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

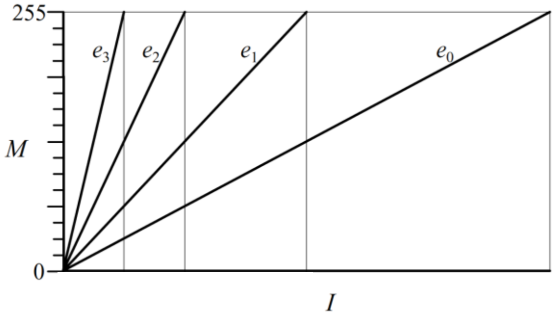
Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, Jia-Bia Huang

CVPR '20

2022.05.09 윤주열

High Dynamic Range Image

Brightness of a scene is almost infinite while we can only measure a finite range.



(a) Exposure: T e3

(b) Exposure: 4T e2

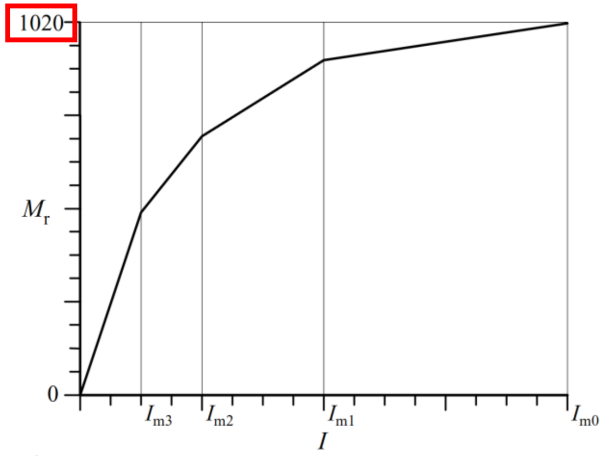
(c) Exposure: 16T e1

(d) Exposure: 64T e0

< Fixed scene with difference exposures >

Simple Summation

Try to capture sufficient amount of light from all brightness.



Tone Mapping

COMPUTED IMAGE

High Dynamic Range Image

Single Image High Dynamic Range Imaging (Inverse Tone Mapping)

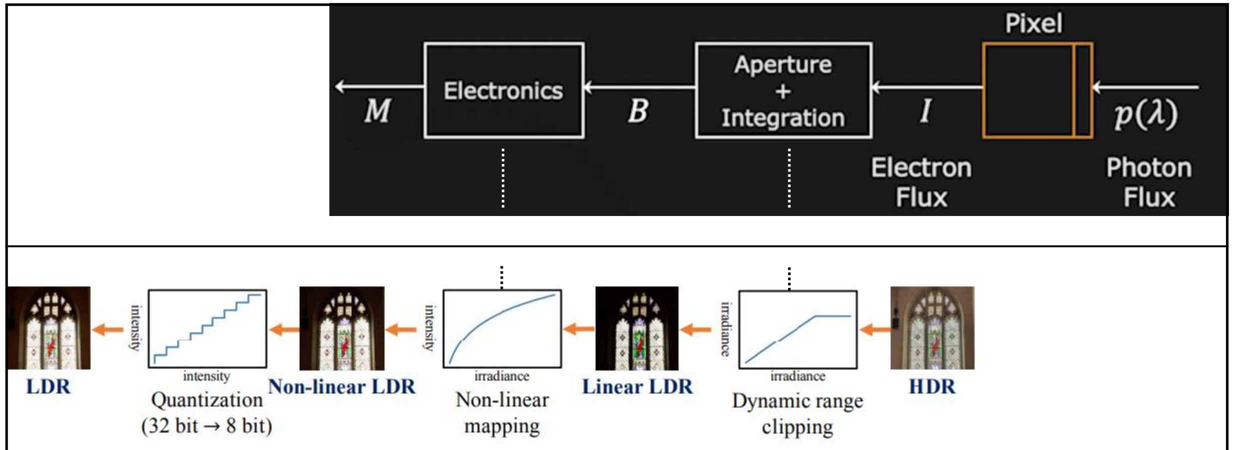
Can be viewed as a channel-wise outpainting.

Camera Pipeline

How exactly are we obtaining LDR images?

I : Scene brightness

M : Measured brightness



(a) LDR Image formation pipeline

Scene brightness is first clipped according to the exposure. (Long exposure → more clipping)

The camera response function (CRF) maps the brightness with a non-linear curve.

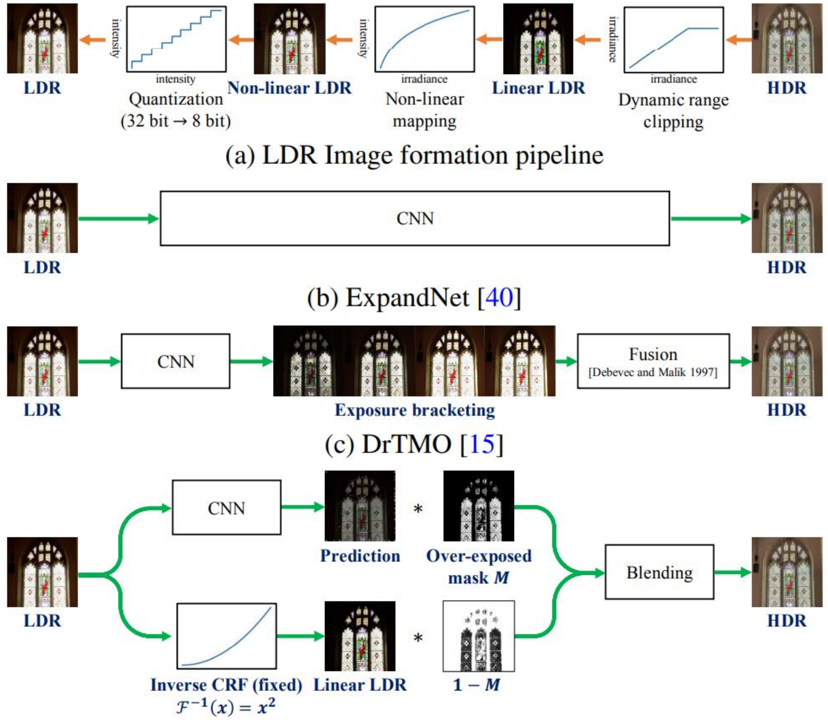
Brightness is quantized 8 bit value.

Method

Idea: Reverse the camera pipeline.

Given a HDR image, create images created within the camera pipeline.

Train each component separately and later fine-tune it.

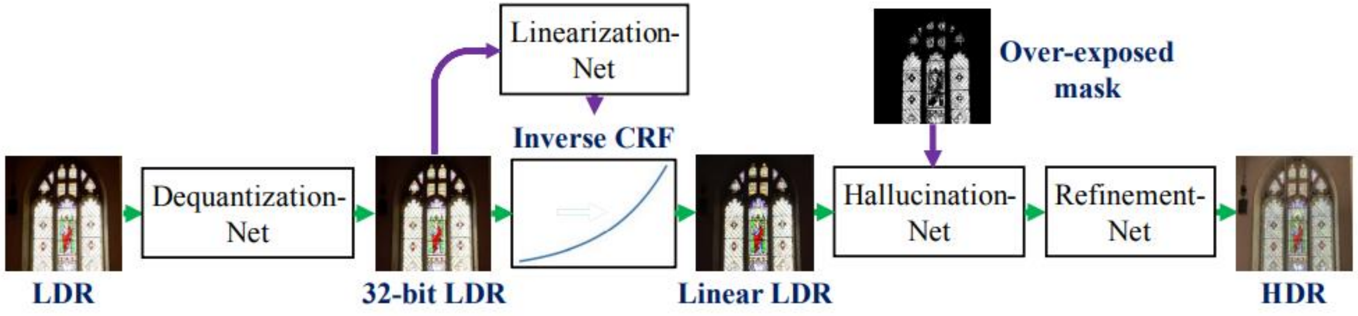


(a) LDR Image formation pipeline

(b) ExpandNet [40]

(c) DrTMO [15]

(d) HDRCNN [14]

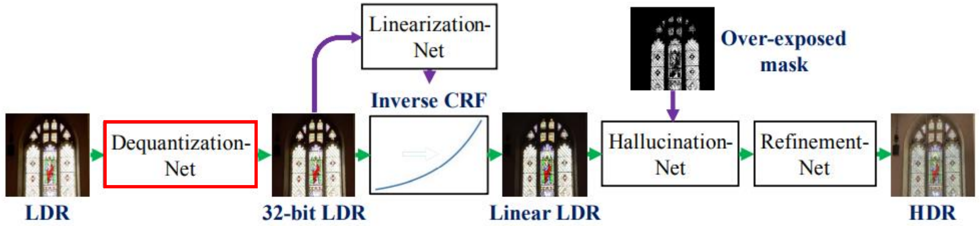


(e) Proposed method

Dequantization

Our input LDR image is an 8-bit image.

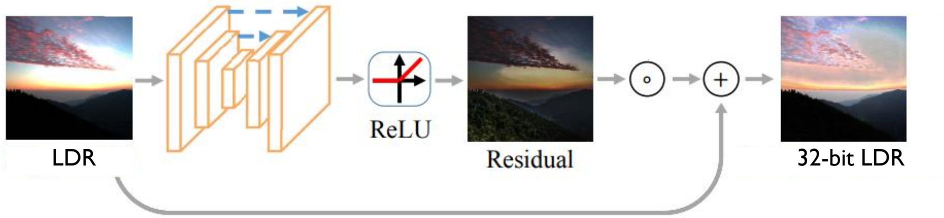
Quantized images contain noise and banding artifacts.

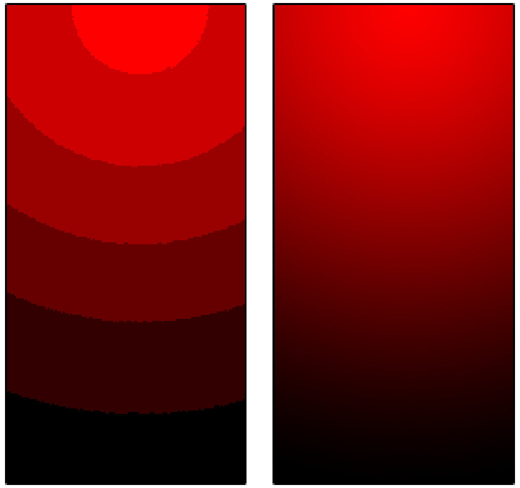


(e) Proposed method

Use a simple Unet to reduce these artifacts.

The output of the Unet is added to the original LDR image.

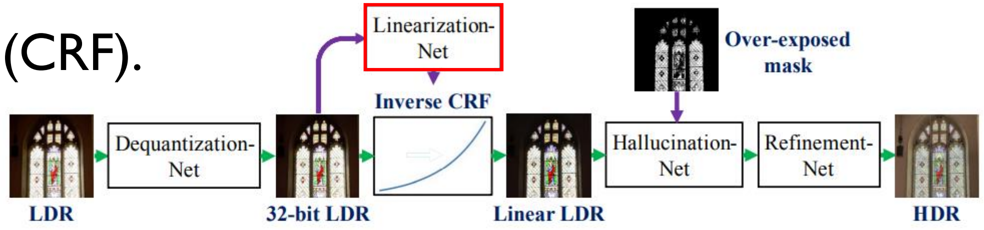




8-bit gradient 24-bit gradient

Linearization

A linear image is mapped with a camera response function (CRF).



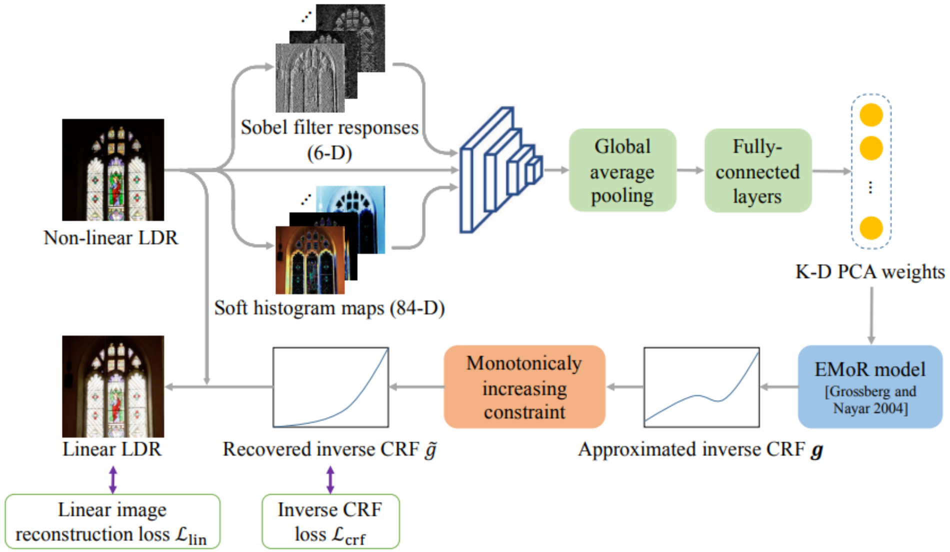
(e) Proposed method

We need to inverse the CRF to obtain a linear image from a LDR image.

Represent a CRF with 1024-dimensional vector. (since CRF is a 1D function with a bounded domain)

Most existing CRF can be approximated by a weighted sum of K PCA components from the EMoR (Empirical Model of Response). → Predict the PCA weights.

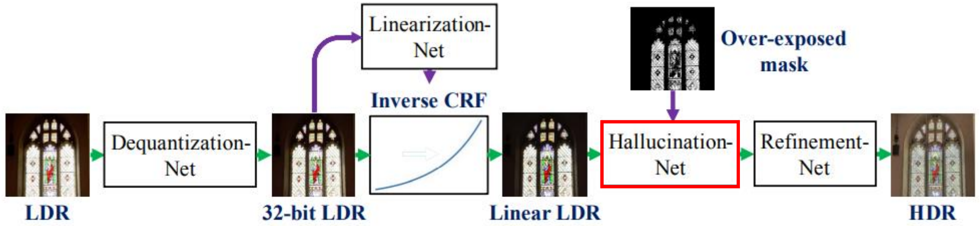
Force the monotonically increasing constraint on the predicted CRF.



Hallucination

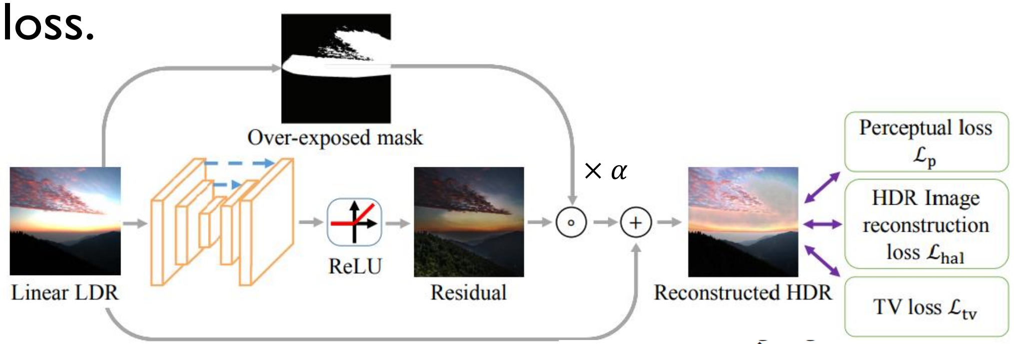
Some of the brightness of the original scene is clipped.

Predict the missing details of the overexposed region.



(e) Proposed method

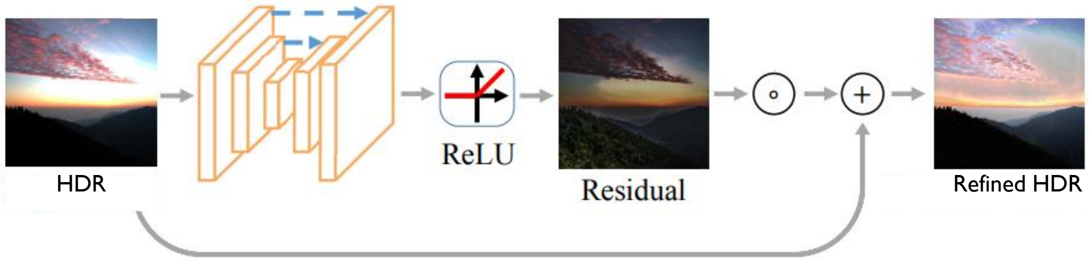
Details are recovered using the perceptual loss and the L2 loss.



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

Refinement step

pass an additional Unet same as the dequantization step to predict the final HDR image.



Results

PSNR of dequantization and linearization.

Table 2: **Comparisons on Dequantization-Net.** Our Dequantization-Net restores the missing details due to quantization and outperforms existing methods.

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)	$\mathbf{35.87 \pm 6.11}$	$\mathbf{0.9955 \pm 0.0070}$

Table 3: **Analysis on alternatives of Linearization-Net.**

Method	L2 error (↓) of inverse CRF	PSNR (↑) of linear image
Pre-defined x^2 [6]	11.64 ± 12.47	24.81 ± 6.47
Pre-defined $x^{2.2}$	9.06 ± 10.74	25.82 ± 6.04
Average inverse CRF	7.36 ± 7.03	25.24 ± 4.82
CRF-Net [19]	4.23 ± 4.37	30.61 ± 6.82
CRF-Net* [19]	2.71 ± 4.10	32.84 ± 6.85
Linearization-Net (ours)	$\mathbf{1.56 \pm 2.52}$	$\mathbf{34.64 \pm 6.73}$

Final HDR quantitative analysis.

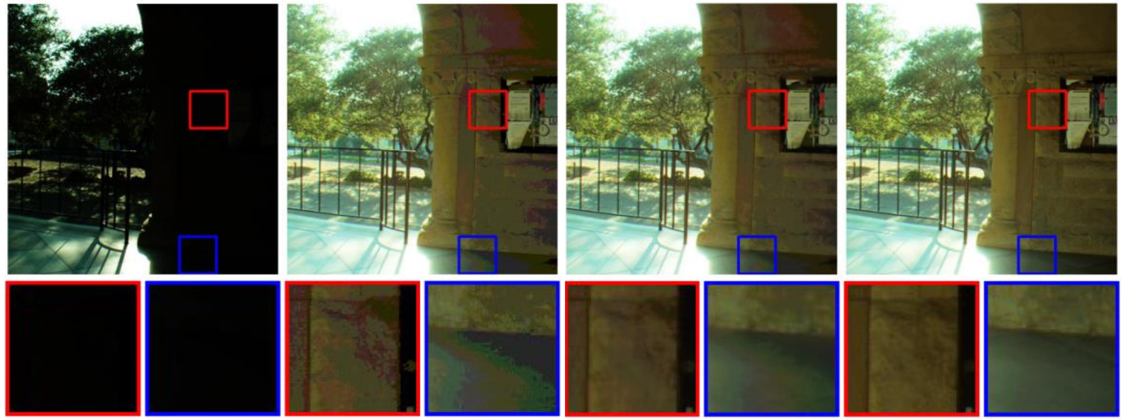
Table 1: **Quantitative comparison on HDR images with existing methods.** * represents that the model is re-trained on our synthetic training data and + is fine-tuned on both synthetic and real training data. **Red** text indicates the best and blue text indicates the best performing state-of-the-art method.

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	$\mathbf{60.11 \pm 6.10}$	51.59 ± 7.42	58.80 ± 3.91	52.66 ± 5.64
Ours+	HDR-SYNTH + HDR-REAL	59.52 ± 6.02	$\mathbf{53.16 \pm 7.19}$	$\mathbf{59.21 \pm 3.68}$	$\mathbf{53.16 \pm 5.92}$

Results

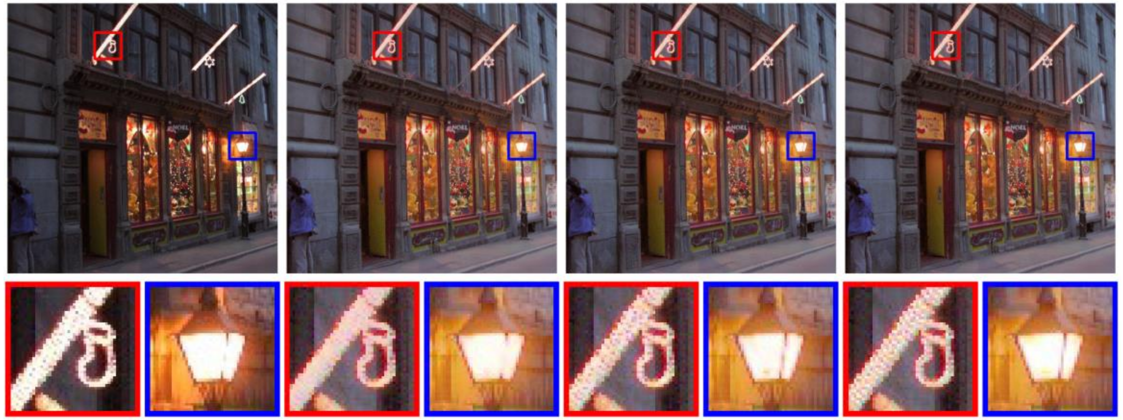
Qualitative results

Dequantization



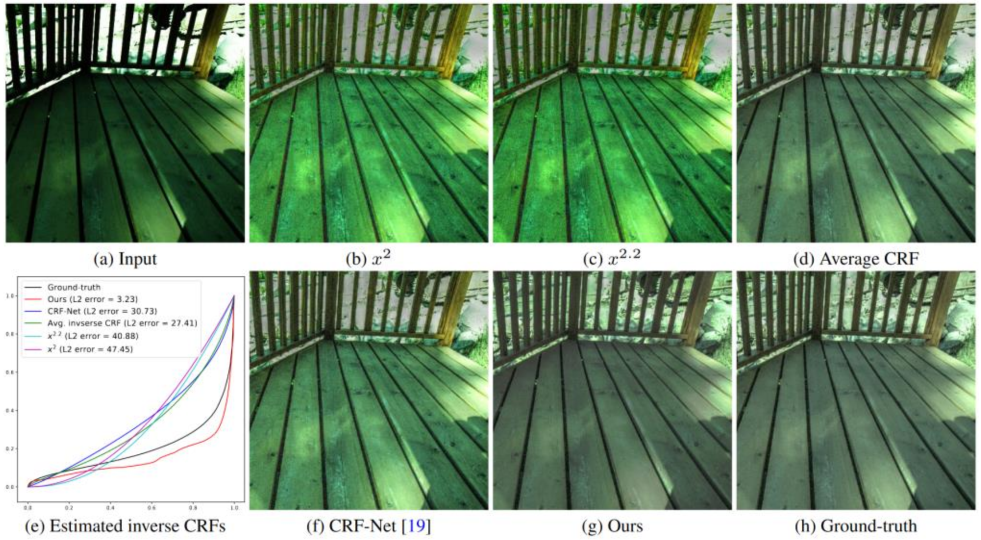
Input | Tone-mapped input | Ours | Ground-truth

Hallucination



Input | w/o perceptual loss | w/ perceptual loss | Ground-truth

Linearization



Results

Failure case

(a) Input LDR (b) Ours (c) Ground truth HDR

Figure 13: **Failure case.** The scene outside the window is severely over-exposed. Existing methods and our model cannot reconstruct plausible content.