UnModNet: Learning to Unwrap a Modulo Image for High Dynamic Range Imaging

Chu Zhou[1], Hang Zhao[2], Jin Han[1], Chang Xu[3], Chao Xu[1], Tiejun Huang[1], Boxin Shi[1*]

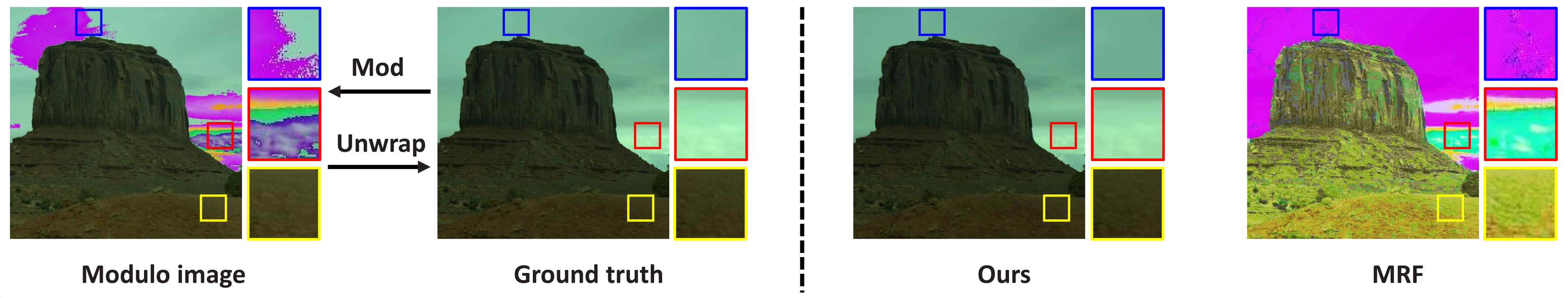
[1]Peking University [2]Massachusetts Institute of Technology [3]University of Sydney

zhou_chu@pku.edu.cn, shiboxin@pku.edu.cn *Corresponding author

Contributions

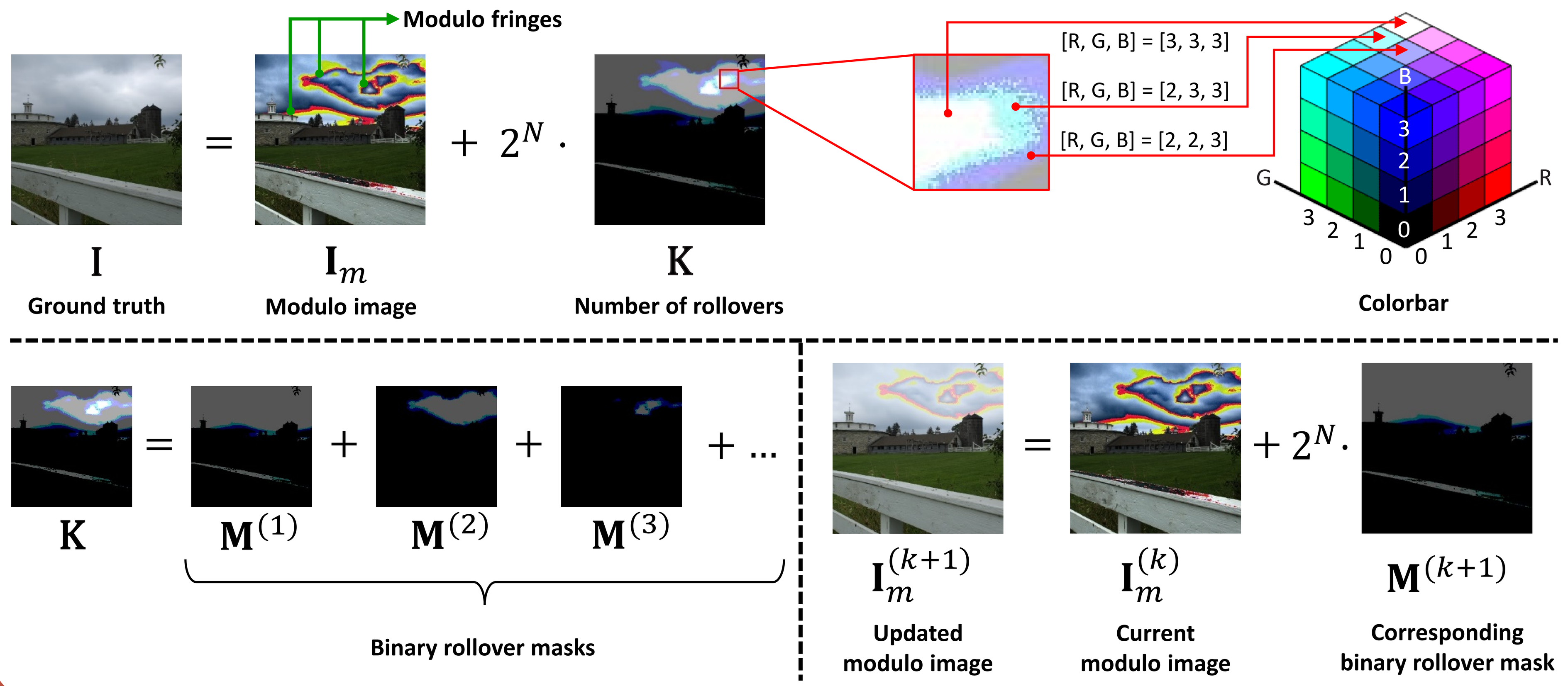
- **UnModNet**: a learning-based modulo image unwrapping framework to recover the original scene radiance from its modulo counterpart for high dynamic range imaging.

  Three customized model designs are proposed to solve the three issues in the previous MRF-based unwrapping algorithm [Zhao *et al.*, ICCP2015] respectively:
  - Modulo edge separator ⟶ Modulus-intensity ambiguity.
  - Rollover mask predictor ⟶ Strong local contrast.
  - Consistent color prediction ⟶ Color misalignment.



Problem Formulation

- Goal: Restore the ground truth HDR image **I** by unwrapping a single modulo image $\mathbf{I}_m$ captured by a modulo camera.
  - Equivalent to estimate the number of rollovers **K**.
- Overall pipeline: Reformulate the unwrapping of a modulo image into a series of binary labeling problems to relieve the ill-posedness.
  - Iteratively update the input modulo image $\mathbf{I}_m$ by predicting the corresponding binary rollover mask **M**.
  - Output the HDR result **I** until the algorithm terminates at $\mathbf{M}^{(k+1)} = \mathbf{0}$.



Method

Network Architecture

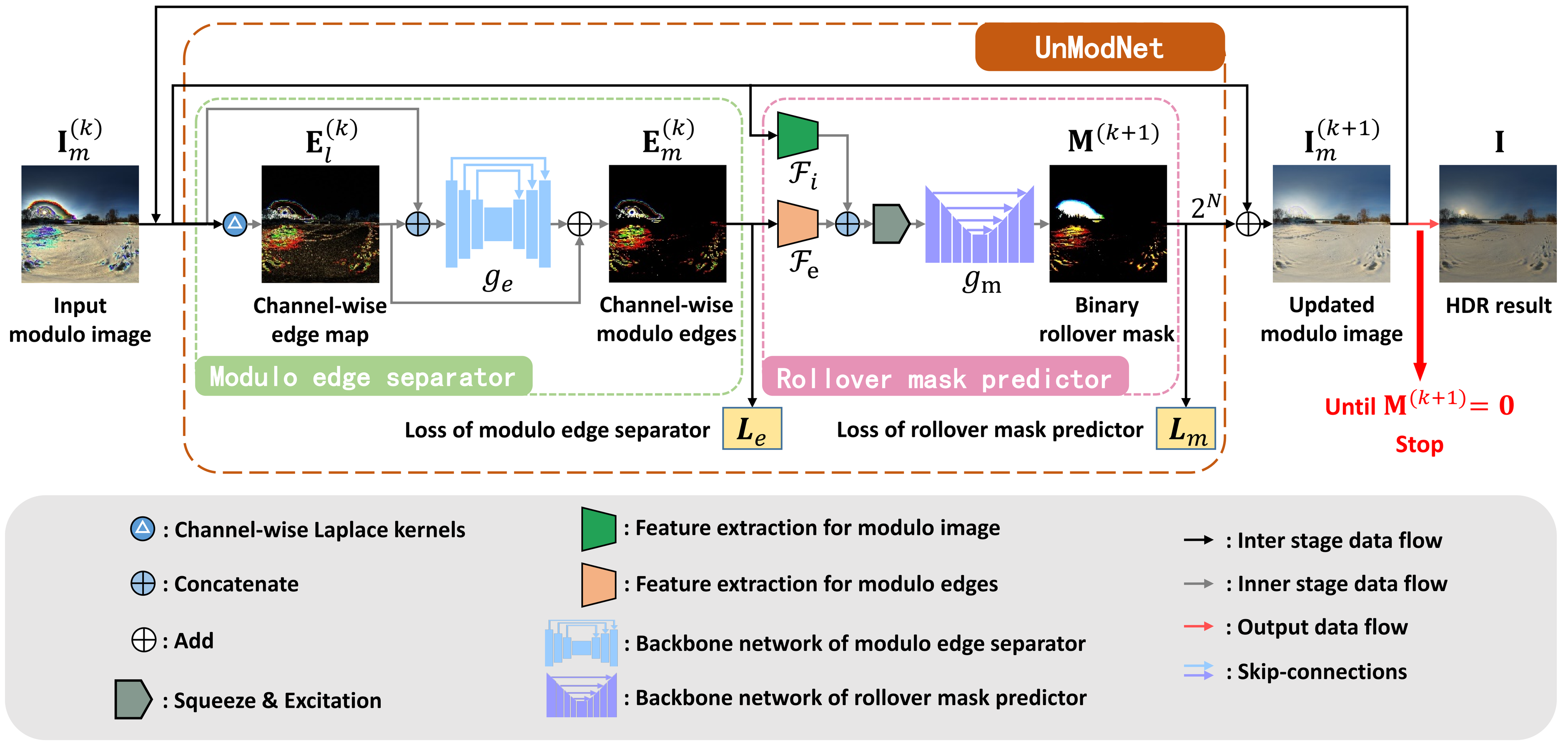
Based on the unique features of modulo pixels and edges, we design UnModNet to be two-stage:

(1) **Modulo edge separator** (to predict channel-wise modulo edges $\mathbf{E}_m$):

$$\mathbf{E}_m = \mathbf{E}_l + g_e(\mathtt{cat}(\mathbf{E}_l, \mathbf{I}_m)).$$

(2) **Rollover mask predictor** (to predict the binary rollover mask **M**):

$$\mathbf{M} = g_m(\mathtt{SE}(\mathtt{cat}(\mathcal{F}_i(\mathbf{I}_m), \mathcal{F}_e(\mathbf{E}_m)))).$$



Each unwrapping iteration can be written as follow (g represents the proposed UnModNet):

$$\mathbf{I}_m^{(k+1)} = \mathbf{I}_m^{(k)} + 2^N \cdot \mathbf{M}^{(k+1)} = \mathbf{I}_m^{(k)} + g(\mathbf{I}_m^{(k)}).$$

Loss Function

- The total loss function of UnModNet: $\mathcal{L} = \alpha \cdot \mathcal{L}_e + \mathcal{L}_m$.
  - The binary cross entropy loss is used for both $\mathcal{L}_e$ and $\mathcal{L}_m$.

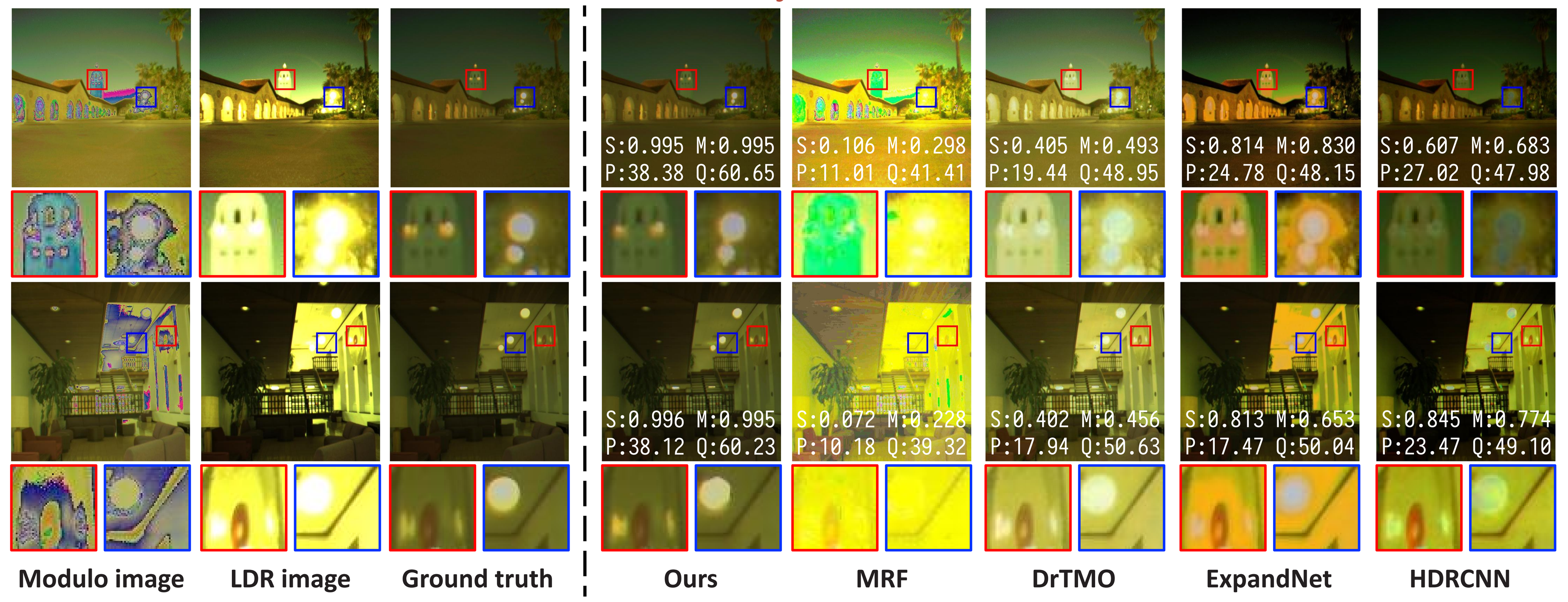
Dataset preparation

- Ground truth HDR image **I**: $\mathbf{I} = \lfloor (2^B - 1) \cdot \mathtt{clip}(\mathcal{E} \cdot \Delta t, [0,1]) \rfloor$.
  - B: quantization bit depth.
  - $\mathcal{E}$: relative irradiance values of each raw HDR image ($\mathcal{E} \in [0,1]$).
  - Δt: appropriate exposure time to control the over-exposure rate.
- Modulo image $\mathbf{I}_m$, the number of rollovers **K**, and the binary rollover masks **M**:

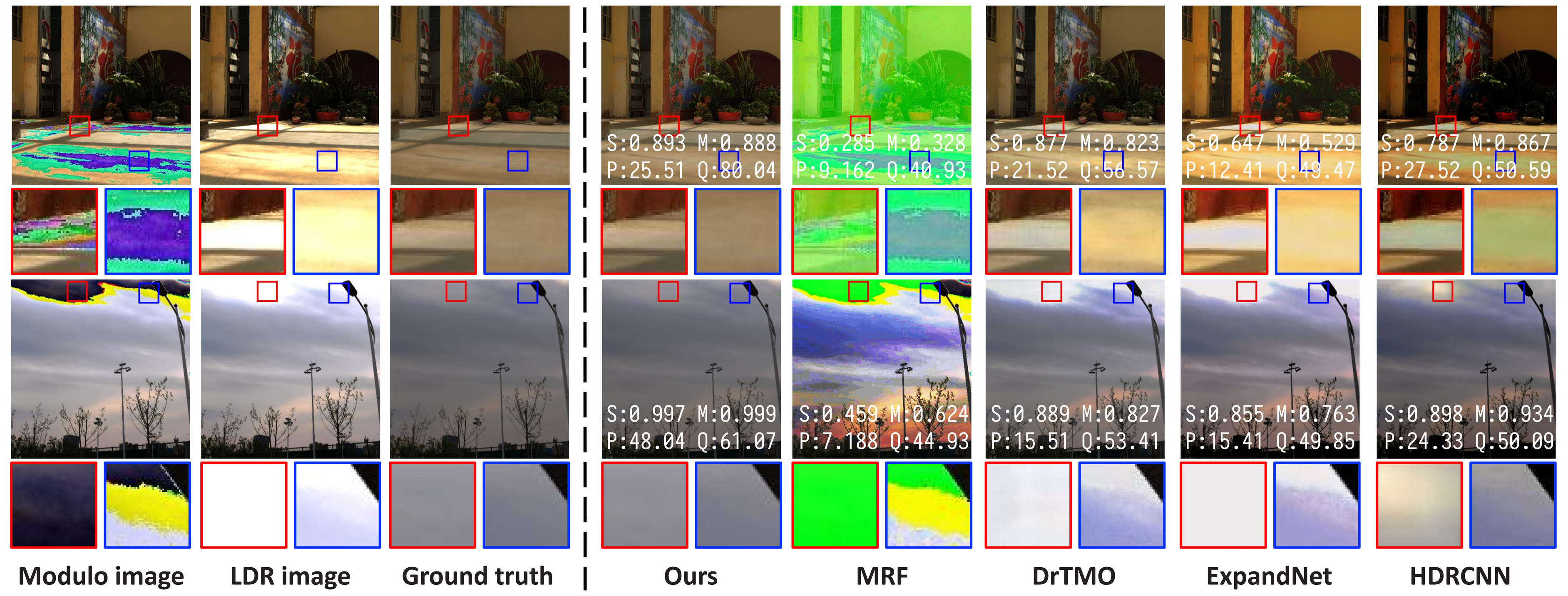
$$\mathbf{I} = \mathbf{I}_m + 2^N \cdot \mathbf{K} \quad \text{and} \quad \mathbf{K} = \sum_{k=1}^{\infty} \mathbf{M}^{(k)}.$$

Experiments

Results on synthetic data

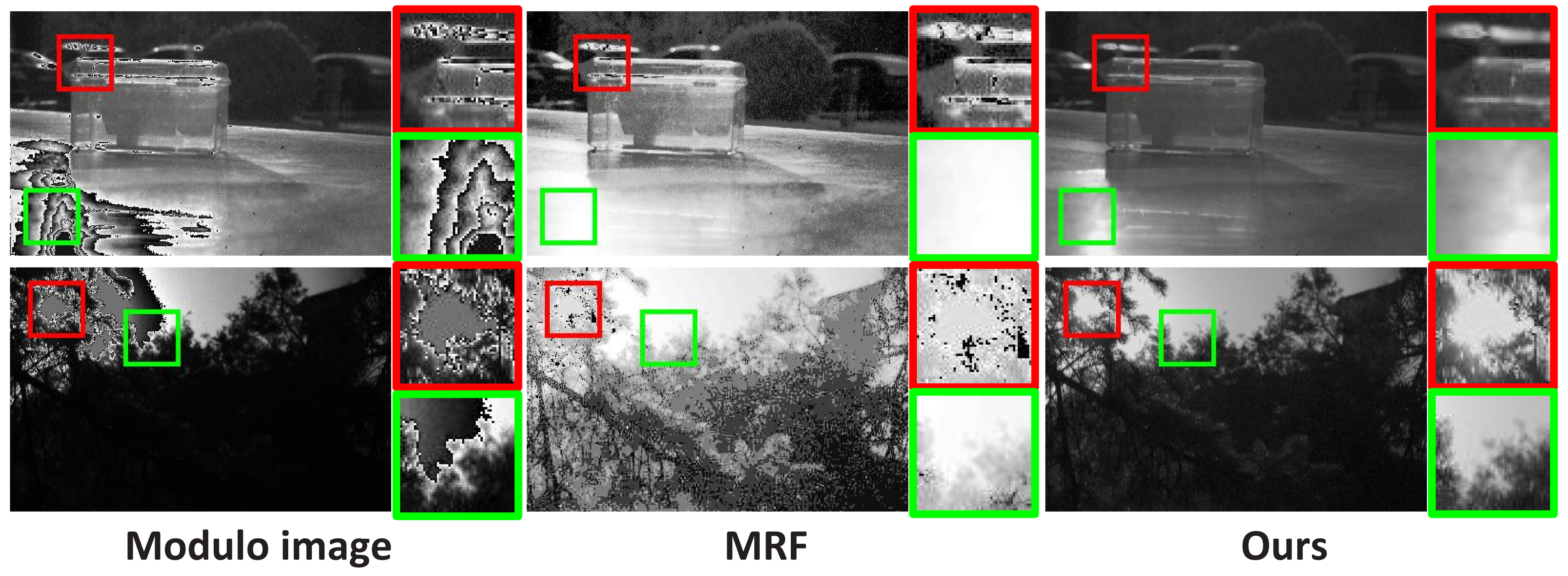


Results on real RGB images



- S: SSIM M: MS-SSIM P: PSNR Q: Q-Score (produced by HDR-VDP-2.2)

Results on images captured by SpiCam-Mod



- SpiCam-Mod: a retina-inspired fovea-like sampling model (FSM) based spike camera [Zhu *et al.*, ICME2019] configured in modulo mode.