Learning Distortion Invariant Representation for Image Restoration from A Causality Perspective

Xin Li[1,*], Bingchen Li[1,*], Xin Jin[2], Cuiling Lan[3,†], Zhibo Chen[1,†]

[1]University of Science and Technology of China [2]Eastern Institute for Advanced Study

[3]Microsoft Research Asia

Poster Tag: TUE-AM-163

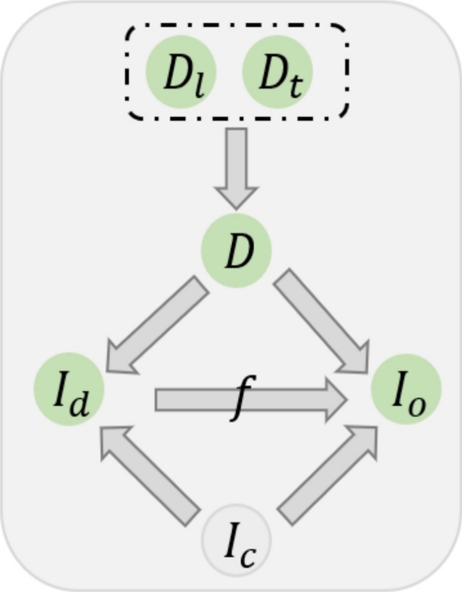
Learning Distortion Invariant Feature Representation for Image Restoration From A Causality Perspective

Microsoft Research 微软亚洲研究院

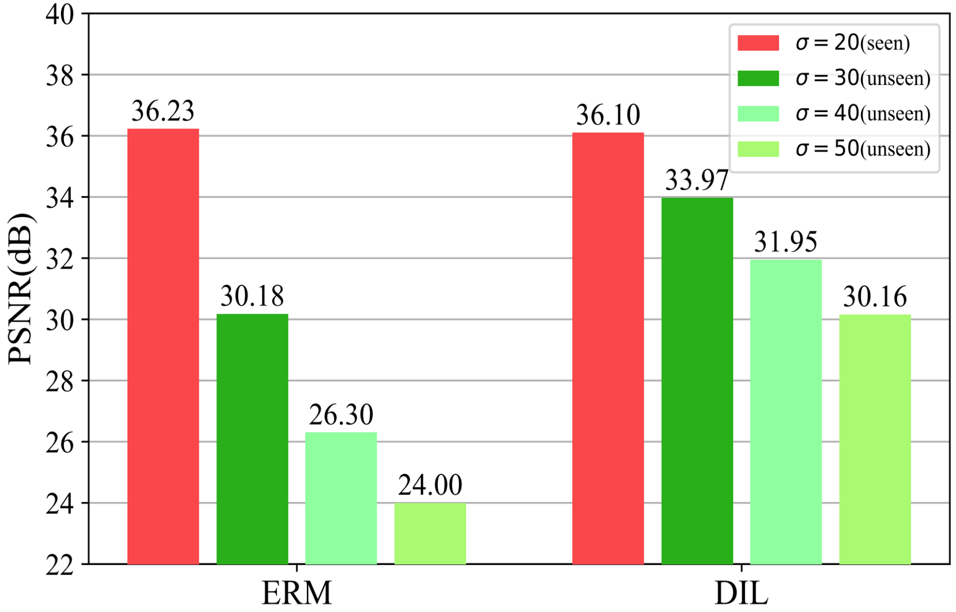
东方理工高等研究院 EASTERN INSTITUTE FOR ADVANCED STUDY

Motivation

- Existing works on image restoration cannot generalize well to real-world degradations with different degrees and types.
- A spurious correlation between I_d and I_o is captured by ERM, which introduce the bad confounding effects of specific distortion d and destroy the generalization capability of IR network.



D Distortion Set
D_t Distortion Types
D_l Distortion Degrees
I_d Distorted Images
I_o Ideal Reconstructed Images
I_c Clean Images
f Restoration Network
d Distortion sampled from D

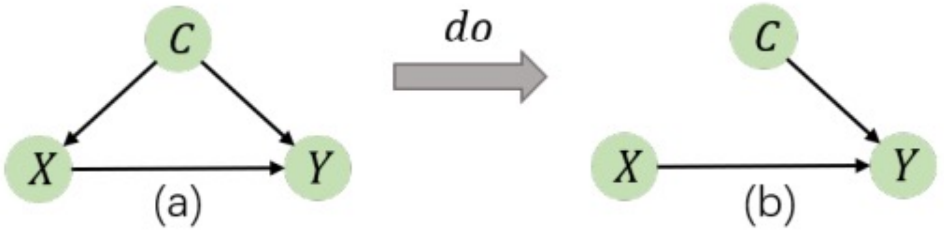


Contributions

- We revisit the image restoration task from a causality view and pinpoint that the reason for the poor generalization of the restoration network, is that the restoration network is not independent to the distortions in the training dataset.
- Based on the back-door criterion in causality, we propose a novel training paradigm, Distortion Invariant representation Learning (DIL) for image restoration, where the intervention is instantiated by a virtually model updating under the counterfactual distortion augmentation and is eliminated with the optimization based on meta-learning.

Methods

- Backdoor criterion in causal inference



$$P(Y|do(X)) = \sum_c P(Y|X, C=c)P(C=c)$$

- Image restoration with back-door criterion

$$P(I_o|do(I_d)) = \sum_{i=1}^{n} P(I_o|I_d, D=d_i)P(D=d_i)$$

- Modeling $\boldsymbol{P(I_o|I_d, D=d_i)}$ in optimization process

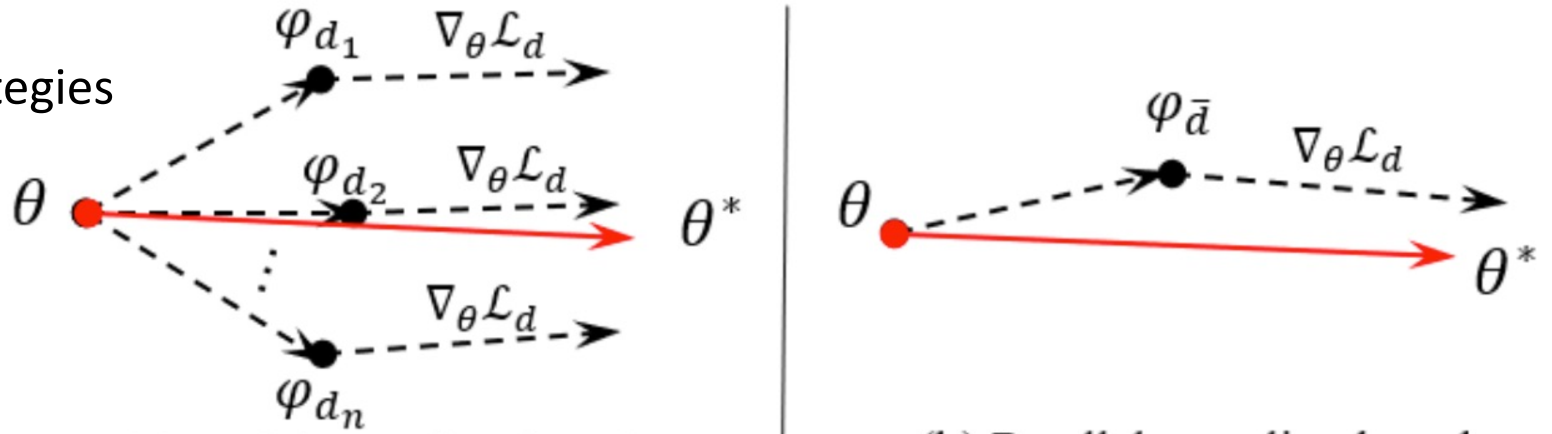
$$\phi_{d_i} = \theta - \alpha\nabla_\theta \mathcal{L}(f_\theta(I_{d_i}), I_c) \Rightarrow \mathcal{L}(f_{\phi_{d_i}}(I_d), I_c)$$

- Modeling $\boldsymbol{P(I_o|do(I_d))}$ with meta learning

$$\theta^* = \arg\min_\theta \mathbb{E}_{(I_d, I_c)\sim\mathcal{D}}\Big[\sum_{d_i \in D} \mathcal{L}(f_{\phi_{d_i}}(I_d), I_c)\Big]$$

- Four Variants:
  - $\boldsymbol{DIL_{pf}}$: parallel sampling
    first-order gradient
  - $\boldsymbol{DIL_{ps}}$: parallel sampling
    second-order gradient
  - $\boldsymbol{DIL_{sf}}$: ***serial*** sampling
    first-order gradient
  - $\boldsymbol{DIL_{ss}}$: ***serial*** sampling
    second-order gradient

- Two sampling strategies



(a) Serial sampling based

(b) Parallel sampling based

$$
\begin{aligned}
\theta^* = \arg\min_{\theta} \mathbb{E}_{(I_d, I_c)\sim\mathcal{D}}\{\frac{1}{n}\sum_{d_i\in D}[\mathcal{L}(f_\theta(I_d), I_c) \\
-\alpha\nabla_\theta\mathcal{L}(f_\theta(I_{d_i}), I_c)\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c) \\
+ o(\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c))]\} \\
= \arg\min_{\theta} \mathbb{E}_{(I_d, I_c)\sim\mathcal{D}}[\\
\mathcal{L}(f_{(\theta-\alpha\nabla_\theta\sum_{d_i\in D}\frac{1}{n}\mathcal{L}(f_\theta(I_{d_i}), I_c))}(I_d), I_c)]
\end{aligned}
$$

$$
\theta^* = \arg\min_{\theta} \mathbb{E}_{(I_d, I_c)\sim\mathcal{D}}[\mathcal{L}(f_{\phi_{\bar{d}}}(I_d), I_c)],
$$

$$
where\ \phi_{\bar{d}} = \theta - \alpha\nabla_\theta\sum_{d_i\in D}\frac{1}{n}\mathcal{L}(f_\theta(I_{d_i}), I_c)
$$

Experiments

- Image Denoising

Datasets	Levels	Methods				
		ERM	DIL_{sf}	DIL_{pf}	DIL_{ss}	DIL_{ps}
CBSD68 [40]	30 (*unseen*)	24.90/0.581	**30.29**$_{(5.39\uparrow)}$**/0.866**	29.92$_{(5.02\uparrow)}$/0.858	27.48$_{(2.58\uparrow)}$/0.809	29.14$_{(4.24\uparrow)}$/0.802
	40 (*unseen*)	21.12/0.400	**28.35**$_{(7.23\uparrow)}$**/0.825**	28.10$_{(6.98\uparrow)}$/0.812	25.90$_{(4.78\uparrow)}$/0.746	25.74$_{(4.62\uparrow)}$/0.629
	50 (*unseen*)	18.96/0.307	**26.64**$_{(7.68\uparrow)}$**/0.779**	26.61$_{(7.65\uparrow)}$/0.766	24.63$_{(5.67\uparrow)}$/0.686	23.34$_{(4.38\uparrow)}$/0.501
Kodak24 [16]	30 (*unseen*)	25.12/0.533	**31.39**$_{(6.27\uparrow)}$**/0.867**	30.87$_{(5.75\uparrow)}$/0.858	27.92$_{(2.80\uparrow)}$/0.801	29.86$_{(4.74\uparrow)}$/0.782
	40 (*unseen*)	21.22/0.352	**29.49**$_{(8.27\uparrow)}$**/0.831**	29.15$_{(7.93\uparrow)}$/0.817	26.46$_{(5.24\uparrow)}$/0.738	26.13$_{(4.91\uparrow)}$/0.588
	50 (*unseen*)	19.02/0.263	**27.76**$_{(8.74\uparrow)}$**/0.788**	27.67$_{(8.65\uparrow)}$/0.775	25.24$_{(6.22\uparrow)}$/0.677	23.60$_{(4.58\uparrow)}$/0.457
McMaster [75]	30 (*unseen*)	25.65/0.569	**31.70**$_{(6.05\uparrow)}$**/0.873**	31.04$_{(5.39\uparrow)}$/0.853	28.15$_{(2.50\uparrow)}$/0.794	30.09$_{(4.44\uparrow)}$/0.800
	40 (*unseen*)	21.73/0.373	**29.81**$_{(8.08\uparrow)}$**/0.831**	29.07$_{(7.34\uparrow)}$/0.802	26.59$_{(4.86\uparrow)}$/0.728	26.24$_{(4.51\uparrow)}$/0.605
	50 (*unseen*)	19.47/0.278	**28.02**$_{(8.55\uparrow)}$**/0.783**	27.31$_{(7.84\uparrow)}$/0.749	25.20$_{(5.73\uparrow)}$/0.664	23.60$_{(4.13\uparrow)}$/0.466
Urban100 [22]	30 (*unseen*)	25.46/0.648	**30.93**$_{(5.47\uparrow)}$**/0.898**	30.26$_{(4.80\uparrow)}$/0.884	26.95$_{(1.49\uparrow)}$/0.825	29.73$_{(4.27\uparrow)}$/0.841
	40 (*unseen*)	21.53/0.479	**28.82**$_{(7.29\uparrow)}$**/0.866**	28.32$_{(6.79\uparrow)}$/0.848	25.26$_{(3.73\uparrow)}$/0.767	26.25$_{(4.72\uparrow)}$/0.691
	50 (*unseen*)	19.28/0.389	**26.88**$_{(7.60\uparrow)}$**/0.829**	26.63$_{(7.35\uparrow)}$/0.811	23.85$_{(4.57\uparrow)}$/0.710	23.71$_{(4.43\uparrow)}$/0.575
Manga109 [41]	30 (*unseen*)	26.62/0.653	**31.97**$_{(5.35\uparrow)}$**/0.910**	31.14$_{(4.52\uparrow)}$/0.901	26.02$_{(-0.6\uparrow)}$/0.833	31.05$_{(4.43\uparrow)}$/0.858
	40 (*unseen*)	22.34/0.442	**29.02**$_{(6.68\uparrow)}$**/0.888**	28.53$_{(6.19\uparrow)}$/0.875	24.31$_{(1.97\uparrow)}$/0.784	27.29$_{(4.95\uparrow)}$/0.704
	50 (*unseen*)	19.95/0.342	**26.52**$_{(6.57\uparrow)}$**/0.860**	26.34$_{(6.39\uparrow)}$/0.846	22.82$_{(2.87\uparrow)}$/0.734	24.47$_{(4.52\uparrow)}$/0.564

Experiments

- Image Deblurring

Datasets	Methods	Levels				
		4.2 (*unseen*)	4.4 (*unseen*)	4.6 (*unseen*)	4.8 (*unseen*)	5.0 (*unseen*)
Set5 [4]	ERM	29.31/0.844	26.55/0.776	24.43/0.709	22.96/0.648	22.00/0.602
	DIL	29.58$_{(0.27\uparrow)}$/0.848	27.52$_{(0.97\uparrow)}$/0.802	25.66$_{(1.23\uparrow)}$/0.751	24.38$_{(1.42\uparrow)}$/0.708	23.46$_{(1.46\uparrow)}$/0.671
Set14 [71]	ERM	27.22/0.781	24.93/0.726	23.16/0.671	21.89/0.624	20.88/0.583
	DIL	27.24$_{(0.02\uparrow)}$/0.778	25.78$_{(0.85\uparrow)}$/0.746	24.35$_{(1.19\uparrow)}$/0.708	23.23$_{(1.34\uparrow)}$/0.672	22.37$_{(1.49\uparrow)}$/0.640
BSD100 [40]	ERM	27.20/0.784	25.17/0.732	23.50/0.682	22.24/0.639	21.28/0.602
	DIL	27.37$_{(0.17\uparrow)}$/0.781	26.16$_{(0.99\uparrow)}$/0.753	24.91$_{(1.41\uparrow)}$/0.719	23.86$_{(1.62\uparrow)}$/0.686	23.02$_{(1.74\uparrow)}$/0.658
Urban100 [22]	ERM	24.95/0.797	22.41/0.723	20.59/0.657	19.33/0.606	18.40/0.565
	DIL	24.97$_{(0.02\uparrow)}$/0.793	23.26$_{(0.85\uparrow)}$/0.743	21.76$_{(1.17\uparrow)}$/0.693	20.70$_{(1.37\uparrow)}$/0.651	19.92$_{(1.52\uparrow)}$/0.618
Manga109 [41]	ERM	28.16/0.865	23.96/0.791	21.21/0.713	19.63/0.652	18.63/0.606
	DIL	28.09$_{(-0.07\uparrow)}$/0.867	25.41$_{(1.45\uparrow)}$/0.822	23.15$_{(1.94\uparrow)}$/0.771	21.69$_{(2.06\uparrow)}$/0.726	20.72$_{(2.09\uparrow)}$/0.691

Experiments

- Hybrid-distorted Image Restoration

Datasets	Methods	Distortion level		
		Mild (unseen)	Moderate (unseen)	Severe (seen)
BSD100 [40]	ERM	25.31/0.687	24.62/0.642	25.27/0.617
	DIL	26.37/0.691	25.23/0.645	25.22/0.613
Urban100 [22]	ERM	23.97/0.736	22.51/0.674	23.38/0.655
	DIL	25.00/0.747	23.13/0.682	23.20/0.645
Manga109 [41]	ERM	27.43/0.863	24.85/0.808	26.50/0.815
	DIL	28.41/0.868	25.30/0.810	26.19/0.766
DIV2K [2]	ERM	26.19/0.766	25.94/0.744	27.42/0.742
	DIL	27.84/0.785	26.89/0.756	27.38/0.737

- RealSR

Methods	Datasets	
	RealSR V3 [5] (unseen)	DrealSR [61] (unseen)
Real-ESRNet [57]	26.19/0.7989	28.22/0.8470
BSRNet [74]	27.46/0.8082	29.45/0.8579
ERM	27.65/0.8098	29.73/0.8628
DIL_{sf}	$27.94_{(0.29\uparrow)}$/0.8098	$29.99_{(0.26\uparrow)}$/0.8648
DIL_{ps}	$28.12_{(0.47\uparrow)}$/0.8067	$30.58_{(0.85\uparrow)}$/0.8712

- Cross different backbones

Models	Methods	Datasets		
		CBSD68 [40]	Kodak24 [16]	Urban100 [22]
RRDB	ERM	24.90/0.581	25.12/0.533	25.46/0.648
	DIL	30.28/0.866	31.39/0.867	30.93/0.898
SwinIR	ERM	24.22/0.551	24.22/0.493	24.73/0.618
	DIL	29.08/0.798	29.71/0.774	29.72/0.834

Experiments

- Real Image Denoising & Image Deraining

Methods	Datasets (Real Denoising)		Datasets (Deraining)		
	SIDD [1]	DND [47]	Rain100L [64]	Rain12 [33]	Rain800 [73]
ERM	38.90/0.9379	38.67/0.9549	27.61/0.8577	31.44/0.8947	23.36/0.8199
DIL_{sf}	39.96 (1.06↑) /0.9410	39.16 (0.49↑) /0.9531	28.15 (0.54↑) /0.8679	32.43 (0.99↑) /0.9163	23.41 (0.05↑) /0.8261
DIL_{ps}	39.92 (1.02↑) /0.9385	39.03 (0.36↑) /0.9553	28.37 (0.76↑) /0.8739	33.07 (1.63↑) /0.9266	23.52 (0.16↑) /0.8281

Learning Distortion Invariant Feature Representation for Image Restoration from A Causality Perspective

Subjective Comparison

- Image Denoising

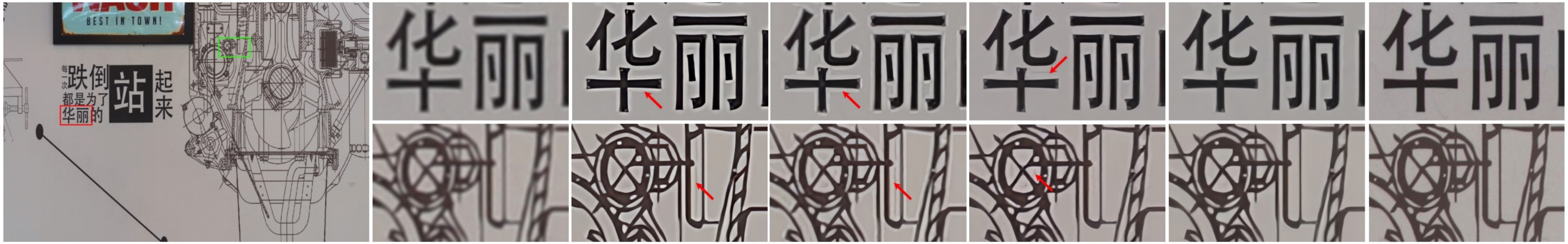
Noisy | **ERM** | **DIL**$_{ps}$ | **DIL**$_{ss}$ | **DIL**$_{pf}$ | **DIL**$_{sf}$

Subjective Comparison

- Hybrid-distorted Image Restoration

Input **Hybrid distortion** **ERM** **DIL** **Ground Truth**

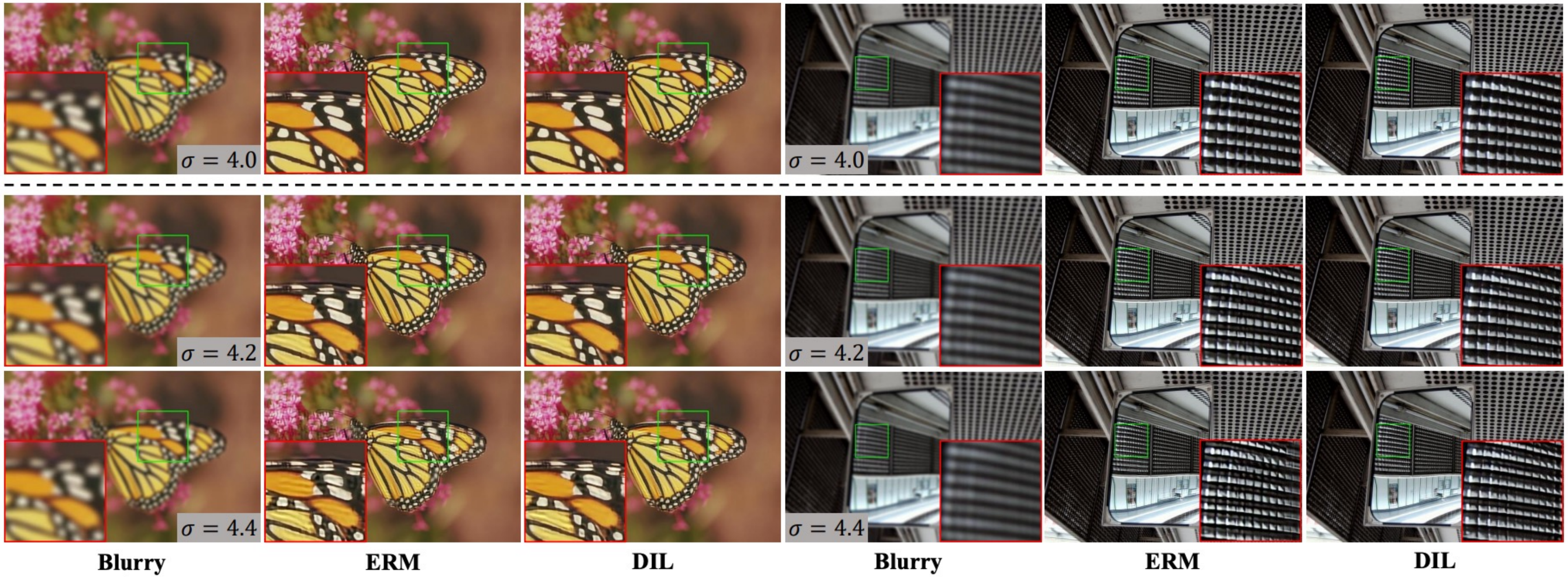
- RealSR



Input **Bicubic** **Real-ESRNet** **BSRNet** **ERM** **DIL$_{ps}$** **Ground Truth**

Subjective Comparison

➢ Image Deblurring



THANK YOU