

# YuZu: Neural-Enhanced Volumetric Video Streaming

*Anlan Zhang*<sup>1</sup> *Chendong Wang*<sup>1\*</sup> *Bo Han*<sup>2</sup> *Feng Qian*<sup>1</sup>

<sup>1</sup>*University of Minnesota, Twin Cities* <sup>2</sup>*George Mason University*



April, 2022

# Volumetric Video

- Volumetric Video
  - Immersion
  - Telepresence
  - VR/AR



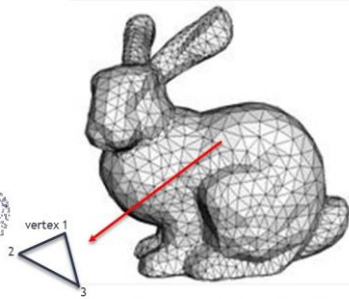
Video source: [https://www.youtube.com/watch?v=aO3TAke7\\_MI](https://www.youtube.com/watch?v=aO3TAke7_MI)

# Volumetric Video

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)

- Representation
  - Point Cloud
  - 3D Mesh
- Streaming over the Internet: high bandwidth consumption
  - Example:  $720\text{Mbps} = 200\text{K points} * 15 \text{ bytes} * 30 \text{ FPS} * 8 / 1000 / 1000$



# Leveraging 3D SR (Super Resolution)

- Improve QoE (quality-of-experience)
- SR for static point cloud
  - SR model (DNN): low-resolution (LR) → high-resolution (HR)
  - Resolution: point density
- SR for VoD (video-on-demand)
  - Offline model training: leveraging overfitting
  - Online streaming: LR content & SR model
  - Bandwidth reduction or QoE improvement



# Motivation: A Case Study

- SR model: PU-GAN [1]
  - SR ratio: 4, 25K → 100K points
- Test video: *Lab*, 2 min
  - ~100K points per frame
- NVIDIA 2080Ti GPU

## Positive Findings 😊

- Good upsampling accuracy
- Significant bandwidth saving, ~74%

## Challenges 😞

- No generic QoE model
- Poor runtime performance
  - < 0.1 FPS, 7GB memory
- No color support

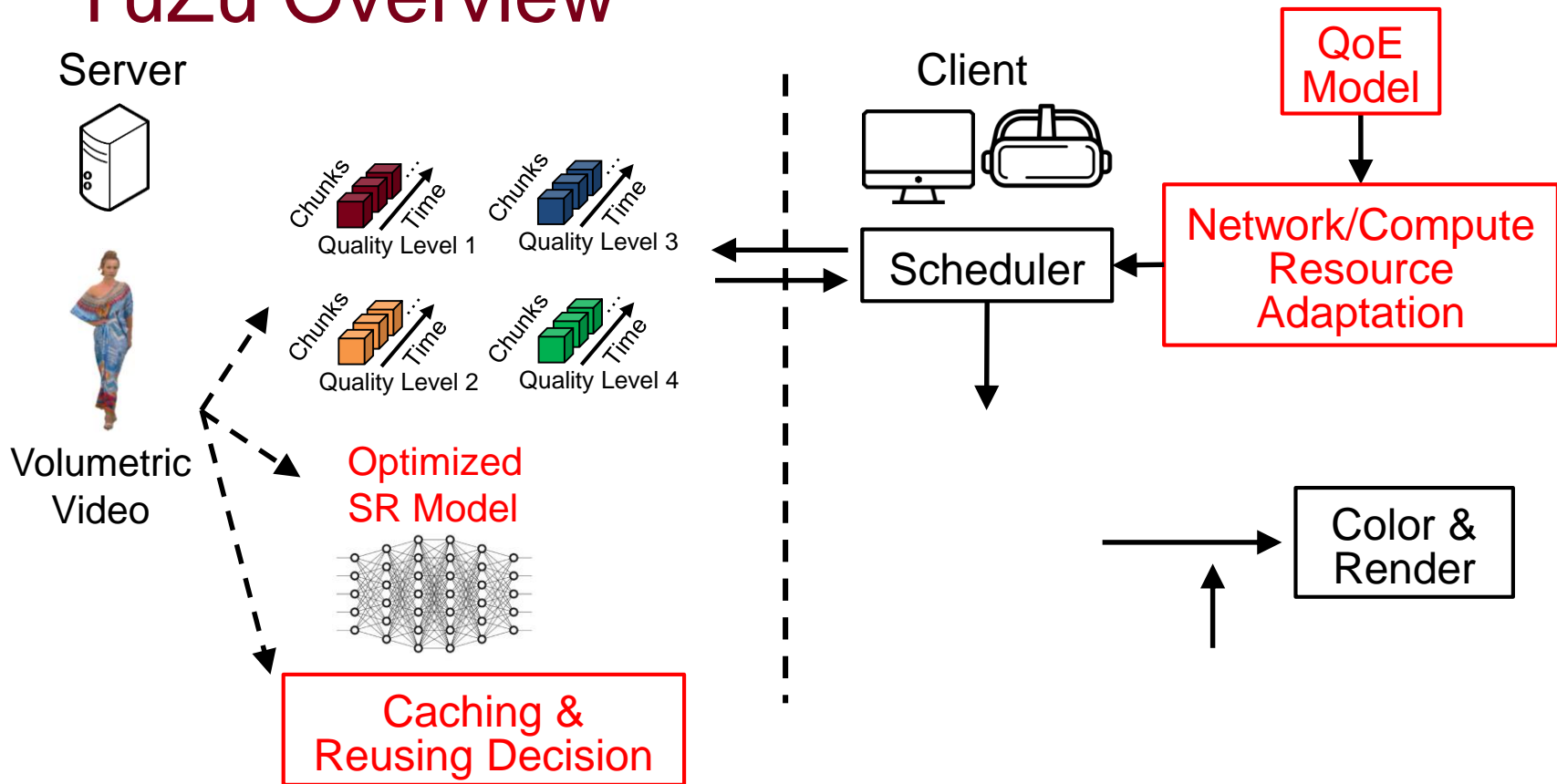


# Our Approach: YuZu

- An empirical QoE model
  - Large-scale (1,446 participants) user studies
- YuZu system design & implementation
  - Intra-frame SR
  - Inter-frame SR
  - Network/Compute resource adaptation
- YuZu evaluation
  - QoE improvement
  - Runtime performance



# YuZu Overview



# QoE Model & User Studies

- An empirical QoE model
  - Point density, viewing distance, SR ratio, visibility, quality switch, stall, etc.
- User studies
  - 4 volumetric videos of human portraits
  - Our optimized PU-GAN [1] model
  - 1,446 participants from 40 countries
  - 10-fold cross validation & cross-video validation
- Takeaways
  - Median QoE prediction error: 12.49%
  - Generic for volumetric videos of the same genre (human portraits)





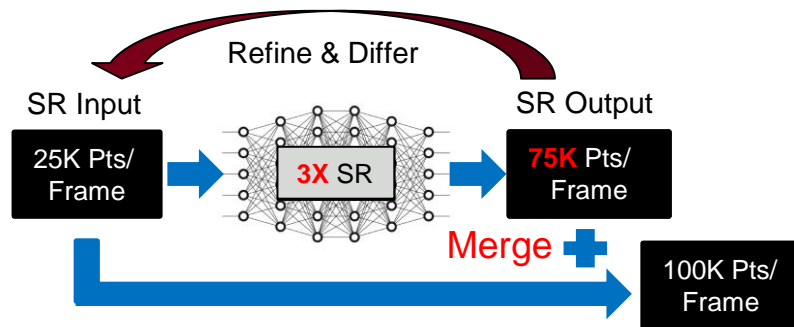
# System Design of YuZu

- Intra-frame SR
  - Speed up SR upsampling within a frame
- Inter-frame SR
  - Cache and reuse SR results across frames
- Network/Compute resource adaptation



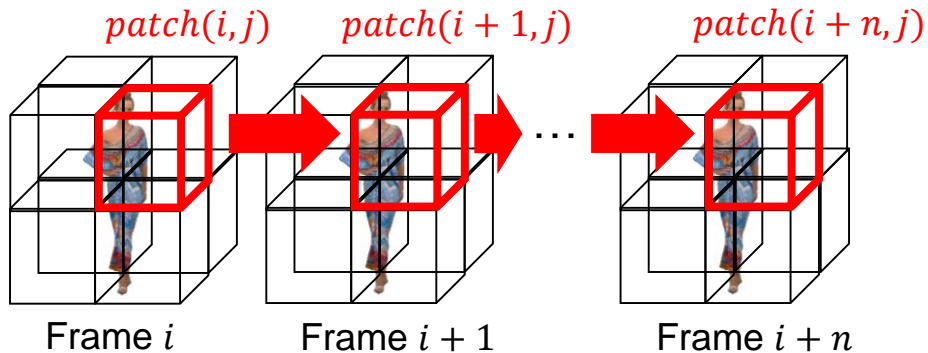
# Intra-frame SR

- Speed up SR a single frame
- Optimize patch generation
  - 3D SR: per-patch basis
  - Trim pre- and post-processing
- Optimize model structure
  - Pruning through layer-by-layer profiling
  - More efficient feature extraction
- Data reduction
  - Merge SR input with SR output
  - Maintain the same SR ratio with less computation overhead



# Inter-frame SR

- Speed up SR across consecutive frames
  - Similarity across consecutive frames
- Cache & reuse SR results
  - Per-patch basis
  - Similarity between patches
  - Only patches at same location
  - Dynamic programming
    - Minimize # of patches to be upsampled
- Offline
  - Pre-compute caching & reusing decisions for VoD content



# Network/Compute Resource Adaptation

- Trade-off
  - Download HR content: high network resource usage
  - Download LR content and Upsample it: high compute resource usage
- QoE-driven, two-stage adaptation
  - Before download each chunk
    - Coarse-grained Search
    - Search quality/SR-ratio assignment of to-be-downloaded chunk
  - Before upsample each frame
    - Fine-grained Search
    - Fine-tune SR ratios



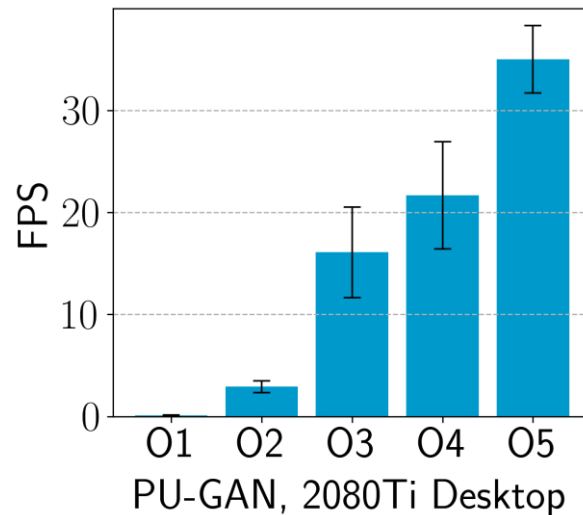
# Evaluation

- Implementation
  - 10,848 LoC (lines of code) in C/C++
- SR performance breakdown
  - Effectiveness of each optimization for 3D SR
- QoE improvement of YuZu
- End-to-end performance of YuZu
- YuZu vs. viewport-adaptive streaming



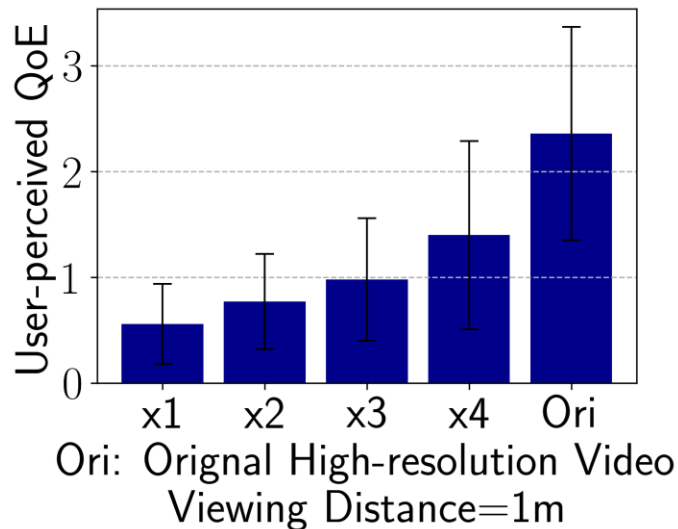
# SR Performance Breakdown

- O1 (Baseline): vanilla PU-GAN [1] model
  - 2080Ti desktop, SR ratio: 4
- Cumulative optimizations
  - O2: O1 + optimize patch generation
  - O3: O2 + optimize model structure
  - O4: O3 + merge SR input with SR output
  - O5: O4 + cache & reuse SR results
- Takeaways
  - Significantly speed up upsampling (up to 307x)
  - Huge GPU memory usage reduction (up to 87%)
  - No accuracy degradation



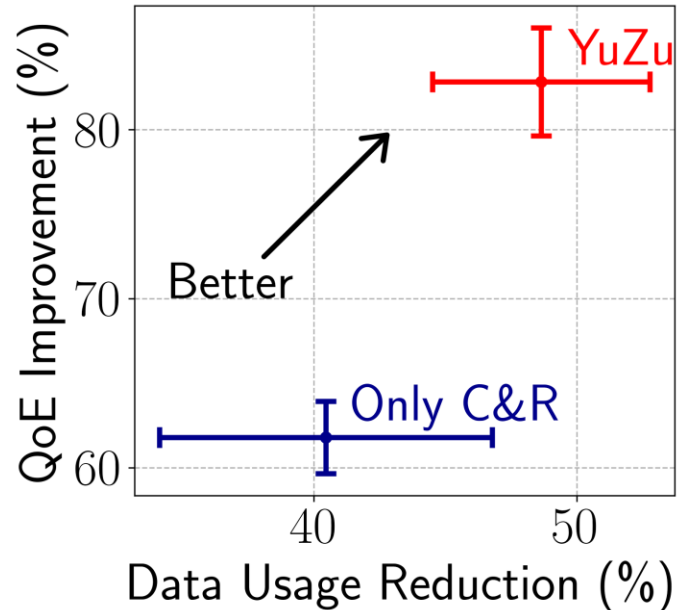
# YuZu's QoE Improvement

- 4 volumetric videos
  - Downsample to 25% # of points
- Optimized PU-GAN [1] model (YuZu)
  - 4 SR ratios: x1 (no SR), x2, x3, x4
- Subjective ratings
  - 512 participants
- Takeaways
  - YuZu boost QoE by up to 150%
  - Positive correlation between QoE improvement and SR ratio



# YuZu's End-to-end Performance

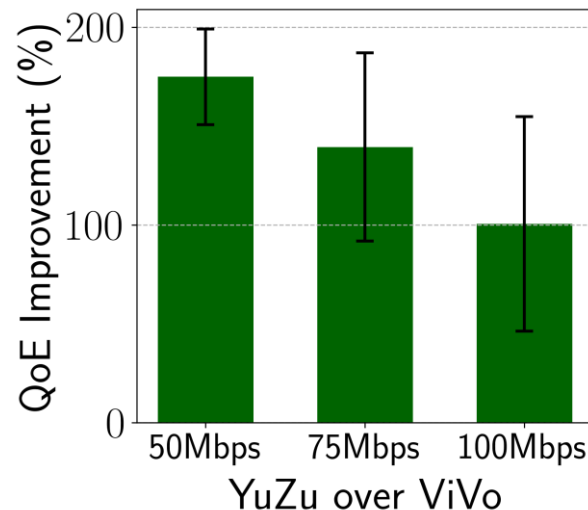
- Fluctuating bandwidth
  - 12 LTE Traces
- Baseline
  - 100% points, x1 SR
- Only C&R
  - 100% points, x1 SR
  - Cache & resume SR results
- Full-fledged YuZu
- Takeaways
  - Significant QoE improvement (83%/62%) and data usage reduction (49%/40%) for YuZu/Only C&R





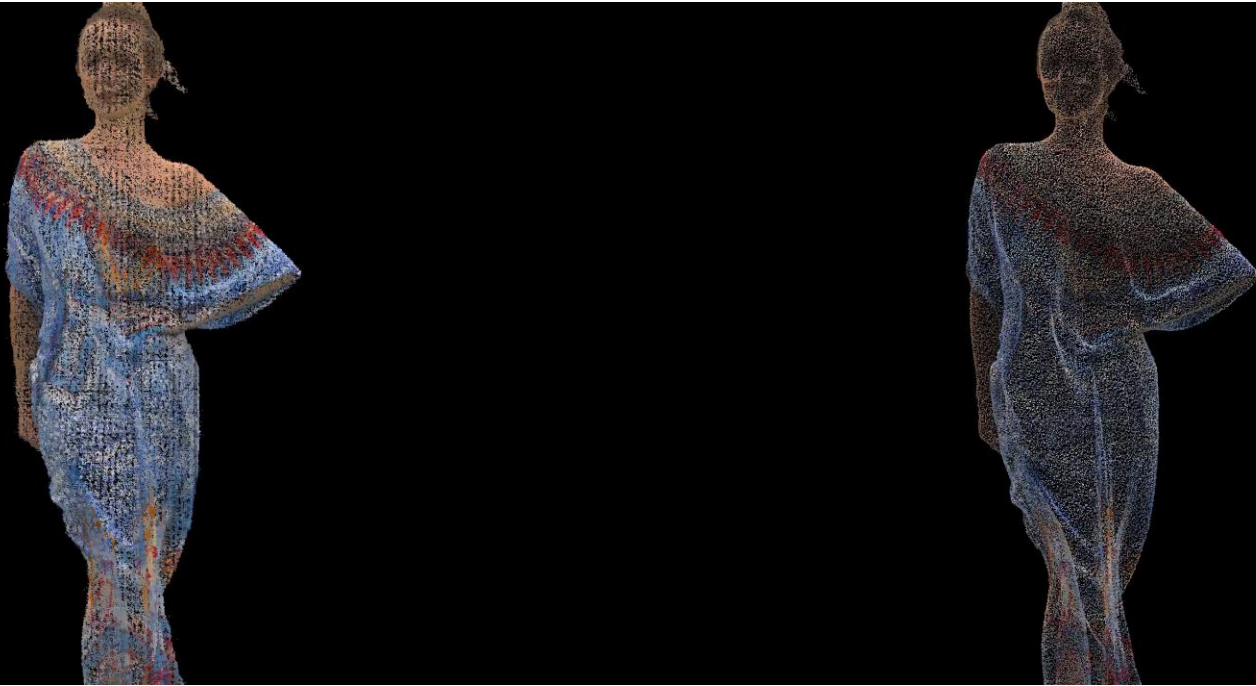
# YuZu vs. Viewport-Adaptive Streaming

- ViVo [1]
  - 6-DoF motion prediction
  - Content in viewport
- Wired network with stable bandwidth
  - 50, 75, 100 Mbps
- 4 videos with 32 users' motion traces
- YuZu outperforms ViVo by 101% to 175% on QoE



# Demo

- Left: 25% points x4 SR, right: 25% points x1 SR



# Conclusion

- An empirical QoE model
  - Large-scale (1,446 participants) user studies
- YuZu system design & implementation
  - Intra-frame SR
  - Inter-frame SR
  - Network/Compute resource adaptation
- YuZu evaluation
  - QoE improvement
  - Runtime performance



# YuZu: Neural-Enhanced Volumetric Video Streaming

*Anlan Zhang*<sup>1</sup> *Chendong Wang*<sup>1\*</sup> *Bo Han*<sup>2</sup> *Feng Qian*<sup>1</sup>

<sup>1</sup>*University of Minnesota, Twin Cities* <sup>2</sup>*George Mason University*



April, 2022