Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline

박 예 인

Vision & Display Systems Lab.

Dept. of Electronic Engineering, Sogang University

Outline

- Introduction
- Background
  - Dynamic range / LDR / HDR
  - Multi-image HDR reconstruction
  - Single-image HDR reconstruction
- Learning to Reverse the Camera Pipeline
  - LDR image formation
  - Dequantization
  - Linearization
  - Hallucination
  - Refinement
  - Experimental Results
- Conclusion

Introduction

- Recovering a HDR image from a single LDR input image
- This paper propose a method to reverse the LDR image formation pipeline.
  - [HDR] → dynamic range clipping → non-linear mapping with a CRF → quantization → [LDR]
  - [LDR] → dequantization → linearization → hallucination → [HDR]

Background

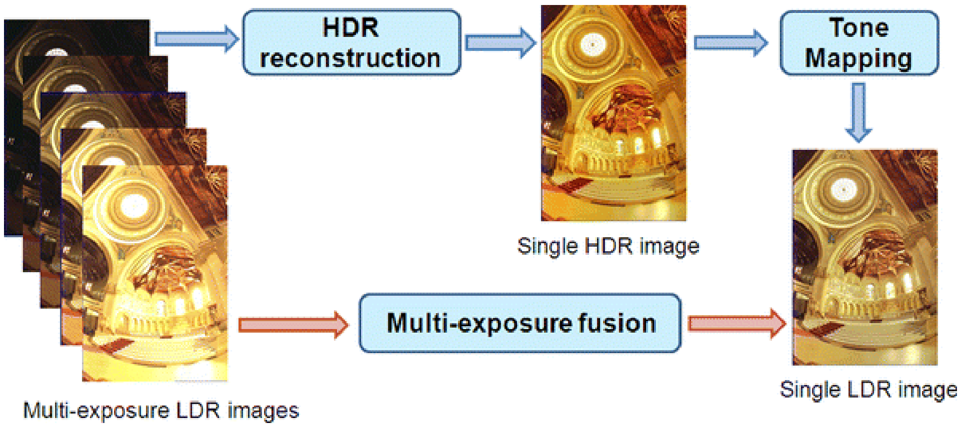
- Dynamic range
  - The ratio of the maximum and minimum values of contrast
    - The ratio of brightness between the brightest and darkest areas
- LDR (Low dynamic range) : Small dynamic range
- HDR (High dynamic range) : Large dynamic range
  - Compared to the LDR image, the details in the dark and bright areas are alive, which has the advantage of adding realism to the screen.

LDR

HDR

Background

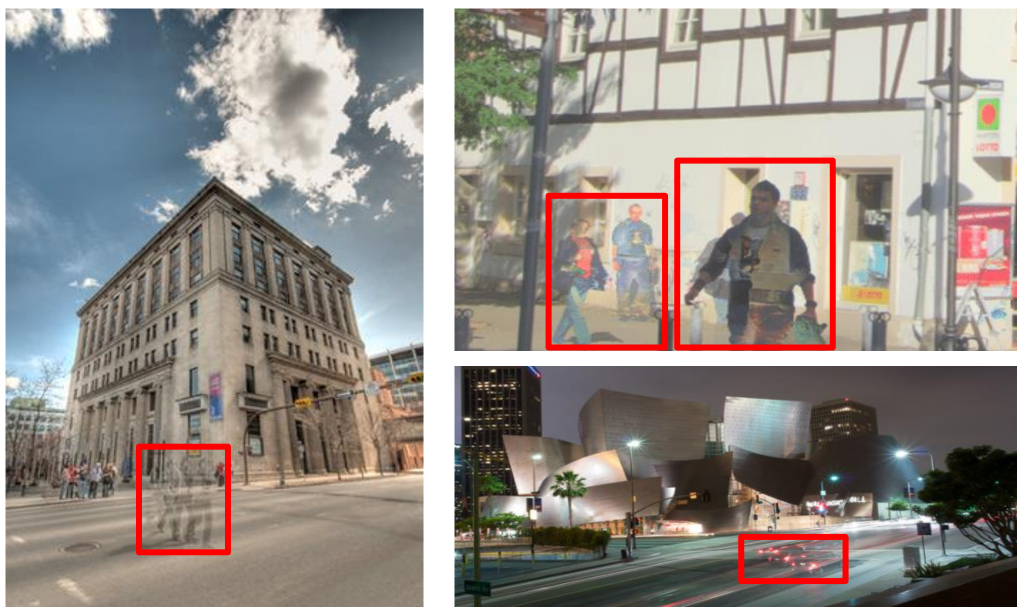
- Multi-image HDR reconstruction
  - Better performance than single-image
    - Since images with various exposure values are input, there is a lot of information given, so resilience is relatively high in terms of detail and color.
  - Ghost artifacts present
    - Ghost artifacts
      - Afterimages that occur when the camera or object moves in the process of sequentially shooting multiple images



Multi-image HDR reconstruction pipeline

Background

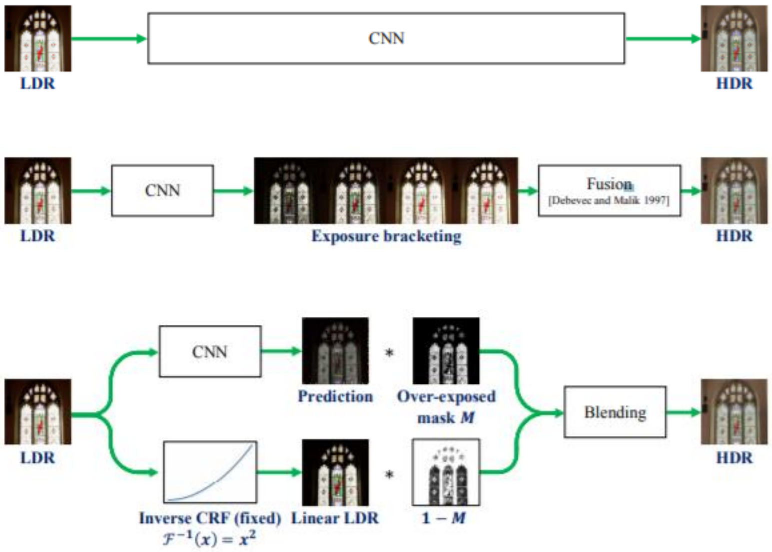
- Multi-image HDR reconstruction
  - Ghost artifacts present
    - When such an afterimage effect occurs, an afterimage problem and a discoloration problem appear on the HDR restored image, resulting in deterioration of image quality.
    - Therefore, a lot of research is being conducted to solve this problem.



Examples of ghost artifacts

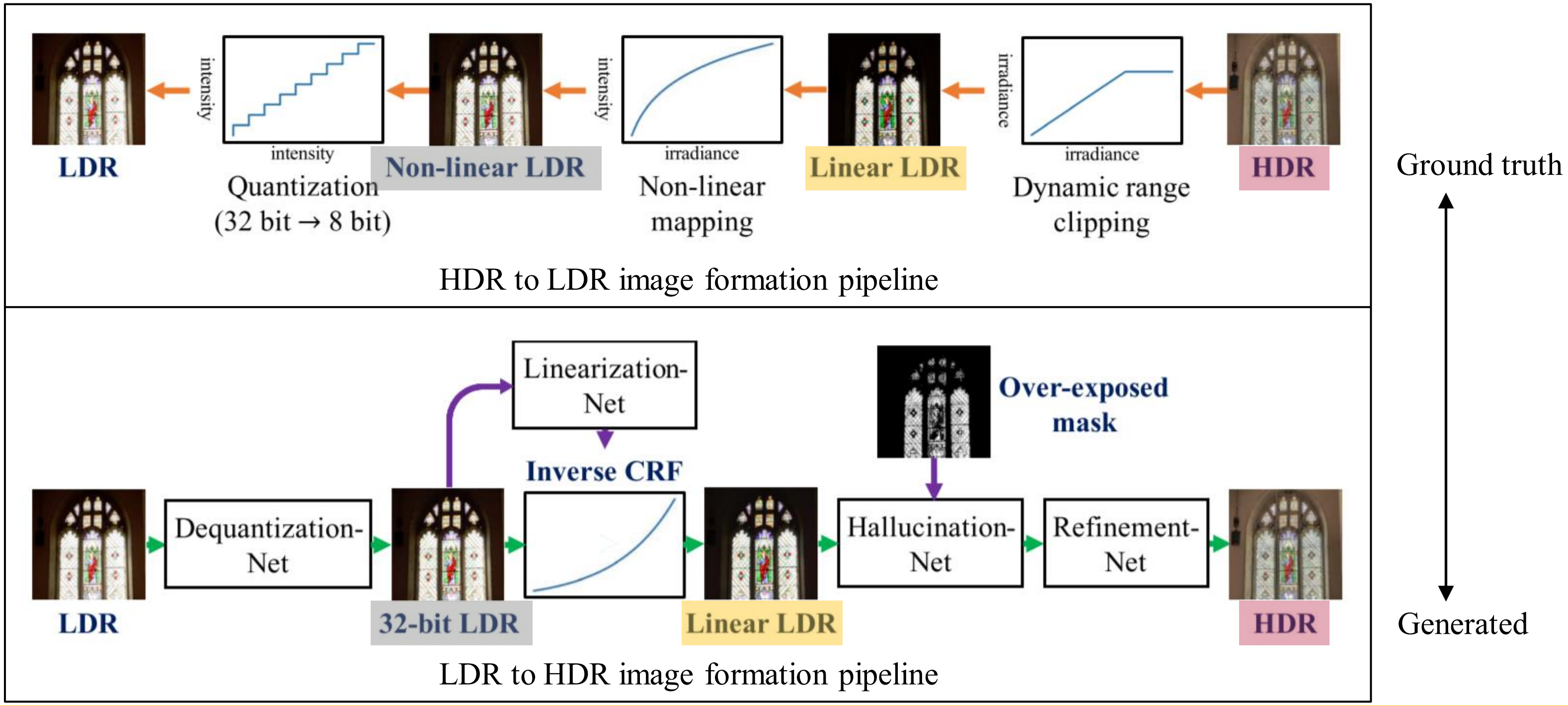
Background

- Single-image HDR reconstruction
  - It can be implemented without images with various exposure values
  - It does not suffer from ghosting artifacts
  - More challenging compared to multi-image
    - Less information is given, resulting in poor resilience in terms of detail and color



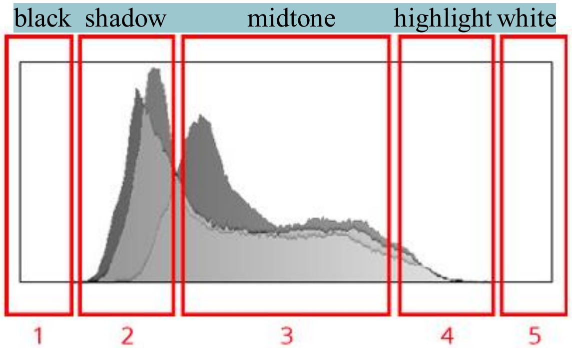
Learning to Reverse the Camera Pipeline

- HDR to LDR image formation pipeline
  - [HDR] → dynamic range clipping → non-linear mapping with a CRF → quantization → [LDR]
- Inverse function of the HDR to LDR image formation pipeline
  - [LDR] → dequantization → linearization → hallucination → [HDR]

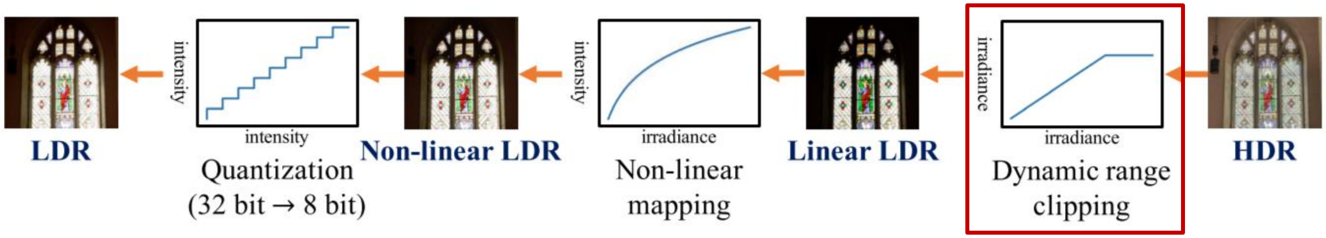


Learning to Reverse the Camera Pipeline

- HDR to LDR image formation pipeline
  - Dynamic range clipping
    - The camera first clips the pixel values of an HDR image to a limited range.
    - Due to the clipping operation, there is information loss for pixels in the over-exposed regions.



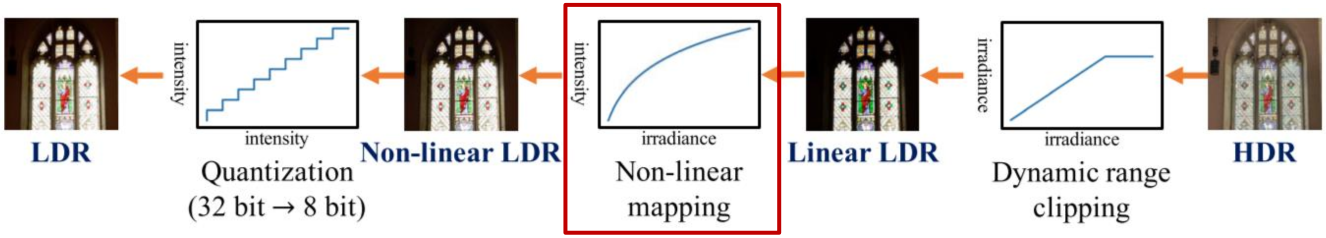
Histogram of the scene



HDR to LDR image formation pipeline

Learning to Reverse the Camera Pipeline

- HDR to LDR image formation pipeline
  - Non-linear mapping from a camera response function (CRF)
    - Non-linear mapping of linear LDR image generated by applying dynamic range clipping
    - A function that maps the irradiance of sensor to the pixel intensity of the image
      - Cameras apply nonlinear CRF mapping to adjust the contrast of the captured image.
        - ✓Calibration, tone mapping
      - A CRF is unique to the camera model.
        - ✓Non-linear characteristics are obtained through internal processing steps such as gamma correction and automatic white balance.
    - CRF estimation is a basic and necessary step in generating high dynamic range images.



HDR to LDR image formation pipeline

Learning to Reverse the Camera Pipeline

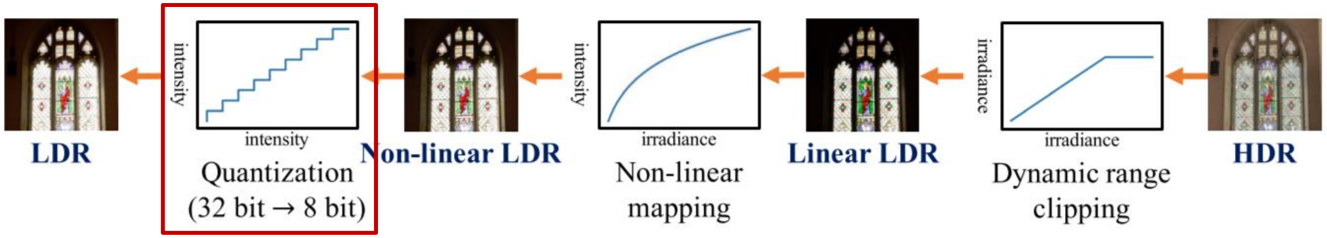
- HDR to LDR image formation pipeline
  - Quantization
    - After the non-linear mapping, the pixel values are quantized to 8 bits by

$$\mathcal{Q}(I_n) = \lfloor 255 \times I_n + 0.5 \rfloor / 255.$$

    - The quantization process leads to errors in the under-exposed regions.
    - LDR image is formed by:

$$L = \Phi(H) = \mathcal{Q}(\mathcal{F}(\mathcal{C}(H)))$$

    - Φ denotes the pipeline of dynamic range clipping, non-linear mapping, and quantization steps..



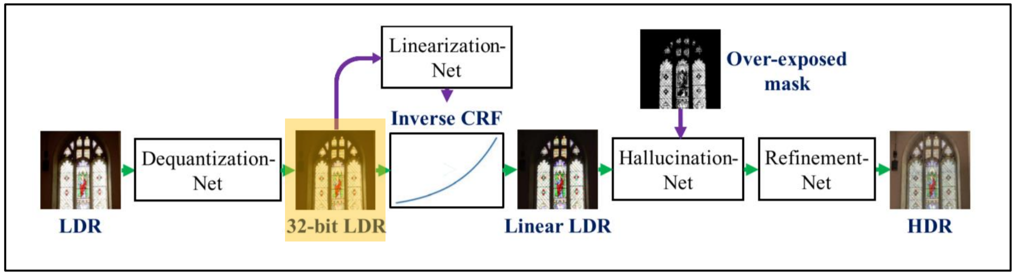
HDR to LDR image formation pipeline

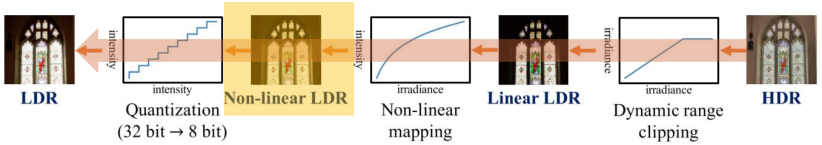
Learning to Reverse the Camera Pipeline

- Dequantization-Net
  - 6-level U-Net architecture with 2 conv layers followed by a leaky ReLU ($\alpha = 0.1$)
  - Tanh layer is used to normalized the output of the last layer to [-1.0, 1.0].
  - The output of the Dequantization-Net is added to the input LDR image.
    - Dequantized LDR image is generated.
  - L2 loss $$\mathcal{L}_{\text{deq}} = \|\hat{I}_{\text{deq}} - I_n\|_2^2$$
    - Ground truth HDR image is constructed by dynamic range clipping and non-linear mapping.

$$f(x) = \begin{cases} x & (x>0) \\ \alpha x & (x \le 0) \end{cases}$$

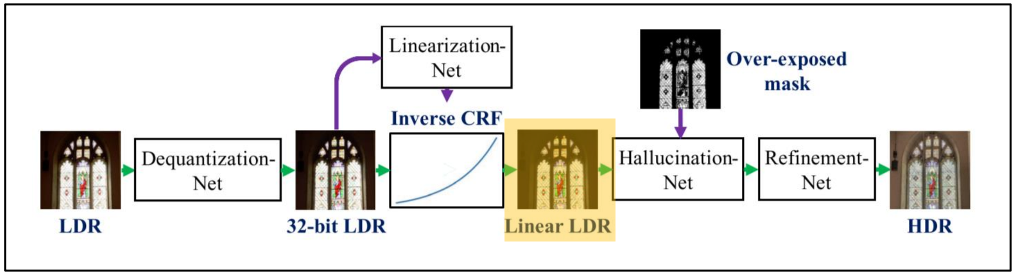
Leaky ReLU

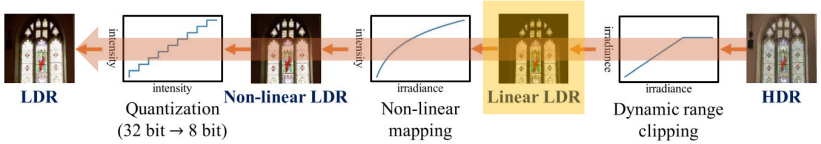




Learning to Reverse the Camera Pipeline

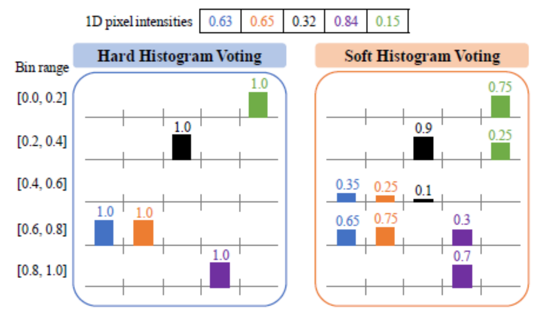
- Linearization-Net
  - The goal is to predict CRF and convert non-linear LDR images to linear irradiance.
  - CRF is unique for each camera, but all CRFs have the following properties.
    - The function should increase monotonically.
    - The minimum and maximum input values must be mapped to the minimum and maximum output values, respectively.
    - Since it is a one-to-one mapping function, the inverse function also has the above features.





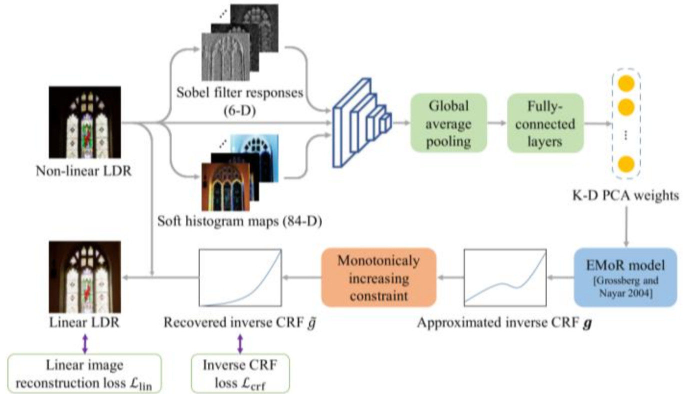
Learning to Reverse the Camera Pipeline

- Linearization-Net
  - Edge and color histogram are effective in predicting inverse CRF
    - Extract edge and color histogram features from non-linear LDR images
      - Sobel filter
      - Spatial-aware soft-histogram layer
  - ResNet-18
    - 2 FC layer → k-dim PCA weights
  - Empirical Model of Response (EMoR) model applied



$$h(i,j,c,b) = \begin{cases} 1 - d \cdot B\,, & \text{if } d < \frac{1}{B} \\ 0\,, & \text{otherwise} \end{cases}$$

i, j : horizontal, vertical pixel positions
c : the index of color channels
b $\in \{1, \cdots, B\}$: the index for the histogram bin
d = $|I(i, j, c) - (2b - 1)/(2B)|$: the intensity distance to the center of the bth bin



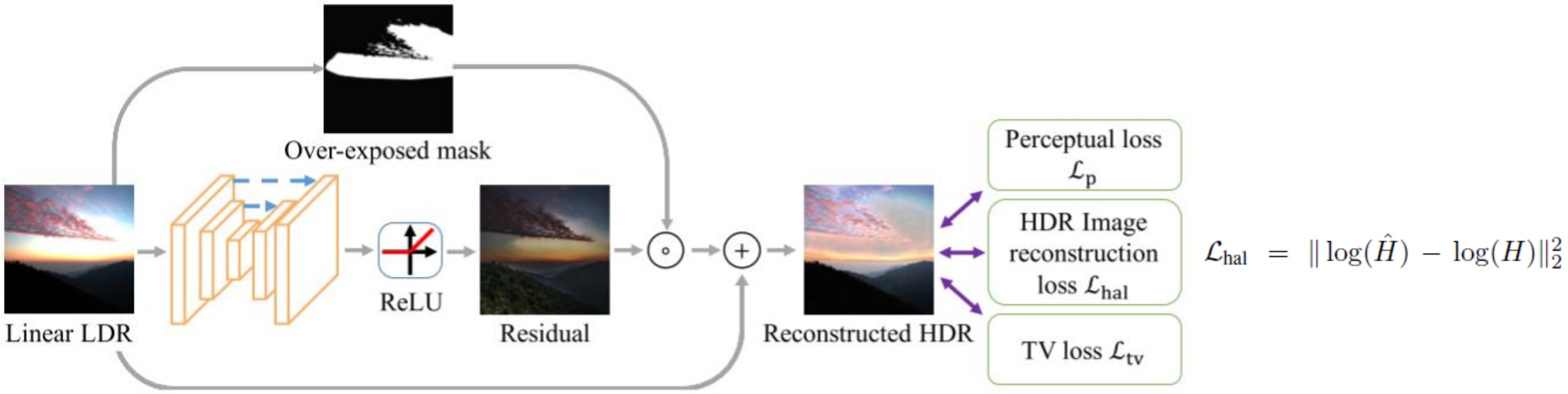
$\mathcal{L}_{\text{lin}} = \|\hat{I}_{\text{lin}} - I_c\|_2^2 \quad \mathcal{L}_{\text{crf}} = \|\tilde{g} - g\|_2^2 \longrightarrow \boxed{\mathcal{L}_{\text{lin}} + \lambda_{\text{crf}} \mathcal{L}_{\text{crf}}}$

Learning to Reverse the Camera Pipeline

- Hallucination
  - The goal is to recover missing details due to dynamic range clipping.
    - Over-exposed regions
  - Encoder-decoder architecture with skip connections
  - Reconstruction HDR image

$$\hat{H} = \hat{I}_{\text{lin}} + \alpha \cdot \mathcal{C}^{-1}(\hat{I}_{\text{lin}})$$

Over-exposed mask $\alpha = \max(0, \hat{I}_{\text{lin}} - \gamma)/(1-\gamma)$ ($\gamma = 0.95$)



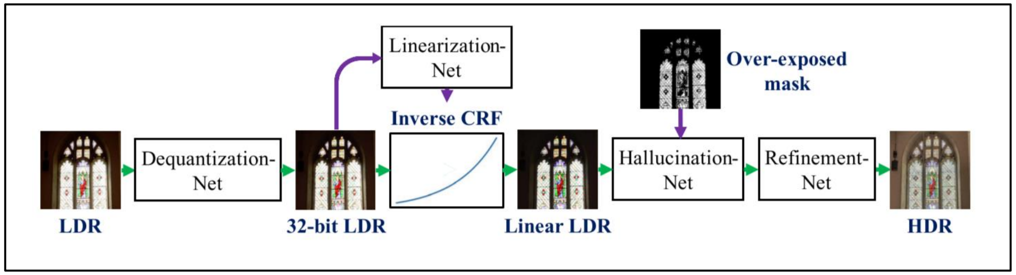
Learning to Reverse the Camera Pipeline

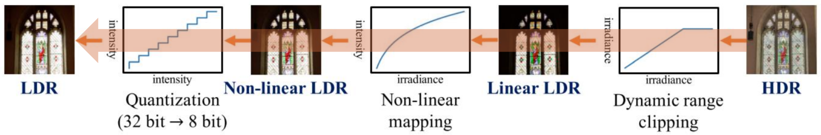
- Combination of loss functions

$$\lambda_{deq}\mathcal{L}_{deq} + \lambda_{lin}\mathcal{L}_{lin} + \lambda_{crf}\mathcal{L}_{crf} + \lambda_{hal}\mathcal{L}_{hal} + \lambda_{p}\mathcal{L}_{p} + \lambda_{tv}\mathcal{L}_{tv}$$

($\lambda_{deq} = 1$, $\lambda_{lin} = 10$, $\lambda_{crf} = 1$, $\lambda_{hal} = 1$, $\lambda_{p} = 0.001$, $\lambda_{tv} = 0.1$)

- Refinement
  - Same U-Net architecture as the Dequantization-Net
  - Refine the output of the Hallucination-Net





Learning to Reverse the Camera Pipeline

- Experimental Results
  - Quantitative comparison
    - HDR-VDP-2 score
    - Proposed method has the highest performance.

Method	Training dataset	HDR-SYNTH	HDR-REAL	RAISE [10]	HDR-EYE [42]
HDRCNN+ [14]	HDR-SYNTH + HDR-REAL	55.51 ± 6.64	51.38 ± 7.17	56.51 ± 4.33	51.08 ± 5.84
DrTMO+ [15]	HDR-SYNTH + HDR-REAL	56.41 ± 7.20	50.77 ± 7.78	57.92 ± 3.69	51.26 ± 5.94
ExpandNet [40]	Pre-trained model of [40]	53.55 ± 4.98	48.67 ± 6.46	54.62 ± 1.99	50.43 ± 5.49
Deep chain HDRI [29]	Pre-trained model of [29]	-	-	-	49.80 ± 5.97
Deep recursive HDRI [30]	Pre-trained model of [30]	-	-	-	48.85 ± 4.91
Ours*	HDR-SYNTH	**60.11 ± 6.10**	51.59 ± 7.42	58.80 ± 3.91	52.66 ± 5.64
Ours+	HDR-SYNTH + HDR-REAL	59.52 ± 6.02	**53.16 ± 7.19**	**59.21 ± 3.68**	**53.16 ± 5.92**

Learning to Reverse the Camera Pipeline

- Experimental Results
  - Quantitative comparison

Method	PSNR (↑)	SSIM (↑)
w/o dequantization	33.86 ± 6.96	0.9946 ± 0.0109
Hou et al. [18]	33.79 ± 6.72	0.9936 ± 0.0110
Liu et al. [35]	34.83 ± 6.04	0.9954 ± 0.0073
Dequantization-Net (Ours)	$\mathbf{35.87 \pm 6.11}$	$\mathbf{0.9955 \pm 0.0070}$

Comparisons on Dequantization-Net

Image	Edge	Histogram	Monotonically increasing	L2 error (↓) of inverse CRF	PSNR (↑) of linear image
✓	-	-	-	2.00 ± 3.15	33.43 ± 7.03
✓	✓	-	-	1.66 ± 2.93	34.31 ± 6.94
✓	-	✓	-	1.61 ± 3.03	34.51 ± 7.14
✓	✓	✓	-	1.58 ± 2.73	34.53 ± 6.83
✓	✓	✓	✓	$\mathbf{1.56 \pm 2.52}$	$\mathbf{34.64 \pm 6.73}$

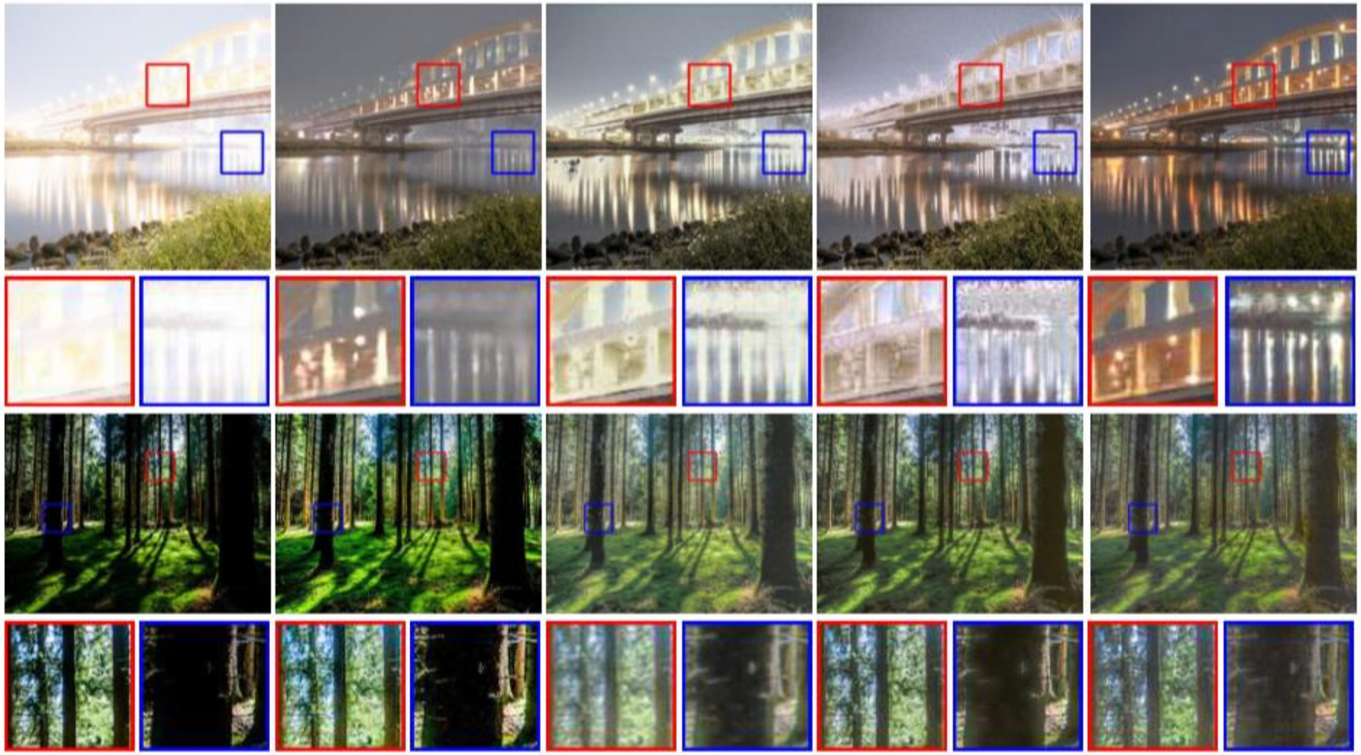
Analysis on alternatives of Linearization-Net

Positive residual	Resize convolution	Perceptual loss	HDR-VDP-2 (↑)
-	-	-	63.60 ± 15.32
✓	-	-	64.79 ± 15.89
✓	✓	-	64.52 ± 16.05
✓	✓	✓	$\mathbf{66.31 \pm 15.82}$

Analysis on alternatives of Hallucination-Net

Learning to Reverse the Camera Pipeline

- Experimental Results
  - Visual comparison



(a) Input LDR (b) HDRCNN+ (c) DrTMO+ (d) ExpandNet (e) Ours+

Sogang University

Conclusion

- Recovering a HDR image from a single LDR input image

- This paper propose a method to reverse the LDR image formation pipeline.
  - [HDR] → dynamic range clipping → non-linear mapping with a CRF → quantization → [LDR]
  - [LDR] → dequantization → linearization → hallucination → [HDR]

- Experimental results validate the effectiveness of proposed method to restore visually pleasing details for a wide variety of challenging scenes.