HDR image reconstruction from a single exposure using deep CNNs

SIGGRAPH ASIA 2017 BANGKOK

- Gabriel Eilertsen, Linköping university
- Joel Kronander, Linköping university
- Gyorgy Denes, University of Cambridge
- Rafał K. Mantiuk, University of Cambridge
- Jonas Unger, Linköping university

LINKÖPING UNIVERSITY

UNIVERSITY OF CAMBRIDGE

Introduction

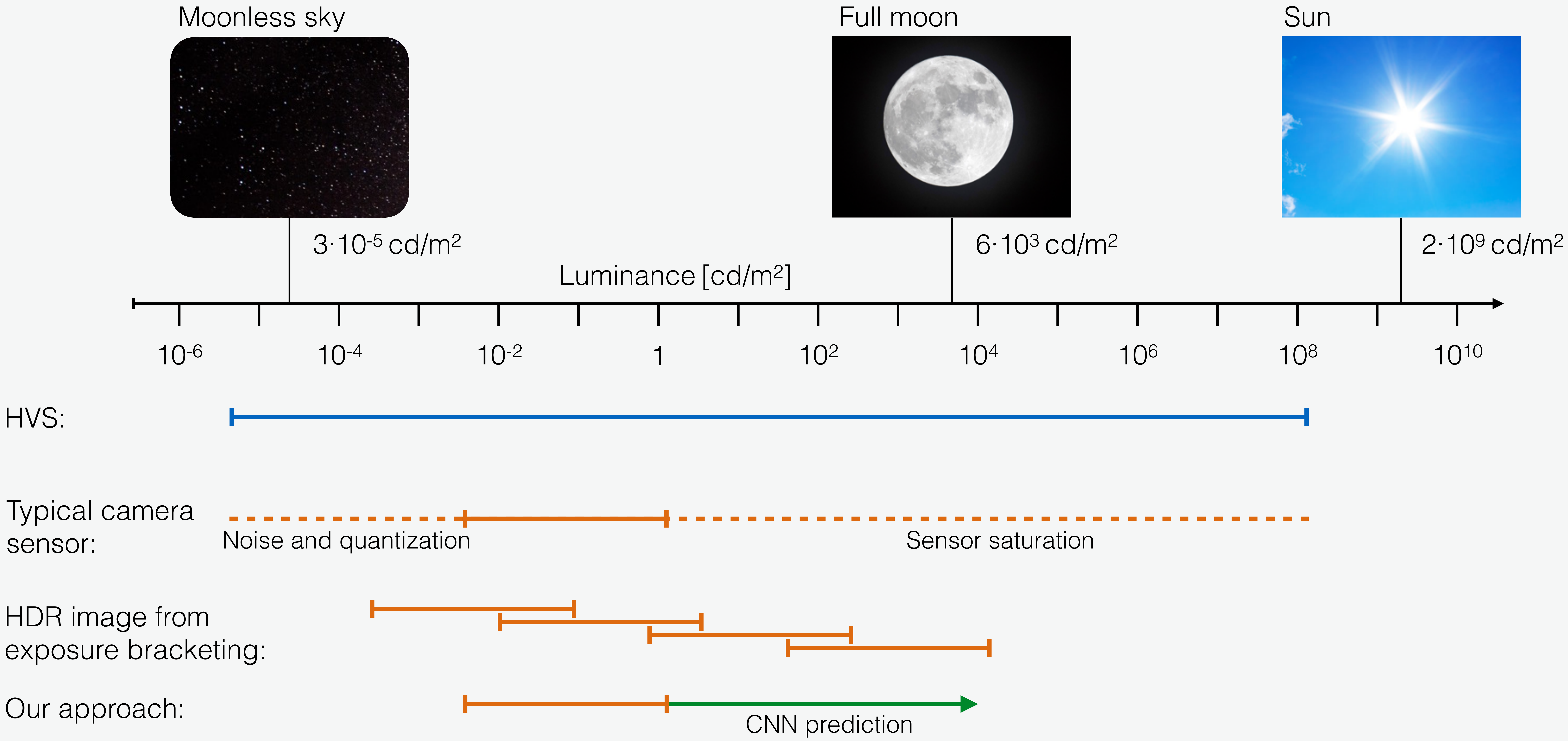
- Reconstruction of high dynamic range (HDR) data
- 8-bit input, captured in one exposure
- Model
  - Convolutional neural net (CNN)
  - Auto-encoder
  - HDR specific design
- Training
  - HDR specific loss function
  - Large set of collected HDR images
  - Pre-training using MIT places dataset
- Evaluation on HDR display

Input LDR image

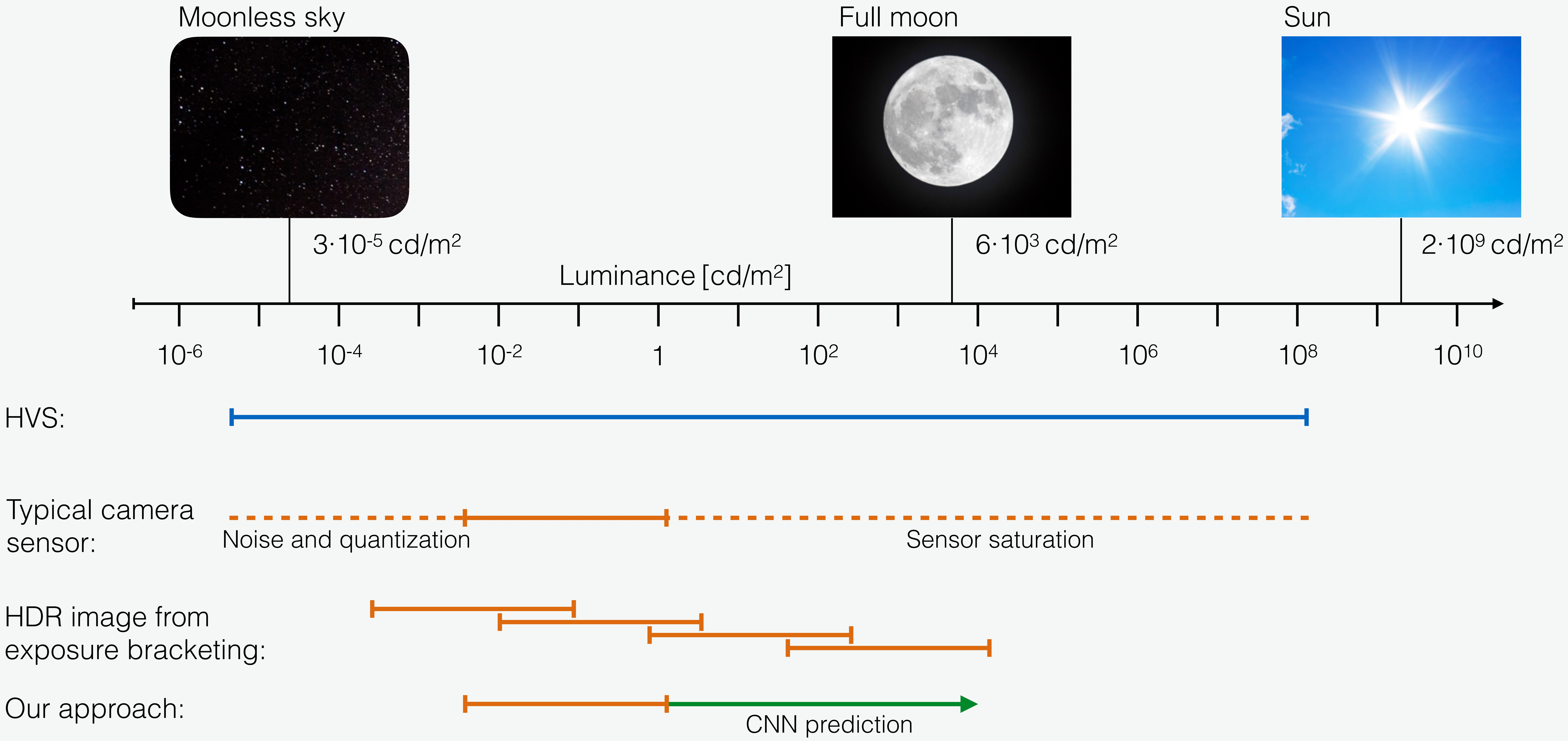
Introduction

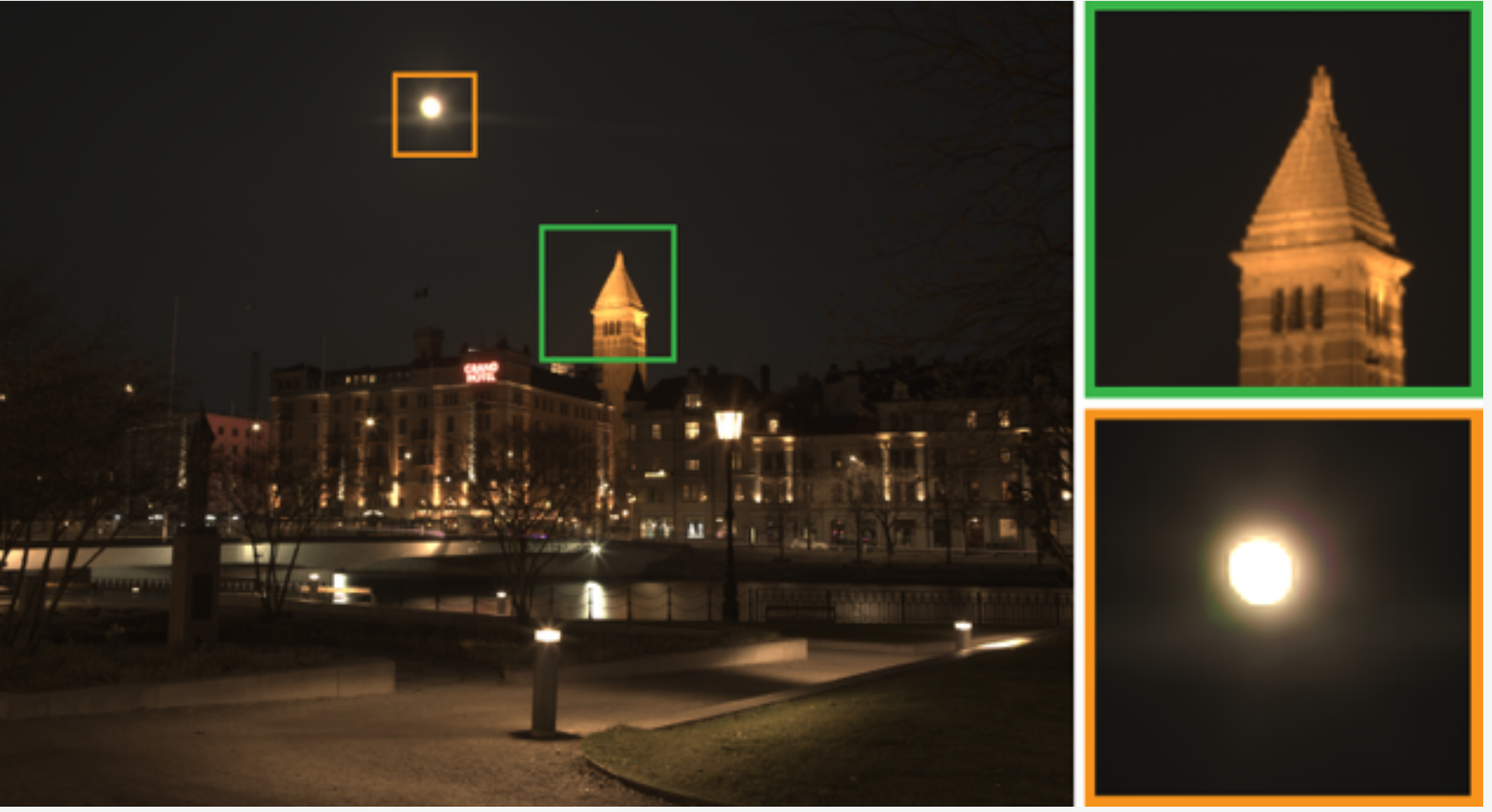
- Reconstruction of high dynamic range (HDR) data
- 8-bit input, captured in one exposure
- Model
  - Convolutional neural net (CNN)
  - Auto-encoder
  - HDR specific design
- Training
  - HDR specific loss function
  - Large set of collected HDR images
  - Pre-training using MIT places dataset
- Evaluation on HDR display

Reconstructed HDR image

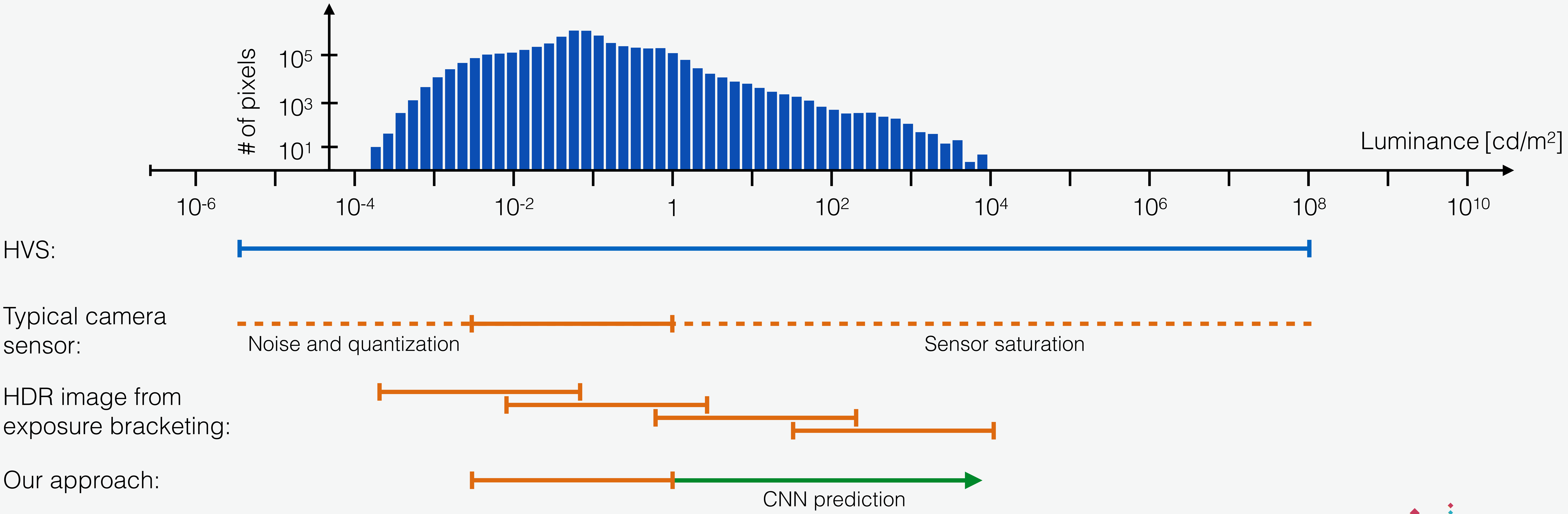


SIGGRAPH ASIA 2017 BANGKOK

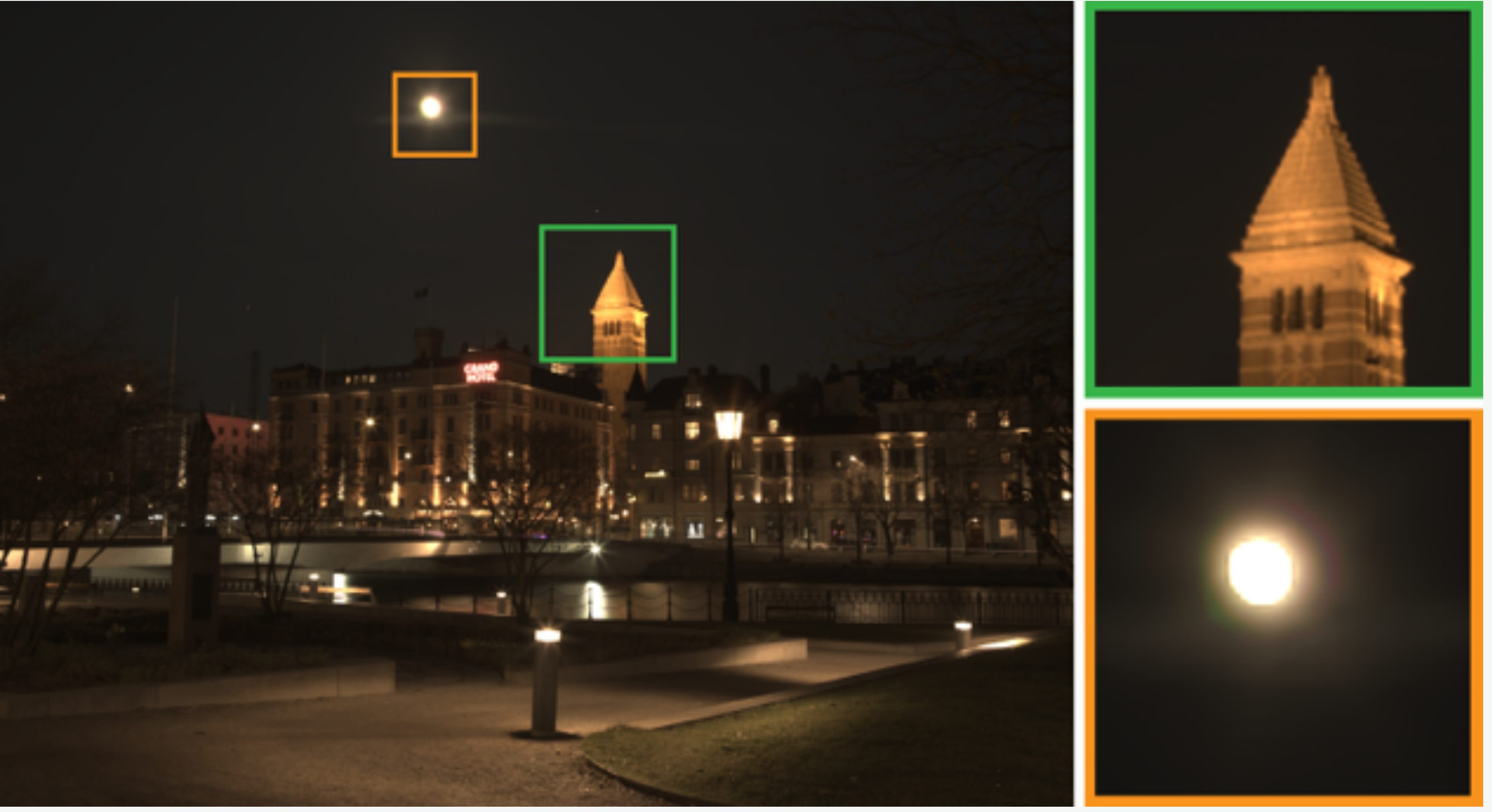

Moonless sky
Full moon
Sun
3·10-5 cd/m2
6·103 cd/m2
2·109 cd/m2
Luminance [cd/m2]
10-6
10-4
10-2
1
102
104
106
108
1010
HVS:
Typical camera sensor:
Noise and quantization
Sensor saturation
HDR image from exposure bracketing:
Our approach:
CNN prediction



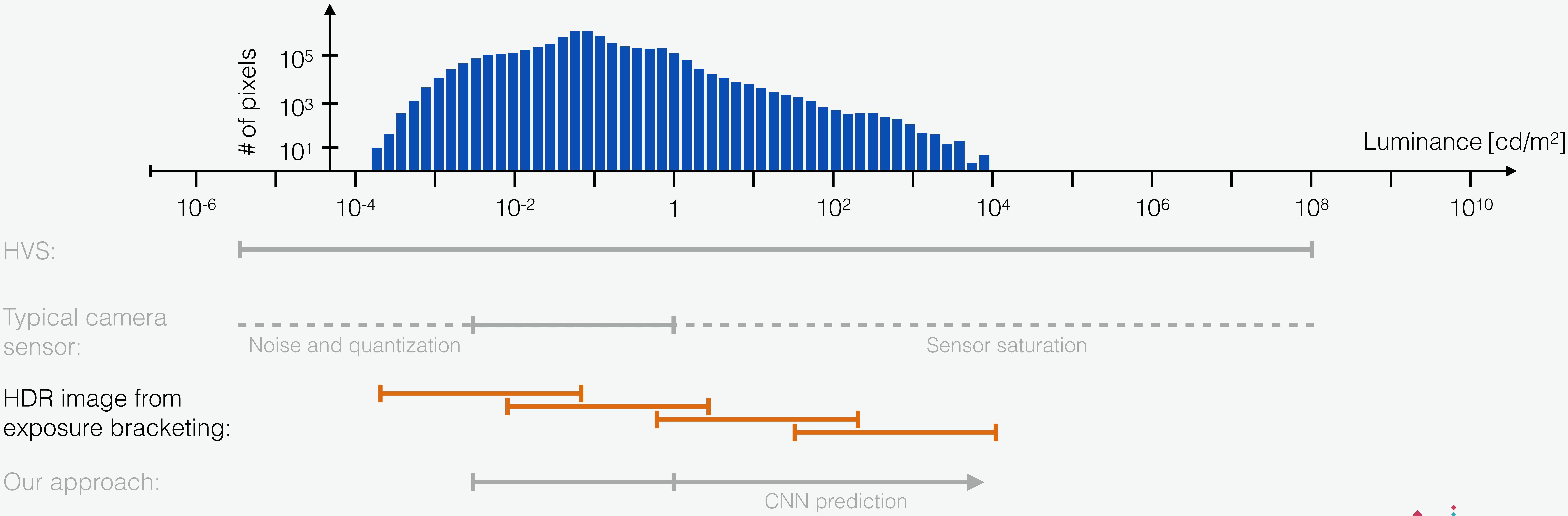
HDR scene



SIGGRAPH ASIA 2017 BANGKOK



HDR scene



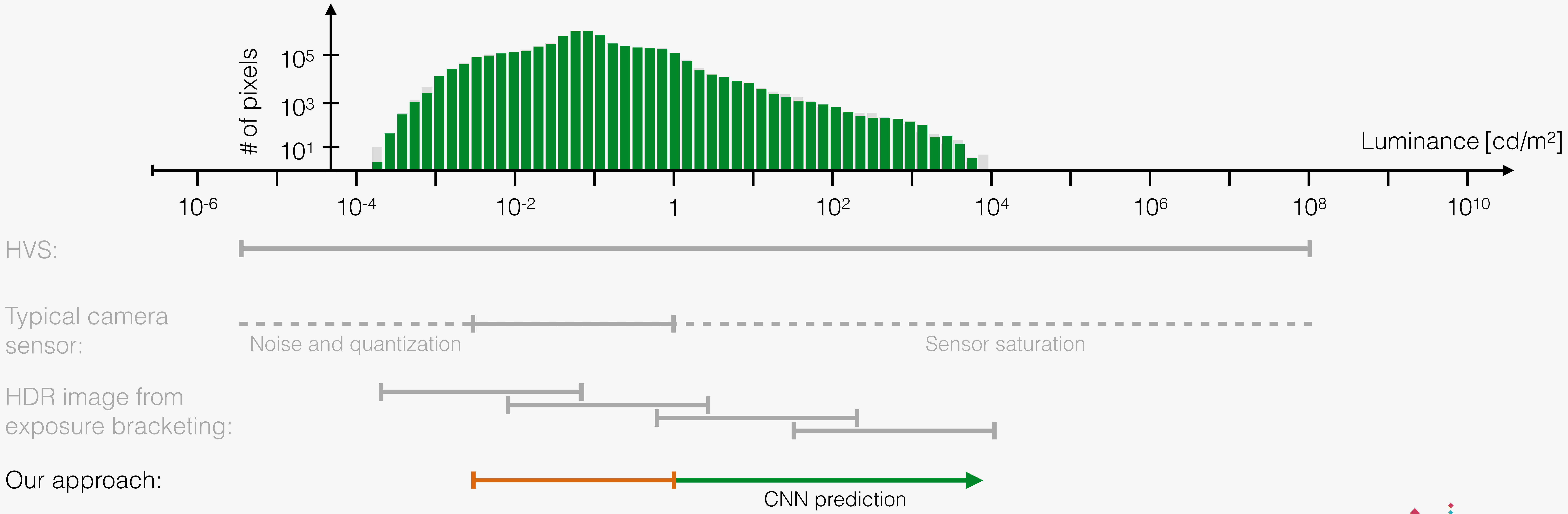
SIGGRAPH ASIA 2017 BANGKOK

HDR scene
Single exposure LDR image
“Lost in noise”
# of pixels
10^5
10^3
10^1
Luminance [cd/m^2]
10^-6
10^-4
10^-2
1
10^2
10^4
10^6
10^8
10^10
HVS:
Typical camera sensor:
Noise and quantization
Sensor saturation
HDR image from exposure bracketing:
Our approach:
CNN prediction

HDR scene

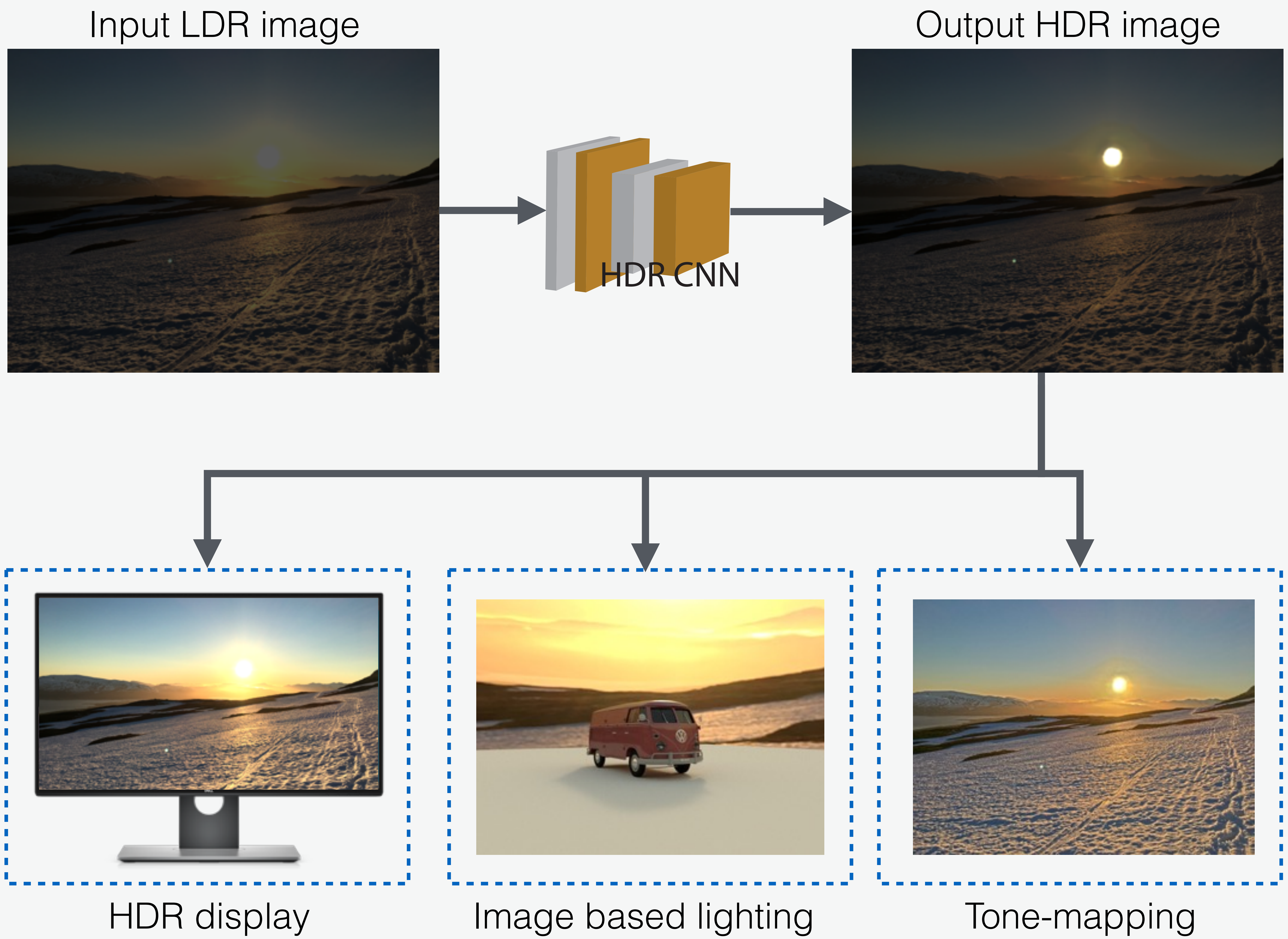
Single exposure LDR image

Reconstructed HDR scene



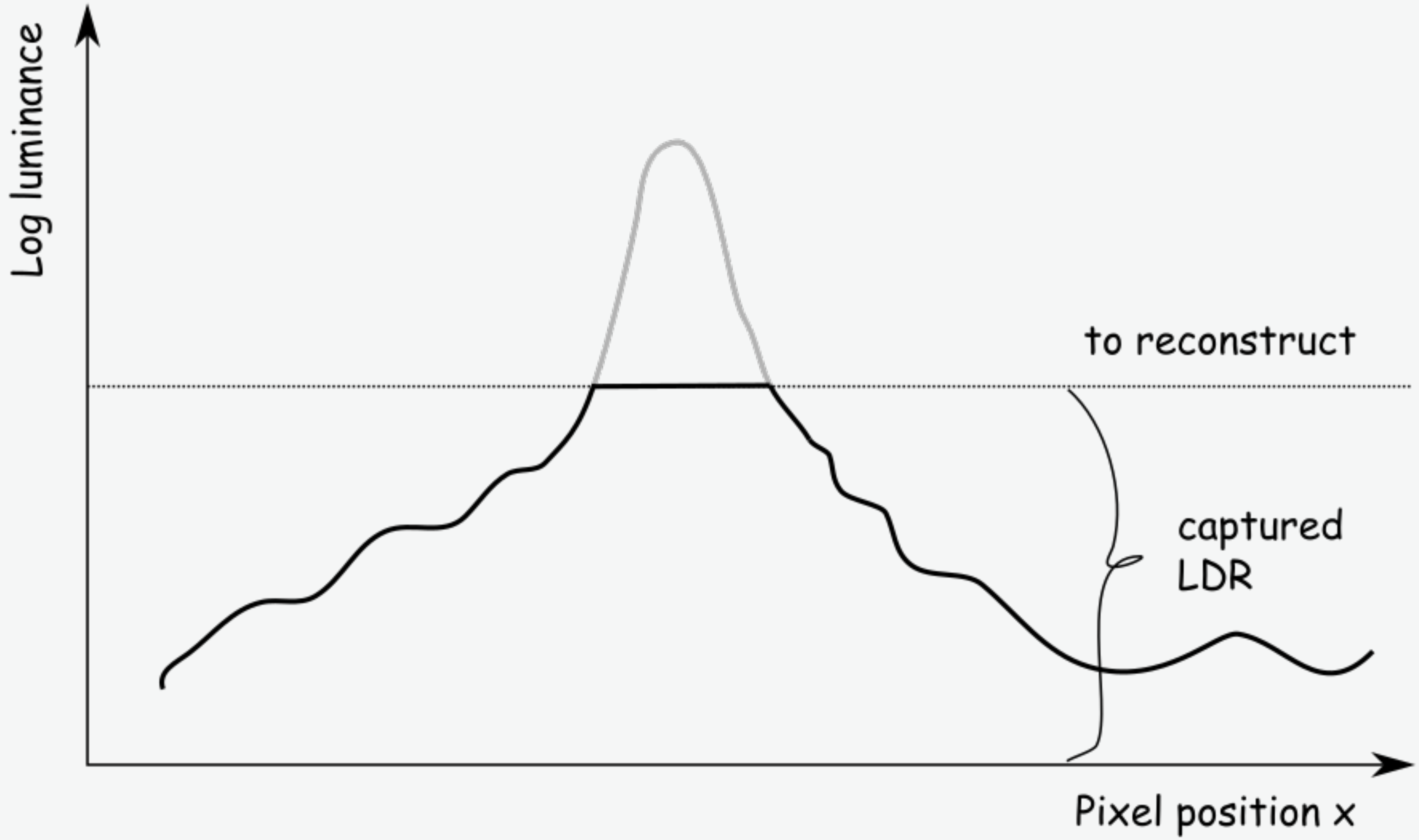
Motivation

- Reconstruction of saturated image regions
  - Most prominent feature of HDR images
  - Most HDR applications require this
- Many applications can benefit from HDR
  - HDR display
  - Image Based Lighting (IBL)
  - Post-processing (exposure correction, tone-mapping, glare simulation etc.)
- Majority of images captured with single exposure



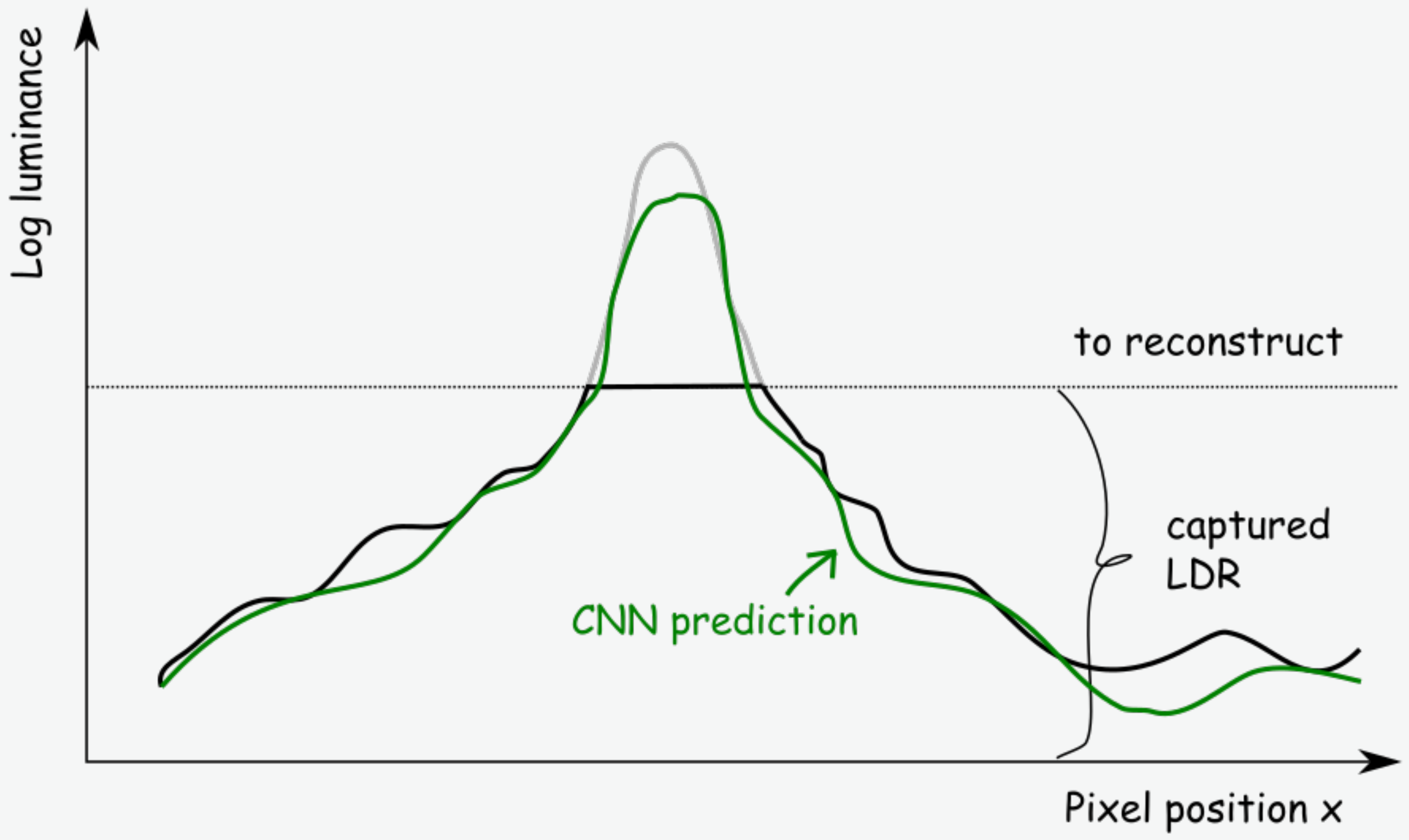
Method

Reconstruction formulation

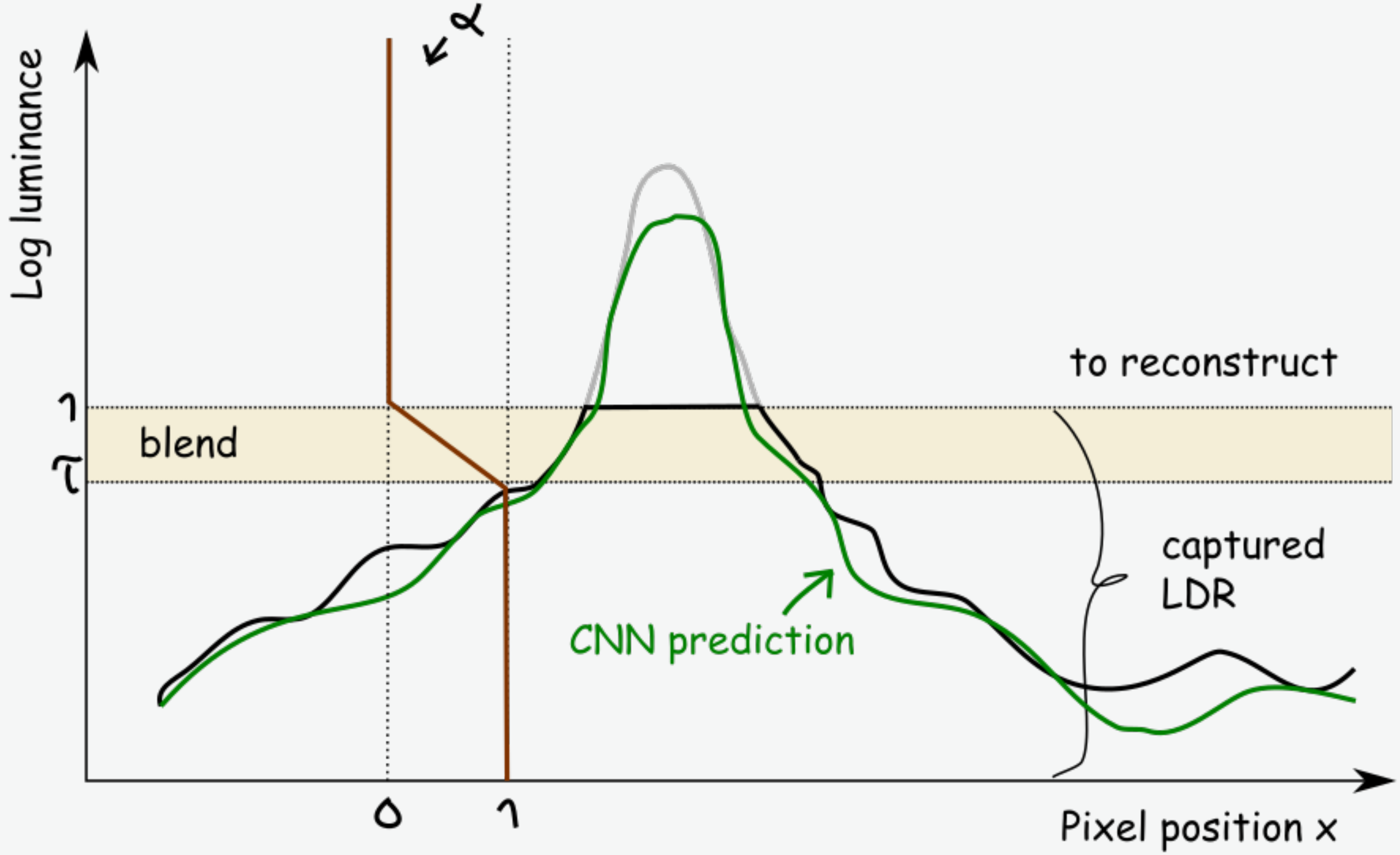


SIGGRAPH ASIA 2017 BANGKOK

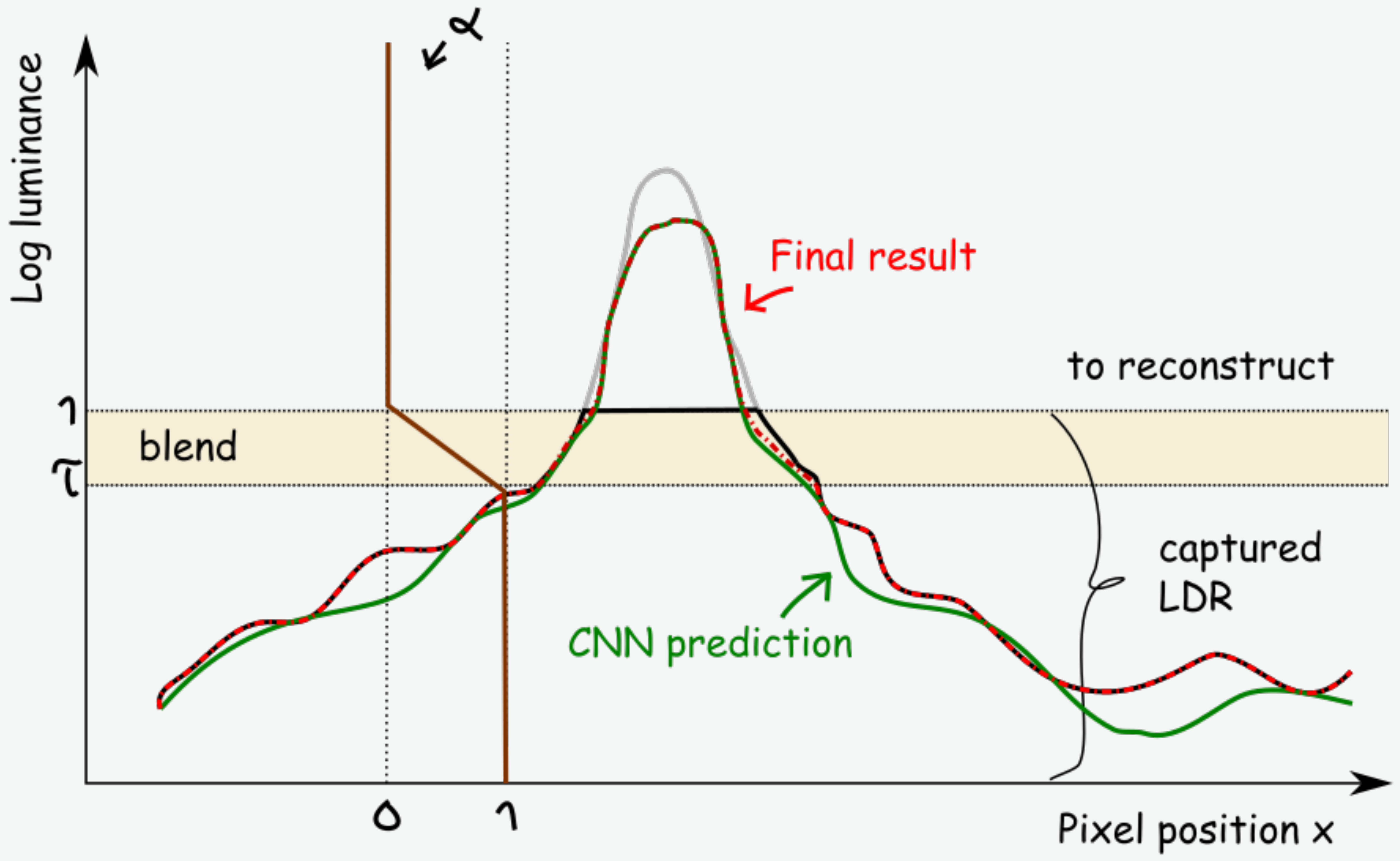
Reconstruction formulation



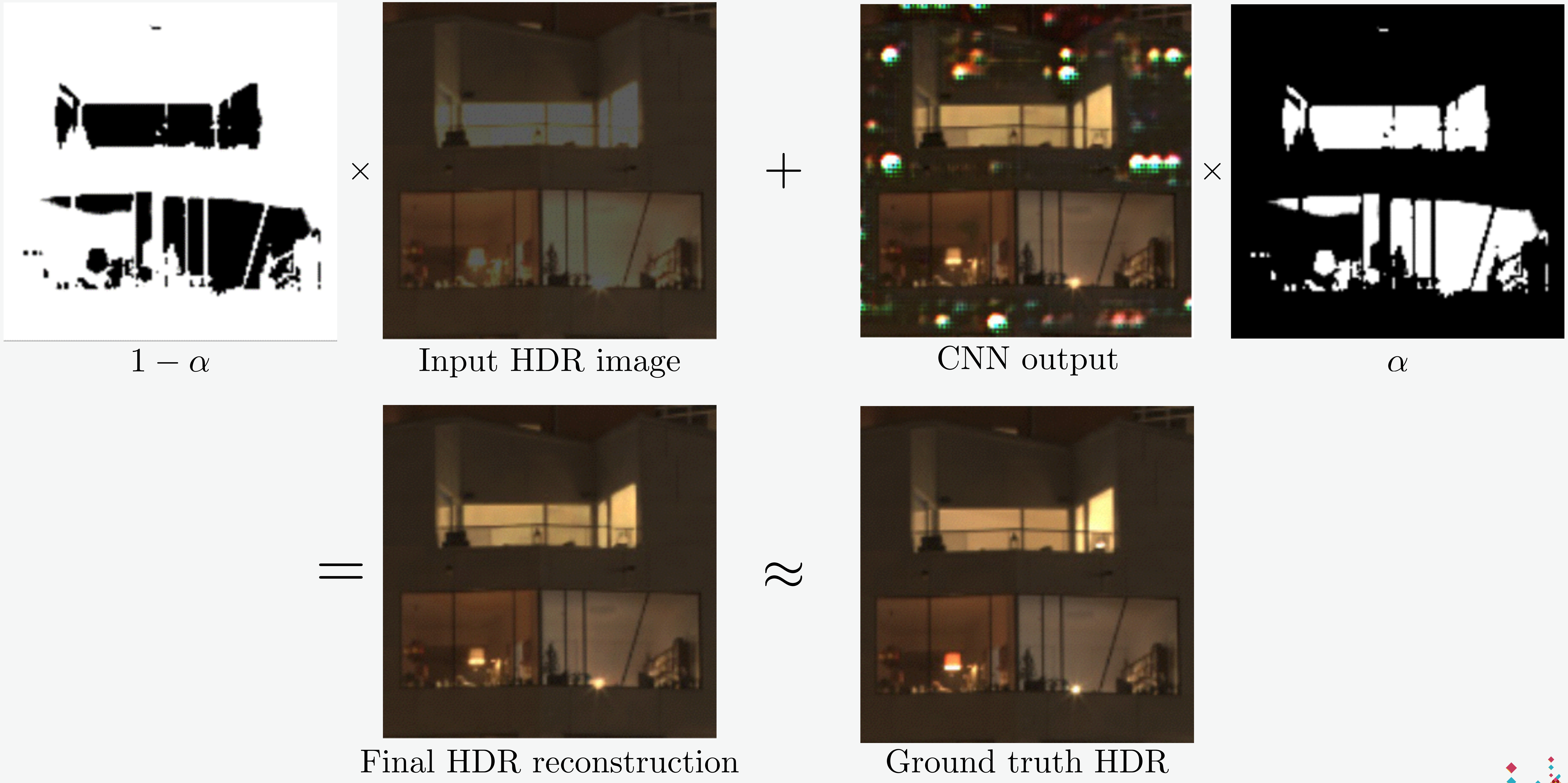
Reconstruction formulation



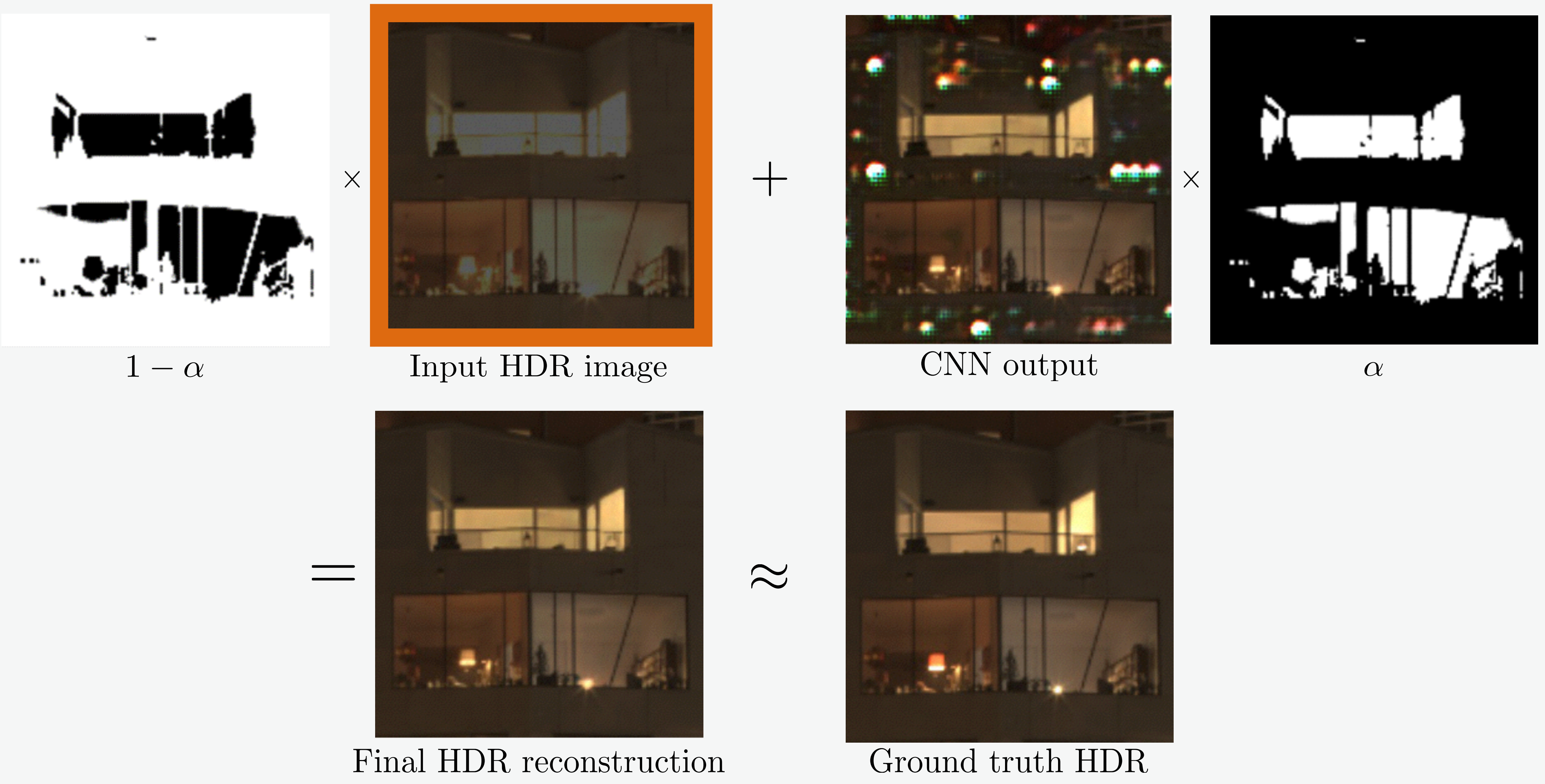
Reconstruction formulation



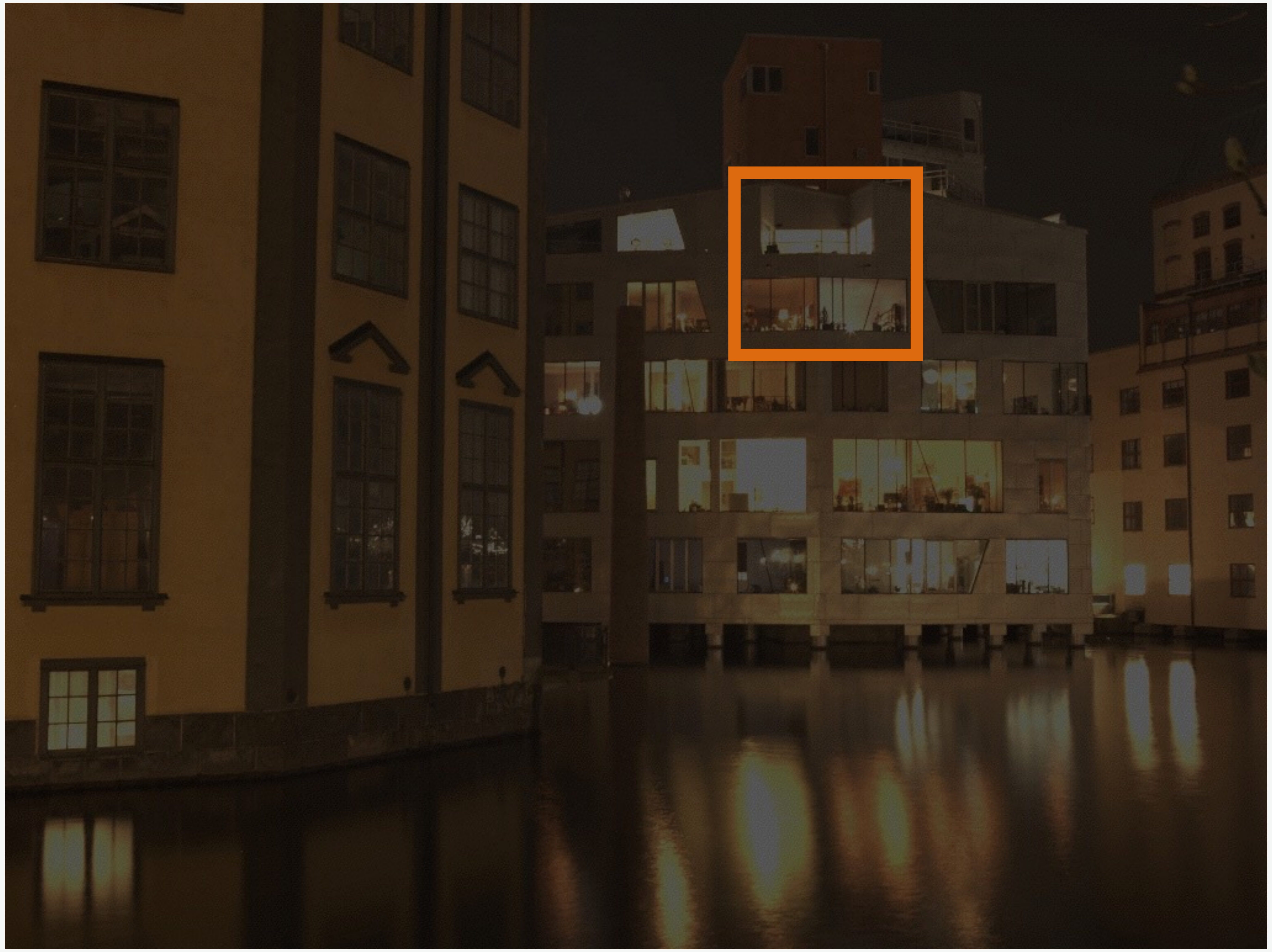
HDR blending



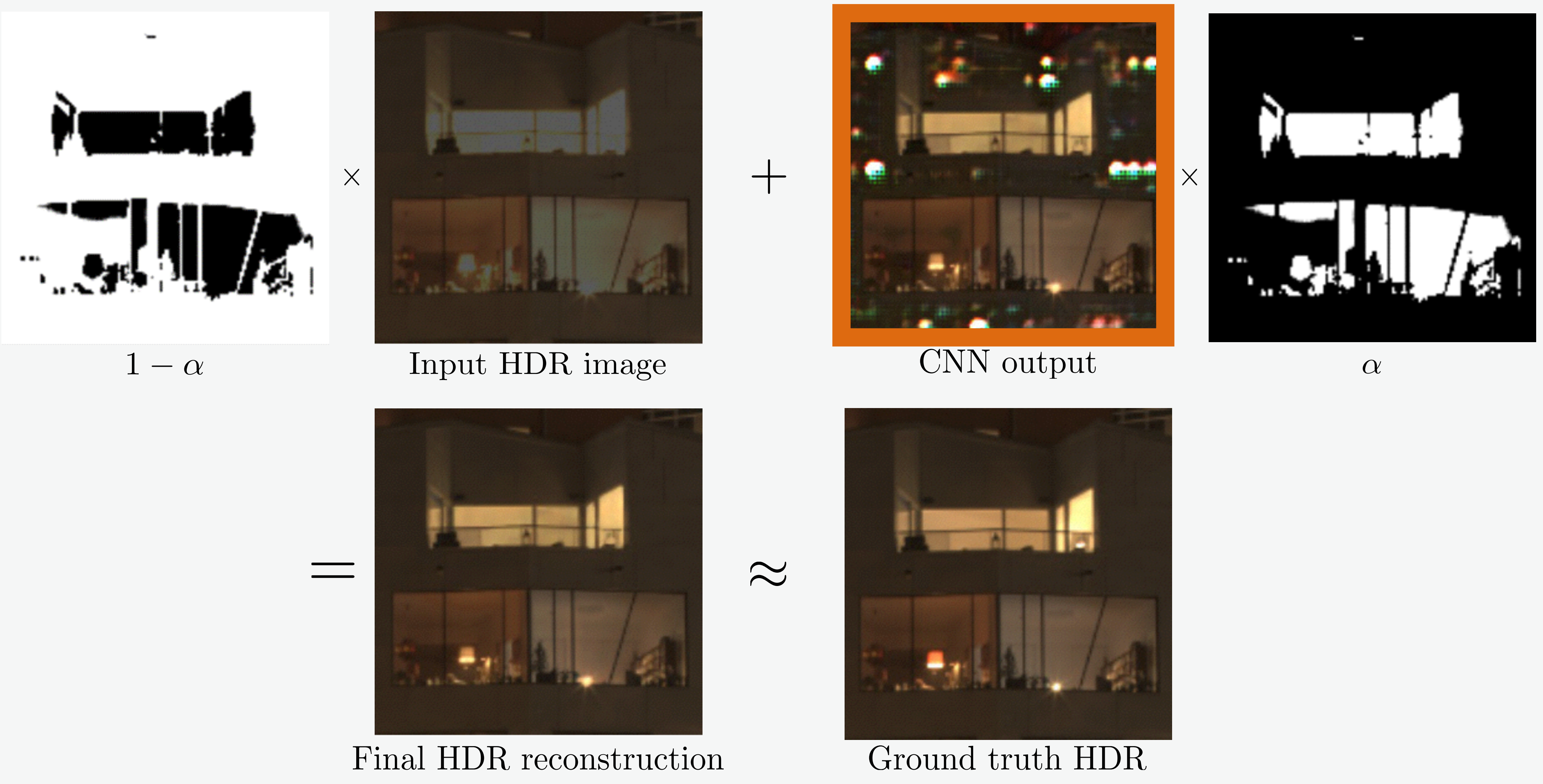
HDR blending



Input HDR image



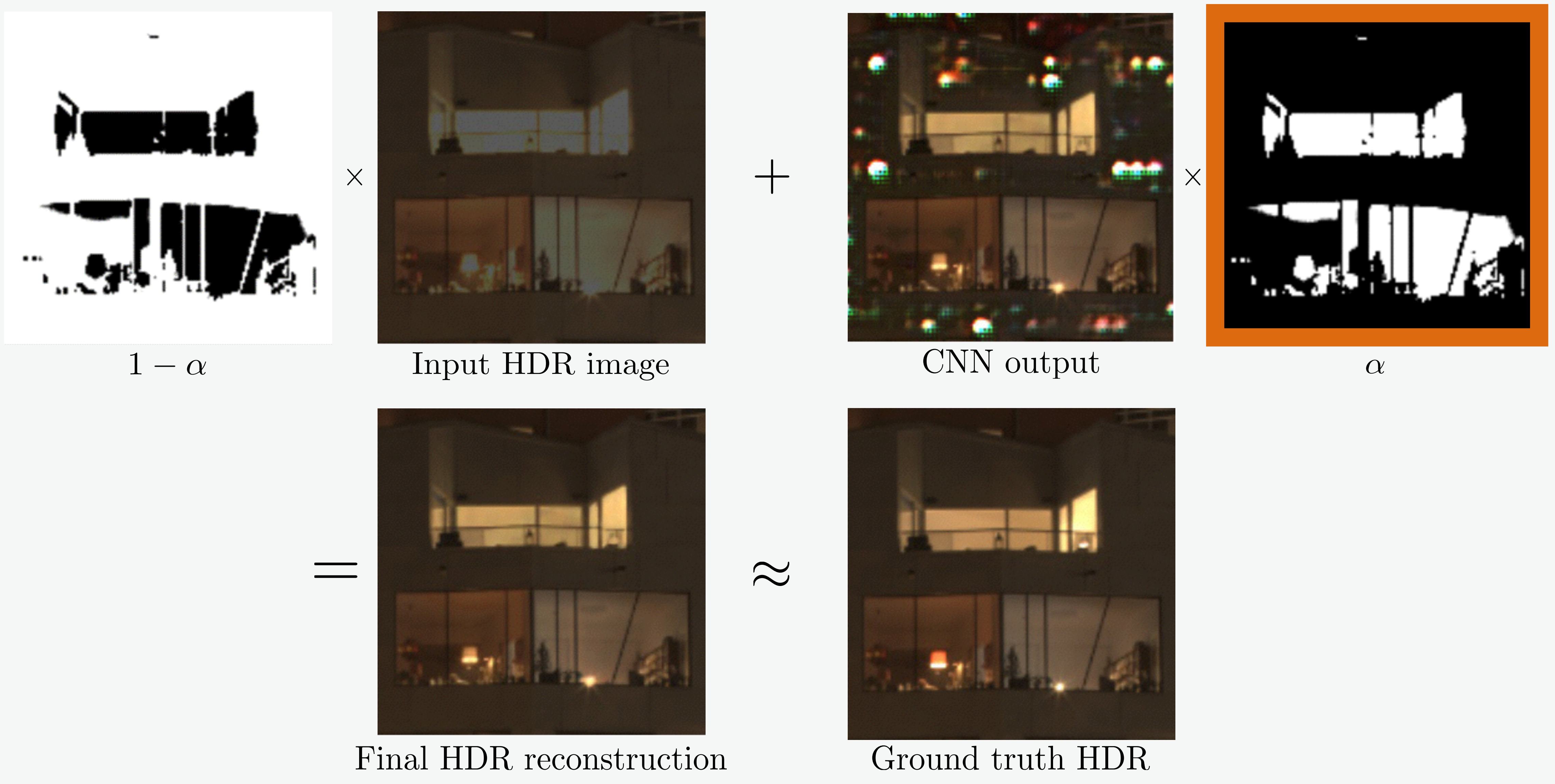
HDR blending



CNN output

SIGGRAPH ASIA 2017 BANGKOK

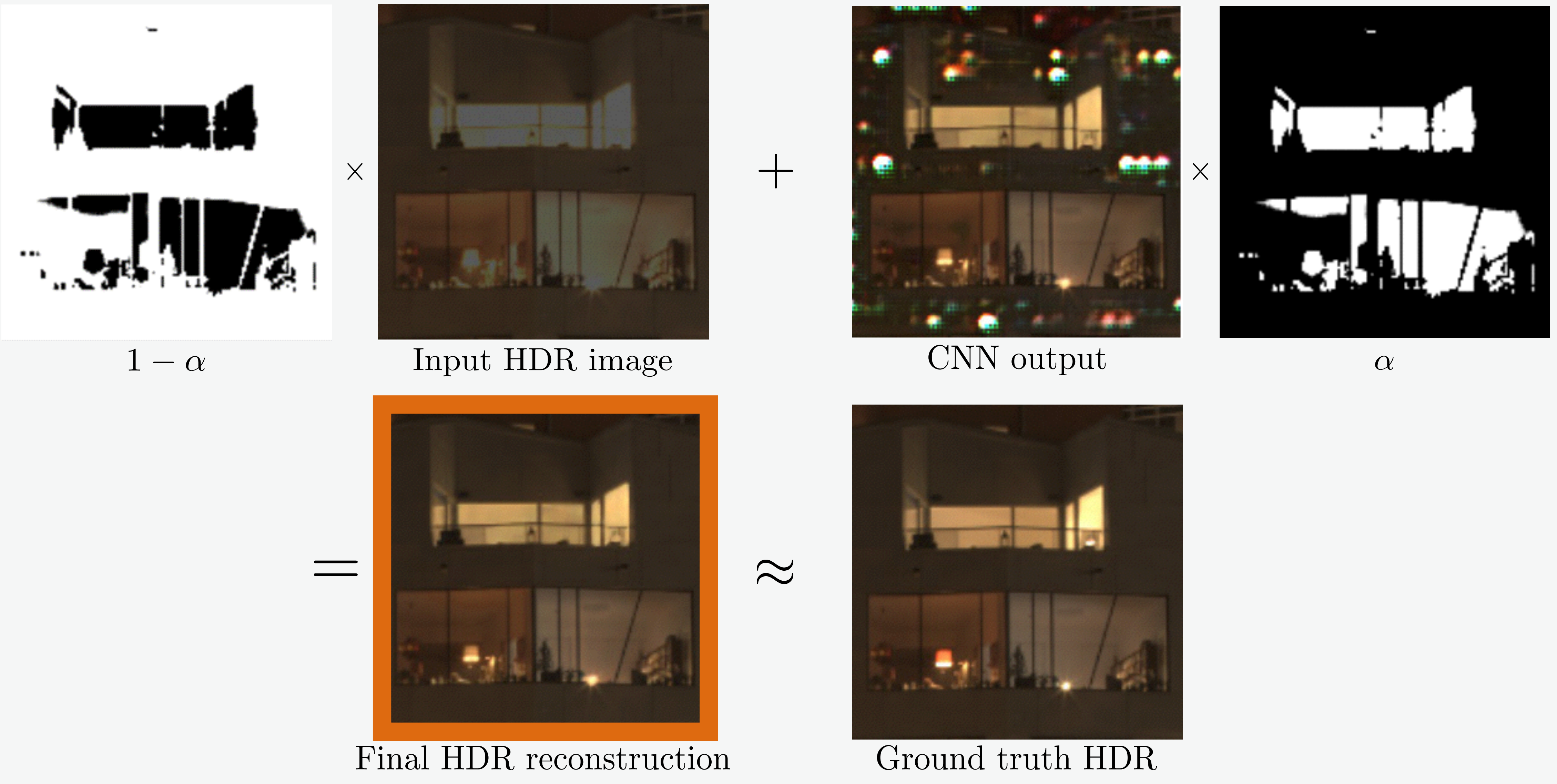
HDR blending



α

SIGGRAPH ASIA 2017 BANGKOK

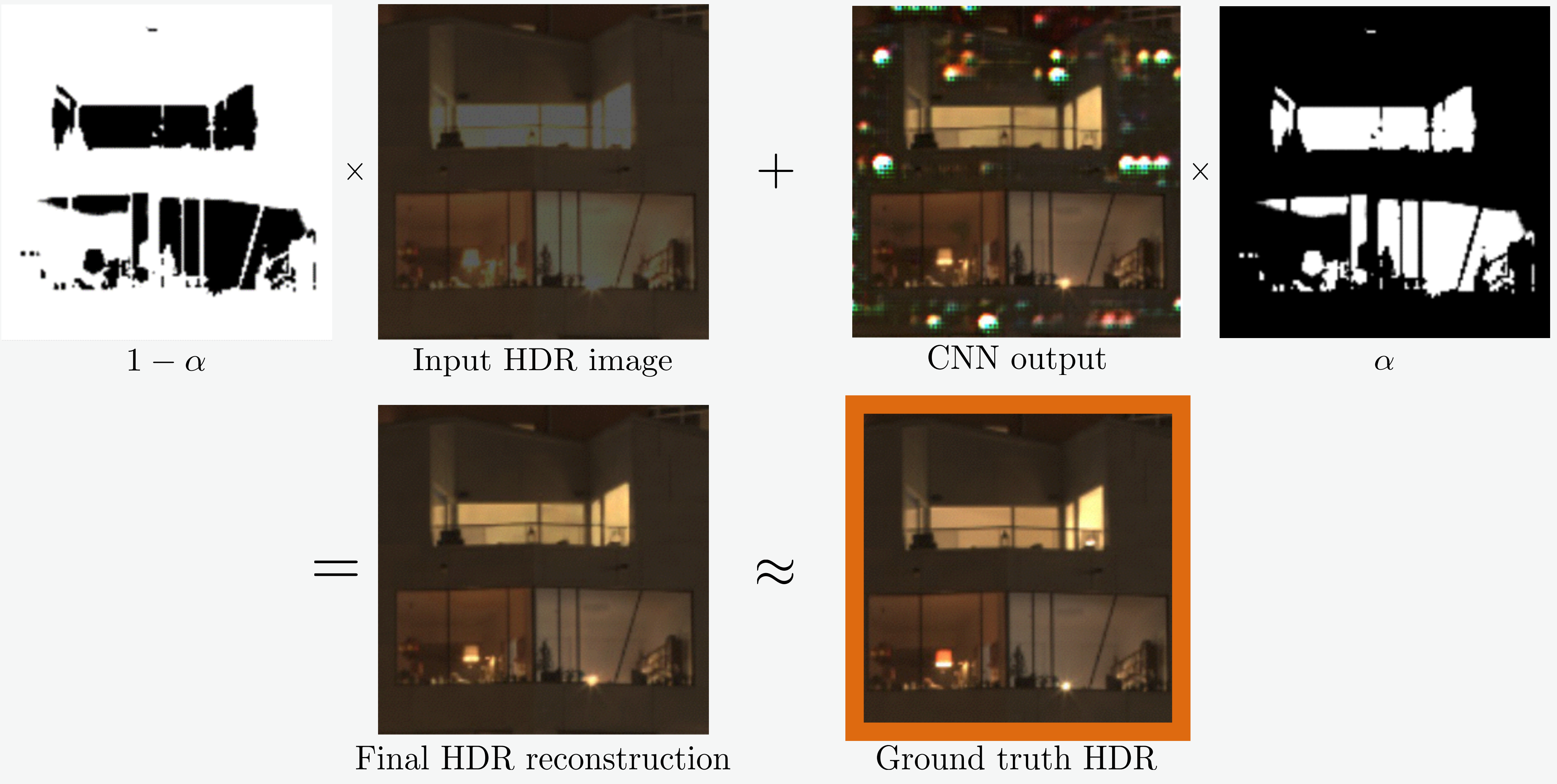
HDR blending



Final HDR reconstruction

SIGGRAPH ASIA 2017 BANGKOK

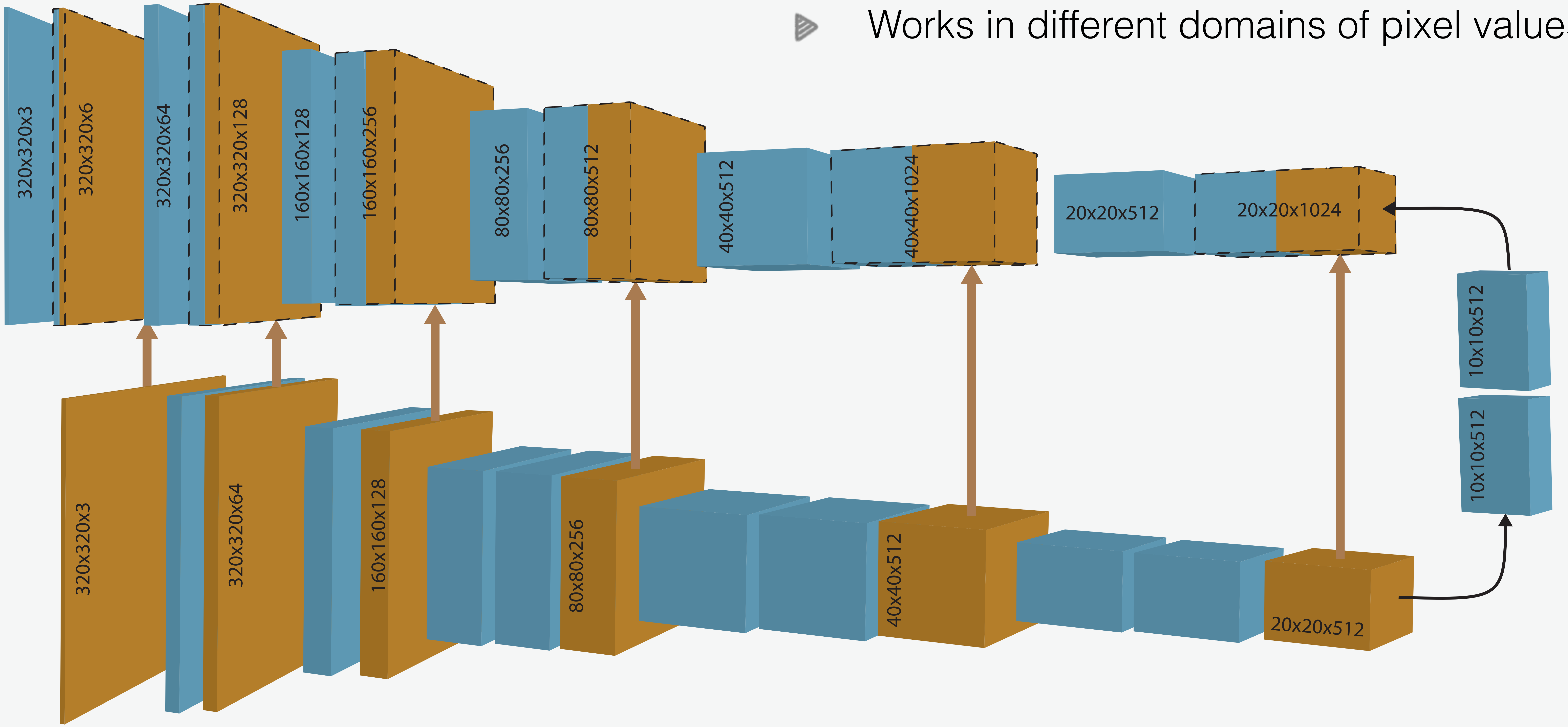
HDR blending



Ground truth HDR

Model

- HDR reconstruction autoencoder CNN
- Fully convolutional
- Works in different domains of pixel values

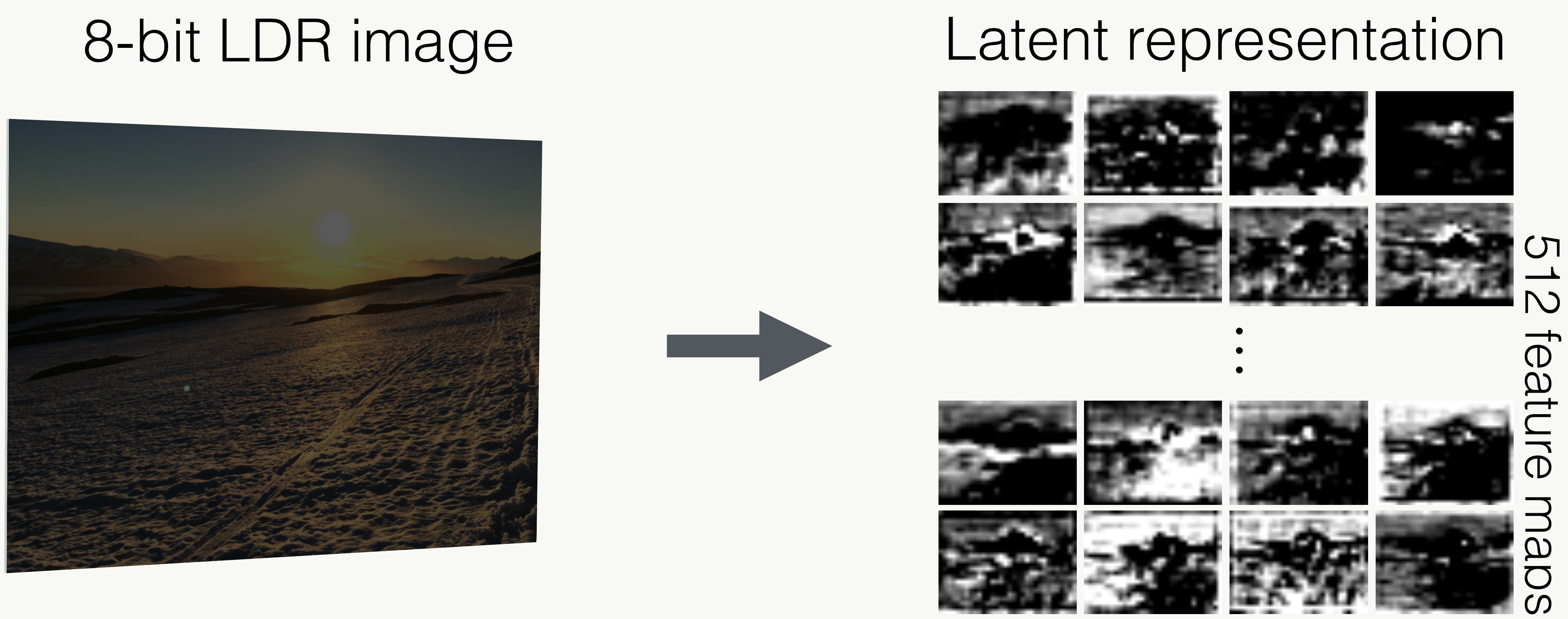


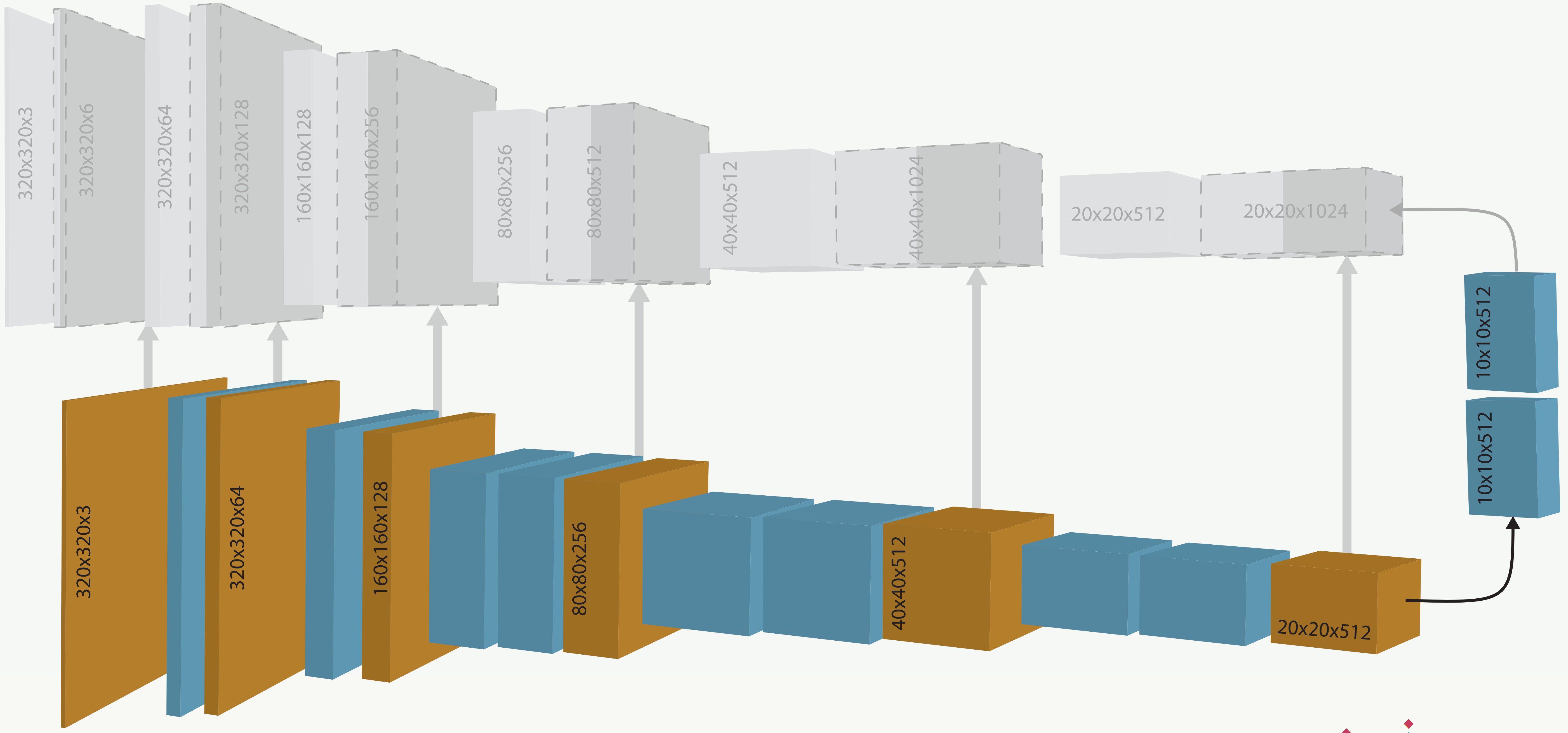
SIGGRAPH ASIA 2017 BANGKOK

Model

Encoder

- VGG16 architecture[1], without fully connected layers
- Processes LDR display values
- Convolution: 3x3 kernels
- Max-pooling for down-sampling
- ReLU activation

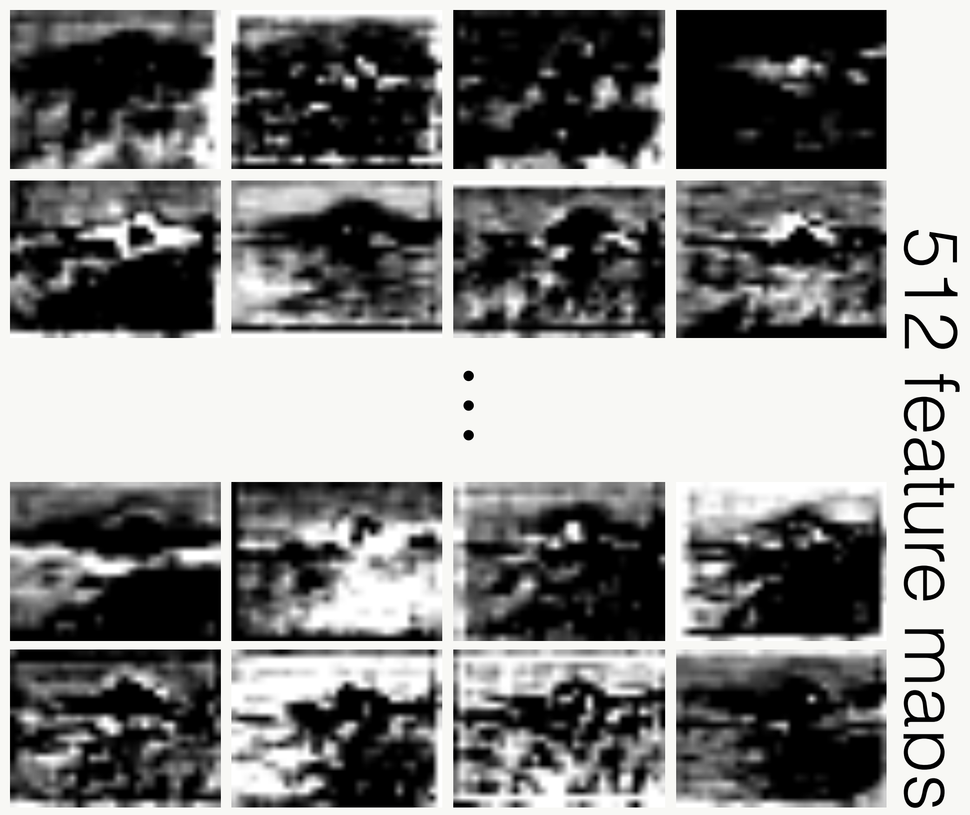




(1) K. Simonyan and A. Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR abs/1409.1556 (2014).

Model

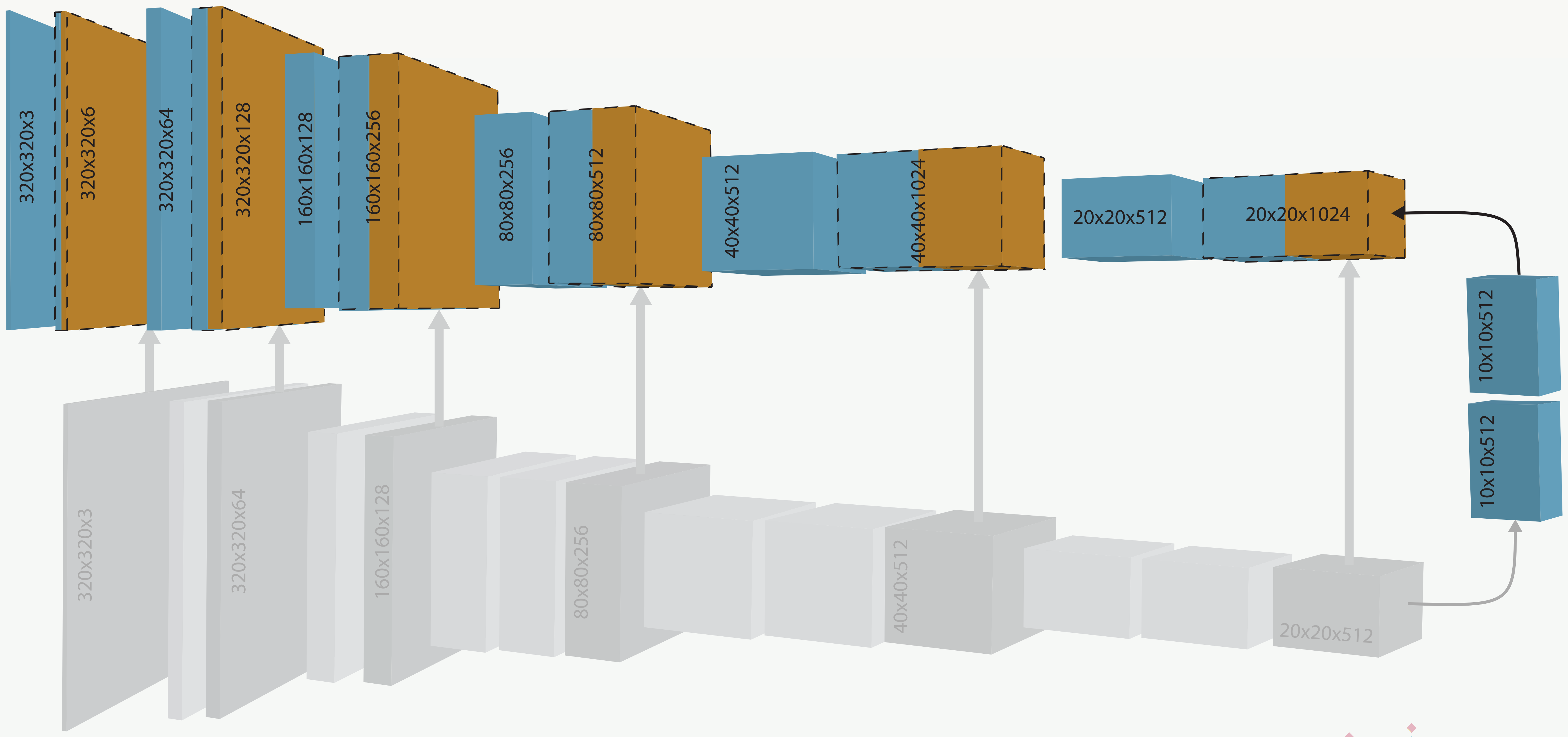
Latent representation



Log HDR image

Decoder

- Log domain processing
- De-convolutional up-sampling: 4x4 kernels
- Kernels initialized for bilinear up-sampling
- ReLU activation
- Batch normalization after each layer

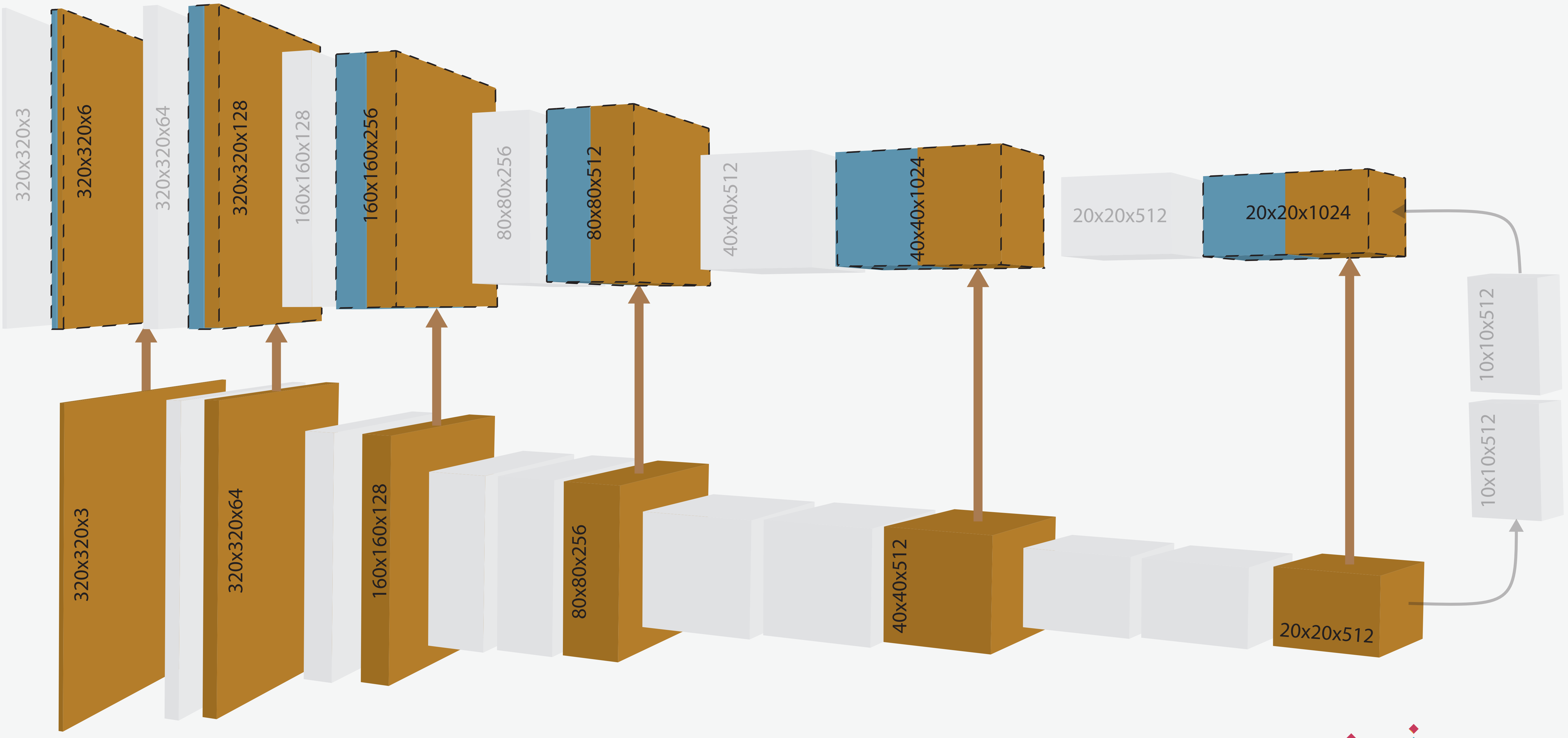


Model

Skip-connections

- Skip-connections from encoder to decoder, at each level of resolution
- Concatenated encoder and decoder layers
- Fusion: 1x1 convolution
- Skip includes domain transformation
- Improves reconstructed details

LDR input | HDR, without skip | HDR, with skip | Ground truth



Loss function

- Loss function separated in illuminance and reflectance terms
- Illuminance
  - Describes global variations
  - Responsible for the high dynamic range
  - Single channel (monochromatic)
  - Estimated from low-pass filtering of luminance
- Reflectance
  - Details and colors
  - Low dynamic range

Luminance Illuminance Reflectance

$$L = IR$$

Assumes all Lambertian surfaces!

$$log(I) = G_\sigma * log(L)$$

Assumes no sharp boundaries!

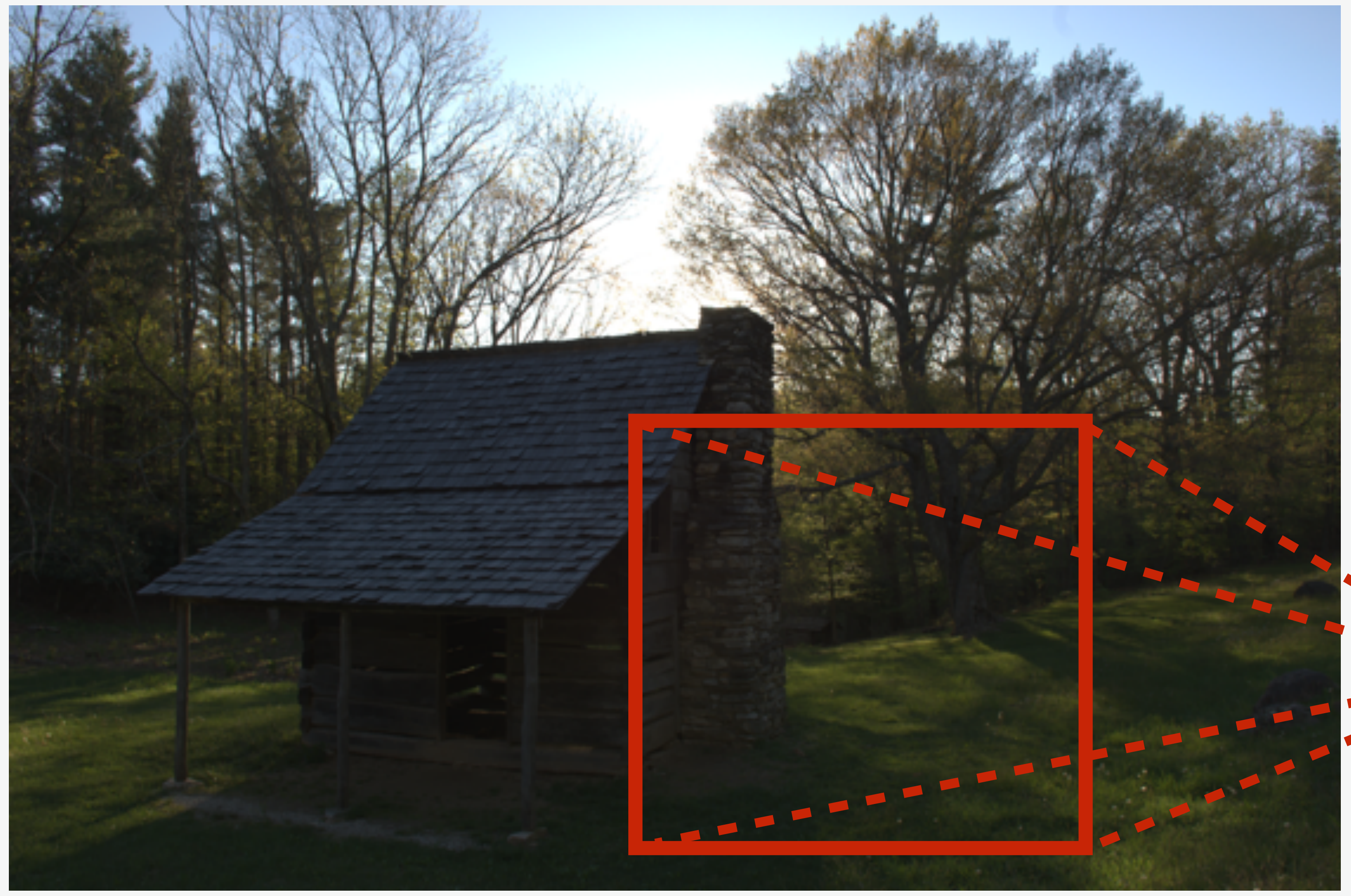
Loss function

$$\mathcal{L}_{IR} = \lambda \sum \left| \alpha \left(log(\hat{I}) - log(I) \right) \right|^2 + (1-\lambda) \sum \left| \alpha \left(log(\hat{R}) - log(R) \right) \right|^2$$

Scalar, controlling relative importance of illuminance/reflectance

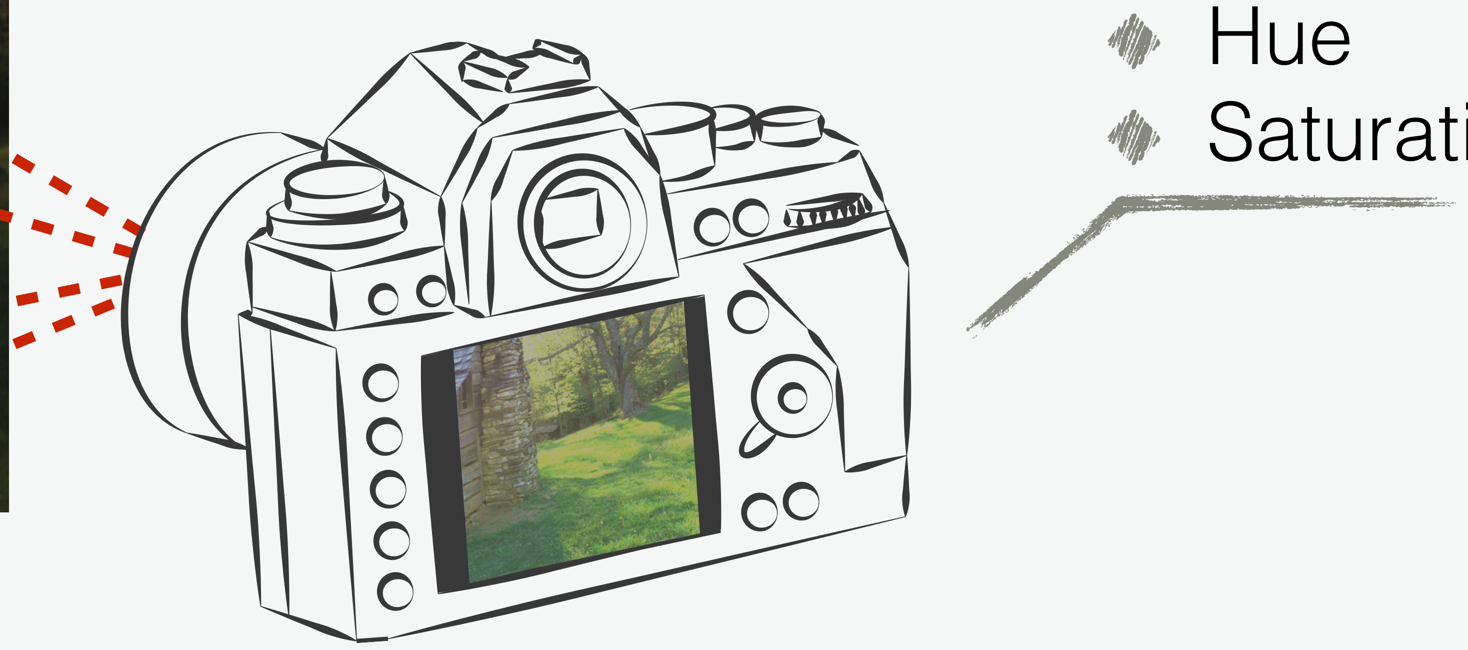
Training data

- HDR image data
  - Mostly online resources
  - 1121 images + 67 videos
- Database
  - Every 10th frame from HDR video sequences
  - In total ~3700 high resolution HDR images
- Training data
  - Augmentation from virtual camera
  - In total ~125K images used in training

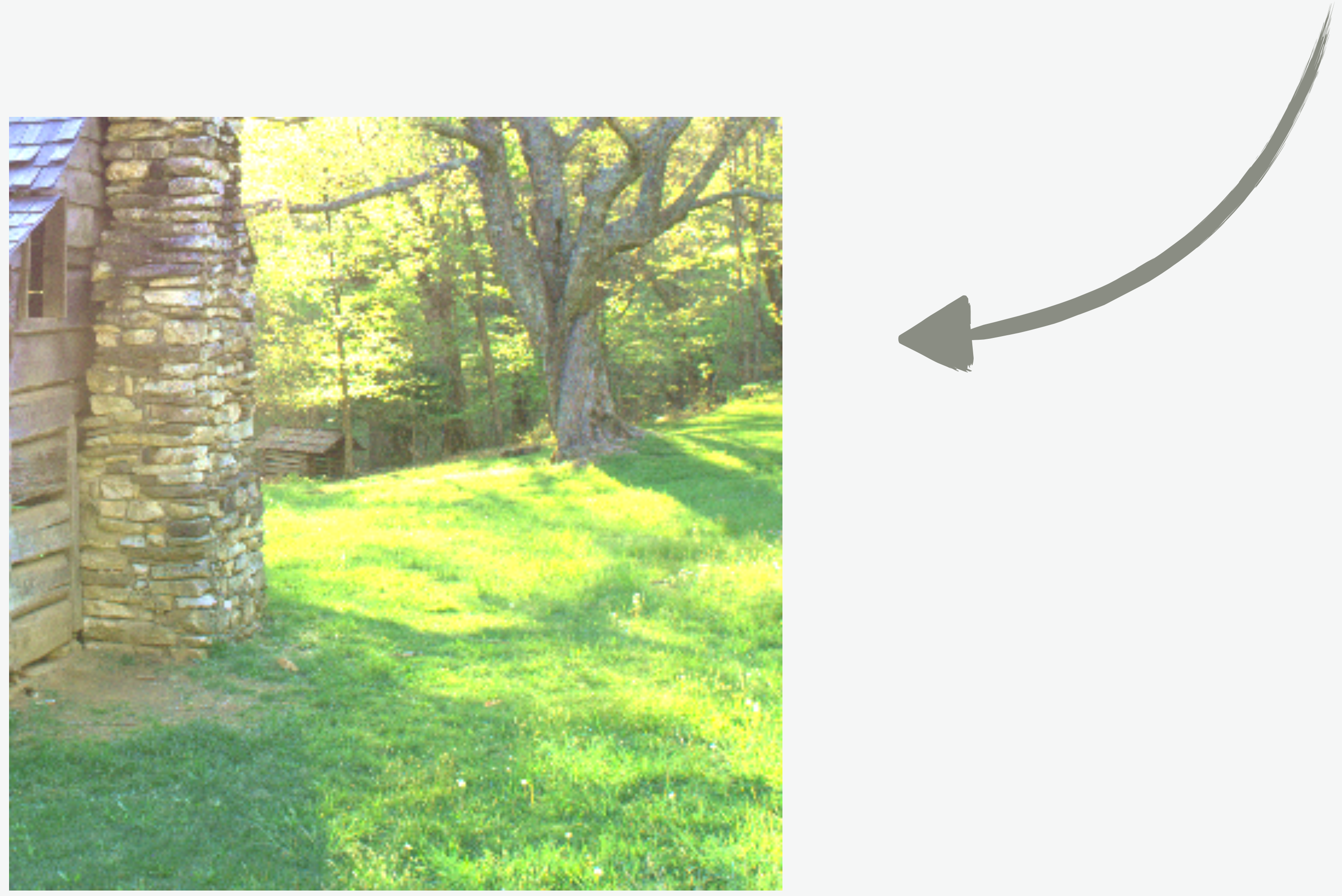


Input HDR image

Virtual camera



- Exposure
- Camera curve
- Crop
- Flip
- Noise
- Hue
- Saturation

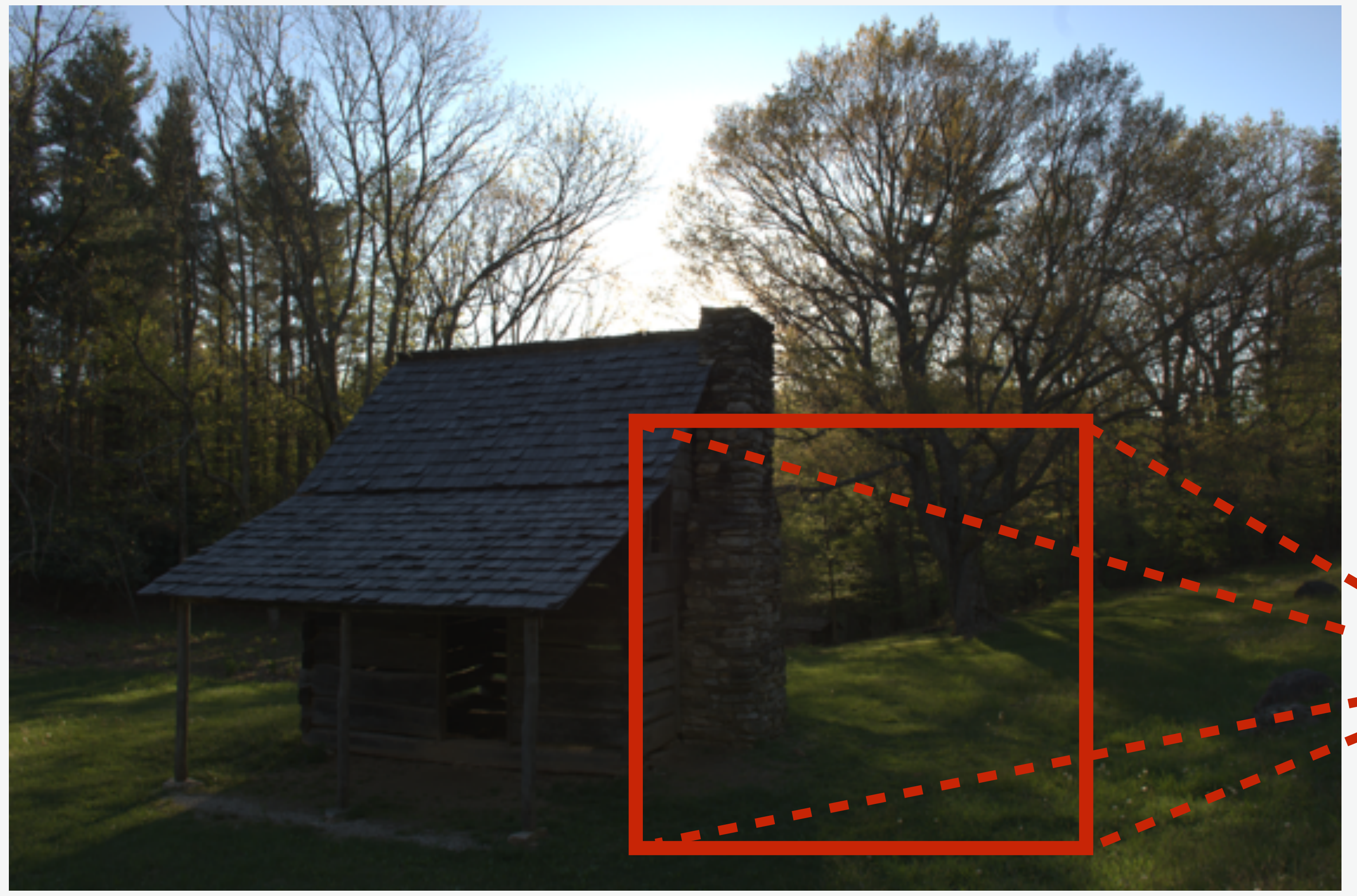


Training image

SIGGRAPH ASIA 2017 BANGKOK

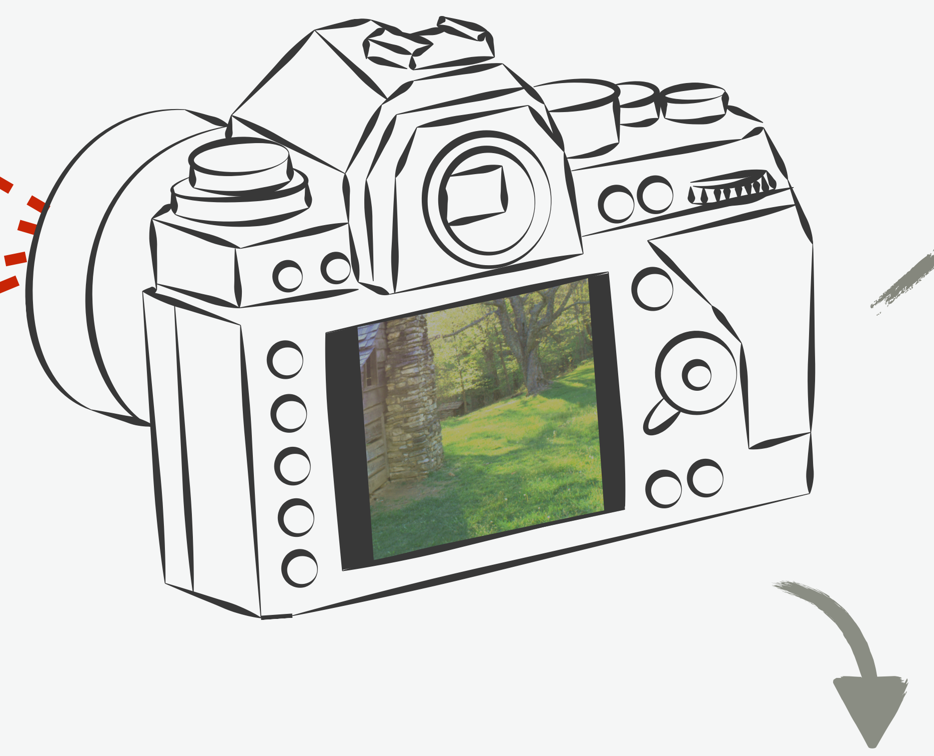
Training data

- HDR image data
  - Mostly online resources
  - 1121 images + 67 videos
- Database
  - Every 10th frame from HDR video sequences
  - In total ~3700 high resolution HDR images
- Training data
  - Augmentation from virtual camera
  - In total ~125K images used in training



Input HDR image

Virtual camera

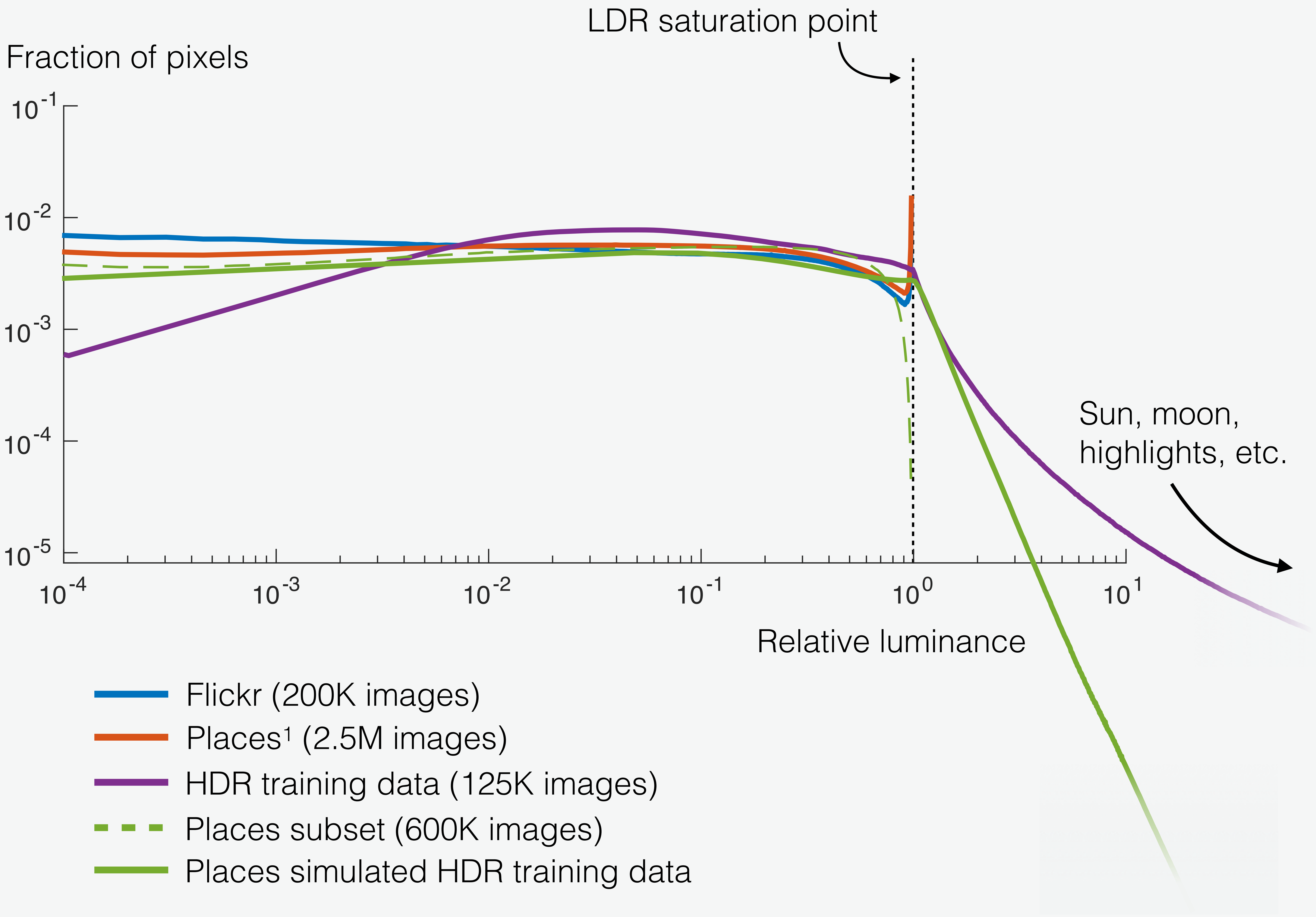


- Exposure
- Camera curve
- Crop
- Flip
- Noise
- Hue
- Saturation

Training images

Training data

- HDR vs. LDR image databases
- Simulated HDR data
    - Subset of Places[1] images, without saturated image regions
    - Linearized to create simulated HDR data
    - Captured with the virtual camera
    - For pre-training, cannot replace true HDR data

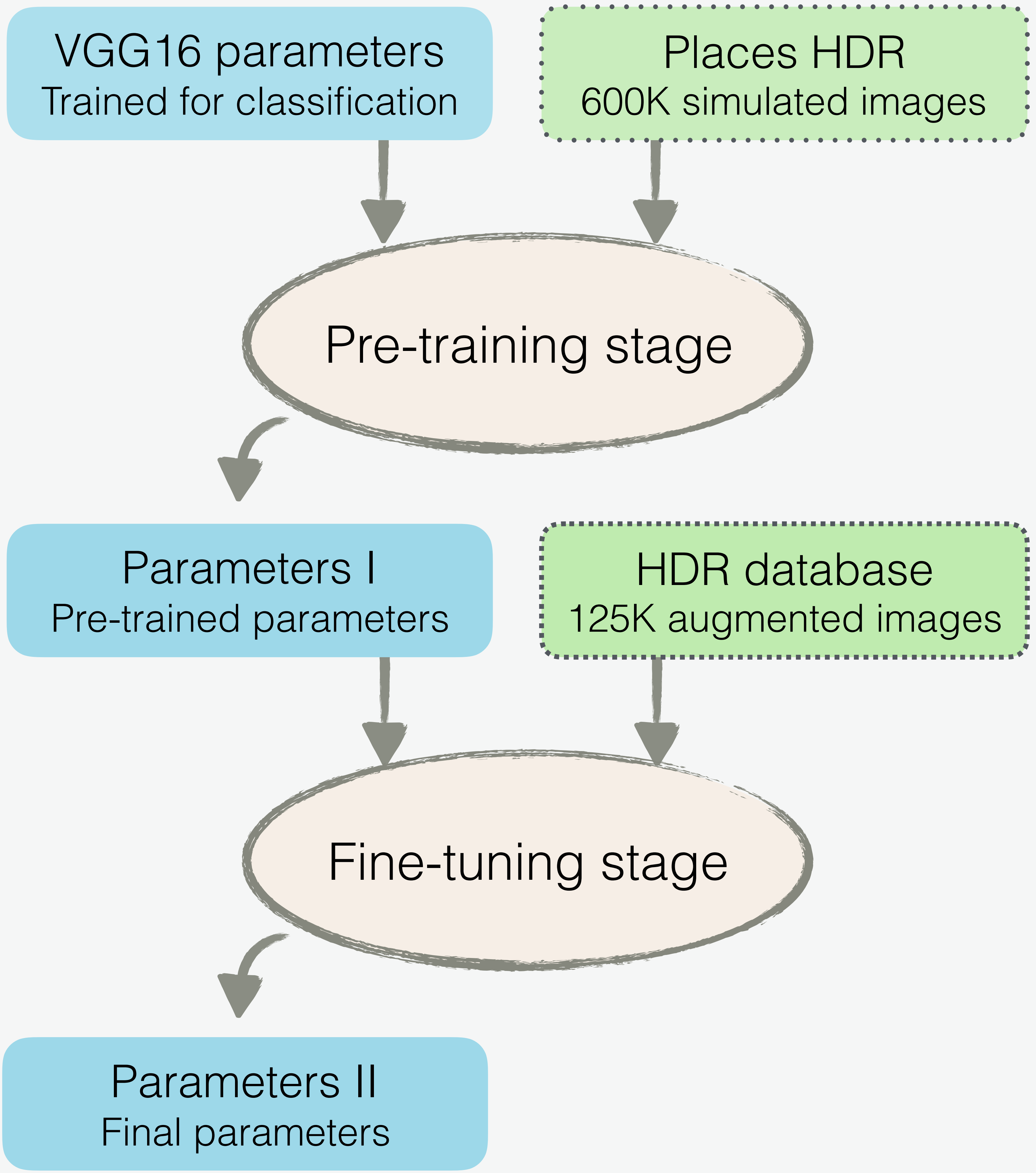


(1) B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. 2014. Learning Deep Features for Scene Recognition using Places Database. In Advances in Neural Information Processing Systems 27. 487–495.

SIGGRAPH ASIA 2017 BANGKOK

Training

- 2-stage training
- 1st stage
  - VGG16 weights
    - For initializing encoder
    - Trained for classification on Places database
  - Places simulated HDR database (224x224)
- 2nd stage
  - Fine-tuning using the gathered HDR database (320x320)
- ADAM[1] optimizer



(1) D. P. Kingma and J. Ba. 2014. Adam: A Method for Stochastic Optimization. CoRR abs/1412.6980 (2014). http://arxiv.org/abs/1412.6980

SIGGRAPH ASIA 2017 BANGKOK

Results

Comparison to ground truth

Reconstruction from image captured by virtual camera

SIGGRAPH ASIA 2017 BANGKOK

Comparison to ground truth

Reconstruction from image captured by virtual camera

Comparison to ground truth

Reconstruction from image captured by virtual camera

SIGGRAPH ASIA 2017 BANGKOK

Comparison to ground truth

Reconstruction from image captured by virtual camera

SIGGRAPH ASIA 2017 BANGKOK

Comparison to ground truth

Reconstruction from image captured by virtual camera

SIGGRAPH ASIA 2017 BANGKOK

Comparison to ground truth

Reconstruction from image captured by virtual camera

SIGGRAPH ASIA 2017 BANGKOK

Reconstruction with real-world cameras

Camera: Canon EOS 5D Mark II

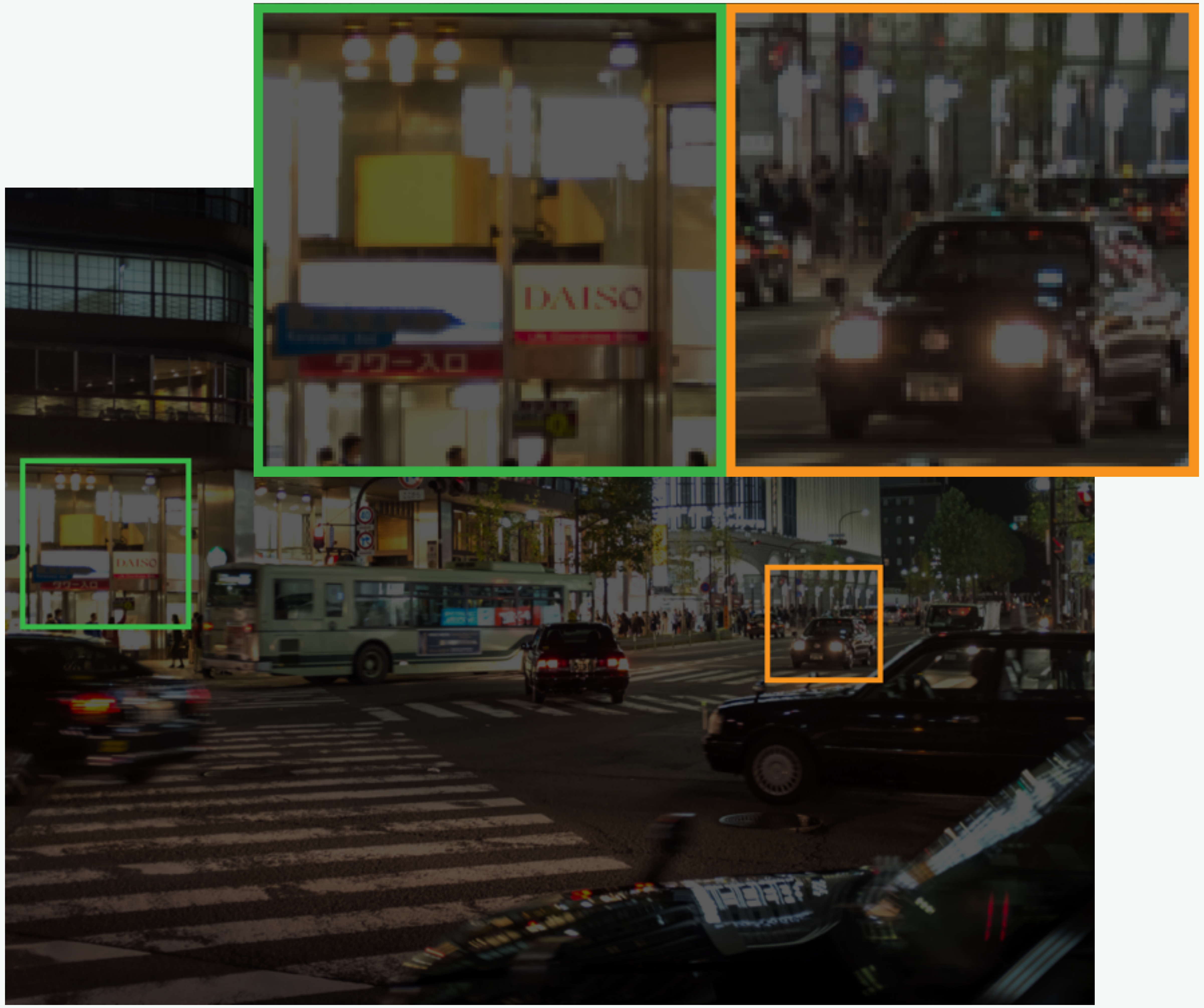
Input

Reconstruction

SIGGRAPH ASIA 2017 BANGKOK

Reconstruction with real-world cameras

Camera: Fuji X100S



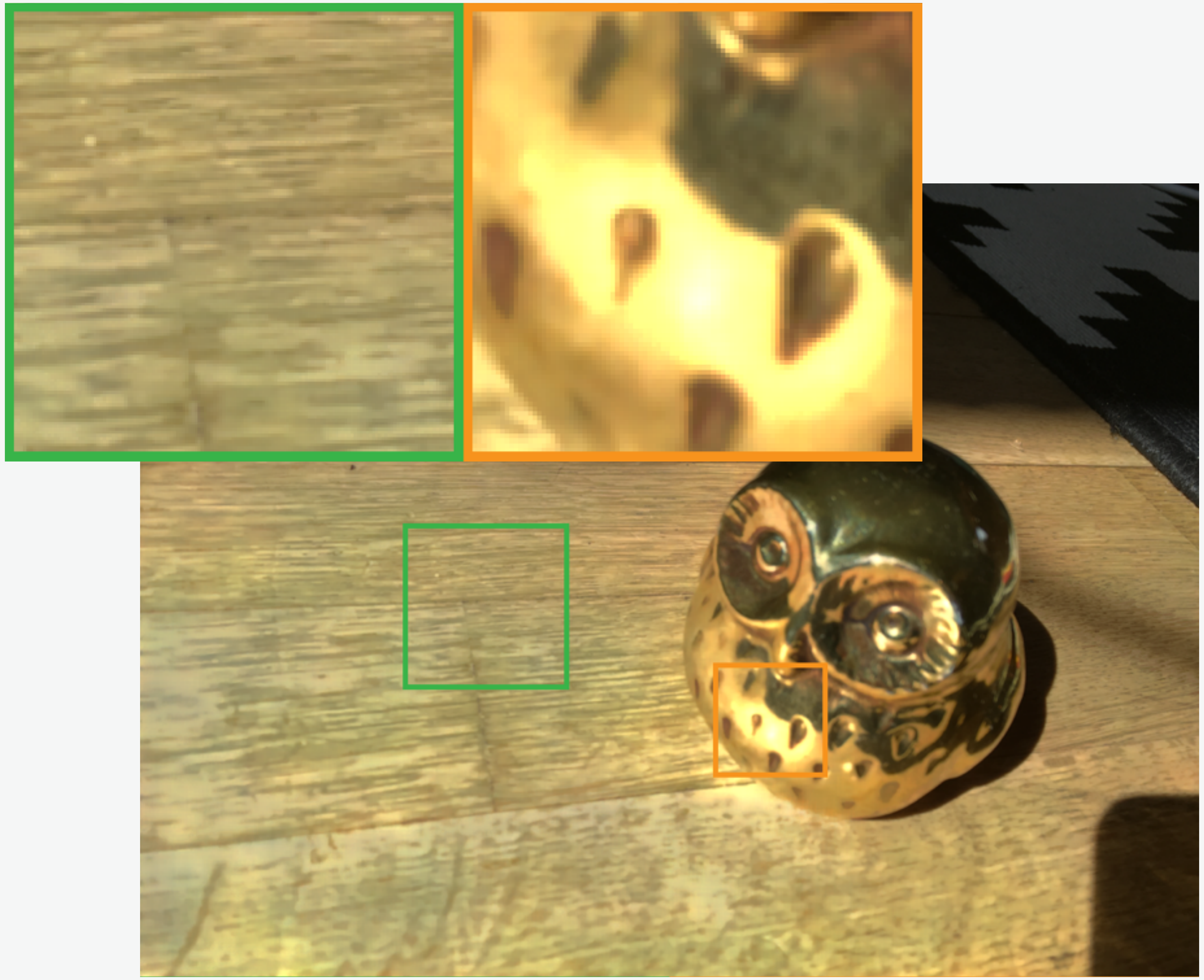
Input

Reconstruction

Reconstruction with real-world cameras

Camera: iPhone 6S

Input



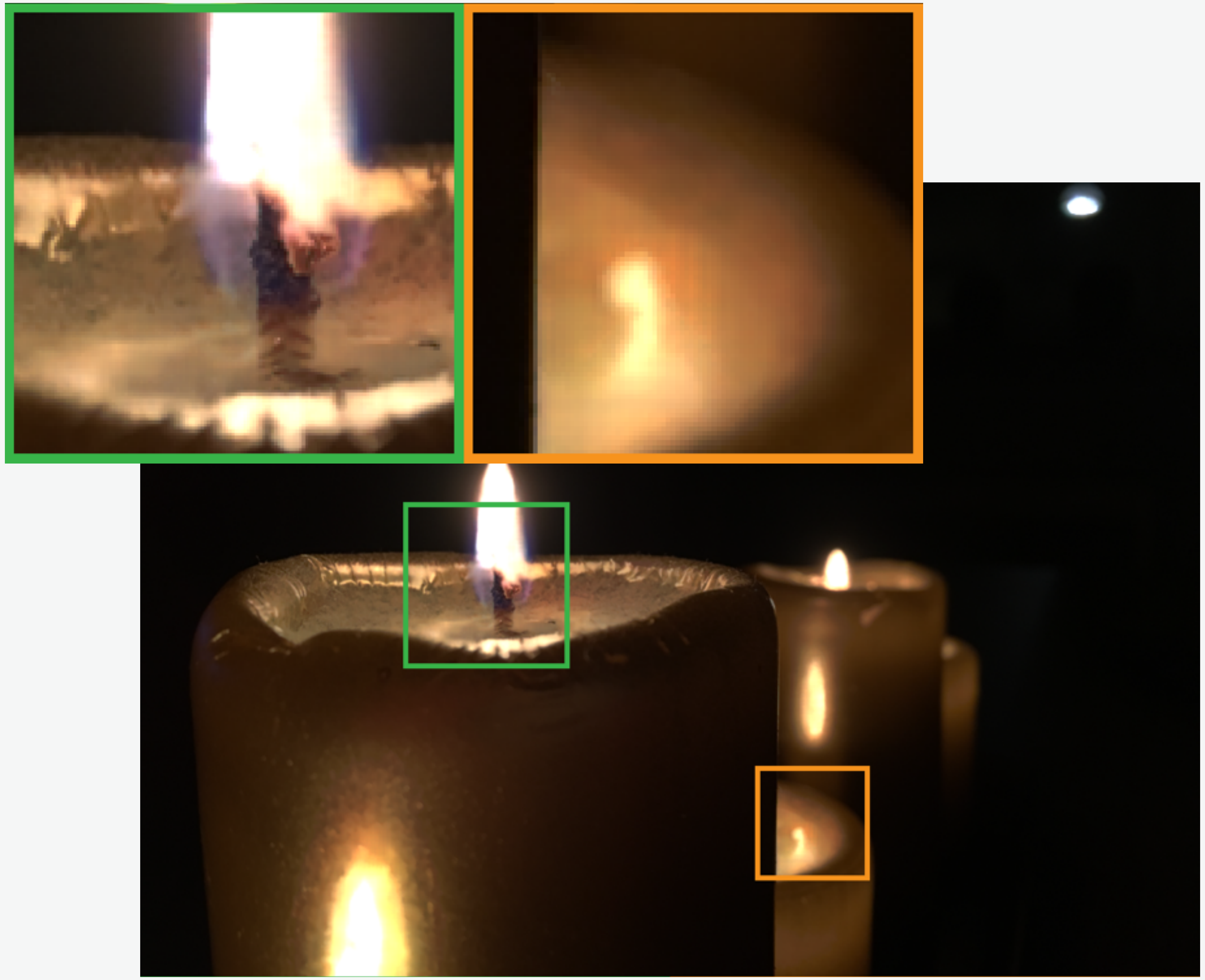
Reconstruction

SIGGRAPH ASIA 2017 BANGKOK

Reconstruction with real-world cameras

Camera: iPhone 6S

Input

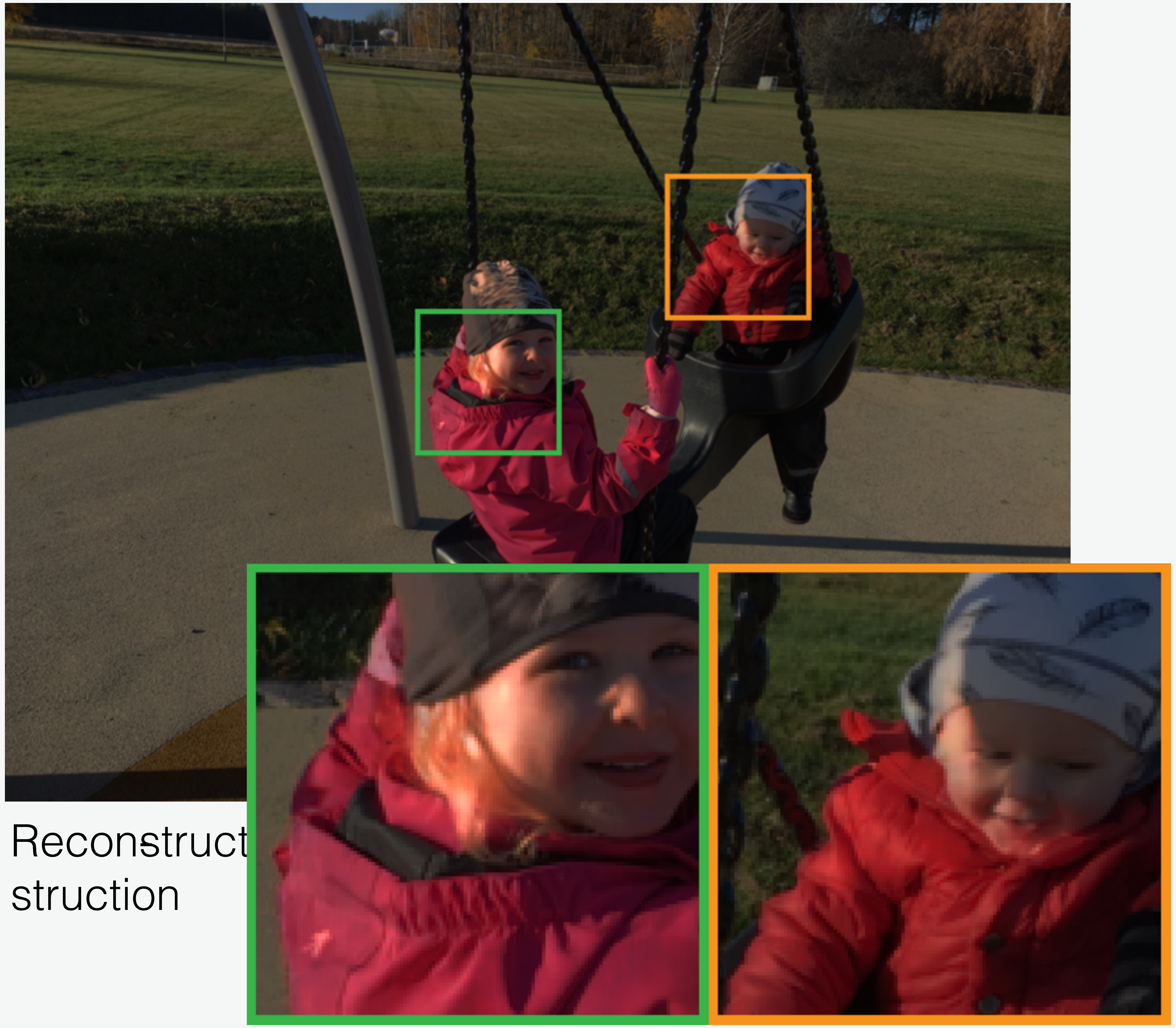


Reconstruction

Reconstruction with real-world cameras

Camera: iPhone 6S

Input



Reconstruct
struction

SIGGRAPH ASIA 2017 BANGKOK

Reconstruction with real-world cameras

Camera: Unknown

Input

Reconstruction

SIGGRAPH ASIA 2017 BANGKOK

HDR video reconstruction

- Example video in HDR video player
- Software: Luma HDRv open source HDR video codec
- Frame-by-frame predictions
- Some minor temporal issues

Library and API: http://lumahdrv.org/

SIGGRAPH ASIA 2017 BANGKOK

Comparison to iTMOs

Input | Meylan et al.[1] | Ground truth

(1) L. Meylan, S. Daly, and S. Süsstrunk. 2006. The Reproduction of Specular Highlights on High Dynamic Range Displays. Color and Imaging Conference 2006, 1 (2006), 333–338.

Comparison to iTMOs

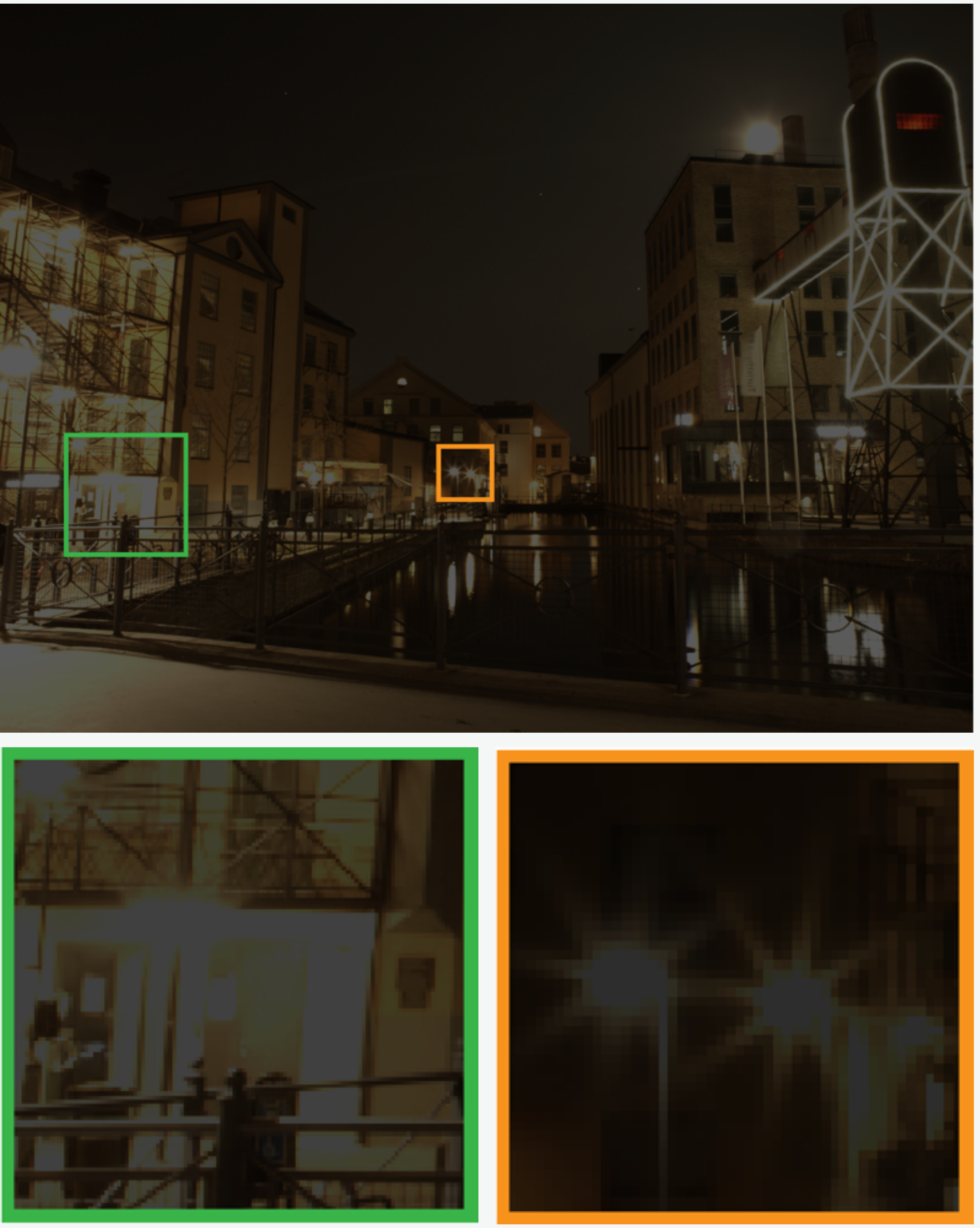
Input | Rempel et al.[1] | Ground truth

(1) A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward. 2007. Ldr2Hdr: On-the-fly Reverse Tone Mapping of Legacy Video and Photographs. ACM Trans. Graph. 26, 3, Article 39 (2007).

SIGGRAPH ASIA 2017 BANGKOK

Comparison to iTMOs

Input



Banterle et al.[1]

Ground truth

(1) F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. 2008. Expanding Low Dynamic Range Videos for High Dynamic Range Applications. In Proceedings of the 24th Spring Conference on Computer Graphics (SCCG '08). ACM, 33–41.

Comparison to iTMOs

SIGGRAPH ASIA 2017 BANGKOK

Image based lighting

Reconstruction of environment lighting panorama

Image based lighting

Rendering with reconstructed panorama

Image based lighting

Reconstruction of environment lighting panorama

Input | Reconstruction | Ground truth

SIGGRAPH ASIA 2017 BANGKOK

Image based lighting

Rendering with reconstructed panorama

SIGGRAPH ASIA 2017 BANGKOK

Subjective evaluation

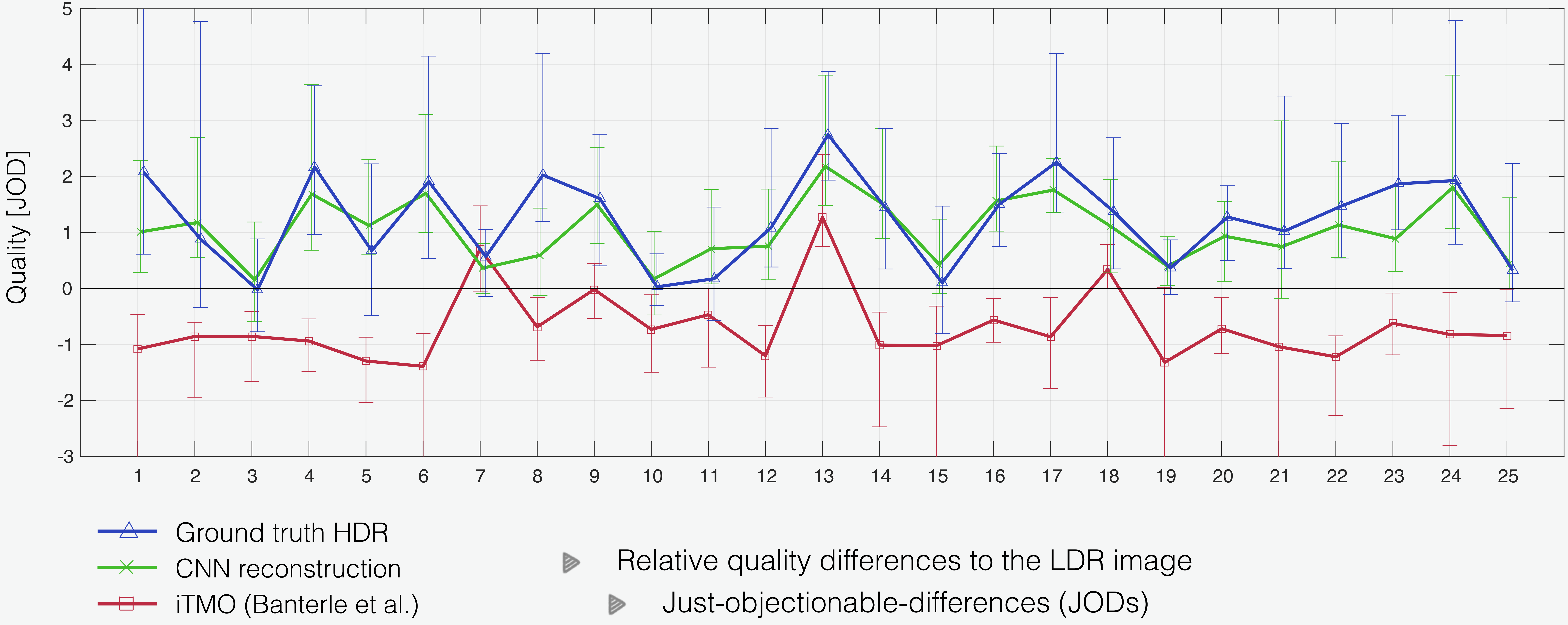
- Setup
  - Full pair-wise comparison, forced choice: "Which looks most natural?"
  - 15 participants, performed each comparison 3 times
  - 25 images, selected randomly from the test set (95 images)
  - 90th percentile pixel value anchored to 180 cd/m^2
- Methods
  - LDR image (captured by virtual camera)
  - Ground truth HDR image
  - iTMO method by Banterle et al.[A]
  - Our CNN HDR reconstruction

- Projector-based HDR display[B]
- iPad retina display
- Back-lit by projector
- Dynamic range, 0.1-5000 cd/m^2

(A) F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. 2008. Expanding Low Dynamic Range Videos for High Dynamic Range Applications. In Proceedings of the 24th Spring Conference on Computer Graphics (SCCG '08). ACM, 33–41.

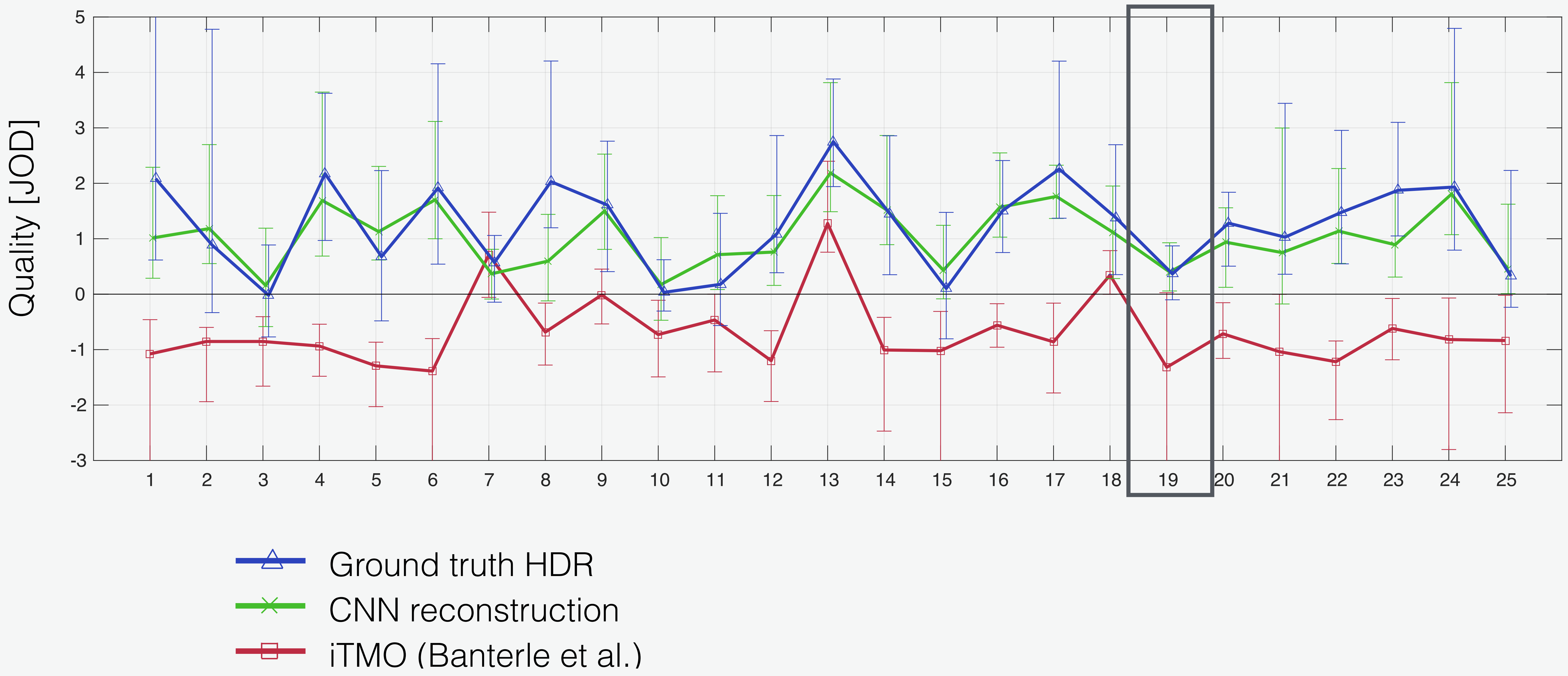
(B) H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. 2004. High Dynamic Range Display Systems. ACM Trans. Graph. 23, 3 (2004), 760–768.

Subjective evaluation



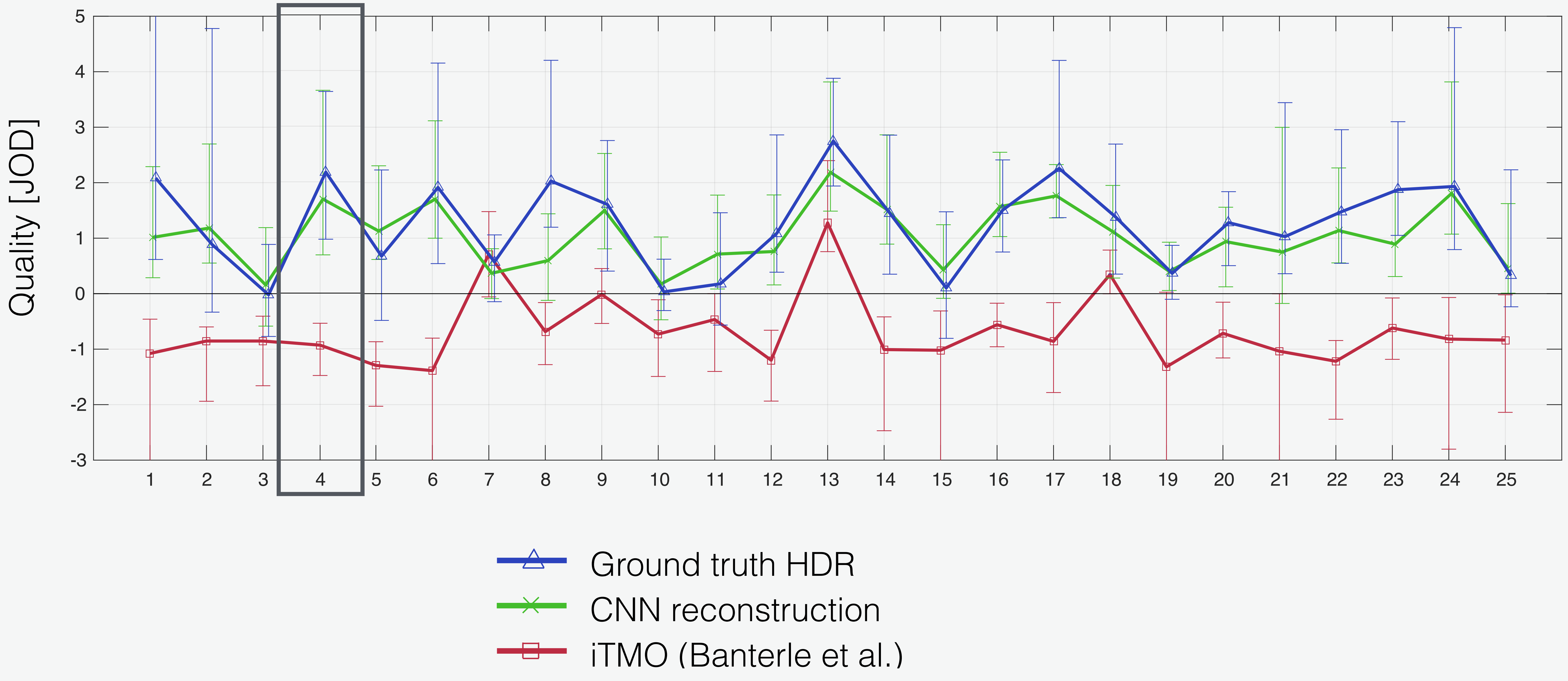
- Relative quality differences to the LDR image
  - Just-objectionable-differences (JODs)
  - 1 unit means ≈75% are expected to see an objectionable quality difference
- CNN reconstruction most often comparable to ground truth!

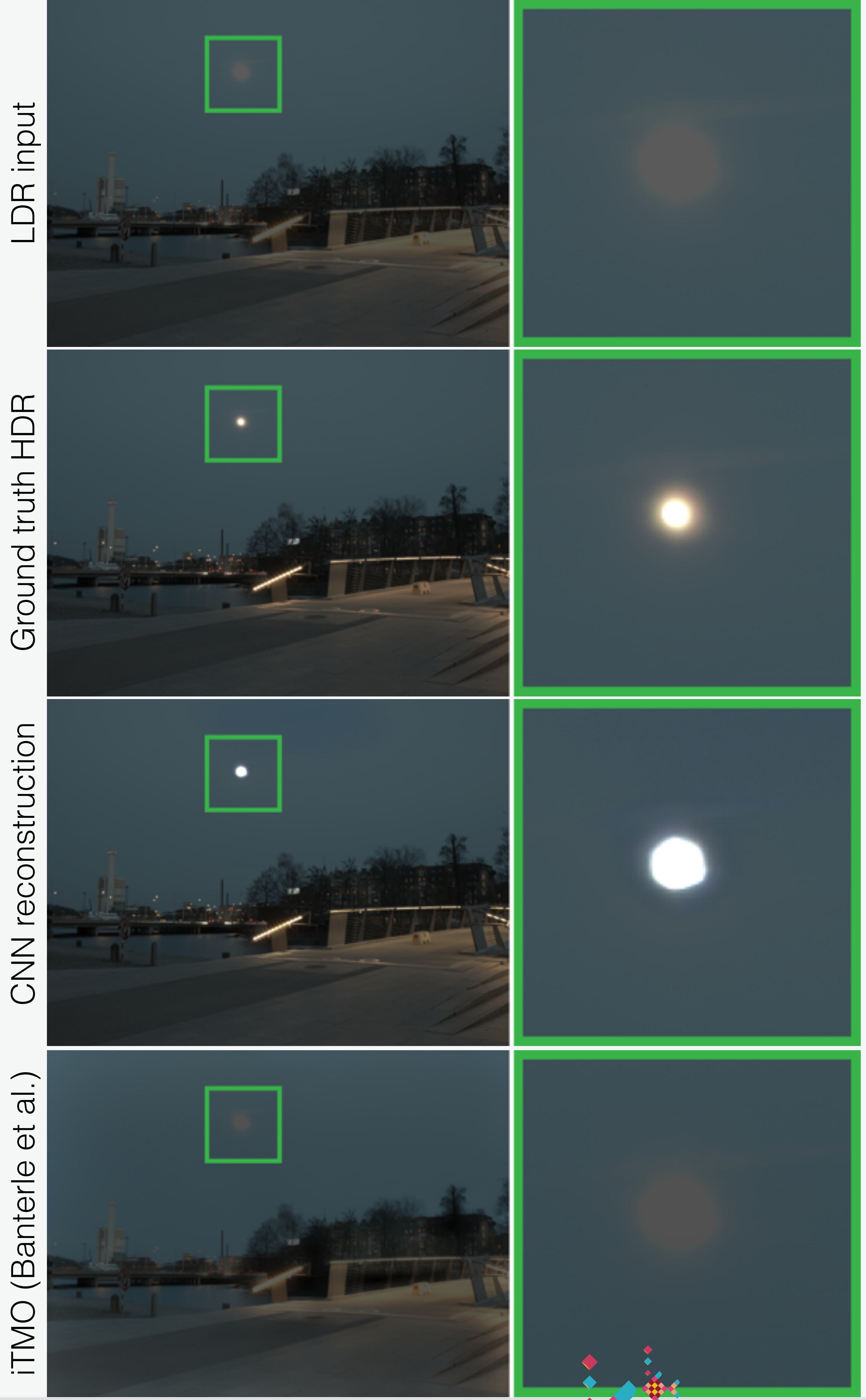
Subjective evaluation



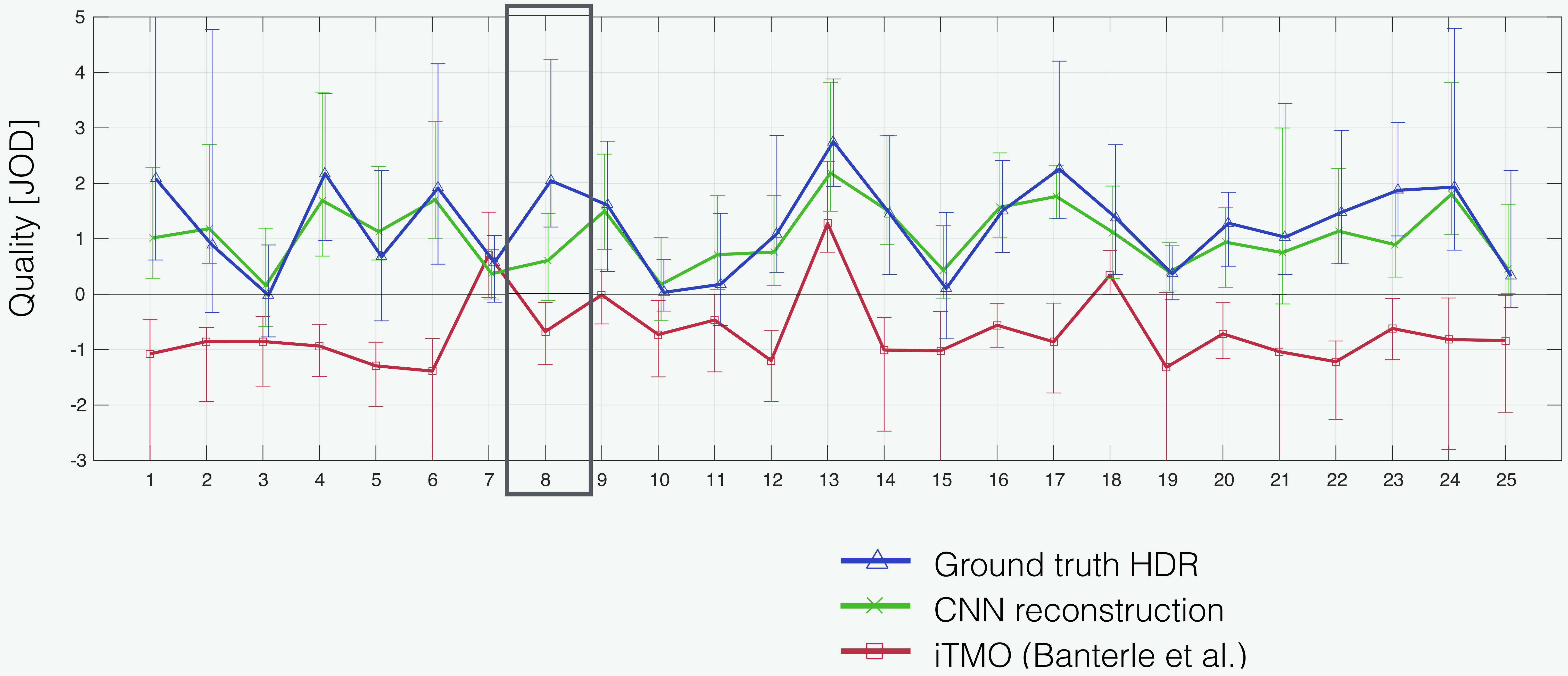
SIGGRAPH ASIA 2017 BANGKOK

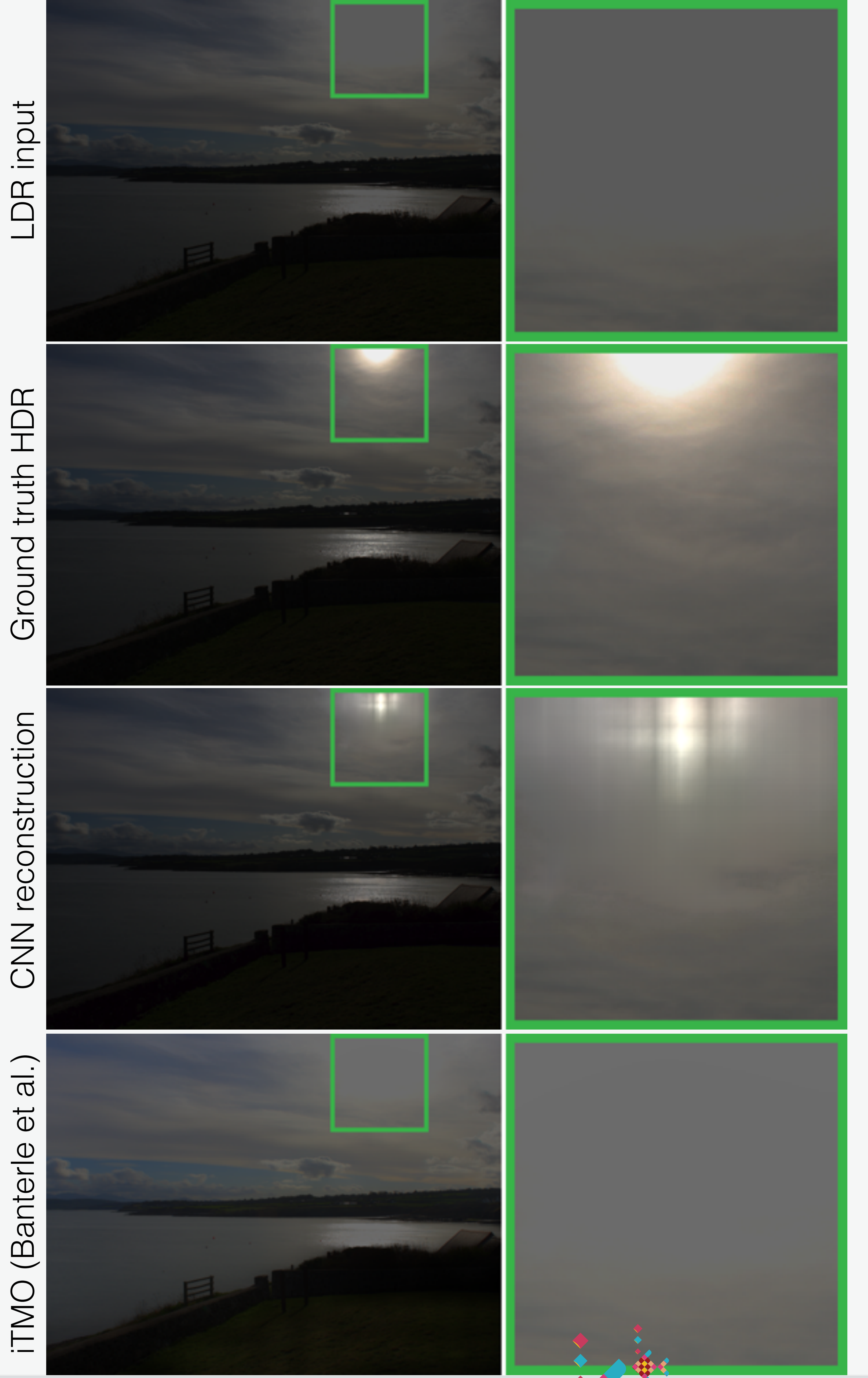
Subjective evaluation





Subjective evaluation

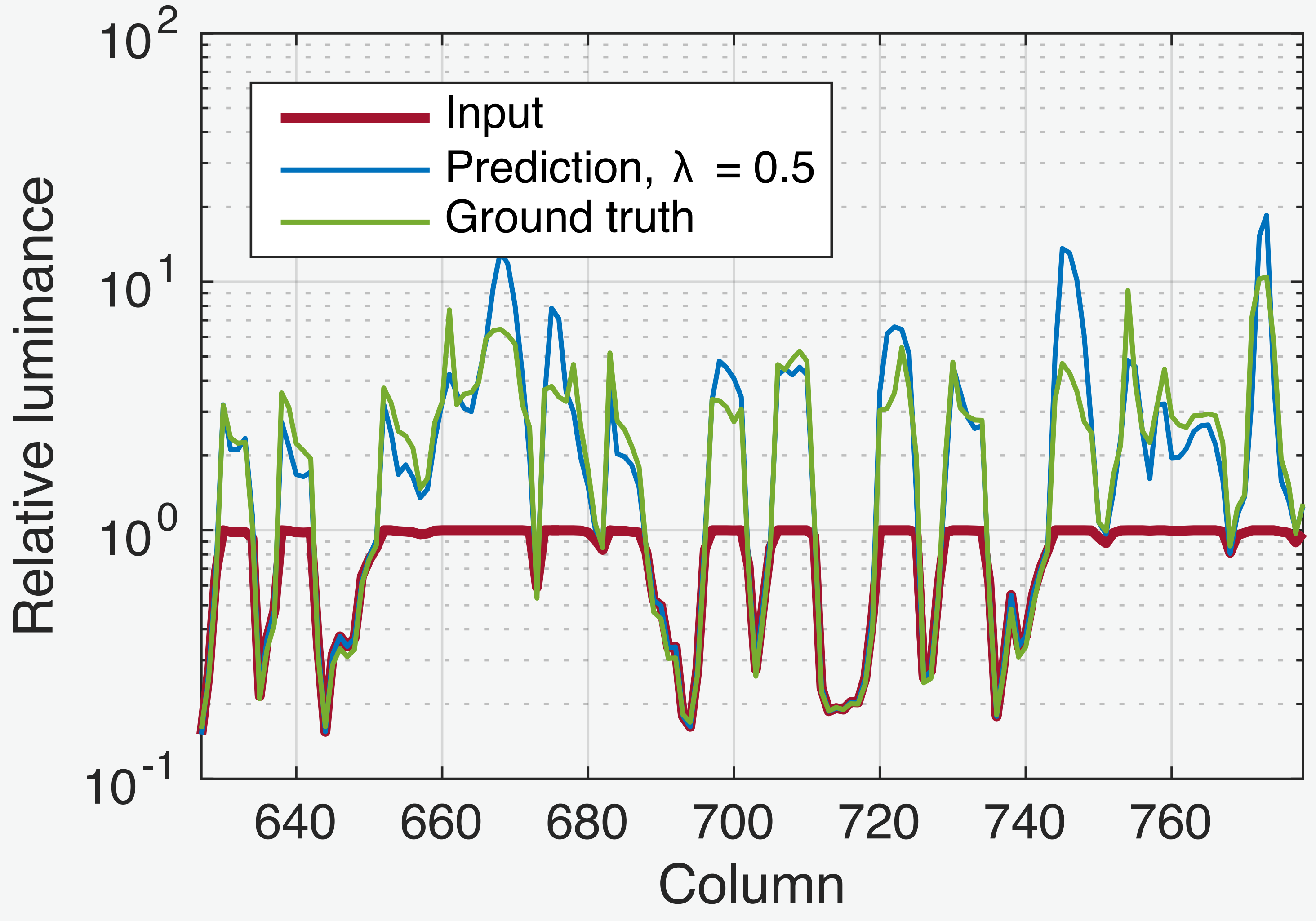


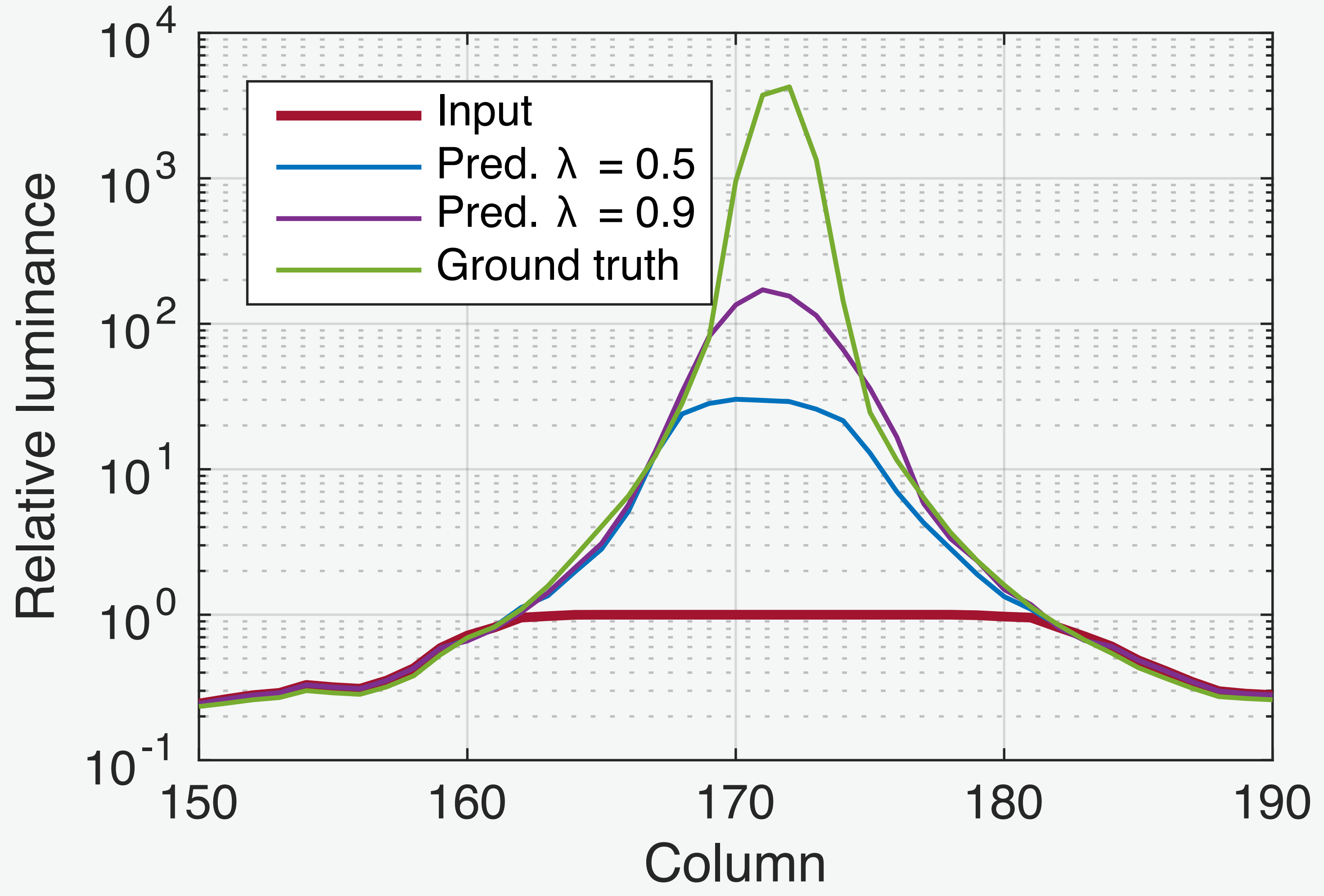


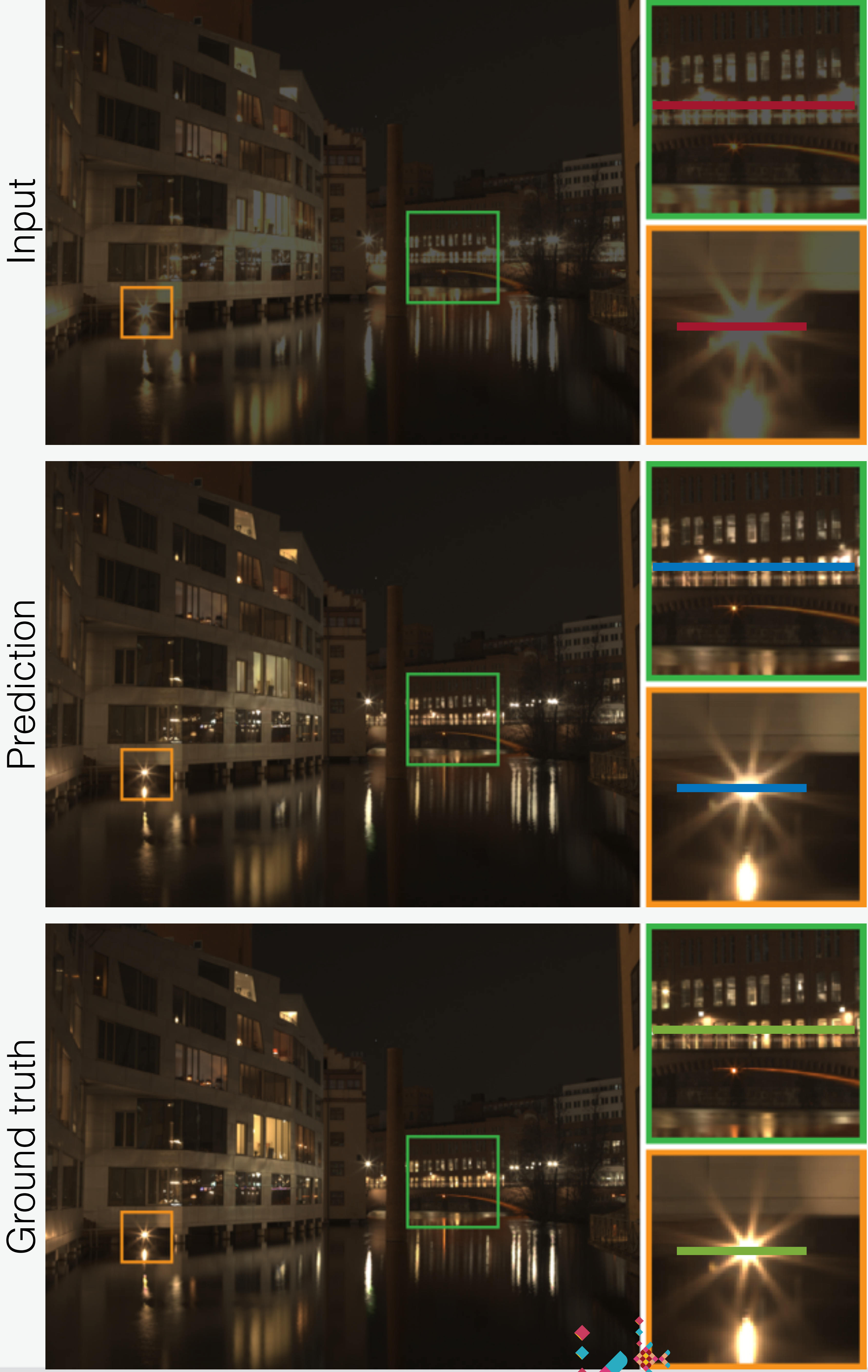
SIGGRAPH ASIA 2017 BANGKOK

Limitations

- Under-estimation of very bright light sources
- Compression artifacts (partly)
- Large saturated image regions







SIGGRAPH ASIA 2017 BANGKOK

Failure cases

Input LDR

Reconstructed HDR

SIGGRAPH ASIA 2017 BANGKOK

Failure cases

Input LDR

Reconstructed HDR

SIGGRAPH ASIA 2017 BANGKOK

Failure cases

Input LDR

Reconstructed HDR

SIGGRAPH ASIA 2017 BANGKOK

Conclusion

- HDR reconstruction from a single exposed LDR image
- Hybrid dynamic range autoencoder
  - Encoder processing LDR display values
  - Decoder reconstructing HDR in log domain
- HDR specific design choices
  - Architecture
  - Training data
  - Optimization
- Quality and robustness confirmed in subjective evaluation

Thank you!

- Supplementary document
- Web gallery with entire testset and reconstructions
- GitHub repository with source code
  - TensorFlow model for inference
  - Trained parameters
  - Updated with parameters trained with compression

http://hdrv.org/hdrcnn/