ICCV HONOLULU HAWAII OCT 19-23, 2025

Robust Unfolding Network for HDR Imaging with Modulo Cameras

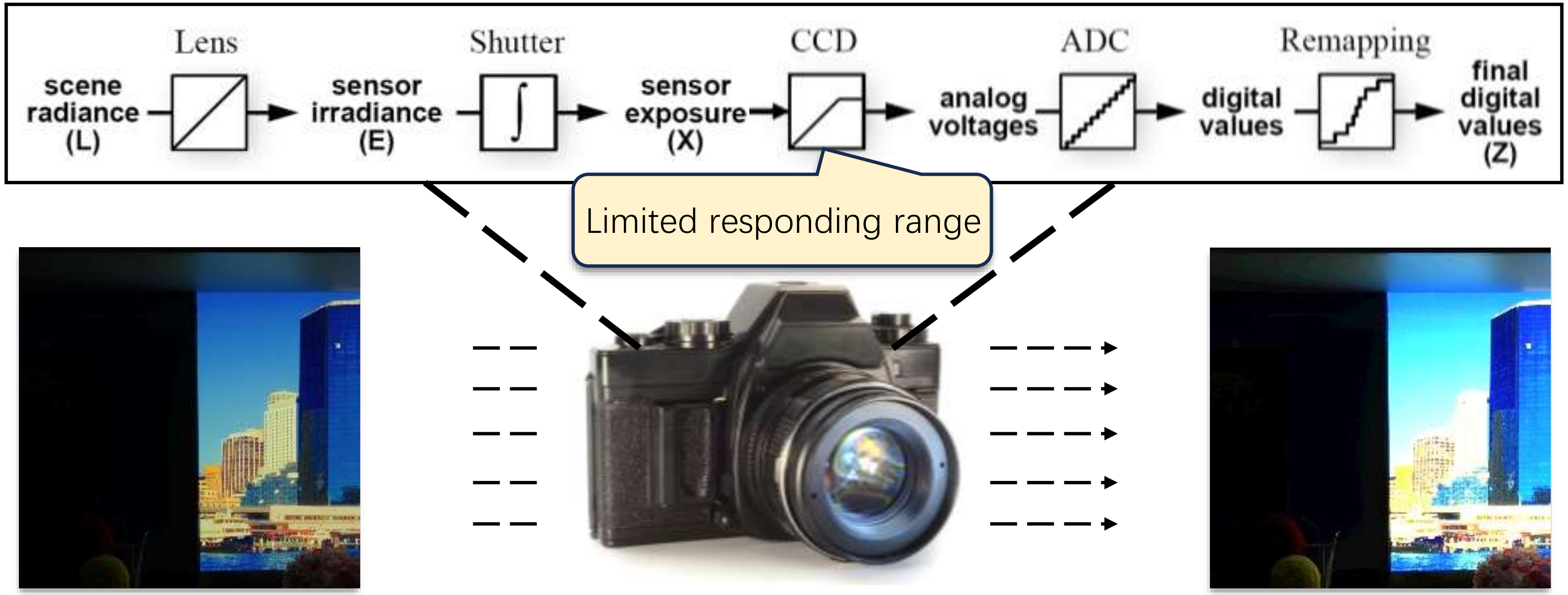
Zhile Chen

Hui Ji

华南理工大学
South China University of Technology

NUS National University of Singapore

High Dynamic Range (HDR) Imaging



Scene with high dynamic range

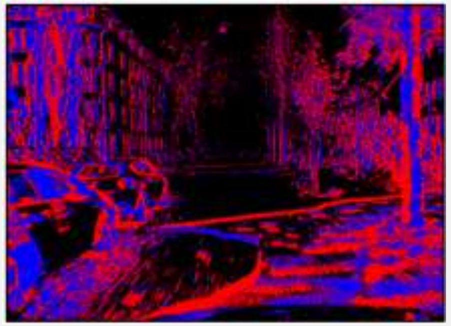
Low dynamic range (LDR) Image

Due to the physical property of standard camera, natural scene with high dynamic range (contrast) will be captured as the over(under)-exposure areas in the captured photo.

High Dynamic Range (HDR) Imaging

Multi-exposure HDR

Single-exposure HDR



Event data

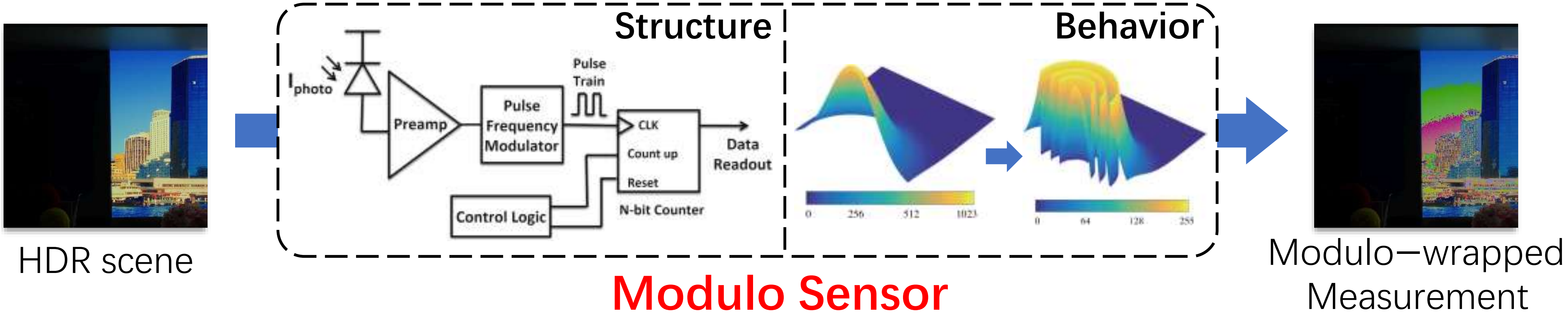
Infrared radiation

Modulo sample

Captured by
modulo camera

Specific sensing techniques for HDR

Modulo Camera for HDR Imaging



Mechanism: Once the accumulated radiance value reaches pre-defined threshold, *e.g.*, 256 for an 8-bit sensor, it resets to zero and resumes counting, enabling a theoretically unbounded dynamic range.

Formulation: Relationship between the HDR image and modulo measurement is

$$\boldsymbol{X} = \mathcal{M}(\boldsymbol{Y}), \qquad \mathcal{M}: \mathcal{M}(y) = y \bmod a, \;\; a = 2^b,$$

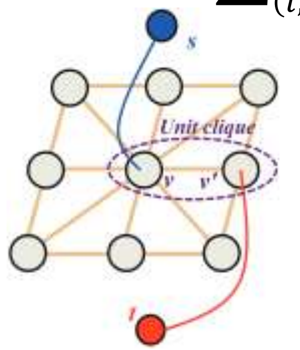
- $\boldsymbol{X} \in [0, a)^{H\times W\times C}$: Modulo-wrapped measurement; $\boldsymbol{Y} \in [0, 2^B)^{H\times W\times C}$: HDR image.

Goal: Reconstruct the HDR Image $\boldsymbol{Y}$ from its modulo-wrapped measurement $\boldsymbol{X}$.

Related Works

- Optimization using handcrafted prior.
  E.g., MRF [1] uses graph-cut optimization with smoothness prior enforced.
  ※ **Limited performance on real-world images with complex patterns.**

$$C(k|I_m) = \sum_{(i,j)\in\mathbb{G}} V(|\hat{I}_i - \hat{I}_j|)$$



Seek 2D binary sets $\delta \in \{0,1\}$ that make $C(k+\delta|I_m) < C(k|I_m)$.

- Optimization using trained deep denoiser.
  E.g., PnP-UA [2] solves a variational regularization model using a pre-trained denoising network.
  ※ **Effectiveness is constrained due to the domain shift.**

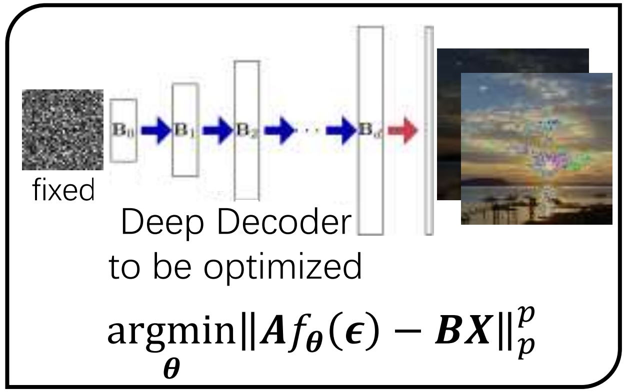
Iterative Optimization

$$\underset{\boldsymbol{Q}}{\operatorname{argmin}} \|\boldsymbol{AQ} - \boldsymbol{BX}\|_p^p + \lambda\mathcal{R}_{\boldsymbol{\theta}}(\boldsymbol{Q})$$

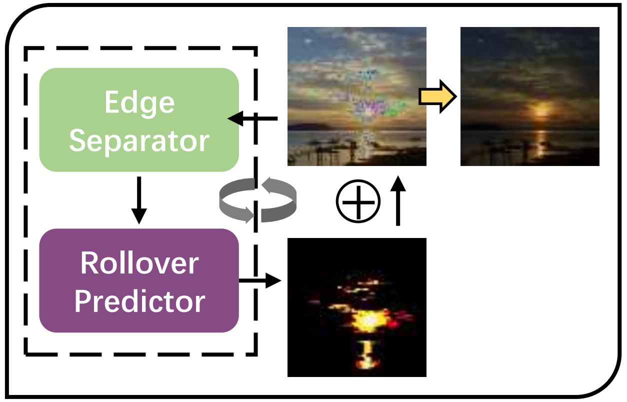
[1] Zhao Hang, Shi Boxin, Fernandez-Cull Christy, et al. Unbounded High Dynamic Range Photography Using a Modulo Camera. ICCP, 2015.
[2] Bacca Jorge, Monroy Brayan, Arguello Henry. Deep Plug-and-Play Algorithm for Unsaturated Imaging. ICASSP, 2024.

Related Works

- Optimization leveraging untrained NN prior.
  E.g., [3] optimizes a TV-regularized variational model using an untrained decoder, leveraging deep image prior.
  ※ **High computational cost and inferior performance.**



- Supervised End-to-end Network.
  E.g., UnModNet [4] formulates modulo inversion as a rollover classification problem.
  ※ **Insufficient generalization as it disregards the inherent properties of the physical imaging process.**



[3] Jagatap Gauri, Hegde Chinmay. High Dynamic Range Imaging Using Deep Image Priors. ICASSP, 2020.
[4] Zhou Chu, Zhao Hang, Han Jin, et al. UnModNet: Learning to Unwrap a Modulo Image for High Dynamic Range Imaging. NeurIPS, 2020.

Contributions

- Reformulation of the HDR reconstruction with modulo cameras in the gradient space, leveraging Itoh's continuity condition.
- The **1st deep unfolding network for HDR reconstruction with modulo cameras**, derived from a variational model using the reformulation, with an auxiliary variable to effectively handle outliers.
- A spiking neuron-based module to learn a sparsity-related prior for outliers, enhancing the performance of HDR reconstruction.
- Superior performance over existing methods on synthetic & real datasets.

Reformulation of Modulo HDR

Proposition 1. *[Extension of 2D Itoh's continuity condition [5]]*
Given an HDR image $\boldsymbol{Y}$ and its modulo counterpart $\boldsymbol{X} = \mathcal{M}(\boldsymbol{Y})$. Let $\widetilde{\mathcal{M}}$ denote another modulo operator defined by $\widetilde{\mathcal{M}}(y) = [(y + a/2) \bmod a] - a/2$. For any point with index (i, j) satisfying $\|\nabla \boldsymbol{Y}_{i,j}\|_\infty = \max\{|\nabla_x \boldsymbol{Y}_{i,j}|, |\nabla_y \boldsymbol{Y}_{i,j}|\} < a/2$, we have

$$\widetilde{\mathcal{M}}(\nabla \boldsymbol{X}_{i,j}) = \nabla \boldsymbol{Y}_{i,j}.$$

This property simplifies the HDR reconstruction by allowing direct inversion of the modulo operation in the gradient domain for pixels within this bound.

[5] Zhile Chen, Yuhui Quan, Hui ji. Deep Unsupervised Unrolling Networks for Phase Unwrapping. CVPR 2024.

Optimization Model

- Introducing an auxiliary variable $\boldsymbol{V}$ to mitigate the impact of outliers, define an optimization problem for modulo HDR reconstruction:

$$\min_{\boldsymbol{Y},\boldsymbol{V}} \left\| \nabla \boldsymbol{Y} - (\widetilde{\mathcal{M}}(\nabla \boldsymbol{X}) - \boldsymbol{V}) \right\|_{\mathrm{F}}^{2} + \Phi(\boldsymbol{Y}) + \Theta(\boldsymbol{V}).$$

sparsity regularization

- Unfold its iterative scheme of sub-problems using proximal gradient descent solver.

In k^{th} iteration,

$\boldsymbol{Q}_k^{(0)} = \boldsymbol{P}_k^{(0)} = \boldsymbol{Y}_{k-1}$, $\beta^{(0)} = 1$.

For $n = 1$ to N,

$$\begin{cases} \boldsymbol{Q}_k^{(n)} = \boldsymbol{P}_k^{(n-1)} + \eta_k \mathcal{G}_{\boldsymbol{P}}[\nabla \boldsymbol{P}_k^{(n-1)} - (\widetilde{\mathcal{M}}(\nabla \boldsymbol{X}) - \boldsymbol{V}_{k-1})], \\ \beta^{(n)} = (1 + \sqrt{1 + 4\beta^{(n-1)} \cdot \beta^{(n-1)}})/2, \\ \boldsymbol{P}_k^{(n)} = \boldsymbol{Q}_k^{(n)} + \dfrac{\beta^{(n-1)} - 1}{\beta^{(n)}} \cdot (\boldsymbol{Q}_k^{(n)} - \boldsymbol{Q}_k^{(n-1)}), \end{cases}$$

#Accelerated gradient descent for better convergence.

$$\boldsymbol{Y}_k = \mathrm{UNet}_{\boldsymbol{Y}}\left(\boldsymbol{Q}_k^{(N)}, \boldsymbol{X}\right), \qquad (1)$$

$$\boldsymbol{E}_k = \widetilde{\mathcal{M}}(\nabla \boldsymbol{X}) - \nabla \boldsymbol{Y}_k,$$

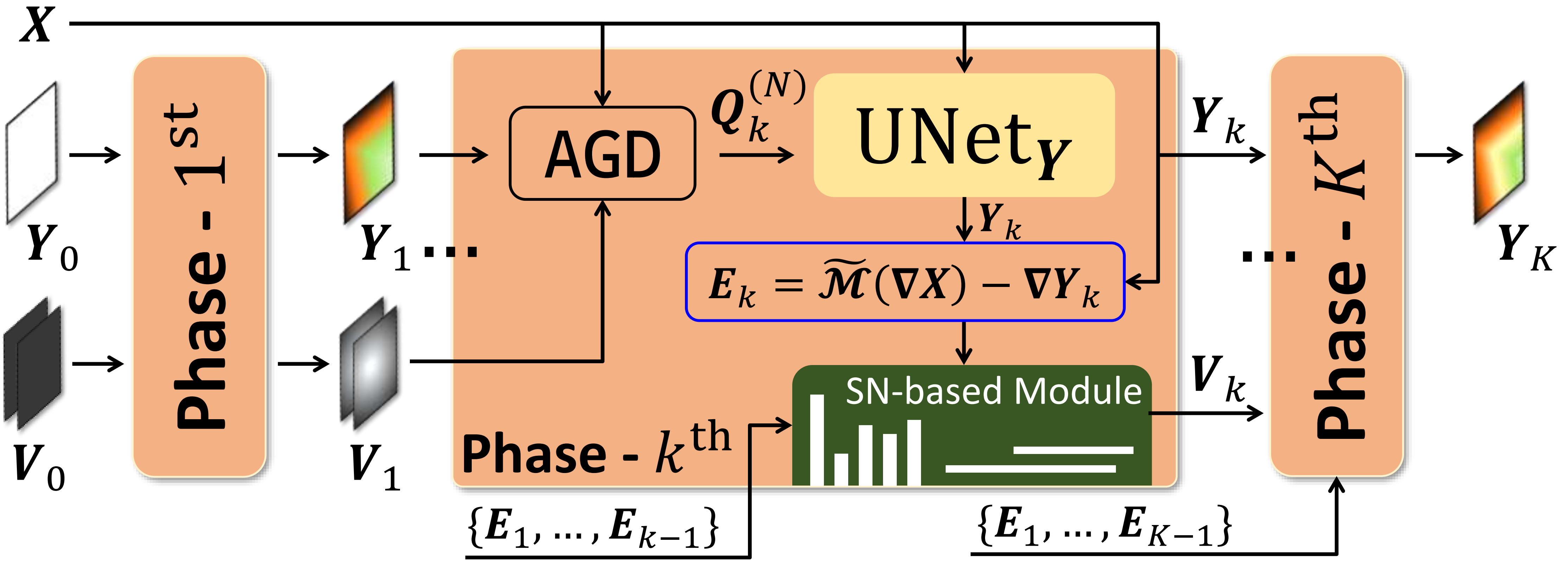
$$\boldsymbol{V}_k = \mathrm{SNM}(\boldsymbol{E}_1, \ldots, \boldsymbol{E}_k); \qquad (2)$$

#(1) Replace the step for $\Phi(\boldsymbol{Y})$ by UNet.

#(2) Replace the step for $\Theta(\boldsymbol{V})$: Learn a spiking neuron-based module (SNM) for capturing the structures of sparse outliers.

Deep Unfolding Network

- Implement each iteration as a phase block in the proposed DUN:



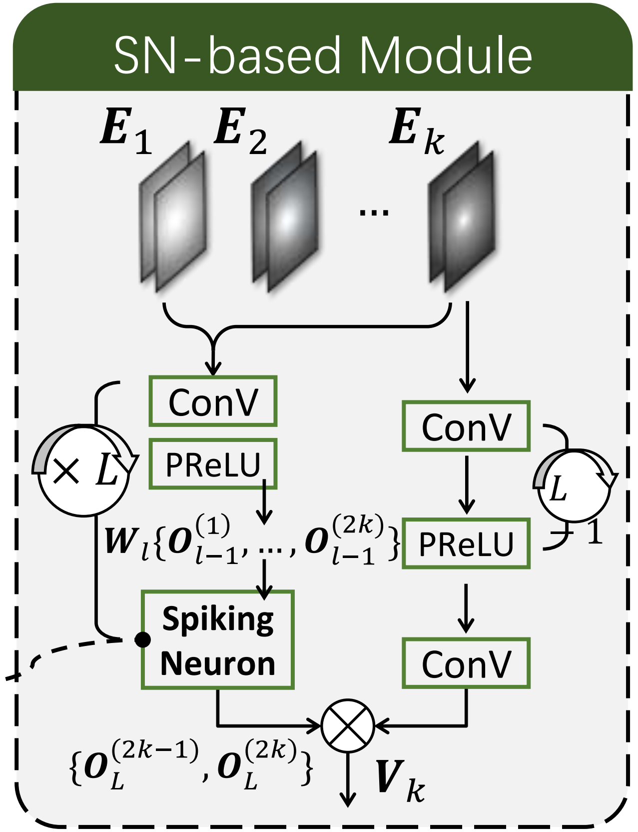
Spiking Neuron-based Module

- Employ the Leaky Integrate-and-Fire (LIF) model [6].

The spiking neuron aggregates outlier information across unfolding phases, enhancing the model's ability to adaptively suppress sparse erroneous measurements and improve reconstruction accuracy.

l^{th} Spiking Neuron

Initialize $\boldsymbol{U}_l^{(0)} = \boldsymbol{O}_l^{(0)} = \mathbf{0}$.
For $t = 1, \ldots, 2k$,
$\boldsymbol{U}_l^{(t)} = \kappa \boldsymbol{U}_l^{(t-1)} \left(1 - \boldsymbol{O}_l^{(t-1)}\right) + \boldsymbol{W}_l \cdot \boldsymbol{O}_{l-1}^{(t)}$,
$\boldsymbol{O}_l^{(t)} = \mathcal{T}\left(\boldsymbol{U}_l^{(t)}, \tau\right)$.
#Thresholding function for spiking.
Output $\{\boldsymbol{O}_l^{(1)}, \ldots, \boldsymbol{O}_l^{(2k)}\}$.



[6] Wang Siqi, Cheng Tee Hiang, Lim Meng-Hiot. LTMD: Learning Improvement of Spiking Neural Networks with Learnable Thresholding Neurons and Moderate Dropout. NeurIPS 2022.

Training Strategy

- The outputs of all phases in the DUN ($k = 1, \dots, K$) are supervised using the overall loss function:

$$\mathcal{L}_{\mathrm{all}} \coloneqq \sum_{k=1}^{K} \gamma_k \left(\mathcal{L}_{\mathrm{g}}(\boldsymbol{Y}_k, \boldsymbol{Y}) + \rho \mathcal{L}_{\mathrm{s}}(\boldsymbol{Y}_k, \boldsymbol{Y}) \right),$$

where $\rho \in \mathbb{R}^+, \gamma_k = 1/(K - k + 1)$, and

$$\mathcal{L}_{\mathrm{g}}(\boldsymbol{Y}_k, \boldsymbol{Y}) \coloneqq \frac{1}{2HWC} \|\boldsymbol{\nabla} \boldsymbol{Y}_k - \boldsymbol{\nabla} \boldsymbol{Y}\|_1,$$

$$\mathcal{L}_{\mathrm{s}}(\boldsymbol{Y}_k, \boldsymbol{Y}) \coloneqq \frac{1}{HWC} \sum_{i,j,c} \log\left(\cosh(\boldsymbol{Y}_k - \boldsymbol{Y})_{i,j,c} \right).$$

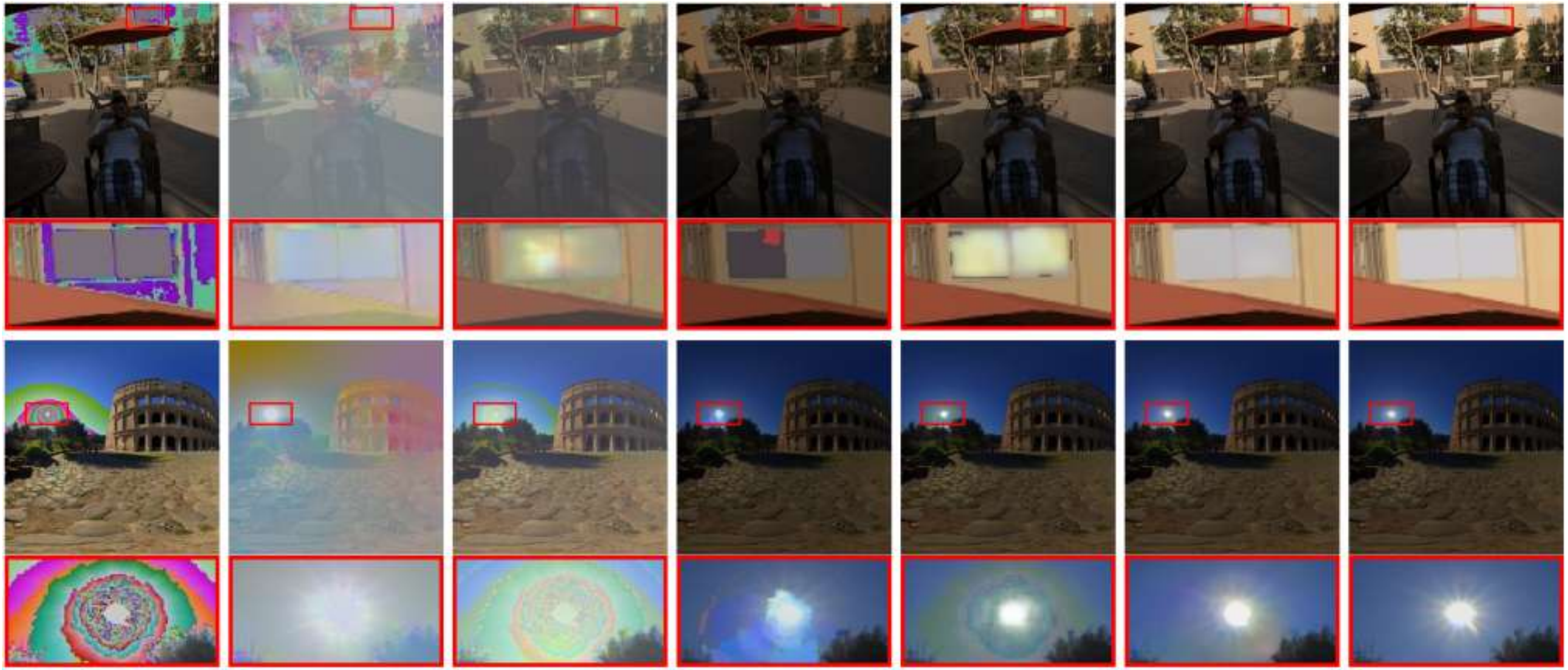
Evaluation on Synthetic Dataset

Boldfaced: best results.

Method	NRMSE(%)	PSNR(dB)	SSIM	MS-SSIM	Param(M)	FLOPs(G)
MRF	45.15	25.08	0.38	0.55	/	/
PnP-UA	28.35	29.38	0.34	0.51	32.64	574.44
UnModNet	6.64	40.42	0.98	0.97	40.83	310.46
ExpandNet†	11.74	21.20	0.55	0.59	0.46	53.80
ExpandNet	10.18	23.57	0.80	0.78	0.46	53.80
Santos *et al.*†	19.52	10.89	0.21	0.32	51.54	75.77
Santos *et al.*	7.85	39.09	0.98	0.97	51.54	75.77
UFormer	7.15	42.34	0.98	0.98	20.60	164.36
Ours	**5.56**	**42.56**	**0.99**	**0.98**	4.15	62.73

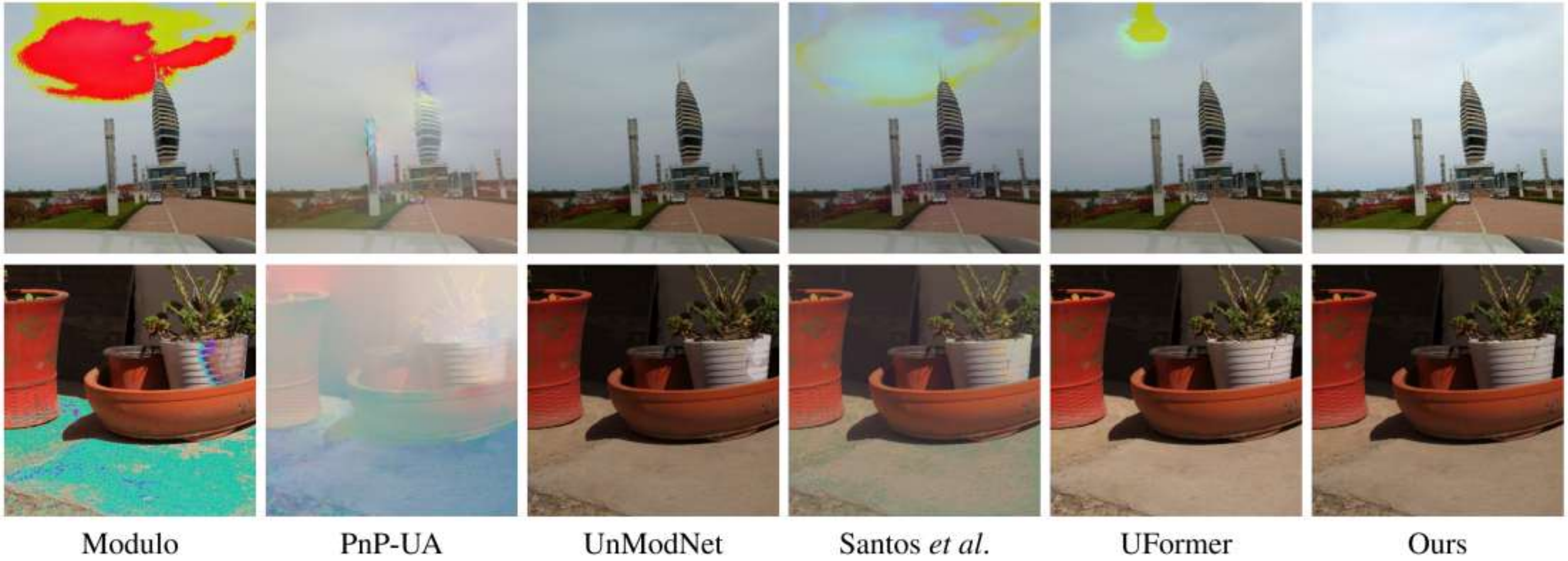
Our DUN achieves the highest performance, while maintaining the second-lowest parameter count and FLOPs.

Visualization on Synthetic Dataset



Our DUN provides the HDR image results with best visual quality.

Evaluation on Real-RGB Dataset



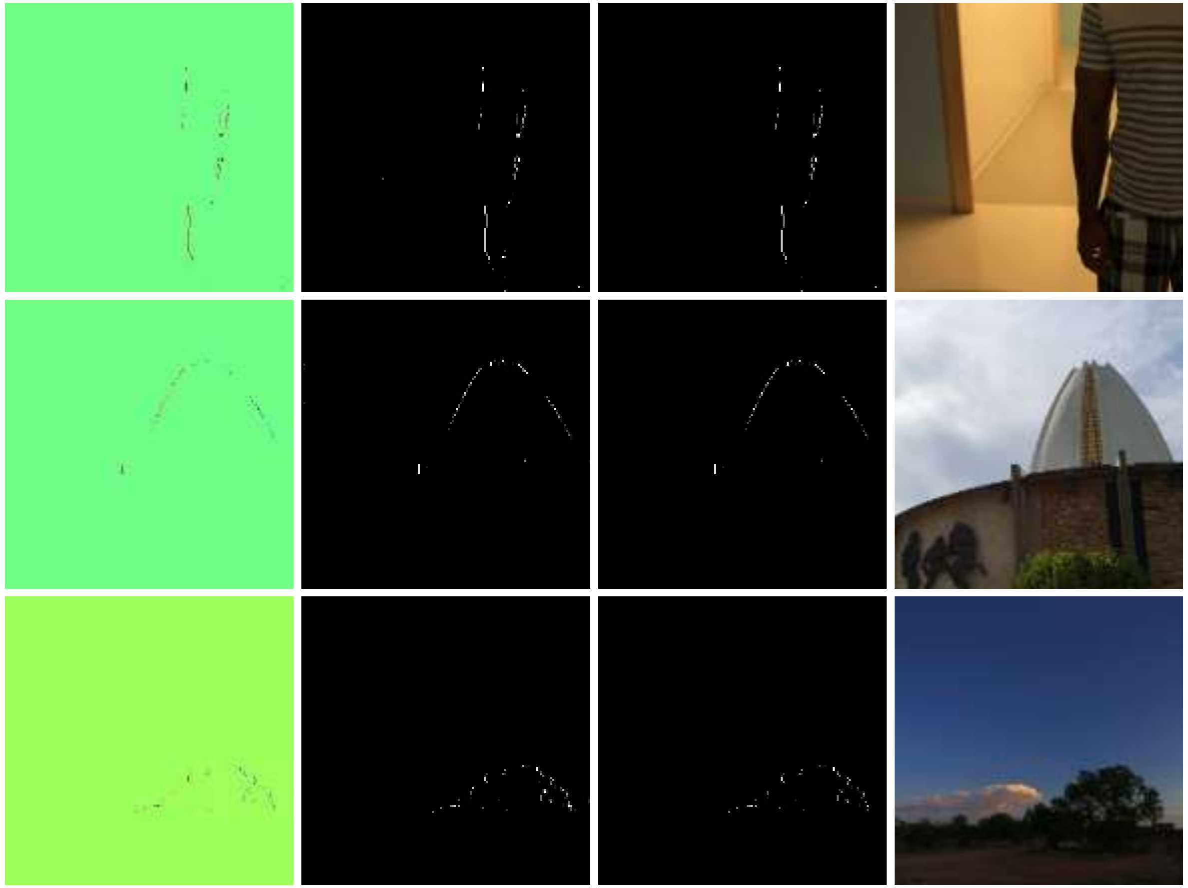
Our DUN yields the most authentic visual results for both samples.

Ablation Study and Analysis

- Main components

Baseline	NRMSE(%)	PSNR(dB)	SSIM	MS-SSIM
w/o $\boldsymbol{V}$	6.56	41.43	0.98	0.98
SN→Th	6.30	41.79	0.98	0.98
AGD→GD	5.97	41.55	0.98	0.98
LogCosh→ℓ_1	5.72	42.28	0.99	0.98
Original	5.56	42.56	0.99	0.98

- Visualization of $\boldsymbol{V}$



Predicted $\boldsymbol{V}$ | Binary map | GT outliers | GT HDR image

Conclusion and Future Work

- To conclude
  - ✓ Reformulating the HDR reconstruction with modulo cameras, simplifying the reconstruction as direct inversion in gradient space for most pixels.
  - ✓ Constructing the first deep unfolding network for modulo HDR, with an auxiliary variable to effectively handle outliers.
  - ✓ Integrating spiking neuron-based modules to learn sparsity-related prior.
- In future
  - ➢ Enhancing the robustness mechanism against outliers.
  - ➢ Extending the DUN for other inverse problems involving modulo operators.

Take home messages

- HDR reconstruction with modulo cameras can be simplified by utilizing the modulo-wrapped gradients of measurement.
- Incorporating physics into NN yields better performance and reduced complexity.
- A spiking neuron-based module is capable of enforcing sparsity regularization with cumulative memory characteristics.

Thank you!