Single Image HDR Reconstruction

박우준

Vision & Display Systems Lab.

Sogang University

Outline

- What to Expect From This Seminar
- HDR in 2020
- HDR Problem Formulation
- Camera Response Function Estimation
  - CRF-net: Single Image Radiometric Calibration using CNNs
  - Linearization-Net
  - Analyzing Modern Camera Response Functions
- Saturated Region Restoration
  - Single Image HDR Reconstruction Using a CNN w/ Masked Features and Perceptual Loss

What to Expect From This Seminar

- Broad *(but shallow)* understanding of the HDR problem

- Major CV conference topics in HDR
  - Special camera
    - Lens, sensor
  - Common camera
    - Single image HDR reconstruction

- Two major problems in single image HDR reconstruction using deep learning
  - Camera response function estimation
  - Saturated region restoration

HDR in 2020

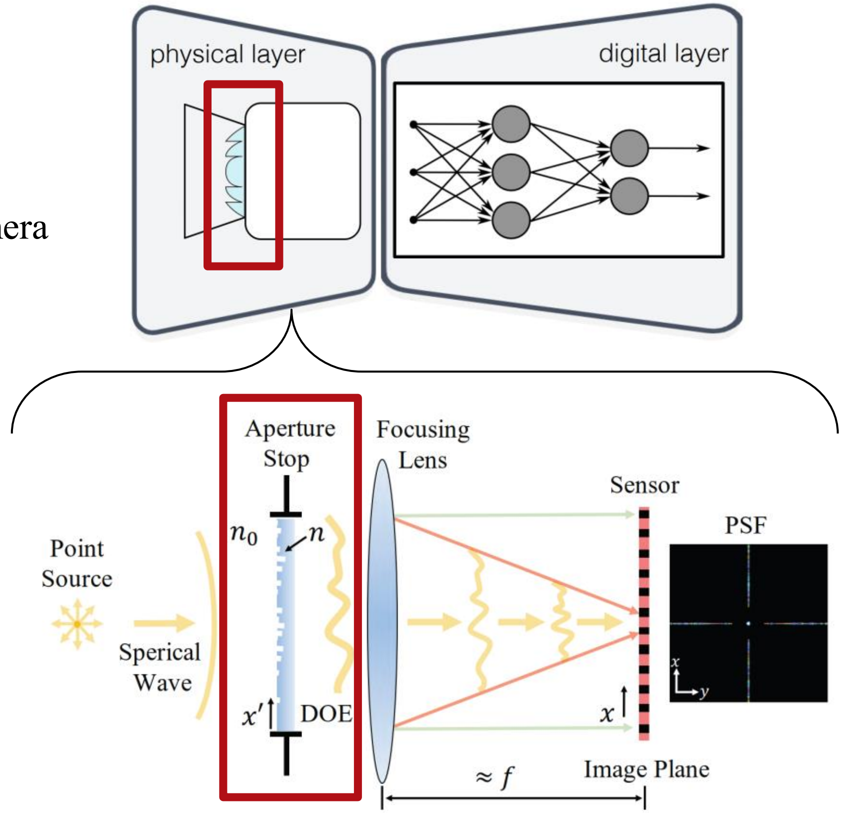
HDR in 2020

: Overview

- Deep optics (lens)
  - Learning Rank-1 Diffractive Optics for Single-Shot High Dynamic Range Imaging [CVPR 2020]
  - Deep optics for single-shot high-dynamic-range imaging [CVPR 2020]
- Special camera (sensor)
  - UnModNet: Learning to Unwrap a Modulo Image for High Dynamic Range Imaging [NIPS 2020]
  - Neuromorphic Camera Guided High Dynamic Range Imaging [CVPR 2020]
- Single image HDR
  - Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline [CVPR 2020]
  - Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss [SIGGRAPH, TOG 2020]
  - End-to-End Differentiable Learning to HDR Image Synthesis for Multi-exposure Images [AAAI 2021]

HDR in 2020
: (1) Deep optics (lens)

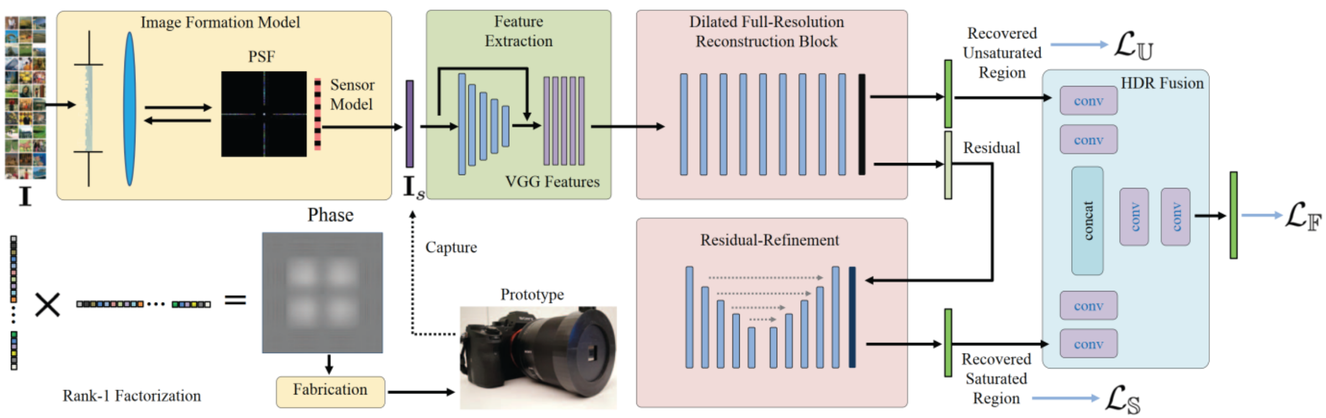
- Diffractive Optics Element (DOE)
  - Pipeline
    - Optical encoder → electronical decoder
  - Optical encoder
    - Additional special lens in front of normal camera
  - Electronical decoder
    - Tailerd neural network
  - Case study
    - Image Classification in Low Light
    - Monocular Depth Estimation
    - Neural Sensors
    - …
    - **HDR Imaging**



HDR in 2020

: (1) Deep optics (lens)

- Learning Rank-1 Diffractive Optics for Single-Shot High Dynamic Range Imaging [CVPR 2020]
  - Learning an optical HDR encoding in a single image
    - **Optical encoder** : DOE maps **saturated highlights** into neighboring unsaturated areas
    - Electronical decoder : reconstruction network tailored to images from a DOE
  - Propose a novel rank-1 parameterization of the DOE
    - Drastically reduces the optical search space
    - Efficiently encode high-frequency detail



HDR in 2020

: (2) Special HW (sensor)

- UnModNet: Learning to Unwrap a Modulo Image for High Dynamic Range Imaging [NIPS 2020]
  - Reconstruction network tailored to images from a modulo camera

Normal camera output Modulo camera output Reconstruction output

HDR in 2020

: (3) Single image HDR

- Single-Image HDR Reconstruction by Learning to **Reverse the Camera Pipeline** [CVPR 2020]
- **End-to-End Differentiable** Learning to HDR Image Synthesis for Multi-exposure Images [AAAI 2021]
- Single Image HDR Reconstruction Using a CNN with **Masked Features** and Perceptual Loss [SIGGRAPH, TOG 2020]

HDR in 2020

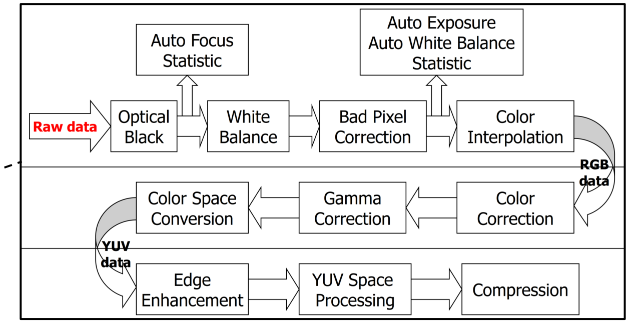
: (3) Single image HDR

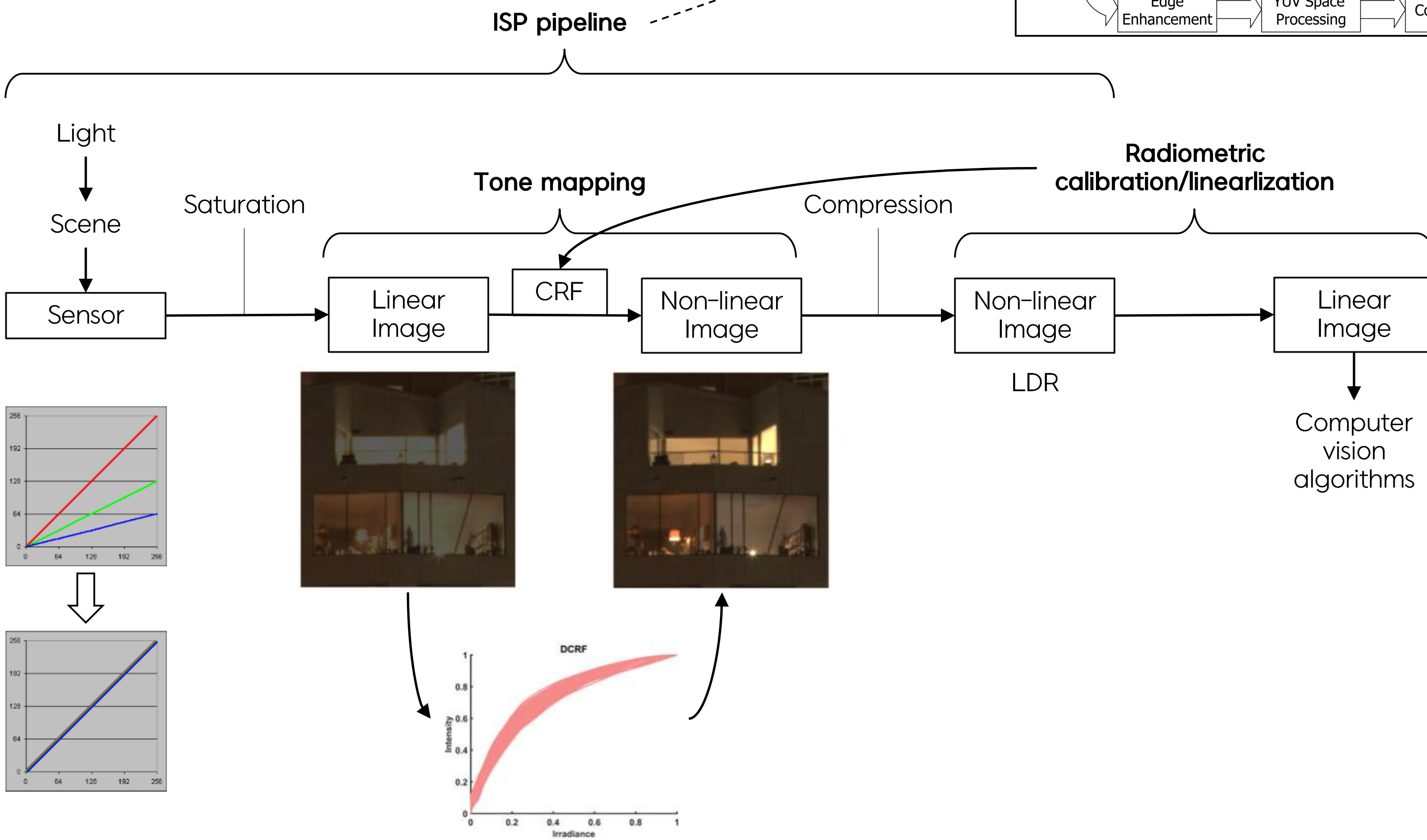
- ~~Single-Image HDR Reconstruction by Learning to **Reverse the Camera Pipeline** [CVPR 2020]~~
- ~~**End-to-End Differentiable** Learning to HDR Image Synthesis for Multi-exposure Images [AAAI 2021]~~
- Single Image HDR Reconstruction Using a CNN with **Masked Features** and Perceptual Loss [SIGGRAPH, TOG 2020]

HDR Problem Formulation

HDR Problem Formulation

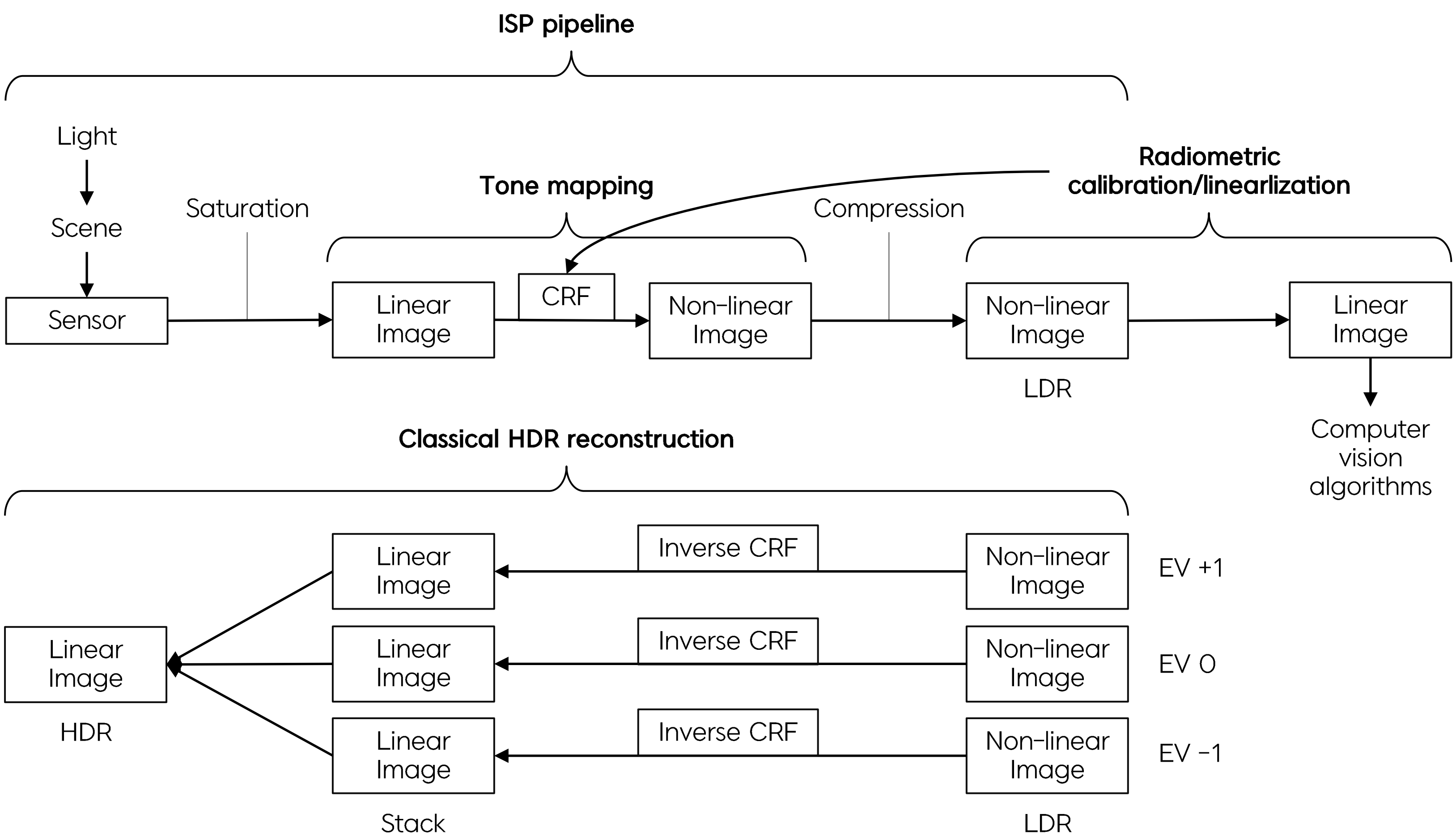
: ISP pipeline





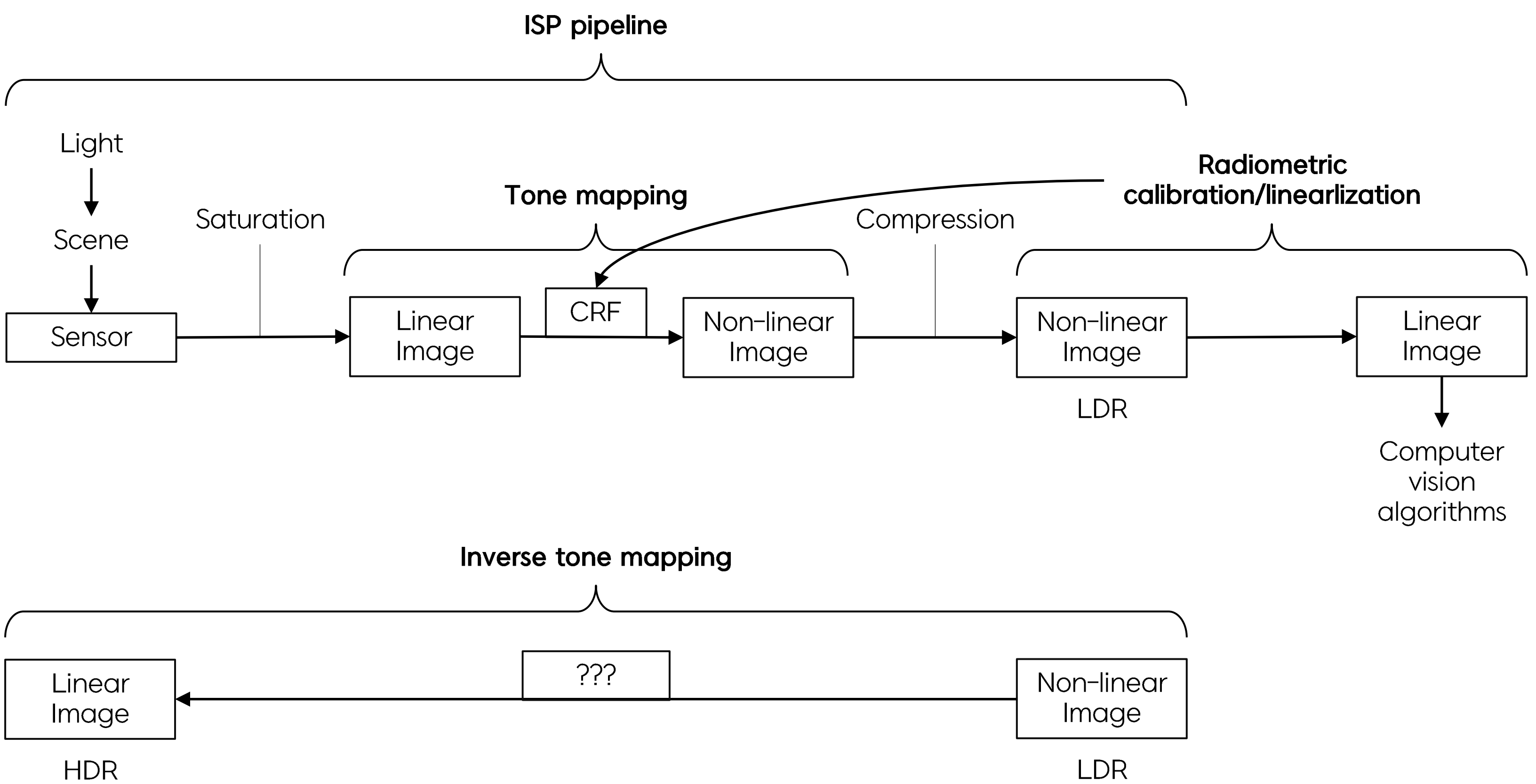
HDR Problem Formulation

: Classical HDR reconstruction



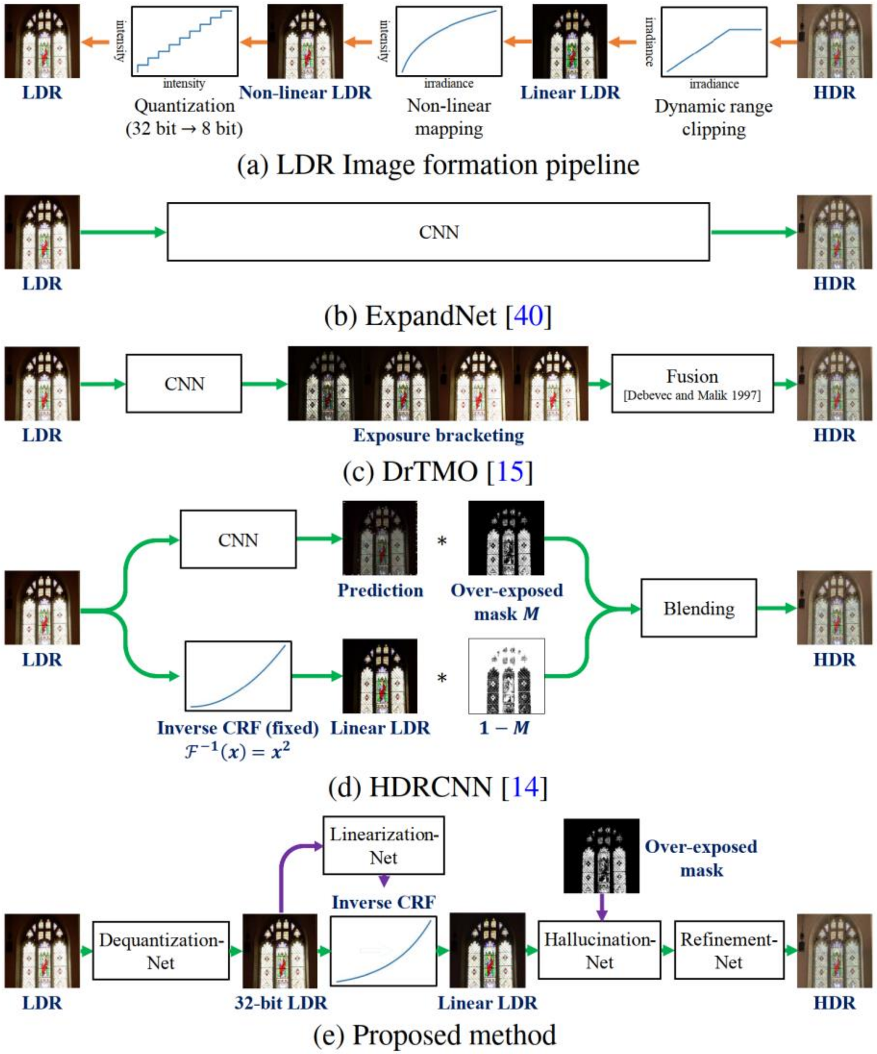
HDR Problem Formulation

: Inverse tone mapping



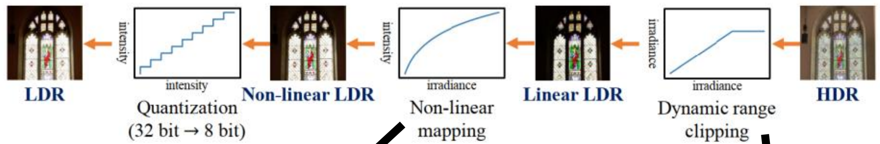
HDR Problem Formulation

: Inverse tone mapping, deep learning methods



HDR Problem Formulation

: Inverse tone mapping, deep learning methods



1. Camera Response Function Estimation

2. Saturated Region Restoration

Camera Response Function Estimation

Camera Response Function Estimation
: Overview

- What
  - Input brightness → output brightness curve
  - Sensor irradiance → pixel intensity curve
  - Light energy incident on image sensors → output of a camera

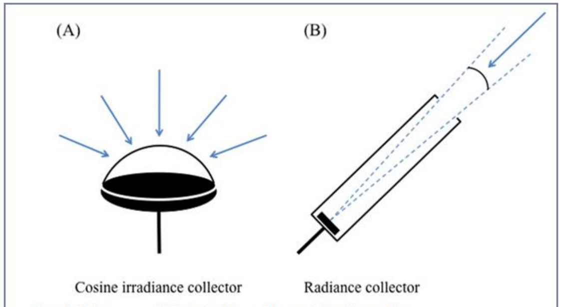
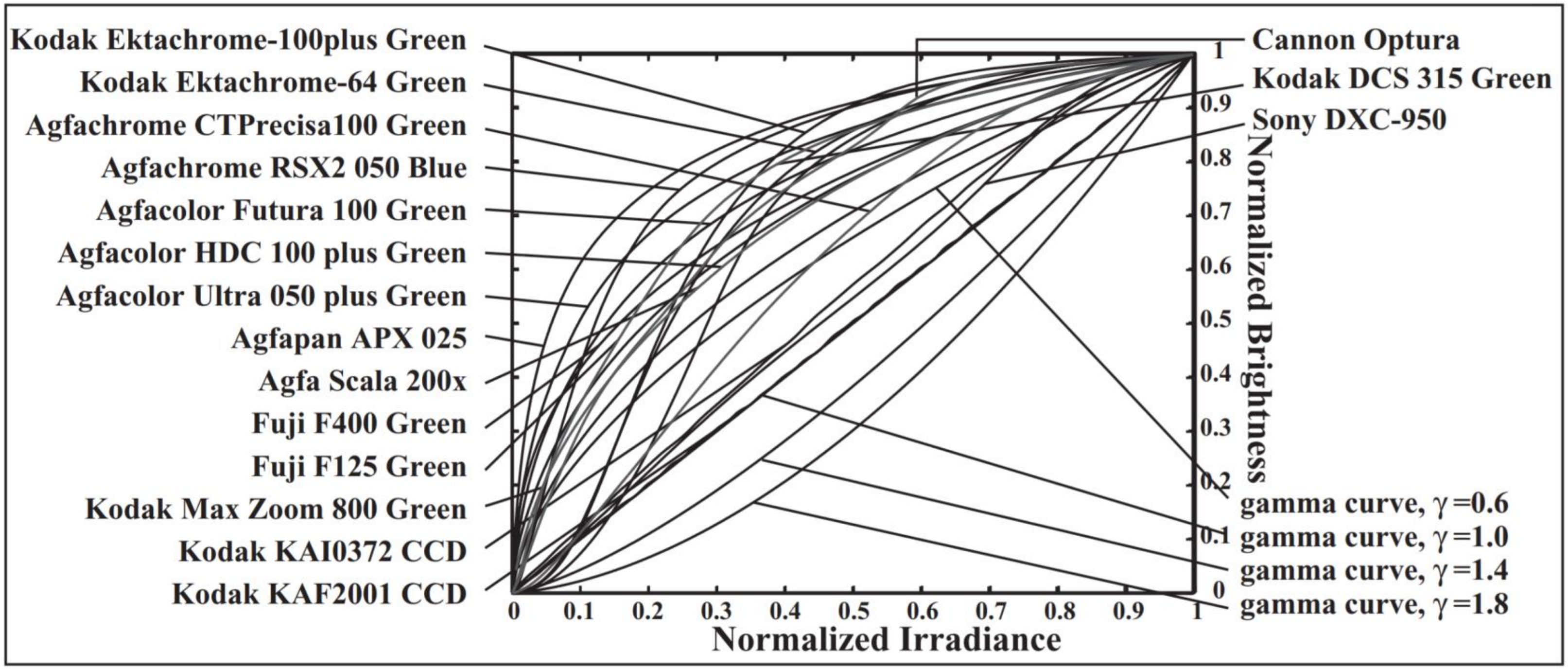
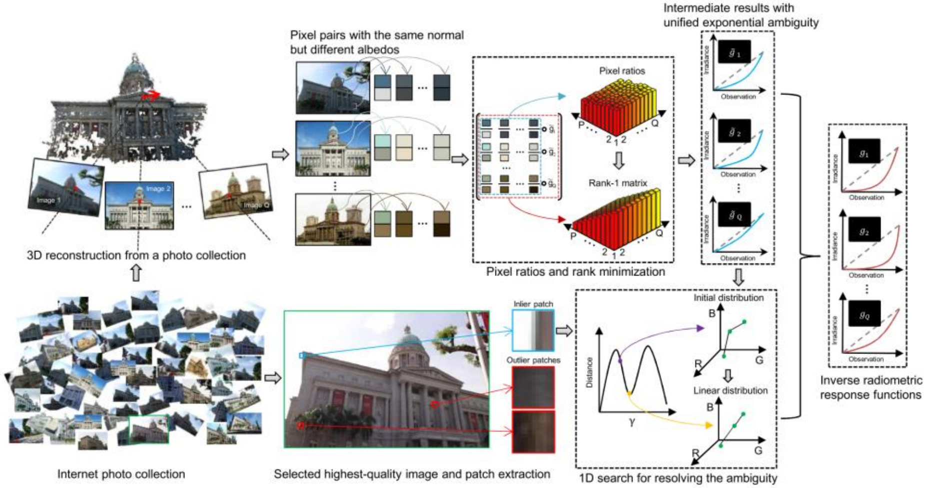


Figure 1: **Light sensors**. (A) Cosine irradiance collector and (B) radiance collector.



Camera Response Function Estimation
: Overview

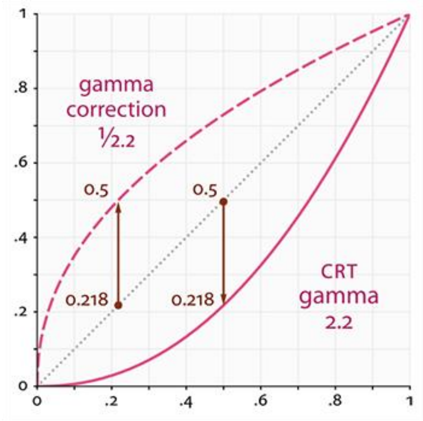
- Why
  - Computer vision algorithms require image irradiance
    - Low level vision tasks
      - De-blurring
    - Handling intensities from different exposure settings
      - Image enhancement : HDR imaging
      - 3D reconstruction : photometric stereo, shape from shading
      - ~~Image authentication : a natural watermark~~

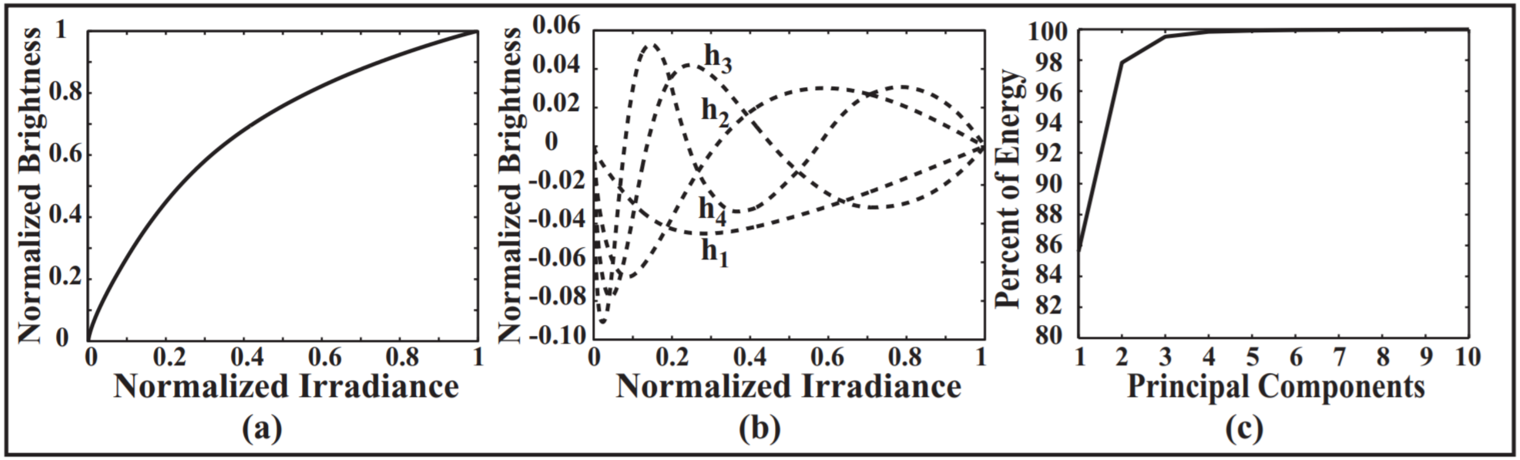


Camera Response Function Estimation

: Overview

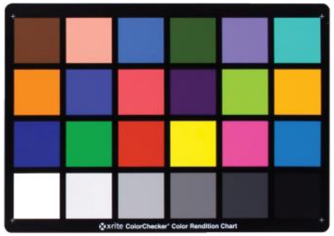
- Models
  - Gamma curves
    - 1 parameter
  - Polynomials
  - Generalized gamma curves
    - Higher-order
  - Empirical model of response (EMoR)
    - Data-driven model (from database of real-world camera response functions; DoRF)





Camera Response Function Estimation
: Overview

- Assumptions
  - Spatially uniform irradiance distribution in a scene → CRF → deviation
    - CRF as a function best restores the assumptions
- How
  - w/ known reflectance patches (Macbeth chart)
  - w/o known reflectance patches (= Automatic CRF estimation)
    - Methods (by inputs)
      - Multiple same-scene images
        - ✓Exposure ratio among images → relationship between irradiance images
      - Single channel image
        - ✓Gamma curve
          - Insufficient for real-world CRF
      - Single RGB image
        - ✓Linearly blended edges



CRF-net

: Single Image Radiometric Calibration using CNNs

- Problem statement
  - Single RGB image CRF estimation
    - Formulated as 11 EMoR model parameter estimation

- Main contribution
  - Conditioned sampling
    - (1) Random patch sample → patch-wise CRF estimation
    - (2) Select on condition
      - # (R+G+B) pixel value histogram bin > 220
    - (3) Aggregate predicted CRFs for whole image
      - Outlier removal & average
  - Pre-training
    - CRF classification

CRF-net

: Single Image Radiometric Calibration using CNNs

- Proposed method
  - Overview

Training

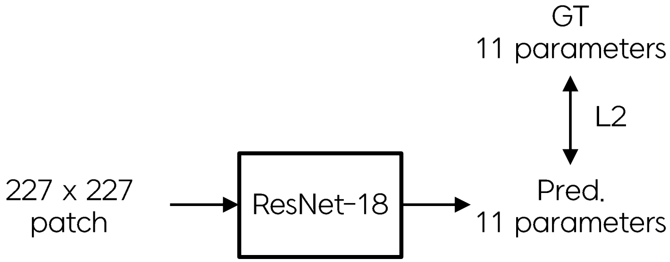
Well-exposed, linear RAW photos

↓

Tonemapped by CRFs in DoRF

↓

1 image → (227 x 227) random patch x 10



Pre-training

CRF classification
(201 likelihoods instead of 11)

Output Size	Configuration	Short-cut
114×114	$\begin{bmatrix} 7 \times 7 \times 64 \end{bmatrix}$, stride 2	
57×57	max pool 3×3, stride 2	
57×57	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1 \times 256 \end{bmatrix}$
57×57	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \end{bmatrix} \times 2$	identity
29×29	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1 \times 512 \end{bmatrix}$
29×29	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 1$	identity
23×23	average pool 7×7, stride 1	
11	fully connected	

CRF-net
: Single Image Radiometric Calibration using CNNs

- Assumptions
  - Input
    - Well-exposed, correctly white balanced
    - Ignore over/underexposed pixels
  - CRF
    - EMoR CRF model
      - Weights of 11 PCA components
    - Same CRF for each color channel
    - CRF is the only source of non-linear transformation
    - Sampled image patches are enough
- Limitations
  - Doesn't works well in outlier cases
    - Oversaturated
    - High contrast

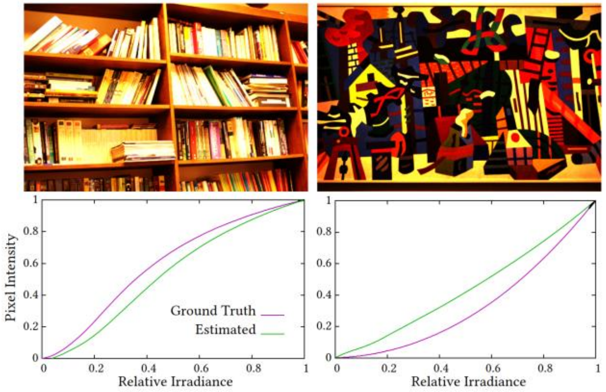


Figure 2: Examples of suboptimal radiometric calibration. The left image exhibits many oversaturated pixels, whereas the right exhibits a very high contrast. In both cases, it is difficult to find good windows that sufficiently (and uniformly) cover the full pixel range. The respective estimation (and linearization $(\times 10^{-2})$**) errors are:** 2.365 (3.037) **and** 3.925 (7.485).

Linearization-Net

: from '~ Learning to Reverse the Camera Pipeline'

- Problem statement
  - Single RGB image CRF estimation
    - Formulated as 11 EMoR model parameter estimation

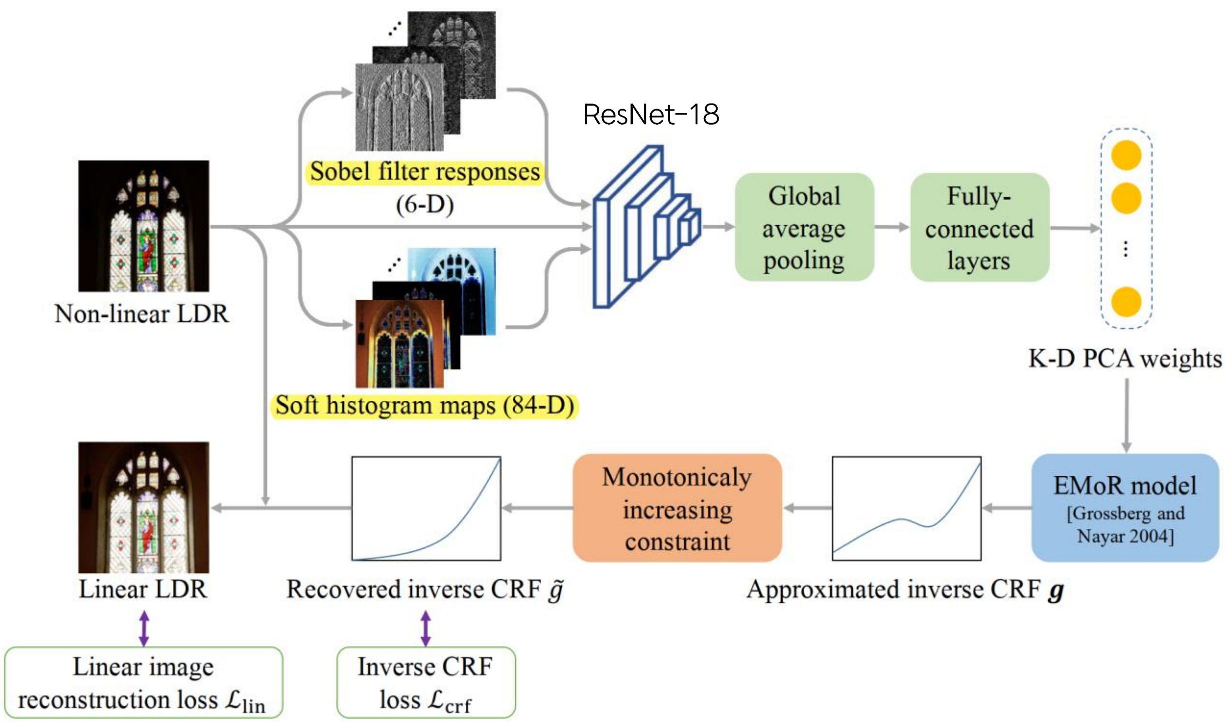
- Main contribution
  - An extension of CRF-net
    - **CRF-net + {input features + constraint}**
  - Additional Priors
    - Inspirations
      - from classical computer vision papers
    - Input features
      - Edge information, histogram
    - Constraint term

Linearization-Net

: *from '~ Learning to Reverse the Camera Pipeline'*

- Proposed method
  - Overview

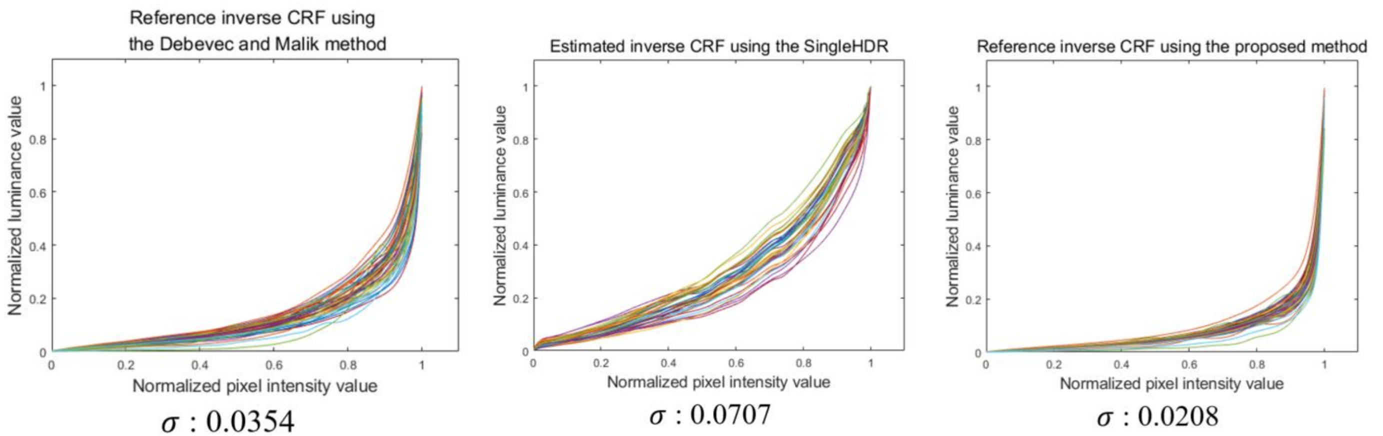
Edge information, intensity histogram



Linearization-Net

: from '~ Learning to Reverse the Camera Pipeline'

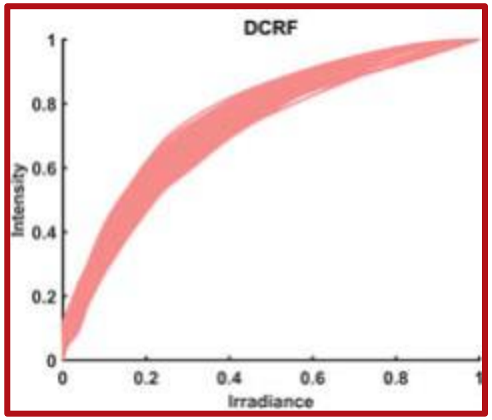
- Limitations
  - Doesn't works so well
    - Baseline method vs. Linearization-Net vs. E2E Differentiable Learning to HDR

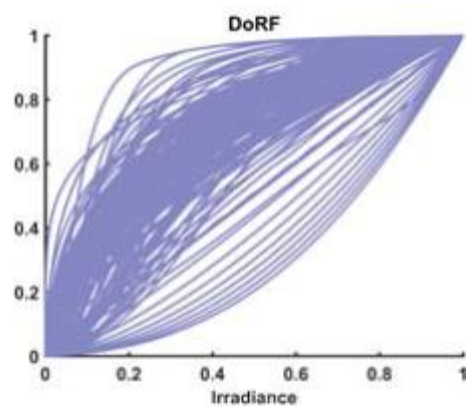


Analyzing Modern CRFs

: is CRF estimation necessary?

- Main contribution
  - A new dataset of 178 CRFs from modern digital cameras (DCRF dataset)
    - From camera review color chart images available online
  - CRF estimation method for/from the proposed dataset
  - Answer question about modern CRFs
    - Which mathematical models are best for CRF estimation?
    - How have CRFs changed over time?
    - And how unique are CRFs from camera to camera?

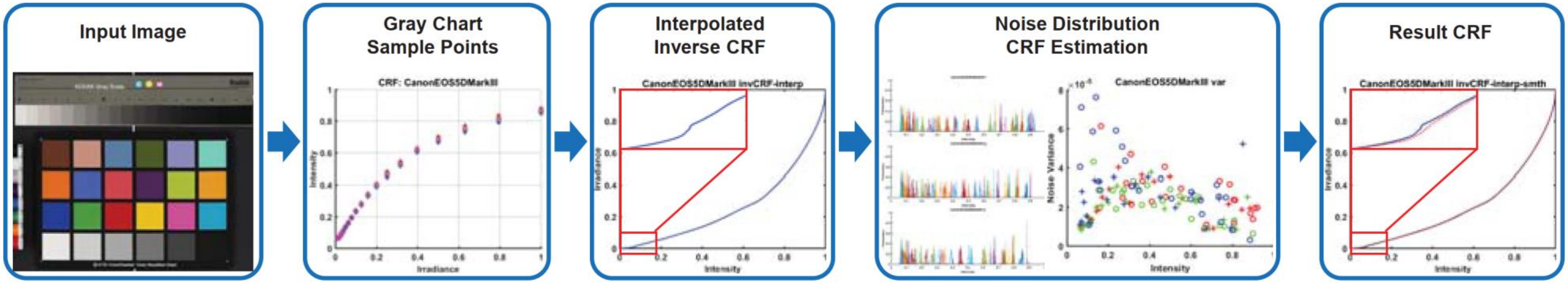




Analyzing Modern CRFs

: is CRF estimation necessary?

- Proposed method



Internet camera review images | Initial sparse CRF estimation | Coarse interpolation | Noise distribution measure | Refining using noise distribution

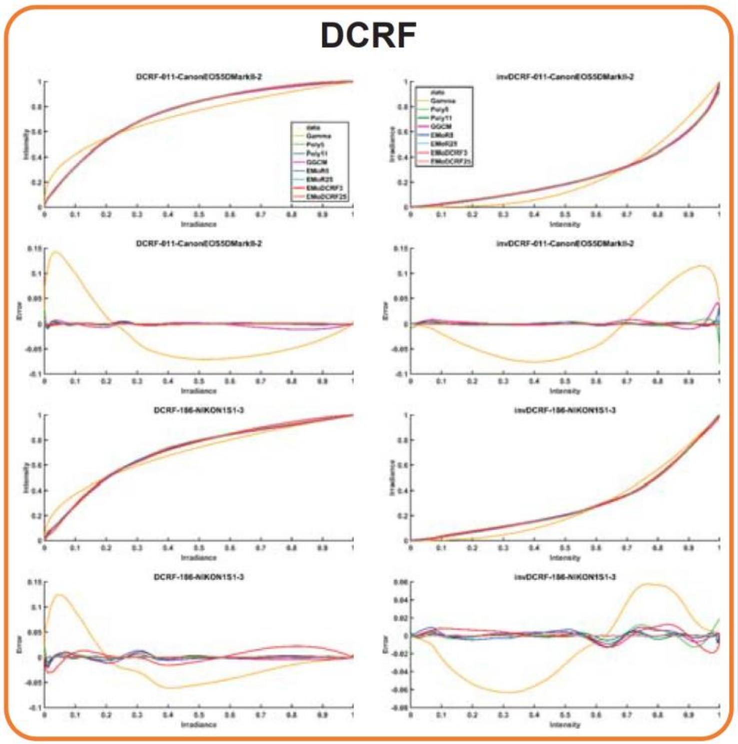
Analyzing Modern CRFs

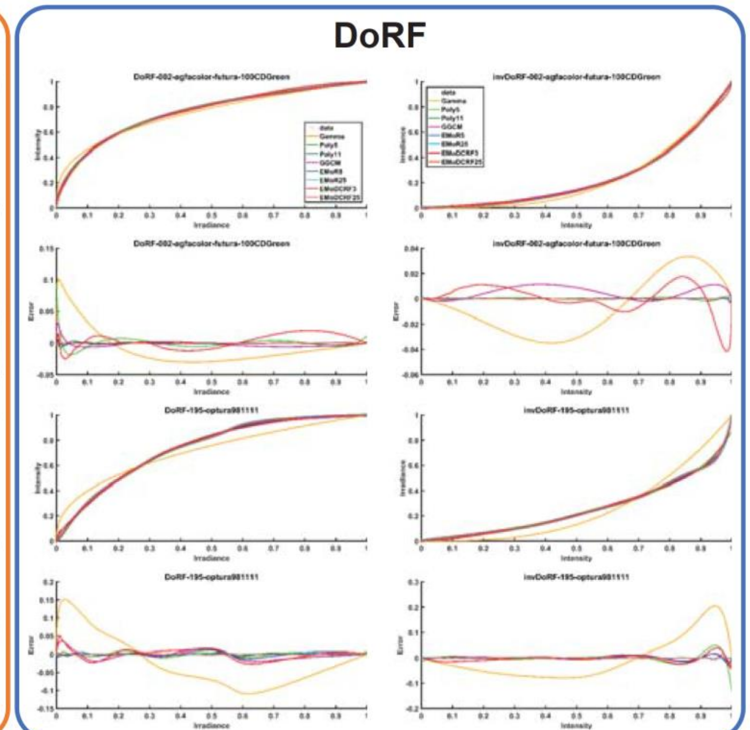
: is CRF estimation necessary?

- Which mathematical models are best for CRF estimation?

Avg. RMSE over different datasets

Data	**Gamma**	**Poly-5**	**Poly-11**	**GGCM**	**EMoR-5**	**EMoDCRF-3**	**EMoR-25**	**EMoDCRF-25**	
DCRF	0.056394	0.005150	0.001860	0.026221	0.003119	0.002362	0.000608	**0.000008**	
invDCRF	0.057726	0.002796	0.001908	0.006732	0.002485	0.002870	0.000320	**0.000065**	
DoRF	0.061654	0.006353	0.001981	0.008556	0.002328	0.018937	**0.000114**	0.003128	
invDoRF	0.054723	0.005829	0.001655	0.023942	0.001790	0.016689	**0.000154**	0.000641	
	1 [13]	5 [19]	11 [15]	2 [17]	5 [20]	3	25 [5]	25	# parameters

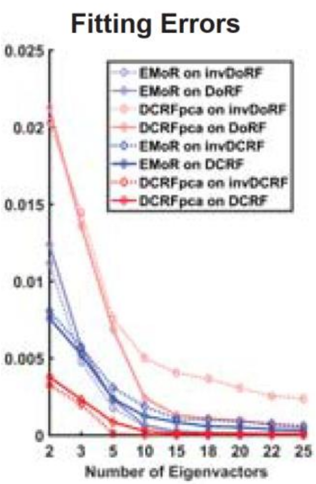




Analyzing Modern CRFs

: is CRF estimation necessary?

- Limitation
  - Purpose
    - CRF estimation as a measure of camera characteristics
      - Originally white color → arbitrary pixel value
    - CRF estimation as a preprocessing for HDR image reconstruction (→ linearlization → HDR image reconstruction)
      - Originally arbitrary color (but too much) → arbitrary pixel value
  - Proposed CRF estimation method
    - Dataset overfitted method
      - Can be justified if their dataset better represents ideal distribution
      - But do they? (online images)
  - Experiments
    - Insufficient comparison with baseline models
      - # parameters : 5 vs. 5, 11 vs. 11



Camera Response Function Estimation

: Conclusion

- Accurate CRF is required for better inverse tone mapping
  - As a preprocessing for HDR reconstruction pipeline
  - Can be considered as a domain generalization problem
- CRFs are camera dependent characteristics
  - There's no single gamma parameter fits all
  - Calls for accurate CRF estimation method
- Modern digital cameras *may* exhibit similar CRFs (than film cameras)
  - But not exactly the same
- Deep learning-based CRF estimation methods have been proposed
  - But not extensively explored

Saturated Region Restoration

Masked Features and Perceptual Loss
: Focus on saturated region restoration

- Problem statement
  - Recovering the missing information in the saturated highlights

- Main contribution
  - Network
    - Feature masking & mask update
      - *Same filters can be used to compute the contribution of the valid pixels in the features*
  - Training
    - Inpainting pre-training
  - Input
    - Patch sampling

Masked Features and Perceptual Loss

: Focus on saturated region restoration

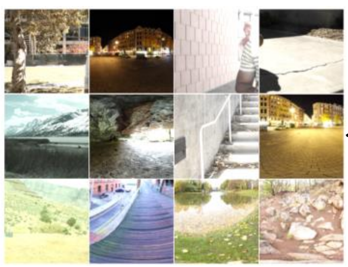
- Proposed method

Pre-training

Random masked
MIT Places [2014]

Image inpainting for irregular holes using partial convolutions [ECCV 2018]
Learning deep features for scene recognition using places database [NIPS 2014]

Fine-tuning

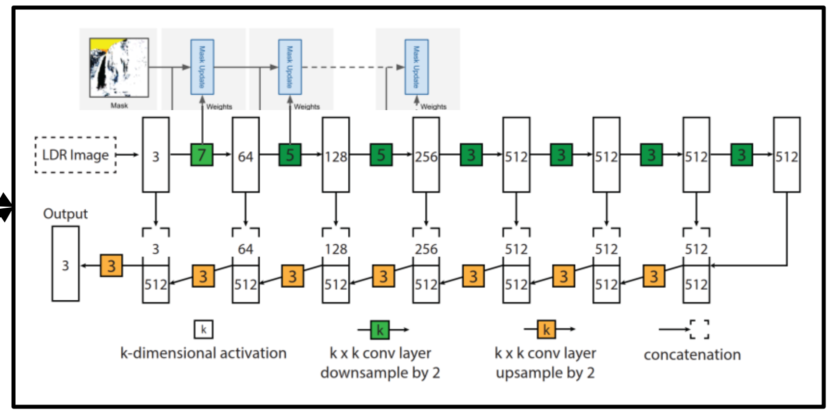


1 image → (512 x 512) random patch x 250

Algorithm 1 Patch Sampling

```
procedure PatchMetric(H, M)
    H: HDR image, M: Mask
    σ_c = 100.0                     ▷ Bilateral filter color sigma
    σ_s = 10.0                      ▷ Bilateral filter space sigma
    I = RgbToGray(H)
    L = log(I + 1)
    B = bilateralFilter(L, σ_c, σ_s)        ▷ Base Layer
    D = L - B                               ▷ Detail Layer
    G_x = getGradX(D)
    G_y = getGradY(D)
    G = abs(G_x) + abs(G_y)
    return mean(G ⊙ (1 − M))
```

U-Net like



Leaky ReLU

ReLU

Loss = L1 (log scale + masked)
+ VGG + style

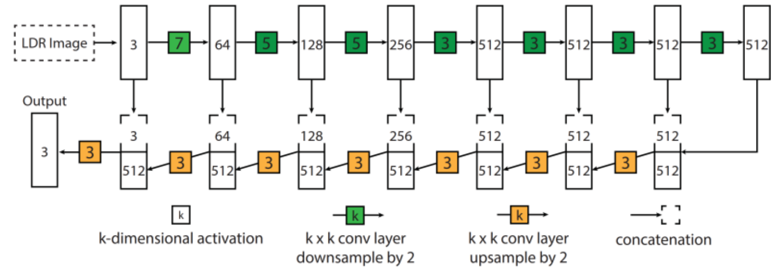
$$\hat{H} = M \odot T^{\gamma} + (1 - M) \odot [\exp(\hat{Y}) - 1]$$

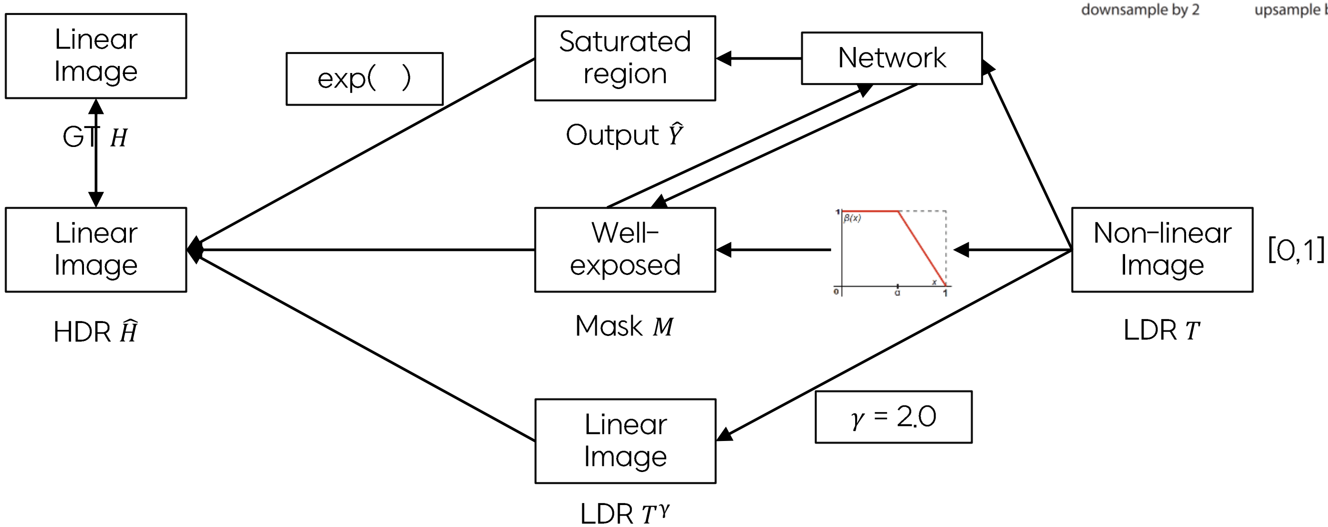
Pred H = final HDR
M = Mask [0,1]
T = input LDR [0,1]
Γ = gamma for linearlization (2.0)
Pred Y = network output (in log)

Masked Features and Perceptual Loss

: Focus on saturated region restoration

- Proposed method
  - Overview





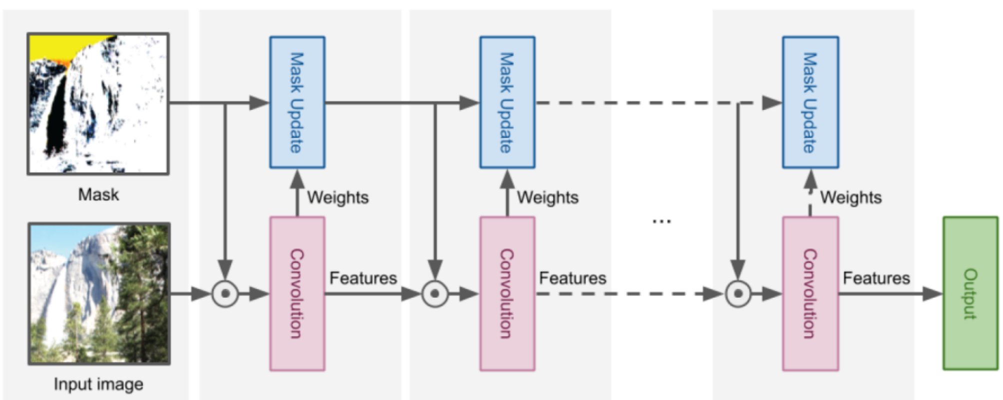
$$\hat{H} = M \odot T^{\gamma} + (1 - M) \odot [\exp(\hat{Y}) - 1]$$

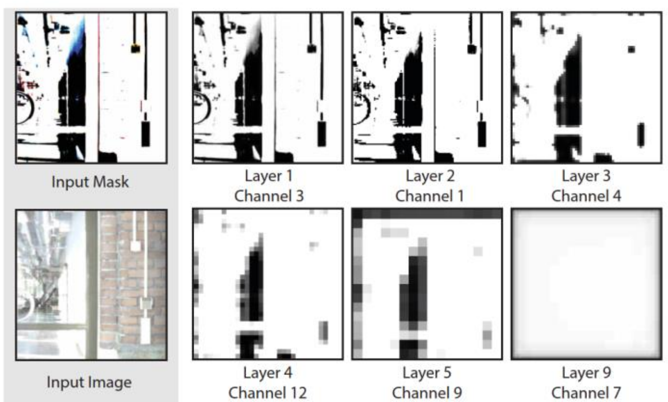
Pred H = final HDR
M = Mask [0,1]
T = input LDR [0,1]
Γ = gamma for linearlization = 2.0
Pred Y = network output (in log scale)
⊙ = Element-wise multiplication

Masked Features and Perceptual Loss

: Focus on saturated region restoration

- Proposed method
  - Feature masking & mask update
    - Soft mask [0,1]
      - Features from weakly saturated regions are not discarded
    - **Feature masking** : reduce magnitude of the features generated from the saturated content
      - Element-wise multiplication of feature map & mask
    - **Mask update** : update contribution of valid mask with same conv. layer
      - Also convolve mask with conv. layer weights





Masked Features and Perceptual Loss

: Focus on saturated region restoration

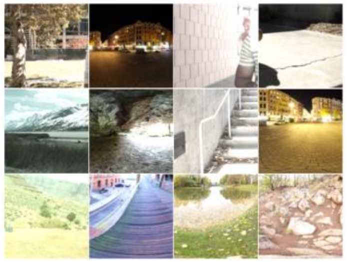
- Proposed method
  - Inpainting pre-training
    - Limited dataset
      - Prior methods
        - ✓Pre-train : simulated HDR (from standard images)
        - ✓Fine-tune : real HDR
          - Didn't worked!
      - Proposed method
        - ✓Pre-train : **inpainting dataset**
          - Learn to create plausible **textures**
          - Binary mask
        - ✓Fine-tune : HDR dataset
          - Adapt to HDR domain
          - … and adapt to saturated region
            - Smooth & textureless

Pre-training

Random masked
MIT Places [2014]

Image inpainting for irregular holes using partial convolutions [ECCV 2018]
Learning deep features for scene recognition using places database [NIPS 2014]

Fine-tuning



1 image → (512 x 512) random patch x 250

Algorithm 1 Patch Sampling

```
procedure PatchMetric(H, M)
    H: HDR image, M: Mask
    σ_c = 100.0                          ▷ Bilateral filter color sigma
    σ_s = 10.0                           ▷ Bilateral filter space sigma
    I = RgbToGray(H)
    L = log(I + 1)
    B = bilateralFilter(L, σ_c, σ_s)     ▷ Base Layer
    D = L - B                            ▷ Detail Layer
    G_x = getGradX(D)
    G_y = getGradY(D)
    G = abs(G_x) + abs(G_y)
    return mean(G ⊙ (1 − M))
```

Masked Features and Perceptual Loss

: Focus on saturated region restoration

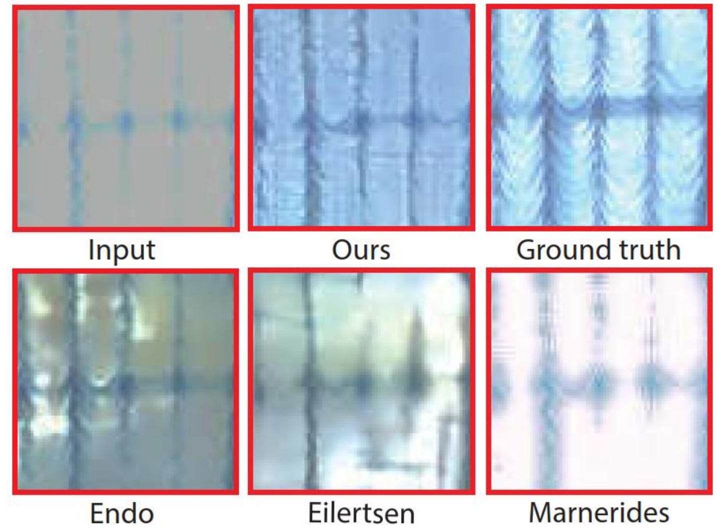

Base Detail Color

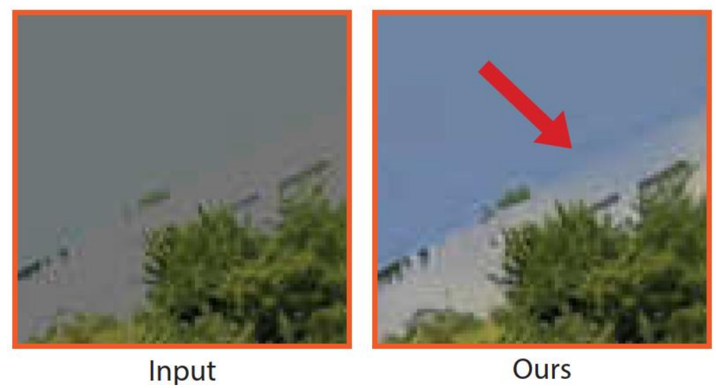
- Proposed method
  - Patch sampling
    - Problem statement
      - How to effectively learn **textures** of **saturated** region
        - ✓Learn from patches with textures and saturated region
      - How to detect & **measure** textured patches
    - Proposed method
      - HDR image decomposition → base layer + **detail layers**
        - ✓"Fast bilateral filtering for the display of high-dynamic-range images", *Siggraph 2002*
          - On how to diplay HDR images on displays with limited dynamic range
            - How to reduce the contrast while preserving detail
          - Two-scale decomposition of the image
            - Base layer : encoding large-scale variations → reduce contrast
            - Detail layer : preserve details
      - Saturated area classification
        - ✓Avg. of the gradients (Sobel operator) of the detail layer > threshold (0.85) → textured

Masked Features and Perceptual Loss

: Focus on saturated region restoration

- Limitations & conclusion
  - Overexposed/satrated region restoration is hard
    - Detailed areas often fail
    - Often input lacks any information at all
  - Color distortion
    - Blend nearby colors
      - Gray buliding + blue sky = blue (building + sky)
  - Temporally unstable
    - Not applicable for HDR videos





What to Expect From This Seminar

- Broad *(but shallow)* understanding of the HDR problem

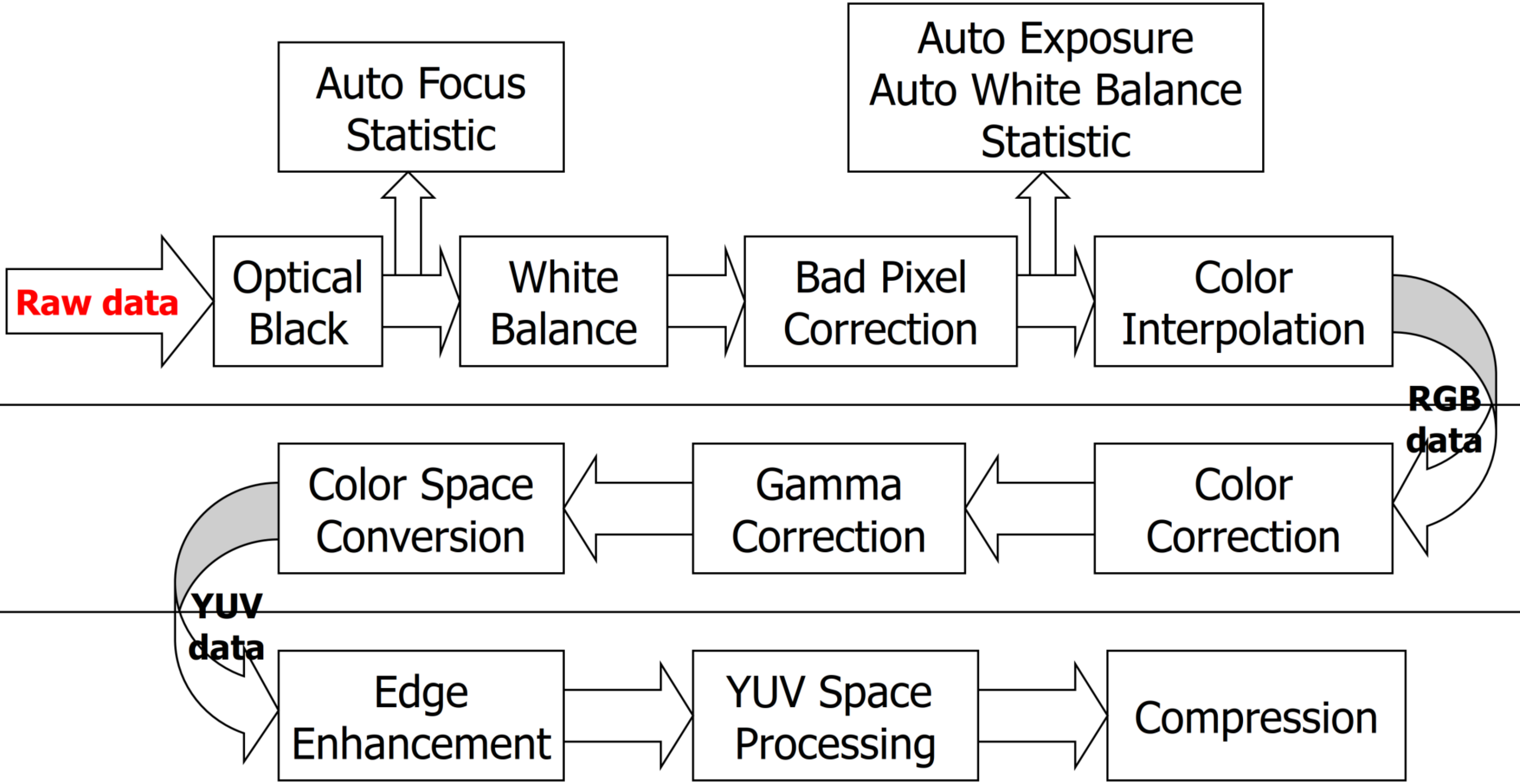
- Major CV conference topics in HDR
  - Special camera
    - Lens, sensor
  - Common camera
    - Single image HDR reconstruction

- Two major problems in single image HDR reconstruction using deep learning
  - Camera response function estimation
  - Saturated region restoration

Supplementary

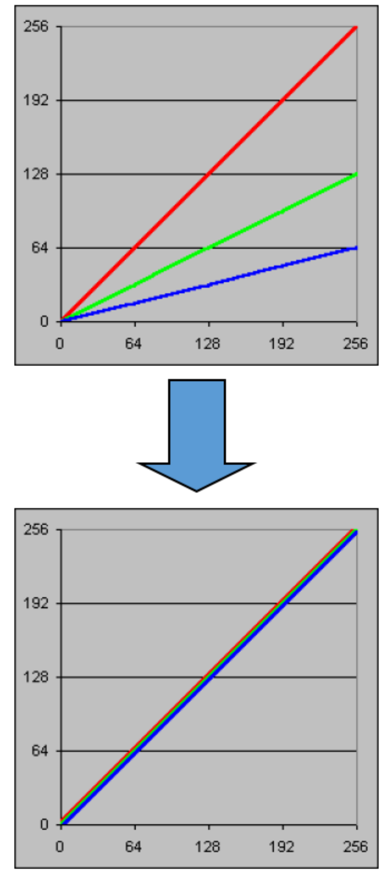
ISP pipeline

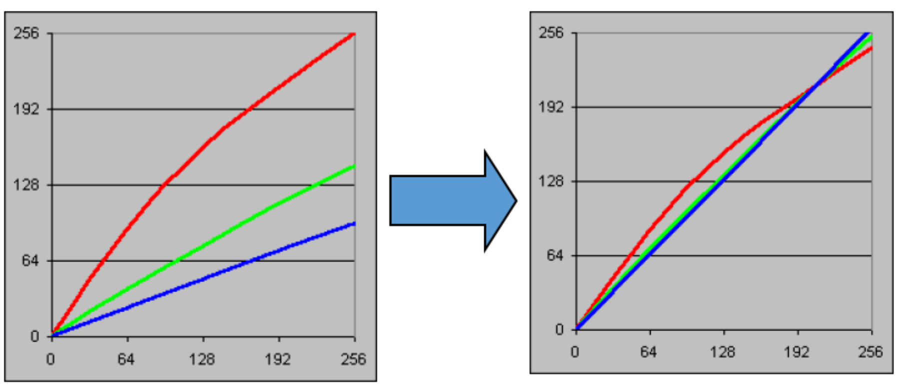
: in signal processing, optics (practical)



White Balance: Matching Human Perception

- **To simulate human eyes white balance**: adjusting colors so that the image looks more natural
- **Adjustable channel gain** for each color channel
- General approaches
  - Gray world assumption
  - Perfect reflector assumption
  - Calibration based approaches
- What if data are nonlinear?

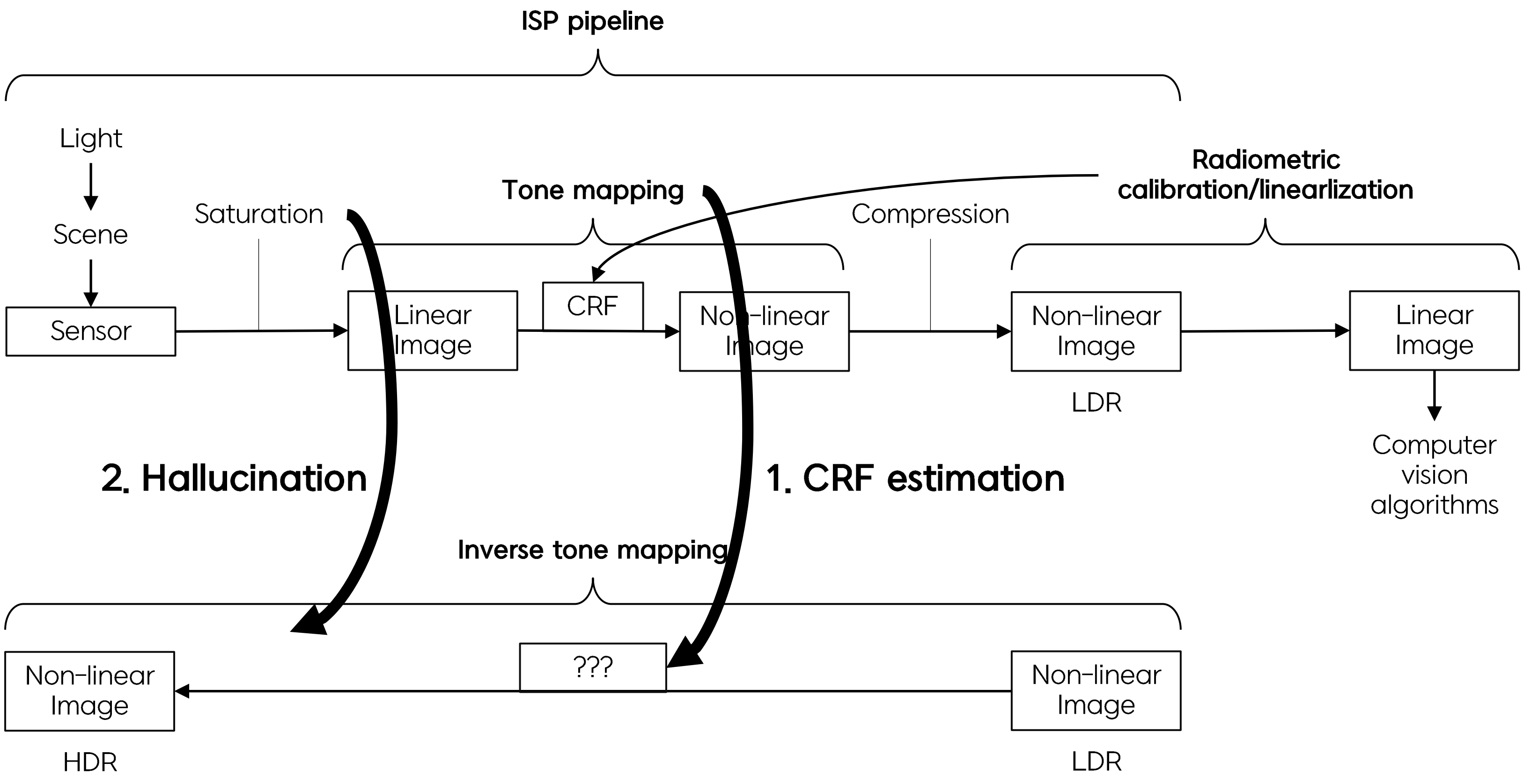




Tone Mapping

- Map tone curve to **get better image**
- Similar to histogram adjustment or Photoshop's curve function
- For Y channel only

HDR problem formulation
: Inverse tone mapping, deep learning methods



Masked Features and Perceptual Loss

: Focus on saturated region restoration

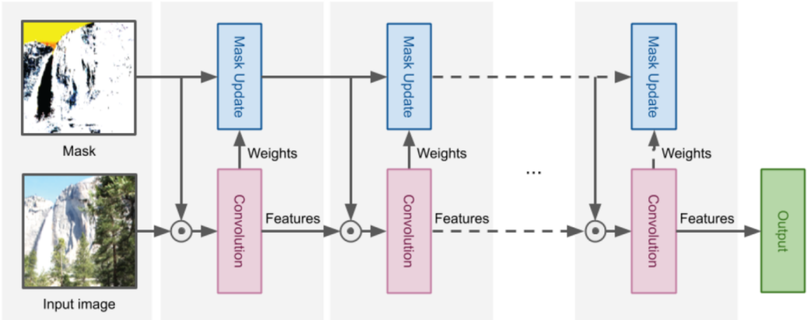
- Proposed method
  - Feature masking
    - Soft mask
      - Features from weakly saturated regions are not discarded
    - **Feature masking** : reduce magnitude of the features generated from the saturated content
      - **Element-wise multiplication** of feature map & mask [0,1]

$$Z_l = X_l \odot M_l \qquad X_l \in \mathbb{R}^{H\times W\times C} \qquad M_l \in [0,1]^{H\times W\times C}$$

    - **Mask update** : update contribution of valid mask with same conv. layer
      - Also **convolve** mask with conv. layer weights

$$M_{l+1} = \left(\frac{|W_l|}{\|W_l\|_1 + \epsilon}\right) * M_l \qquad |W_l| \in \mathbb{R}^{H\times W\times C} \qquad \|W_l\|_1 \in \mathbb{R}^{1\times 1\times C}$$

Replicated to fit H x W x C



Masked Features and Perceptual Loss
: Focus on saturated region restoration

- Proposed method
  - Loss
    - L1 loss $L_r = \|(1-M) \odot (\hat{Y} - \log(H+1))\|_1$
    - VGG loss $L_v = \sum_l \|\phi_l(\mathcal{T}(\tilde{H})) - \phi_l(\mathcal{T}(H))\|_1$ $\quad \tilde{H} = M \odot H + (1-M) \odot \hat{Y} \quad \mathcal{T}(H) = \frac{\log(1+\mu H)}{\log(1+\mu)}$
    - Style loss $L_s = \sum_l \|G_l(\mathcal{T}(\tilde{H})) - G_l(\mathcal{T}(H))\|_1$

$$\underbrace{G_l(X)}_{C_l \times C_l} = \underbrace{\frac{1}{K_l}}_{\substack{\text{Normalization factor} \\ C_l H_l W_l}} \phi_l(X)^T \underbrace{\phi_l(X)}_{(H_l W_l) \times C_l}$$