ICIP2022 BORDEAUX

IEEE Signal Processing Society

IEEE

Learning-based Lossless Point Cloud Geometry Coding using Sparse Tensors

Dat Thanh Nguyen and André Kaup

Friedrich-Alexander-Universität
Technische Fakultät

LMS

EmpkinS

October 16, 2022

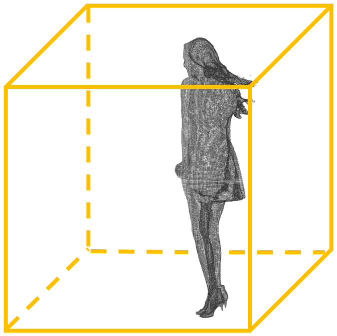
1. Point Cloud and Point Cloud Compression

- Point Cloud: a set of points representing 3D scenes (Human bodies, objects, 3D maps,…).
- Each point represented by:
  - Geometry coordinates: x,y,z
  - Attributes: RGB, reflectance, etc.
- Our work focuses on geometry compression

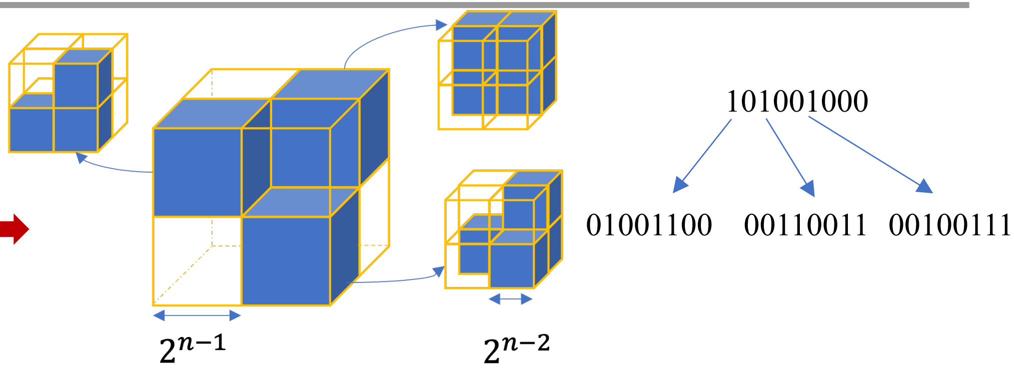
Point cloud in the video: https://www.mpeg.org/

1. Point Cloud and Point Cloud Compression

x	y	z
16	10	59
71	06	95
11	60	20
31	80	02
…		



$2^n \times 2^n \times 2^n$ voxels



Point representation

- Directly acquired from sensors/cameras
- Tabular data

Voxel representation

- Can be directly process by conventional CNN
- Huge computational power

Octree representation

- Eliminate the sparsity
- Lower uncompressed geometry bitrate
- Geometric information lost

Sparse-tensor $\{\boldsymbol{C}, \boldsymbol{F}\}$ [1]

[1] Choy, C., Gwak, J., & Savarese, S. (2019). 4d spatio-temporal convnets: Minkowski convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3075-3084).

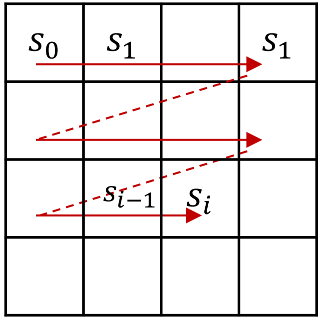
2. SparseVoxelDNN – Lossless geometry compression

❖ **Arithmetic coder:** Based on the probability density function p of each source symbol s to be encoded

- Average bit per symbol: $L = \sum_{s \in S} p(s) l(s)$
- Shannon theorem: $L \geq H(p) = \sum_{s \in S} p(s) \log(p(s))$ [2]

❖ p is unknown, how do we learn the probability density estimation $\hat{p}$?

→ Factorize the joint distribution p(S) as a product of conditional distributions [3]



$$p(S) = \prod_{i=1}^{n} p(s_i | s_{i-1}, \ldots, s_1, s_0)$$

[2]Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.
[3] Van den Oord, Aaron, et al. "Conditional image generation with pixelcnn decoders." Advances in neural information processing systems 29 (2016).

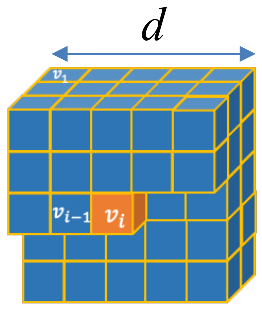
2. SparseVoxelDNN – Lossless geometry compression

❖ Factorize the joint occupancy probability of a voxel block v

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

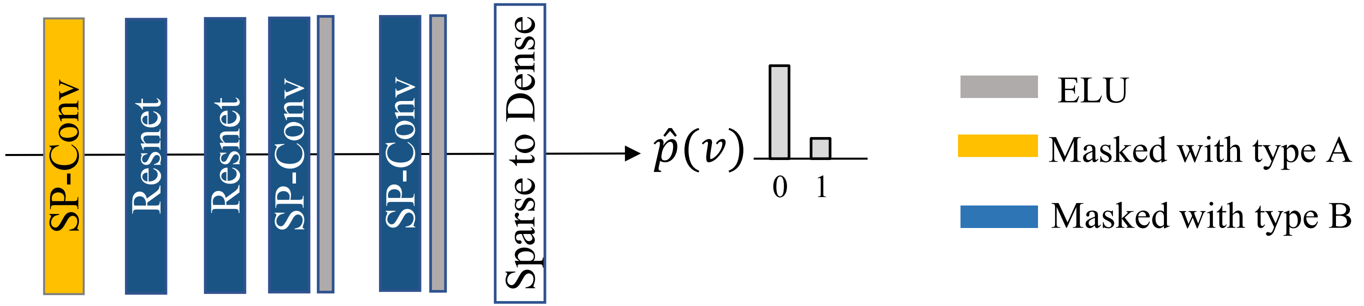
❖ A sparse-tensor based neural network estimates $p(v)$ by minimizing $\tilde{L}$:

→ Sub-optimal lower bound of bitrate: $\tilde{L} = \sum_{v_i \in S} p(v_i) \log(\hat{p}(v_i))$



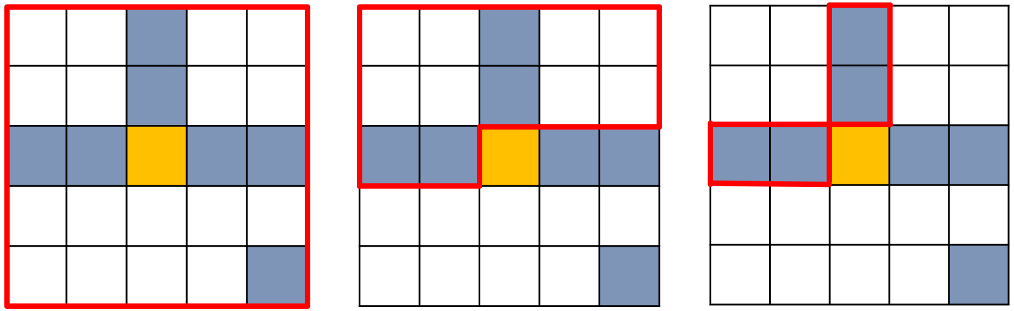
A voxel block $\boldsymbol{v}$

x	y	z
16	10	59
71	06	95
11	60	20
31	80	02
…		



2. SparseVoxelDNN – Lossless geometry compression

- To enforce the causality constraint, we employ two sparse masks inspired by [4]



No masks, violating causality | Applying type A and type B mask from [4] | Applying sparse type A and type B mask

1	1	1
1	0	0
0	0	0

Type A mask

1	1	1
1	1	0
0	0	0

Type B mask

[4] Nguyen, Dat Thanh, et al. "Lossless coding of point cloud geometry using a deep generative model." IEEE Transactions on Circuits and Systems for Video Technology 31.12 (2021): 4617-4629.

2. SparseVoxelDNN – Lossless geometry compression

❖ Training Dataset:

Dataset	**8i**	**CAT1**	**Owlii**	**MVUB**	**Total**
Content	Dynamic full human bodies, dense	Static, culture heritage, sparse	Dynamic human mesh sequences, dense	Dynamic upper bodies, dense	
No blocks 64	235.369	2499	358.834	317763	914.465



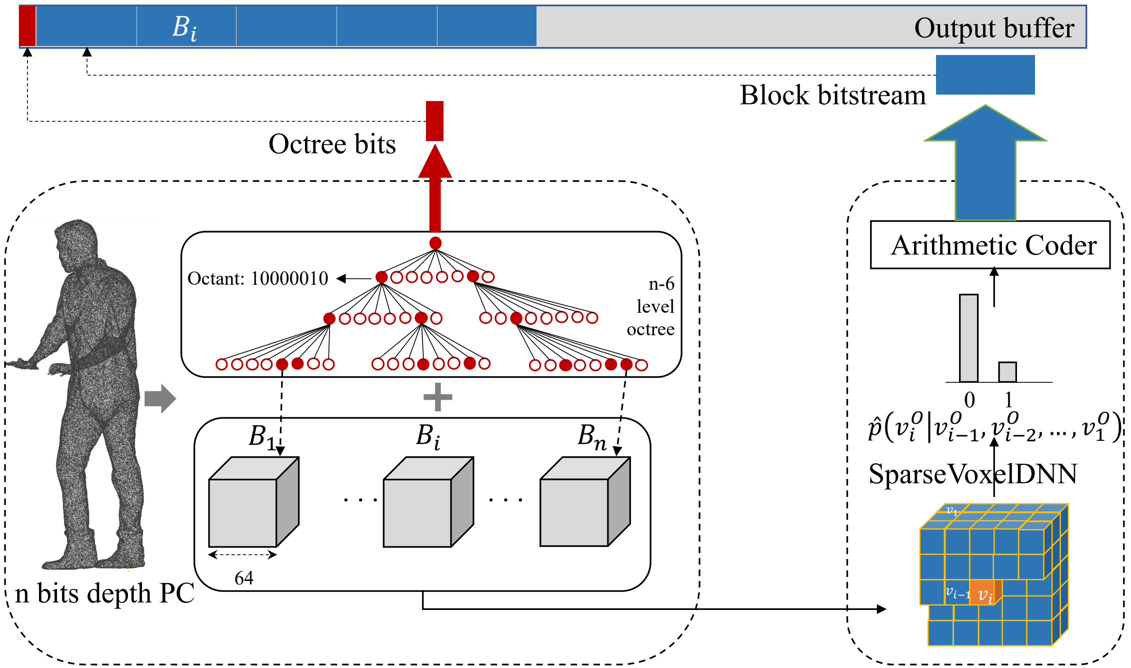

MPEG 8i: http://plenodb.jpeg.org/pc/8ilabs
MPEG CAT1: https://mpegfs.int-evry.fr/mpegcontent
MPEG Owlii: https://mpeg-pcc.org/index.php/pcc-content-database/owlii-dynamic-human-textured-mesh-sequence-dataset/
MVUB: http://plenodb.jpeg.org/pc/microsoft

2. SparseVoxelDNN – Lossless geometry compression

❖ Encoder:

➢ PC are first partitioned in to blocks of size 64

➢ Encode voxel by voxel using SparseVoxelDNN context model



2. SparseVoxelDNN – Lossless geometry compression

- Bpov: bit per occupied voxel/bit per point
- GPCC v14: PCC method from MPEG
- VoxelDNN: Deep generative based method [4]
- Test PCs from four widely used datasets

		G-PCC	VoxelDNN		SparseVoxelDNN	
	Point Cloud	bpov	bpov	Gain over G-PCC	bpov	Gain over G-PCC
MVUB	Phil	1.15	0.82	-28%	0.40	-66%
	Ricardo	1.07	0.75	-29%	0.35	-67%
	Average	**1.11**	**0.79**	**-28.5%**	**0.37**	**-66.4%**
8i	Redandblack	1.09	0.70	-36%	0.30	-64%
	Loot	0.95	0.61	-36%	0.33	-65%
	Thaidancer	1.00	0.66	-34%	0.32	-68%
	Boxer	0.95	0.59	-38%	0.30	-68%
	Average	**1.00**	**0.64**	**-36.0%**	**0.34**	**-66.5%**
CAT1	Frog	1.92	1.70	-10%	1.23	-36%
	Arco Valentino	4.85	4.99	+3%	3.03	-38%
	Shiva	3.67	3.50	-4%	2.66	-28%
	Average	**3.48**	**3.40**	**-3.8%**	**2.31**	**-33.6%**
USP	BumbaMeuBoi	5.41	5.07	-6%	1.81	-67%
	RomanOiLight	1.86	1.62	-12%	1.98	+6%
	Average	**3.64**	**3.49**	**-9.5%**	**1.90**	**-30.5%**

[4] Nguyen, Dat Thanh, et al. "Lossless coding of point cloud geometry using a deep generative model." IEEE Transactions on Circuits and Systems for Video Technology 31.12 (2021): 4617-4629.

2. SparseVoxelDNN – Lossless geometry compression

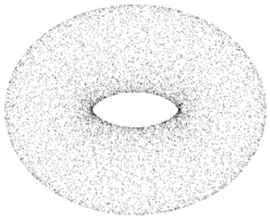
- Computational complexity
  - Encoder 50 times faster than VoxelDNN
  - Encoder 4.7 times slower than G-PCC (v14)
  - Long decoding time

	G-PCC	VoxelDNN	SparseVoxelDNN
(Enc)	1.6	355	7.2
(Dec)	1.1	330	229

3. Conclusions

❖ Conclusions

- ➢ Representing the point cloud using both octree, voxel and point representation
- ➢ Estimating voxel occupancy using point-based neural network with 3D sparse convolution masks
- ➢ Average rate saving of 50% over G-PCC while avoiding expensive computation



Question?