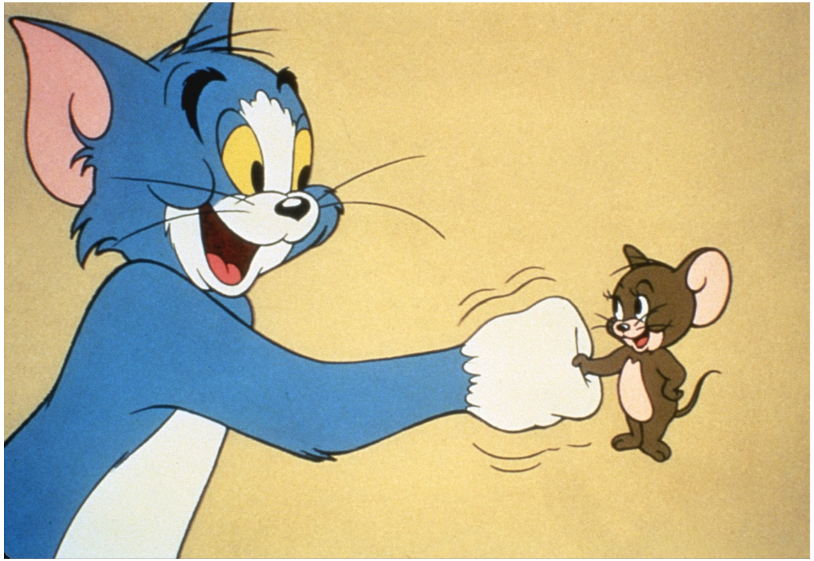
Efficient Volumetric Video Streaming Through Super Resolution

Anlan Zhang[1], **Chendong Wang**[1], Bo Han[2], Feng Qian[1]

[1] University of Minnesota, Twin Cities
[2] George Mason University

Video is Evolving

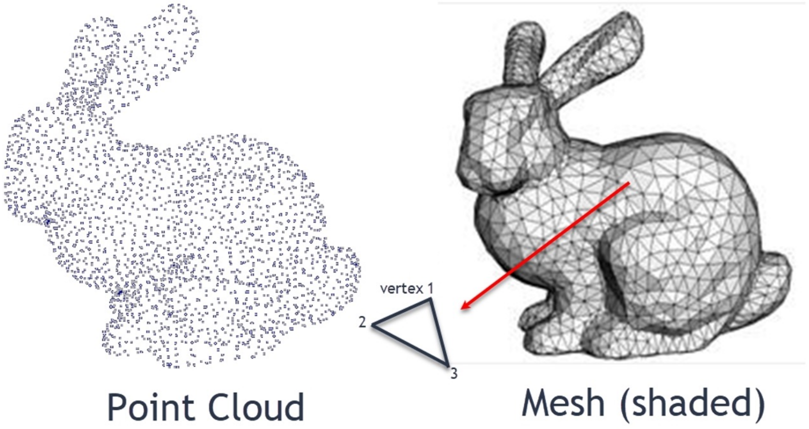


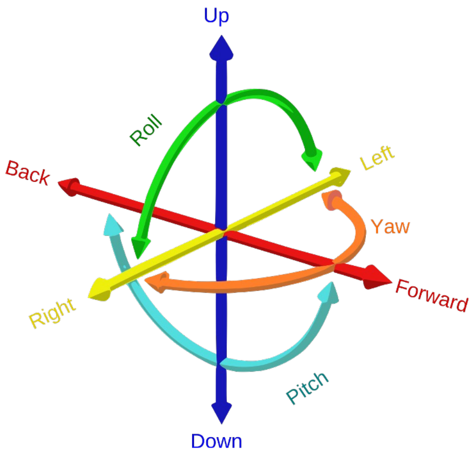
Traditional video
2-dimensional
0-DoF(Degree of Freedom)

360^{o} video
2-dimensional (spherical)
3-DoF

Image source: https://learnex.com.mx/home/principal/more-360-degree-video-content/,https://www.animationkolkata.com/blog/2017/03/06/the-evolution-of-tom-jerry-2d-animation/

Volumetric Video





- 3D animated point cloud or mesh
- 6-DoF
- Immersive telepresence experience

Image source: “Learn ARCore - Fundamentals of Google ARCore”

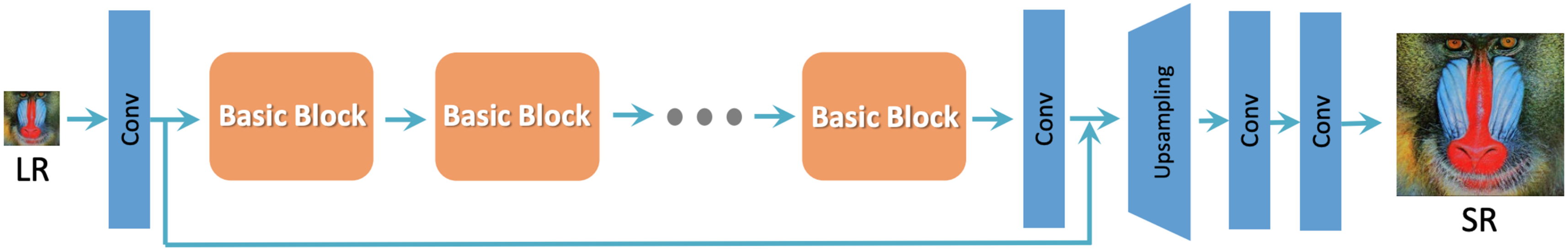
Video source: https://www.youtube.com/watch?v=aO3TAke7_MI

Volumetric Video Streaming

- Challenge:
  - High bandwidth utilization
- Current solutions:
  - DASH-PC[1]: bitrate adaptation
    - Network condition
  - ViVo [2]: visibility-aware volumetric video streaming
    - User viewport prediction
- Can we leverage the computation resource to mitigate the bandwidth consumption?

[1]M. Hosseini and C. Timmerer. Dynamic Adaptive Point Cloud Streaming. In ACM Packet Video, 2018.
[2]B. Han, Y. Liu, and F. Qian. ViVo: Visibility-Aware Mobile Volumetric Video Streaming. In ACM MobiCom, 2020.

Image Super Resolution

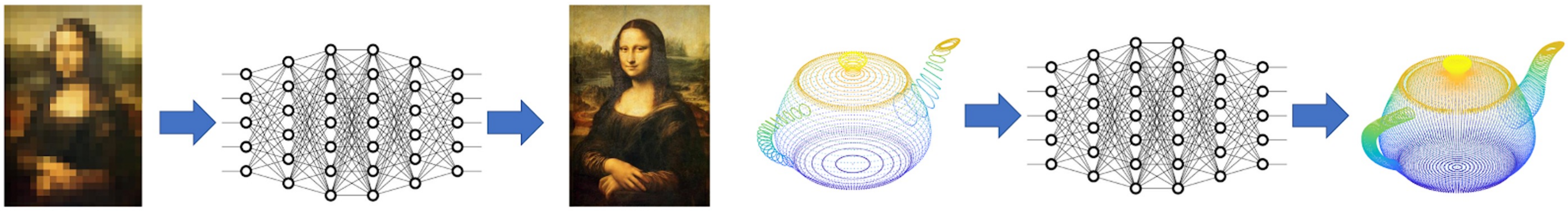


- Enhance the visual quality（resolution）of Images with Conv DNN
- Inference speed: ~30FPS

Image source: “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks”

Point Cloud Super Resolution

- Enhance the point cloud density with Conv DNN
- Advantages:
  - Reduce the bandwidth by ~ 75% (when SR ratio = 4)
  - Achieve reasonable reconstruction error

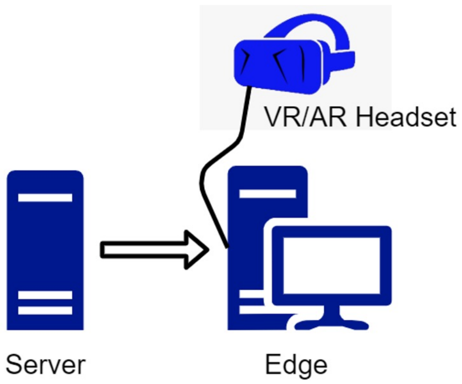


Pilot Study

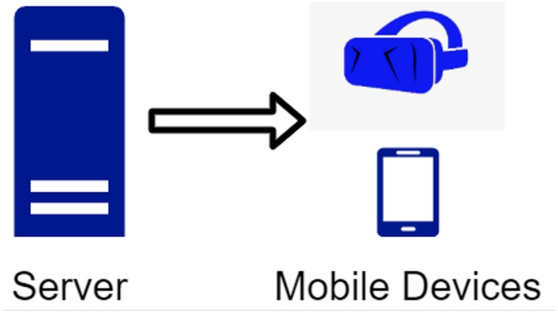
- Experiment Setup:
  - Video: A person giving a talk, 3600+ frames with ~100k points/frame
  - Device: A PC with an NVIDIA 2080 Ti GPU
  - Model: PU-GAN[1]
  - SR ratio: x4, 25k->100k
- Limitation:
  - Off-the-shelf 3D SR model achieve 0.1 FPS
  - Visual inconsistency
  - Lack of color support

[1] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. PU-GAN: A Point Cloud Upsampling Adversarial Network. In ICCV, 2019.

Target Scenario



Edge-assisted VR/AR system



Standalone mobile VR/AR

VoluSR: Efficient Volumetric Video Streaming Through Super Resolution

- Speeding up 3D SR
- Improving Visual Consistency across Frames
- Coloring SR Results
- Adapting to Network and Computation Resource

Speeding up 3D SR

- Simplifying 3D SR model
  - Profile the inference time of an off-the-shelf 3D SR model (e.g. PU-GAN[1])
  - Principle: reduce inference time, maintain good accuracy
    - Prune the layers that contribute little to the final result
    - Replace with more efficient operations

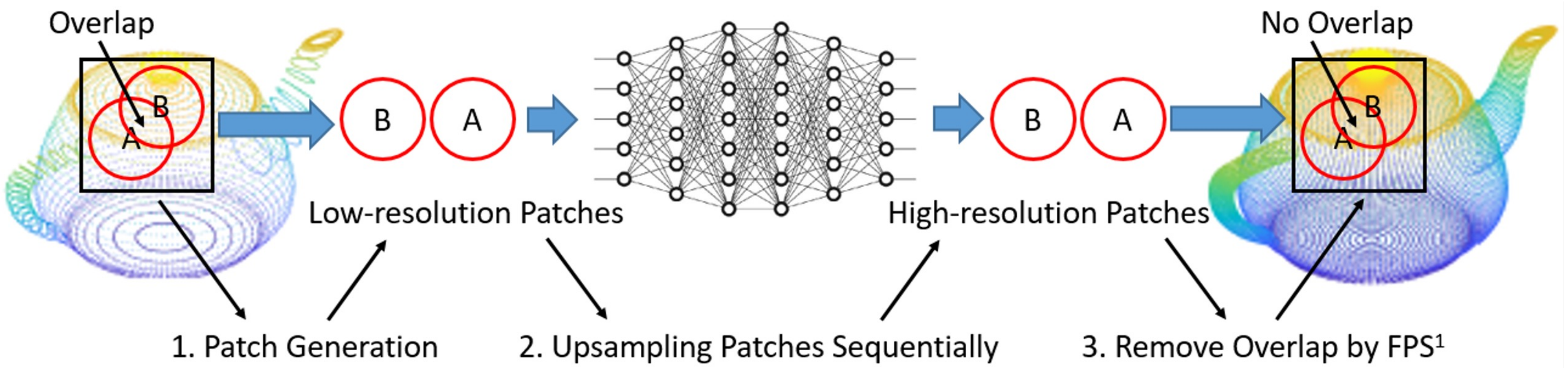
	Feature Extraction	Feature Expansion	Point Set Generation
% Time	78.3%	19.3%	2.4%

Table: Profile the Inference time of the PU-GAN[1] model

[1] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. PU-GAN: A Point Cloud Upsampling Adversarial Network. In ICCV, 2019.

Speeding up 3D SR

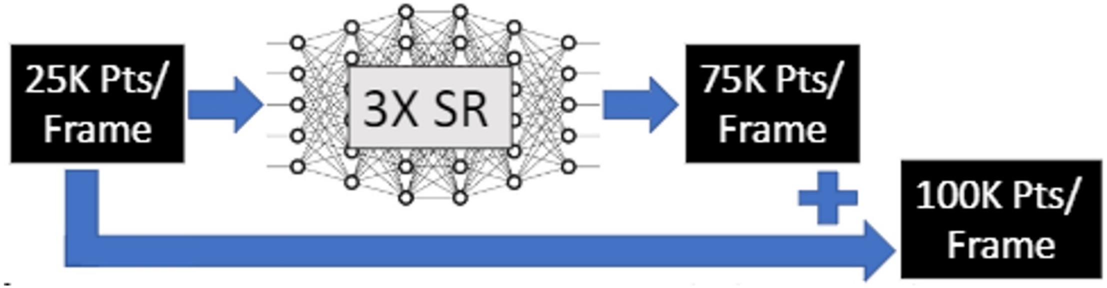
- Optimizing patch generation
  - Patch generation in an off-the-shelf 3D SR process
    - Overhead comes from overlaps
  - Method: Simplify the patches’ geometry shapes to remove overlaps
    - E.g. dividing the space into small cubic grids



FPS[1]: Farthest Point Sampling

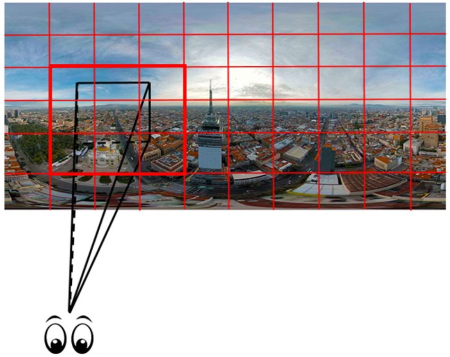
Speeding up 3D SR

- Overlaying SR input on SR output
  - A point cloud is a set of unstructured points → easy to merge two point clouds
  - SR's output points are different from the input

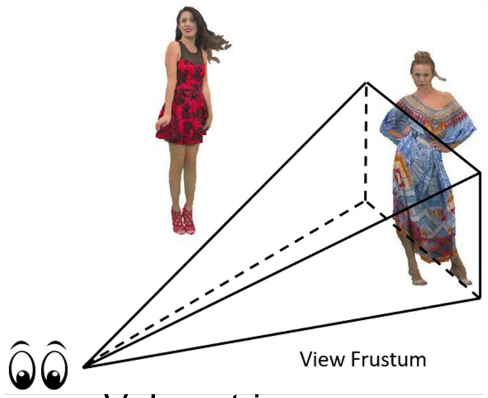


Speeding up 3D SR

- Adapting to user's viewport
  - 360-degree video: longitude/latitude
    - Only streaming cells in/around the predicted viewport
  - Volumetric video: 6-DoF, more challenging
    - Only upsample patches are predicted to fall into view frustum[1]



360-degree Video

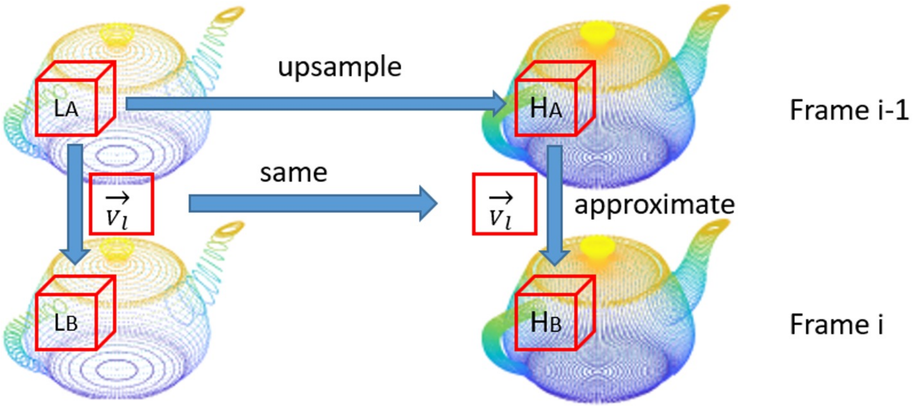


Volumetric Video

360-degree video image source: https://360labs.net/blog/vr-video-formats-explained
Volumetric video model source: http://plenodb.jpeg.org/pc/8ilabs
View Frustum[1]: In 3D computer graphics, the view frustum is the region of space in the modeled world that may appear on the screen.

Improving Visual Consistency across Frames

- Consider visual consistency in the loss function of SR model
- Reuse SR results
  - A “motion vector” to approximate the transition
    - Patches in the same position in consecutive frames
    - Smooth
    - Less computation overhead
    - Patches should have similar geometry shapes



Coloring SR Results

- Modify the SR model
  - Inferring the color components
  - Accurate but heavyweight
- Approximate a point’s color, e.g. through interpolation
  - Using the low-resolution frame
  - Lightweight but less accurate

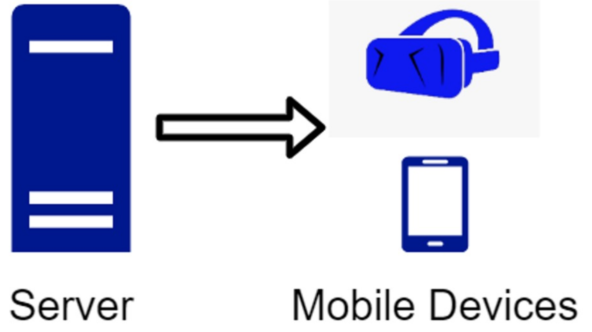
Adapting to Network and Computation Resource

- Traditional video streaming
  - Adaptive bitrate (ABR) algorithm
  - Adapt video quality to network condition
- SR-enhance volumetric video streaming
  - Not only network condition, but also computation resource
    - Trade-off
  - How to model user’s Quality of Experience (QoE)?
    - To guide our adaptation algorithm
    - Involve more factors compared to 2D videos’ QoE
      - Point density
      - Viewing distance
      - SR distortion
      - ...

Evaluation A Part of Our Optimizations

- Scenario: Mobile-only streaming
- Evaluation setup
  - A Jetson TX2 embedded system board with a 256-core NVIDIA Pascal GPU
  - Upsample a video from 5k to 20k points/frame (SR ratio = 4)
    - The video used in our pilot study
  - 3D SR model: PU-GAN[1]
    - Simplifying the model and patch generation
- Results

	Vanilla	Optimized
Memory Usage	4760	1138
Frames per second (FPS)	0.2	10.7
Average Accuracy (cm)	3.50	2.55



Mobile-only Streaming

[1] *R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. PU-GAN: A Point Cloud Upsampling Adversarial Network. In ICCV, 2019.*

Recap & On-going Work

- *VoluSR*: Efficient Volumetric Video Streaming Through Super Resolution
  - Speeding up 3D SR
  - Improving Visual Consistency across Frames
  - Coloring SR Results
  - Adapting to Network and Computation Resource
- Preliminary results
  - Runtime performance is improved greatly
  - Good SR inference accuracy is achieved
- Currently we are developing the whole system

University of Minnesota
Driven to Discover

George Mason University

Efficient Volumetric Video Streaming Through Super Resolution

Anlan Zhang[1], **Chendong Wang**[1], Bo Han[2], Feng Qian[1]

[1] University of Minnesota, Twin Cities
[2] George Mason University