Visual Computing Institute

RWTH AACHEN UNIVERSITY

Octree Transformer: Autoregressive 3D Shape Generation on Hierarchically Structured Sequences

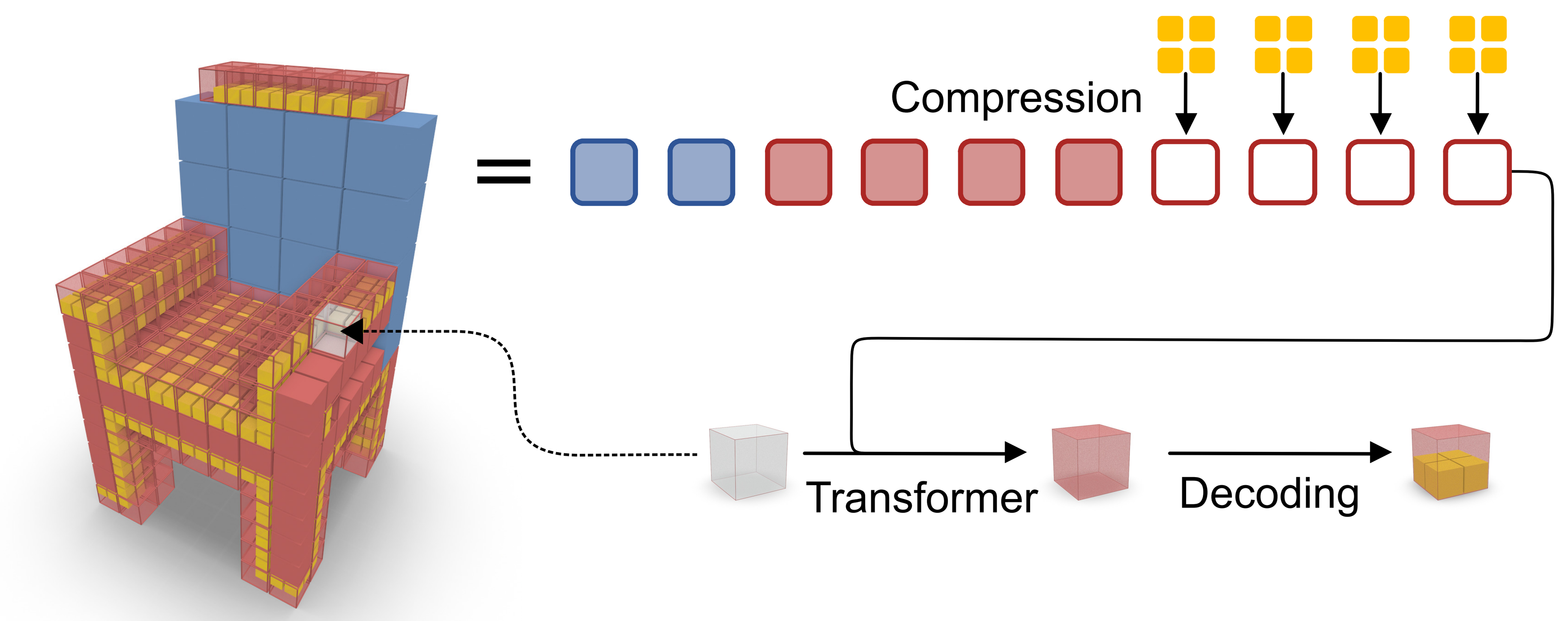
Moritz Ibing, Gregor Kobsik, Leif Kobbelt

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Overview

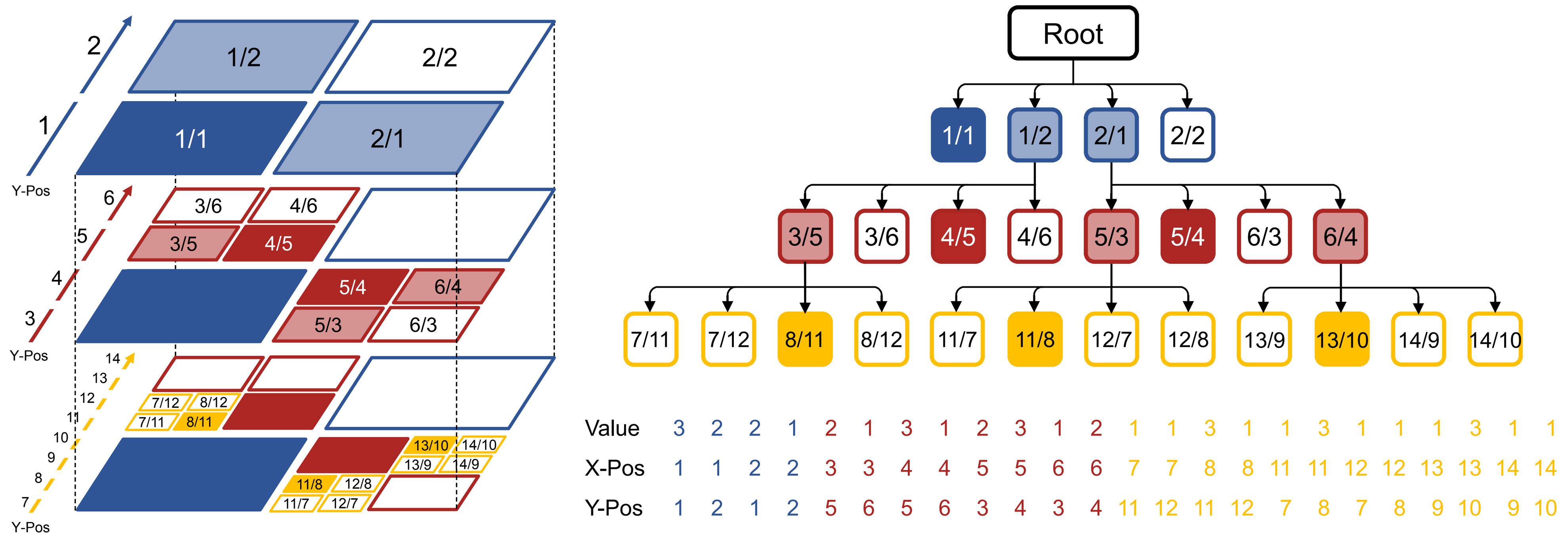
Autoregressive models are very powerful in NLP, but work on linearized data, which is not trivial to obtain for 3D data. We model **shape generation** as a **sequence generation** task:

1. Linearize octree encoded 3D shapes to a 1D sequence
2. Introduce different compression schemes to embed multiple voxels into a single token
3. Propose a fully autoregressive decoding scheme for generating octrees



Octree Linearization & Positional Encoding

- Linearize octree by **breadth-first traversal**
- Introduce novel **hierarchical positional encodig** as follows:



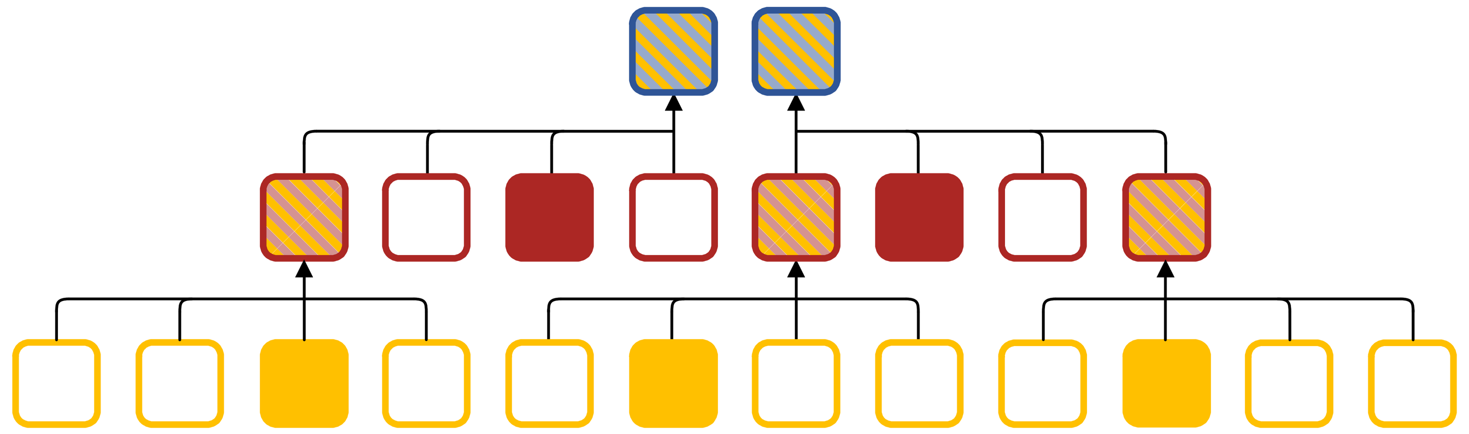
Sequence Compression

The compression follows the **hierarchical octree structure**:

- by compressing **siblings** we achive compression rates of up to 8 (4 in quadtree example)
- for higher compression rates we consider **cousins** (of higher order) and replace parent tokens by their compressed children's representation

Sequence Length with Different Compression Schemes

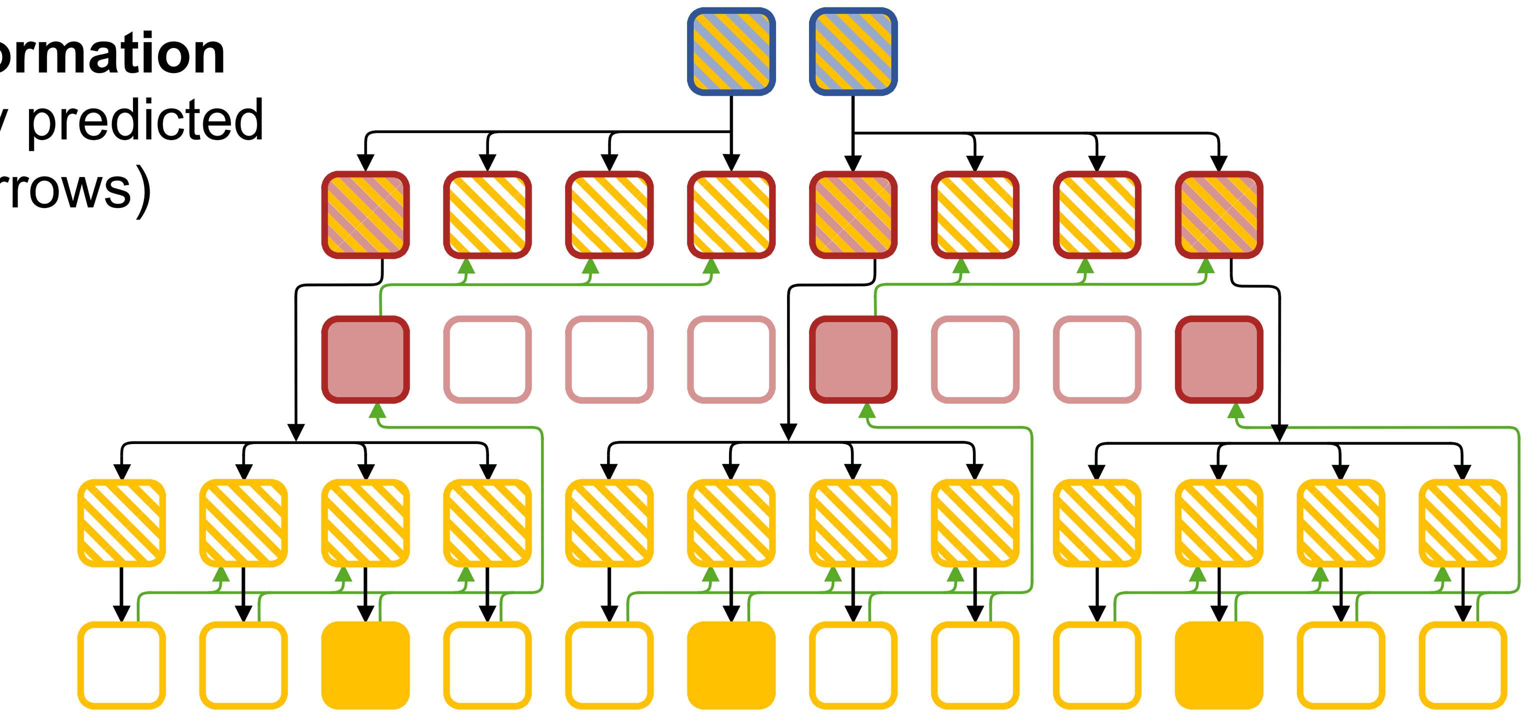
Res	Octree tokens	0/4	0/8	1/8	2/8
2	8	2	1	-	-
4	64	16	8	1	-
8	224	56	28	8	1
16	816	204	102	28	8
32	3352	838	419	102	28
64	13264	3316	1658	419	102
128	48530	12132	6066	1658	419
256	144096	36024	18012	6066	1658



Autoregressive Decoding

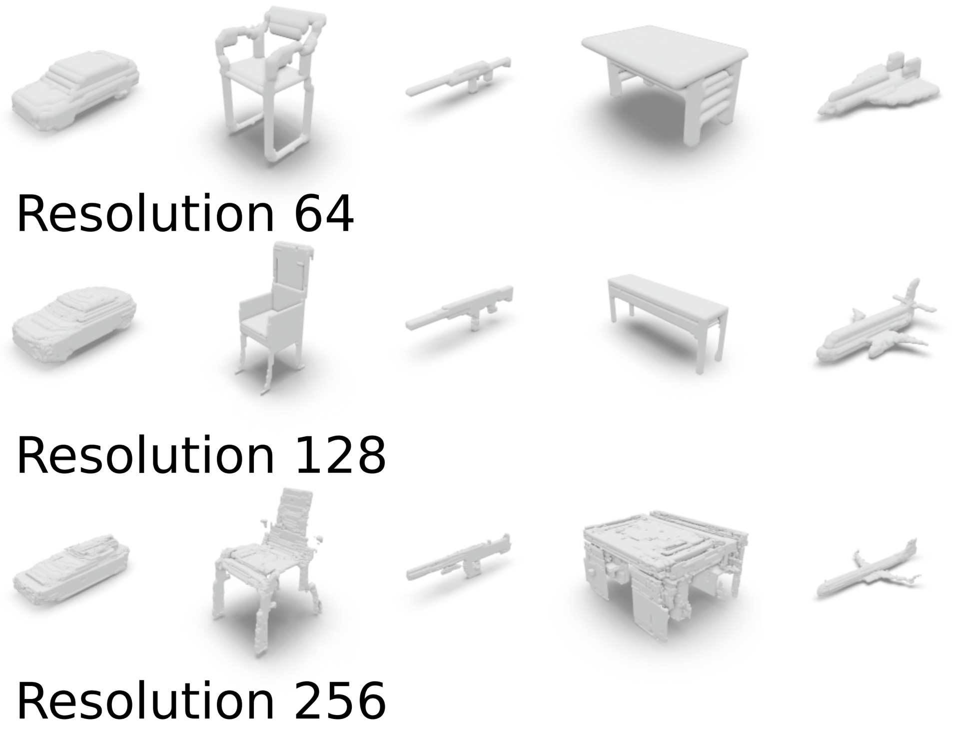
The decoding of predicted tokens needs to take the compression scheme into consideration to generate every single voxel in a **fully autoregressive** manner. We alternate between:

- **Upsampling of embedding** vectors (black arrows) and
- **Forwarding information** about previously predicted tokens (green arrows)



Results

Shape Generation

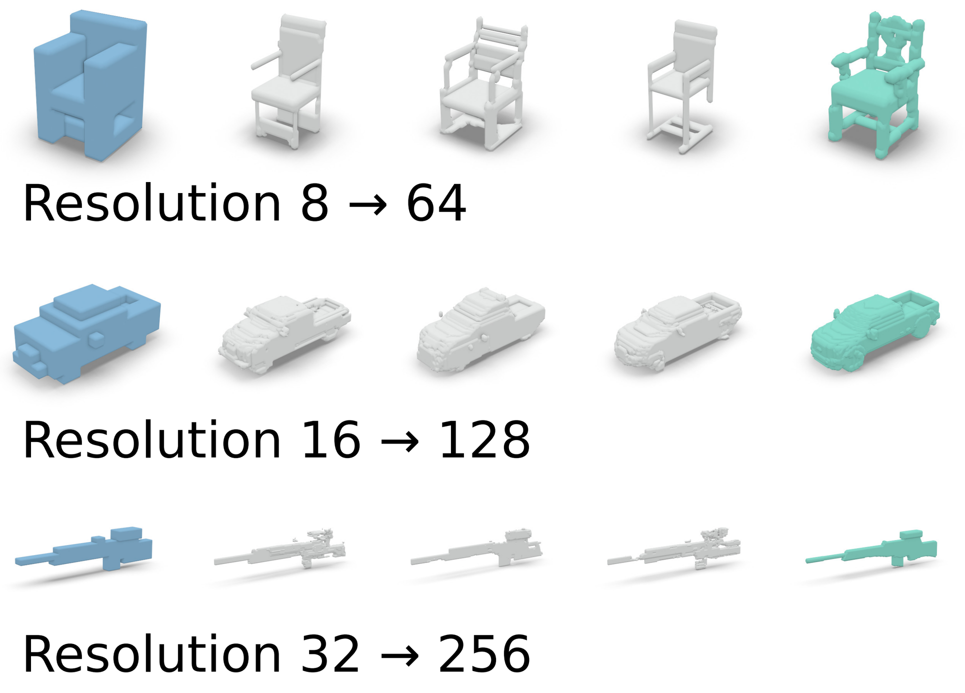


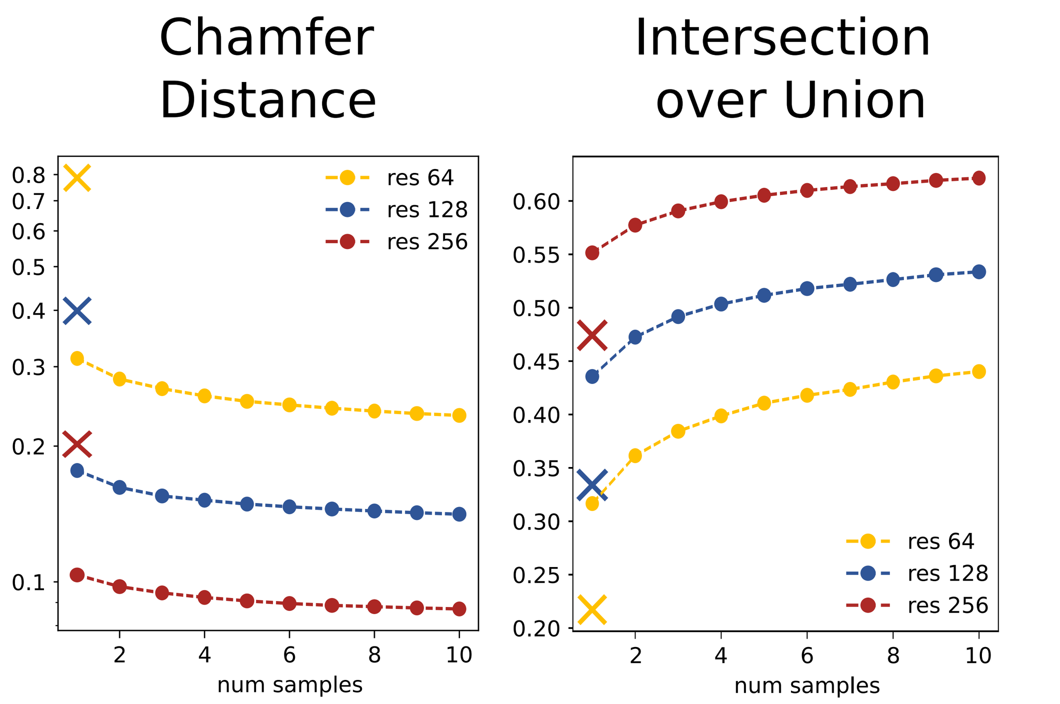
Comparison to Related Work

		Car	Chair	Rif e	Table	Plane
↑COV %	3DGAN	12.13	25.07	62.32	18.80	
	IM-GAN	69.33	75.44	65.26	86.43	70.33
	Grid IM-GAN	80.67	82.08	81.47	86.19	81.58
	Octree Transformer	60.26	76.47	60.21	80.55	73.05
	Train Set	85.67	84.73	84.00	87.13	85.04
↓MMD	3DGAN	1993	4365	4476	5208	
	IM-GAN	1287	2893	3760	2527	3689
	Grid IM-GAN	1225	2768	3366	2396	3226
	Octree Transformer	1363	2958	3582	2496	3664
	Train Set	984	2317	3085	2066	2225
↓ECD	3DGAN	28855	26279	6495	32116	
	IM-GAN	20606	2553	3288	1018	6543
	Grid IM-GAN	1062	144	94	188	355
	Octree Transformer	8563	1889	1835	1098	2573
	Train Set	11	1	2	5	1

Shape Super-resolution

We **improve the quality** by up-sampling shapes compared to the low-resolution input (**X**) over multiple non-deterministic runs.





References

- [3DGAN] Wu et al., *"Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling" in NeurIPS 2016*
- [IM-GAN] Chen et al., *"Learning Implicit Fields for Generative Shape Modeling" in CVPR 2019*
- [Grid IM-GAN] Ibing et al., *"3D Shape Generation with Grid-based Implicit Functions" in CVPR 2021*