ErA: Error-Aware Deep Unrolling Network for Single Image Defocus Deblurring

Tu Vo **Chan Y. Park**
KC Machine Learning Lab, Seoul, South Korea

Problem Definition and Contributions

Goal: Restore sharp images from spatially varying camera optics blur.
Key Contributions:

- Blind, spatially varying deblurring beyond old assumed fixed-kernels.
- Joint PSF prediction with error-aware regularization in a constrained optimization model.
- Augmented-Lagrangian–based unrolling network achieving SOTA on DPDD, RealDOF, and CUHK.

Problem Formulation

We model defocus deblurring as:

$$\min_{\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}} \frac{1}{2}\|\boldsymbol{\mathcal{H}} \otimes \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}}\|_2^2$$

Method

We introduce a sparse error term and 2 regularization terms:

$$\underset{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}},\mathcal{E}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{\mathcal{H}} \otimes \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}} + \mathcal{E}\|_2^2 + \|\mathcal{E}\|_1 + \phi(\boldsymbol{\mathcal{X}}) + f(\mathcal{E}) \qquad (1)$$

ALM formulation:

$$\begin{aligned} \underset{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}},\mathcal{E},\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{P}},\boldsymbol{\mathcal{Z}}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{\mathcal{U}} - \boldsymbol{\mathcal{Y}} + \mathcal{E}\|_2^2 + \phi(\boldsymbol{\mathcal{Z}}) + \lambda_3\|\mathcal{E}\|_1 + f(\boldsymbol{\mathcal{P}}) \\ \text{subject to} \quad & \boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{H}} \otimes \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{P}} = \mathcal{E}, \boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{X}} \end{aligned} \qquad (2)$$

Optimization Updates

Closed-form updates:

$$\boldsymbol{\mathcal{U}}_{t+1} = \frac{\lambda_1(\boldsymbol{\mathcal{H}} \otimes \boldsymbol{\mathcal{X}}_t) + \boldsymbol{\Gamma}_t + \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{E}}_t}{1+\lambda_1}$$

$$\boldsymbol{\mathcal{X}}_{t+1} = \boldsymbol{\mathcal{F}}^{-1}\left\{\frac{\boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{H}}^T(-\boldsymbol{\Gamma}_t + \lambda_1\boldsymbol{\mathcal{U}}_{t+1}) - \boldsymbol{\Omega}_t + \lambda_2\boldsymbol{\mathcal{Z}}_{t+1})}{\lambda_1\boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{H}})^2 + \lambda_2}\right\}$$

Soft-thresholding:

$$\boldsymbol{\mathcal{E}}_{t+1} = \text{soft-thresh}\left(-\frac{\boldsymbol{\Delta}_t + \boldsymbol{\mathcal{U}}_{t+1} - \boldsymbol{\mathcal{Y}} - \lambda_3\boldsymbol{\mathcal{P}}_t}{1+\lambda_3}\right)$$

CNN-based:

$$\boldsymbol{\mathcal{P}}_{t+1} = \boldsymbol{\mathcal{D}}_f\left(\boldsymbol{\mathcal{P}}_t - \frac{\boldsymbol{\Delta}_t + \lambda_3\boldsymbol{\mathcal{E}}_{t+1}}{\lambda_3}\right)$$

Multipliers:

$$\begin{aligned} \boldsymbol{\Gamma}_{t+1} &= \boldsymbol{\Gamma}_t + \lambda_1(\boldsymbol{\mathcal{H}} \otimes \boldsymbol{\mathcal{X}}_{t+1} - \boldsymbol{\mathcal{U}}_{t+1}) \\ \boldsymbol{\Omega}_{t+1} &= \boldsymbol{\Omega}_t + \lambda_2(\boldsymbol{\mathcal{X}}_{t+1} - \boldsymbol{\mathcal{Z}}_{t+1}) \\ \boldsymbol{\Delta}_{t+1} &= \boldsymbol{\Delta}_t + \lambda_3(\boldsymbol{\mathcal{E}}_{t+1} - \boldsymbol{\mathcal{P}}_{t+1}) \end{aligned}$$

Network Architecture

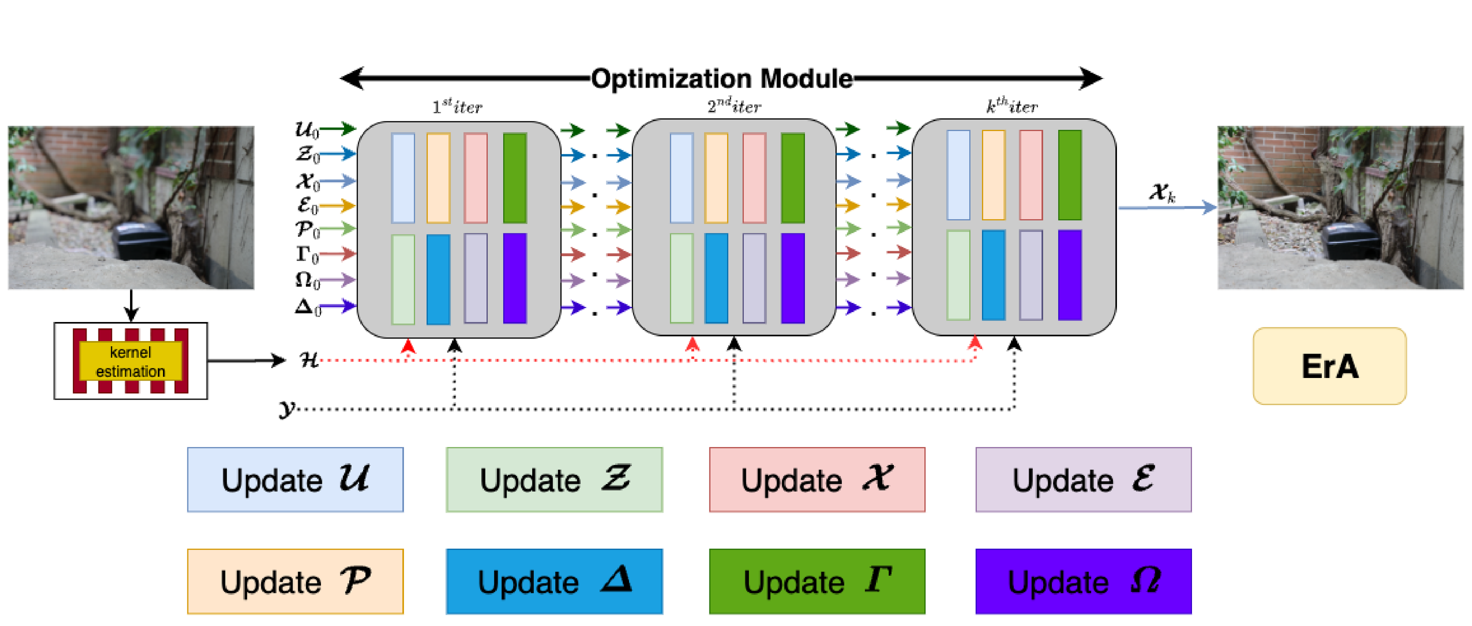

Figure: ErA architecture with closed-form and CNN-based update modules.

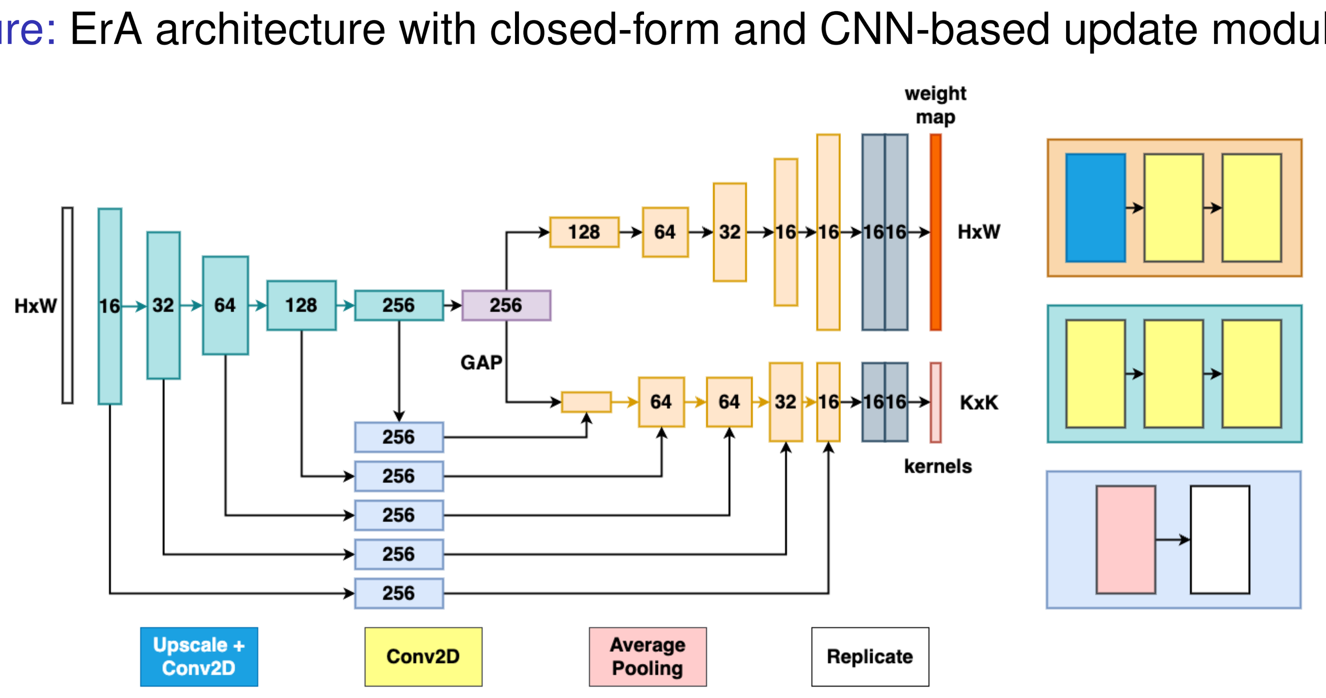

Figure: Kernel Estimation Module predicting global kernel and spatial weight map.

Experiments & Results

Dataset: DPDD 16-bit dual-pixel dataset, 500 samples in 16-bit format, split into 350 training, 74 validation, and 76 test samples.
Loss:

$$L = \omega\|\boldsymbol{\mathcal{X}}_{\text{pred}} - \boldsymbol{\mathcal{X}}_{\text{gt}}\|_1 + (1-\omega)\|\boldsymbol{\mathcal{H}} \otimes \boldsymbol{\mathcal{X}}_{\text{pred}} - \boldsymbol{\mathcal{Y}}\|_1$$

first term ensures the image reconstruction ability, the second term is in-charge of the kernel estimation consitency. ω is set to 0.5.

Quantitative Results

Method / Metric	DPDNet [1]	IFANet [2]	Restormer [3]	INIKNet [4]	NRKNet [5]	P2IKT [6]	IRNexT [7]	ErA
DPDD								
PSNR ↑	24.348	25.366	25.980	26.113	26.110	26.280	26.300	**26.687**
SSIM ↑	0.747	0.789	0.811	0.804	0.810	0.807	0.814	**0.815**
LPIPS ↓	0.277	0.217	**0.178**	0.183	0.223	0.191	0.206	0.219
RealDOF								
PSNR ↑	22.870	23.500	25.091	25.382	25.027	25.480	25.660	**25.747**
SSIM ↑	0.670	0.681	0.762	0.767	0.766	0.762	0.755	**0.772**
LPIPS ↓	0.425	0.444	**0.285**	0.287	0.338	0.306	0.336	0.319
RTF								
PSNR ↑	23.608	24.041	24.212	25.467	**25.929**	25.260	25.333	25.502
SSIM ↑	0.591	0.758	0.821	0.832	0.829	0.819	**0.854**	0.823
LPIPS ↓	0.296	0.289	0.224	0.215	0.218	**0.207**	0.249	0.215

Qualitative Results

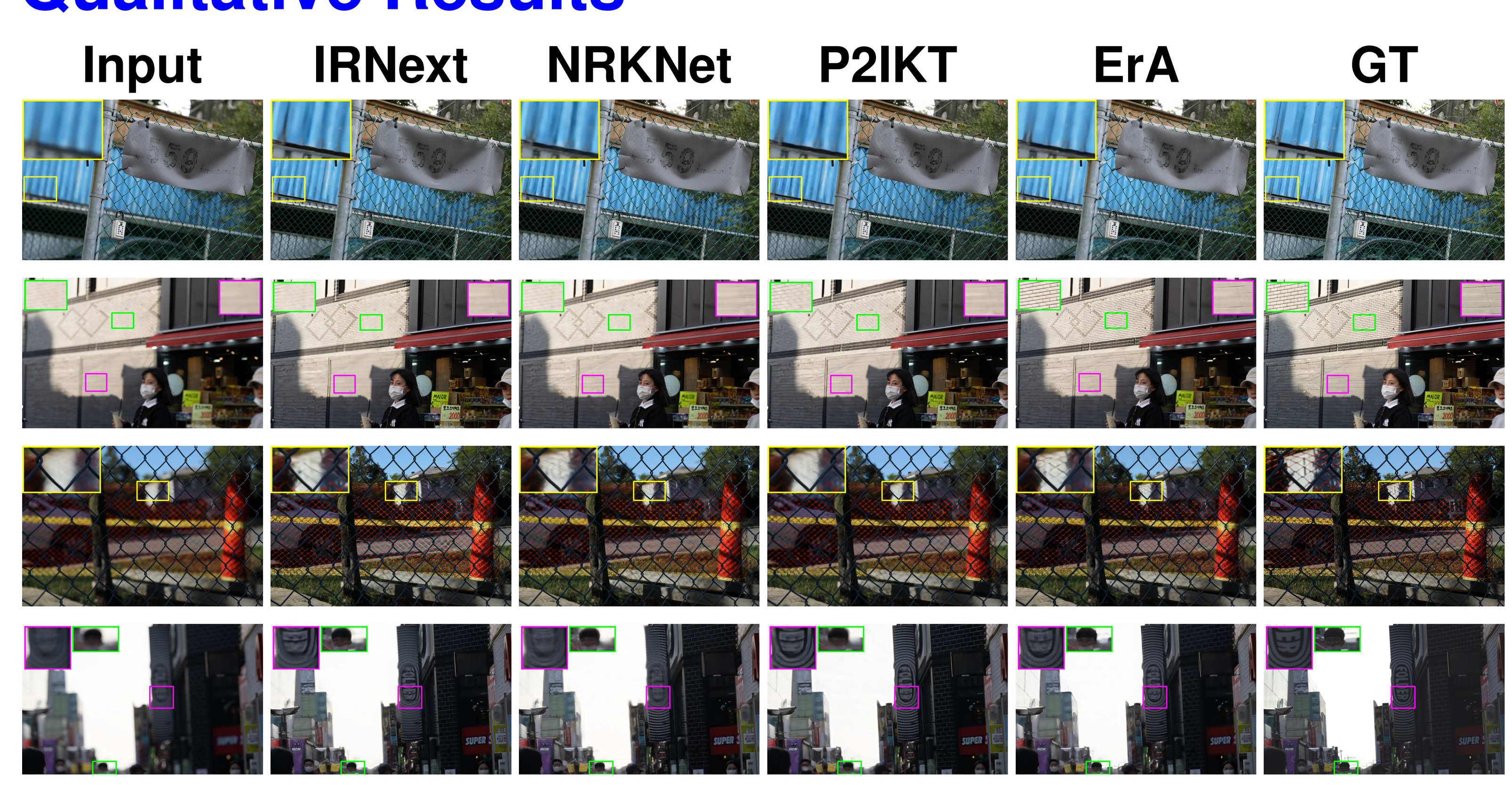

Figure: Comparison on DPDD [1] and RealDOF samples across methods.

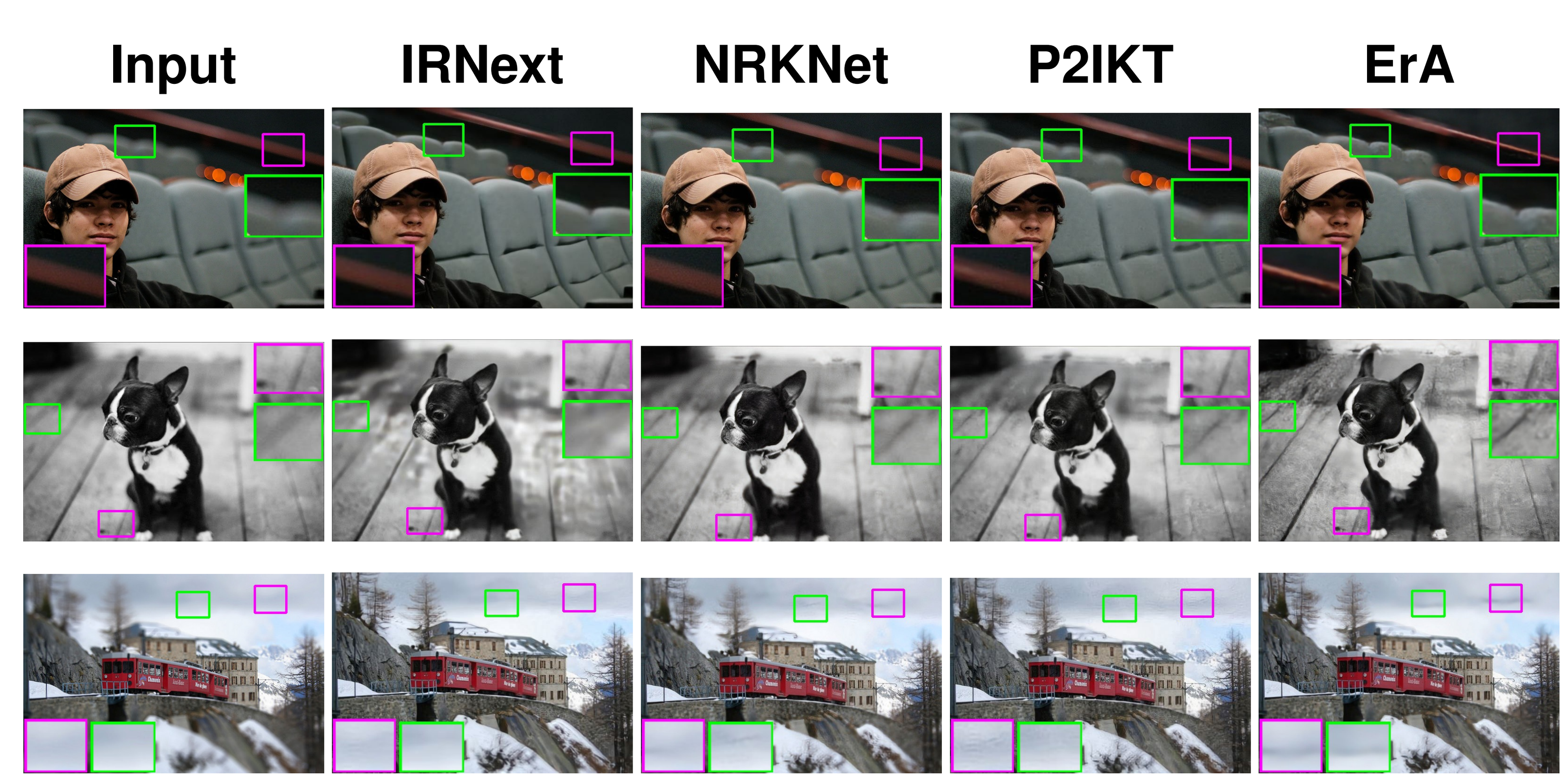

Figure: Comparison of results on the CUHK [8] dataset.

Future Work

- Improve reconstruction via better optimization and physical priors.
- Replace CNN update module with a lighter, faster, memory-efficient version.
- Generalize framework to handle mixed or compound defocus scenarios.

References

[1] Abuolaim et al., DPDD, ECCV 2020
[2] Lee et al., IFAN, CVPR 2021
[3] Zamir et al., Restormer, CVPR 2022
[4] Quan et al., INIKNet, ICCV 2023
[5] Wu et al., NRKNet, CVPR 2023
[6] Tang et al., P2IKT, AAAI 2024
[7] Cui et al., IRNeXt, ICML 2023
[8] Shi et al., CUHK, CVPR 2014