Learning-based lossless compression of 3D point cloud geometry

Dat Thanh Nguyen, Maurice Quach, Giuseppe Valenzise, Pierre Duhamel

December 4, 2020
Contents

1. Introduction
2. Proposed solutions
3. Performance evaluation
4. Conclusion and future works
1. Introduction – Point Cloud

- Point Cloud (PC) is the preferred data structure of 3D applications (VR, AR, etc.)
- Each point in 3D space is represented by geometry information \((x,y,z)\) and attributes
- Point clouds are sparse
- 800,000 points -> 1 000 Mbps (uncompressed)
- We focus on **geometry coding**
1. Introduction – PC geometry

- Representations of PC geometry:
  - Voxels: quantize PC into a pre-defined precision, PC → D x D x D voxels, less than 2% voxels are occupied
  - Octree: recursively split voxelized PC into 8 sub-cubes, PC → sequence of octets

- Losslessly encode voxel value using arithmetic coding? **Probability model**
1. Introduction – Problem definition

- Partition point cloud into high level octree + non-empty 64x64x64 blocks
- High level octree is transmitted as side information
- We focus on encoding **64x64x64 blocks**
2. VoxelDNN

- Context adaptive arithmetic coding on voxel domain
- Probability model \( p(v) \) for each block \( d \times d \times d \) voxels:

\[
p(v) = \prod_{i=1}^{d^3} p(v_i | v_{i-1}, v_{i-2}, \ldots, v_1)
\]

- Estimate each term using a Deep Neural Network:
  - Input: block of \( d \times d \times d \) voxels (\( v_1 \) to \( v_{d^3} \))
  - Output: conditional probability distribution of each \( v_i : \hat{p}(v_i | v_{i-1}, v_{i-2} \ldots v_1) \)
  - Causality constraints: \( \hat{p}(v_i) \) only depend on previous voxels: \( v_{i-1}, v_{i-2}, \ldots, v_1 \)
2. VoxelDNN

- Causality is enforced by using masked filters in each convolutional layer: type A and type B mask [1]
- 2 residual connections to avoid vanishing gradient and speed up convergence
- Cross-Entropy (CE) loss: \( H(p, \hat{p}) = \mathbb{E}_{v \sim p(v)} \left[ \sum_{i=1}^{d^3} - \log \hat{p}(v_i) \right] \).
2. Multi-resolution encoder

- Encode voxels in each block sequentially from the first voxel to the last voxel
- Partition blocks into multiple child blocks at different size to eliminate the sparsity
- Rate-optimized multi-resolution algorithm: select the partitioning solution that has minimum bitrate.

![Diagram](image)

From left to right: 1, 2, 3, and 4 partitioning level
3. Experimental setup

- The training data is mixed from:
  - MPEG 8i*: longdress10, soldier10: 4297 blocks 64
  - Microsoft MVUB*: andrew10, david10, sarah10: 4820 blocks 64
  - ModelNet*: 200 heaviest PCs: 11147 blocks 64
  - Total: 18291 training blocks, 1973 blocks for validation

- Training:
  - 50 epochs with early stopping, batch = 8
  - Adam optimizer, lr=1e-3

- Test dataset: sequences from MPEG 8i and Microsoft MVUB

*MPEG 8i: http://plenodb.jpeg.org/pc/8ilabs
Microsoft MVUB: http://plenodb.jpeg.org/pc/microsoft
ModelNet40: https://modelnet.cs.princeton.edu
3. Experimental setup

- Some testing Point Clouds

Phil 10 bits
Ricardo 10 bits
Loot 10 bits
Redandblack 10 bits
3. Experimental results

- Average bits per occupied voxel (bpov) for 4 partitioning level: G-PCC[2], P(PNI)[3]

Table 1: Average rate in bpov of VoxelIDNN at different partitioning levels compared with MPEG G-PCC and P(PNI).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Point Cloud</th>
<th>P(PNI) bpov</th>
<th>G-PCC bpov</th>
<th>block 64 bpov</th>
<th>Gain over G-PCC</th>
<th>block 64 + 32 bpov</th>
<th>Gain over G-PCC</th>
<th>block 64 + 32 + 16 bpov</th>
<th>Gain over G-PCC</th>
<th>block 64 + 32 + 16 + 8 bpov</th>
<th>Gain over G-PCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVUB</td>
<td>Phil9</td>
<td>1.88</td>
<td>1.2284</td>
<td>0.9819</td>
<td>20.07%</td>
<td>0.9317</td>
<td>24.15%</td>
<td>0.9203</td>
<td>25.08%</td>
<td>0.9201</td>
<td>25.10%</td>
</tr>
<tr>
<td></td>
<td>Ricardo9</td>
<td>1.79</td>
<td>1.0422</td>
<td>0.7910</td>
<td>24.10%</td>
<td>0.7276</td>
<td>30.19%</td>
<td>0.7175</td>
<td>31.16%</td>
<td>0.7173</td>
<td>31.17%</td>
</tr>
<tr>
<td></td>
<td>Phil10</td>
<td>-</td>
<td>1.1617</td>
<td>0.8941</td>
<td>23.04%</td>
<td>0.8381</td>
<td>27.86%</td>
<td>0.8308</td>
<td>28.48%</td>
<td>0.8307</td>
<td>28.49%</td>
</tr>
<tr>
<td></td>
<td>Ricardo10</td>
<td>-</td>
<td>1.0672</td>
<td>0.8108</td>
<td>24.03%</td>
<td>0.7596</td>
<td>28.82%</td>
<td>0.7539</td>
<td>29.36%</td>
<td>0.7533</td>
<td>29.41%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>1.84</td>
<td>1.1248</td>
<td>0.8694</td>
<td><strong>22.71%</strong></td>
<td>0.8142</td>
<td><strong>27.61%</strong></td>
<td>0.8056</td>
<td><strong>28.38%</strong></td>
<td>0.8053</td>
<td><strong>28.41%</strong></td>
</tr>
<tr>
<td>MPEG 8i</td>
<td>Loot10</td>
<td>1.69</td>
<td>0.9524</td>
<td>0.7016</td>
<td>26.33%</td>
<td>0.6464</td>
<td>32.13%</td>
<td>0.6400</td>
<td>32.80%</td>
<td>0.6387</td>
<td>32.94%</td>
</tr>
<tr>
<td></td>
<td>Redandblack10</td>
<td>1.84</td>
<td>1.0889</td>
<td>0.7921</td>
<td>27.26%</td>
<td>0.7383</td>
<td>32.20%</td>
<td>0.7317</td>
<td>32.80%</td>
<td>0.7317</td>
<td>32.80%</td>
</tr>
<tr>
<td></td>
<td>Boxer9</td>
<td>-</td>
<td>1.0815</td>
<td>0.8034</td>
<td>25.71%</td>
<td>0.7620</td>
<td>29.54%</td>
<td>0.7558</td>
<td>30.12%</td>
<td>0.7560</td>
<td>30.14%</td>
</tr>
<tr>
<td></td>
<td>Thaidancer9</td>
<td>-</td>
<td>1.0677</td>
<td>0.8574</td>
<td>19.70%</td>
<td>0.8145</td>
<td>23.71%</td>
<td>0.8091</td>
<td>24.22%</td>
<td>0.8078</td>
<td>24.34%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>1.77</td>
<td>1.0476</td>
<td>0.7886</td>
<td><strong>24.75%</strong></td>
<td>0.7403</td>
<td><strong>29.34%</strong></td>
<td>0.7341</td>
<td><strong>29.92%</strong></td>
<td>0.7334</td>
<td><strong>29.99%</strong></td>
</tr>
</tbody>
</table>
3. Experimental results

- Percentage of encoded points in each block size
4. Conclusion

- VoxelDNN:
  - Hybrid octree/voxel-based lossless compression method
  - The first deep generative model in voxel space
  - Multi-resolution encoder

- 28% gain over MPEG G-PCC standard

- Future works:
  - More powerful generative model
  - Jointly encode geometry and attributes
References


Q&A