Point Cloud Processing & Compression

Zhu Li

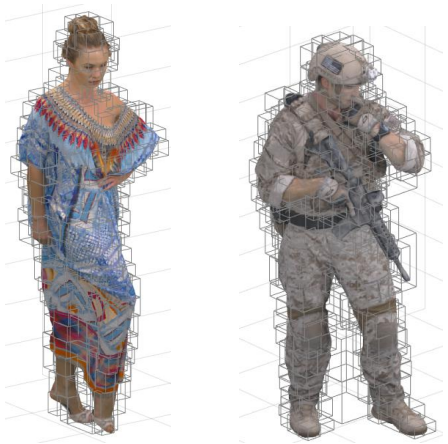
Director, UMKC NSF Center for Big Learning

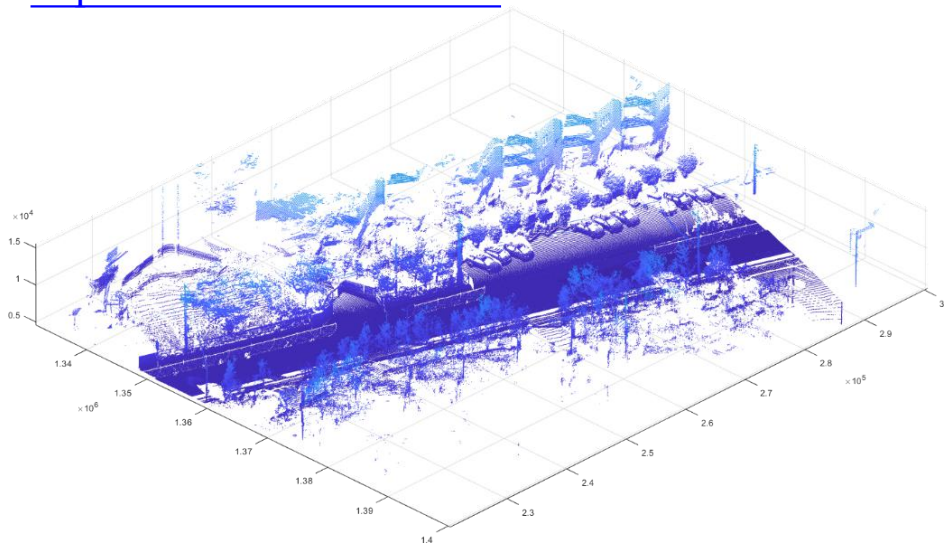
Dept of Computer Science & Electrical Engineering

University of Missouri, Kansas City

Email: zhu.li@ieee.org, lizhu@umkc.edu

Web: http://l.web.umkc.edu/lizhu





Outline

- Short Self Intro
- Research Motivation and Highlights
- Sparse Conv Engine based PCC work
- Video based PCC work
- Summary

Short Bio:

Research Interests:

- Immersive Media Communication: light field, point cloud and 360 video capture, coding and low latency communication.
- Data & Image Compression: video, medical volumetric data, DNA sequence, and graph signal compression with deep learning
- Remote Sensing & Vision: vision problem under low resolution, blur, and dark conditions, hyperspectral imaging, sensor fusion
- Edge Computing & Federated Learning: gradient compression, light weight inference time engine, retina features, fast edge cache for video CDN

NSF I/UCRC Center *for* Big Learning
Creating Intelligence

signal processing and learning

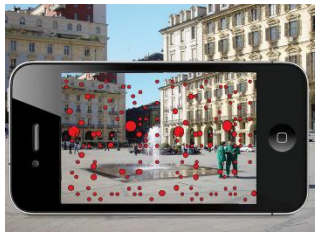
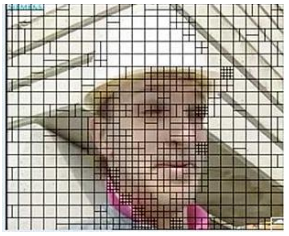
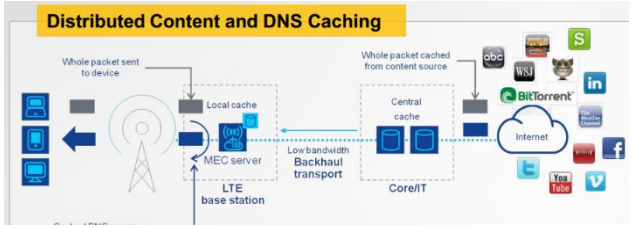


image understanding



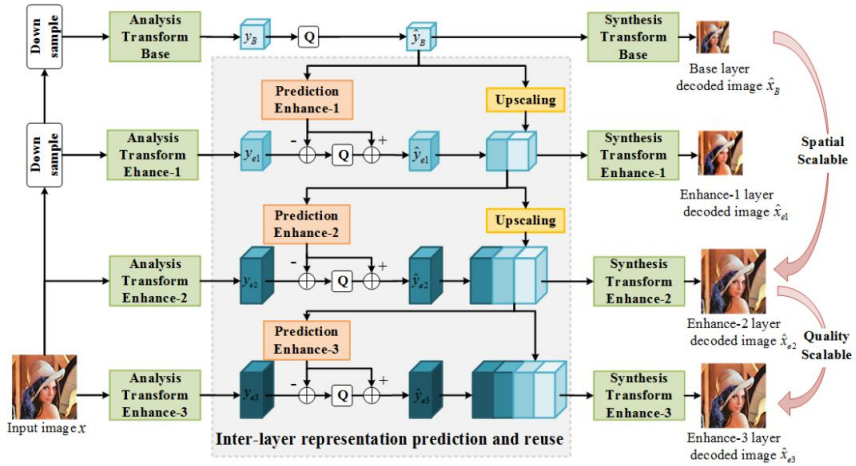
visual communication

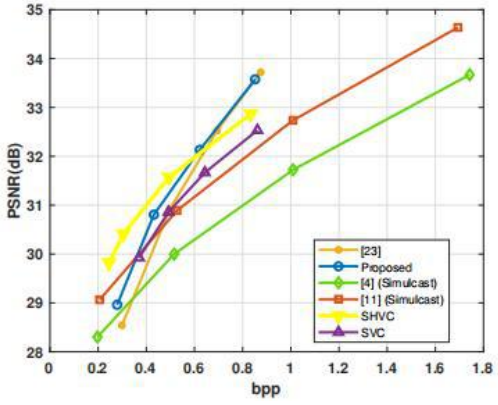


mobile edge computing & communication

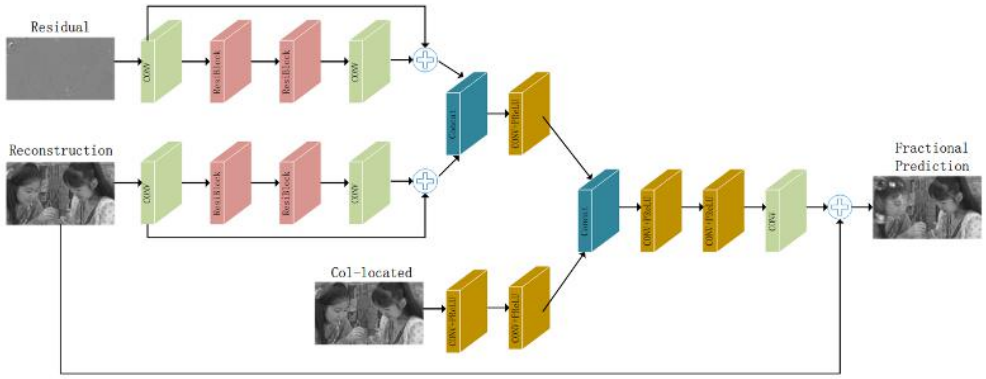
Data & Image Compression Highlights

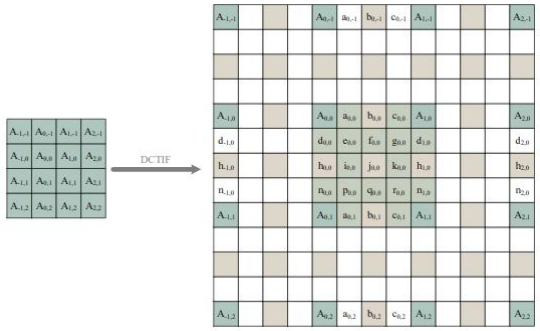
- Y. Mei, L. Li, Z. Li, and F. Li, "Learning-Based Scalable Image Compression with Latent-Feature Reuse and Prediction", *IEEE Trans on Multimedia* (T-MM), 2021.





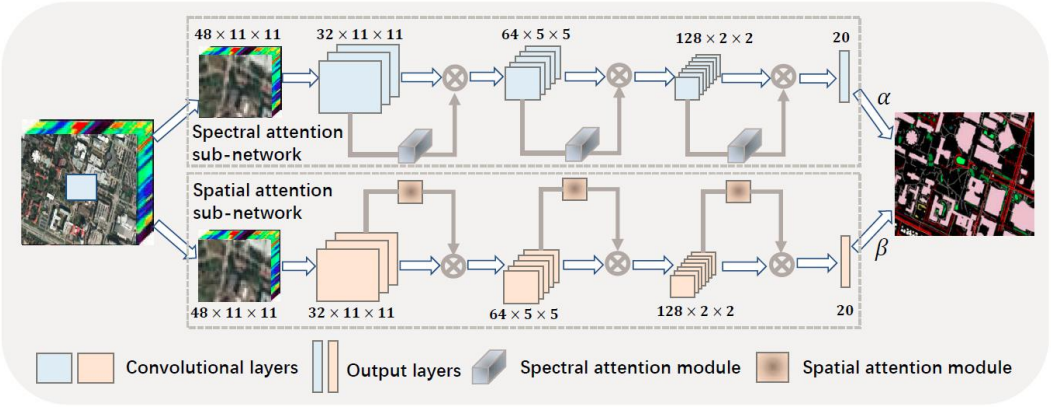
- H. Zhang, L. Song, L. Li, Z. Li, and X.K. Yang"Compression Priors Assisted Convolutional Neural Network for Fractional Interpolation", *IEEE Trans on Circuits and Systems for Video Tech,* 2020





Remote Sensing & Vision Highlights

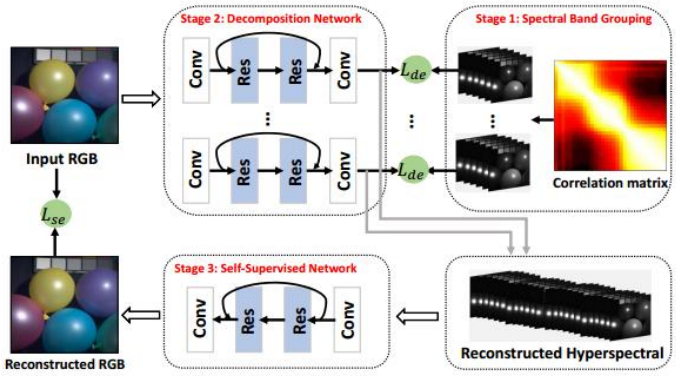
- R. Hang, Z. Li, Q. Liu, P. Ghamisi and S. Bhattacharyya, "**Hyperspectral Image Classification with Attention Aided CNNs**", *IEEE Trans. on Geoscience & Remote Sensing* (T-GRS), 2020.



Attention CNN for Hyperspectral Image Classification

- Introducing a dual stream network architecture with separate attention model for spatial and spectral feature maps
- Achieving the SOTA performance.

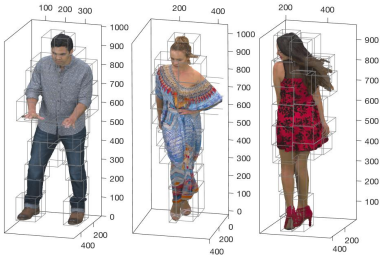
- R. Hang, Q. Liu, and Z. Li, "Spectral Super-Resolution Network Guided by Intrinsic Properties of Hyperspectral Imagery", *IEEE Trans on Image Processing*(T-IP), 2021



PRINET: Spectral Super Resolution

- Super-resolve hyper-spectral info from RGB inputs
- A dual loss network that learn a correlation decomposed HSI images
- Achieving the new SOTA performance.

Immersive Media Coding & Communication (NSF/IUCRC)




- "Deep Learning Geometry Compression Artifacts Revomal for Video Based Point Cloud Compression", *Int'l Journal on Computer Vision* (IJCV), 2021.
- "Video-based Point Cloud Compression Artifact Removal", *IEEE Trans on Multimedia* (T-MM), 2021.
- "Efficient Projected Frame Padding for Video-based Point Cloud Compression", *IEEE Trans on Multimedia*(T-MM), 2020.
- "Rate Control for Video-based Point Cloud Compression", *IEEE Transactions on Image Processing* (T-IP), 2020.
- "λ-domain Perceptual Rate Control for 360-degree Video Compression", *IEEE Journal of Selected Topics in Signal Processing* (JSTSP), 2020.
- "Advanced 3D Motion Prediction for Video Based Dynamci Point Cloud Compression", *IEEE Trans on Image Processing*(T-IP), 2019.
- "Quadtree-based Coding Framework for High Density Camera Array based Light Field Image", *IEEE Trans on Circuits and Systems for Video Tech*(T-CSVT), 2019.
- "Advanced Spherical Motion Model and Local Padding for 360 Video Compression", *IEEE Trans on Image Processing* (T-IP) vol. 28, no. 5, pp. 2342-2356, May 2019.
- "Scalable Point Cloud Geometry Coding with Binary Tree Embedded Quadtree", *IEEE Int'l Conf. on Multimedia & Expo* (ICME) ,San Diego, USA, 2018.
- "Pseudo sequence based 2-D hierarchical coding structure for light-field image compression", *IEEE Journal of Selected Topics in Signal Processing* (JSTSP), Special Issue on Light Field, 2017.

Edge Media Computing & Federated Learning

- "Referenceless Rate-Distortion Modeling with Learning from Bitstream and Pixel Features", *ACM Multimedia* (MM), Seattle, 2020.

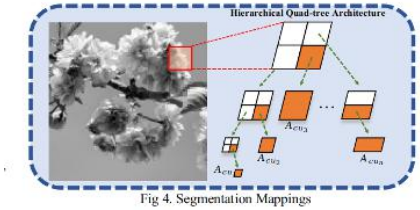


Fig 4. Segmentation Mappings

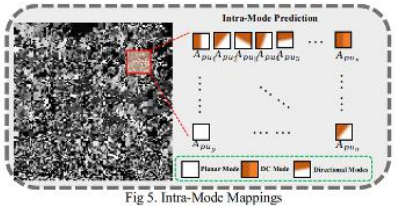
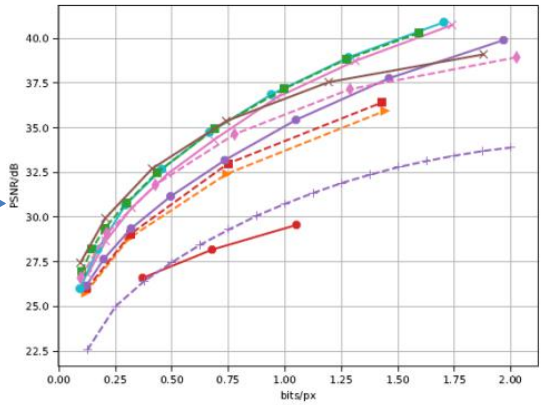


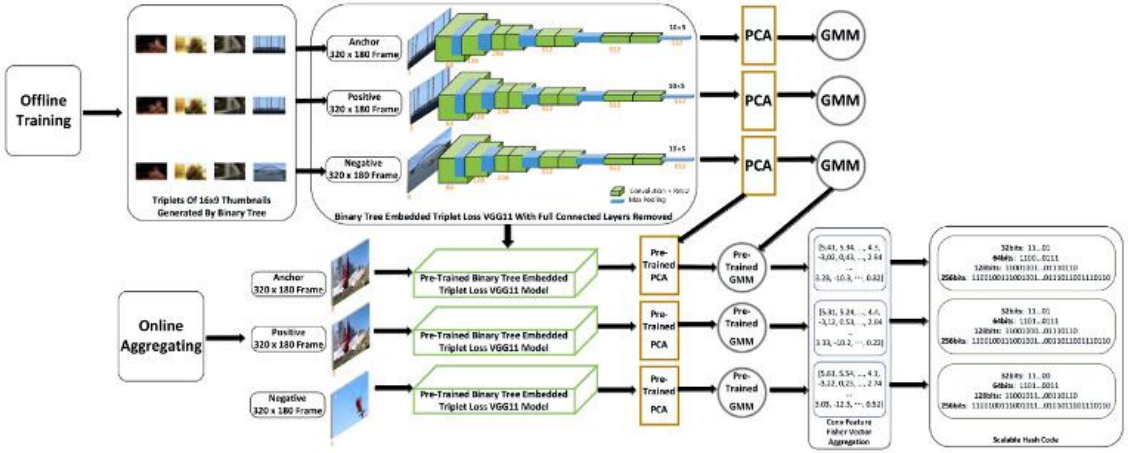
Fig 5. Intra-Mode Mappings

learn from one encoding

ref-less R-D modeling



- "Scalable Hash From Triplet Loss Feature Aggregation for Video De-Duplication", *Journal of Visual Communication & Image Representation* (JVCIR), 2020.



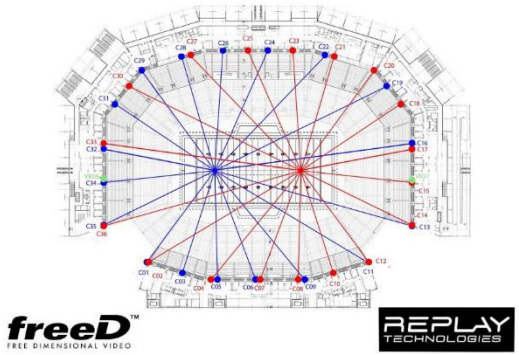
0-FPR deduplication at <10ms latency for very large repository

What is Point Cloud

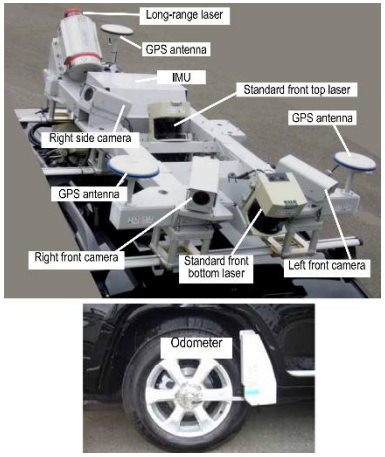
- A collection of Un-ordered points with
  - Geometry: expressed as [x, y, z]
  - Color Attributes: [r g b], or [y u v]
  - Additional info: normal, timestamp, …etc.
- Key difference from mesh: no order or local topology info

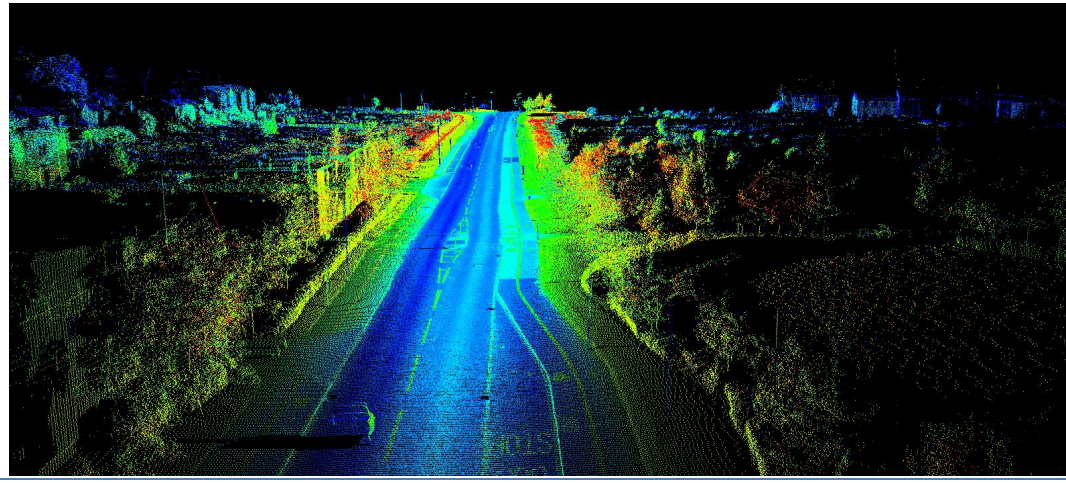
Point Cloud Capture

- Passive: Camera array stereo depth senso



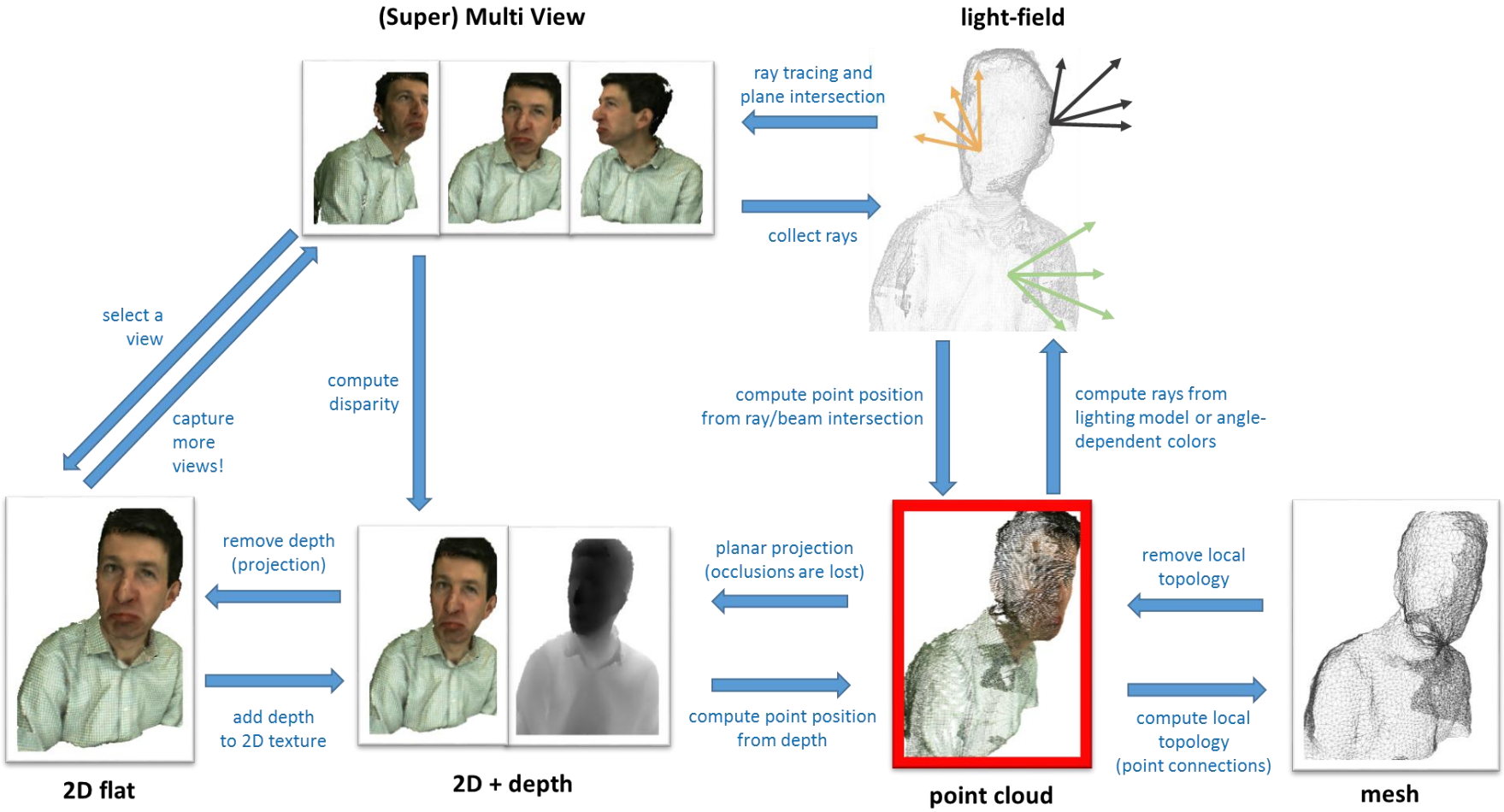
- Active: LiDAR, mmWave, TOF sensors





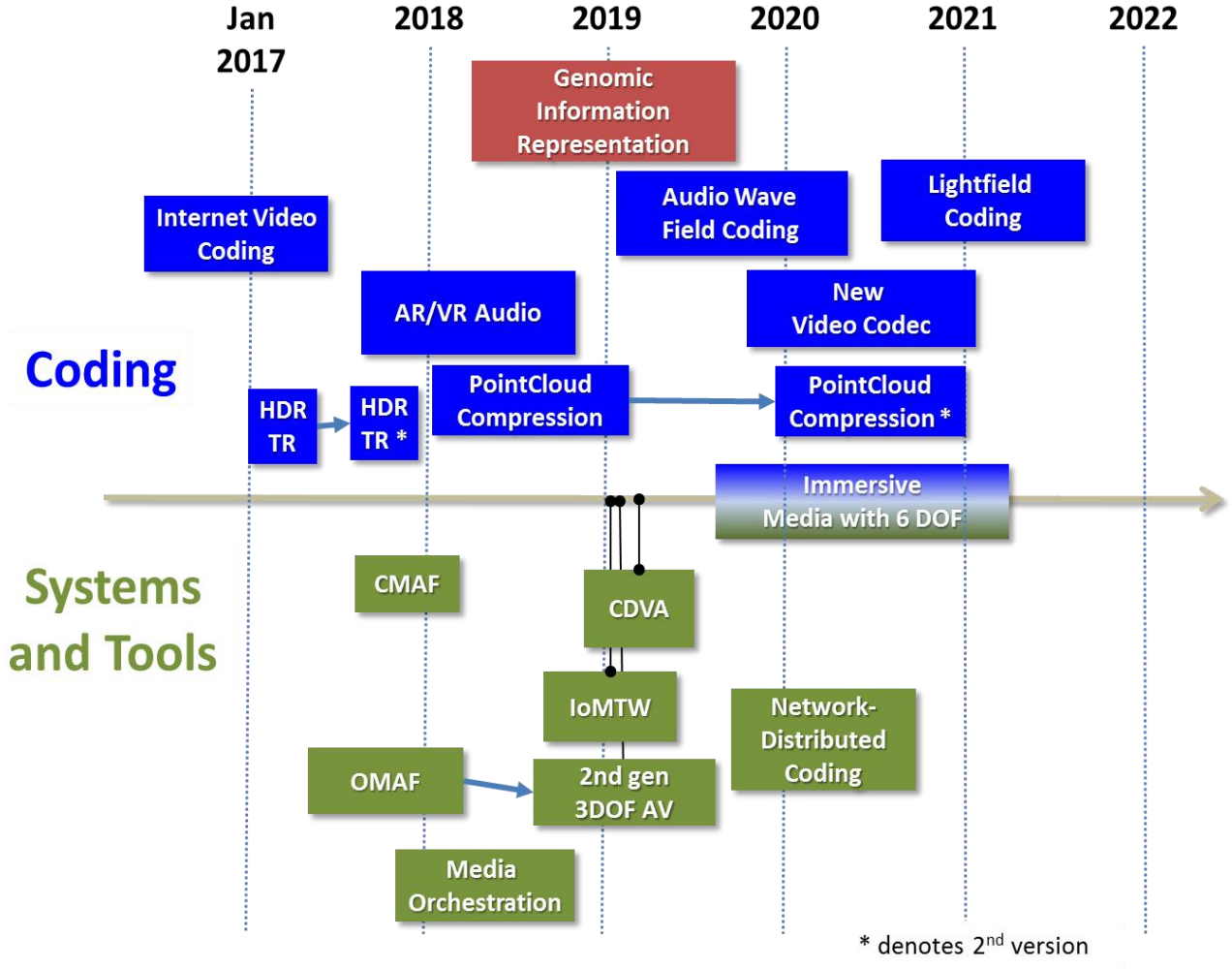
Point Cloud Inter-Operability with Other Formats

- Provide true 6-DoF Content capacity



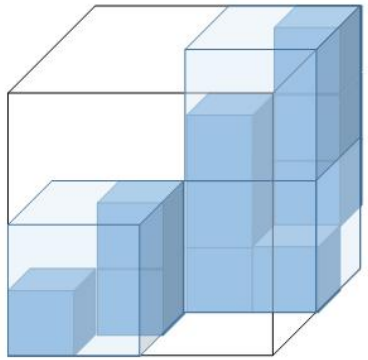
PCC in MPEG

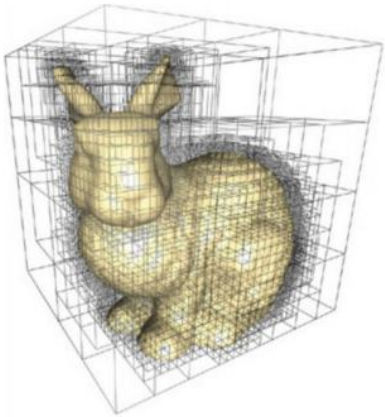
- Part of the MPEG-Immersive grand vision

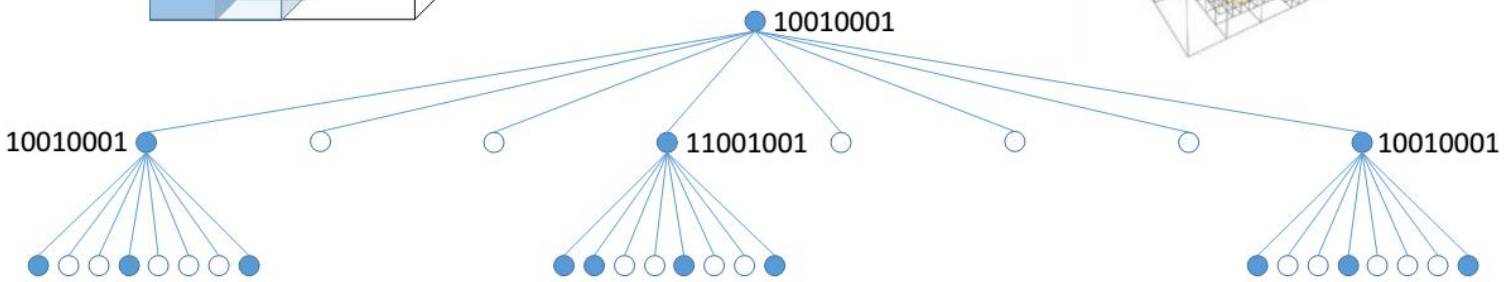


Octree Based Point Cloud Compression

- Octree is a space partition solution
  - Iteratively partition the space into sub-blocks.
  - Encoding: 0 if empty, 1 if contains data points
  - Level of the tree controls the quantization error







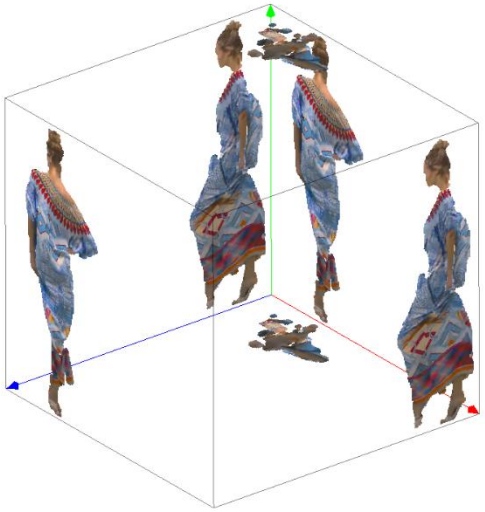
Credit: Phil Chou, PacketVideo 2016

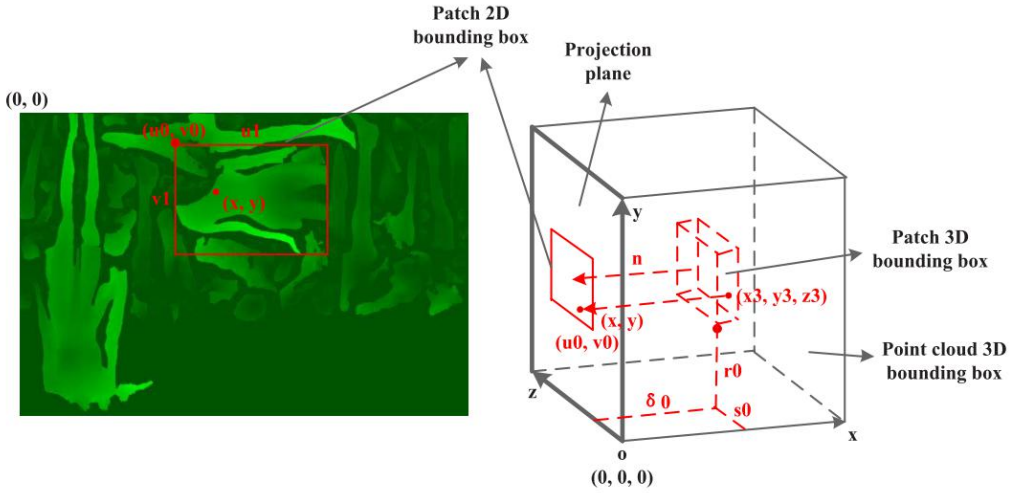
Deep Learning tools for Point Cloud

- Grid-based architectures:
  - Voxelization-based [1][2].
  - Projection-based [3][4][5][6].
- Point-based methods (usually involves knn search):
  - PointNet [7], PointNet++ [8].
  - PointCNN [9].
- New Data Representations:
  - Graph CNNs [10][11][12].
  - Octree-based CNN [13].
  - Sparse CNNs: Submanifold sparse convolutions [14], **SparseConvNet** [15], and **MinkowskiNet** [16]

Video-based point cloud compression

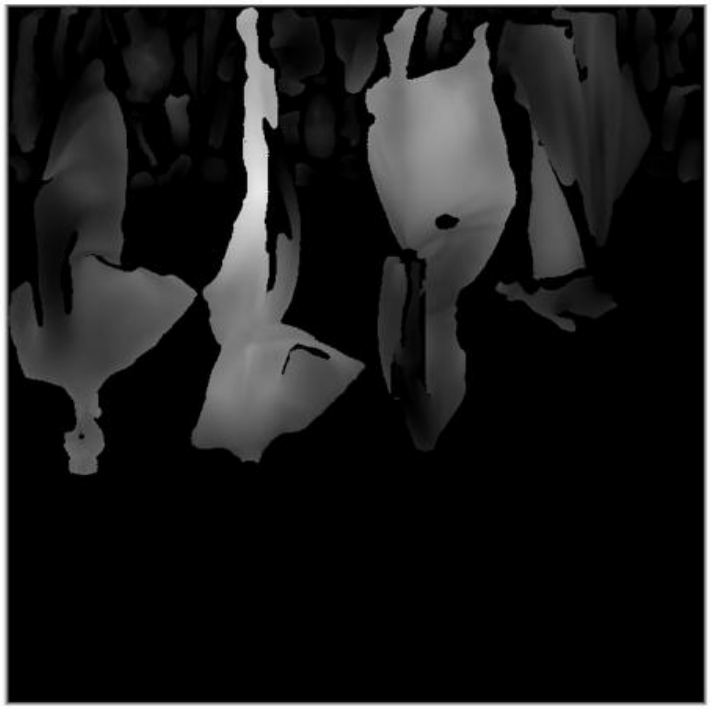
- Basic steps
  - Normal-based projection, frame packing, and frame padding
- Normal-based projection
  - Organize the points with similar normal into a patch
  - Project each patch to the 3D point cloud bounding box





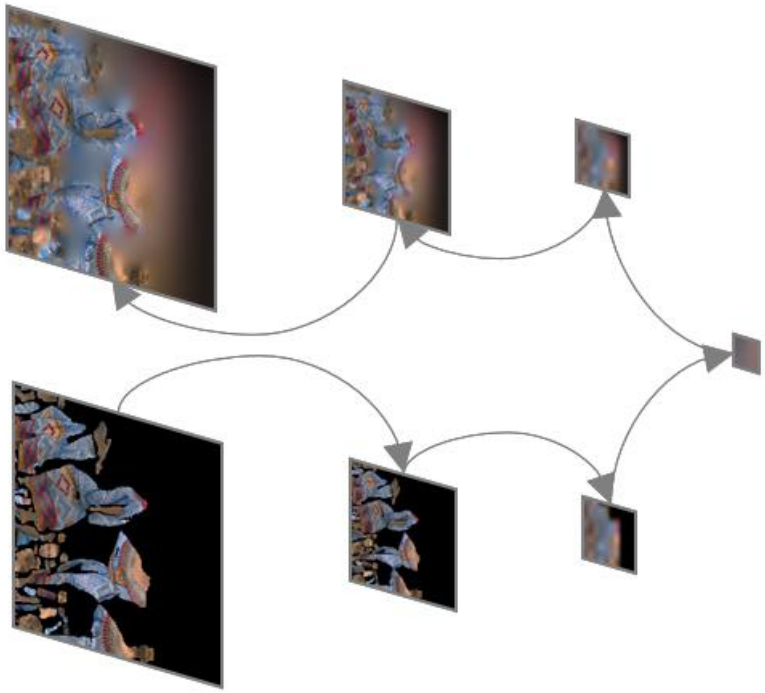
Video-based point cloud compression

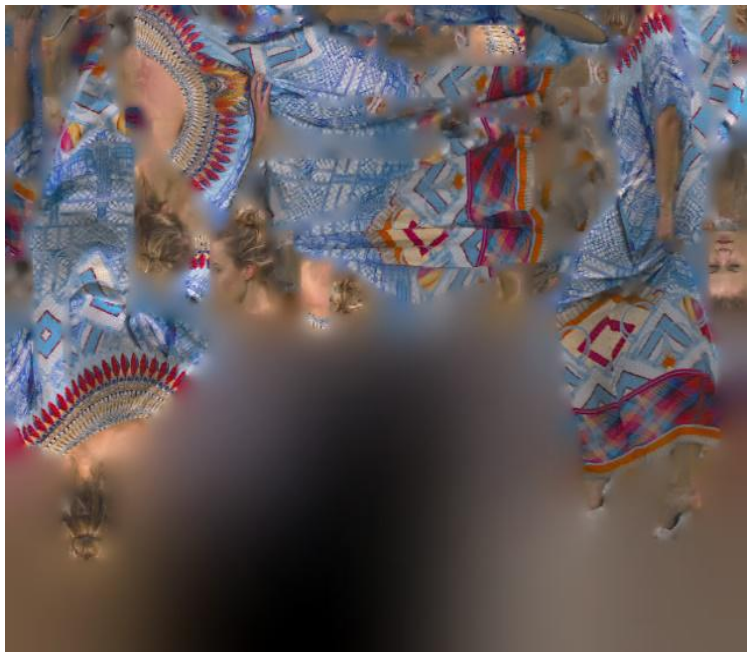
- Frame packing: pack the patches into frames
  - Exhaustive search empty space for the current patch
  - Patch rotation is supported
  - Introduced a lot of sharp edges



VPCC - Texture Padding

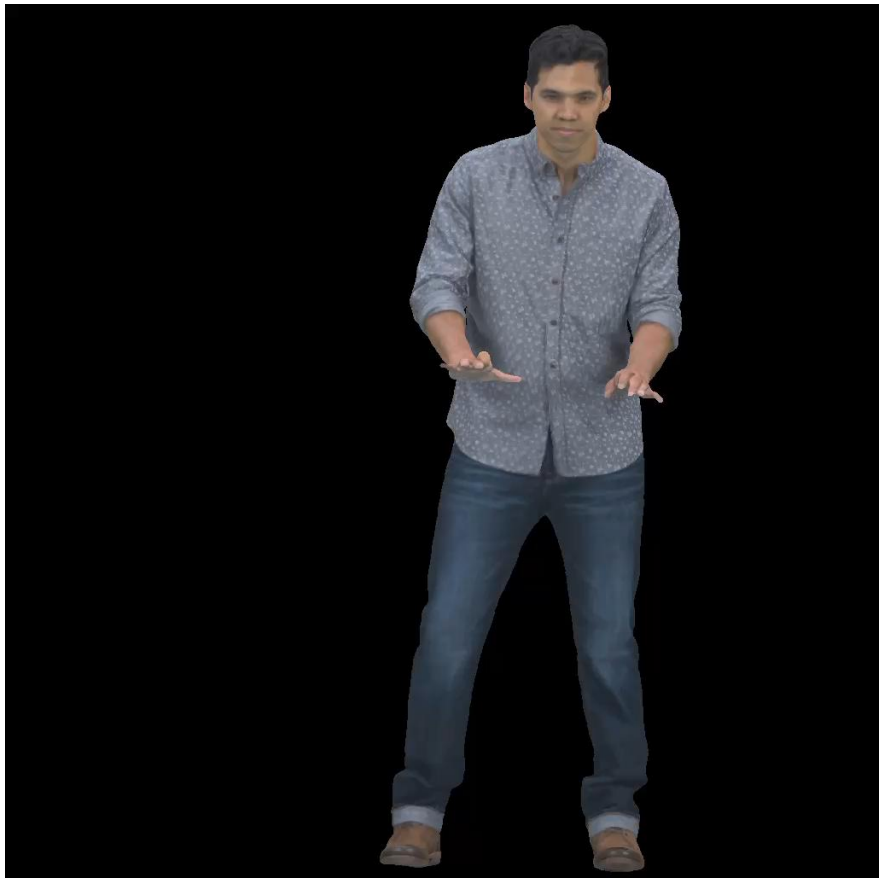
- Texture padding: a number of methods are proposed to minimize the bitrate of the unoccupied pixels
- Using push-pull algorithm as an example, like dilation



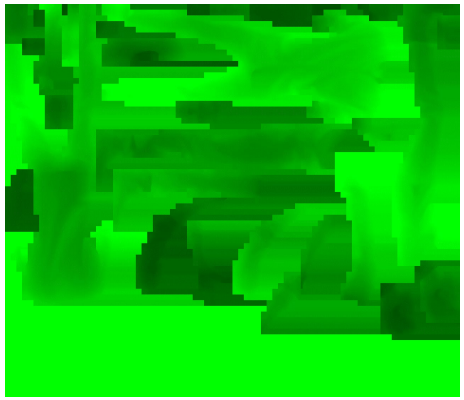


Video-based point cloud compression

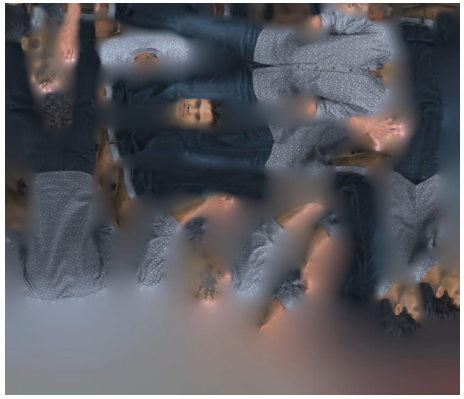
- Basic idea: project a point cloud to a 2-D video for an efficient compression



Geometry



Attribute



Outline

- Short Self Intro
- Research Motivation and Highlights
- Sparse Conv Engine Based PCC
  - PU-Dense: point cloud upscaling work (T-IP)
  - Compression artifacts removal (T-MM)
- Video based PCC
  - Advanced 3D motion for VPCC (T-IP)
  - Depth Field Denoising (IJCV)
- Summary

PU-Dense

PU-Dense: Sparse Tensor-based Point Cloud Geometry Upsampling

Anique Akhtar, *Student Member, IEEE,* Zhu Li, *Senior Member, IEEE,*
Geert Van der Auwera, *Senior Member, IEEE,* Li Li, *Member, IEEE,* Jianle Chen, *Senior Member, IEEE,*

Project Page:

https://aniqueakhtar.github.io/publications/PU-Dense/

Code:

https://github.com/aniqueakhtar/PointCloudUpsampling

Point Cloud Upsampling

- A very relevant problem:
  - lack of scan line density from LiDAR
  - undersampling of the mesh
  - zoom in for more details
- Summary of current SOTA:
  - **PU-Net** [17]: Based on PointNet++
  - **EC-Net** [18]: An edge-aware model by adding point-to-edge distance loss. However, have to manually label the dataset for training.
  - **PPPU** or **3PU** [19]: Patch-based upsampling. Employs a cascaded 2x upsampling networks.
  - **PU-GAN** [20]: Implemented a GAN.
  - **PUGeo-Net** [21]: Geometric-centric approach.
  - **PU-GCN** [22]: Multi-level feature extraction using an inception-based graph convolutional network. They employ shuffling rather than duplicating features to expand the feature space for upsampling.
  - **Others: [23~25]**

Main Issues

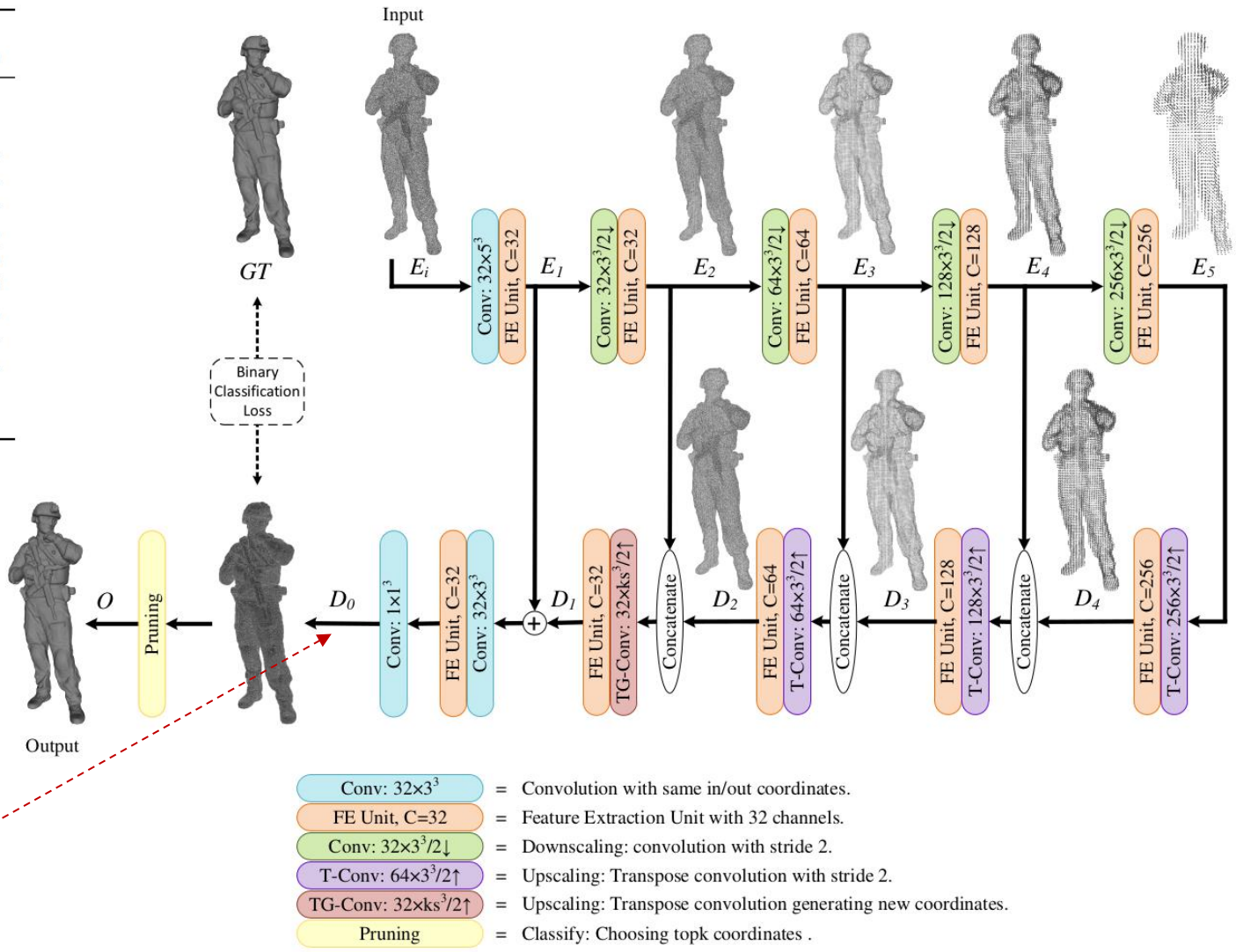
- Backbone network limitations (PointNet based)
  - Patch-based solutions, usually limited to a fix and small number of points input (between few hundreds to few thousands.)
  - Computationally intensive and memory hogging loss functions (e.g. Chamfer Distance).
  - Shallower networks with smaller receptive fields that limit their discriminative power and learning ability.
  - Extensive pre and post processing required. (Farthest point sampling and kNN search)
- Performance and robustness still lagging:
  - Not able to have deep architecture and large receptive field
  - Cannot handle large scale data set like 8i with > 1M points
  - Overfitting with PointNet

PU-Dense: Point Cloud Upsampling

- Sparse Convolution Back Bone (Minkowski Engine):
  - A fully convolutional geometry upsampling network that is translation invariant and has a variable input size.
  - Novel Feature Embedding (FE) with Inception-Residual Block (IRB) and a 3D Dilated Pyramid Block (3D-DPB)
  - Much larger network with more trainable network weights
- New Loss Function:
  - Employs memory efficient binary voxel classification / cross-entropy loss instead of CD
- Memory efficiency:
  - allows processing of millions of points per inference time.
- Robustness:
  - Can generalize to different datasets. It doesn't just work on synthetic point clouds but can also work for real-world scanned LiDAR based datasets as well as dense photo-realistic point clouds.
  - Robust against noise. Faster inference time.

PU-Dense Architecture

Tensor	Size of coordinates (C)	Size of features (F)
GT	830,397	1
E_i	207,599	1
E_1	207,599	32
E_2	139,244	32
E_3	52,612	64
E_4	14,440	128
E_5	3,623	256
D_4	14,440	256
D_3	52,612	128
D_2	139,244	64
D_1	4,379,676	32
D_0	4,379,676	1
O	830,397	1



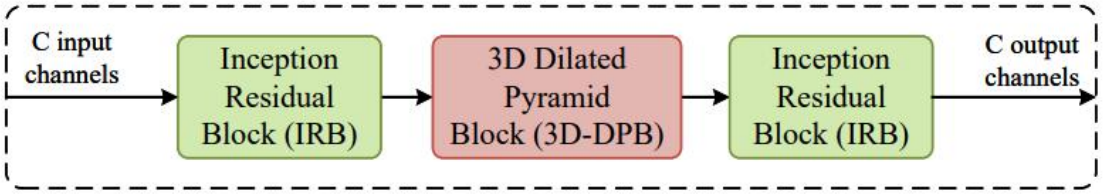
D_0: voxel occupancy prob

PU-Dense Architecture

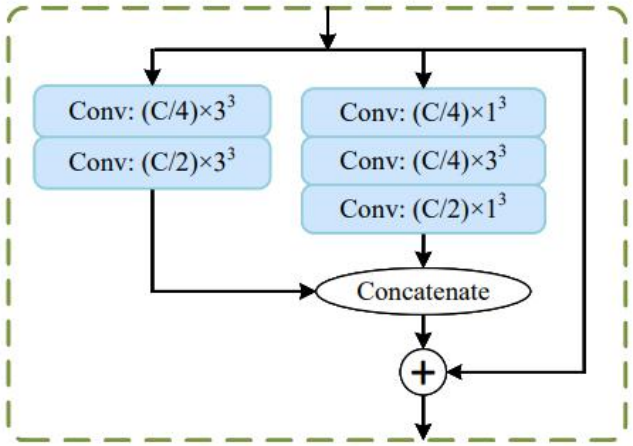
- Overall an U-net like structure
  - Voxilzed point cloud representation : limited variations.
  - 3 Downscaling with stride
  - increasing feature dimension which encodes occupancy for 2^3, 4^3 ,8^3 sized cubes.
  - Novel Feature Embedding (more details later)
  - Decoding into an occupancy prob function for each voxel location via TG-Conv (Transpose Generative Conv) layer
- Loss function:
  - instead of Chamfer loss and other similar distance based, we use occupancy prob loss
  - Binary Cross Entropy (BCE) : this is the key

Feature Embedding (FE) Unit

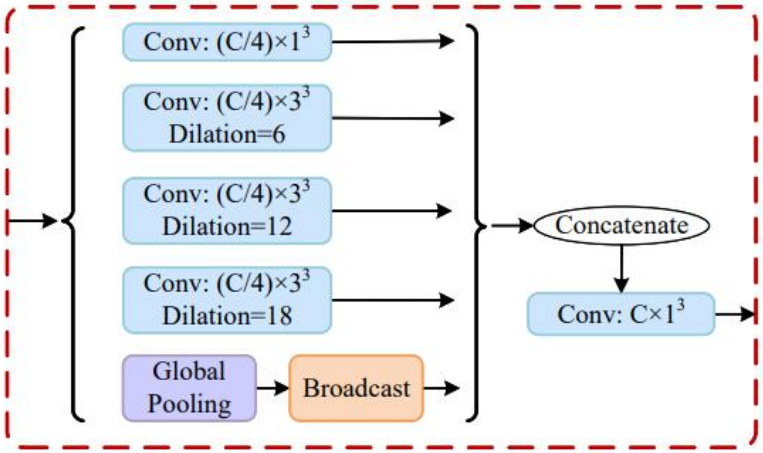
- FE Units.
  - IRB: similar to the inception in image domain, variable kernel size
  - 3D DPB: use dilation in kernel to improve receptive field size



Feature Extraction Unit



Inception-Residual Block (IRB)



3D Dilated Pyramid Block (3D-DPB)

Data Set

- Training from ShapeNet, testing on 8i and Technicolor

(i) Longdress (ii) Loot (iii) Red and black (iv) Soldier

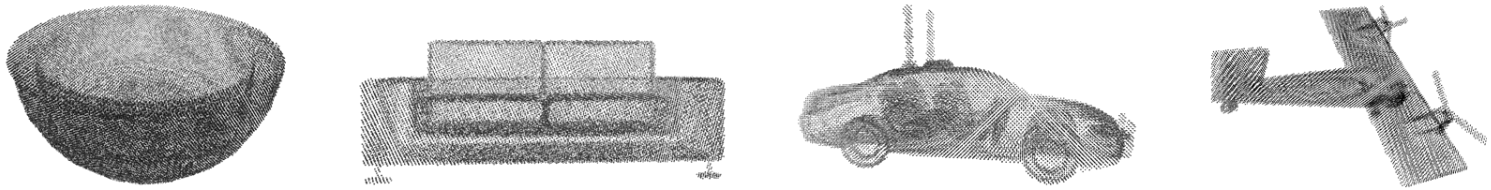
(a) 8iVFB v2

(i) Boxer (ii) Thaidancer

(b) 8iVSLF

(i) Queen

(c) Technicolor



(d) ShapeNet

Performance

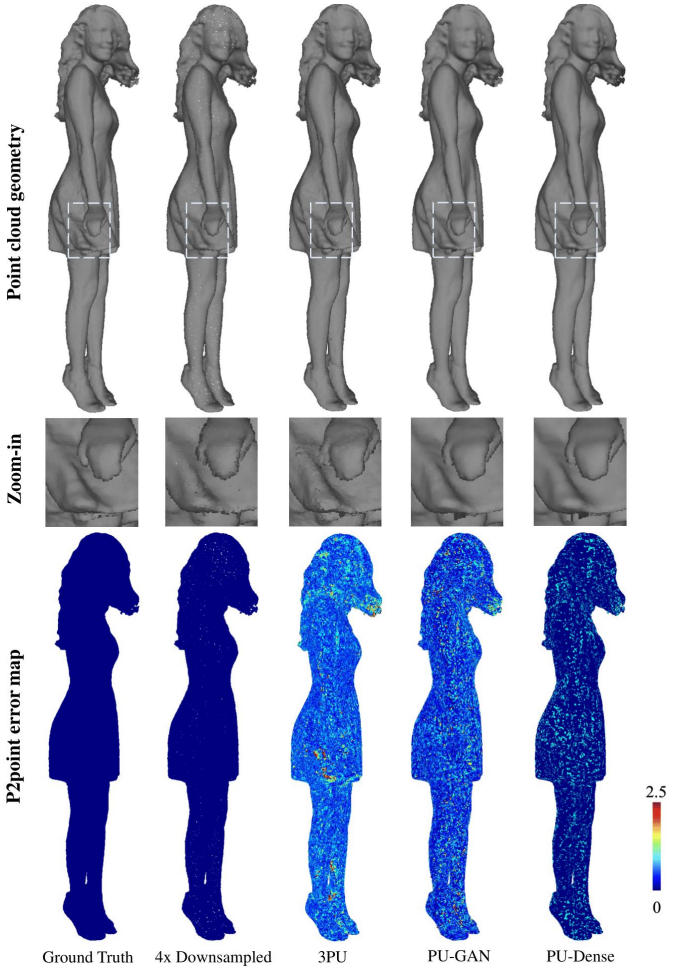
- Point CloudUpsampling
  - very significant performance gains over previous SOTA

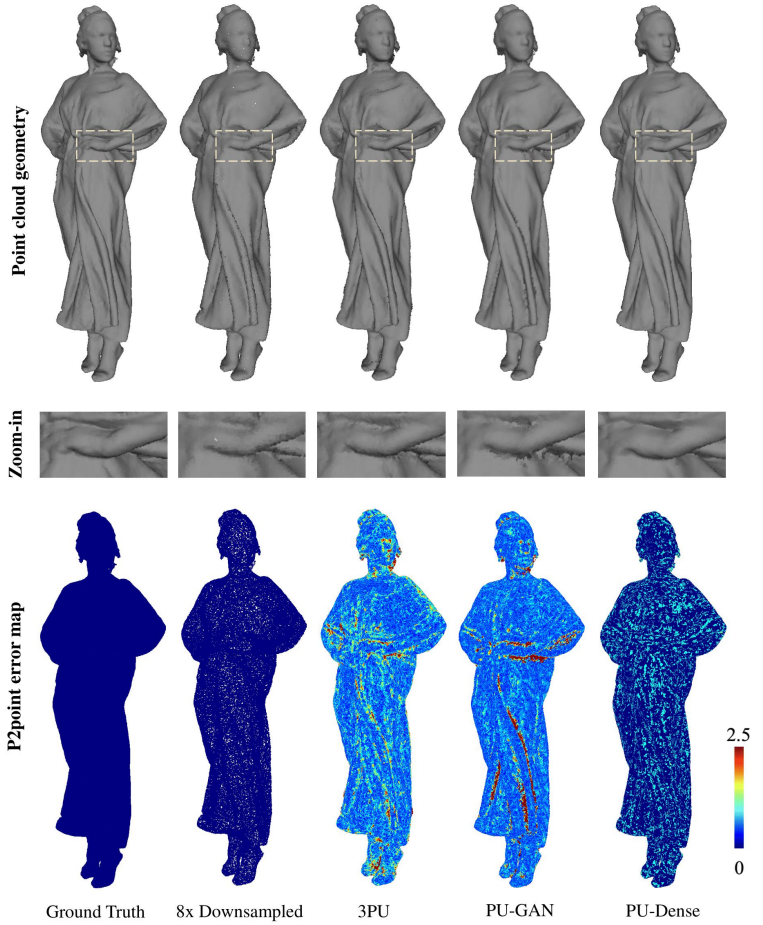
TABLE II
EXTENDED COMPARATIVE RESULTS (CD (10^{-2}) AND PSNR).

Dataset	Upsampling Method	4x		8x	
		CD (10^{-2}) ↓	MSE PSNR (dB) ↑	CD (10^{-2}) ↓	MSE PSNR (dB) ↑
ShapeNet	Downsampled PC	108.18	64.63	199.94	61.96
	3PU	76.36	68.65	149.20	65.37
	PU-GAN	49.41	70.64	174.58	64.88
	PU-Dense	**18.82**	**75.24**	**30.52**	**73.11**
8iVFB	Downsampled PC	114.63	64.38	222.91	61.49
	3PU	67.04	69.41	105.43	66.83
	PU-GAN	45.60	70.92	117.66	66.19
	PU-Dense	**19.38**	**75.05**	**33.18**	**72.57**
8iVSLF	Downsampled PC	286.67	73.17	600.34	70.00
	3PU	204.92	76.98	368.63	74.78
	PU-GAN	156.94	77.18	231.39	75.34
	PU-Dense	**135.41**	**78.92**	**202.82**	**76.79**
Queen	Downsampled PC	106.69	64.69	196.46	62.04
	3PU	57.13	70.19	90.90	67.55
	PU-GAN	41.67	71.43	110.42	66.36
	PU-Dense	**15.76**	**75.93**	**25.45**	**73.76**

Subjective Results

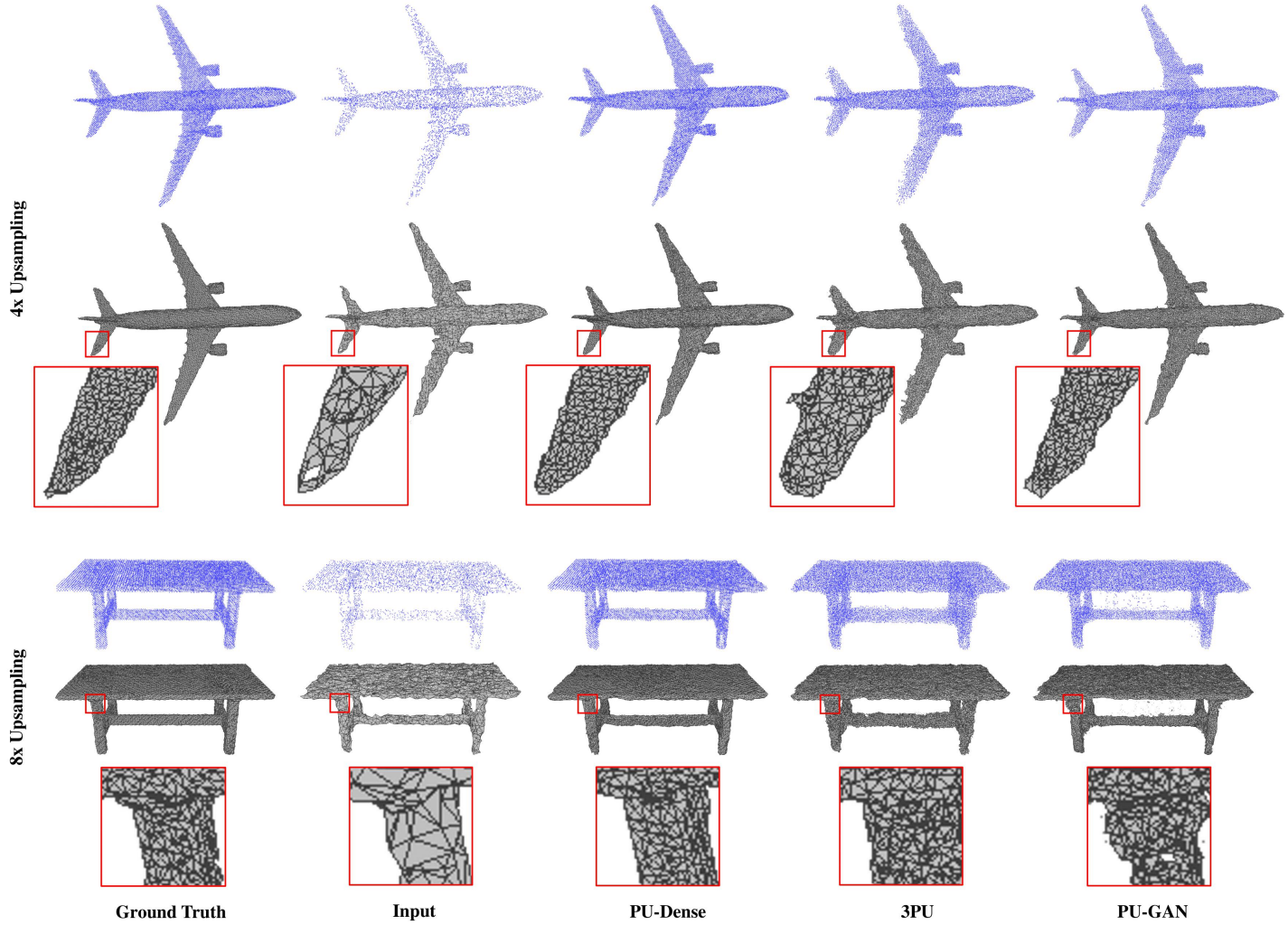
- 8i Sequences 4X upsacling





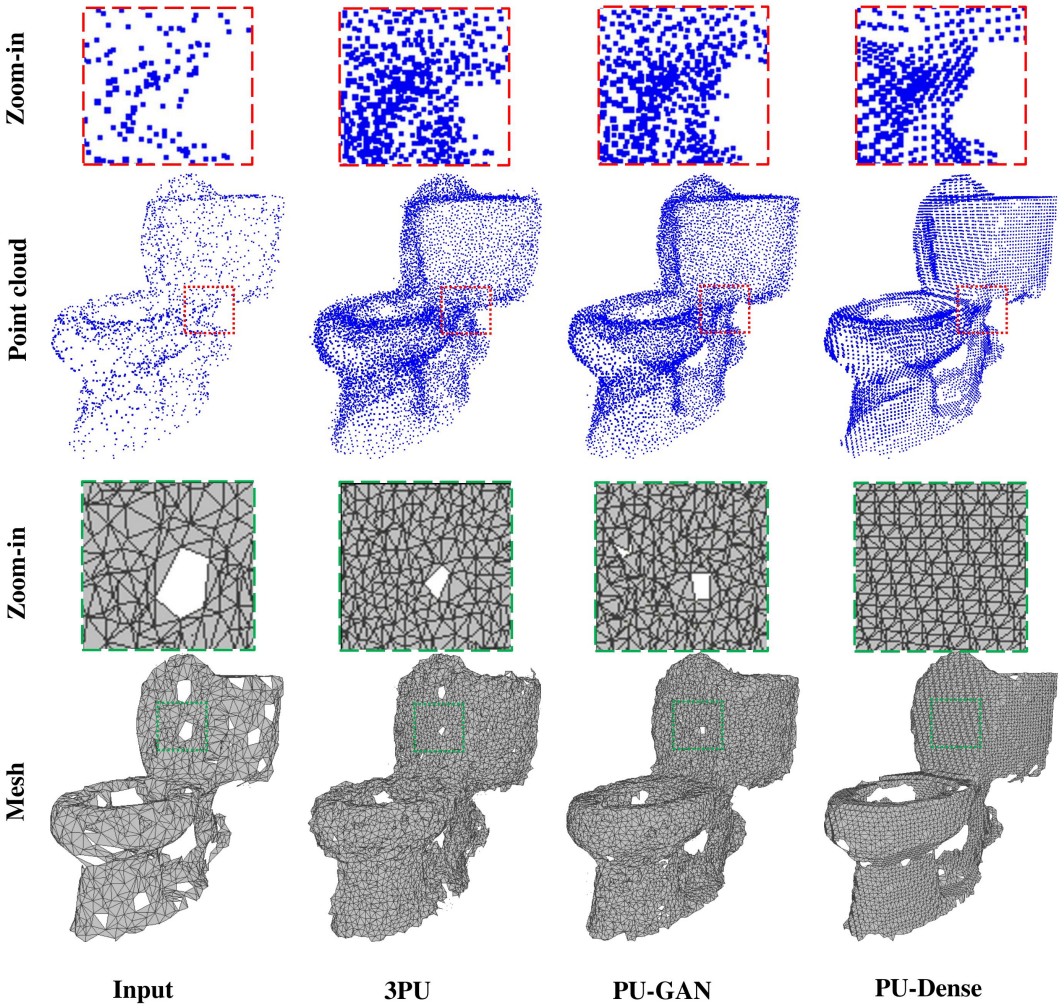
Mesh Quality- Synthetic Objects

- Mesh quality from re-genration: 4x and 8x upscaled



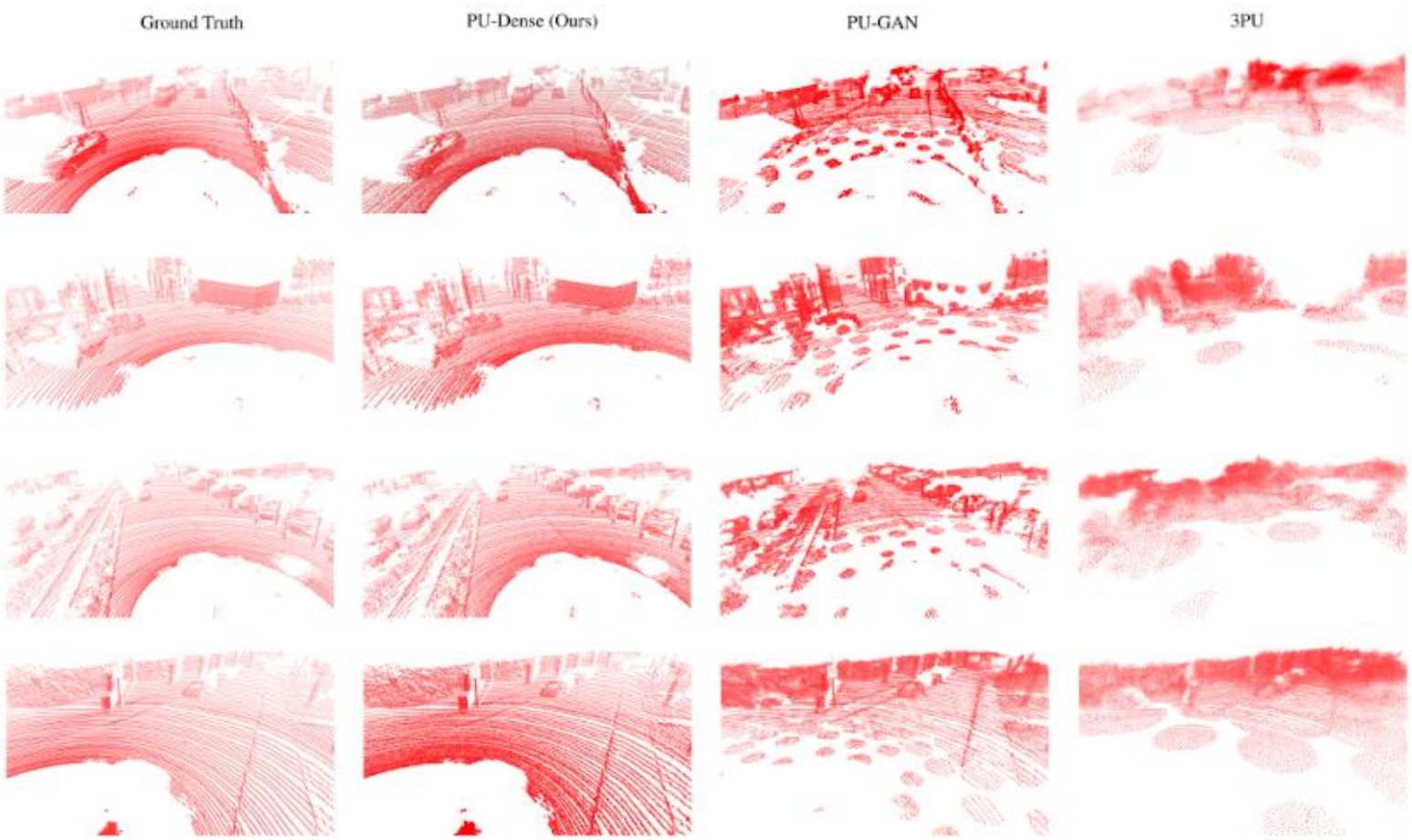
Mesh Quality - Real World

- Mesh quality from re-genration: 4x upscaled points



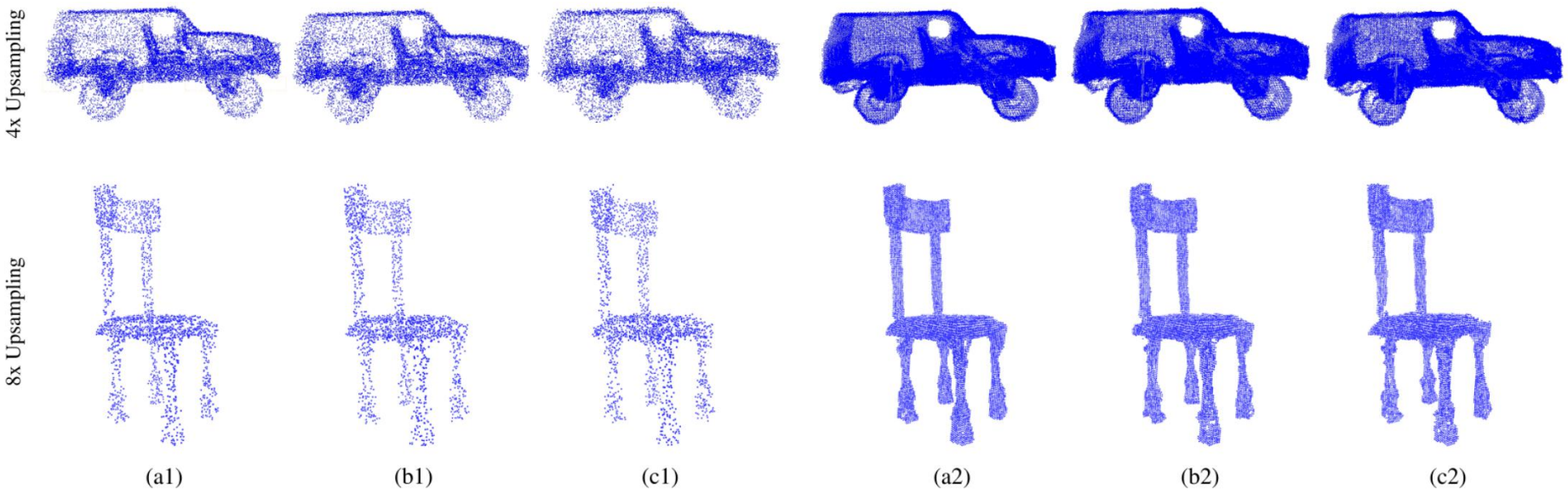
LiDAR data results

- KITTI: 4x upscaling



Robustness to Noise

- Adding noises to the low resolution point cloud:



Visual results for PU-Dense on noisy data for 4x and 8x upsampling.
Left: (a1), (b1), and (c1) are the sparse inputs with 0%, 1%, and 2% Gaussian noise, respectively.
Right: (a2), (b2), and (c2) are the upsampled results from (a1), (b1), and (c1) respectively.

Complexity

- Much more network parameters
- But faster inferences

TABLE IV
QUANTITATIVE COMPARISON: AVERAGE EVALUATION TIME PER POINT CLOUD FOR 4X UPSAMPLING.

Dataset	Upsampling Method	Computation time (min)
8iVFB	3PU	27.49
	PU-GAN	24.78
	PU-Dense	**00.79**

TABLE V
QUANTITATIVE COMPARISON: NUMBER OF TRAINABLE PARAMETERS.

Upsampling Method	Trainable parameters
3PU	152,054
PU-GAN	541,601
PU-Dense	**13,172,441**

References

[1] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D shapenets: A deep representation for volumetric shapes," in Proceedings of the IEEE conference on computer vision and pattern recognition,2015, pp. 1912–1920.

[2] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "Segcloud: Semantic segmentation of 3d point clouds," in 2017 international conference on 3D vision (3DV). IEEE, 2017, pp. 537–547.

[3] A. Kanezaki, Y. Matsushita, and Y. Nishida, "Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5010–5019.

[4] H. Su, V. Jampani, D. Sun, S. Maji, E. Kalogerakis, M.-H. Yang, and J. Kautz, "Splatnet: Sparse lattice networks for point cloud processing," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2530–2539.

[5] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12 697–12 705.

[6] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 945–953.

[7] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652–660.

[8] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," arXiv preprint arXiv:1706.02413, 2017.

[9] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on χ-transformed points," in Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018, pp. 828–838.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[11] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," Acm Transactions On Graphics (tog), vol. 38, no. 5, pp. 1–12, 2019.

[12] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4558–4567.

[13] P.-S. Wang, Y. Liu, Y.-X. Guo, C.-Y. Sun, and X. Tong, "O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis," ACM Transactions on Graphics (TOG), vol. 36, no. 4, pp. 1–11, 2017.

[14] B. Graham and L. van der Maaten, "Submanifold sparse convolutional networks," arXiv preprint arXiv:1706.01307, 2017.

[15] B. Graham, M. Engelcke, and L. Van Der Maaten, "3D Semantic Segmentation with Submanifold Sparse Convolutional Networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9224–9232.

[16] C. Choy, J. Gwak, and S. Savarese, "4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3075–3084.

[17]L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "PU-Net: Point Cloud Upsampling Network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2790–2799.

[18] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "EC-Net: an edge-aware point set consolidation network," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 386–402.

[19] W. Yifan, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung, "Patch-based Progressive 3D Point Set Upsampling," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5958–5967.

[20] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "PU-GAN: a point cloud upsampling adversarial network," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 7203–7212.

[21] Y. Qian, J. Hou, S. Kwong, and Y. He, "PUGeo-Net: A Geometry-Centric Network for 3D Point Cloud Upsampling," in European Conference on Computer Vision. Springer, 2020, pp. 752–769.

[22] G. Qian, A. Abualshour, G. Li, A. Thabet, and B. Ghanem, "PU-GCN: Point cloud upsampling using graph convolutional networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 11 683–11 692.

[23] R. Li, X. Li, P.-A. Heng, and C.-W. Fu, "Point cloud upsampling via disentangled refinement," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 344–353.

[24] S. Ye, D. Chen, S. Han, Z. Wan, and J. Liao, "Meta-pu: An arbitrary-scale upsampling network for point cloud," IEEE Transactions on Visualization and Computer Graphics, 2021.

[25] Y. Qian, J. Hou, S. Kwong, and Y. He, "Deep magnification-flexible upsampling over 3d point clouds," IEEE Transactions on Image Processing, vol. 30, pp. 8354–8367, 2021.

PU-Dense Summary

- Loss function: Point Cloud Upsampling is about voxel occupancy prediction, switching from distance based loss to occupancy prob loss is the main break through
- Network backbone: PointNet and variations are limited in efficiency and performance, sparse conv network backbones like Minkowski Engine allows for much larger data set and deeper network, lead to significantly better performance
- New SOTA: This sparse conv backbone + occupancy prob loss framework gives us new performance in a variety of problems, including, upsampling, denoising (next topic), and inter-prediction coding*.

*Anique Akhtar, Zhu Li, Geert Van der Auwera, Jianle Chen, "Dynamic Point Cloud Interpolation", in *IEEE Int'l Conf on Audio, Speech and Signal Processing* (ICASSP), 2022.

Outline

- Short Self Intro
- Research Motivation and Highlights
- Sparse Conv Engine Based PCC
  - PU-Dense: point cloud upscaling work (T-IP)
  - Compression artifacts removal (T-MM)
- Video based PCC
  - Advanced 3D motion for VPCC (T-IP)
  - Depth Field Denoising (IJCV)
- Summary

Compression Artifacts Removal

Video-based Point Cloud Compression Artifact Removal

Anique Akhtar, *Student Member, IEEE,* Wen Gao, Li Li, *Member, IEEE,* Zhu Li, *Senior Member, IEEE*
Wei Jia, *Student Member, IEEE,* Shan Liu, *Senior Member, IEEE*

```
@ARTICLE{8786879,
  author={Li, Li and Li, Zhu and Zakharchenko, Vladyslav
and Chen, Jianle and Li, Houqiang},
  journal={IEEE Transactions on Image Processing},
  title={Advanced 3D Motion Prediction for Video-Based
Dynamic Point Cloud Compression},
  year={2020},
  volume={29},
  number={},
  pages={289-302},
  doi={10.1109/TIP.2019.2931621}}
```

Point Cloud Compression Artifact Removal

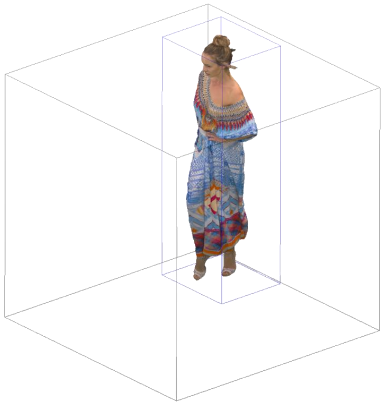
- Distortion from VPCC based coding
  - V-PCC is the current state-of-the-art in dynamic point cloud compression.
  - This work proposes the first deep-learning-based geometry artifact removal algorithm for the V-PCC standard for dynamic point clouds.
  - Ours is a pioneering work in V-PCC artifact removal without any extra bandwidth to the V-PCC standard.

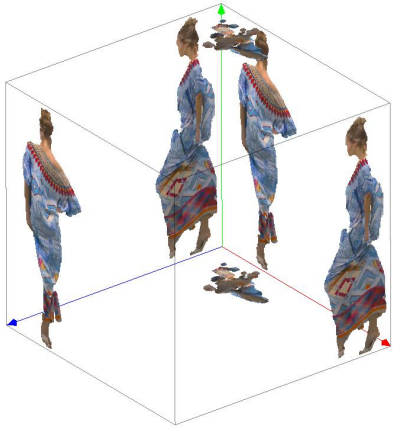
(a) High-bit rate (b) Low-bit rate

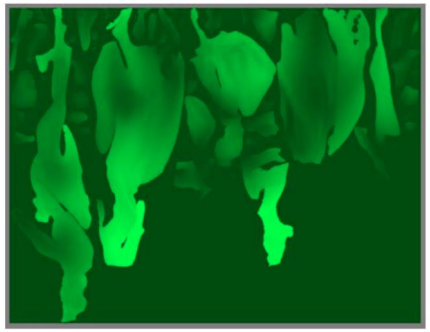
Blocking effects in a point cloud coded at different bitrates using V-PCC encoding.

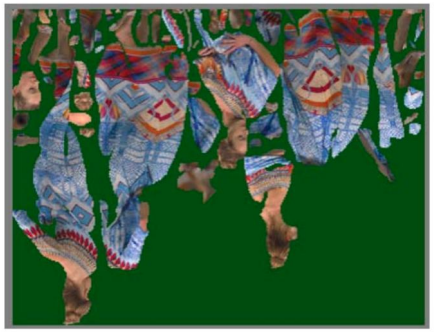
Geometry Distortion in V-PCC

- V-PCC is based on the projection of the point cloud patches to 2D planes and encoding the sequence as 2D texture and geometry patch sequences.
- Afterward, these projected video frames are encoded by leveraging video compression techniques. Since these compression techniques are lossy, compression artifacts are often introduced due to quantization noise affecting the point cloud geometry.



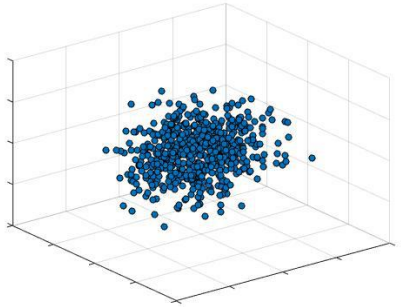




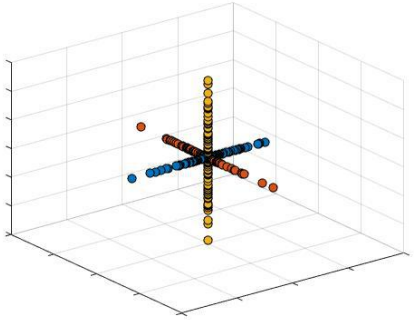


Loss Function Design

- We observe that the geometry distortion of the V-PCC reconstructed point cloud exists only in the direction of the V-PCC projection.
- We exploit this prior knowledge to learn both the direction and level of quantization noise by limiting the degree of freedom of the learned noise.

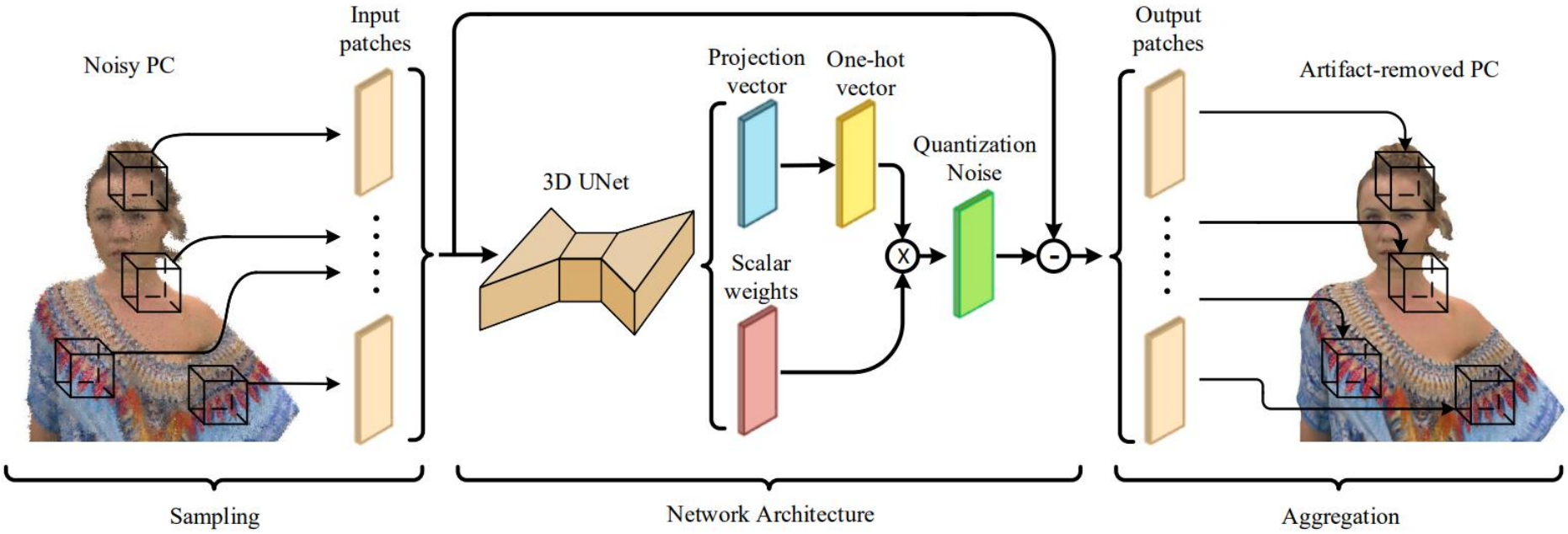


(a) Gaussian noise



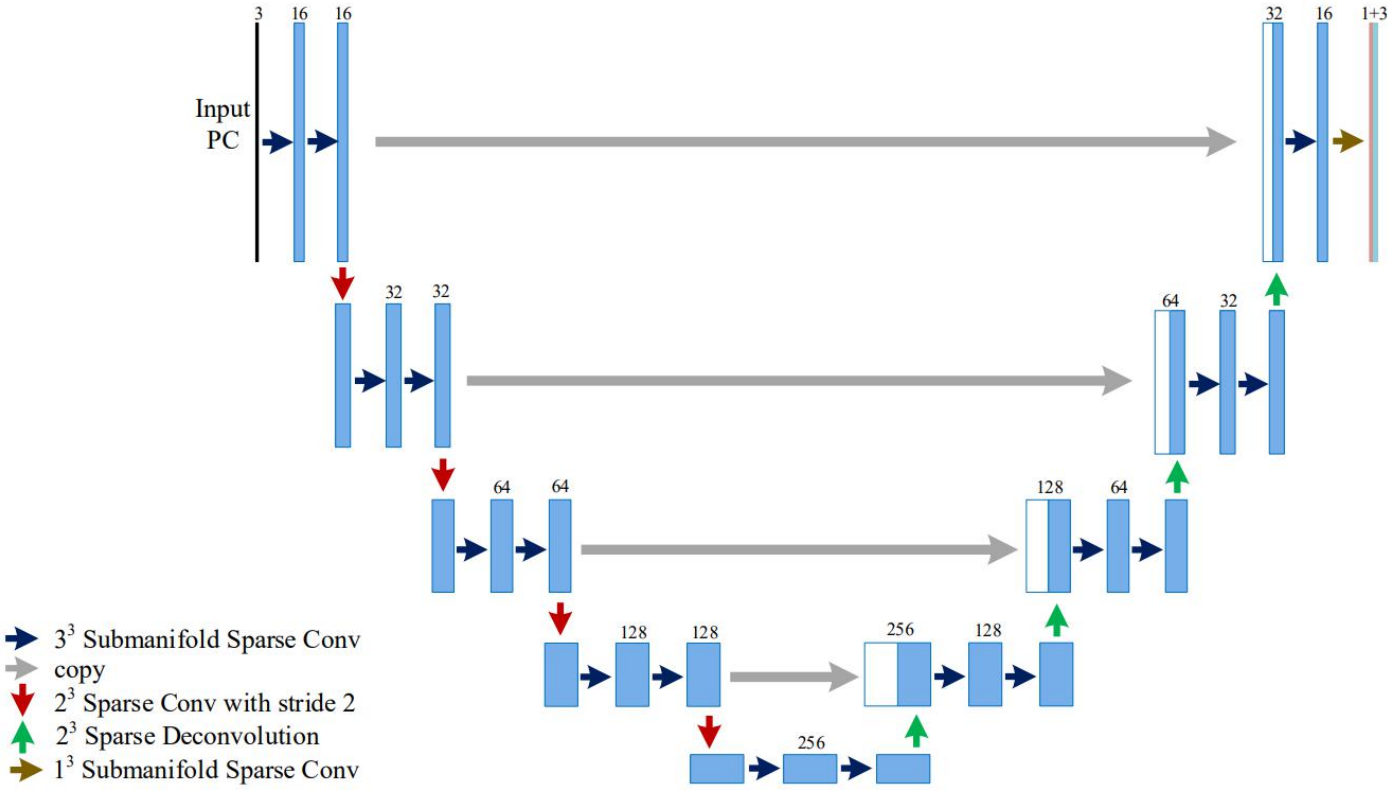
(b) V-PCC quantization noise.

Network Architecture



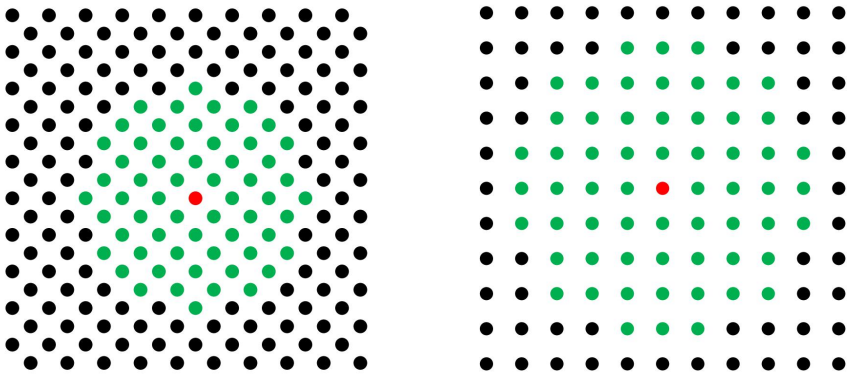
$$L_{CD}(P_O, P_G) = \sum_{x \in P_O} \min_{y \in P_G} ||x - y||_2^2 + \sum_{y \in P_G} \min_{x \in P_O} ||x - y||_2^2$$

UNet Like Network Architecture



Sampling & Aggregation

We employ farthest point sampling (FPS) to sample points on the noisy point cloud and then extract cube patches around the sampled points.



(a) Dense Surface. (b) Sparse Surface.

Fig. Patch correspondence mismatch problem for k = 61. The k-NN search covers a smaller surface on a denser point cloud as compared to a sparser point cloud.

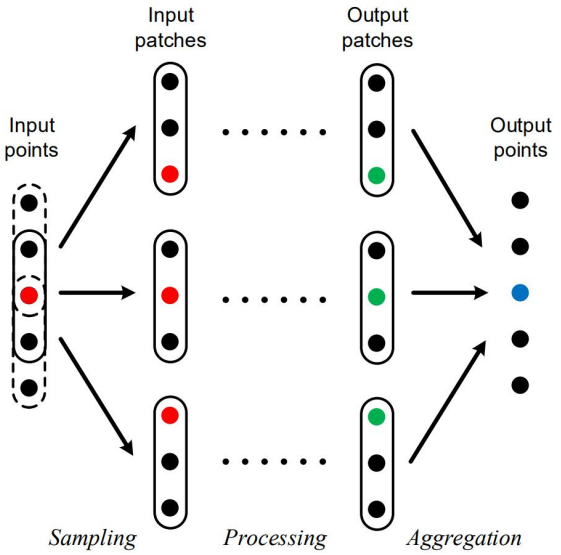


Fig. An example sampling and aggregation scheme.

Artifact Removal: Simulation Results

Bitrate label	Actual bitrate
br1	0.01866 bpp
br2	0.01632 bpp
br3	0.01502 bpp

Table 1. Bitrates used in simulation

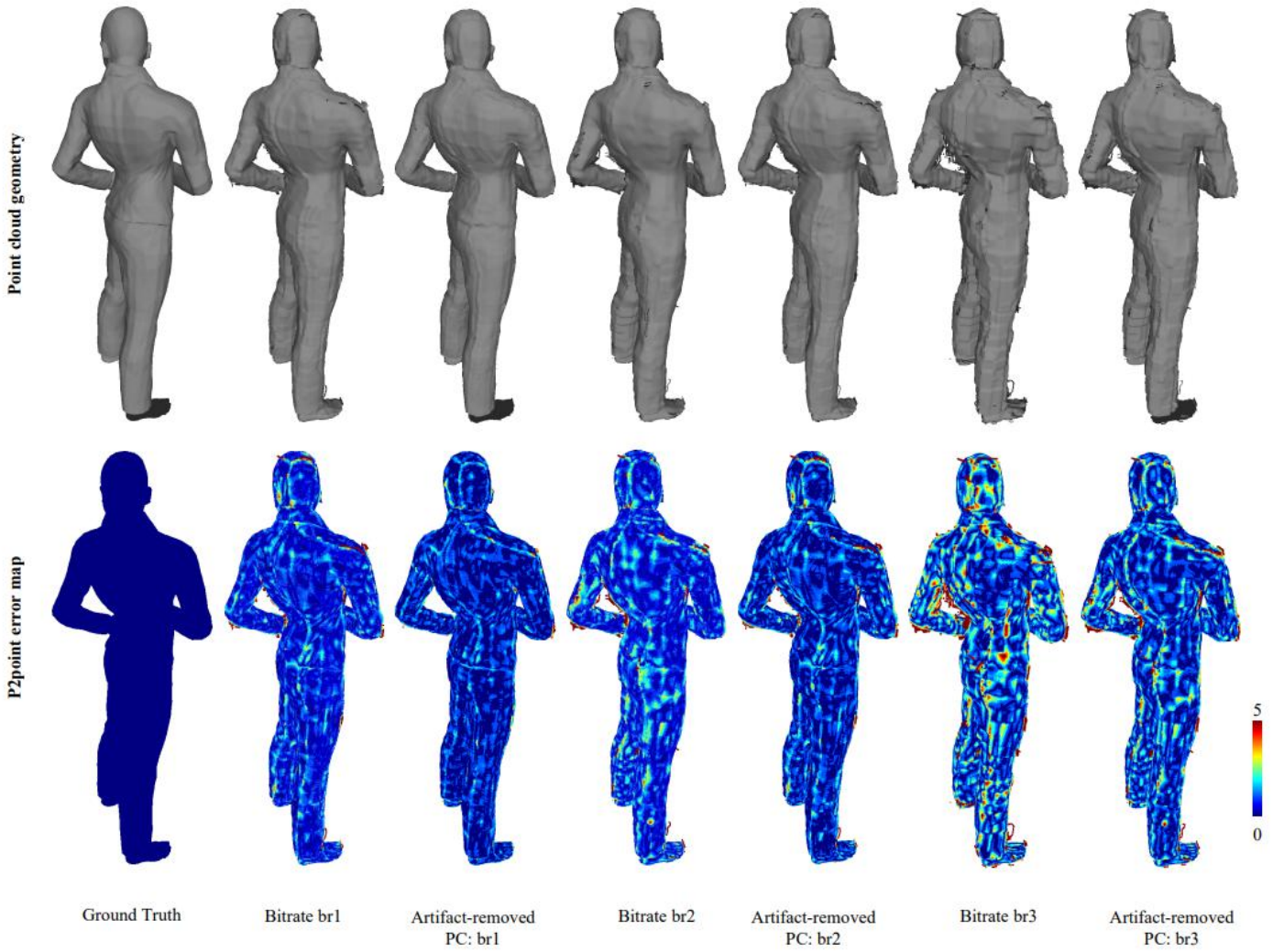
	PSNR (dB)		Hausdorff PSNR (dB)	
Bitrate	Noisy PC	Cleaned PC	Noisy PC	Cleaned PC
br3	59.62	60.47	37.02	37.21
br2	61.84	62.36	39.58	39.64
br1	64.20	64.53	41.15	41.20
BD-rate savings:			11.3 %	

Table 2. Average PSNR results

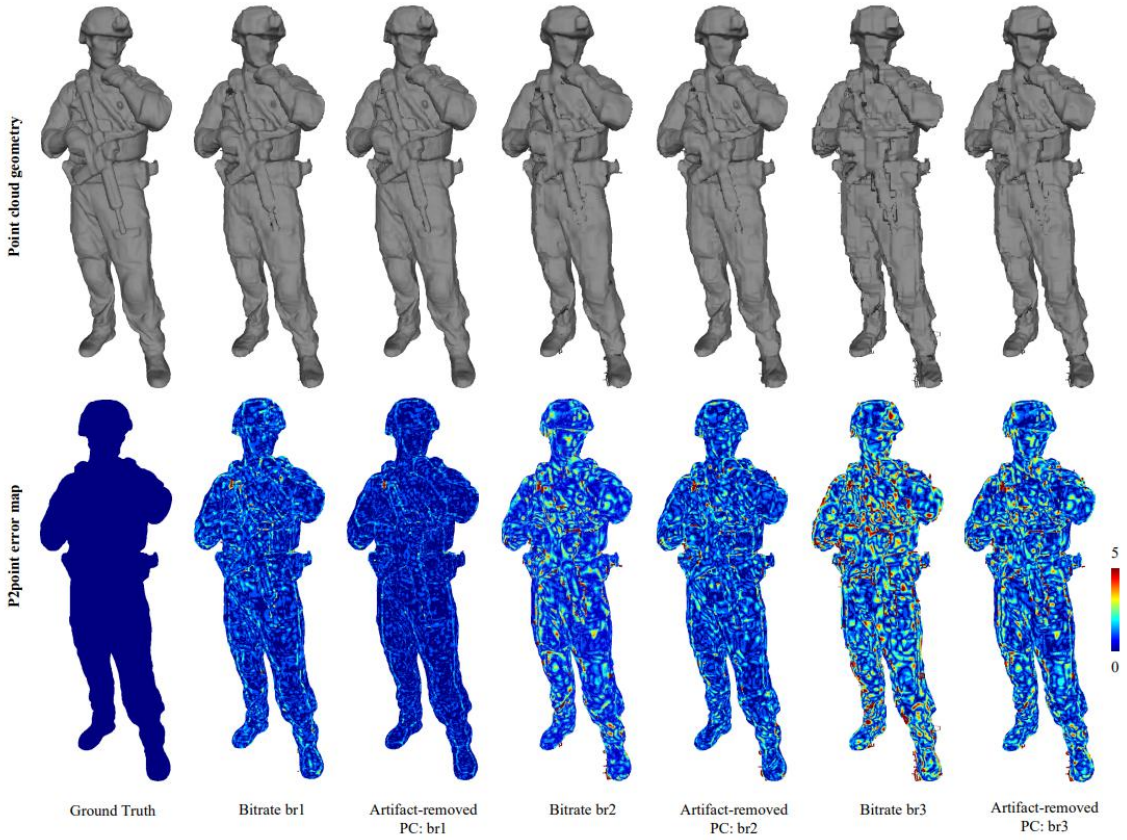
		PSNR (dB)		
Test PC	Bitrate	Noisy PC	Cleaned PC	Improvement
Queen	br3	60.23	60.79	0.56
	br2	62.35	62.90	0.55
	br1	64.94	65.21	0.27
RedAndBlack	br3	59.44	60.38	0.94
	br2	61.62	62.18	0.56
	br1	63.90	64.27	0.37
Soldier	br3	59.20	60.25	1.05
	br2	61.53	62.01	0.48
	br1	63.76	64.12	0.36

Table 3. Simulation results for each individual sequence

Artifact Removal: Visual Results



Artifact Removal: Visual Results



Artifact Removal: Summary

- This work presents a first-of-its-kind deep learning-based point cloud geometry compression artifact removal framework for V-PCC encoded dynamic point clouds.
- We leverage the prior knowledge that during V-PCC, the quantization noise is introduced only in the direction of the point cloud projection.
- We employ a 3D sparse convolutional neural network to learn both the direction and the magnitude of geometry quantization noise.
- To make our work scalable, we propose a cube-centered neighborhood extraction scheme with a sampling and aggregation method to extract small neighborhood patches from the original point cloud.
- Experimental results show that our method considerably improves the V-PCC reconstructed point cloud's geometry quality in both objective evaluations and visual comparisons.

Outline

- Short Self Intro
- Research Motivation and Highlights
- Sparse Conv Engine Based PCC
  - PU-Dense: point cloud upscaling work (T-IP)
  - Compression artifacts removal (T-MM)
- Video based PCC
  - Advanced 3D motion for VPCC (T-IP)
  - Depth Field Denoising (IJCV)
- Summary

Advanced 3D Motion for VPCC

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 29, 2020 289

Advanced 3D Motion Prediction for Video-Based Dynamic Point Cloud Compression

Li Li, *Member, IEEE*, Zhu Li, *Senior Member, IEEE*, Vladyslav Zakharchenko, *Member, IEEE*, Jianle Chen, *Senior Member, IEEE*, and Houqiang Li, *Senior Member, IEEE*


```
@ARTICLE{Li20TMM,
  author={Li, Li and Li, Zhu and Zakharchenko, Vladyslav and Chen,
Jianle and Li, Houqiang},
  journal={IEEE Transactions on Image Processing},
  title={Advanced 3D Motion Prediction for Video-Based Dynamic
Point Cloud Compression},
  year={2020},
  volume={29},
  number={},
  pages={289-302},
  doi={10.1109/TIP.2019.2931621}}
```

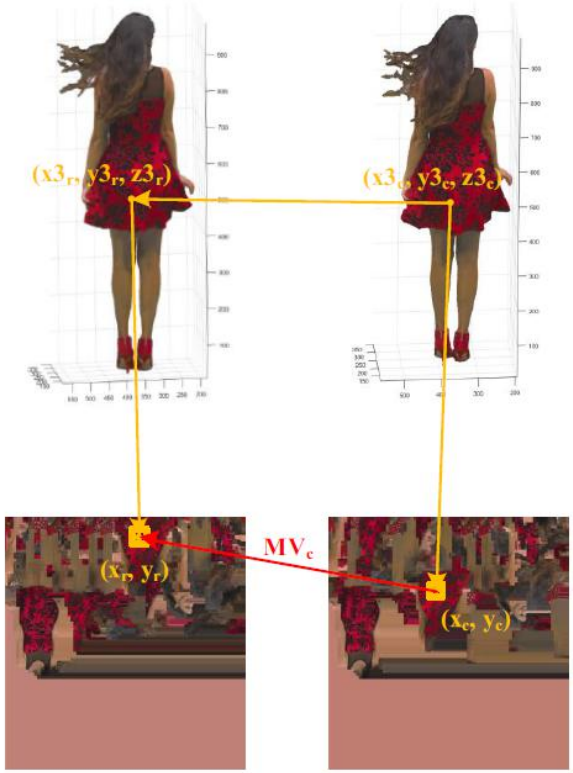
VPCC Motion Model

- The corresponding patches may be put in totally different positions in various frames (Green squares)
  - The current video codec may be unable to find a good motion vector for each block in this case
  - The geometry is encoded before the attribute, we can use the geometry to derive a better motion vector for attribute

General 3D to 2D motion model

- Given the 3D motion and the 3D to 2D correspondence, we can derive the 2D motion
  - g(), f(): 3D to 2D projection in reference and current frames

$$\boldsymbol{MV_c} = \boldsymbol{g}(x3_r, y3_r, z3_r) - \boldsymbol{f}(x3_c, y3_c, z3_c)$$

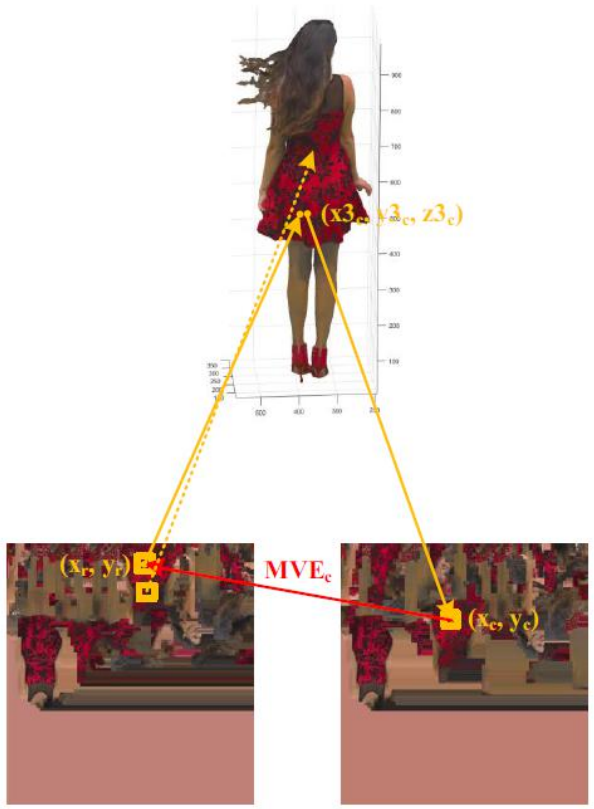


Geometry-based motion prediction

- In the V-PCC, we know the 3D-to-2D correspondence but do not know the 3D motion
- We assume the current frame and the reference frame will not change dramatically

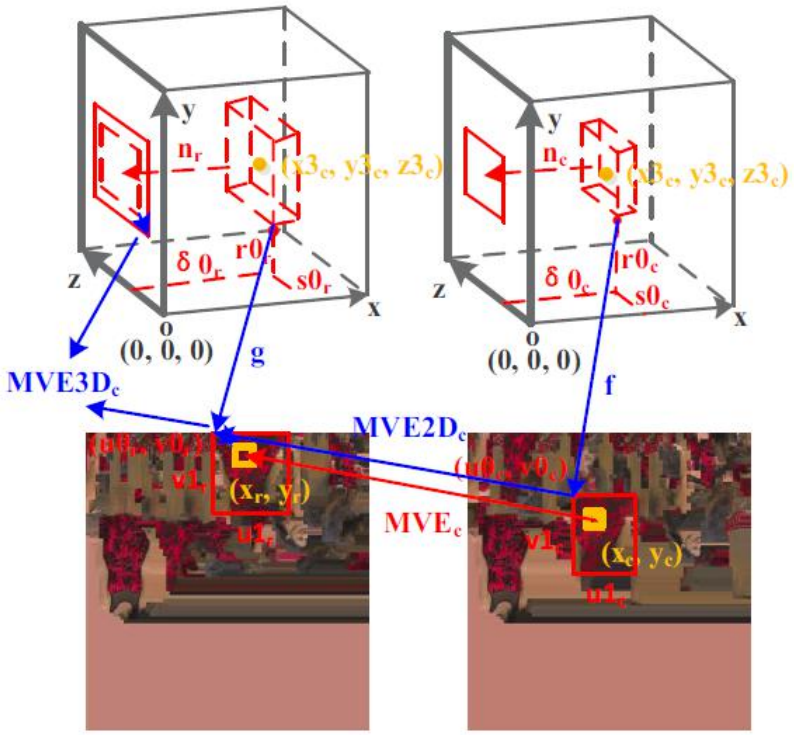
$$\boldsymbol{MVE_c} = \boldsymbol{g}(x3_c, y3_c, z3_c) - \boldsymbol{f}(x3_c, y3_c, z3_c)$$

- The problem is that (x3c,y3c,z3c) may not have a corresponding 2D point in the reference frame
  - We perform motion estimation which will increase the encoder and decoder complexity



Auxiliary information based motion prediction

- The previous method has the following two disadvantages
  - The high encoder and decoder complexity
  - It can only apply to the attribute
- The auxiliary information based motion prediction
  - The auxiliary information basically provides the coarse geometry
  - We use the 3D offset plus the 2D offset



Experiments setup

- The proposed algorithm is implemented in the V-PCC reference software and the corresponding HEVC reference software
- We test the all the dynamic point clouds defined in the common test condition including loot, redandblack, soldier, queen, longdress
- For the geometry, both point-to-point is point-to-plane are used
- For the attribute, the qualities of the luma, Cb, and Cr are considered

Experimental results on the overall scheme

- Overall scheme results

TABLE III
PERFORMANCE OF THE GEOMETRY-BASED MOTION PREDICTION COMPARED WITH THE V-PCC ANCHOR

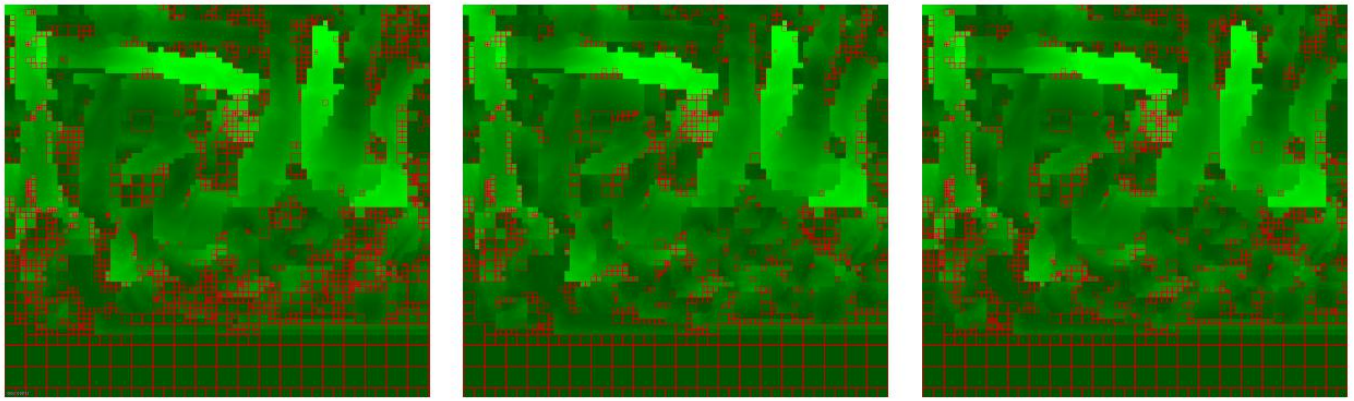
Test point cloud	Geom.BD-GeomRate D1	Geom.BD-GeomRate D2	Attr.BD-AttrRate Luma	Attr.BD-AttrRate Cb	Attr.BD-AttrRate Cr	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	0.0%	0.0%	−18.1%	−31.4%	−30.4%	−3.4%	−6.1%	−8.4%	−17.7%	−16.9%
RedAndBlack	0.0%	0.0%	−16.3%	−25.0%	−15.9%	−4.6%	−4.6%	−8.8%	−15.4%	−8.4%
Solider	0.0%	0.0%	−33.4%	−42.5%	−43.2%	−8.2%	−8.2%	−17.2%	−26.3%	−27.0%
Queen	0.0%	0.0%	−13.7%	−20.5%	−19.2%	−3.5%	−3.6%	−7.8%	−12.7%	−11.6%
LongDress	0.0%	0.0%	−9.8%	−13.5%	−12.3%	−3.7%	−3.7%	−6.4%	−9.5%	−8.4%
Avg.	0.0%	0.0%	−18.2%	−26.6%	−24.2%	−4.7%	−4.7%	−9.7%	−16.3%	−14.5%
Enc. time self	97%									
Dec. time self	98%									
Enc. time child	486%									
Dec. time child	337%									

TABLE IV
PERFORMANCE OF THE AUXILIARY-INFORMATION-BASED MOTION PREDICTION COMPARED WITH THE V-PCC ANCHOR UNDER THE NORMATIVE SOLUTION

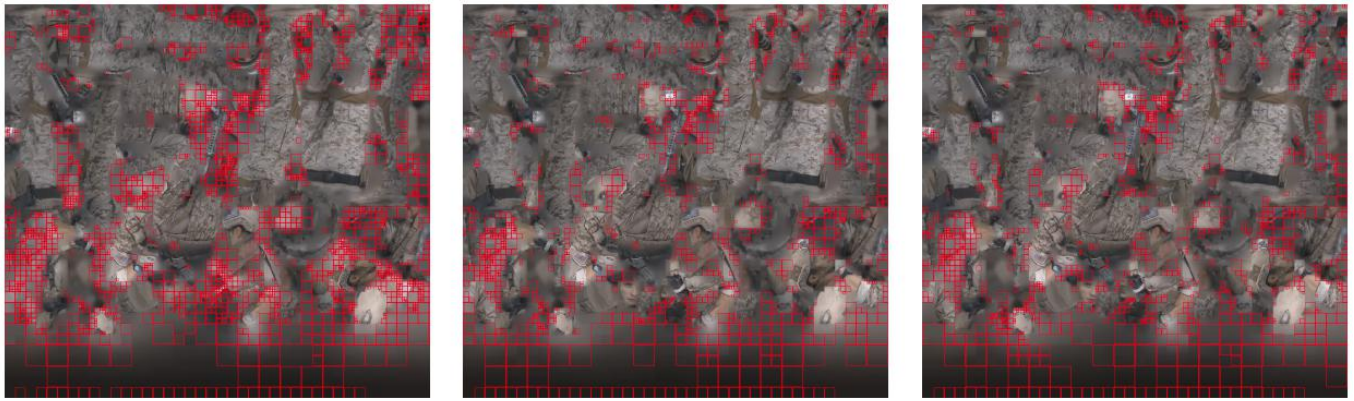
Test point cloud	Geom.BD-GeomRate D1	Geom.BD-GeomRate D2	Attr.BD-AttrRate Luma	Attr.BD-AttrRate Cb	Attr.BD-AttrRate Cr	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	−4.0%	−3.9%	−16.3%	−26.4%	−28.5%	−6.3%	−6.2%	−9.6%	−16.7%	−17.9%
RedAndBlack	−1.0%	−1.1%	−12.2%	−18.9%	−10.9%	−4.0%	−4.1%	−7.2%	−12.1%	−6.2%
Solider	−8.0%	−7.9%	−31.3%	−41.4%	−40.4%	−13.6%	−13.4%	−19.8%	−28.7%	−28.1%
Queen	−5.9%	−5.9%	−11.8%	−17.0%	−15.7%	−7.3%	−7.3%	−9.1%	−12.9%	−11.8%
LongDress	−1.1%	−1.1%	−8.3%	−11.2%	−10.2%	−3.8%	−3.6%	−5.7%	−8.2%	−7.3%
Avg.	−4.0%	−4.0%	−16.0%	−23.0%	−21.1%	−7.0%	−6.9%	−10.3%	−15.7%	−14.3%
Enc. time self	100%									
Dec. time self	100%									
Enc. time child	98%									
Dec. time child	99%									

Performance Analysis

- Intra blocks reduce significantly, resulting in taking adv of inter coding efficiency

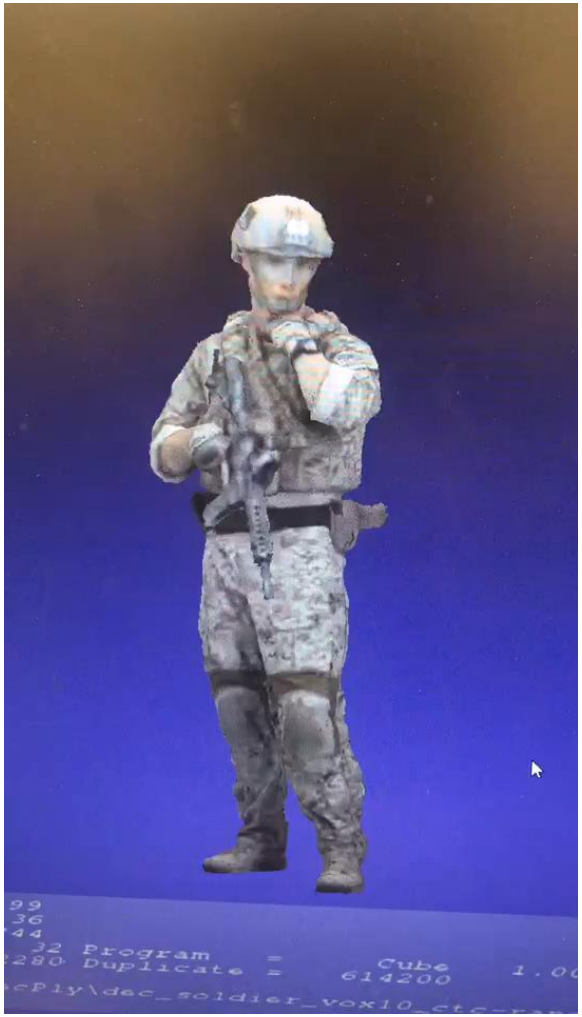


(a) Soldier Geometry Anchor (b) Soldier Geometry Normative (c) Soldier Geometry Non-normative

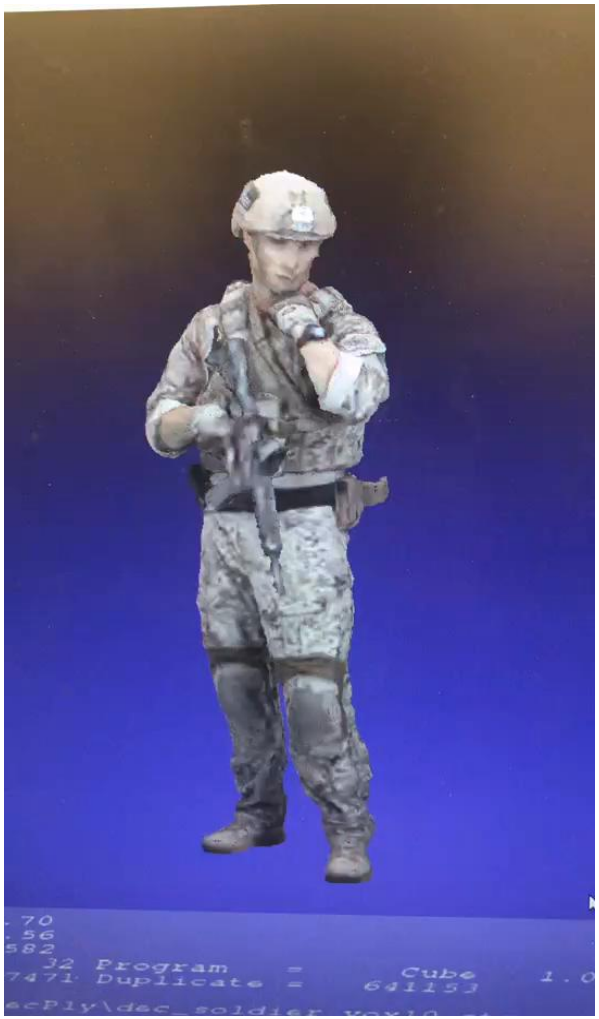


(d) Soldier Geometry Anchor (e) Soldier Geometry Normative (f) Soldier Geometry Non-normative

Subjective quality



Anchor



Proposed

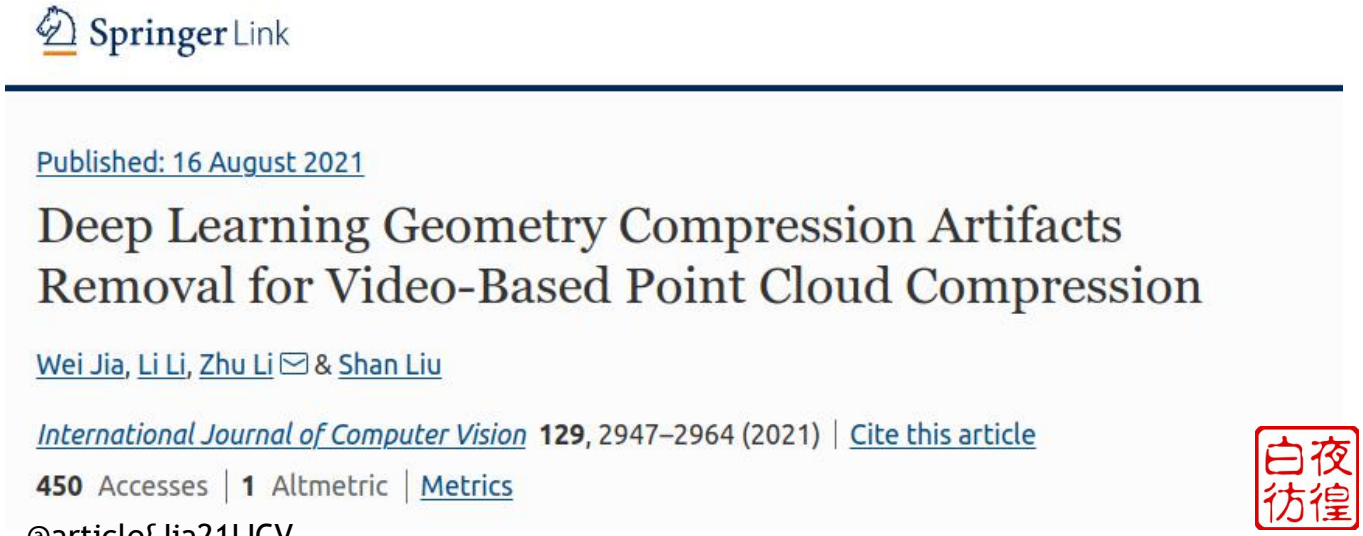
Adv 3D Motion for VPCC Summary

- Pain point: motion coherence is destroyed in the VPCC projection process, leads to poor motion compensation performance
- Key contribution: recover motion coherence in 3D domain, and generate a predictor for 2D motion estimation and compensation in HEVC codec.
- Significance: adopted in the VPCC test model.

Outline

- Short Self Intro
- Research Motivation and Highlights
- Sparse Conv Engine Based PCC
  - PU-Dense: point cloud upscaling work (T-IP)
  - Compression artifacts removal (T-MM)
- Video based PCC
  - Advanced 3D motion for VPCC (T-IP)
  - Depth Field Denoising (IJCV)
- Summary

VPCC Depth Denoise Work

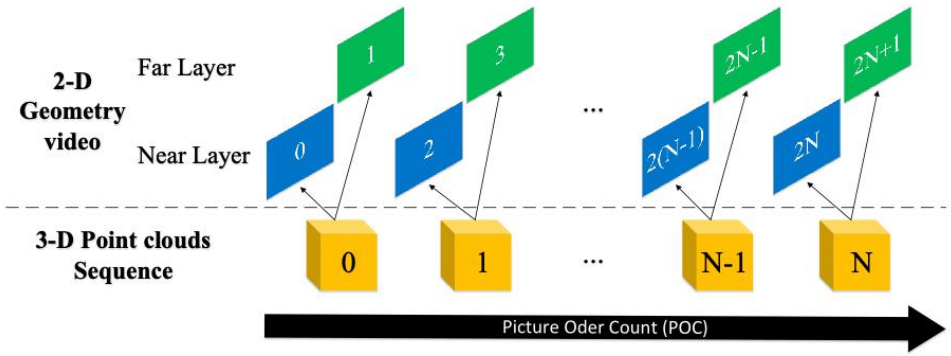


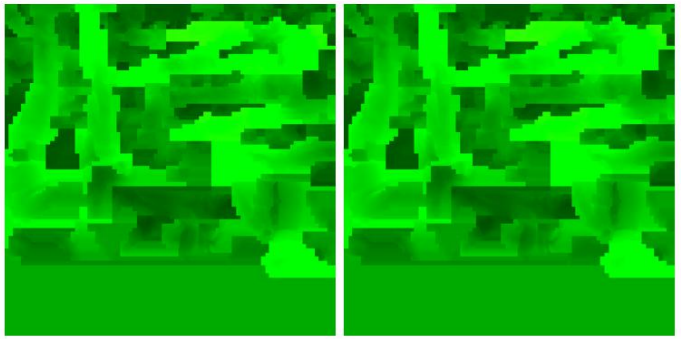

```
@article{Jia21IJCV
  author    = {Wei Jia and
               Li Li and
               Zhu Li and
               Shan Liu},
  title     = {Deep Learning Geometry Compression Artifacts Removal for Video-Based
               Point Cloud Compression},
  journal   = {Int. J. Comput. Vis.},
  volume    = {129},
  number    = {11},
  pages     = {2947--2964},
  year      = {2021},

  doi       = {10.1007/s11263-021-01503-6},
}
```

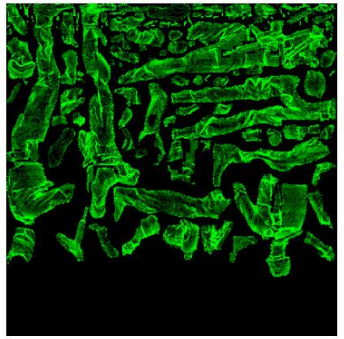
The Geometry in VPCC

- Geometry is projected to 2 layers of depth in VPCC
  - mainly used to handle occulation
  - coding wise is treated as interleaved sequences





(a) Near layer (b) Far layer

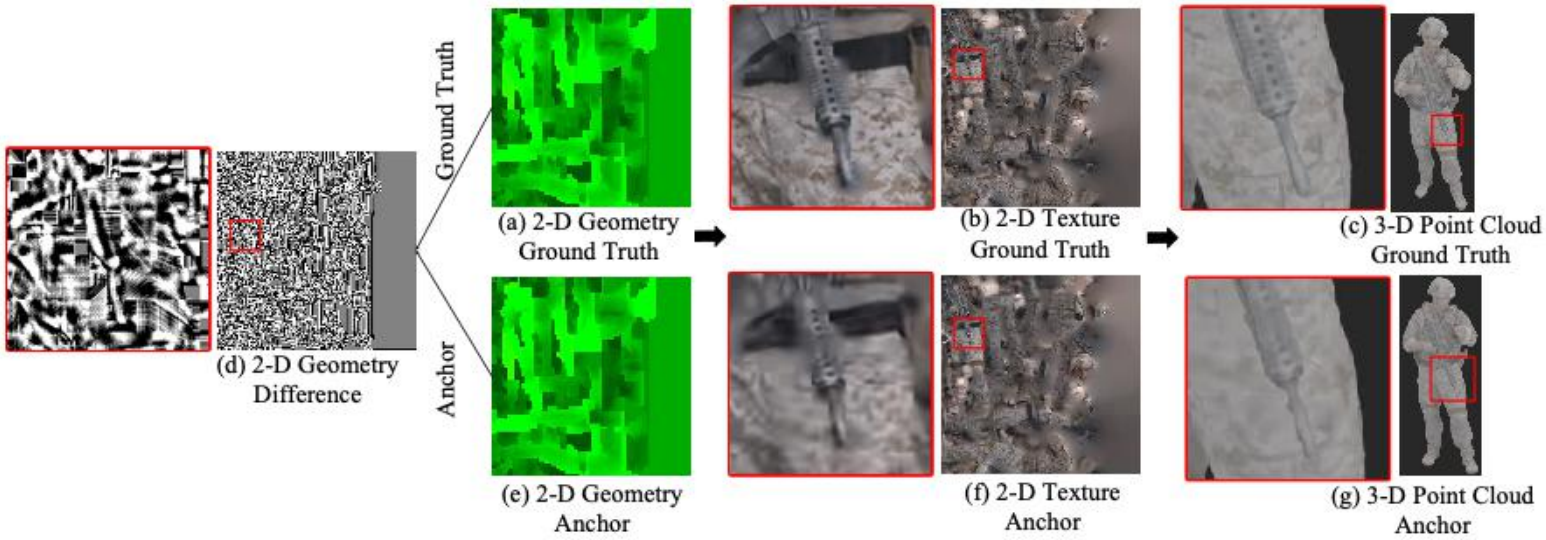


(c) The difference of near and far layer

Near (a) and far (b) layer frames in 2-D geometry video

Geometry Recovery from Depth

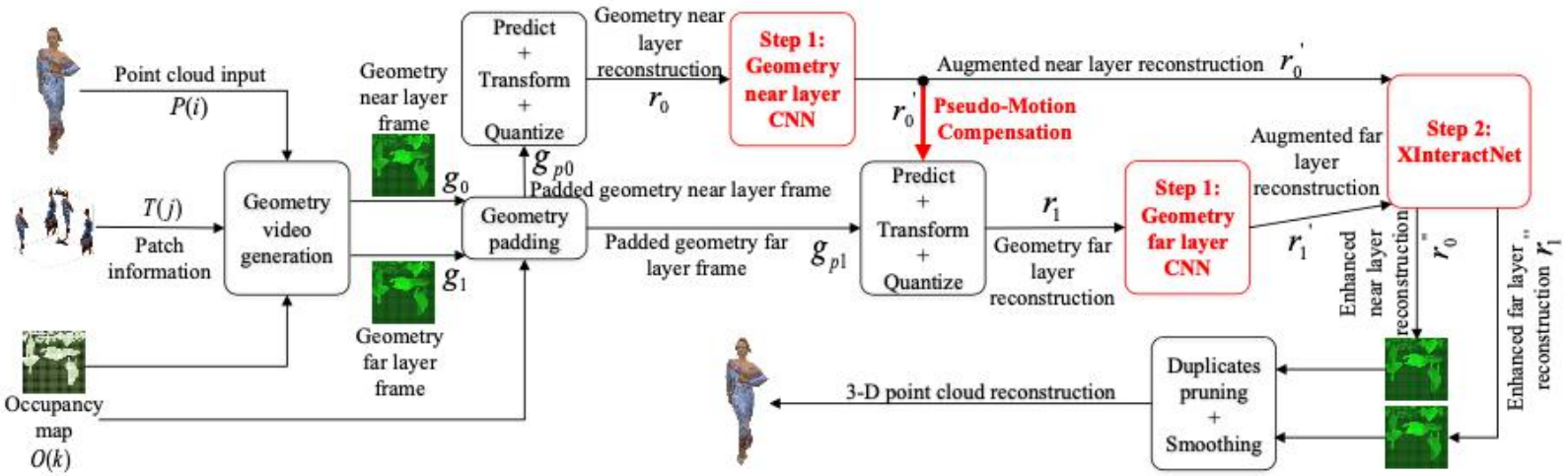
- VPCC use occupancy map to signal patches, error comes from:
  - quantization loss in depth (near and far field)
  - as well as the occupancy resolution loss in [Jia21TMM]* (4x4 by default)



*[Jia21TMM] W. Jia, L. Li, A. Akhtar, Z. Li and S. Liu, "Convolutional Neural Network-based Occupancy Map Accuracy Improvement for Video-based Point Cloud Compression," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3079698.

Joint Near and Far Depth Field Denoising

- A dual path network taking in near and far depth field
- Training strategy:
  - near field denoising network is trained first
  - far field benefit from denoised near field input in a pseudo-motion compensation scheme
  - some regularization in loss of near and far depth field interaction



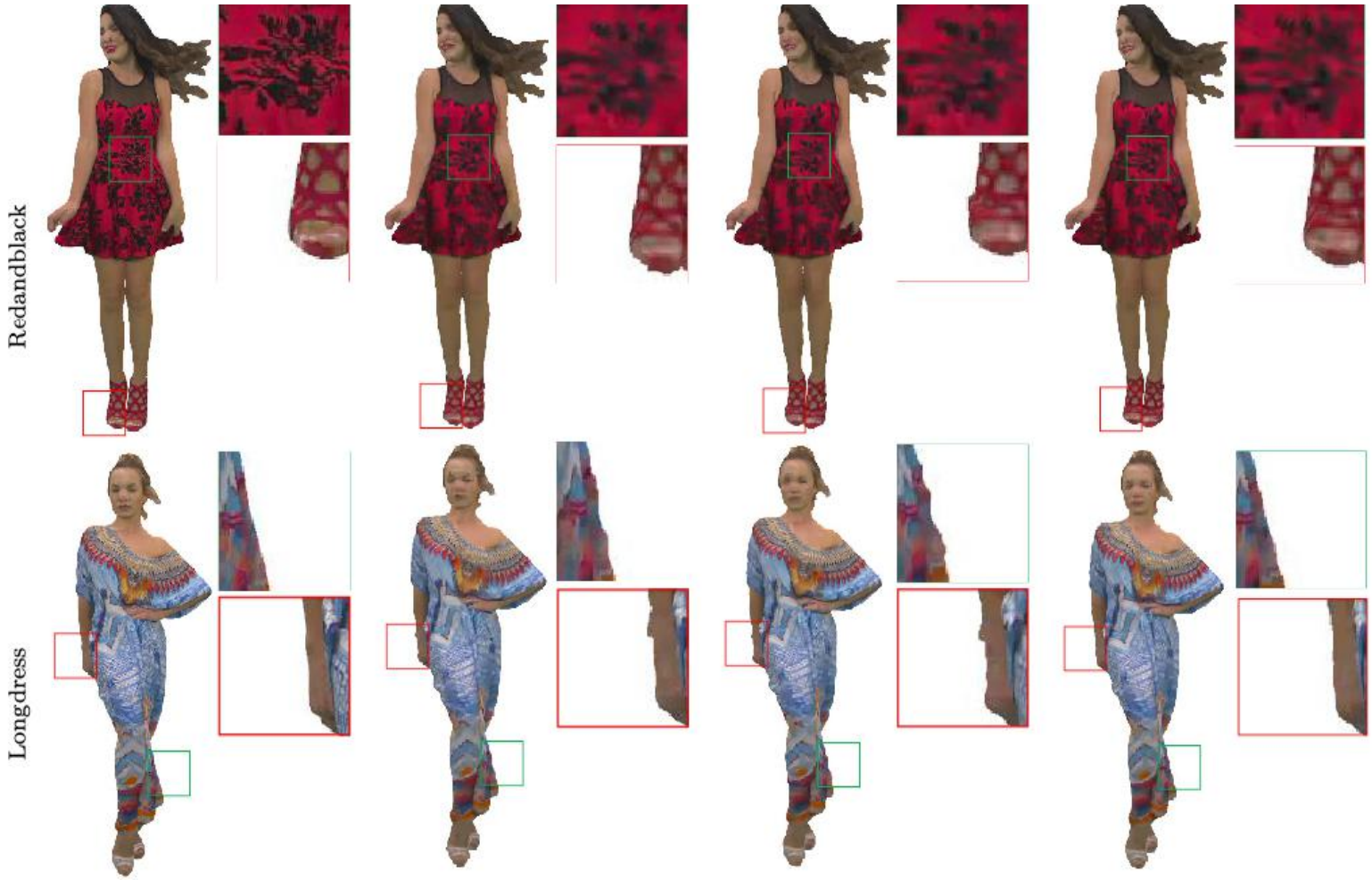
Results

- BD rate for all intra case:
  - significant coding gains :)
  - while very limited decoding complexity cost

Table 3 Proposed two-step method and geometry padding method [18] against the geometry smoothing method [32] respectively on BD-rate and time complexity within the first 32 frames of sequences under all intra

Class	Sequence	V-PCC with geometry padding method (SOTA) [18]					Proposed two-step method				
		Geom.BD-Totalrate		Attr.BD-Totalrate			Geom.BD-Totalrate		Attr.BD-Totalrate		
		D1 ↓	D2 ↓	Luma ↓	Cb ↓	Cr ↓	D1 ↓	D2 ↓	Luma ↓	Cb ↓	Cr ↓
A	Loot	−4.4%	−9.9%	4.4%	4.8%	5.4%	**-20.8%**	**-18.2%**	−1.2%	−1.4%	−0.5%
	Redandblack	−0.8%	−7.5%	4.4%	5.9%	5.0%	**-10.1%**	**-11.3%**	−1.1%	−0.9%	−2.4%
	Soldier	−1.5%	−7.8%	4.4%	7.3%	6.4%	**-13.8%**	**-12.8%**	−1.9%	0.1%	−0.8%
B	Longdress	−1.3%	−8.2%	2.3%	3.5%	3.2%	**-12.8%**	**-13.7%**	−2.9%	−1.4%	−2.3%
	Class a	−2.2%	−8.4%	4.4%	6.0%	5.6%	**-14.9%**	**-14.1%**	−1.4%	−0.7%	−1.2%
	class b	−1.3%	−8.2%	2.3%	3.5%	3.2%	**-12.8%**	**-13.7%**	−2.9%	−1.4%	−2.3%
Avg.	All	−2.0%	−8.3%	3.9%	5.4%	5.0%	**-14.4%**	**-14.0%**	−1.8%	−0.9%	−1.5%
	Enc.Self			102%					103%		
	Enc.Child			101%					101%		
	Dec.Self			102%					121%		
	Dec.Child			101%					212%		

Subjective Results



(a) Ground truth (b) Geometry smoothing method (c) Geometry padding method (d) Proposed two-step method

VPCC Depth Denoising Summary

- VPCC has a lot of room for improvement
  - joint near+far depth field denoising give significant BD rate reduction
  - this shall be combined with the Occupancy map SR work [Jia21TMM]
  - Attributes denoising still has many open opportunities

Acknowledgement

- Thanks for the support from

NSF AFOSR ONR QUALCOMM SAMSUNG oppo

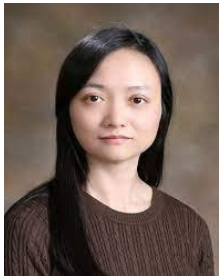
- Many thanks for the hard work and creativity of my students, post-doc and collaborators

Anique Akhtar

Wei Jia
ByteDance

Li Li, USTC

Zhan Ma
Nanjing Univ



Yi-Ling Xu
SJTU

Jianle Chen, Google

Shan Liu,
Tencent

Vlad Zakharchenko
OPPO

Geert van der Auwera,
Qualcomm

Summary

- Q & A