Selective Kernel and Motion-emphasized Loss Based Attention-guided CNN for HDR Imaging of Dynamic Scenes

Graduate School of Information, Production and System

Waseda University

Ikenaga Lab.

Deng Yipeng

Outline

- Background
- Proposals
- Experiment results
- Conclusion

Outline

- Background
  - Ghost Problem in HDR Reconstruction
  - Conventional Work of Yan19
  - Problems and Proposals
- Proposals
- Experiment results
- Conclusion and Future work

Ghost Problem In HDR Reconstruction

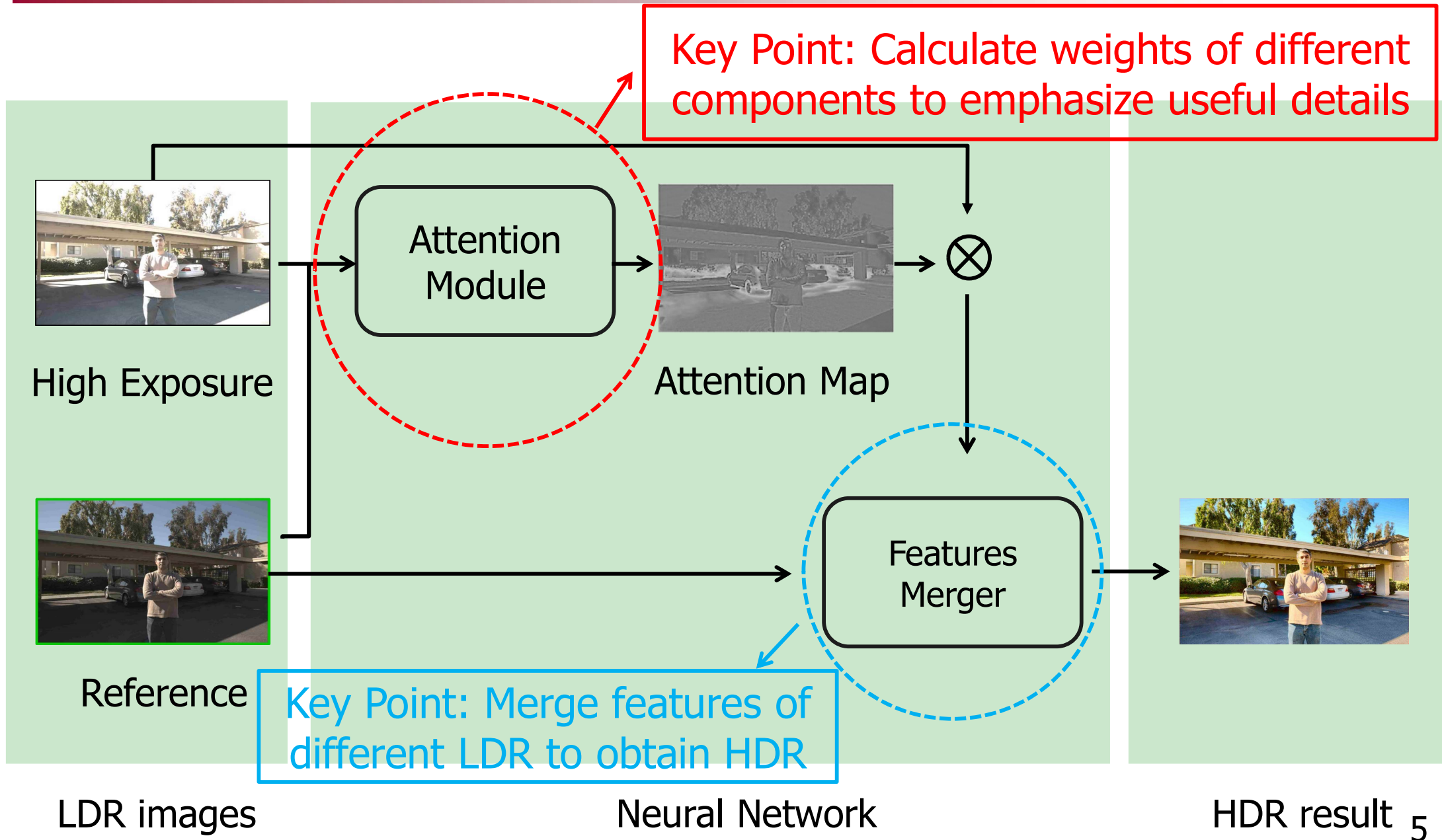LDR images with different exposure

fusion

result with ghost

Dynamic Scenes:
Scenes where foreground motion exist
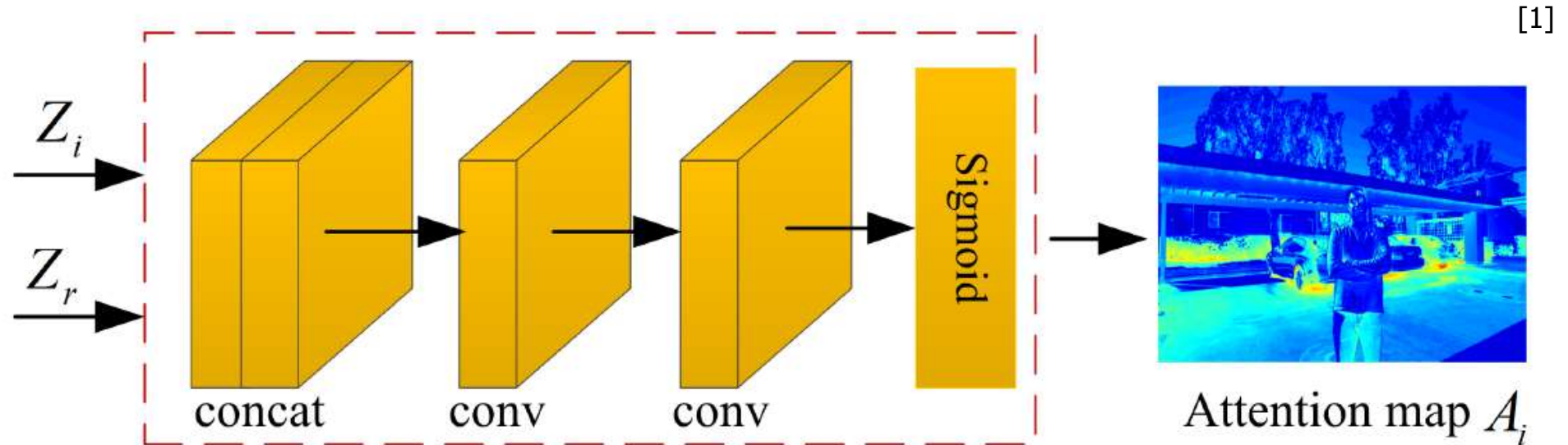
Due to the motion, ghost-like artifacts appear

Conventional Work of Yan19[1]

Key Point: Calculate weights of different components to emphasize useful details

Attention Module

High Exposure

Attention Map

Reference

Key Point: Merge features of different LDR to obtain HDR

Features Merger

LDR images

Neural Network

HDR result

5

[1] Yan et al., "Attention-guided Network for Ghost-free High Dynamic Range Imaging, " *CVPR*, 2019

[1]

Z_i
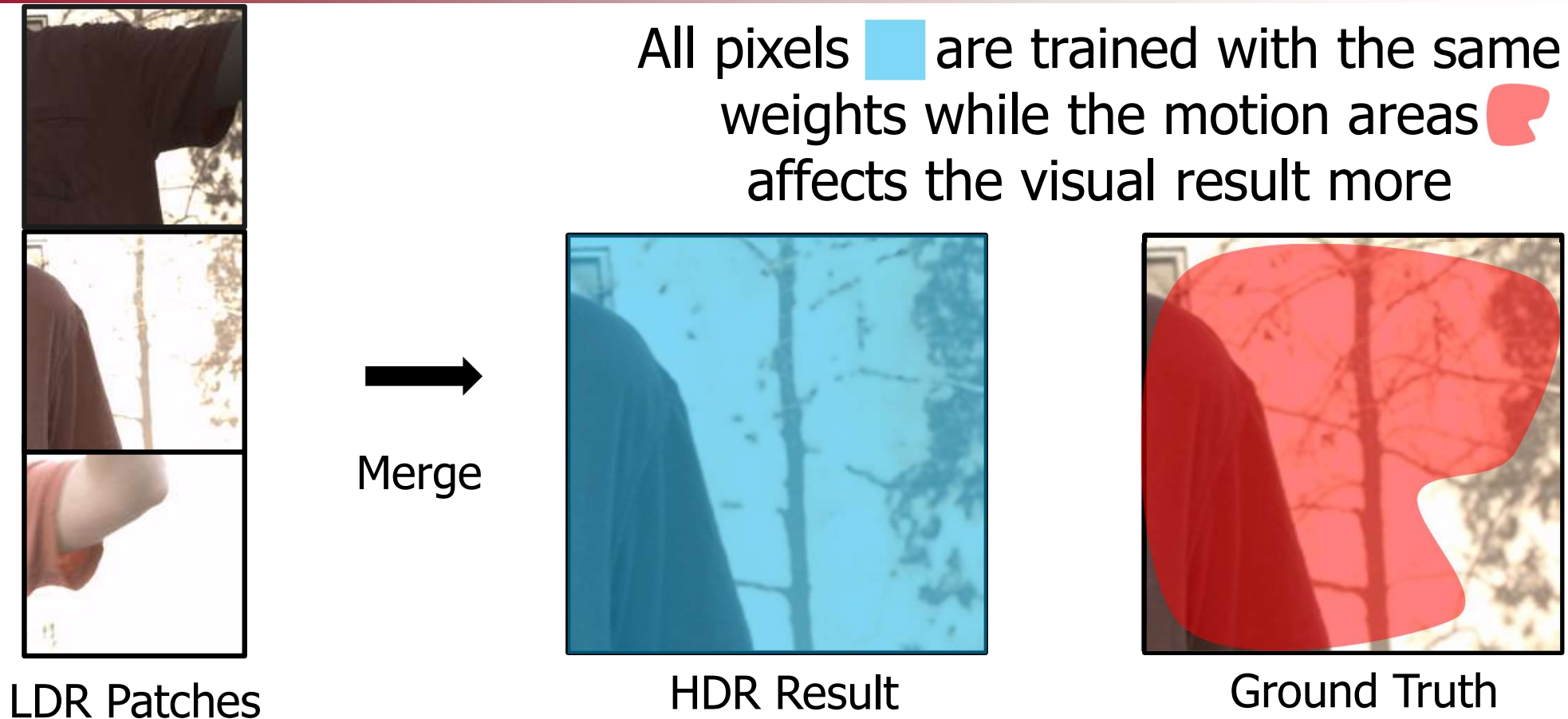
Z_r

concat conv conv Sigmoid

Attention map A_i

Inaccurate attention scores:
- The difference between spatial attention and channel attention is ignored
- Shallow convolution layers limits model's performance
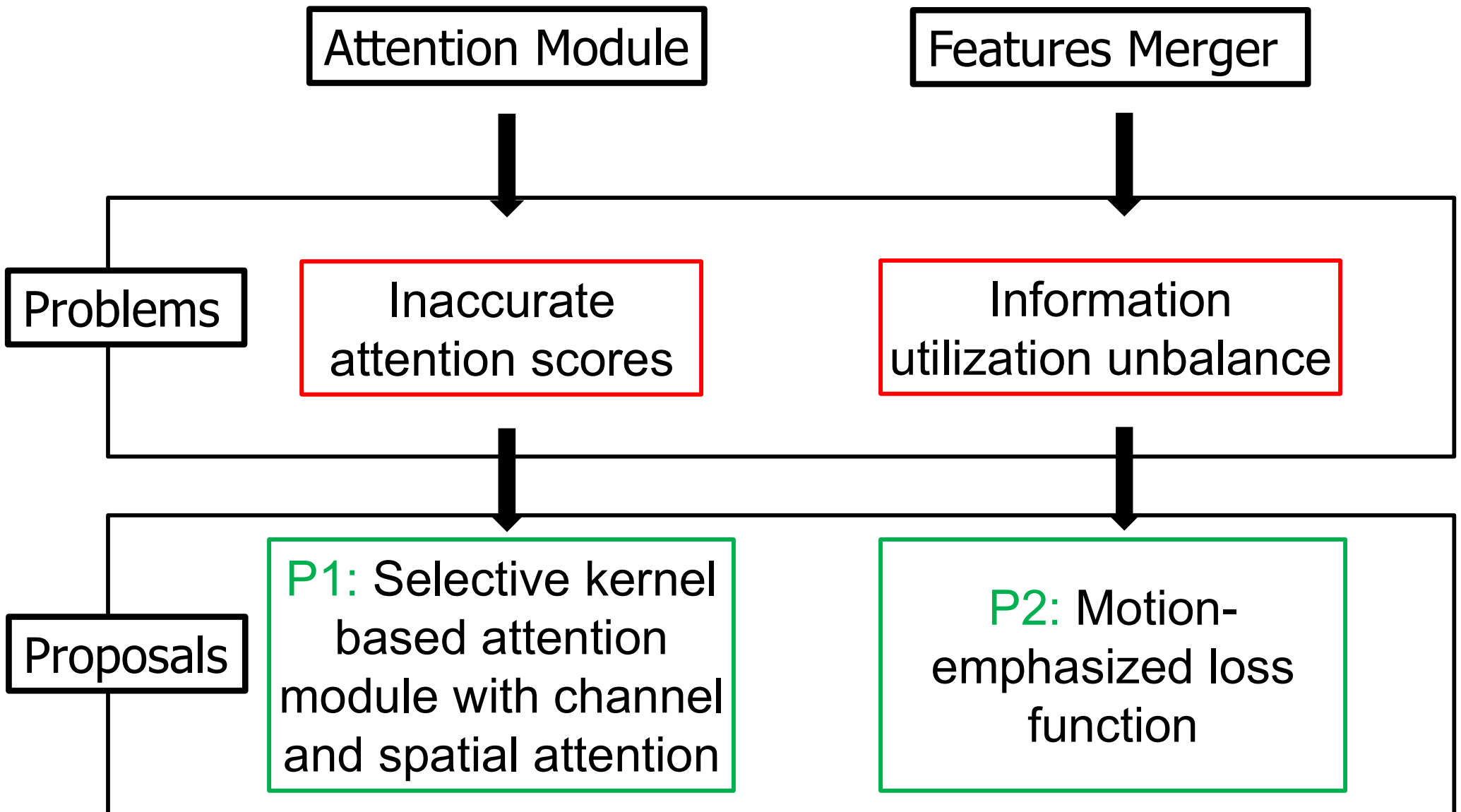
Problem of Features Merger in Conventional Work

All pixels ▮ are trained with the same weights while the motion areas ▮ affects the visual result more

Merge →

LDR Patches

HDR Result

Ground Truth

Information utilization unbalance:
- All pixels are treated as the same in error calculation
- The difference between motion areas and static areas is ignored

Problem Statement and Proposals

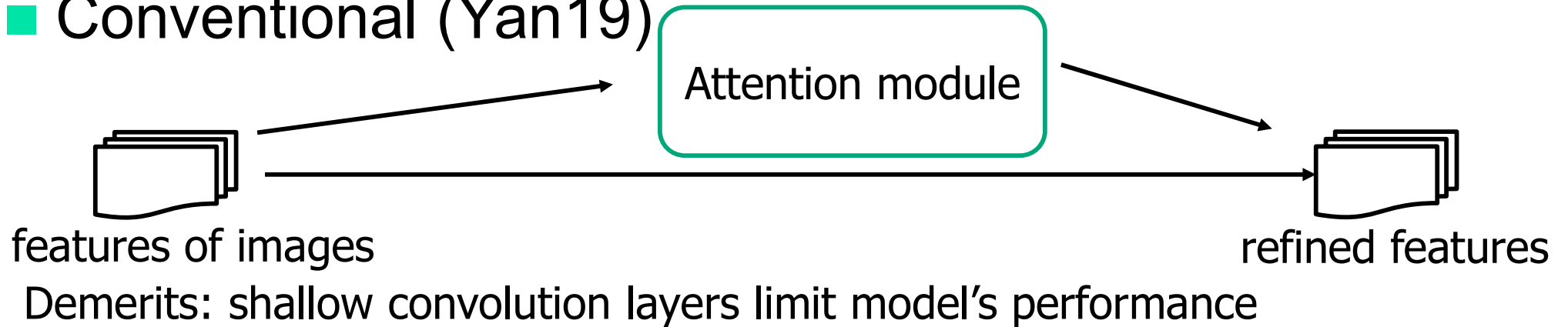Attention Module

Features Merger

Problems

Inaccurate attention scores

Information utilization unbalance

Proposals

P1: Selective kernel based attention module with channel and spatial attention

P2: Motion-emphasized loss function

Outline

- **Background**
- **Proposals**
  - Framework
  - P1: Selective Kernel Based Attention Module with Channel and Spatial Attention
  - P2: Motion-emphasized Loss Function
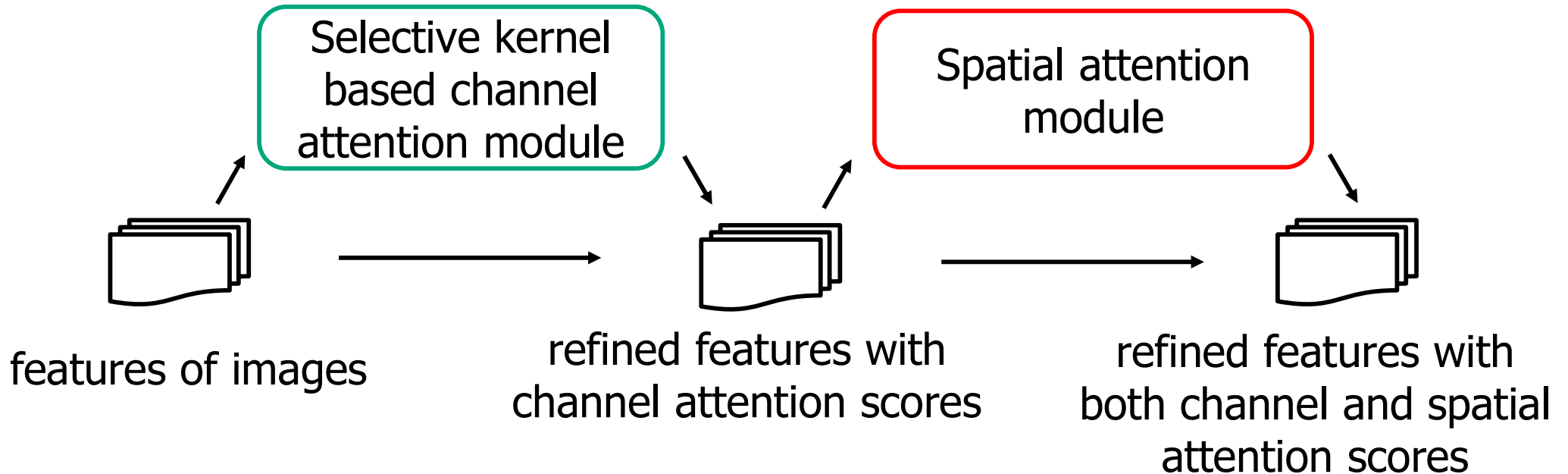- **Experiment results**
- **Conclusion**

# P1:Conceptual Difference

■ Conventional (Yan19)

Attention module

features of images → refined features

Demerits: shallow convolution layers limit model's performance

■ Selective Kernel Based Attention Module

Selective kernel based channel attention module

Spatial attention module

features of images → refined features with channel attention scores → refined features with both channel and spatial attention scores

Merits: accurate attention scores with channel and spatial information
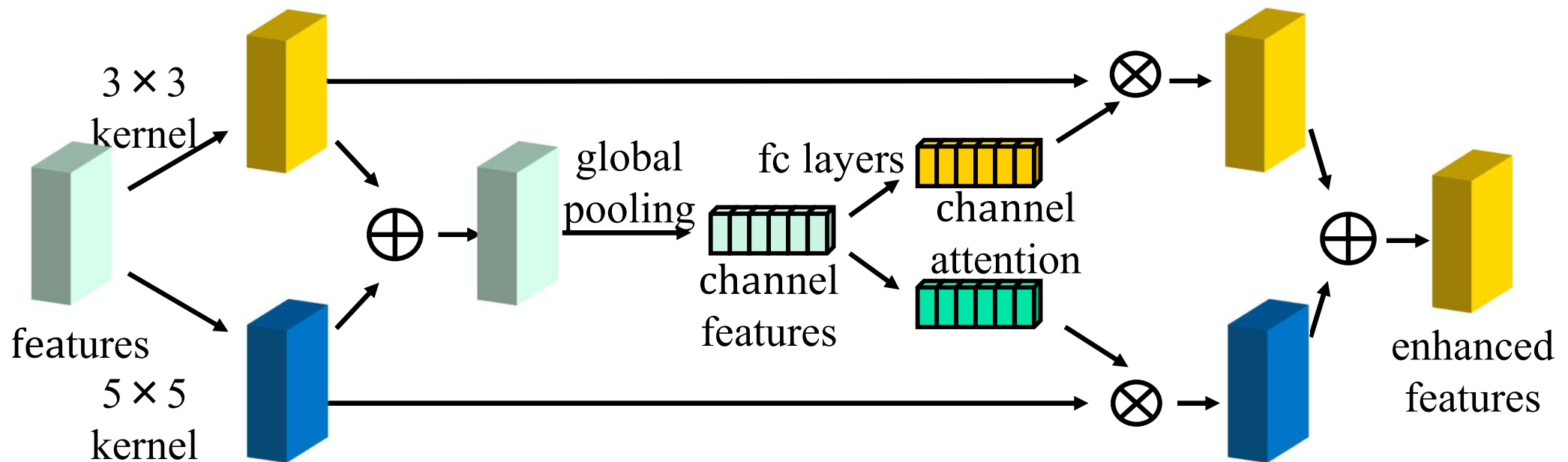
- Channel attention module is designed to learn "what" is important

- Selective kernel (two different kernel sizes) is adopted in our network for robustness

P1: Detail of Spatial Attention Module

$H \times W \times C$ features $H \times W \times 2$ features $H \times W \times 1$ attention scores   enhanced features

average and max pooling
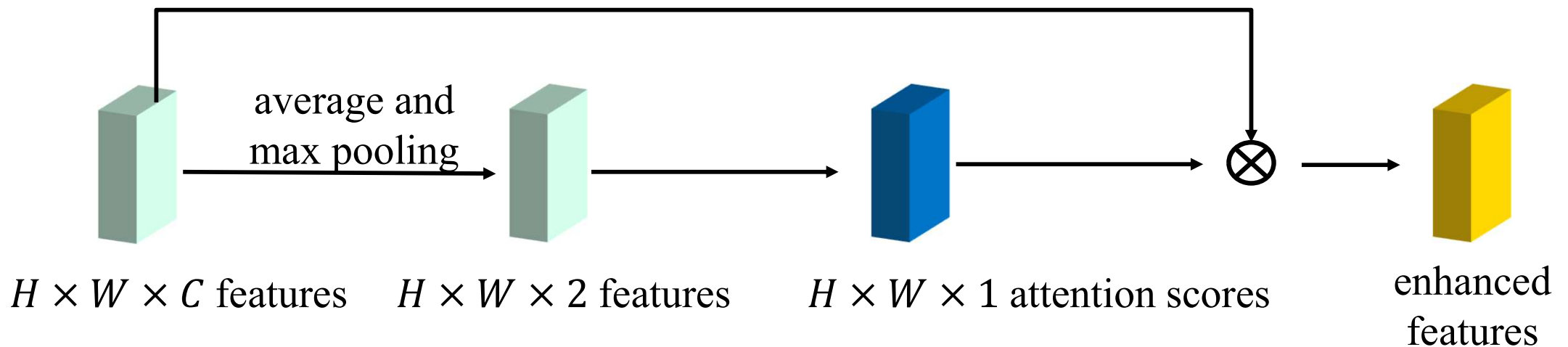
- Spatial attention module is designed to learn "where" is important

P2:Conceptual difference

■ Conventional (Yan19)

All pixels are treated as the same in error calculation

Demerits: the difference between motion area and background is ignored
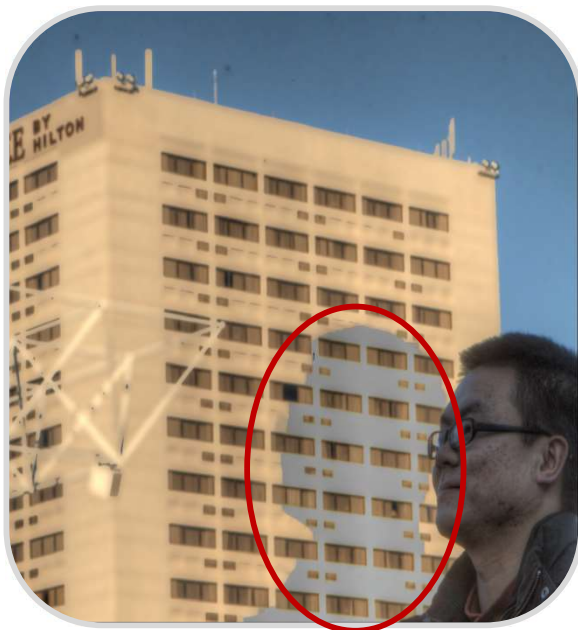
■ Motion-emphasized Loss Function

Motion areas are detected first and larger weights are assigned

Merits: by giving more weights to motion and saturated areas, the useful information is focused

13

Detailed procedure of P2

Ghost-like artifacts appear in motion areas.
Intuitively, the better detail image is recovered in motions areas, the better visual result we can obtain.

$$MAE_{motion} = \frac{1}{N} \sum_{i=1}^{N} w_i \left| \tilde{Y}_i - Y_i \right|$$

By assigning w_i with larger value, the importance of motions areas is emphasized

Outline

- **Background**
- **Proposals**
- **Experiment results**
  - Evaluation Method
  - Experiment Results
- **Conclusion and Future work**

Evaluation Method

- ## PSNR-L
  - PSNR value in linear domain (generated HDR image)
- ## PSNR-T
  - PSNR value in tone-mapped domain
  - The tone-mapping function is shown as $T = \dfrac{\log(1+\mu H)}{\log(1+\mu)}$
- ## HDR-VDP-2[3]
  - A visual metric that compares a pair of images
  - Consider both visibility and quality
  - Image has better quality with higher scores

[3] Mantiuk R, Rempel A G, Heidrich W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions[J]. 2011, 30(4):1-14.

Experiment Results

	Sen12	Kalantari17	Wu18	Yan19	Ours
PSNR(T)	40.9545	42.7423	41.7403	42.9167	43.1733
PSNR(L)	38.3156	41.2158	40.8739	40.1648	40.8990
HDR-VDP-2	56.8968	60.5088	60.5006	60.8320	61.0222

Proposed method achieves state-of-the-art performance in PSNR and HDR-VDP-2 metrics.

17

Experiments on Loss Function

	MAE loss	motion-emphasized MAE loss
PSNR(T)	43.1274	43.1733
PSNR(L)	40.9193	40.8990
HDR-VDP-2	60.7841	61.0222

Results shows slight improvements with metrics calculating image quality based on all pixels, so a new metric which considers motion areas is needed
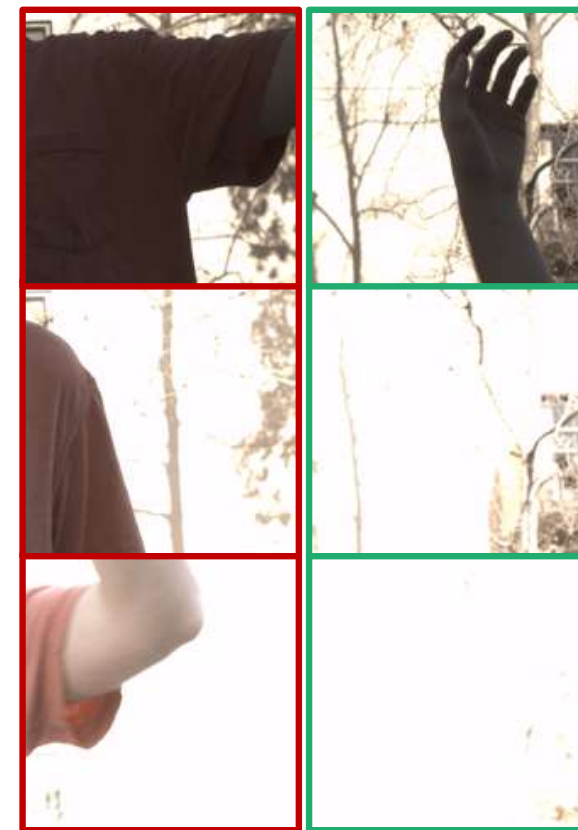
Visual Results

LDR images

Tonmapped our result

LDR patches

Sen
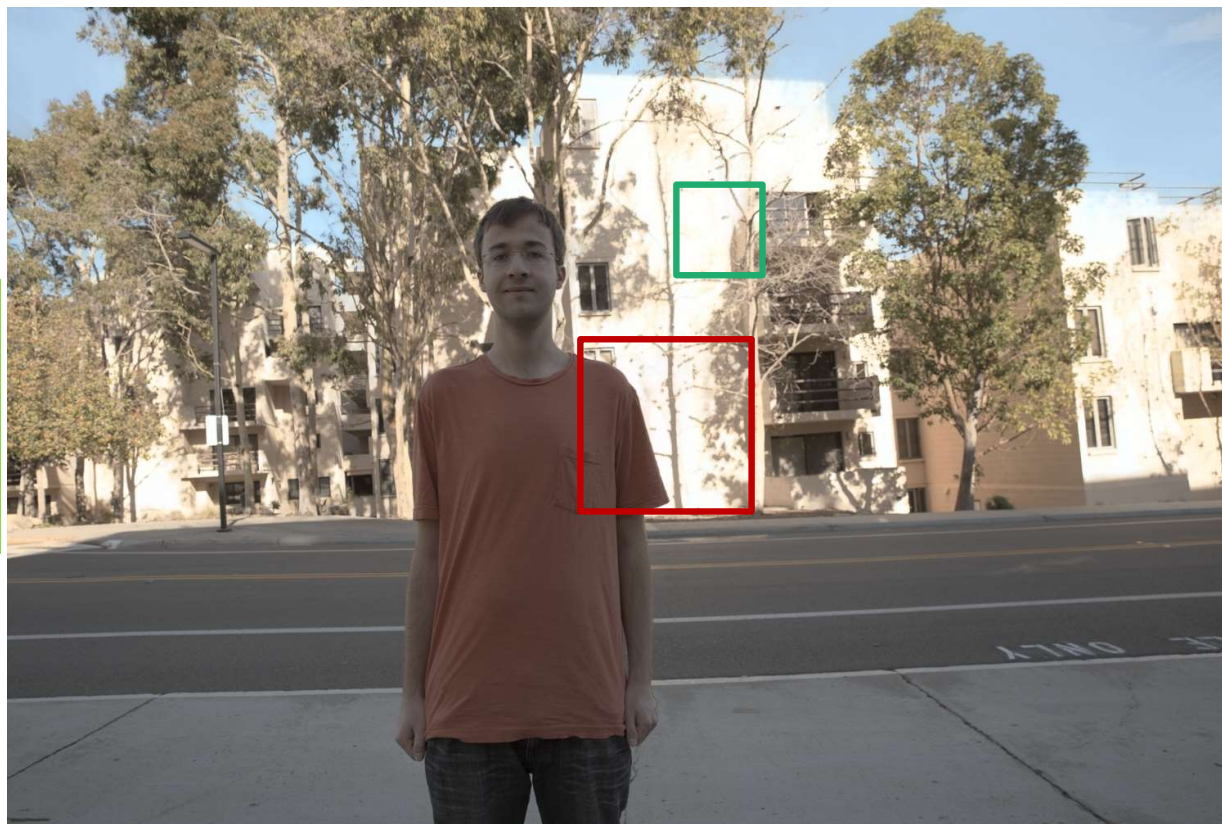
Kalantari

Wu
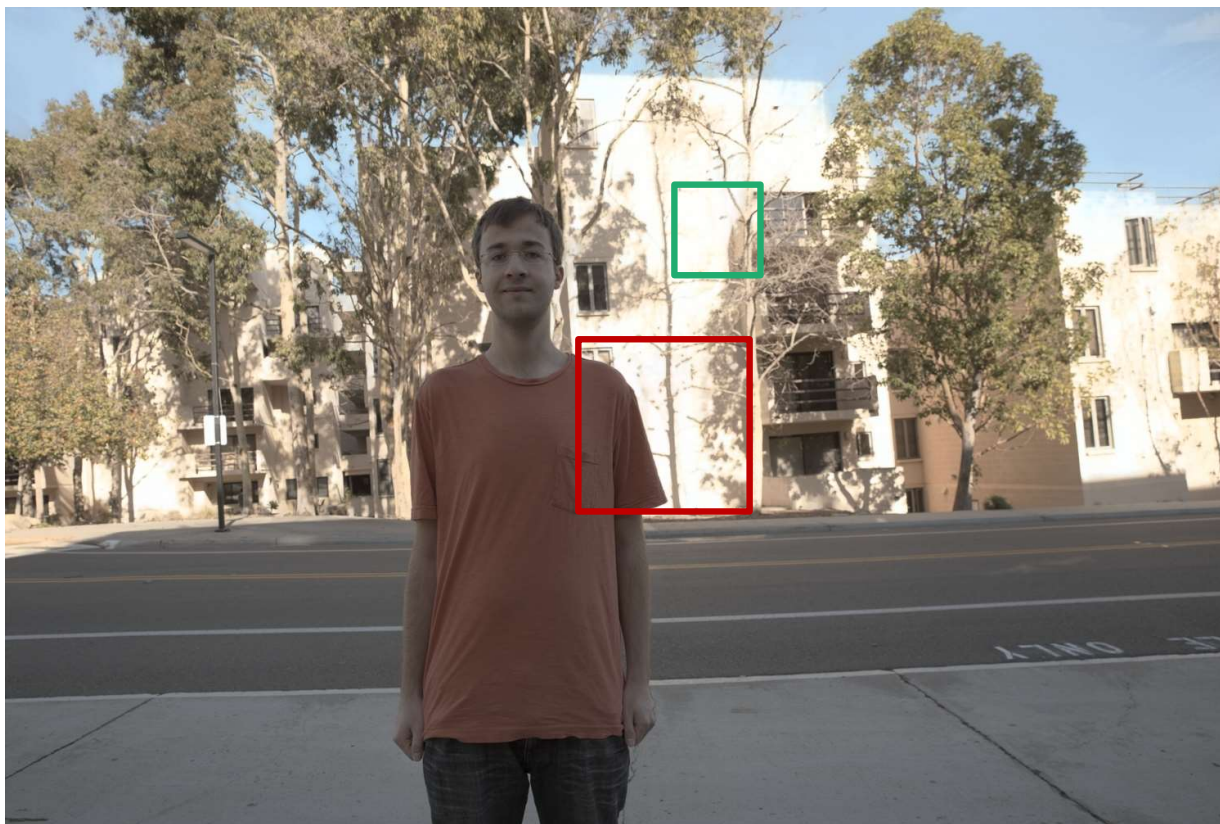
Ours

Ground Truth
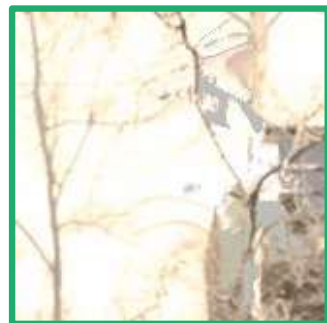
Visual Results

LDR images

Tonmapped our result

LDR patches

Sen

Kalantari

Wu

Ours

Ground Truth

20

Outline

- Background

- Proposals

- Experiment results

- Conclusion

Conclusion

- ## Target
  - Ghost-Free HDRI Reconstruction of Dynamic Scenes

- ## Proposals
  - P1: Selective Kernel Based Attention Module
  - P2: Motion-emphasized Loss Function

- ## Result
  - Proposed method achieves state-of-the-art performance in PSNR and HDR-VDP-2 metrics results
  - The motion-emphasized loss function improves result from 60.78 to 61.02 slightly
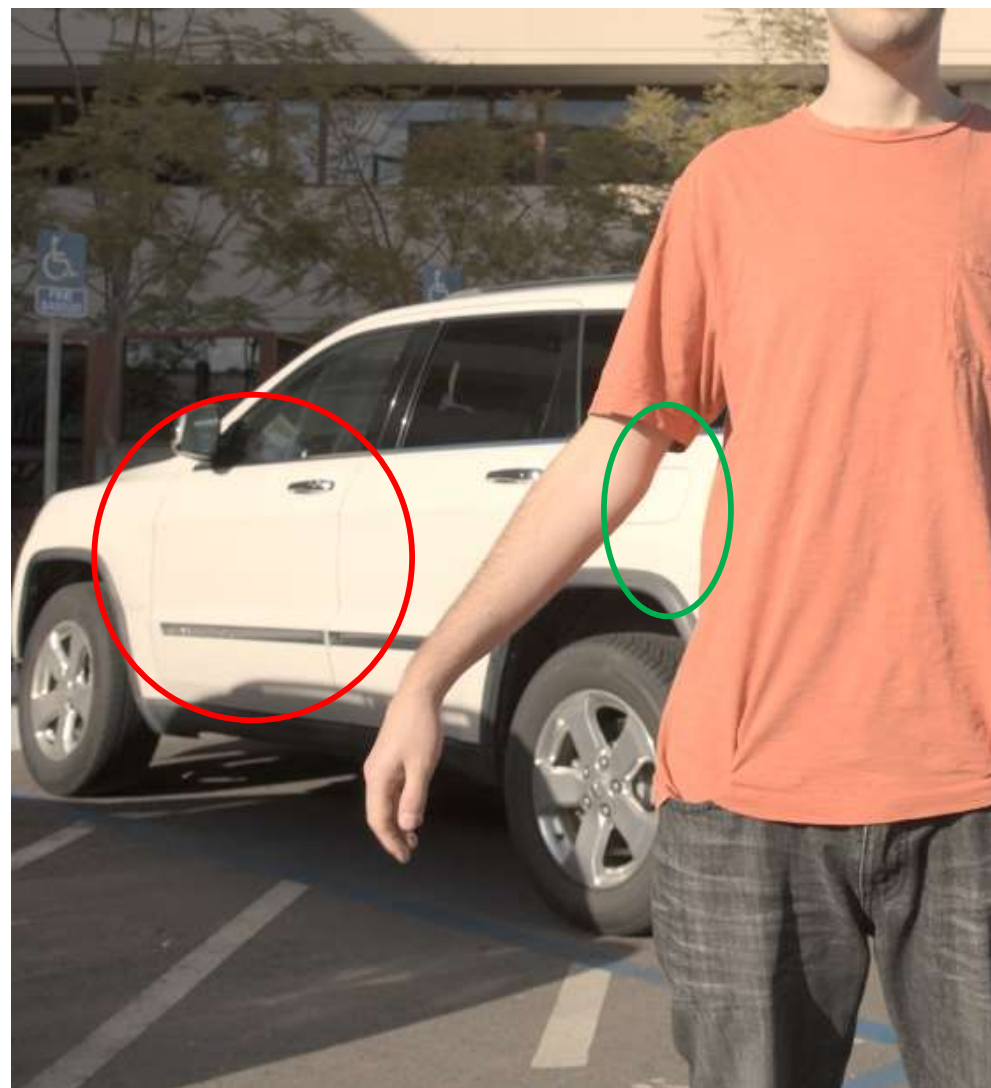
Thank You for Listening!

Challenging Cases with Bad Result

Ours

Ground Truth