VOLUT: EFFICIENT VOLUMETRIC STREAMING ENHANCED BY LUT-BASED SUPER-RESOLUTION

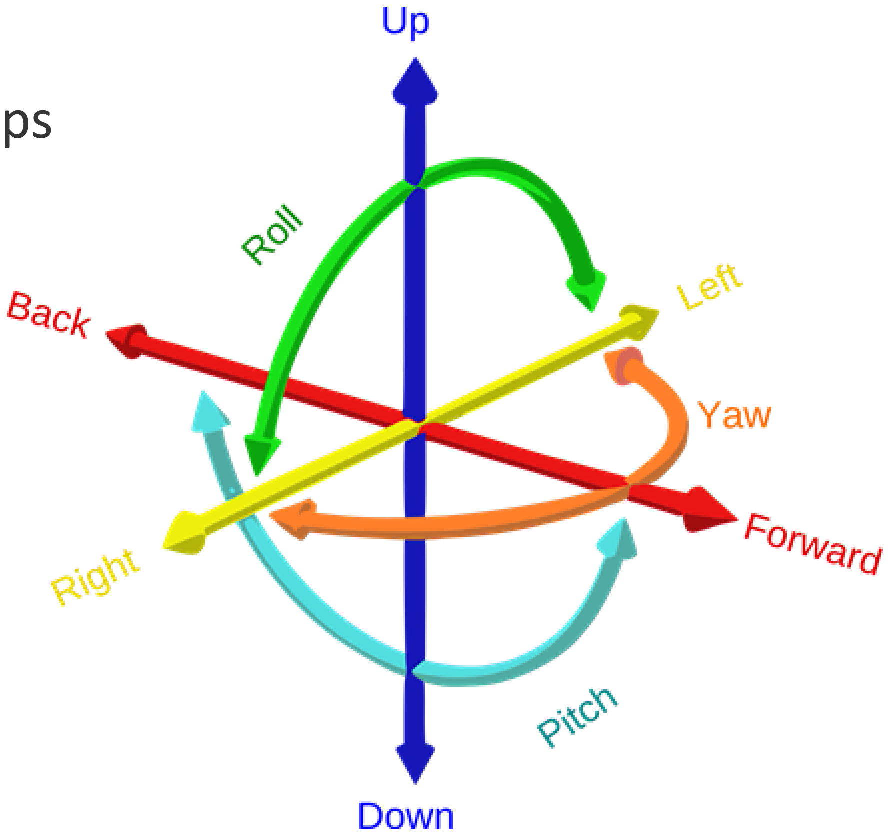
Chendong Wang[1], Anlan Zhang[2], Yifan Yang[3], Lili Qiu[3], Yuqing Yang[3], Xinyang Jiang[3], Feng Qian[2], Suman Banerjee[1]

[1]University of Wisconsin-Madison, [2]University of Southern California, [3]Microsoft Research Asia

MLSys 2025

Volumetric Video

- Immersive: 6-DoF (degree-of-freedom) movement
  - X, Y, Z, yaw, pitch, roll
- Watch through VR/AR/MR headsets or on desktops
- Applications: e.g., telepresence

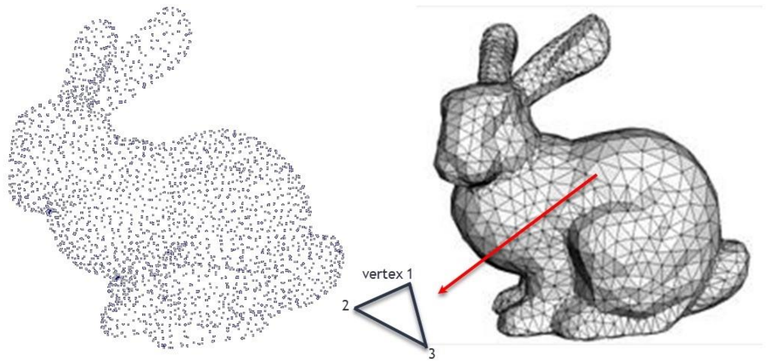


Degree of freedom

Volumetric Video

- Representation: point cloud or 3D mesh
- Challenge of streaming volumetric video over the Internet
  - Bandwidth-intensive: hundreds of Mbps or even Gbps
    - Example: 200k points (15 Bytes for a point) per frame at 30 FPS requires 720 Mbps

Point Cloud: un-sorted set of 3D points with attributes, e.g., color

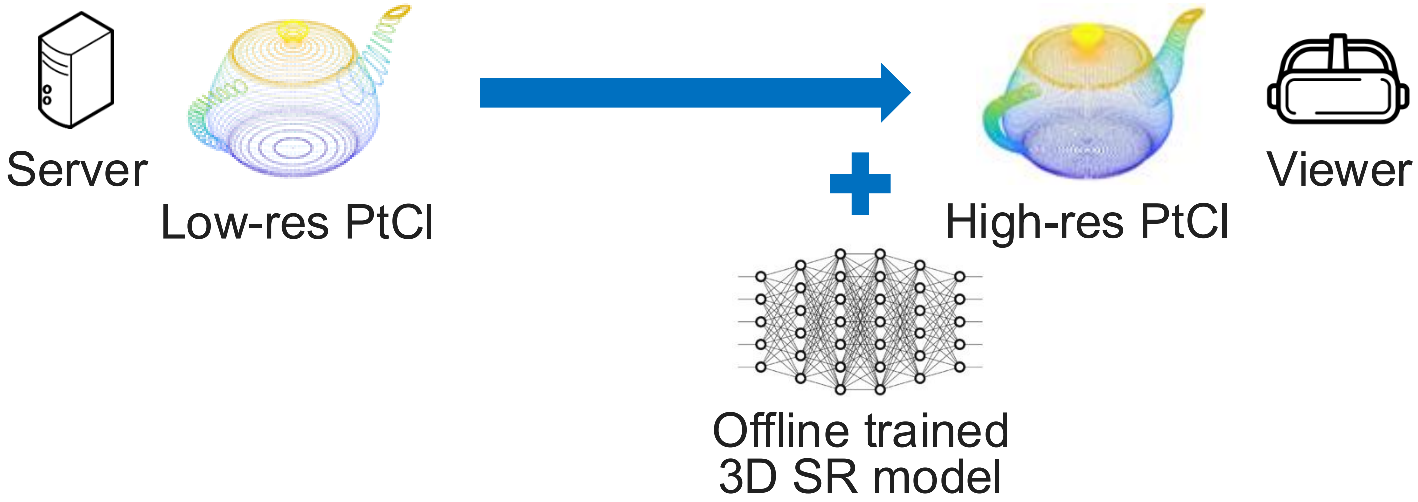


Point Cloud   Mesh (shaded)

3D Mesh: 3D model consisting of polygons (vertices, edges, and faces)

Existing SR-based Solution (YuZu, NSDI'22)

- Improve the QoE of volumetric video streaming through 3D super resolution (SR)
  - Point cloud (PtCl) resolution/quality: point density



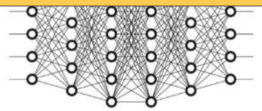
Sizeof(low-res PtCl + SR model) << *Sizeof*(high-res PtCl)

Bandwidth Saving! But …

Existing SR-based Solution (YuZu, NSDI'22)

- Improve the QoE of volumetric video streaming through 3D super resolution (SR)
  - Point cloud (PtCl) resolution/quality: point density

☹
- Still heavyweight after extensive optimization
- Limited upsampling ratios
- Growing offline training cost



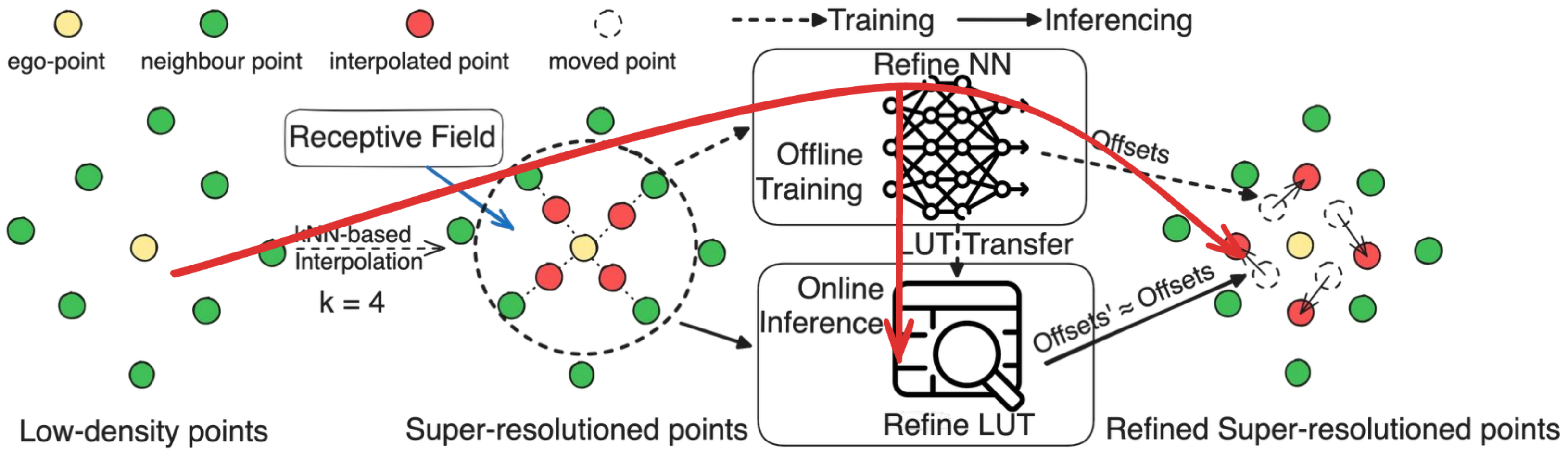
Offline trained
3D SR model

Sizeof(low-res PtCl + SR model) << *Sizeof*(high-res PtCl)

Bandwidth Saving! But …

Our Solution: VoLUT

- Interpolation + Refinement
  - Reduce 3D SR complexity
- Trade on-device memory for efficient model inference
  - NN → Lookup Table (LUT)

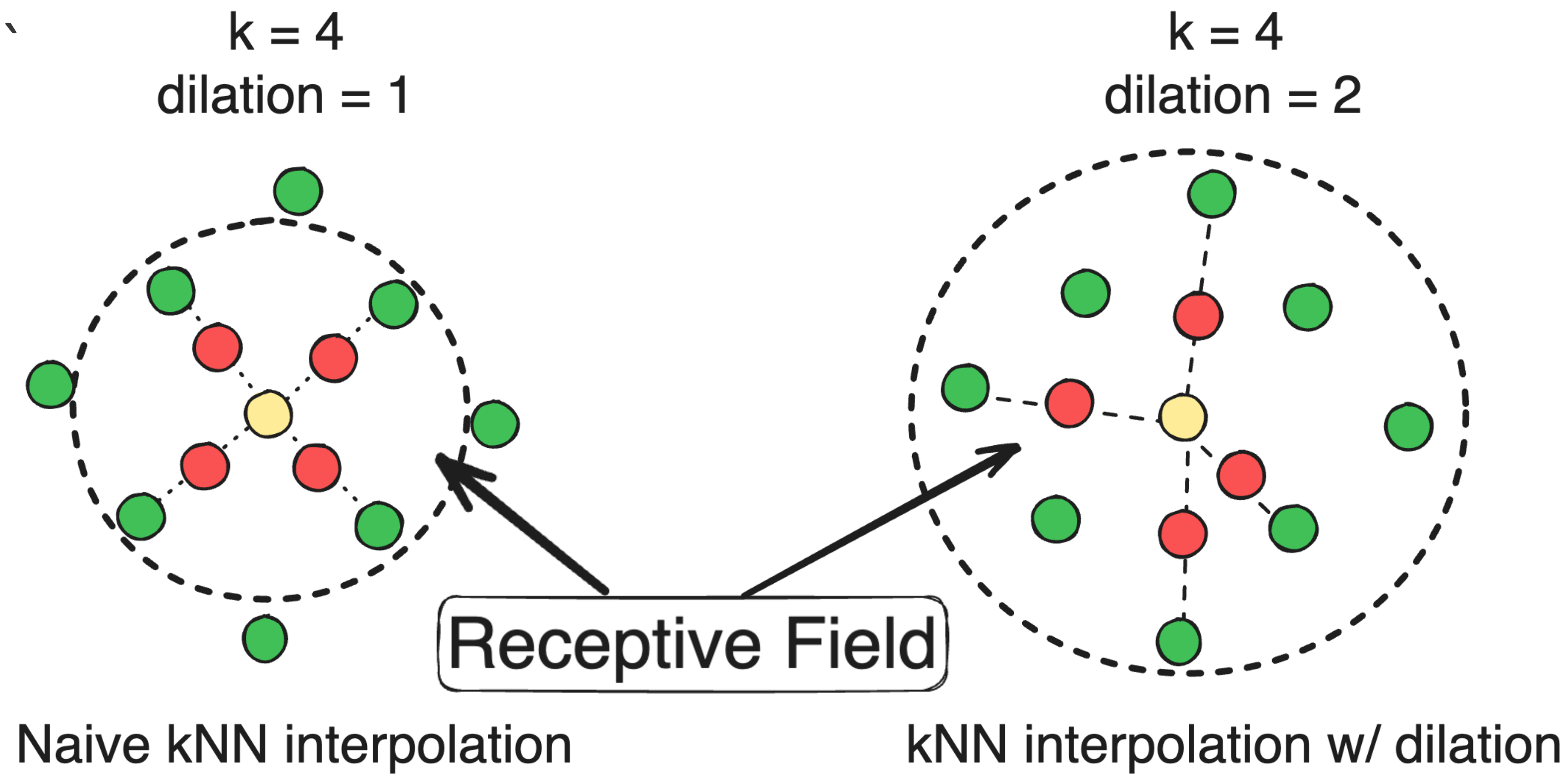


Challenges

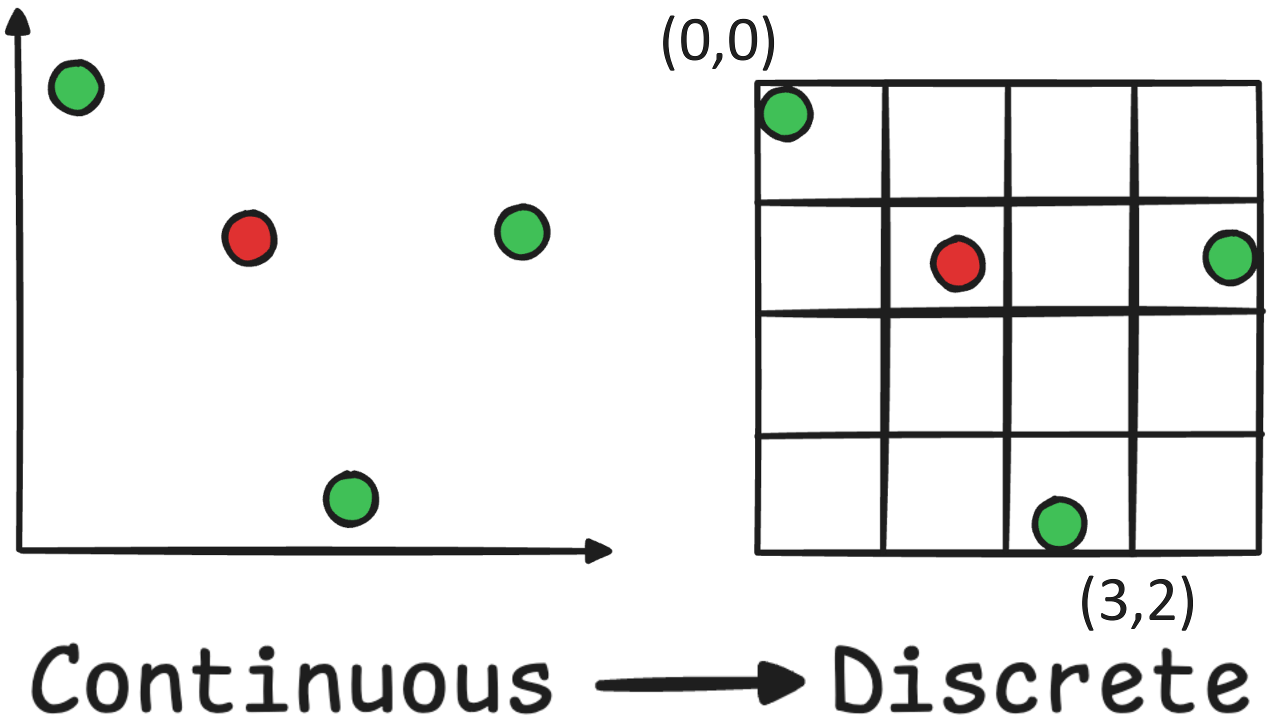
- How to reduce visual distortions and computational complexity
- How to build LUTs for continuous 3D SR
- How to do adaptive streaming for arbitrary SR ratio

Core Technique 1 – Dilated Interpolation

- Add "dilation" to kNN → larger receptive field
- Octree-based parallelism + neighbor reuse → 4× speed-up

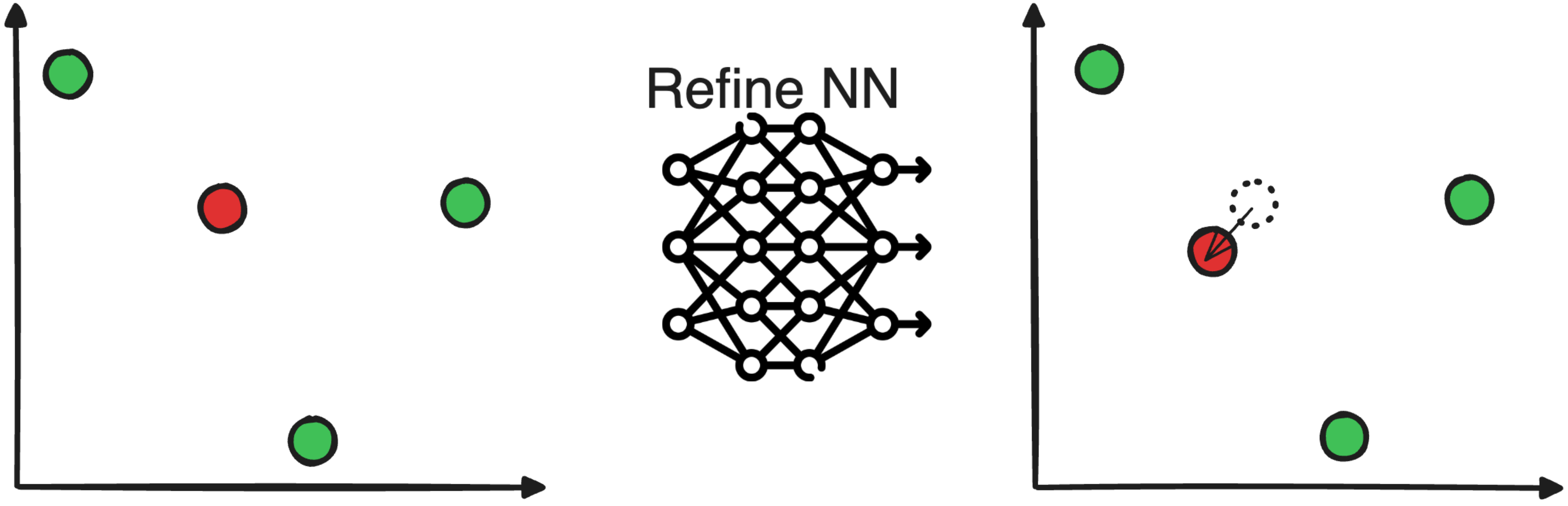


Core Technique 2 – Continuous space LUT conversion



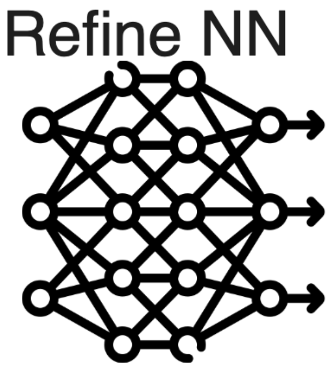
- Raw k = 3 neighborhood
- Normalized to $[0,1]^{d=2}$
- Quantization (b×b), b = 4

Core Technique 2 – Continuous space LUT conversion



Core Technique 2 – Continuous space LUT conversion

Key	Value
(0,0) (0,0) (0,0) (0,0)	
(0,0) (0,0) (0,0) (0,1)	
…	
(1,1) (1,1) (1,1) (1,1)	

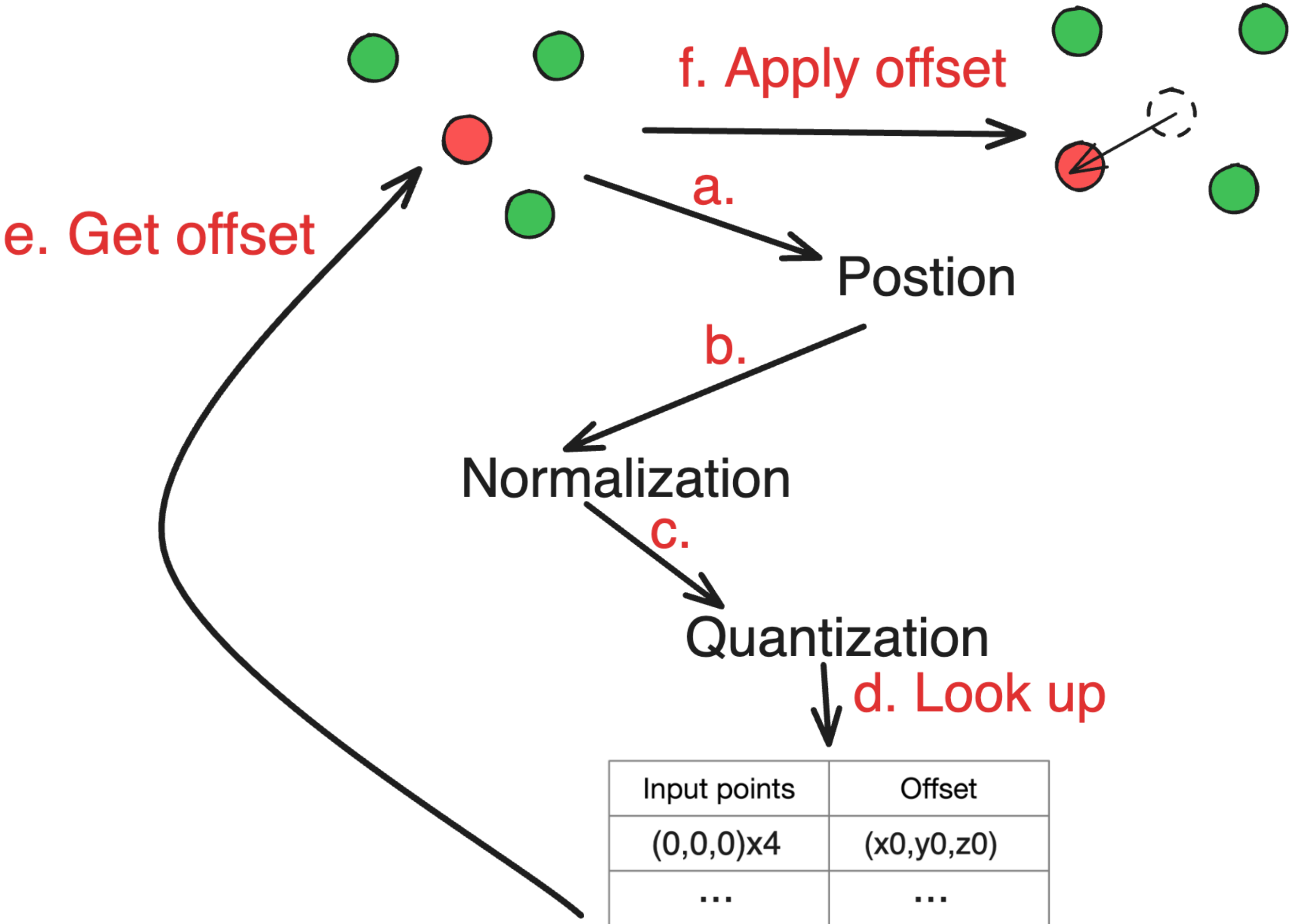


Key	Value
(0,0) (0,0) (0,0) (0,0)	offset_0
(0,0) (0,0) (0,0) (0,1)	offset_1
…	
(1,1) (1,1) (1,1) (1,1)	offset_m

m=b^(nd)

- LUT size tradeoff:
  - n=4,b=64 → ~100 MB;
  - n=4,b=128 → ~1.6 GB

Core Technique 3 – LUT Refinement



Input points	Offset
(0,0,0)x4	(x0,y0,z0)
...	...
(15,15,15)x4	(xk,yk,zk)

Look Up Table

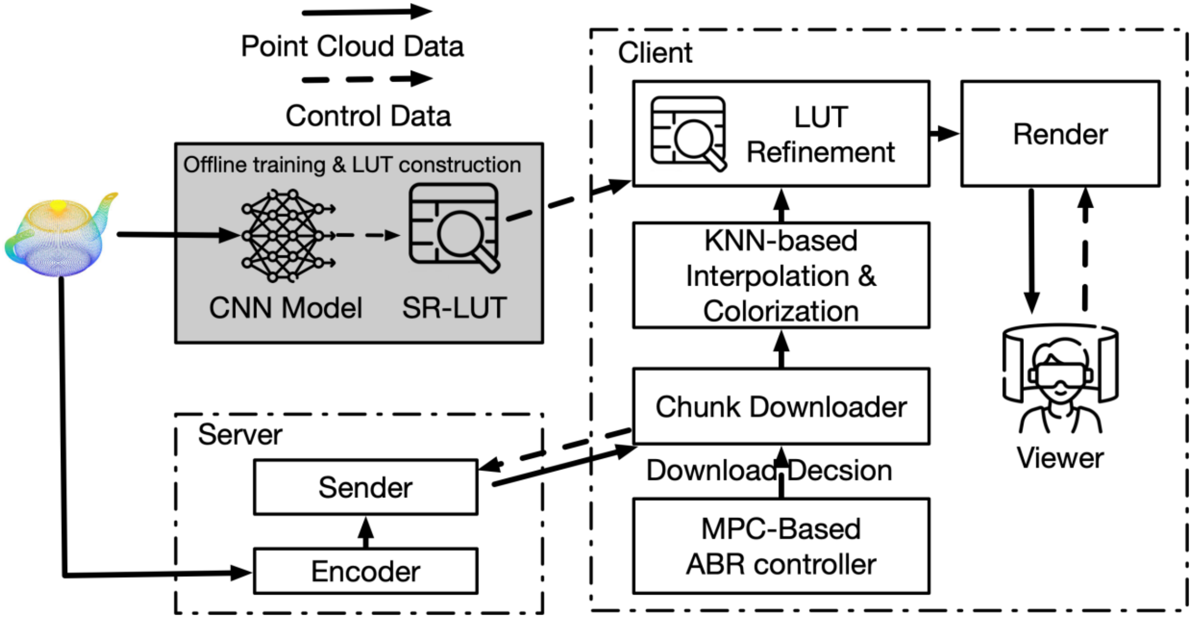
Core Technique 4 – Continuous ABR

- Discrete bitrate ladders → quality jumps (360p, 1080p, 4k …)
- Our MPC formulation:

$$\max_{\rho_t} \sum_t \left(\alpha\, Q(\rho_t) - \beta\, V(\rho_t) - \gamma\, S(\rho_t)\right)$$

- Any down-sampling rate → smooth transitions

VoLUT System Architecture

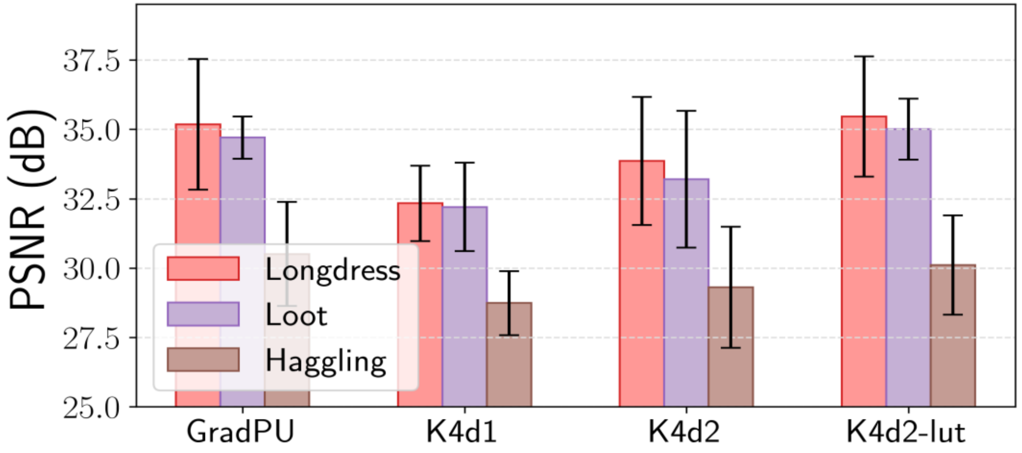


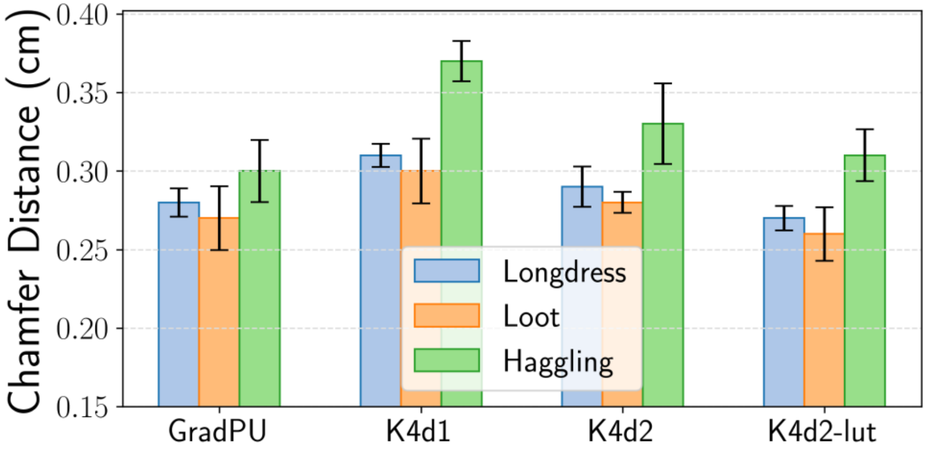
- Server: density-aware downsampling + ABR
- Client:
  - Dilated kNN interpolation
  - LUT-based refinement
  - Rendering

Experimental Setup

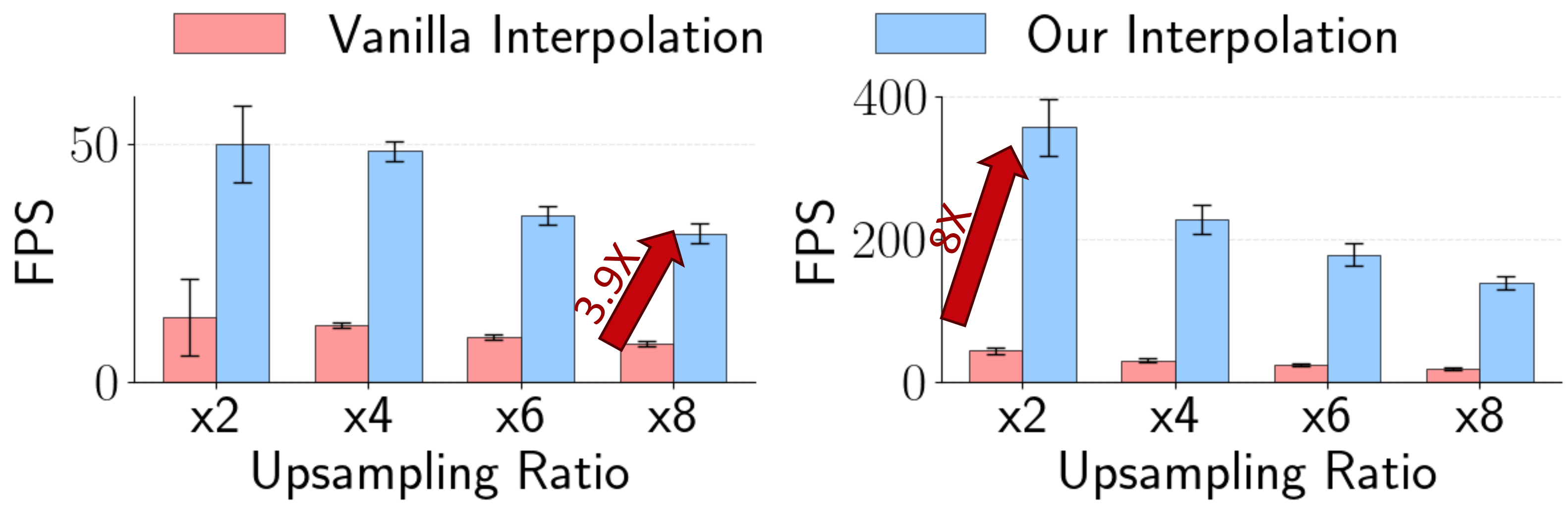
- Datasets: 6 volumetric videos (~100 K pts/frame)
- Networks: fixed 50/75/100 Mbps + LTE traces
- Hardware: RTX 3080Ti, Orange Pi (Quest-class)
- Baselines: GradPU, YuZu, ViVo

Super-Resolution Quality



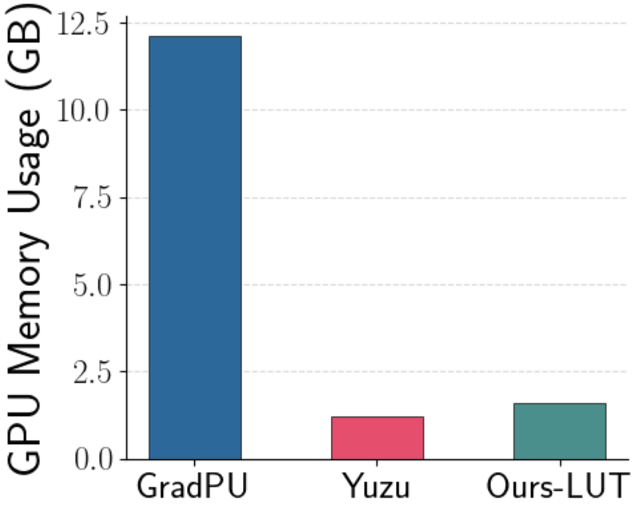


Performance & Resource Usage

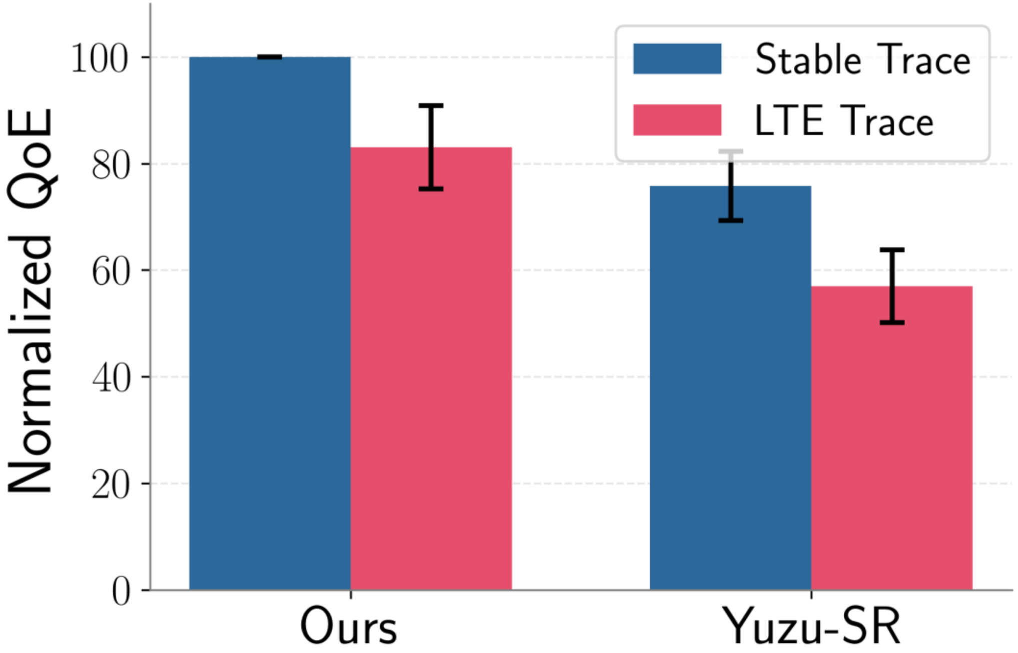


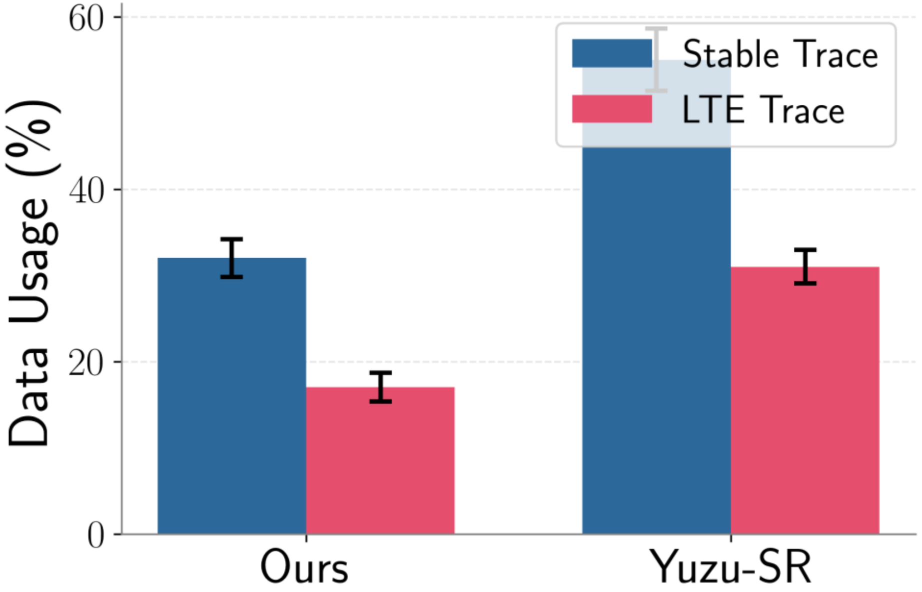
Orange Pi

3080Ti



QoE & Bandwidth Savings





- Under real User traces

Conclusion

- **Key contributions:**
  - Enabling 3D Super-resolution at line rate on mobile level devices
  - Making possible 3D continuous space to benefit from LUTs
  - End-to-end systems that achieves high QoE with low data usage.

THANKS!

VOLUT: EFFICIENT VOLUMETRIC STREAMING ENHANCED BY LUT-BASED SUPER-RESOLUTION

Chendong Wang[1], Anlan Zhang[2], Yifan Yang[3], Lili Qiu[3], Yuqing Yang[3], Xinyang Jiang[3], Feng Qian[2], Suman Banerjee[1]

[1]University of Wisconsin-Madison, [2]University of Southern California, [3]Microsoft Research Asia