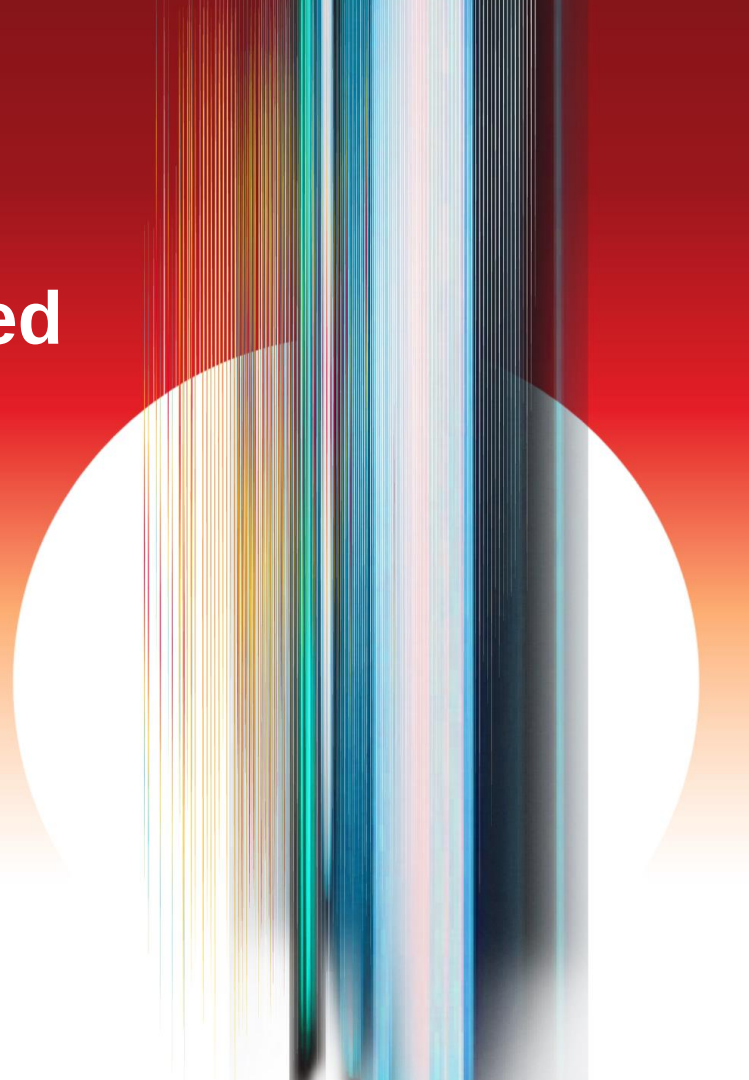
SIGGRAPH ASIA 2021 TOKYO

LIVE

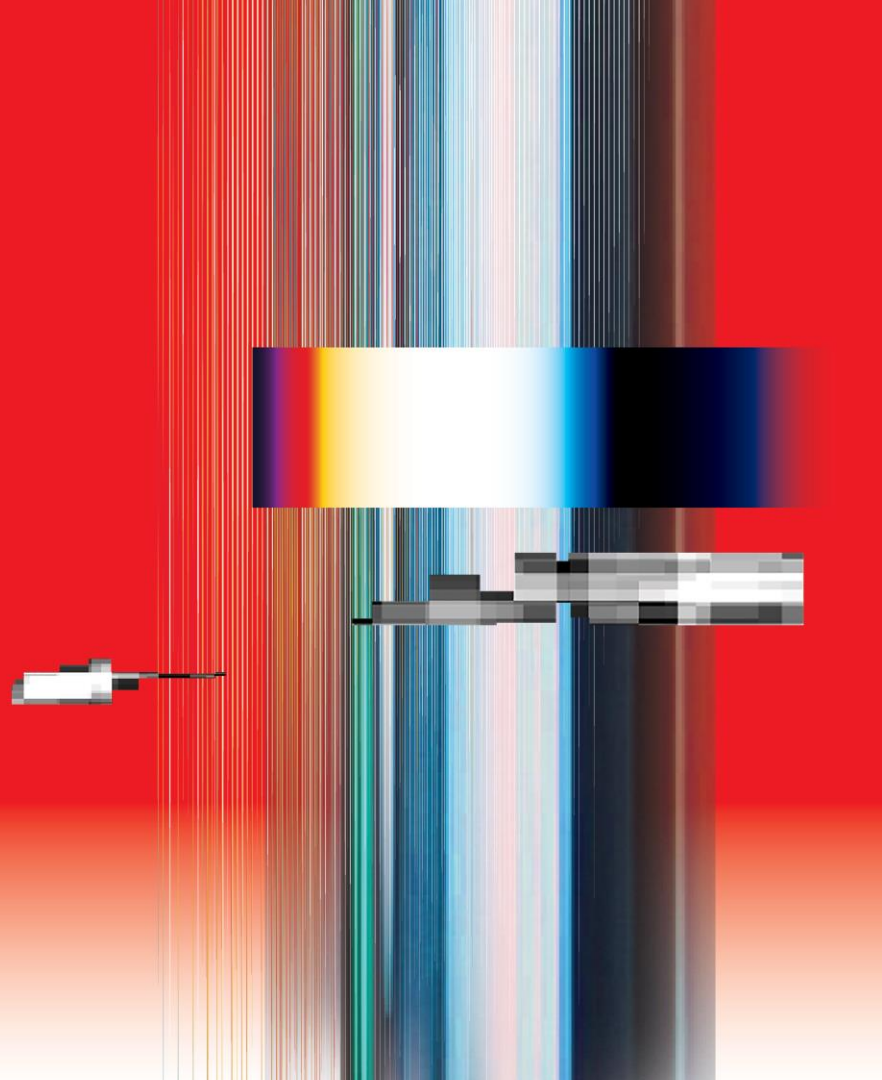
ExtraNet: Real-time Extrapolated Rendering for Low-latency Temporal Supersampling

Jie Guo[1] Xihao Fu[1] Liqiang Lin[1]
Hengjun Ma[1] Yanwen Guo[1]
Shiqiu Liu[2] Ling-Qi Yan[3]

[1]Nanjing University, [2]NVIDIA Corporation
[3]University of California, Santa Barbara



Background



- Modern real-time rendering applications require more and more realistic graphics.
- Real-time ray tracing technology

↓

Maintaining high resolution at 60fps stably with complicated shading is difficult.

↓

Lowering computation cost to increase frame rate

Spatial Supersampling

Rendering at low resolution, and upscaling to high resolution

NVIDIA	AMD
Deep Learning Super Sampling(DLSS)	Fidelity Super Resolution(FSR)
Applying neural network on resolution upsampling	Non-DL Upscaling +Sharpening

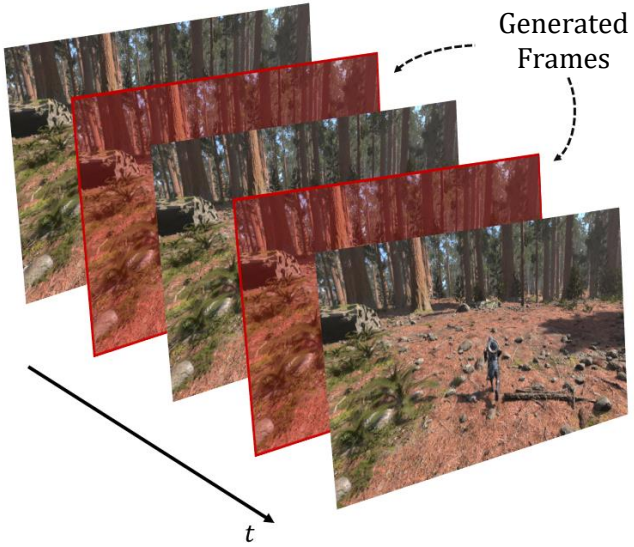
Temporal Denoising/Anti-Aliasing

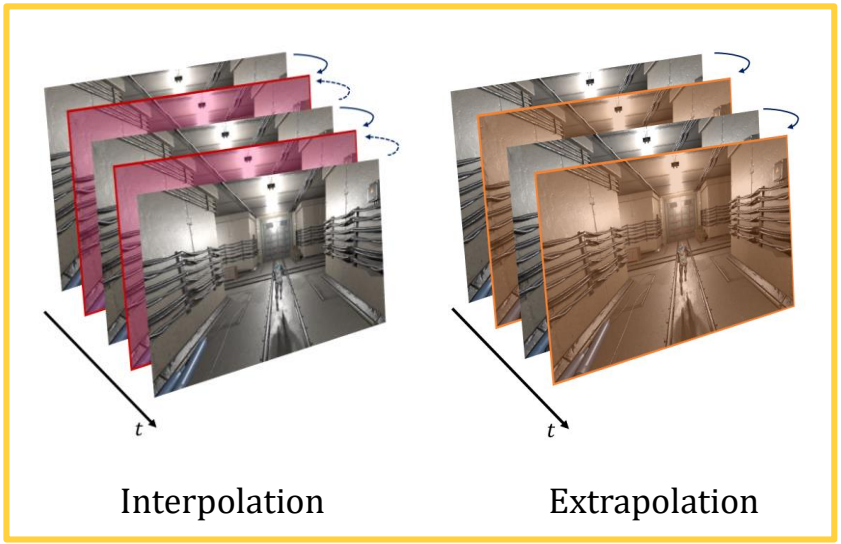
Shading at low samples and upsampling with temporal filtering

- Shading is still necessary, at least 1spp.

Temporal Supersampling

Generating new frames with existing rendered results





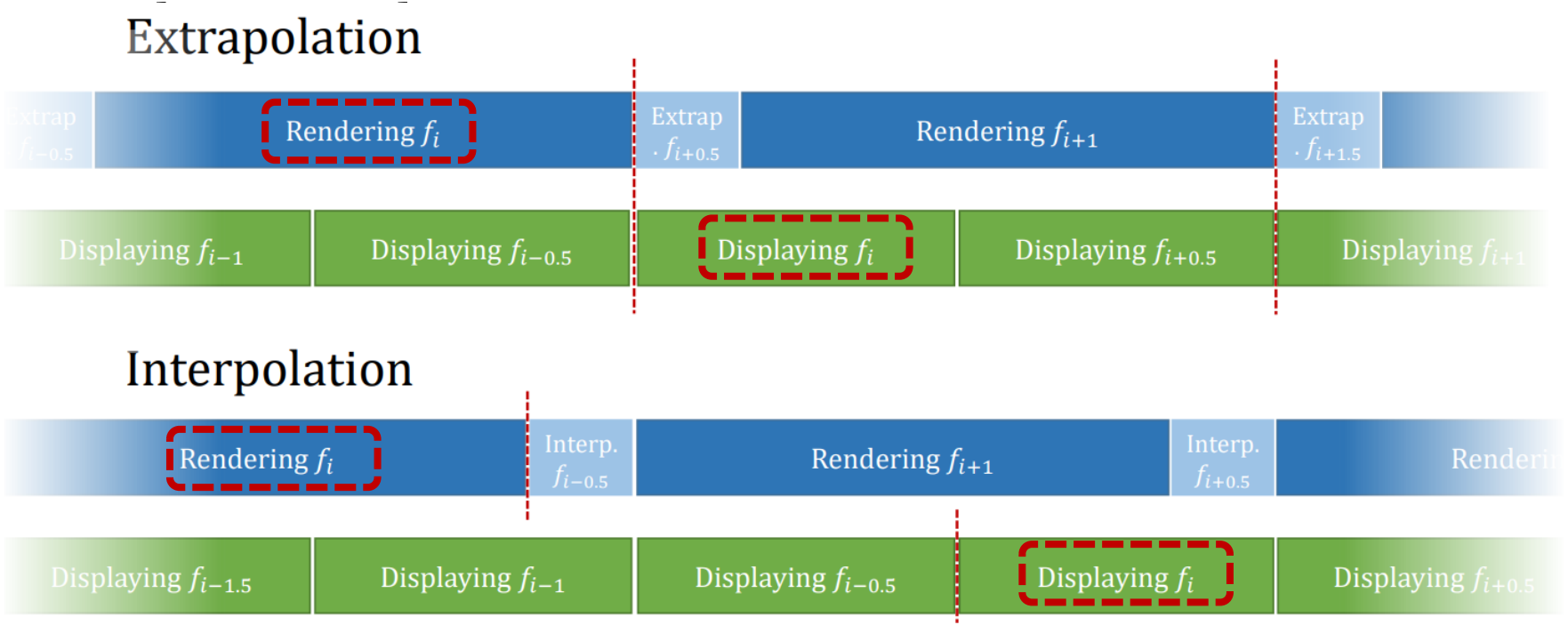
Temporal Supersampling

Image Warping

Uni- or Bi-directional Warping
[Schollmeyer et al. 2017, Yang et al. 2011, Mark et al. 1997]

- Without considering dynamic shading change

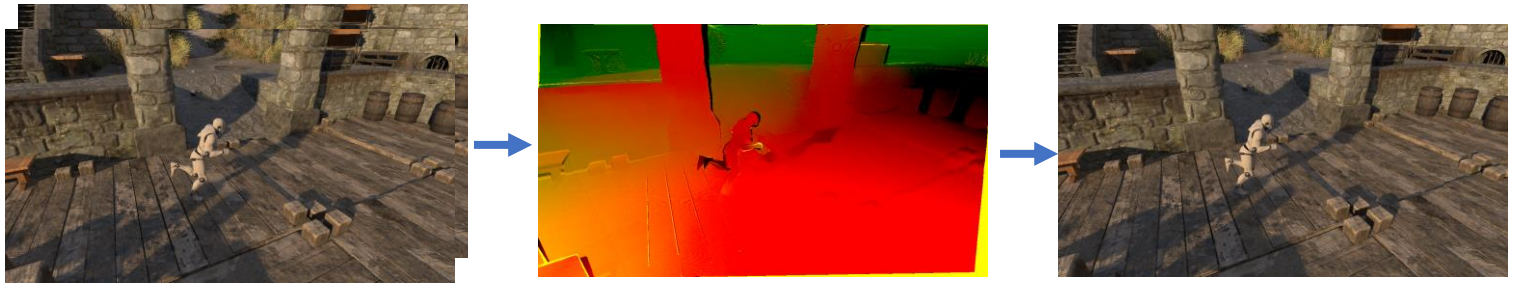
Temporal Supersampling



Temporal Supersampling

ASW(2.0) [Oculus 2016]

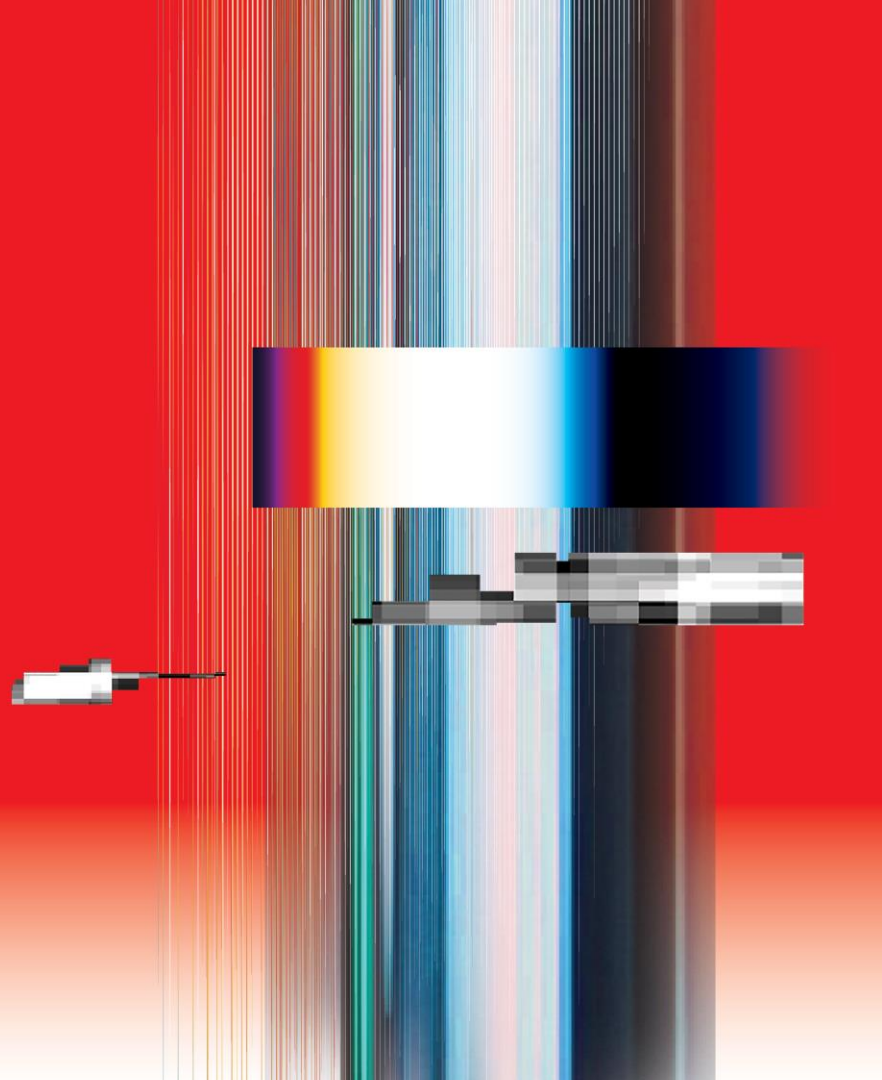
- (MAYBE) using optical flow of previous frames to extrapolate a new frame



$f_{i-1}\&f_i$ OF_i $f_{i+0.5}$

- Distortion when optical flow fail to capture accurate motion

Our Work



Contribution

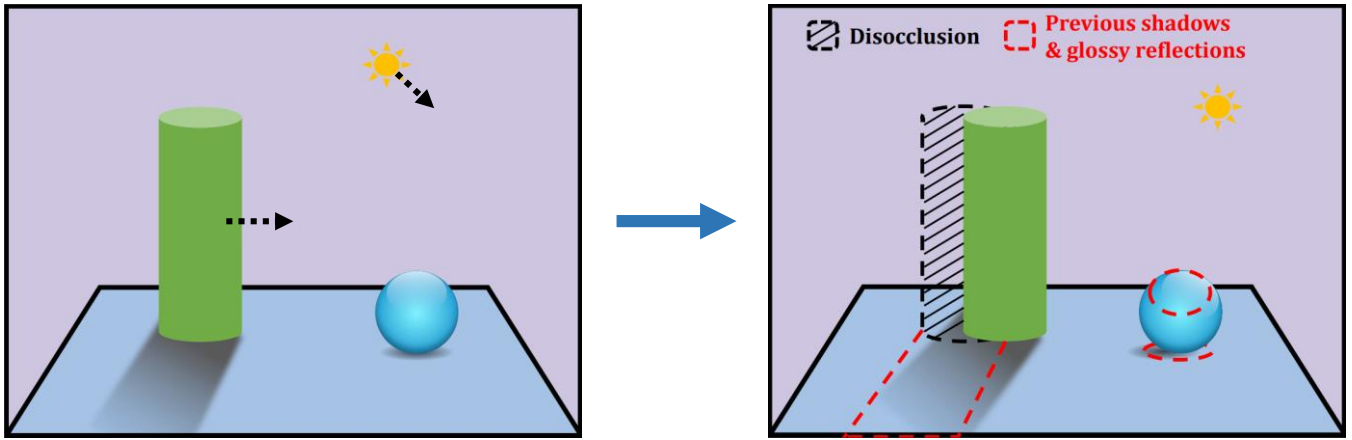
- A real-time rendering and extrapolation architecture for increasing frame rate
- A neural network that utilizes G-buffers and previous rendered results as input to perform extrapolation
- Low performance overhead with high quality output

Versus related work:

- No shading at extrapolated frames (**Minimizing computation**)
- Smooth extrapolated sequences (**High quality**)
- No necessity for future frames (**No extra latency**)
- Using motion vector rather than optical flow (**No distortion**)

Challenges

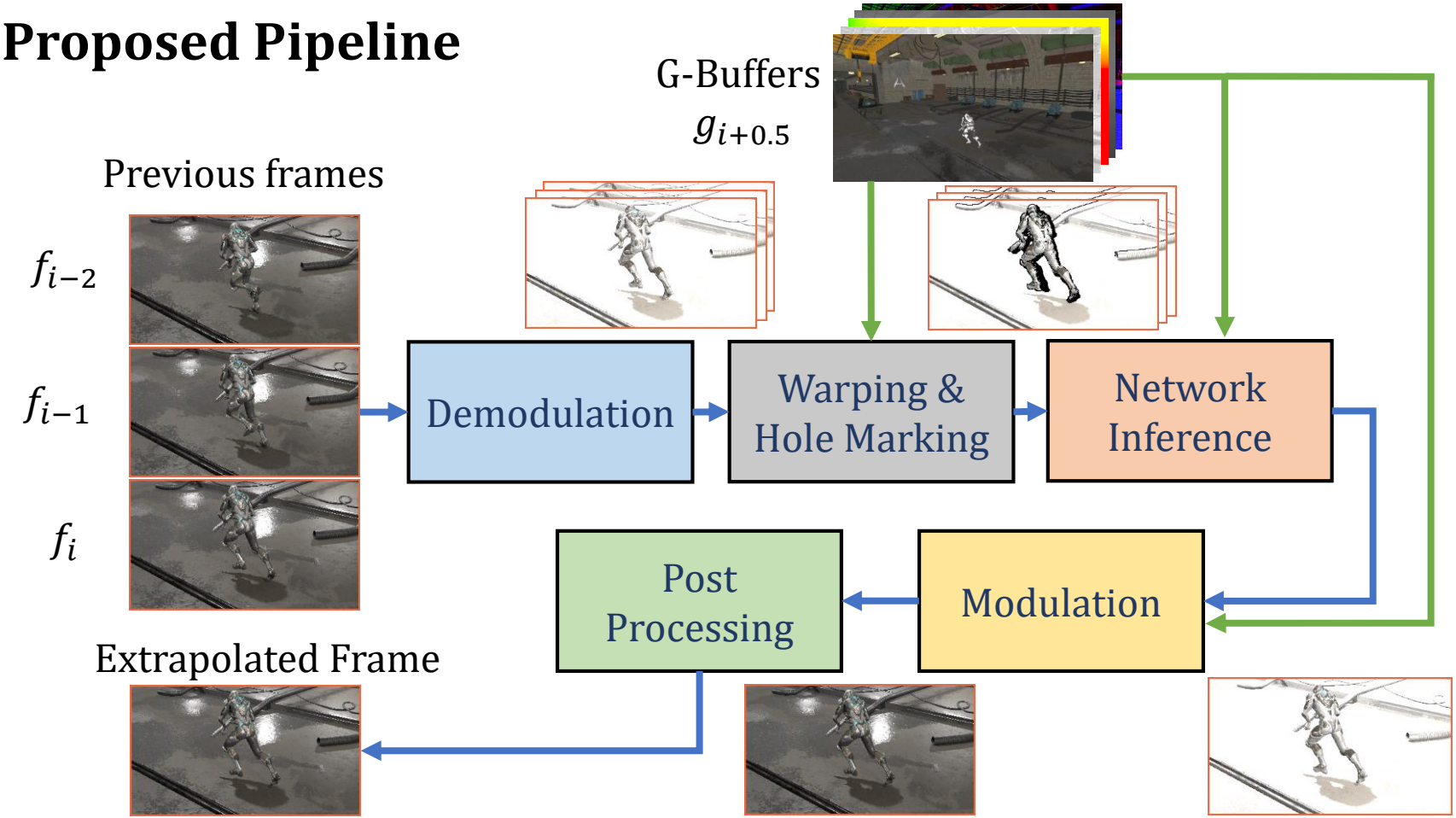
How we do extrapolation:



f_i $f_{i+0.5}$

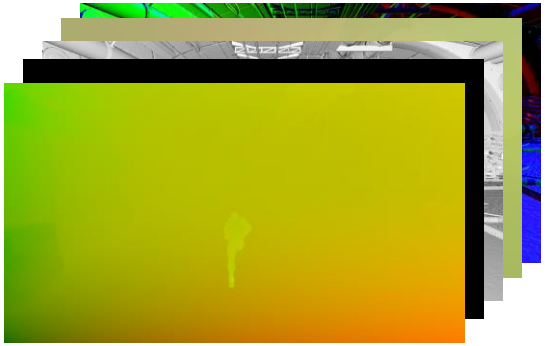
- Disocclusion
- Dynamic Shading Change
- Fast enough for real-time rendering

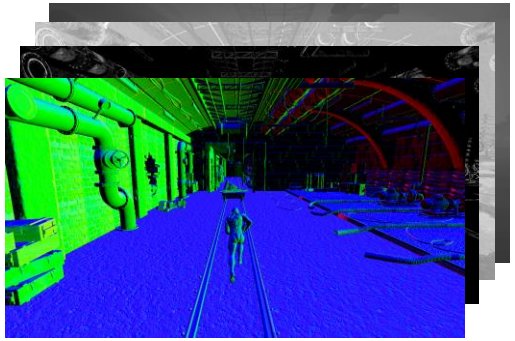
Proposed Pipeline



G-buffers

9 buffers in total





World Normal, Depth, World Position, Metallic, World Normal

used for warping and hole marking

Demodulation & Modulation

- Separate albedo and illumination
- For better disocclusion inpainting


Rendered Result

/


Albedo

=


Demodulated Result


Network Output

×


Albedo

=


Modulated Result

Warping

Using **motion vector** to gather samples from previous frames

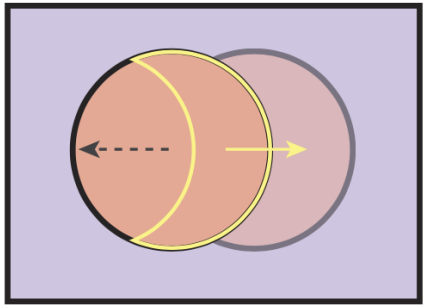
- Ghosting at wrong temporal correlation

GT

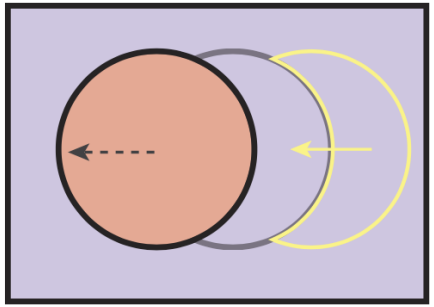
Traditional MV

Occlusion MV

Occlusion Motion Vector [Zeng et al. 2021]



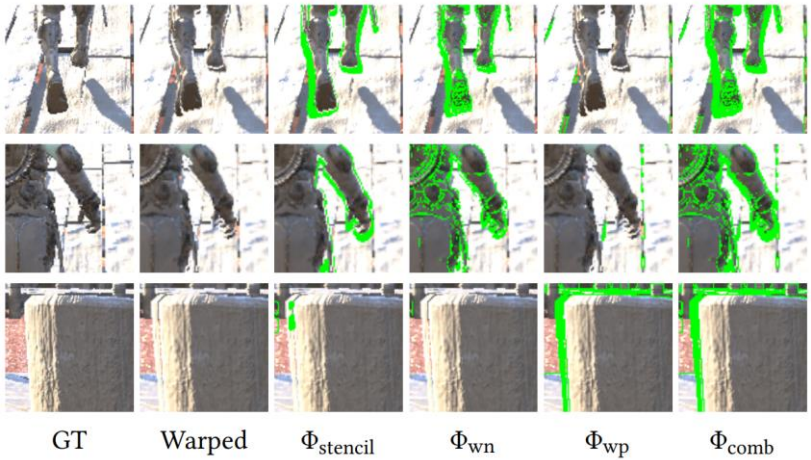
Traditional MV

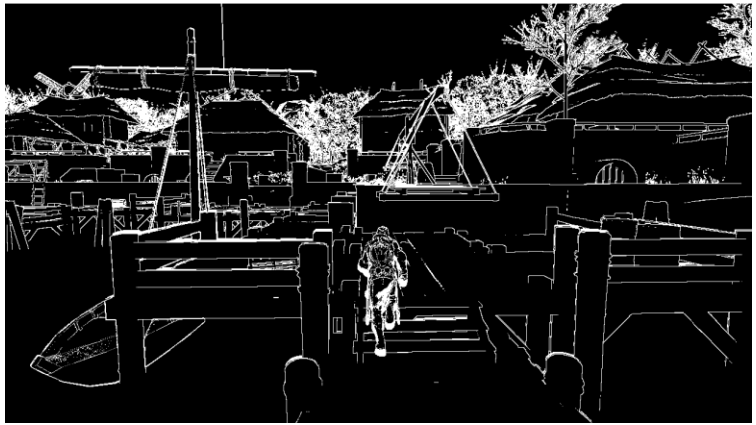


Occlusion MV

Hole Marking

- Why: Marking out the disocclusion
- How: Based on G-buffers information
  - **Stencil** of moving objects (dynamic objects)
  - **World Normal** (self occlusion of dynamic objects)
  - **World Position** (static objects)





Example Mask

Scenes & Data

Four scenes from Unreal Engine 4:

MD

RF

BK

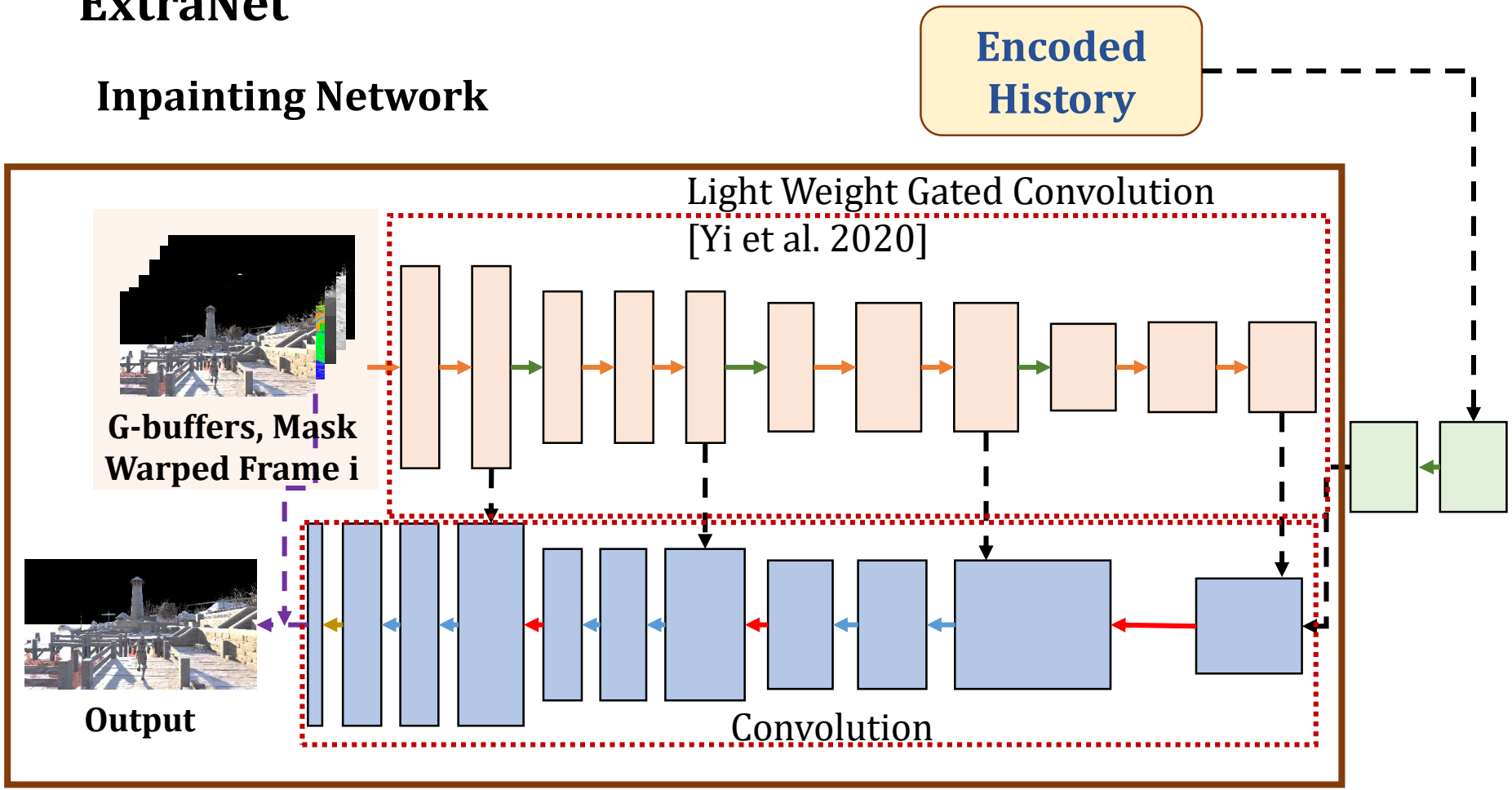
WT

Training and testing setup:

	Model 1	Model 2	Model 3	Model 4
Training Scene(s)	MD	RF	BK	MD+RF+BK
Testing Scene	MD	RF	BK	WT

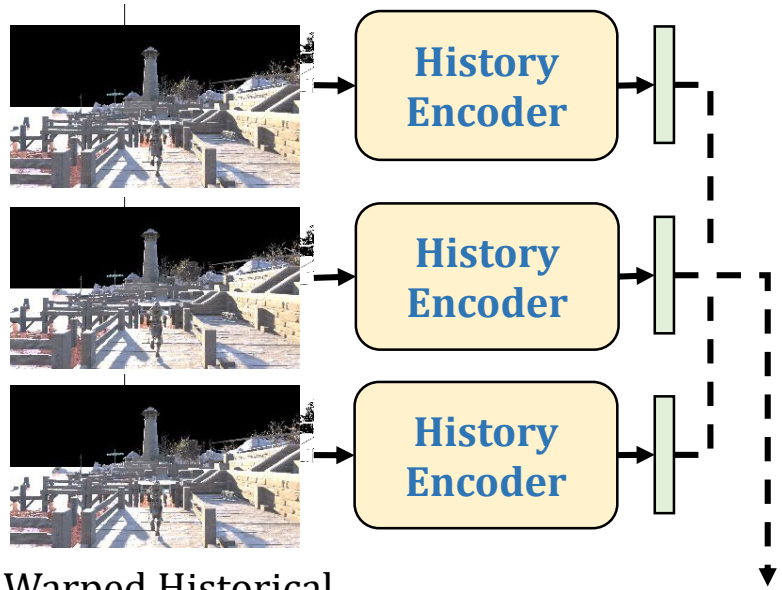
ExtraNet

Inpainting Network

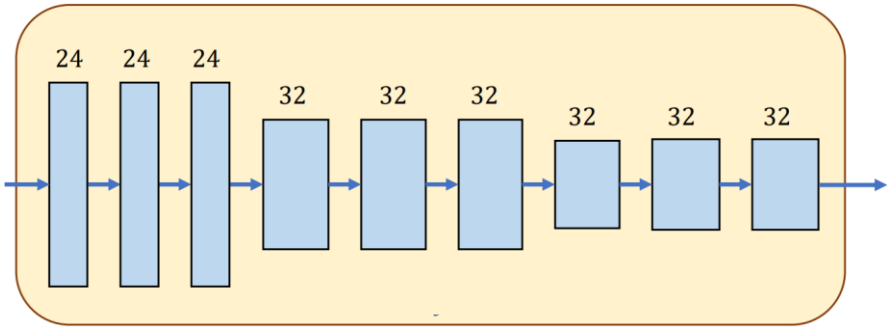


ExtraNet

History Encoder



Warped Historical Frames and Masks (Frame i, i-1 and i-2)



History Encoder Architecture

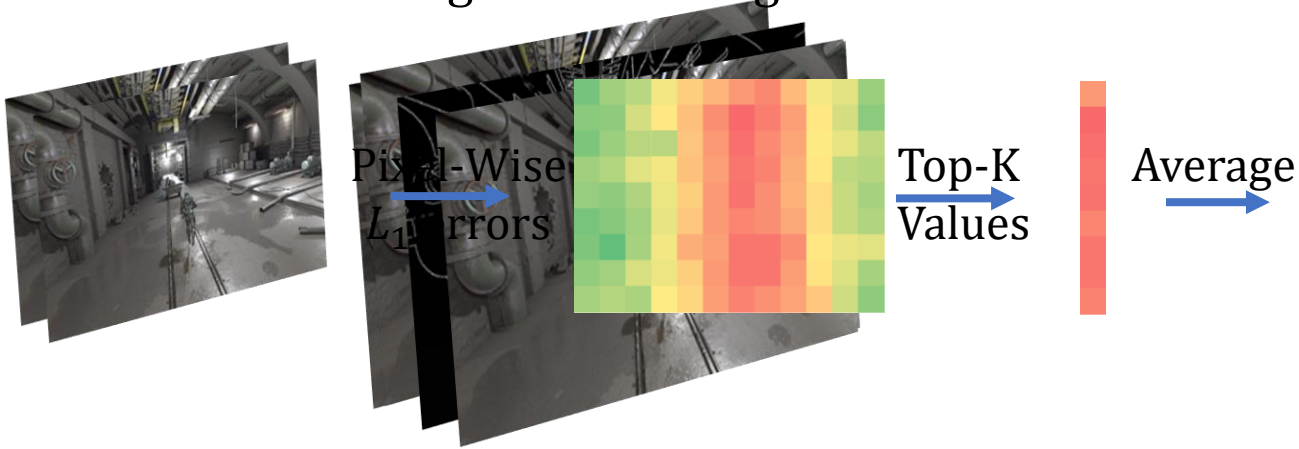
ExtraNet

Training Loss

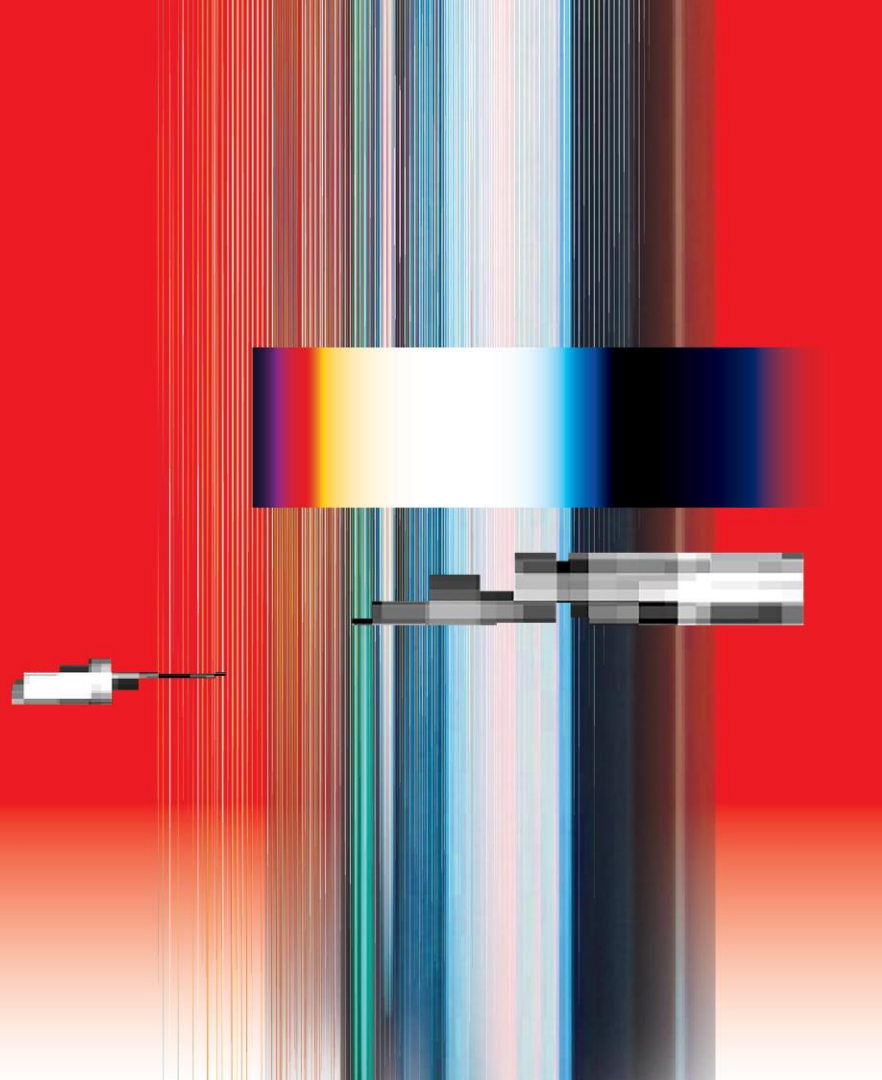
$$\mathcal{L} = \mathcal{L}_{l_1} + \lambda_{\text{hole}}\mathcal{L}_{\text{hole}} + \lambda_{\text{shade}}\mathcal{L}_{\text{shade}}$$

L_1 distance between reference and output

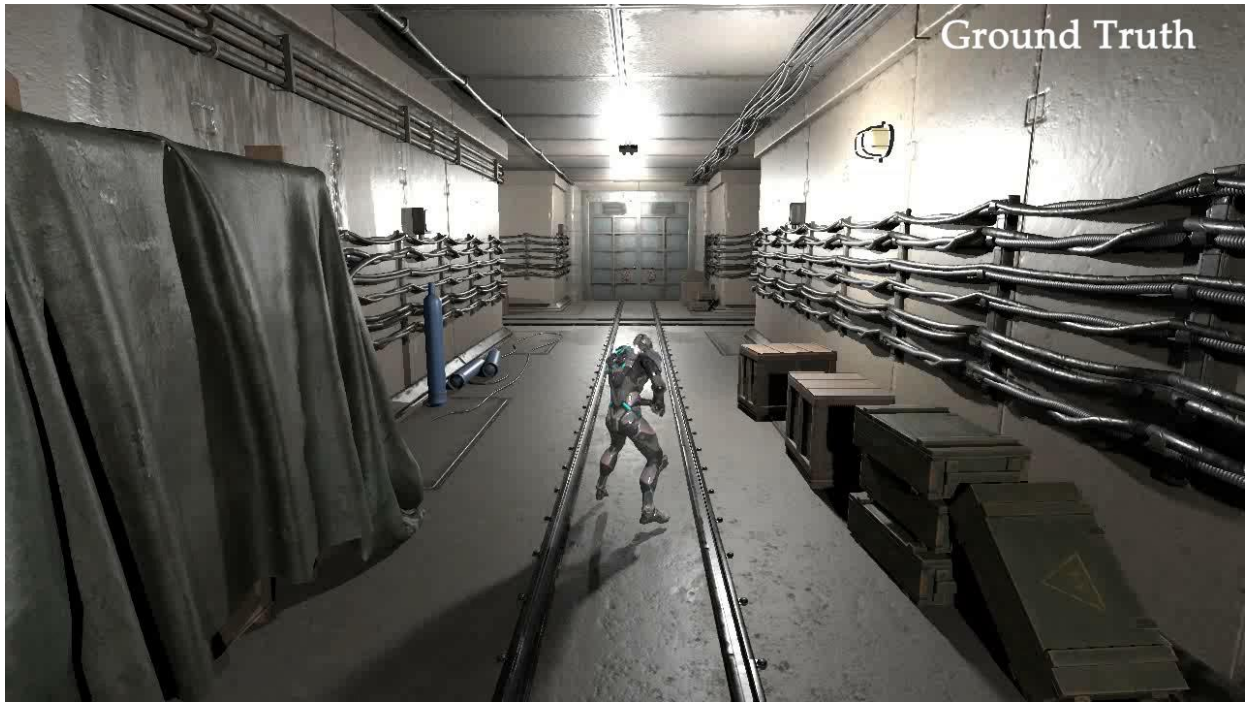
L_1 distance of the top-k largest errors region



Results



Extrapolated Results



Versus Image Warping Extrapolation

HIW[Schollmeyer et al. 2017]

Versus Image Warping Interpolation

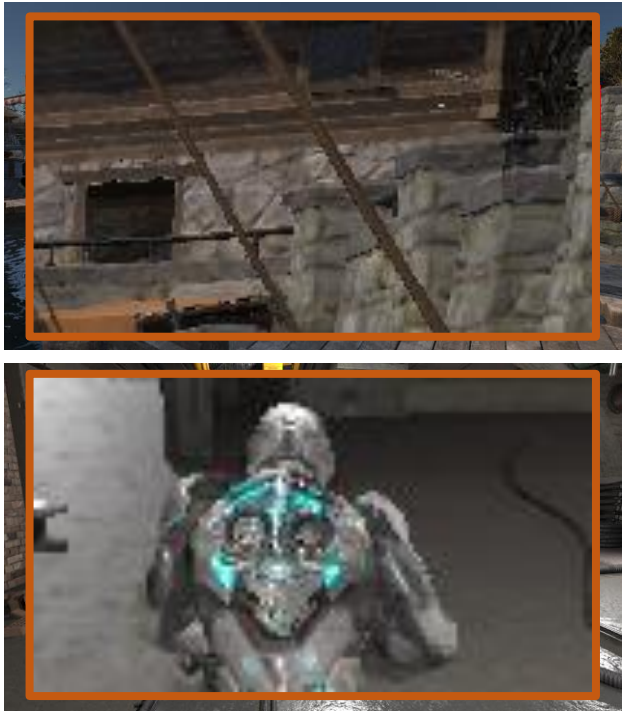
BSR[Yang et al. 2011]

Versus Image Warping Interpolation

3DWarp[Mark et al. 1997]

Versus Video Interpolation

Missing geometry structure



Ours

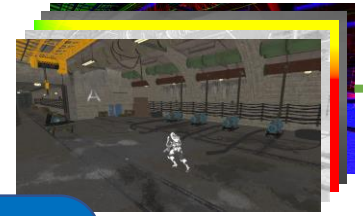
DAIN[Bao et al. 2019]

Versus ASW

ASW fails to compute accurate optical flow under complicated conditions.

ASW GT Ours

Runtime Performance



G-Buffers

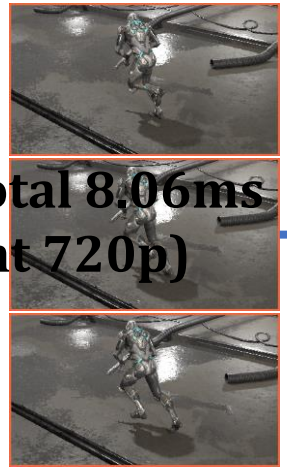
$g_{i+0.5}$

Previous frames

f_{i-2}

f_{i-1}

f_i



Total 8.06ms (at 720p)

G-Buffers Rendering 1.32ms

Demodulation

Warping & Hole Marking

Network Inference

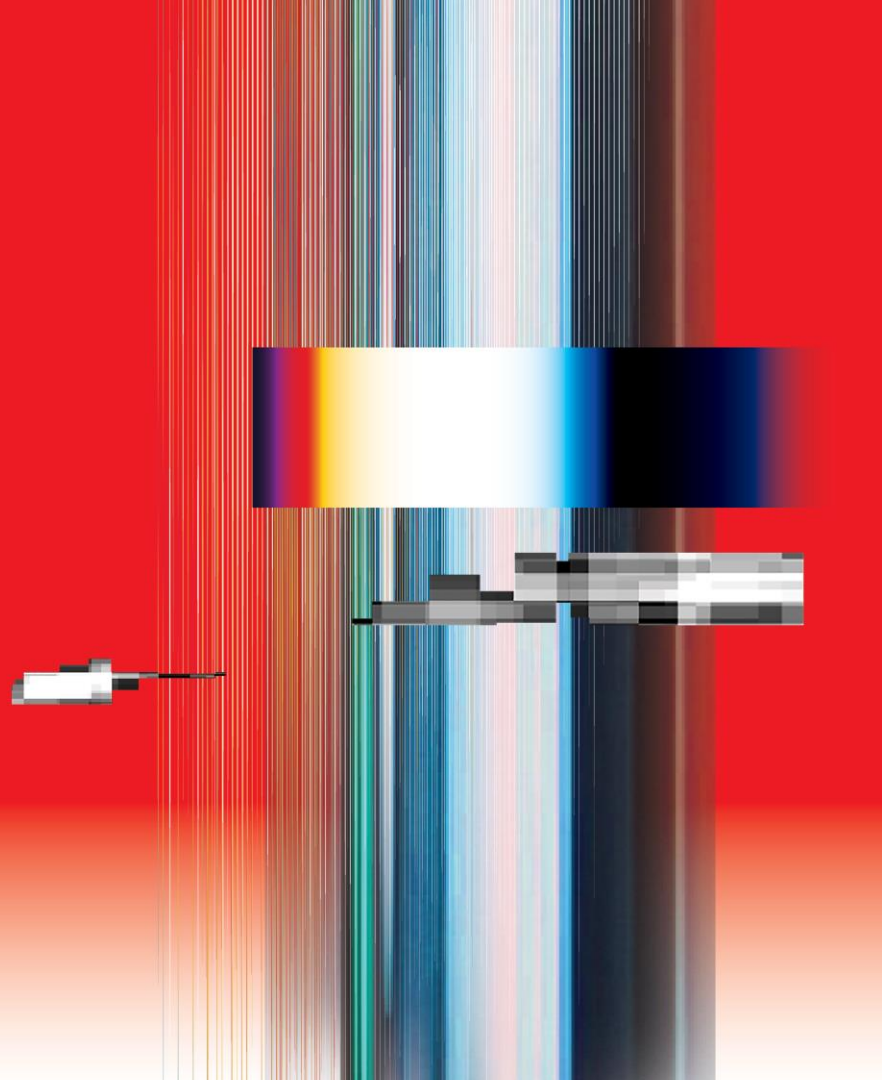
Warping & Hole Marking 1.35ms

Network Inference 5.39ms

Extrapolated Frame

G-Buffers rendering cost is averaged on all of our test scenes

Limitations & Discussions



Out of Screen Disocclusion

When camera rotates at fast speed:

Ground Truth

Ours

Blurry Shadow

The output shading of ExtraNet tends to be blurry.

Ground Truth

Ours

- By applying **temporal** anti-aliasing, it is hard to notice the blurriness in a real game play.

Temporal Discontinuity

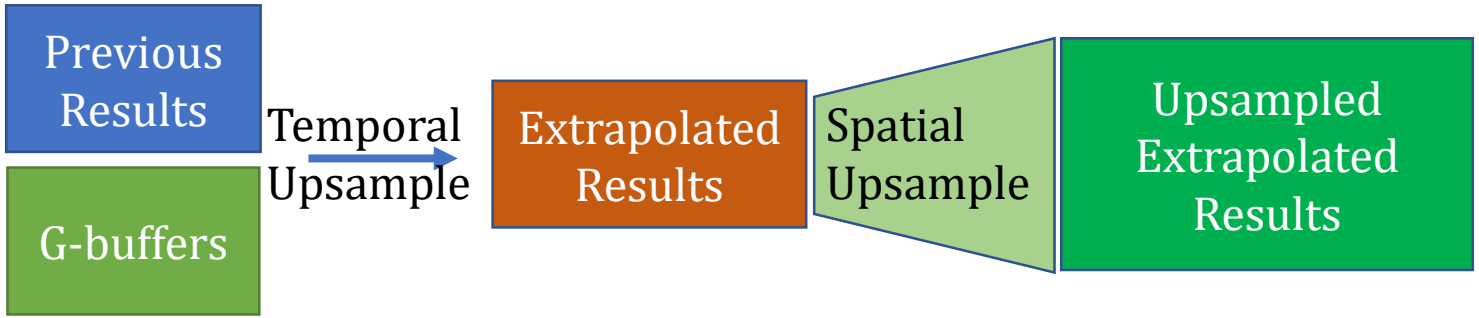
Sudden change of light:

Ground Truth

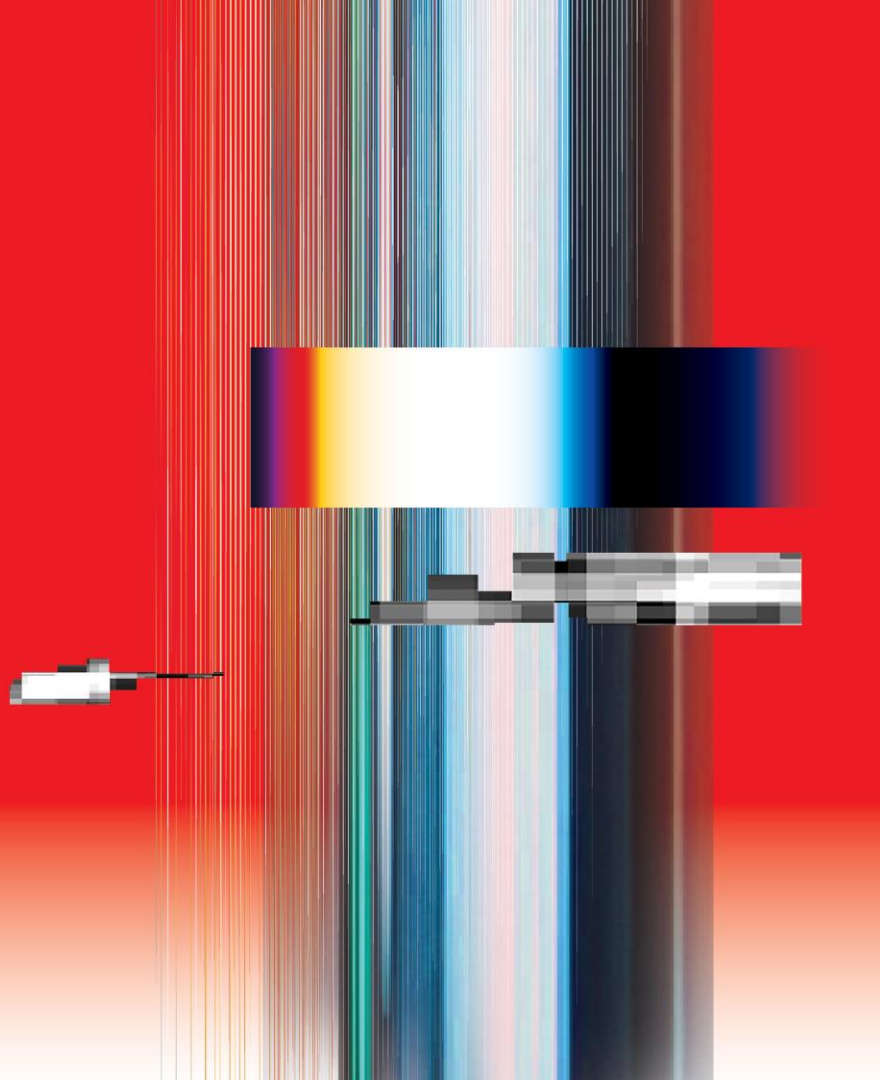
Ours

Boosting Performance

- **CUDA implementation of network inference**
  - Currently TensorRT is used;
  - Handy-crafted CUDA kernels by utilizing tensor cores may be faster.
- **Combining spatial and temporal supersampling**
  - Embedding our pipeline into resolution super-sampling framework;
  - Lowering G-buffers and inference cost.



Conclusion

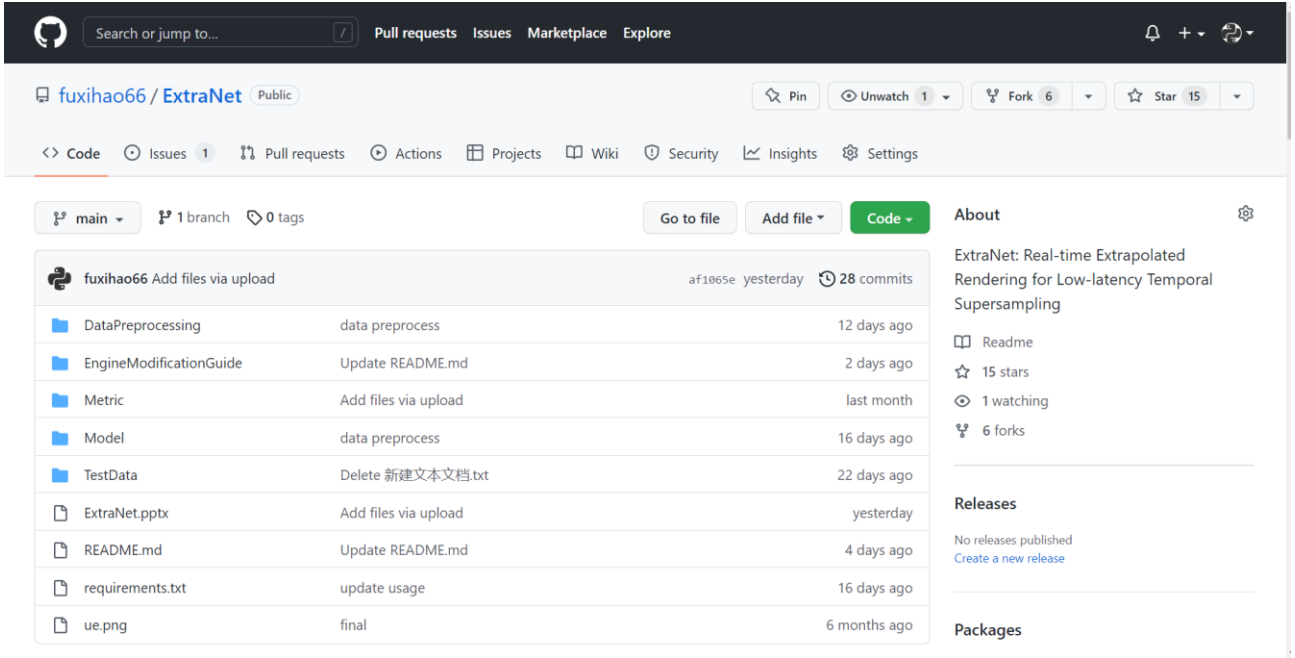


Conclusion

- Necessity for increasing frame rate
- An extrapolation pipeline, low latency
- Disocclusion & Dynamic shading change
- Fast algorithm with high quality results

Source Codes:

https://github.com/fuxihao66/ExtraNet



SIGGRAPH ASIA 2021 TOKYO

CONFERENCE 14 - 17 DECEMBER 2021
EXHIBITION 15 - 17 DECEMBER 2021
TOKYO INTERNATIONAL FORUM, JAPAN

sa2021.siggraph.org

LIVE

Sponsored by

Organized by

koelnmesse
we energize your business | since 1924