ROAD: Learning an Implicit Recursive Octree Auto-Decoder to Efficiently Encode 3D Shapes

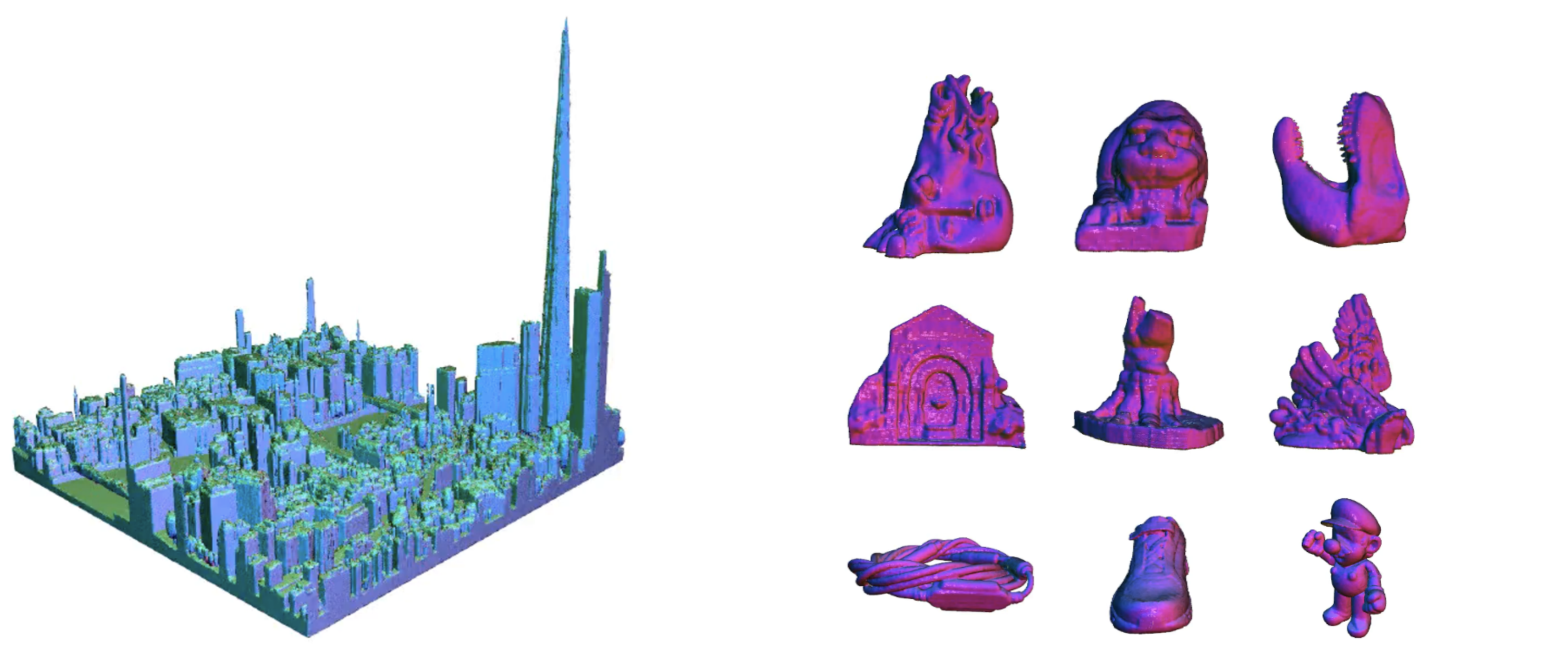
TOYOTA RESEARCH INSTITUTE

Sergey Zakharov, Rareș Ambruș, Katherine Liu, Adrien Gaidon

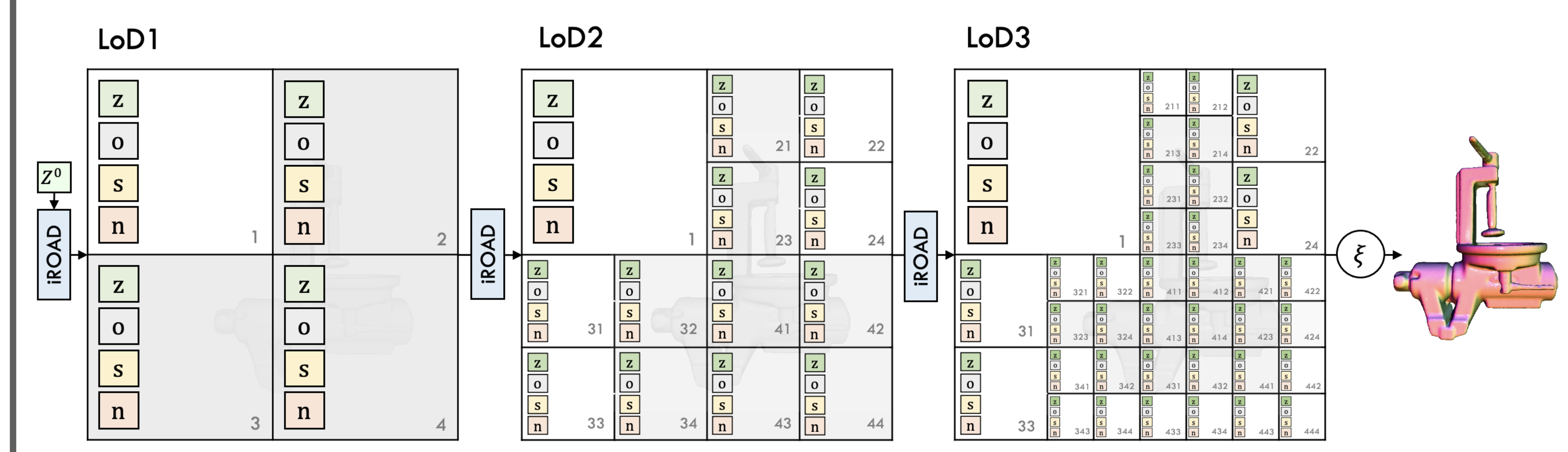
CoRL 2022

Motivation

- Represent an arbitrary number of objects compactly
- Implicit representations of 3D objects scale poorly
- Learn a **recursive hierarchical latent space** which promotes reusability of parts across shapes



Method

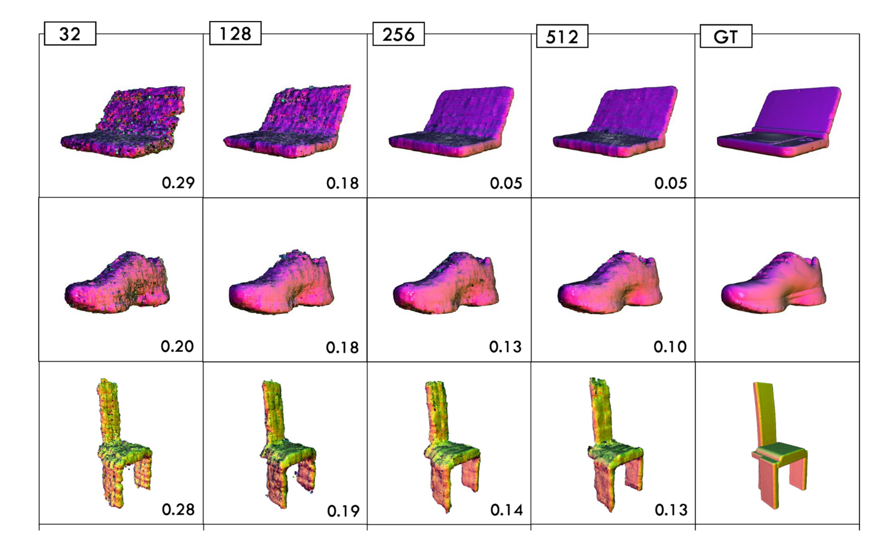


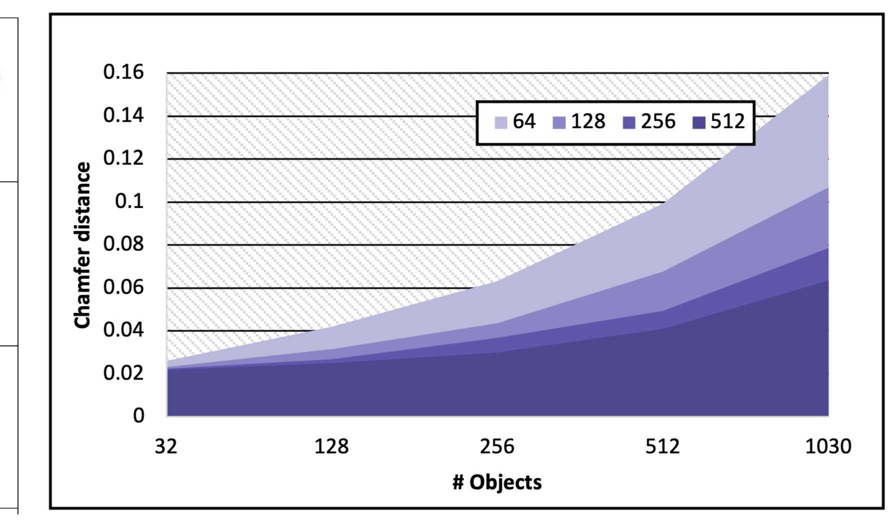
- Our method **extracts object surfaces** by performing an **octree traversal** recursively
- Each cell encodes occupancy, local SDF and normals
- **Large datasets** can be encoded with high surface reconstruction quality and **99% less storage space**

Results

Method	ShapeNet150			Thingi32		
	Storage (MB) (↓)	gIoU (↑)	Chamfer (↓)	Storage (MB) (↓)	gIoU (↑)	Chamfer (↓)
DeepSDF [8]	1052.6	86.9	0.316	224.6	96.8	0.053
FFN [43]	301.6	88.5	0.077	64.3	97.7	0.033
SIREN [58]	151.3	78.4	0.381	32.3	95.1	0.077
Neural Implicits [63]	4.4	82.2	0.500	**0.9**	96.0	0.092
NGLOD [20]	185.4	91.7	0.062	39.6	**99.4**	0.027
Ours / LoD6	**3.8**	86.3	0.175	3.2	96.4	0.138
Ours / LoD7		94.2	0.067		98.4	0.045
Ours / LoD8		94.9	0.041		98.7	0.022
Ours / LoD9		**94.9**	**0.036**		98.7	**0.017**

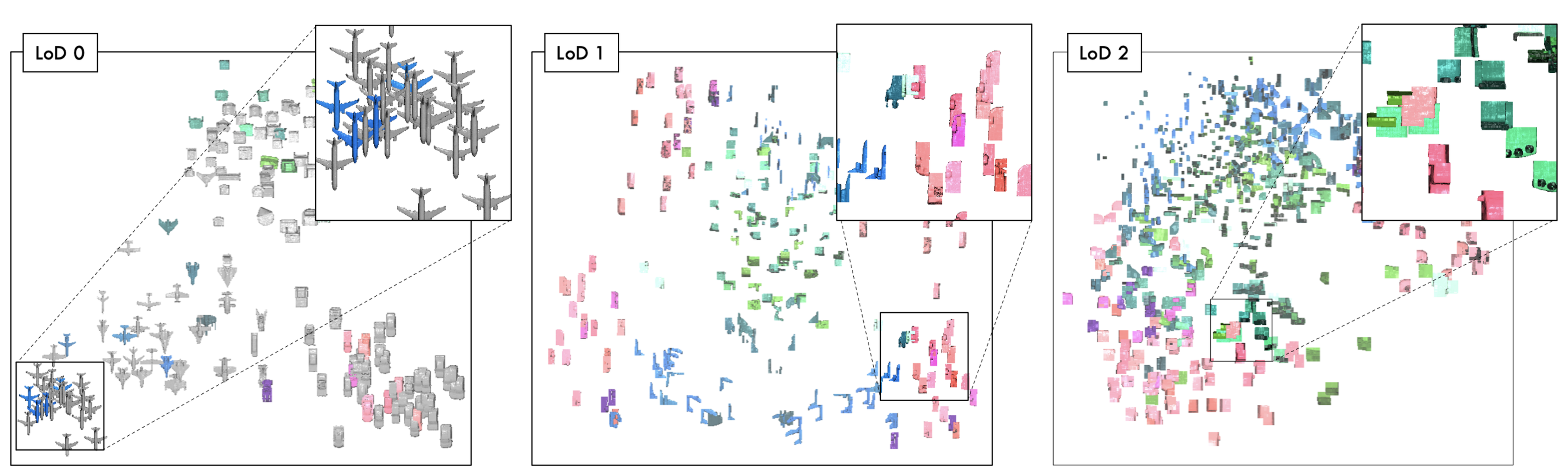
- **Shape Reconstruction:** we train one model for an entire dataset, outperforming even single object baselines



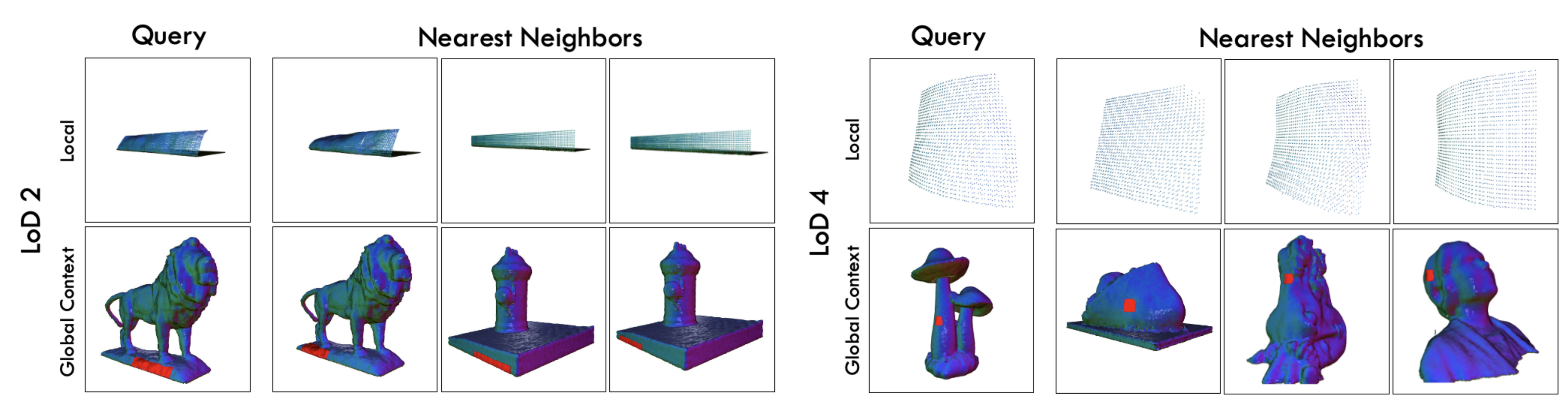


- Generalization and Chamfer distance vs # objects

Latent Space



- LoD 0 - **similar latent vectors encode similar geometries** and show a clear class separation
- Higher LoDs - the projected latent space is increasingly **shared** by different classes



- **Nearest neighbors** in latent space encode similar shapes at the same LoD (based on Euclidean distance)
- Similar shape latents are used at different global positions, as indicated by the colored points
- Our method reuses **common geometric primitives**

Ablation

Surface density	Low	Medium	High
Sphere tracing	5 min	6 min	10 min
Ours	**11 ms**	**13 ms**	**17 ms**

Superposition	Storage (MB)	gIoU	Chamfer
Direct	3.2	98.7	0.017
Addition	3.2	96.9	0.039
Concatenation	17.9	99.4	0.013

- We extract object surfaces in **real time** at inference
- **Latent vector fusion** ablation for LoD propagation