SIGGRAPH ASIA 2023 SYDNEY

FuseSR: Super Resolution for Real-time Rendering through Efficient Multi-resolution Fusion

Zhihua Zhong[*1,4], Jingsen Zhu[*1], Yuxin Dai[3], Chuankun Zheng[1], Yuchi Huo[†2,1], Guanlin Chen[4], Hujun Bao[1], Rui Wang[†1]

[1]State Key Lab of CAD&CG, Zhejiang University, [2]Zhejiang Lab, [3]Zhejiang A&F University, [4]Zhejiang University City College

Connecting STORIES

The 16th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia

CONFERENCE 12 - 15 December 2023
EXHIBITION 13 - 15 December 2023
ICC, Sydney, Australia
ASIA.SIGGRAPH.ORG/2023

Sponsored by acm

Organized by koelnmesse

SIGGRAPH ASIA 2023 SYDNEY

Background

Super Resolution for Real-time Rendering

Connecting STORIES

The 16th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia

CONFERENCE 12 - 15 December 2023
EXHIBITION 13 - 15 December 2023
ICC, Sydney, Australia

ASIA.SIGGRAPH.ORG/2023

Sponsored by

Organized by koelnmesse

Super Resolution for Real-time

Super Resolution

- ill-posed
- poor information input ⇒rich information output

Solution

poor information input $\Rightarrow$ rich information output

poor information input $\Rightarrow$ rich information output

extra input $\Rightarrow$

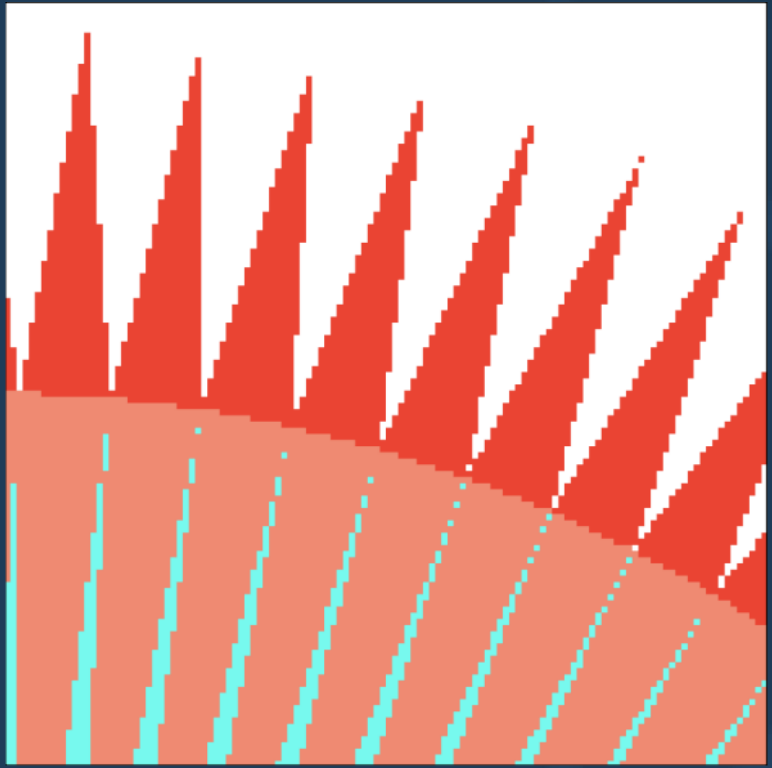
SIGGRAPH ASIA 2023 SYDNEY

Input Information Utilization Efficiency

- Interpolation
- Self-similarity
- Heuristics: sharpen, back-projection, edge-detection
- Neural network

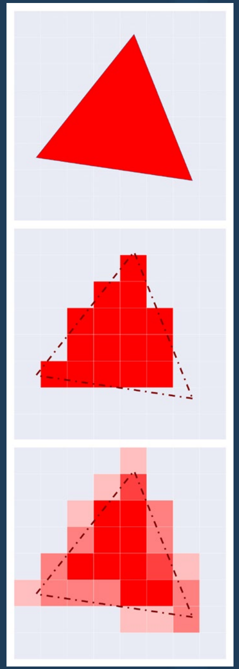
Anti aliasing

- Under sampling
- Reconstruction w/o adding sampling frequency



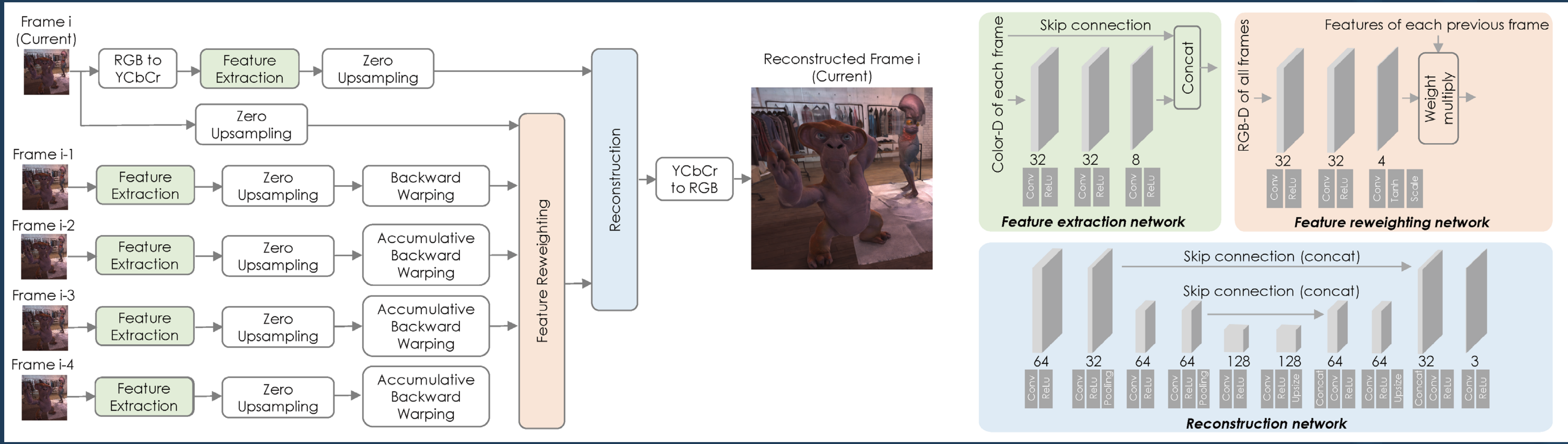
Anti aliasing

- MSAA : adding sampling frequency selectively
- FXAA : edge-detection
- TAA : temporal additional information



Previous Method and Challenge

Previous work



NSRR [SIGGRAPH’20]

SIGGRAPH ASIA 2023 SYDNEY

Our method: FuseSR

Our method: FuseSR

LR image

Auxiliary Features
G-Buffers
Historical Frames
…

SR Network

HR image

Previous Method and Challenge

Neural Supersampling for Real-time Rendering

LEI XIAO, SALAH NOURI, MATT CHAPMAN, ALEXANDER FIX, DOUGLAS LANMAN, and ANTON KAPLANYAN, Facebook Reality Labs

Fig. 1. Results of our real-time, learned 4 × 4 supersampling are shown for four sample scenes. From top to bottom: the rendered low-resolution color input, our reconstruction, and the rendered reference images. Our supersampling method takes the color, depth, and motion vectors of multiple low-resolution frames, and produces high-fidelity reconstructions by reducing aliasing and recovering scene details.

Due to higher resolutions and refresh rates, as well as more photorealistic effects, real-time rendering has become increasingly challenging for video ... scenario, significantly outperforming existing superresolution and temporal antialiasing work.

NSRR [SIGGRAPH'20]

- Previous Method: Utilizing historical frames as auxiliary features
  - ☹ Insufficient information
  - ☹ Struggles in dynamic scenes
- Ours: Utilizing HR G-buffers as auxiliary features
  - ☺ High-frequency details
  - ☺ Easy and fast to acquire

SIGGRAPH ASIA 2023 SYDNEY

Method

Real-time Super-resolution Network

Connecting STORIES

The 16th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia

CONFERENCE 12 - 15 December 2023
EXHIBITION 13 - 15 December 2023
ICC, Sydney, Australia
ASIA.SIGGRAPH.ORG/2023

Sponsored by

Organized by koelnmesse

Problem Formulation

- Input: LR image I^{LR} with auxiliary features
- Output: Upsampled HR image $\hat{I}^{HR}$
- Auxiliary features:
  - LR G-buffer G^{LR}
  - LR historical frames $I^{history}$
  - HR G-buffer G^{HR} (we introduce)

$$\boxed{\hat{I}^{HR} = \mathbf{SR}(I^{LR}, G^{LR}, I^{history}, G^{HR})}$$

Key Designs

- BRDF demodulation

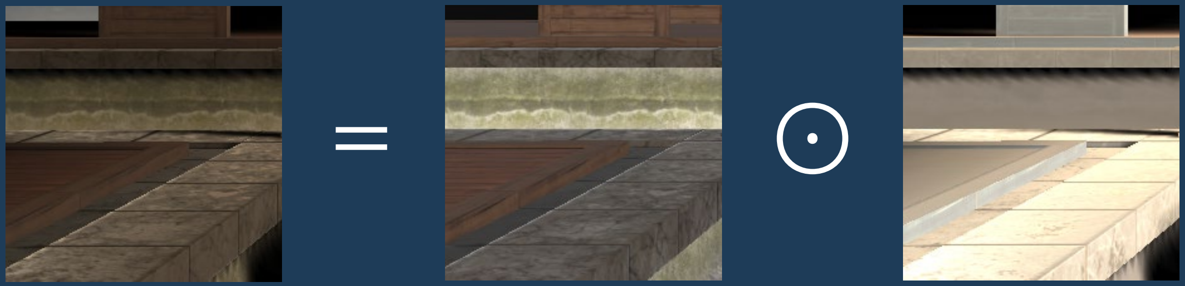
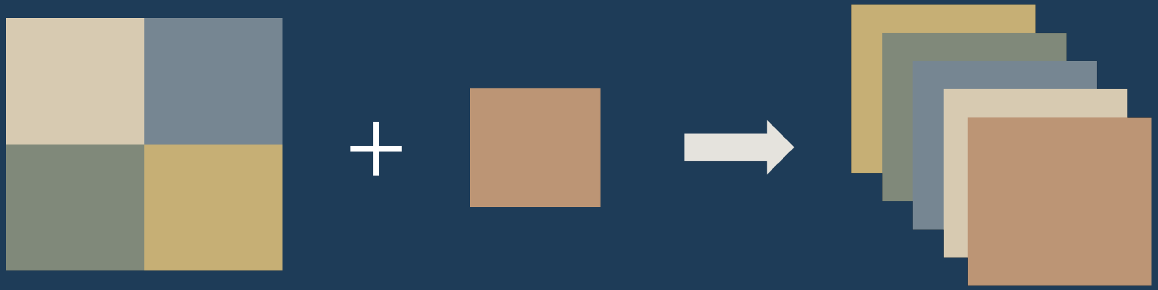


Image = BRDF ⊙ Irradiance

- Multi-resolution fusion network (H-Net)



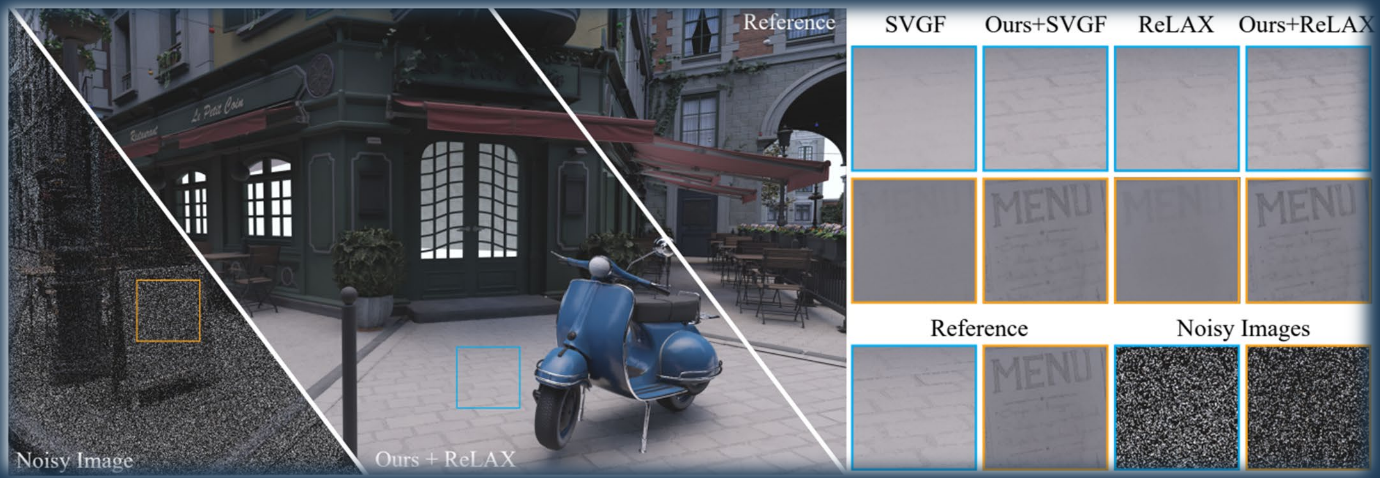
HR + LR → Fused in channels

BRDF Demodulation

- Motivation:
  - Rendering equation: $L_o(\omega_o) = \int_\Omega f_r(\omega_i, \omega_o) L_i(\omega_i) \cos\theta_i \, \mathrm{d}\omega_i$
  - Diffuse material $f_r(\omega_i, \omega_o) = f_{Albedo}$
  - Demodulation: $L_o(\omega_o) = f_{Albedo} \int_\Omega L_i(\omega_i) \cos\theta_i \, \mathrm{d}\omega_i$
  - Limitation: for diffuse material only. Unphysical-based for glossy material.

BRDF Demodulation

- Filter out high-frequency texture and material details
- Zhuang et al.'s Demodulation:
  - BRDF term $F_\beta(\omega_o) = \int_\Omega f_r(\omega_i, \omega_o) \cos\theta_i \, \mathrm{d}\omega_i$
  - Demodulated irradiance term $L_D(\omega_o) = \dfrac{L_o(\omega_o)}{F_\beta(\omega_o)}$

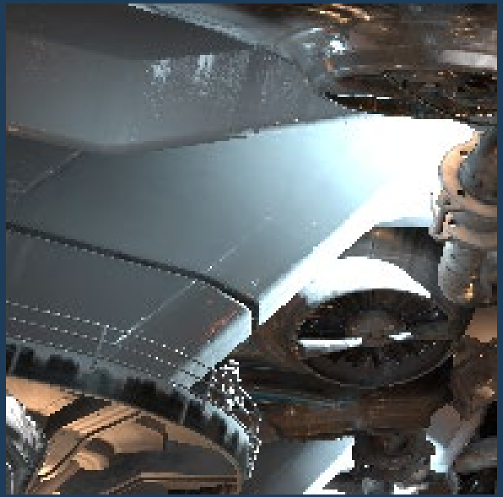


BRDF Demodulation

- High-frequency texture and material details are baked in F_β
- Demodulated irradiance term L_D becomes much smoother

Image

BRDF F_β



Irradiance L_D

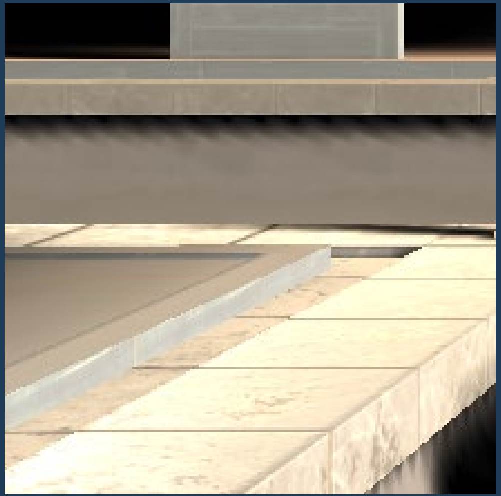
BRDF Demodulation

- High-frequency texture and material details are baked in F_β
- Demodulated irradiance term L_D becomes much smoother

Image

• High-frequency texture and material details are baked in F_β
• Demodulated irradiance term L_D becomes much smoother

BRDF F_β



Irradiance L_D

BRDF Demodulation

- BRDF term F_β can be pre-computed
- Demodulated irradiance term L_D to be predicted by network $\mathbf{\Phi}$

$$\boxed{\hat{L}_D^{HR} = \mathbf{\Phi}\left(L_D^{LR}, G^{LR}, I^{history}, G^{HR}\right), \hat{I}^{HR} = F_\beta^{HR} \odot \hat{L}_D^{HR}}$$

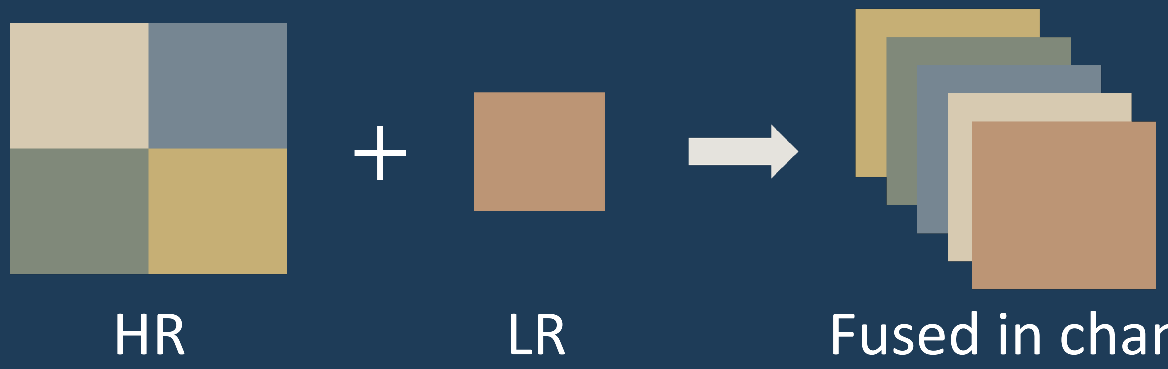
- Estimate smoother $\hat{L}_D^{HR}$ instead of $\hat{I}^{HR}$ by network

H-Net: Multi-resolution Fusion Network

- Network inputs
  - LR inputs: $\mathrm{I}^{LR}, G^{LR}, I^{history}$
  - HR input: G^{HR}
- Challenge: Inputs contain multi-resolution features
  - How to efficiently and effectively fuse them within our network?
- Naïve solutions
  - Upsampling: slow network ☹
  - Pooling: lossy, damage details ☹

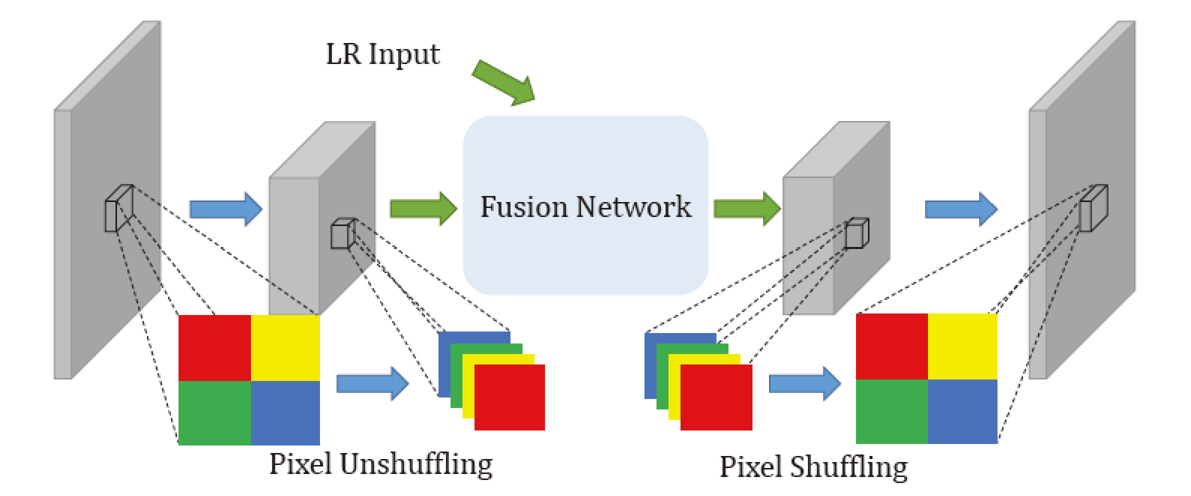
H-Net: Multi-resolution Fusion Network

- Our solution: pixel-shuffling operation
  - Unshuffle: $[C, H * r, W * r] \rightarrow [C * r^2, H, W]$
  - Shuffle: $[C * r^2, H, W] \rightarrow [C, H * r, W * r]$
  - Converting between pixel-wise spatial information and channel-wise deep information without information loss



H-Net: Multi-resolution Fusion Network

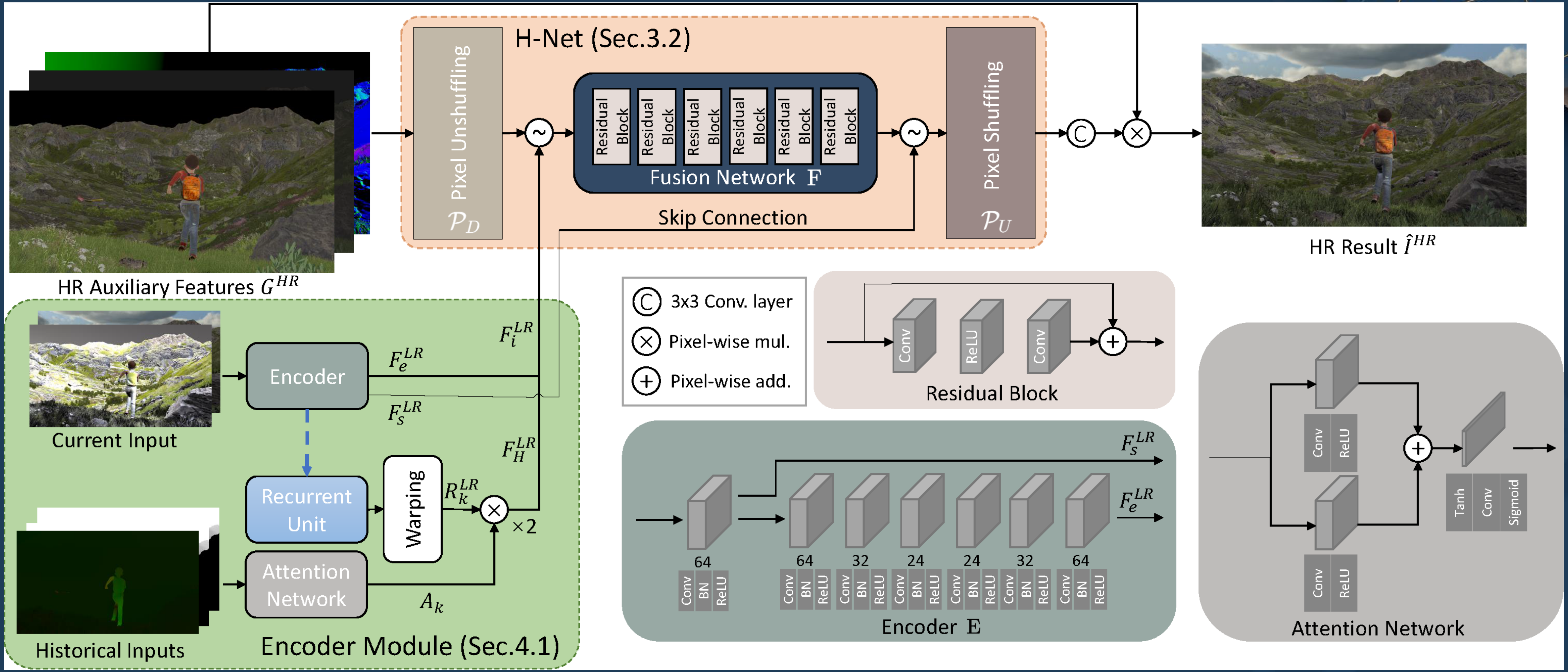
- Our design: H-Net
  1. Pixel-unshuffle G^{HR} into LR space and concatenate with other LR inputs
  2. Concatenated feature processed by fusion network $\mathbf{F}$ at LR level
  3. Pixel-shuffle the output of $\mathbf{F}$ into HR space as the HR output $\hat{L}_D^{HR}$



H-Net: Multi-resolution Fusion Network

- Our design: H-Net
  1. Pixel-unshuffle G^{HR} into LR space and concatenate with other LR inputs
  2. Concatenated feature processed by fusion network $\mathbf{F}$ at LR level
  3. Pixel-shuffle the output of $\mathbf{F}$ into HR space as the HR output $\hat{L}_D^{HR}$
- Pixel-unshuffling G^{HR} can also aggregate neighboring pixels in G^{HR} to obtain a compact implicit representation
  - Outperforming upsampling strategy

Network Details



SIGGRAPH ASIA 2023 SYDNEY

Experiments

High-fidelity Super-resolution Results

Connecting STORIES

The 16th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia

CONFERENCE 12 - 15 December 2023

EXHIBITION 13 - 15 December 2023

ICC, Sydney, Australia

ASIA.SIGGRAPH.ORG/2023

Sponsored by

Organized by koelnmesse

Implementation Details

- Two versions of implementation
  - FuseSR: Full network implementation with optimal quality
  - FuseSR

Dataset Details

- 4 well-designed scenes from Unreal Engine
  - 2 from UE4 and 2 from UE5
  - 4K (3840x2160) resolution
  - Customized shaders to generate pre-computed BRDF and other G-buffers

Dataset Details

- 4 well-designed scenes from Unreal Engine
  - 2 from UE4 and 2 from UE5, 4K resolution

Kite scene from UE4

Dataset Details

- 4 well-designed scenes from Unreal Engine
  - 2 from UE4 and 2 from UE5, 4K resolution

Showdown scene from UE4

Dataset Details

- 4 well-designed scenes from Unreal Engine
  - 2 from UE4 and 2 from UE5, 4K resolution

Slay scene from UE5

Dataset Details

- 4 well-designed scenes from Unreal Engine
  - 2 from UE4 and 2 from UE5, 4K resolution

City scene from UE5

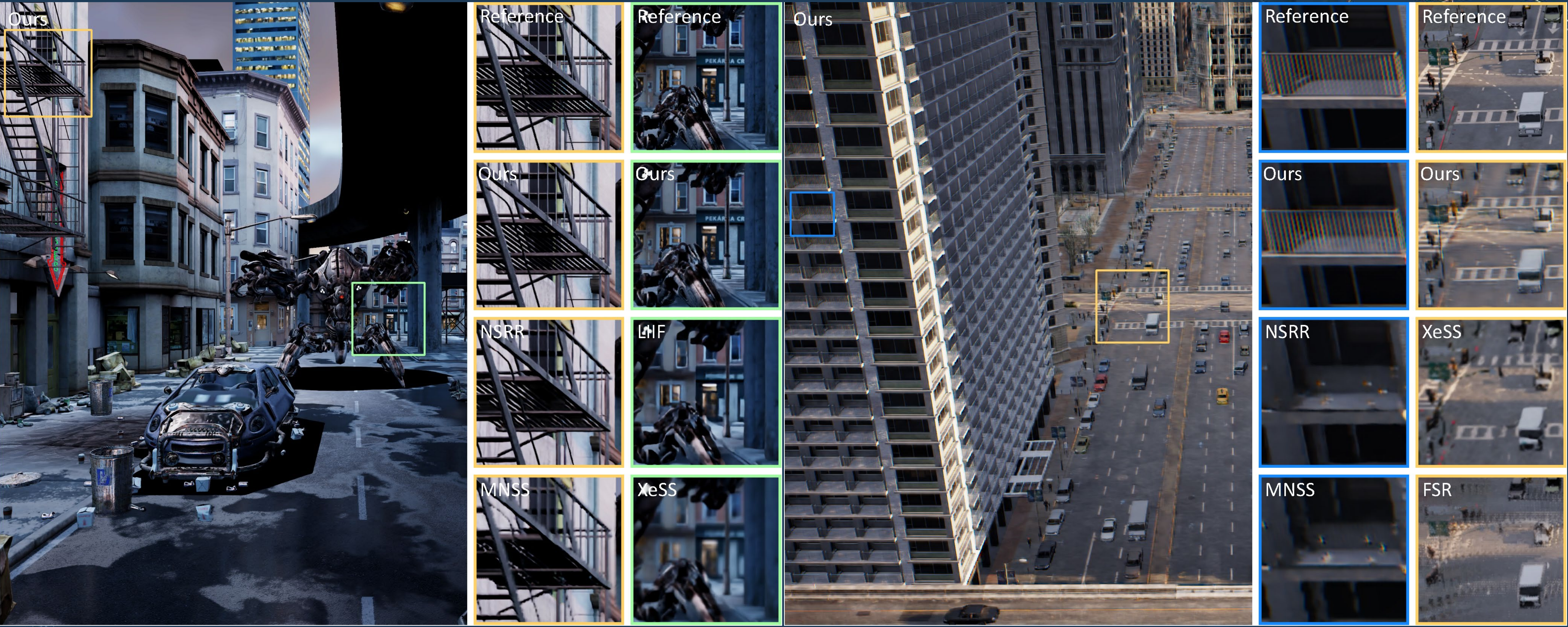
SIGGRAPH ASIA 2023 SYDNEY

Quality Evaluation (Quantitative)

- PSNR and SSIM in 4 scenes
  - "Ours" outperforms baselines with a large margin
  - "Ours

		Ours	Ours ⚡	NSRR	MNSS	LIIF	FSR	XeSS	Ours-8x	NSRR-8x	MNSS-8x
PSNR (dB)	Kite	32.33	31.22	27.74	28.00	26.47	29.12	28.30	30.21	25.00	25.72
	Showdown	36.32	31.42	30.27	29.17	30.33	26.29	29.31	33.61	29.17	25.62
	Slay	37.02	34.41	35.42	35.39	31.12	32.39	34.94	34.26	32.12	33.47
	City	28.94	28.66	27.65	28.23	26.56	26.63	27.15	27.20	25.95	26.46
SSIM	Kite	0.933	0.900	0.832	0.829	0.817	0.887	0.893	0.899	0.765	0.770
	Showdown	0.976	0.949	0.945	0.914	0.942	0.866	0.917	0.955	0.914	0.813
	Slay	0.972	0.958	0.962	0.963	0.962	0.928	0.944	0.957	0.939	0.943
	City	0.921	0.901	0.899	0.896	0.874	0.836	0.888	0.916	0.873	0.873

Quality Evaluation (Qualitative, 4x4)



Quality Evaluation (Qualitative, 4x4)

Quality Evaluation (Qualitative, 8x8)

- FuseSR is the first to succeed in 8x8 super-resolution

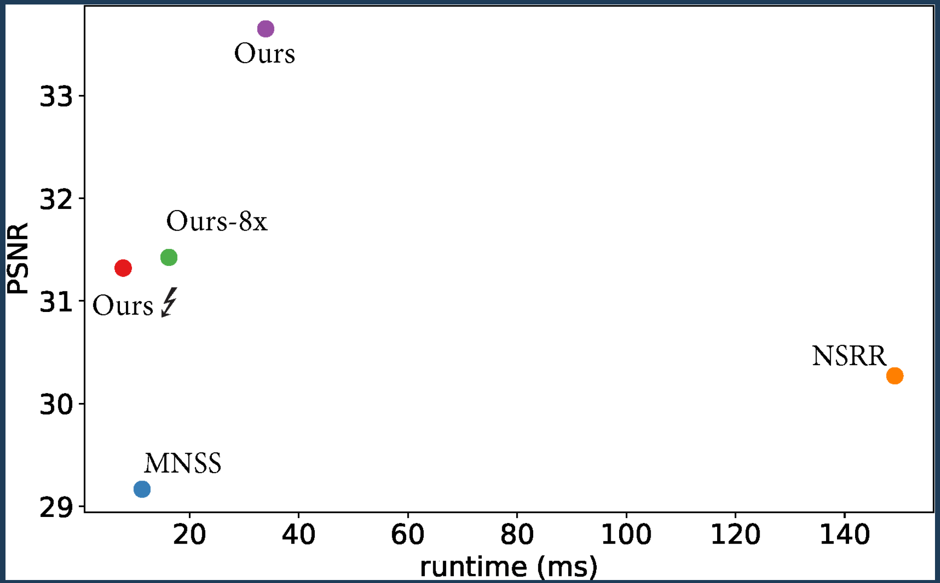
Performance Evaluation

- Our method has better performance in general
  - Especially in high-resolution settings (like 4K)
- Our experiment validates the low cost of HR G-buffer generation

	720p	1080p	2K	4K
HR G-buffer	0.83	0.93	1.97	2.35
Ours	6.21	8.44	15.09	33.93
Ours-8x	6.20	7.57	8.79	16.20
Ours ⚡	2.66	2.88	3.96	7.82
NSRR	13.53	26.29	64.02	149.20
MNSS	2.26	3.57	5.52	11.29

Performance Evaluation

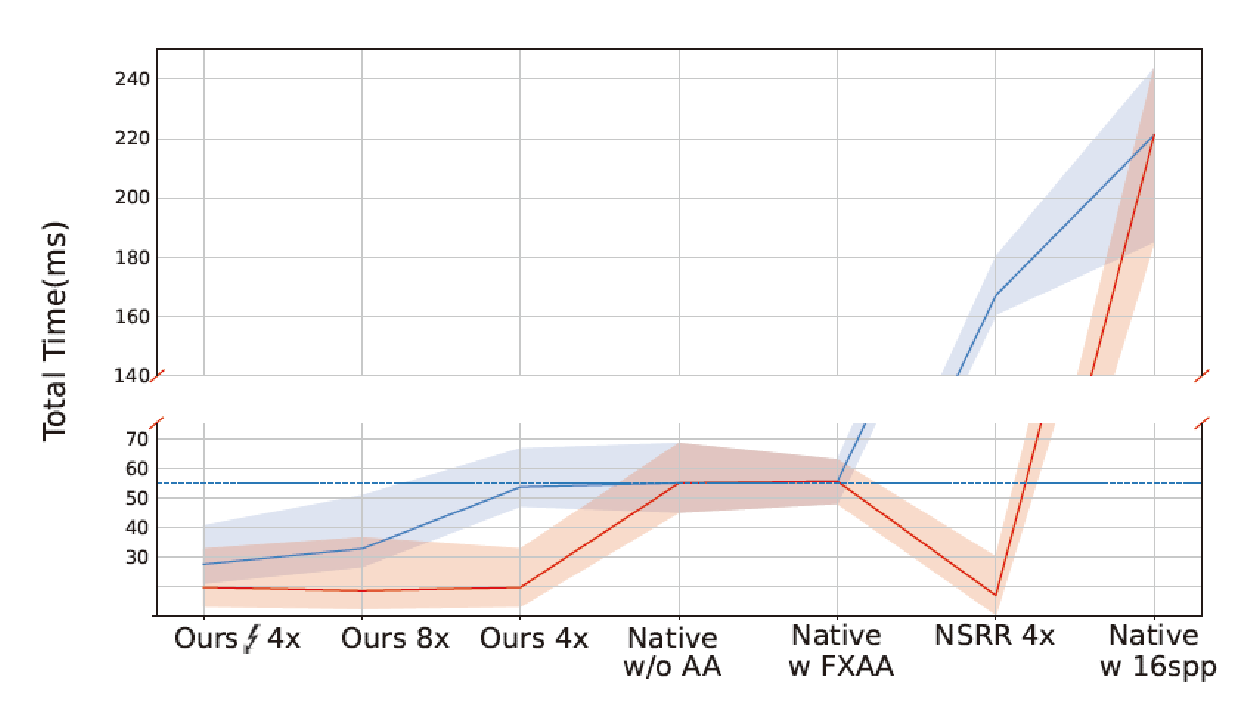
- Our method has better performance in general
  - Especially in high-resolution settings (like 4K)
- Our method has best trade-off between performance and quality



SIGGRAPH ASIA 2023 SYDNEY

Performance Evaluation

- Our method can indeed accelerate the rendering process
  - Rendering LR image & HR G-buffer + Ours >>Faster>> Rendering HR image

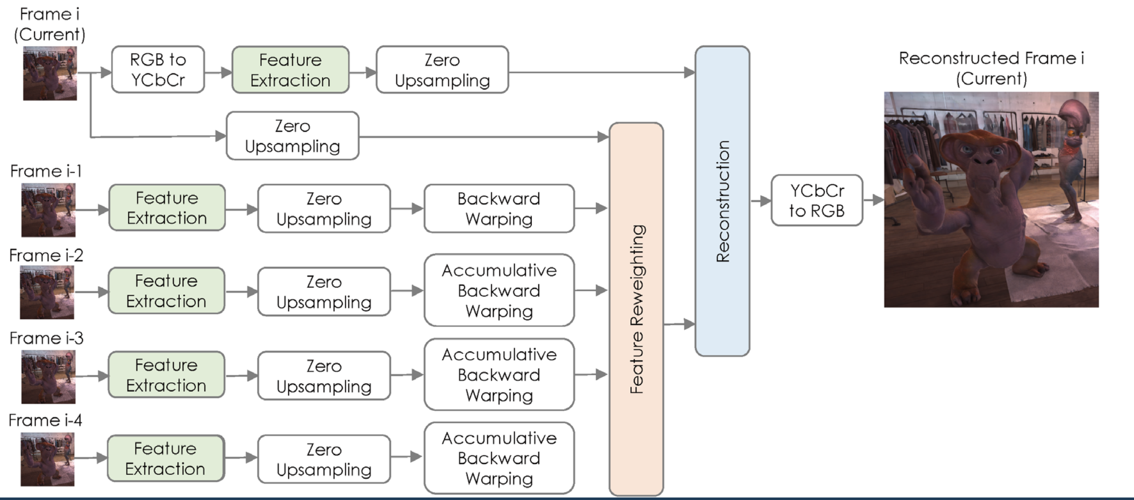


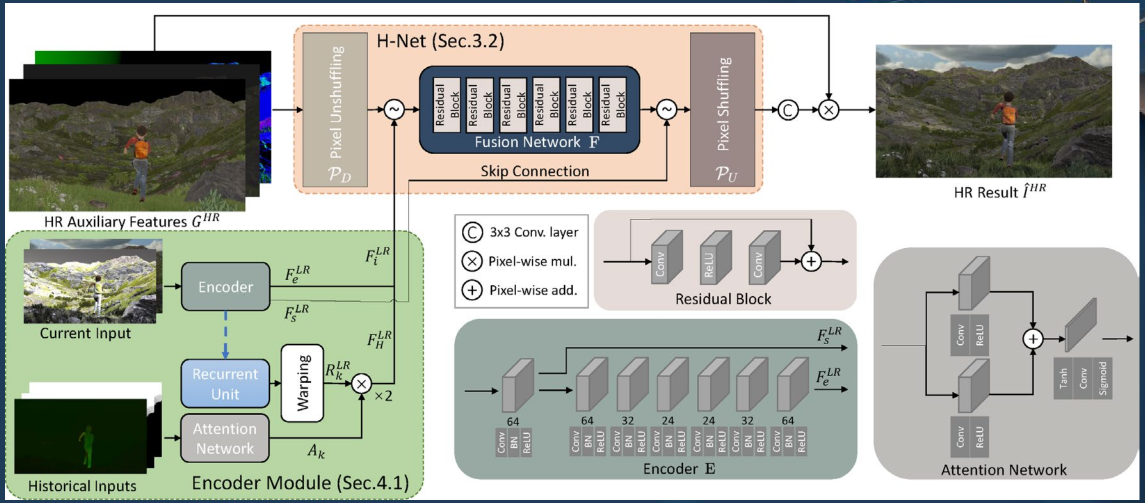
SIGGRAPH ASIA 2023 SYDNEY

Discussion

Design Comparison and Further Understanding

Network Design



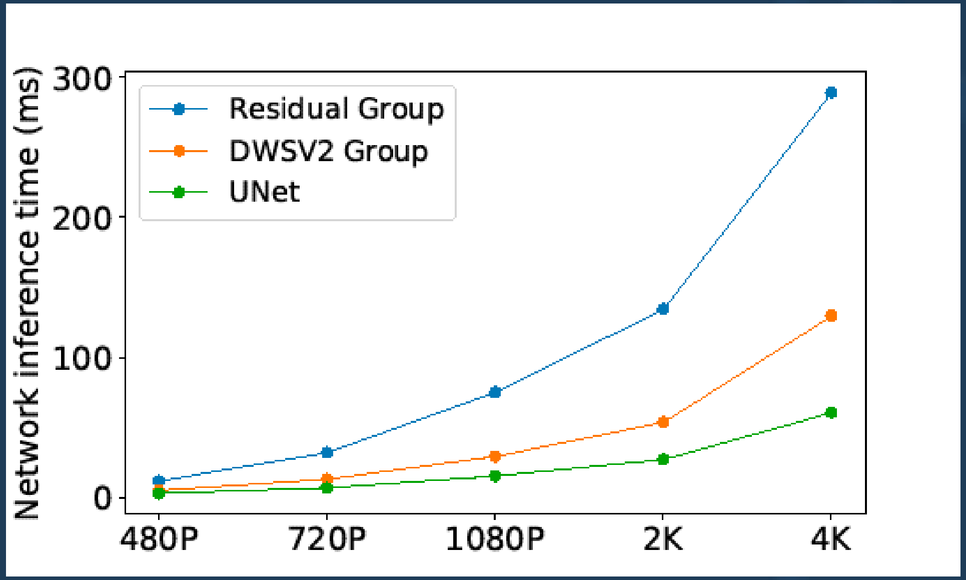


Early Upsamping

Late Upsamping

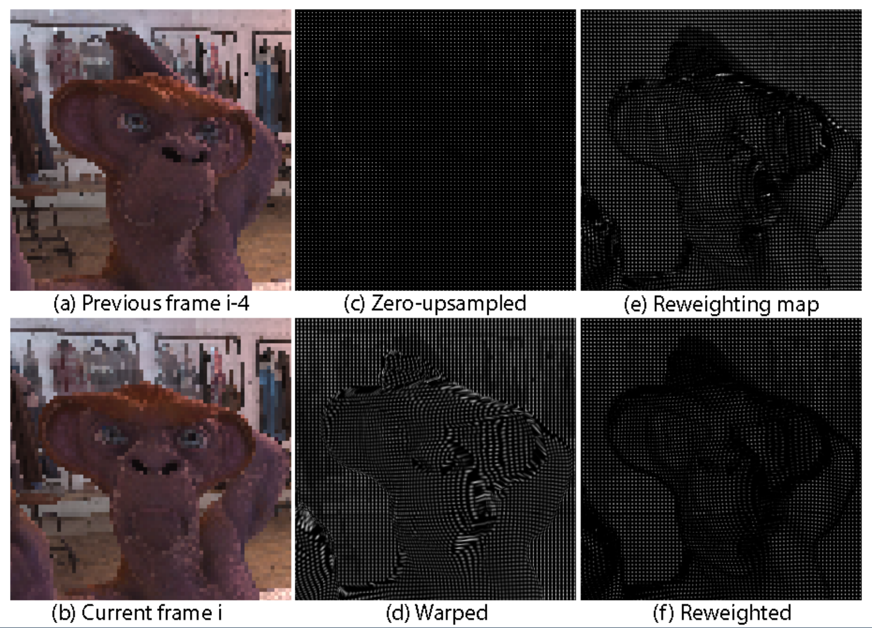
Resolution and Inference Time

- CNN infer time highly related to feature resolution.
- Late upsampling have advantages in speed.

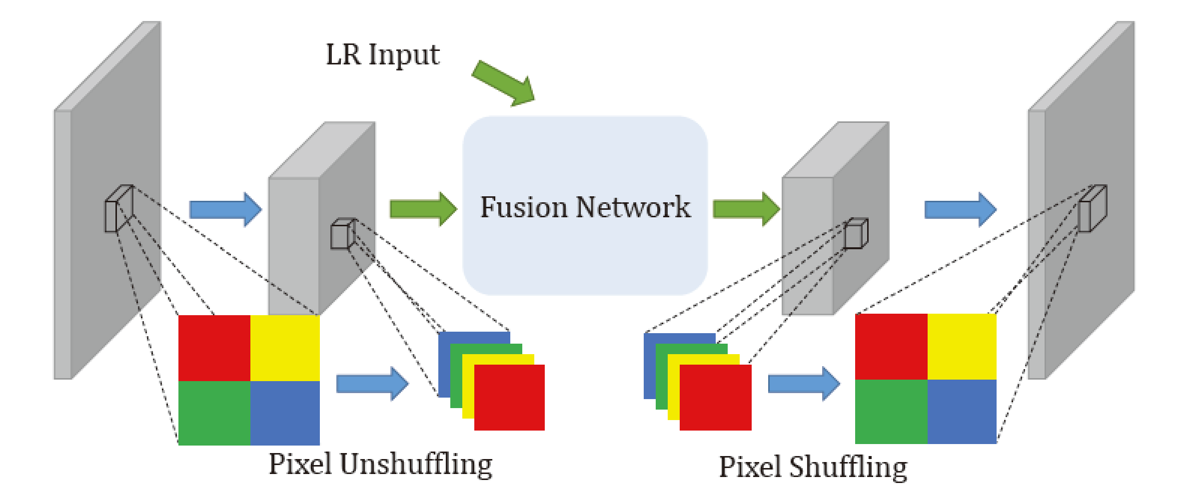


Network Design

High utilization rate of input information



Temporal warping with zero-upsampling



H-Net

Alter Perspective under FuseSR

Resolution

Neural Shading

LR Render Result

Neural AA

HR G-Buffers

SIGGRAPH ASIA 2023 SYDNEY

FuseSR: Super Resolution for Real-time Rendering through Efficient Multi-resolution Fusion

Zhihua Zhong[*], Jingsen Zhu[*], Yuxin Dai, Chuankun Zheng, Yuchi Huo[†], Guanlin Chen, Hujun Bao, Rui Wang[†]

浙江大學
ZHEJIANG UNIVERSITY

Code will be released at Project Page ↓