User-Instructed Disparity-aware Defocus Control

Yudong Han[1], Yan Yang[2], Hao Yang[1], Liyuan Pan[1]
[1]Beijing Institute of Technology [2]BDSI, Australian National University

Australian National University

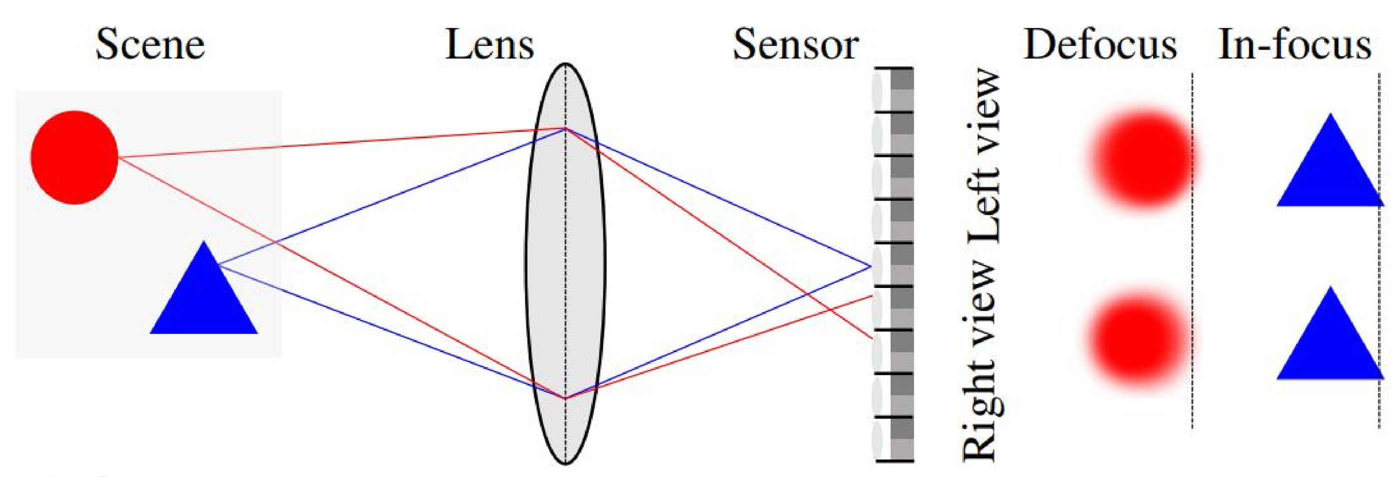
NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

➢ Task Definition of Defocus Control

Users are able to flexibly manipulate sharp and blurred regions in their captured images.

➢ Preliminary: Dual-Pixel (DP) Imagery Model



DP model:

$$\mathbf{D}(i,j) = f \times B/\mathbf{Z}(i,j) + \text{const} \approx f \times B/\mathbf{Z}(i,j) \quad \mathbf{C}(i,j) = \mathcal{J}(f) \cdot \mathbf{D}(i,j)$$

Defocus Control:

$$\mathbf{C}(i,j) = \mathcal{J}(f) \cdot (\mathbf{D}(i,j) - d_{trg}) \quad d_{trg} = \frac{1}{|\Omega|} \sum_{(i',j') \in \Omega : \mathbf{M}(i',j')=1} \mathbf{D}(i',j')$$

Tips: Disparity is directly correlated to the amount of defocus blur. Rays from points of the cone object intersect at the sensor and are in-focused. The resulting regions in the DP pairs exhibit no disparity.

➢ Motivation and Model Design

(i) **Flexibility:** Enable defocus control using simple text prompt.

$$\mathbf{q}_{cls} = \mathcal{F}_m(\hat{\mathbf{I}}, \mathcal{T}, \mathbf{q}_{cls}) \quad \mathbf{M}_{trg} = \mathcal{F}_g(\mathbf{q}_{cls}, \mathbf{F}_{sam})$$

(ii) **Interpretability:** Learns the defocus map and associated blur kernels based on mathematical formulation of the DP model.

How to Control:

$$\mathbf{F}_D = 2 \cdot \text{Tanh}\left(\text{Norm}(\mathbf{F}_G)/\rho\right) - 1 \quad \mathbf{F}_{init} = \mathcal{J}(\mathbf{F}_G) \cdot \mathbf{F}_D$$

$$\mathbf{F}_{coc} = \mathbf{R} \odot \mathbf{F}_G + \mathcal{P}_{feat}\left(\mathcal{P}_{rgb}\left(\mathbf{F}_{init}\right)\right)$$

Kernel Retrieval:

$$\mathbf{K}_{i,j} = \sum_{n=1}^{N} \mathbf{a}_{i,j}(n) \cdot \mathbf{K}_n$$

$$\mathbf{a}_{i,j} = \text{Softmax}\left(\mathcal{P}_k\left(Avg(\{\mathbf{K}_n\}_{n=1}^{N}) \odot \mathcal{P}_{coc}(\mathbf{F}_{coc}(i,j))\right)\right)$$

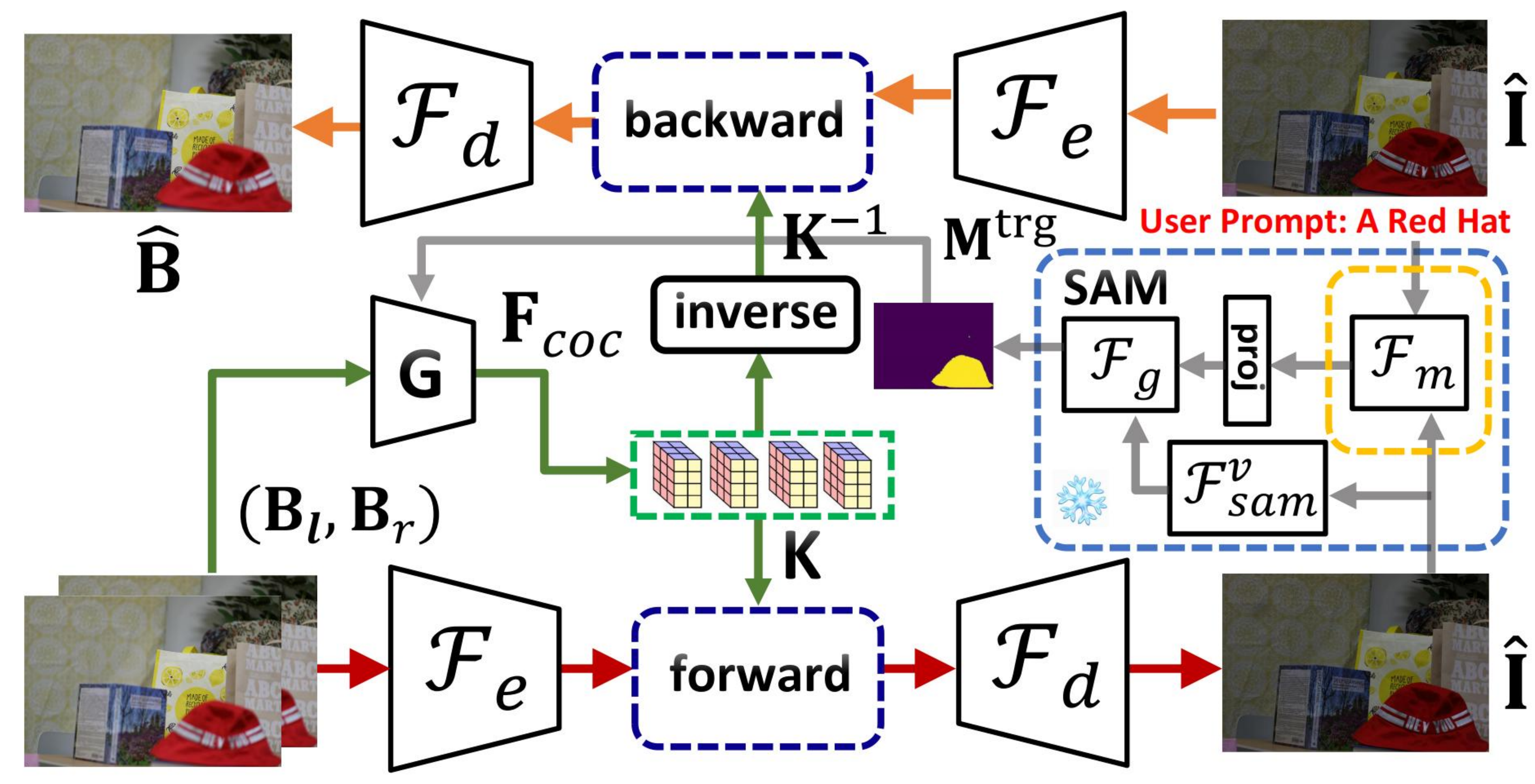
(iii) **Robustness:** An unified invertible deblur-and-reblur framework using the shared kernel.

Forward: $\hat{\mathbf{F}}_I(i,j) = Inv(\sum \mathbf{F}_B(i+\delta i, j+\delta j) \cdot \mathbf{K}_{i,j}(\delta i, \delta j))$

Backward: $\hat{\mathbf{F}}_B(i,j) = \sum_{\delta i, \delta j} Inv'(\mathbf{F}_I)(i+\delta i, j+\delta j) \cdot \mathbf{K}_{i,j}^{-1}(\delta i, \delta j)$

Kernel Inversion: $\mathbf{K}^{-1} = \mathcal{F}^{-1}(\frac{1}{\mathcal{F}(\mathbf{K})})$

Methods



Disparity-aware Kernel Estimation: Disparity information serves as the blur indication to retrieve suitable kernel for defocus control.

Invertible Network: By simple inverse mathematical transformation, we are enable to model the deblurring and reblurring process based on shared kernel.

When Testing:

$$\mathbf{F}_D^{trg}(i,j) = \mathbf{F}_D(i,j) - \frac{1}{|\Omega|} \sum_{(i',j') \in \Omega : \mathbf{M}^{trg}(i',j')=1} \mathbf{F}_D(i',j')$$

Loss Function:

$$\mathcal{L}_{total} = \mathcal{L}_{deb}(\mathbf{I}, \hat{\mathbf{I}}) + \mathcal{L}_{reb}(\mathbf{B}, \hat{\mathbf{B}})$$
$$+ \lambda_{coc} \cdot \mathcal{L}_{coc}(\mathbf{C}, \mathcal{P}_{rgb}(\mathbf{F}_{init})) + \lambda_{grad} \cdot \mathcal{L}_{coc}^{\nabla}(\mathbf{C}, \mathcal{P}_{rgb}(\mathbf{F}_{init}))$$

Experiments

Deblurring Performance on DPD-disp dataset

Method	Computional Cost		Quantitative Metrics			
	Params (M)	Flops (G)	PSNR↑	SSIM↑	$MAE_{(10^{-1})}$ ↓	$MSE\_rel_{(10^{-1})}$ ↓
single-image defocus deblurring						
EBDB	-	-	23.45	0.683	0.49	0.67
DMENet	26.71	4787	23.41	0.714	0.51	0.67
RDPD	24.28	901	25.39	0.772	0.40	0.53
IFAN	10.48	794	25.99	0.804	0.37	0.50
BAMBNet	4.50	1804	26.40	0.821	0.36	0.47
DeepRFT	9.60	3682	25.71	0.801	0.37	0.51
Restormer	26.13	4458	26.66	**0.833**	0.35	0.46
dual-pixel defocus deblurring						
DPDNet	31.03	3150	25.13	0.786	0.41	0.55
DDDNet	6.40	1661	25.36	0.768	0.41	0.54
K3DN	5.00	1033	26.84	0.829	0.35	**0.46**
Ours	**3.12**	**495**	**26.89**	0.829	**0.33**	0.47

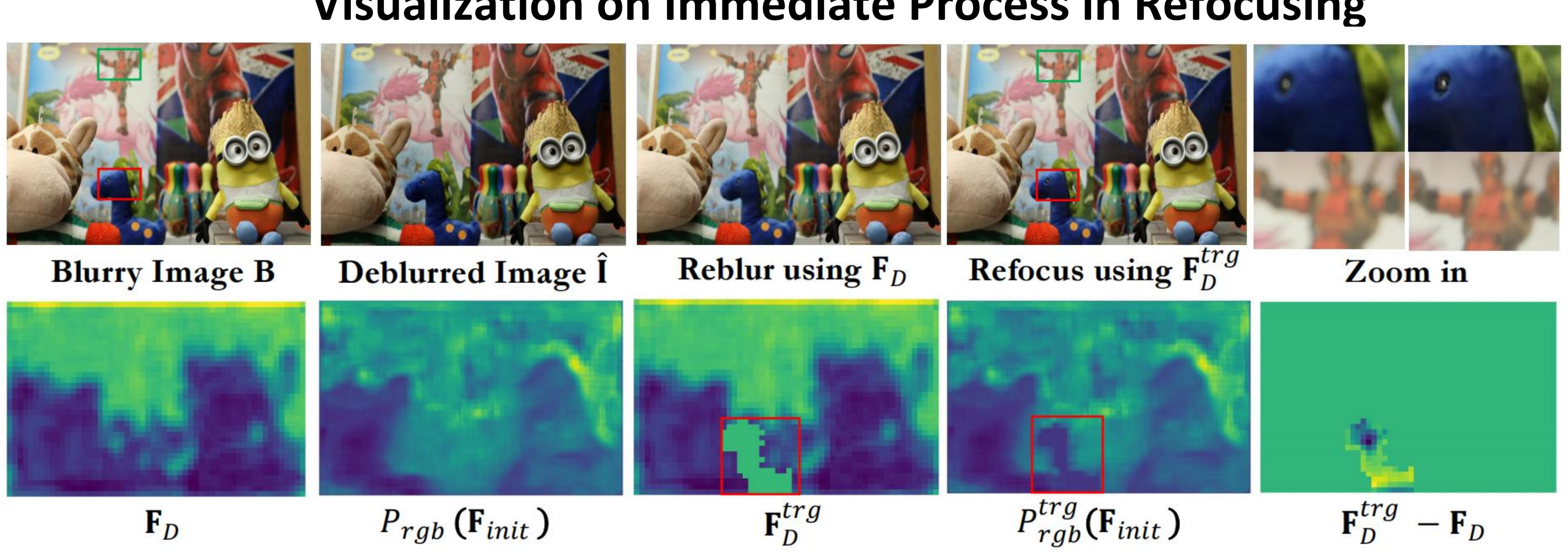
Refocusing Performance on DPD-disp dataset

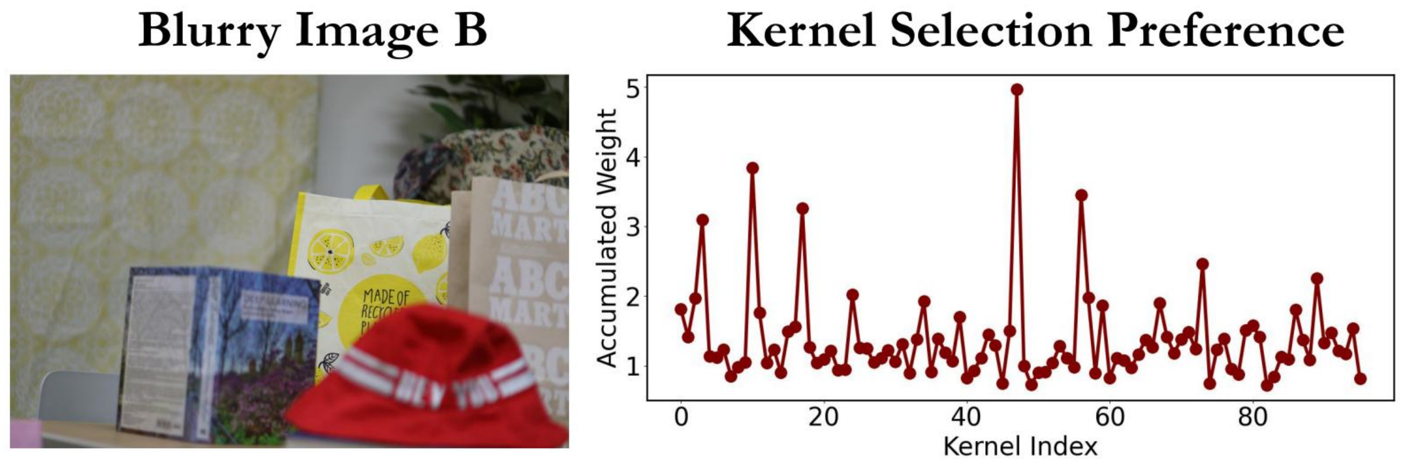
Methods	PSNR↑	SSIM↑	$MAE_{(10^{-1})}$ ↓	LPIPS↓
Omni-Kernel†	18.75	0.692	0.79	31.4
K3DN†	21.72	0.765	0.67	29.7
Ours	21.93	0.772	0.56	25.1

Refocusing Performance on DP5K dataset

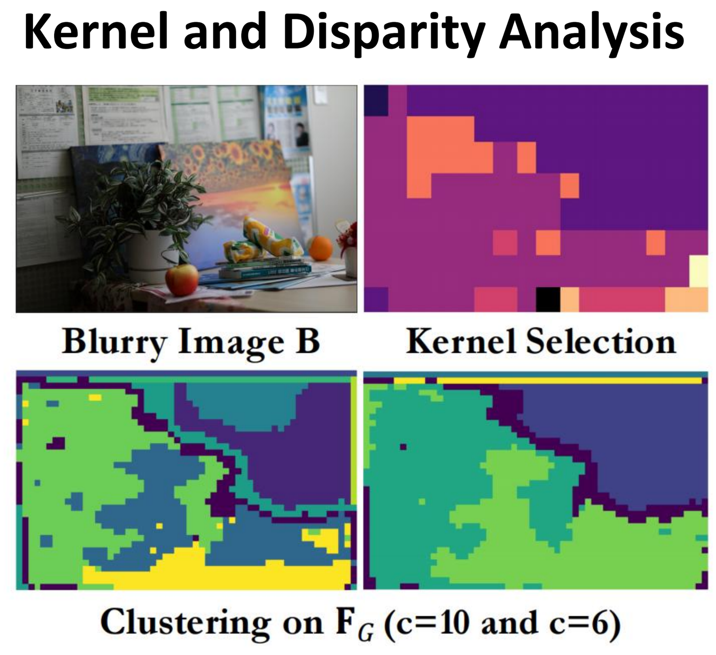
Methods	PSNR↑	SSIM↑	$MAE_{(10^{-1})}$ ↓	LPIPS↓
Omni-Kernel†	25.32	0.816	0.47	23.5
K3DN†	30.76	0.943	0.30	14.4
Ours	31.21	0.953	0.33	13.8

Visualization on Immediate Process in Refocusing





Kernel and Disparity Analysis



Effect of Joint Training

Model Variants	PSNR↑	SSIM↑	$MAE_{(10^{-1})}$ ↓
Deblurring Performance			
w $\mathcal{L}_{reb} + \mathcal{L}_{deb}$	26.89	0.829	0.35
w $\mathcal{L}_{deb}$	26.80	0.823	0.35
Reblurring Performance			
w $\mathcal{L}_{reb} + \mathcal{L}_{deb}$	29.18	0.875	0.22
w $\mathcal{L}_{reb}$	29.06	0.863	0.28

Visualization on self-collected Dataset