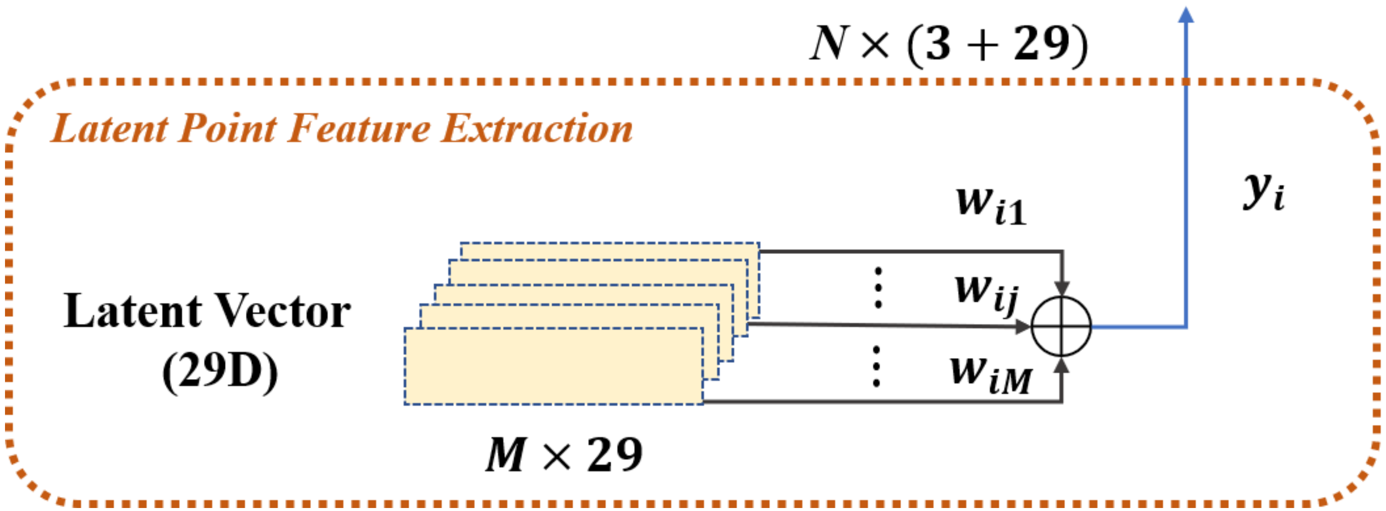


DPFE branch

Key Spheres Feature Extraction and Aggregation

- 29 dimensions latent vector representing each sphere feature
- The same linear weighted aggregation

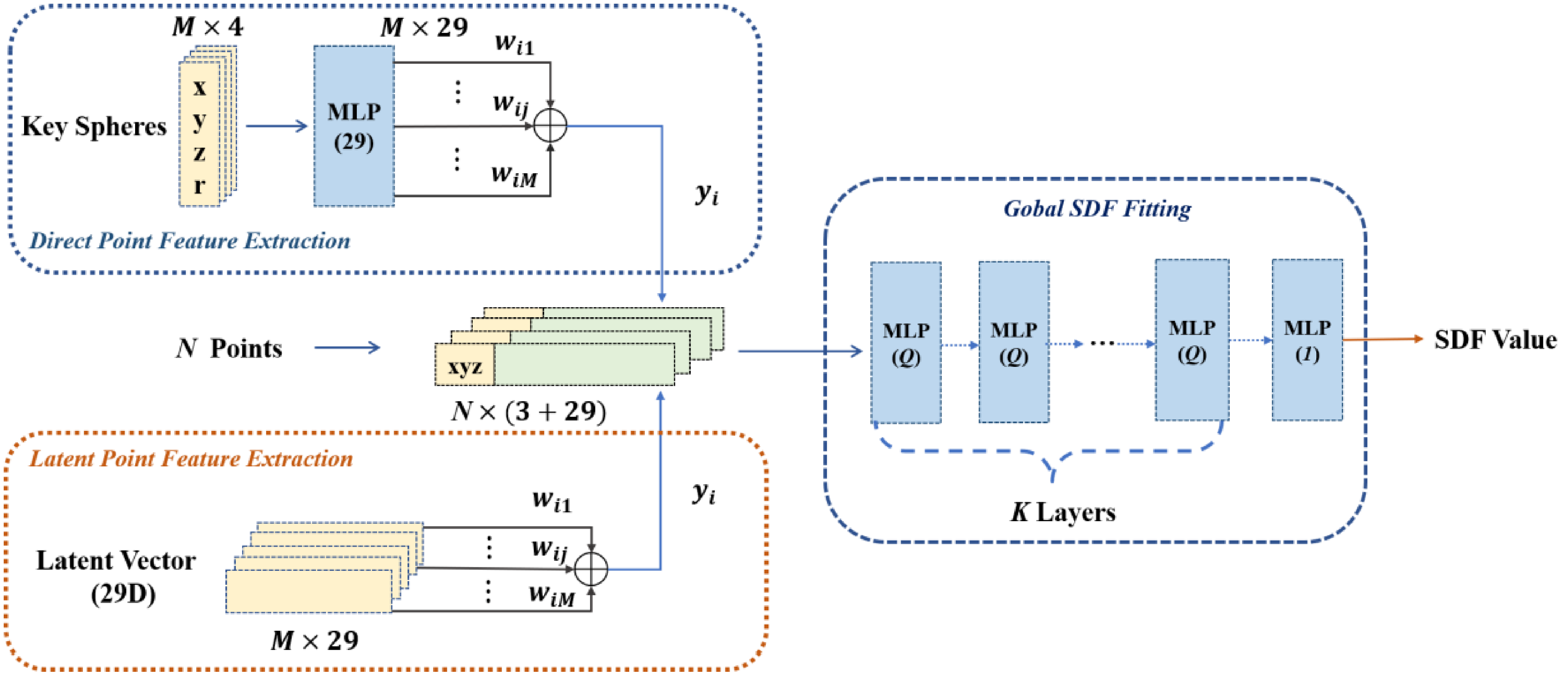
$$\mathbf{y}_i = \sum_{j=1}^{M} (w_{ij} * \mathbf{z}_j), \quad \text{with } w_{ij} = d_{ij} \,/ \sum_{j=1}^{M} d_{ij}, \quad \text{where } d_{ij} = 2 * rad(\mathcal{S}_j) + dis(\mathbf{x}_i, \mathbf{c}_j).$$



LPFE branch

Whole Network Architecture

- Concat of point coordinate and aggregation feature from key spheres
- MLP used for global SDF fitting



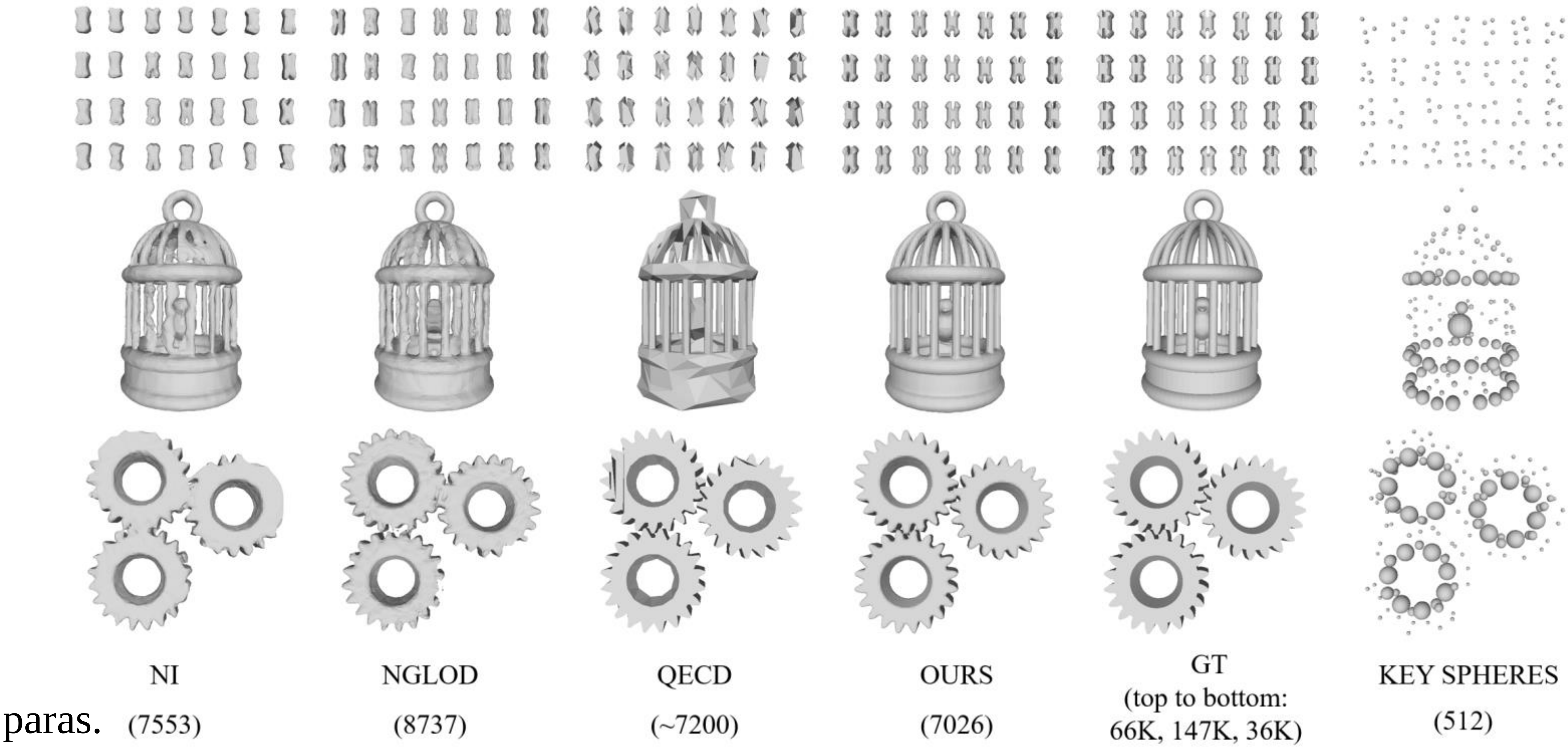
Experiments

Datasets & Metrics

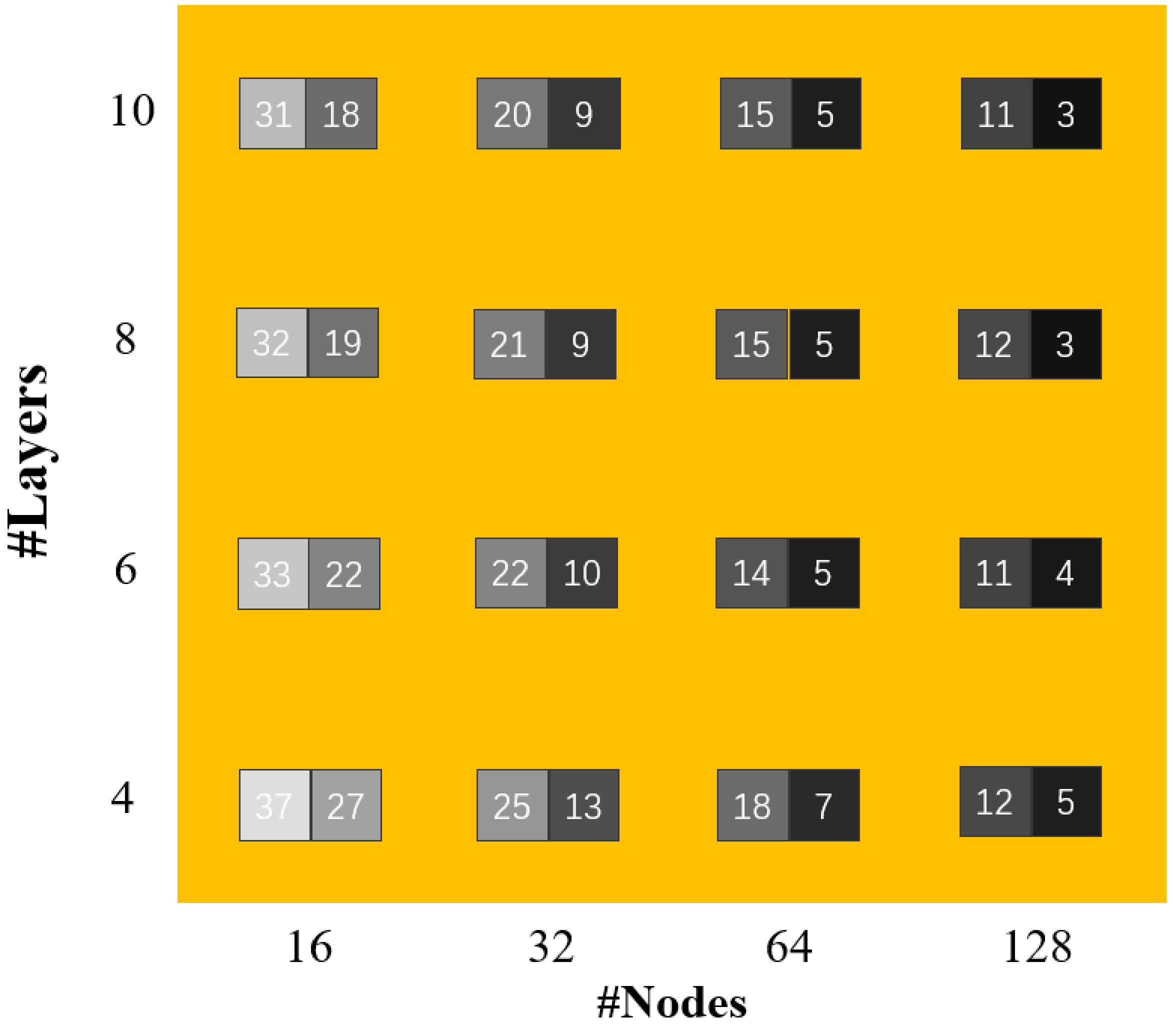
- Thingi32 / Thing10K / ShapeNet150 (planes, chairs, and cars)

- Surface error / importance error / Chamfer distance (CD) / intersection over union (IOU)

Visual Comparison with Three Other Methods



Numerical Comparison with NI

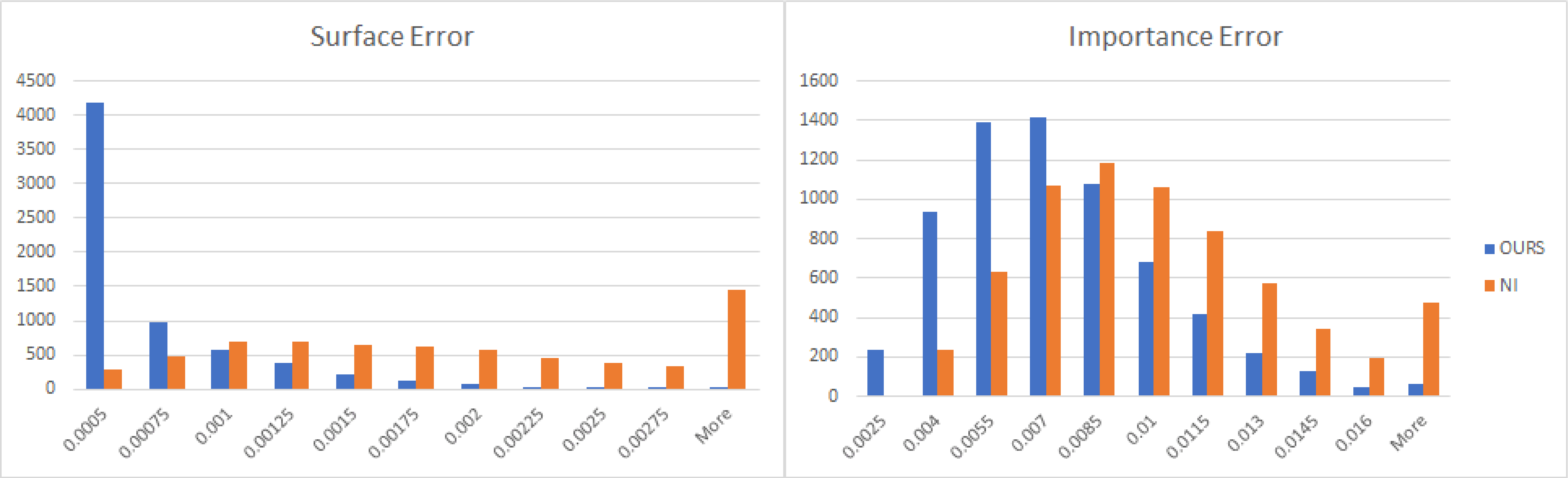


Surface error (times 10,000) on Thingi32

Former: NI

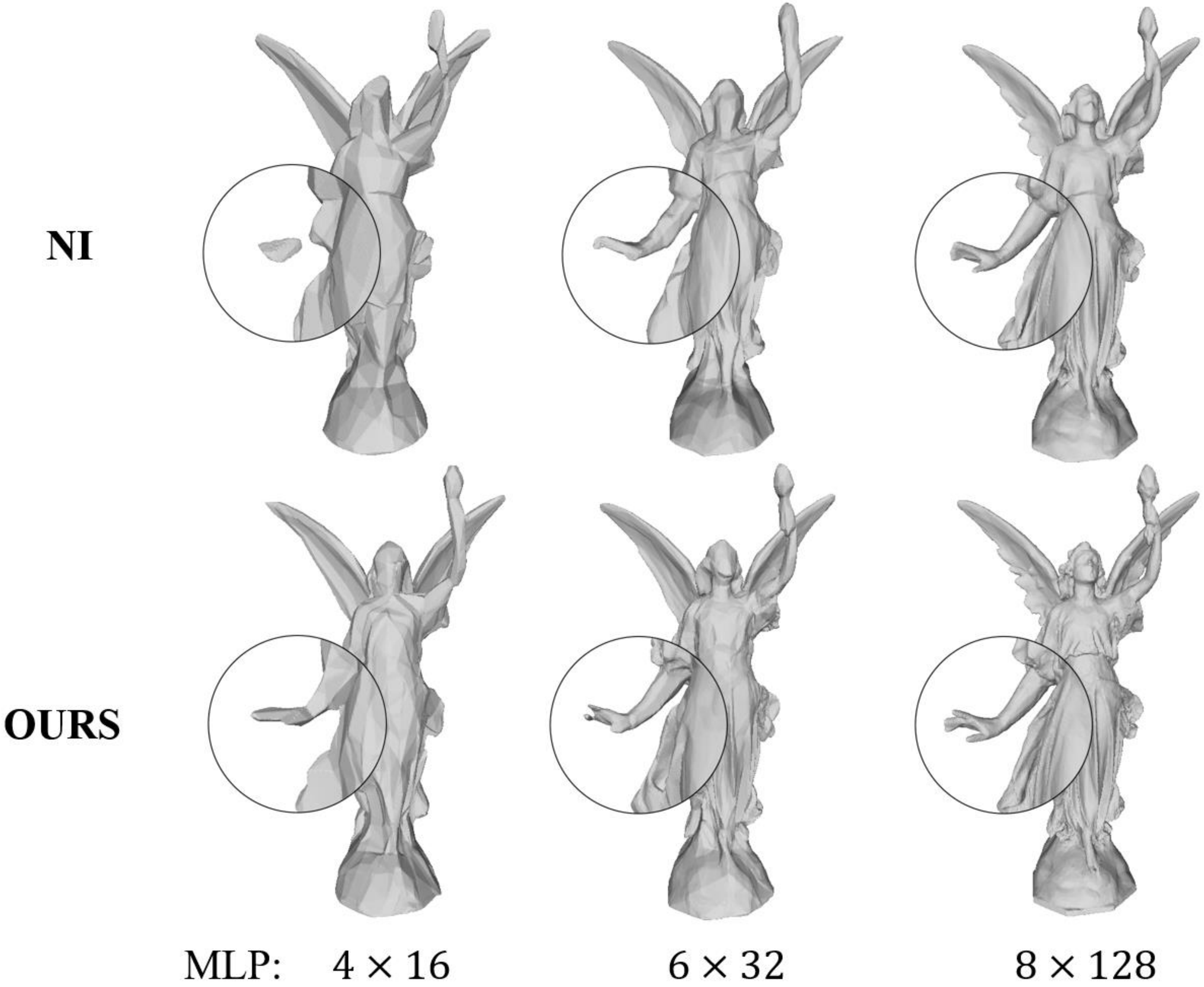
Latter: Ours w. DPFE branch & 128 spheres

Numerical Comparison with NI



Histograms of surface and importance errors of on Thingi10K dataset

Visual Comparison with NI

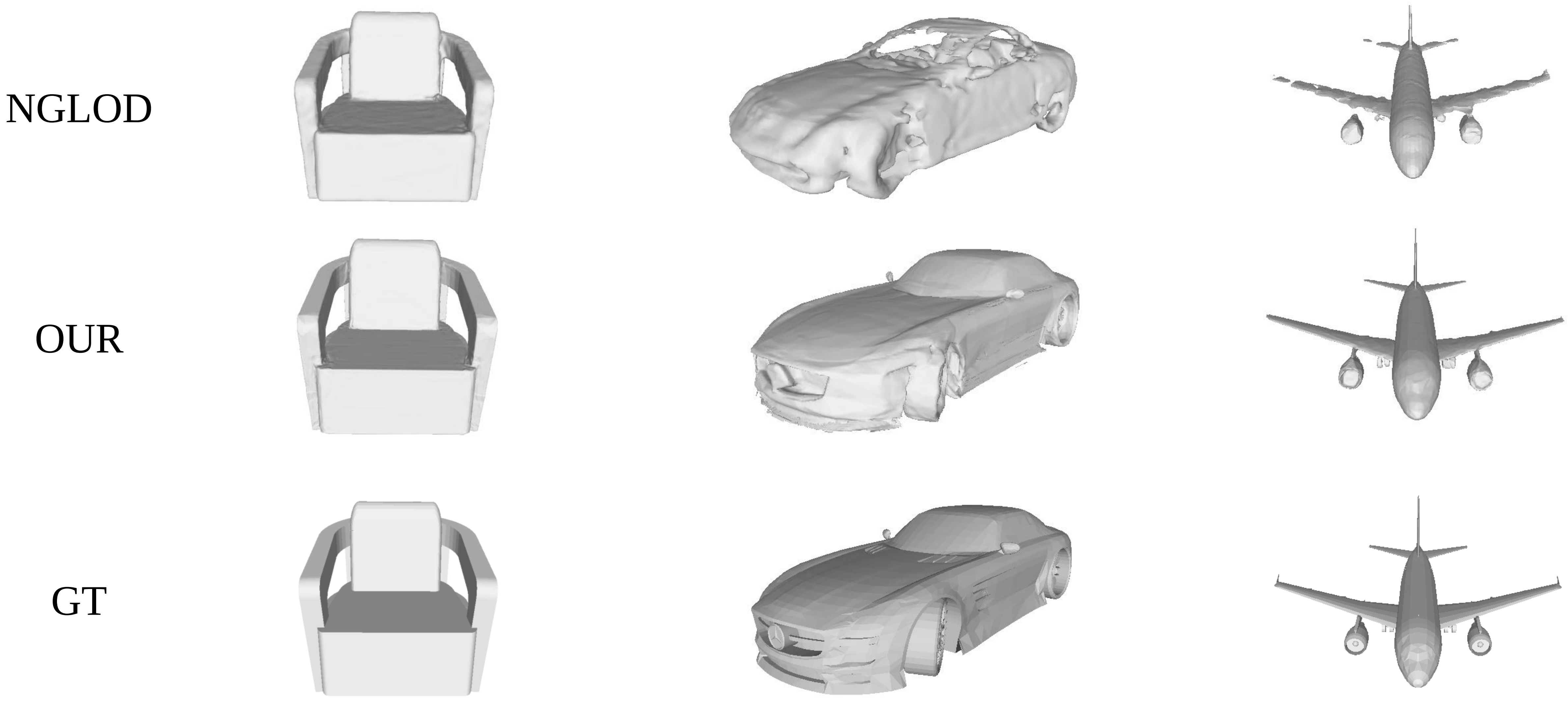


Numerical Comparison with NGLOD and Other Methods

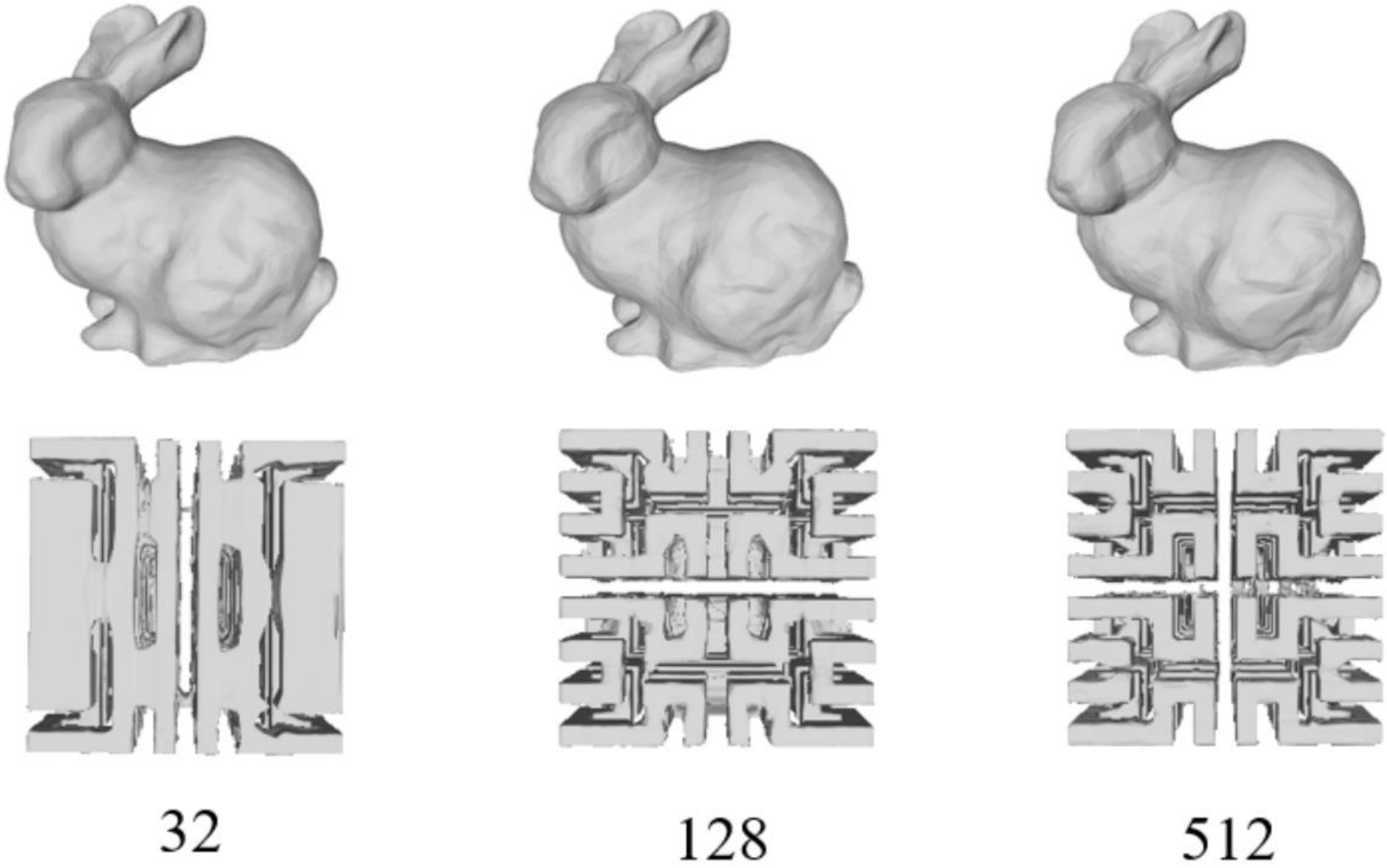
Method	Storage (KB)	ShapeNet150		Thingi32	
		gIoU	CD	gIoU	CD
DeepSDF	7186	86.9	0.316	96.8	0.053
FFN	2059	88.5	0.077	97.7	0.033
SIREN	1033	78.4	0.381	95.1	0.077
NI	30	82.2	0.5	96	0.092
NGLOD: (LOD1)	96	84.6	0.343	96.8	0.079
(LOD2)	111	88.3	0.198	98.2	0.041
(LOD3)	163	90.4	0.112	99	0.030
(LOD5)	1356	91.7	0.062	**99.4**	0.027
OURS: (32)	30	93.8	0.188	98.6	0.028
(128)	42	93.8	0.138	98.6	0.029
(512)	93	94.1	0.124	98.5	0.031
(512*)	158	**95.9**	**0.060**	99.2	**0.025**

Ours with
LPFE branch

Visual Comparison with NGLOD (LOD-1)

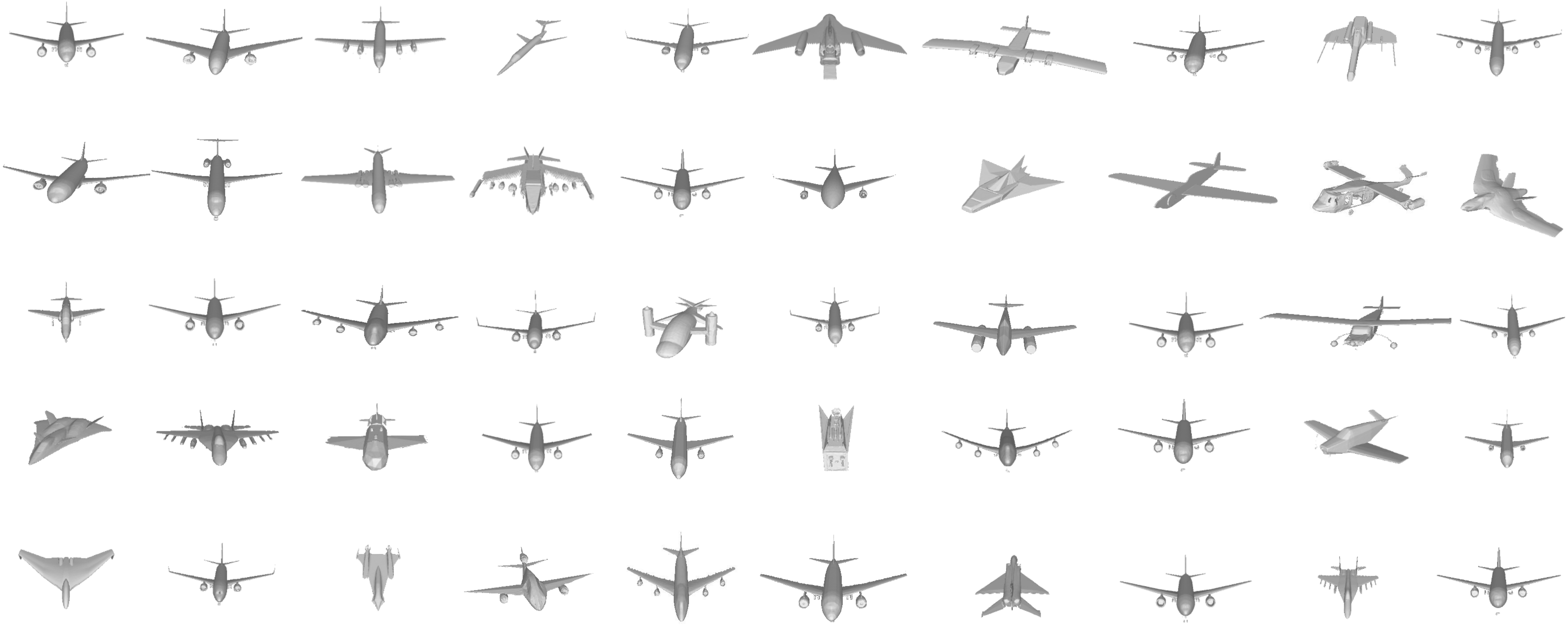


Visual Comparison at Various Resolutions

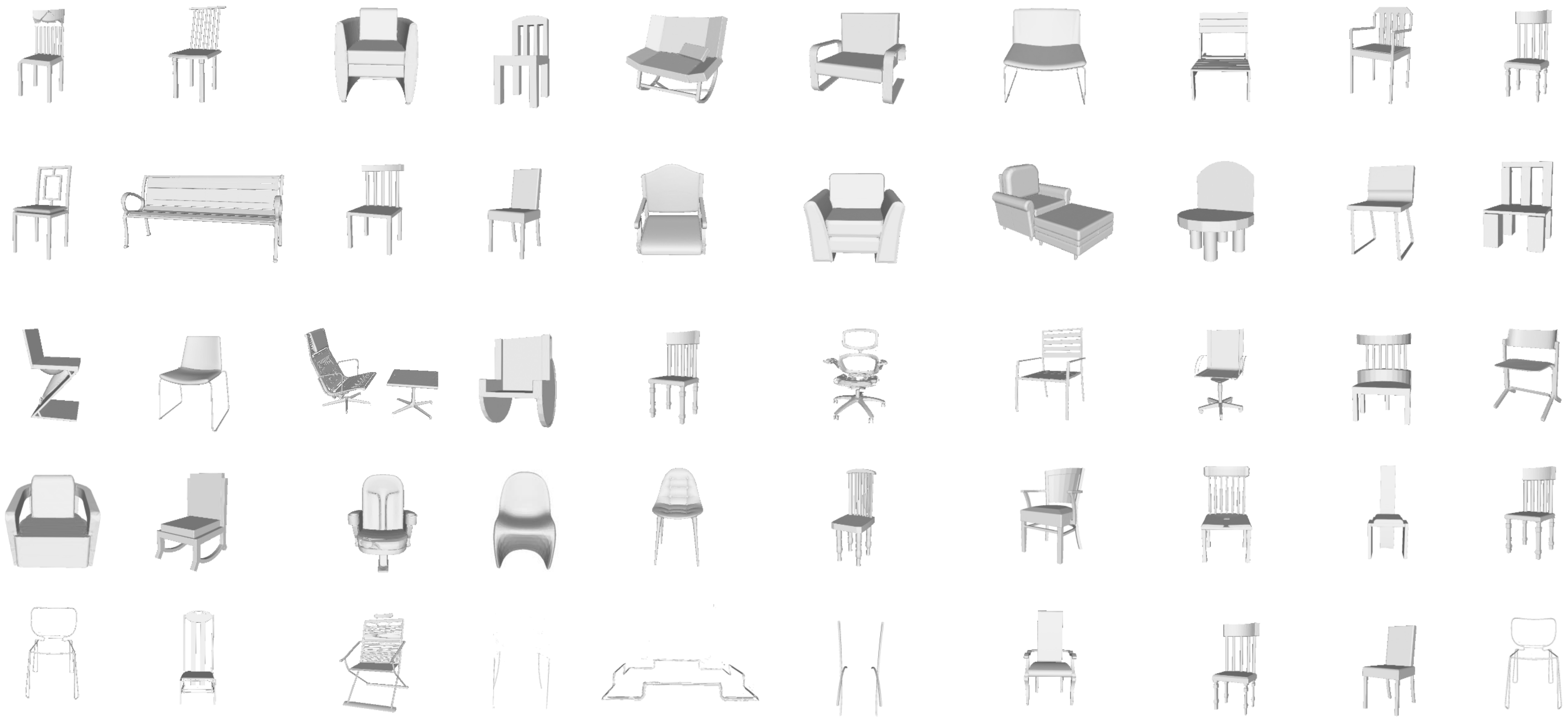


Visualization of Reconstructed Thingi32

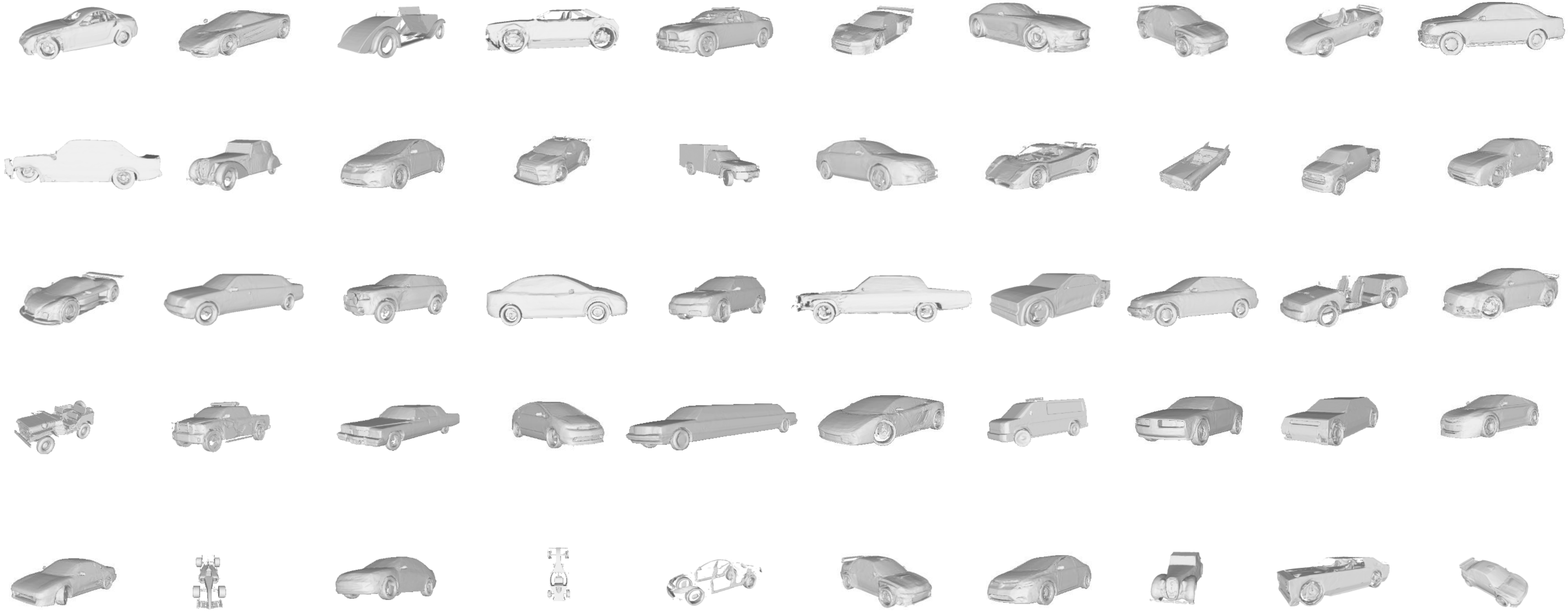
Visualization of Reconstructed Planes in ShapeNet150



Visualization of Reconstructed Chairs in ShapeNet150



Visualization of Reconstructed Cars in ShapeNet150



Compression Rate under High-fidelity Reconstruction

Method	Thingi32 (mean size 221K)	ShapeNet150-Planes (mean size 181K)
NI	29.2:1	23.9:1
NGLOD (LOD-1)	25.2:1	20.7:1
Ours w. DPFE & 32 spheres	**33.2:1 (finer detail)**	27.2:1
Ours w. DPFE & 128 spheres	31.4:1	**25.7:1 (more robust)**
Ours w. LPFE & 32 spheres	29.1:1	23.9:1
Ours w. LPFE & 128 spheres	20.5:1	16.8:1

Conclusion & Future Work

Conclusion

- A hybrid model reconstruction method w. explicit key spheres & implicit MLP

- Prior spheres + MLP global fitting = fine reconstruction quality & high compression rate

- Big network is not proper for some tasks, especially on neural compression. (personal opinion)

Future Work

- Improve the feature aggregation method (e.g., transformer)
- Encode edge information of the rough shape
- Two stages training for very complex objects
- Revise the fitting target to others, rather than SDF
- Adaptive networks for different models

Data Compression Conference (DCC) 2022

graduate students:

Yuanzhan Li

Yuqi Liu

Yujie Lu

Siyu Zhang

advisors:

Shen Cai

Yanting Zhang

Visual and Geometric Perception Lab, Donghua University, Website: www.cscvlab.com

Source code: https://github.com/cscvlab/3D-Objects-Compression