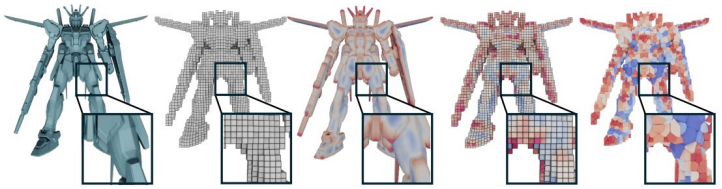
SuperVoxelGPT: Adaptive and Ordered 3D Tokenization for Autoregressive Shape Generation

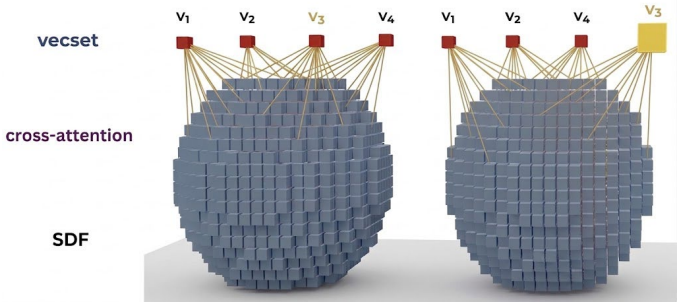
Yuan Li, Congyi Zhang, Xifeng Gao, Xiaohu Guo

ECCV 2026 | MALMÖ | SEPT 8–12

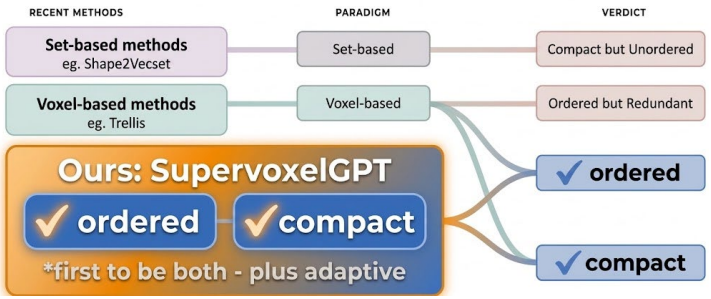
Where to generate, where to allocate tokens



Uniform tokenization spends the same number of tokens on flat regions as on detailed ones. The resulting sequence is ordered, but far too long for autoregressive 3D generation.

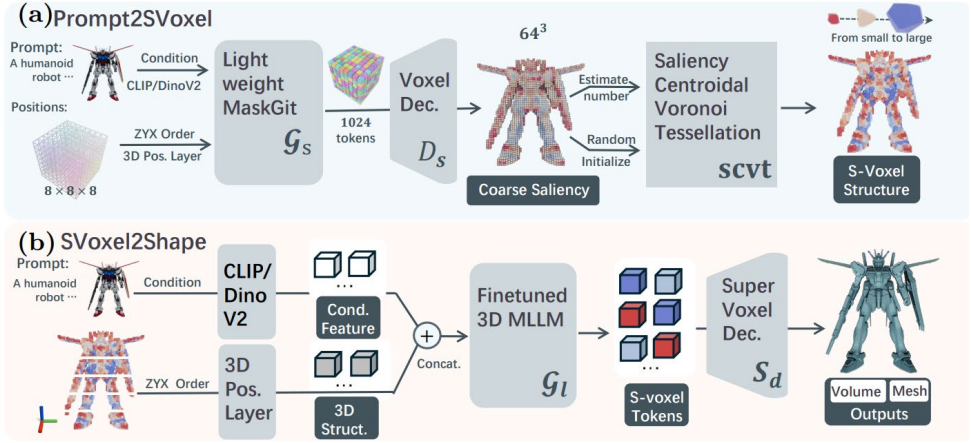


Set-based representations are compact, but unordered: a shape with k tokens admits k! valid orderings, leaving no canonical sequence to autoregressive model.

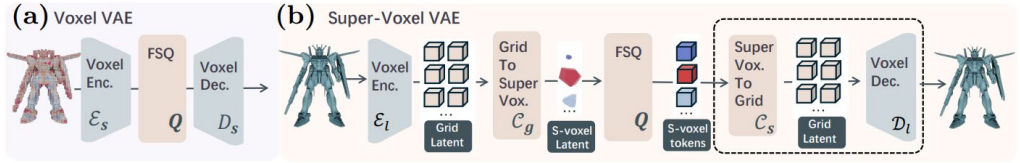


There must be a shape representation in the middle, which Is compacted and ordered.

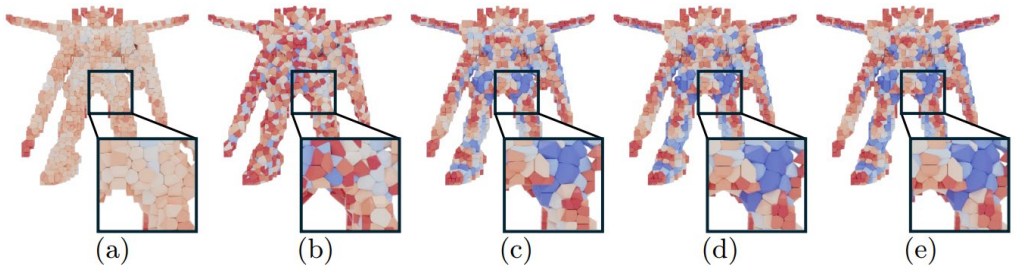
Our insight: use shape saliency to derive token allocation



A two-stage pipeline built on Trellis2: the first stage predicts saliency (occupancy) from the prompt; the second generates the shape from the prompt and saliency.

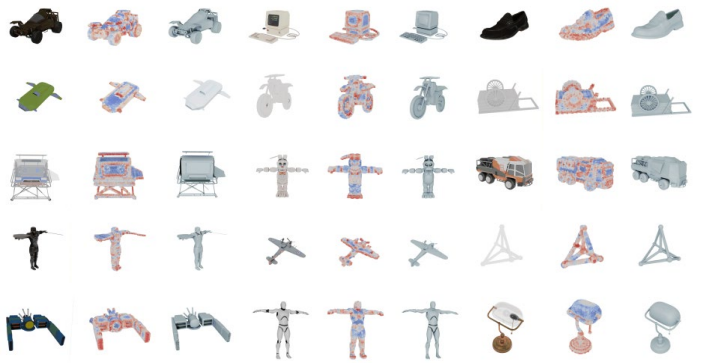


Compression by projecting uniform voxel features onto super-voxel features



Adaptive allocation yields fewer tokens; parallel decoding yields faster inference

Results



Limitation

Our saliency prediction solution may miss fine details that fall below the saliency threshold.

For noisy shapes, saliency becomes nearly uniform, so tokens cannot be compressed and our method loses its efficiency advantage.