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

Moritz Ibing, Gregor Kobsik, Leif Kobbelt

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

References

• [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

Sequence Compression

The compression follows the hierarchical octree structure:

• by compressing siblings we achieve 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

<table>
<thead>
<tr>
<th>Res</th>
<th>Octree tokens</th>
<th>0.5%</th>
<th>1%</th>
<th>1.5%</th>
<th>2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>256</td>
<td>144</td>
<td>94</td>
<td>85</td>
<td>73</td>
</tr>
<tr>
<td>128</td>
<td>256</td>
<td>144</td>
<td>94</td>
<td>85</td>
<td>73</td>
</tr>
<tr>
<td>256</td>
<td>256</td>
<td>144</td>
<td>94</td>
<td>85</td>
<td>73</td>
</tr>
</tbody>
</table>

Octree Linearization & Positional Encoding

• Linearize octree by breadth-first traversal
• Introduce novel hierarchical positional encoding as follows:

Overview

Sequence Compression

Octree Linearization & Positional Encoding

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)

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

Results

Shape Generation

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

References

• [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