JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

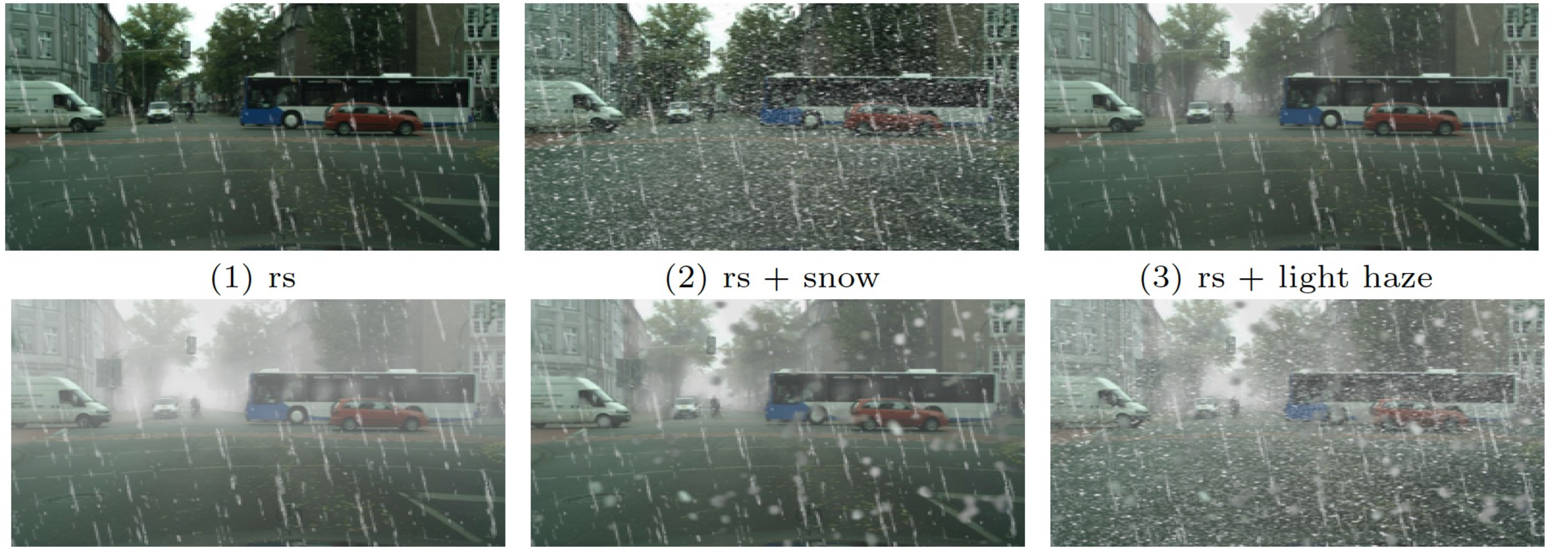
Context-aware Pretraining for Efficient Blind Image Decomposition

Chao Wang[1,2] Zhedong Zheng[3] Ruijie Quan[1] Yifan Sun[2] Yi Yang[1]

[1]ReLER, CCAI, Zhejiang University [2]Baidu Inc.
[3]Sea-NExT Joint Lab, School of Computing, National University of Singapore

Poster Session THU-AM-163

Blind Image Decomposition (BID)



(1) rs (2) rs + snow (3) rs + light haze

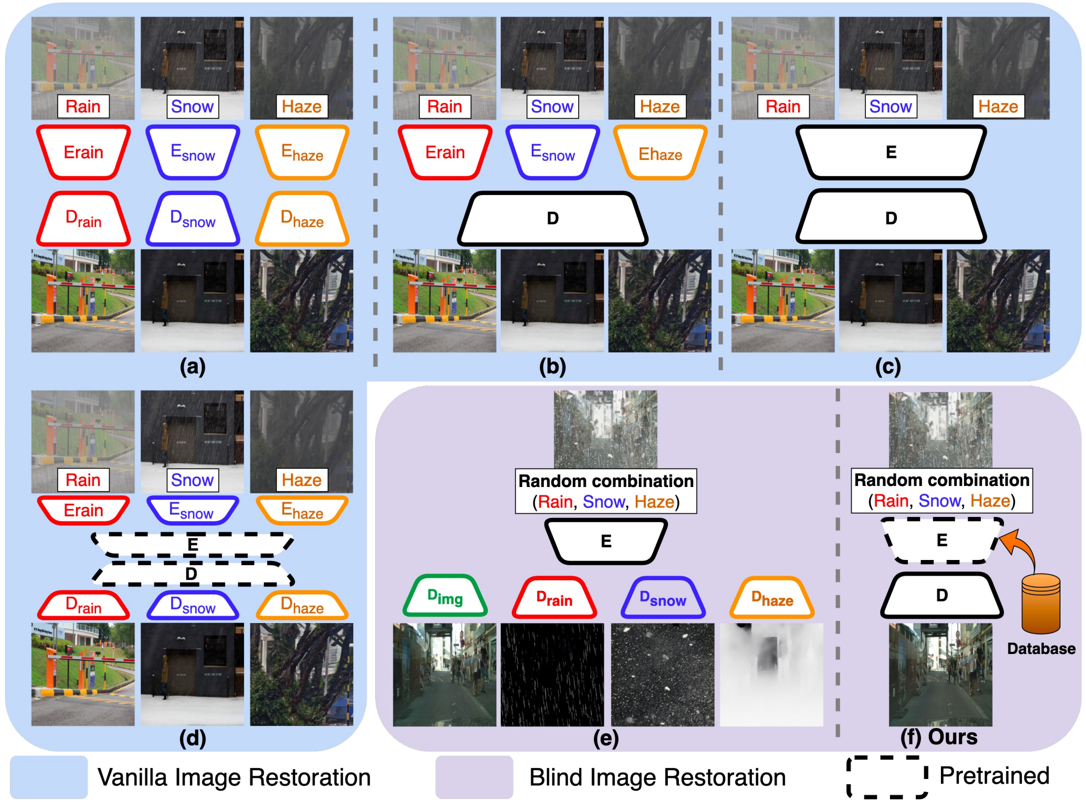
(4) rs + heavy haze (5) rs + mh + raindrop (6) rs + snow + mh + raindrop

rs: rain streak, mh: moderate haze

Separating a superimposed image into constituent underlying images in a blind setting, that is, both the source components involved in mixing as well as the mixing mechanism are unknown.

Motivation

- Existing methods typically require massive data supervision, making them infeasible to real-world scenarios.
- The conventional paradigm usually focuses on mining the abnormal pattern of a superimposed image to separate the noise, which de facto conflicts with the primary image restoration task.
- Pretraining model on ImageNet can efficiently adapt to the high-level representative vision benchmarks such as recognition and detection, yet the pretraining on MAE in low-level vision tasks is still under-explored.

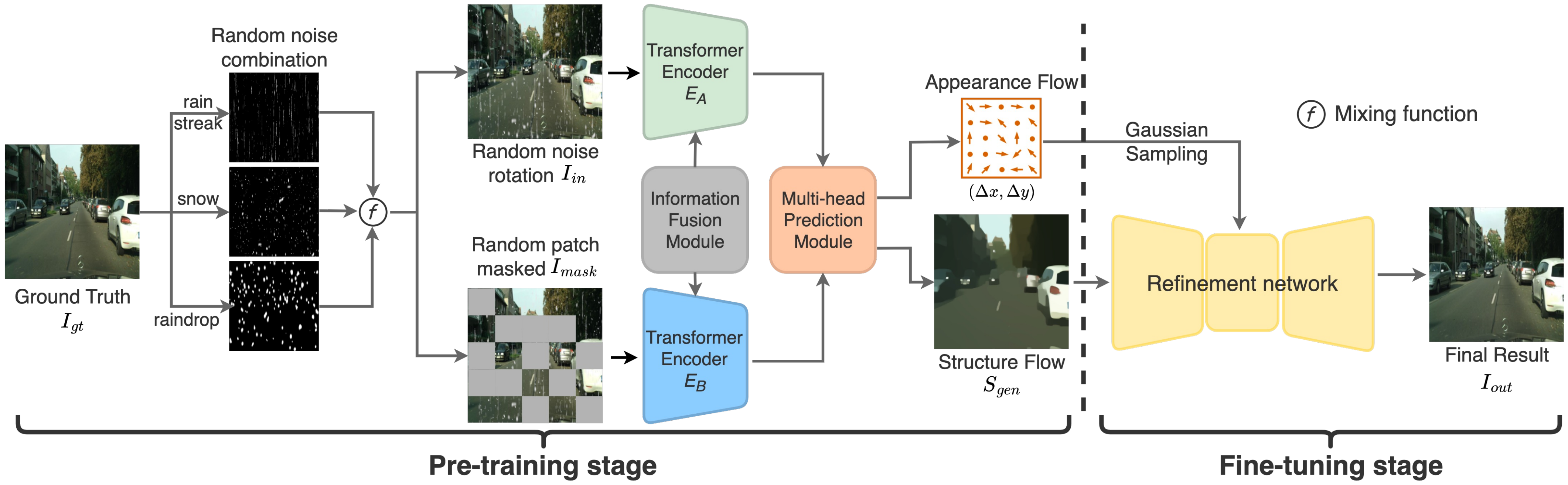


Contribution

- We introduce a new self-supervised learning paradigm, called Context-aware Pretraining with two pretext tasks: mixed image separation and masked image reconstruction.

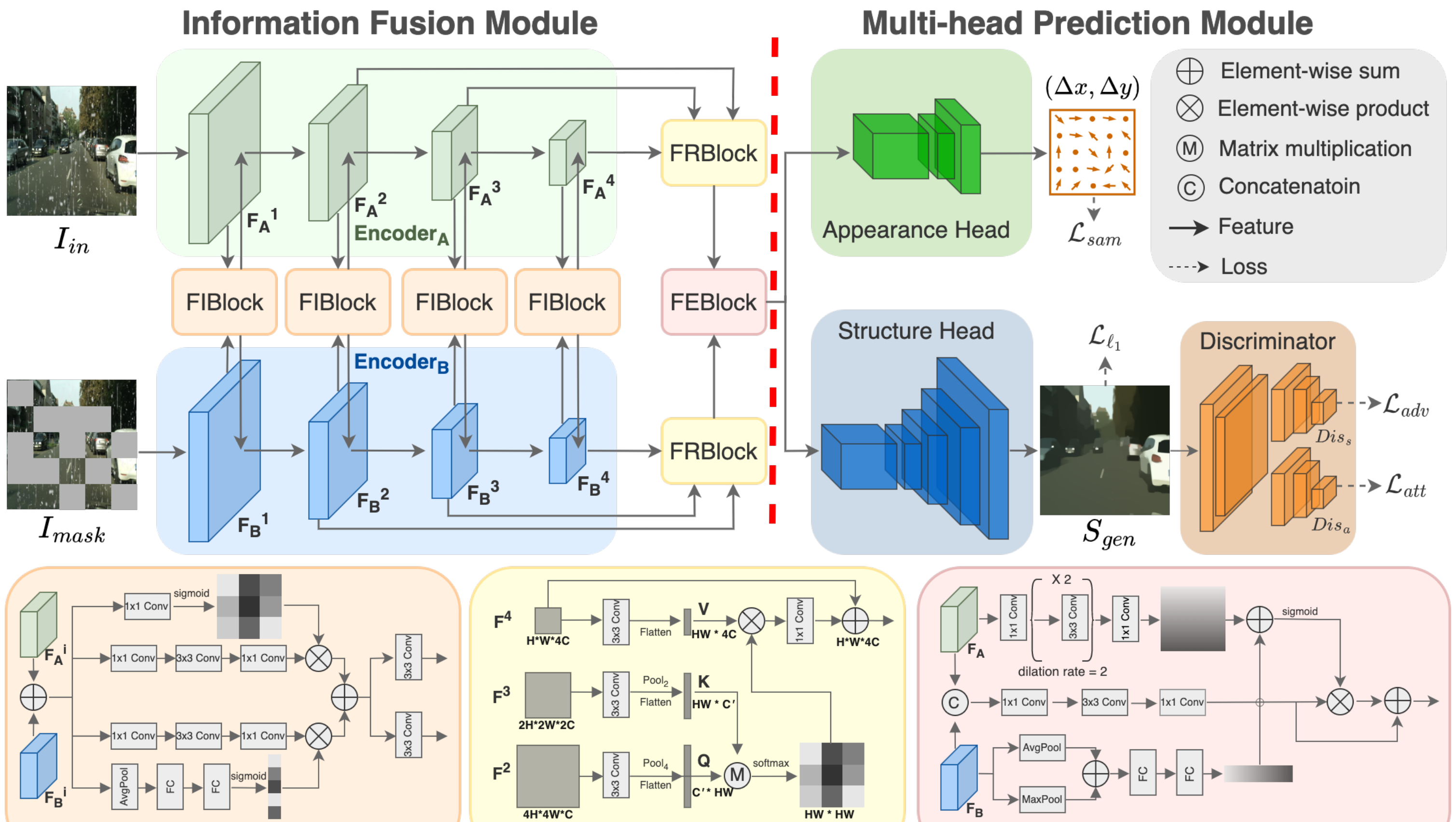
- To facilitate the feature learning, we also propose a Context-aware Pretrained Network (CPNet), which is benefited from the proposed information fusion module and multi-head prediction module for texture-guided appearance flow and conditional attribute label.

Overview



Random noise combination
rain streak
snow
raindrop
Ground Truth I_{gt}
Random noise rotation I_{in}
Random patch masked I_{mask}
Transformer Encoder E_A
Information Fusion Module
Transformer Encoder E_B
Multi-head Prediction Module
Appearance Flow
$(\Delta x, \Delta y)$
Structure Flow S_{gen}
Gaussian Sampling
Mixing function
Refinement network
Final Result I_{out}
Pre-training stage
Fine-tuning stage

Context-aware Pretraining



(a) Feature Interaction Block (FIBlock)

(b) Feature Refinement Block (FRBlock)

(c) Feature Enhancement Block (FEBlock)

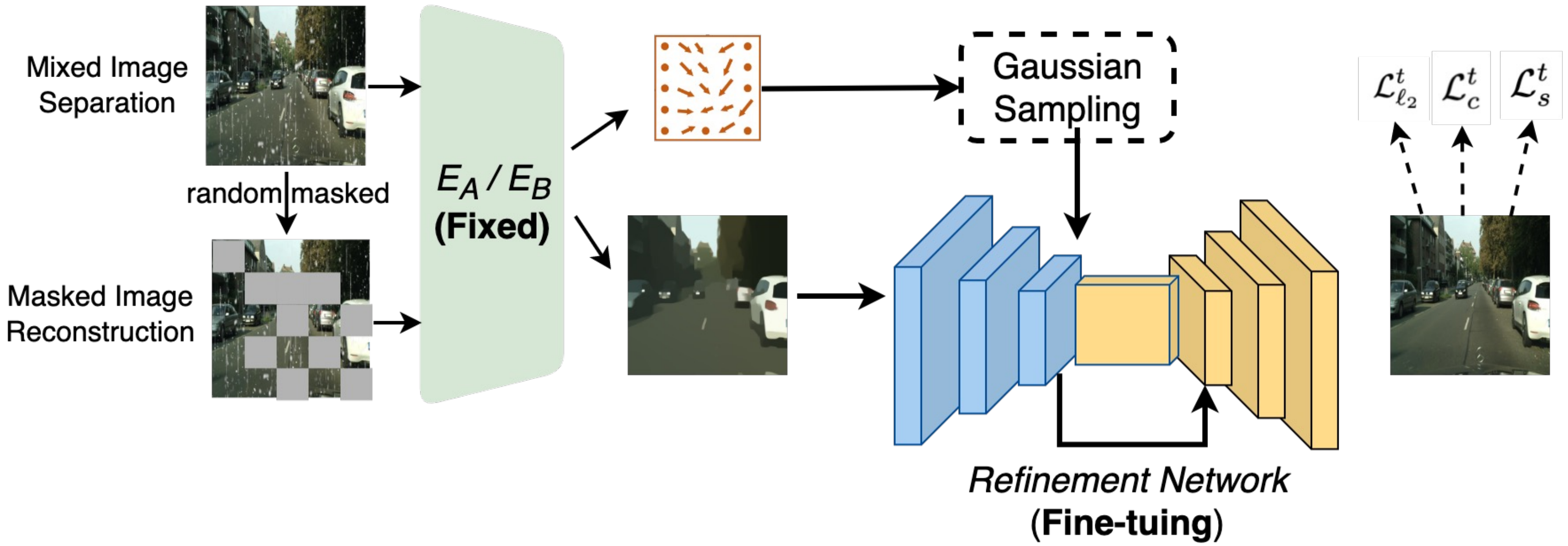
$$\mathcal{L}_{\ell_1} = \left\| S_{gen} - S_{gt} \right\|_1 .$$

$$\mathcal{L}_{adv} = \mathbb{E}\left[\log\left(1 - Dis_s\left(S_{gen}\right)\right)\right] + \mathbb{E}\left[\log Dis_s\left(S_{gt}\right)\right],$$

$$\mathcal{L}_{att} = -\sum_{i=1}^{N} \log P_i\left(S_{gen} \mid \theta_{Dis_a}\right).$$

$$\mathcal{L}_{sam} = \frac{1}{N} \sum_{(x,y)\in\Omega} \exp\left(-\frac{\mu\left(\Phi_{I_{gt}}^{x,y}, \Phi_{I_{in}}^{x+\Delta x, y+\Delta y}\right)}{\alpha\left\|\Phi_{S_{gt}} - \Phi_{S_{gen}}\right\|_1 + \epsilon}\right),$$

Efficient Fine-tuning



$$\mathcal{L}^t_{\ell_2} = \left\| I_{gen} - I_{gt} \right\|_2 .$$

$$\mathcal{L}^t_c = \sum_{i=1}^{N} \frac{1}{HWC} \left| \phi^{gt}_{pool_i} - \phi^{gen}_{pool_i} \right|_1 ,$$

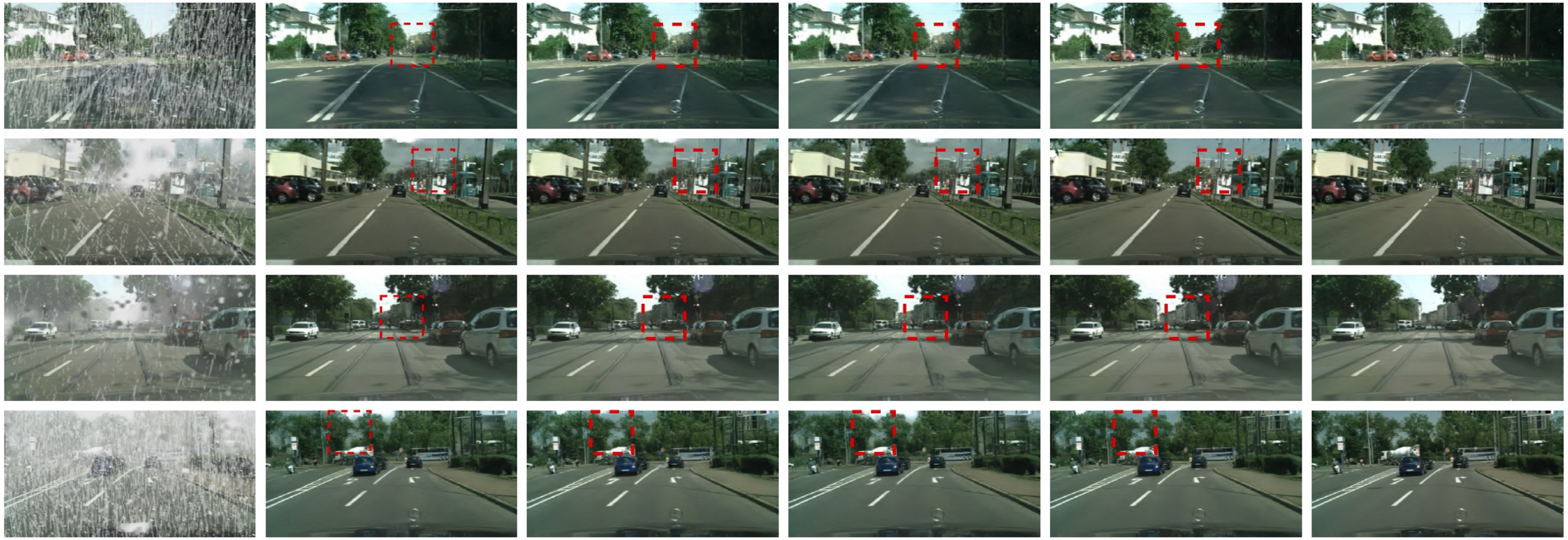
$$\mathcal{L}^t_s = \sum_{i=1}^{N} \frac{1}{C*C} \left| \frac{1}{HWC} \left(\phi^{style_{gt}}_{pool_i} - \phi^{style_{gen}}_{pool_i} \right) \right|_1 ,$$

Experiment Results

Table 1. Quantitative results of Task I in driving scenario. We evaluate the performance in Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) under 6 BID cases, which are (1): rain streak, (2): rain streak + snow, (3): rain streak + light haze, (4): rain streak + heavy haze, (5): rain streak + moderate haze + raindrop, (6) rain streak + snow + moderate haze + raindrop. The best performance under each case is marked in **bold** with the second performance underlined.

Case	Input		MPRNet [62]		Restormer [61]		All-in-one [28]		BIDeN [17]		Ours	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
(1)	25.69	0.786	33.39	0.945	**34.29**	**0.951**	32.38	0.937	30.89	0.932	33.95	0.948
(2)	18.64	0.564	30.52	0.909	30.60	0.917	28.45	0.892	29.34	0.899	**33.42**	**0.937**
(3)	17.45	0.712	23.98	0.900	23.74	0.905	27.14	0.911	28.62	0.919	**32.99**	**0.932**
(4)	11.12	0.571	18.54	0.829	20.33	0.853	19.67	0.865	26.77	0.891	**29.02**	**0.908**
(5)	14.05	0.616	21.18	0.846	22.17	0.859	24.23	0.889	27.11	0.898	**30.07**	**0.925**
(6)	12.38	0.461	20.76	0.812	21.24	0.821	22.93	0.846	26.44	0.870	**29.57**	**0.914**

Experiment Results



Input | Restormer | All-in-One | BIDeN | Ours | GT

Figure 5. Qualitative results of Task I in driving scenario under several mixed cases. Row 1-4 represents the cases (3)-(6) respectively as presented in Table 1. For all cases, our model can produce more precise and faithful images. (Please zoom in to see the details.)

Experiment Results

Table 2. Quantitative results of Task II.B. (1)-(3) indicate that models are trained under different settings. The best performances are marked in **bold** with the second performance underlined. Performance variations between cases (1) and (3) are marked in **blue**.

Method		Input	MPRNet	BIDeN (1)	BIDeN (2)	BIDeN (3)	Ours (1)	Ours (2)	Ours (3)
Rainstreak	NIQE ↓	4.87	**4.10**	4.15	4.28	4.33 **(+0.18)**	4.12	4.12	4.13 **(+0.01)**
	BRISQUE ↓	27.82	28.66	25.76	26.19	26.57 **(+0.81)**	**25.53**	25.57	25.58 **(+0.05)**
Raindrop	NIQE ↓	5.63	4.87	4.55	4.67	4.72 **(+0.17)**	**4.48**	4.59	4.50 **(+0.02)**
	BRISQUE ↓	24.88	29.17	20.29	20.82	21.22 **(+0.93)**	**20.08**	20.11	20.16 **(+0.08)**
Snow	NIQE ↓	4.75	4.48	4.21	4.25	4.31 **(+0.10)**	**4.14**	4.15	4.16 **(+0.02)**
	BRISQUE ↓	22.68	25.78	21.99	22.25	22.42 **(+0.43)**	**21.83**	21.85	21.88 **(+0.05)**

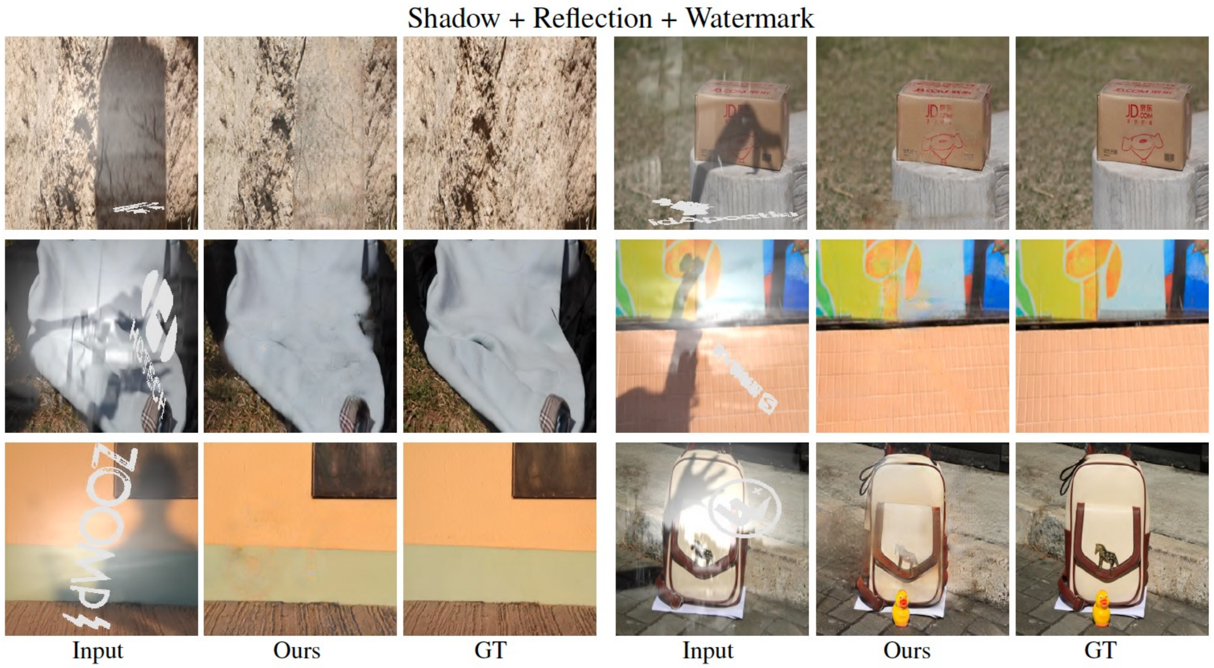
Table 3. Quantitative results of Task III. We evaluate the Root Mean Square Error (RMSE ↓) in LAB color space. The best performances are marked in **bold** with the second performance underlined. Performance variations between cases (1) and (3) are marked in **blue**.

RMSE	DHAN	Auto-Exposure	BIDeN (1)	BIDeN (2)	BIDeN (3)	Ours (1)	Ours (2)	Ours (3)
Shadow	8.94	**8.56**	12.01	14.15	15.49 **(+2.14)**	8.65	8.70	8.76 **(+0.11)**
Non-Shadow	**4.80**	5.75	7.52	8.21	8.93 **(+2.46)**	4.98	4.98	4.99 **(+0.01)**
All	**5.67**	6.51	8.77	9.85	10.69 **(+3.34)**	5.97	5.99	6.03 **(+0.06)**

Experiment Results

Input | BIDeN(3) | Ours | Input | BIDeN(3) | Ours

Experiment Results



Experiment Results

Table 5. Ablation on the pre-training dataset for Task I case (5).

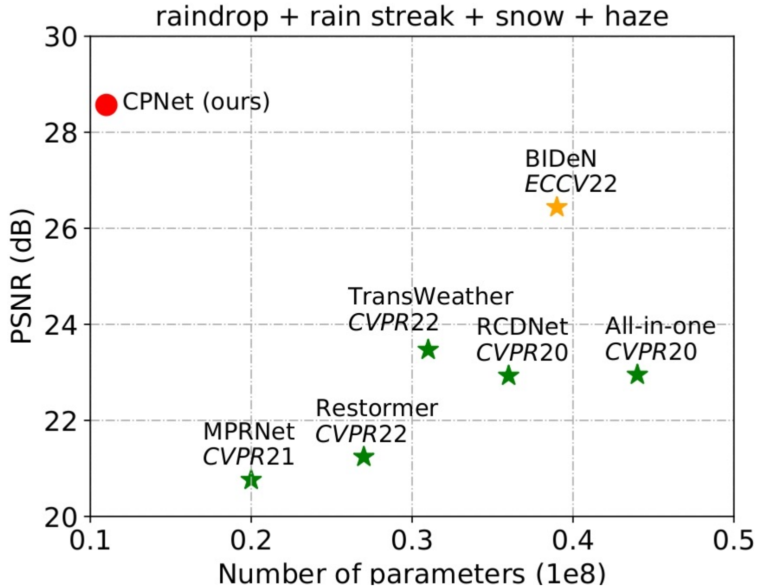
Dataset	BIDeN	only BID	10% ImageNet	Scene class	Object class	Full ImageNet
PSNR↑	27.11	26.80	28.95	28.43	24.89	**30.07**

Table 6. Discussion on the model efficiency. All models are tested under the same environment for fair comparisons.

Method	MPRNet	All-in-one	BIDeN	Ours
Param (M)	21.15	44.26	38.61	11.30
FLOPs (G)	135	350	344	102
Inference time (s)	0.21	0.34	13.21	0.26

Table 9. Performance for Task I case (5) during finetuning.

Fine-tuning epochs	BIDeN	10	20	30	40
PSNR↑	27.11	27.08	29.15	**30.07**	30.05



JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Visualization

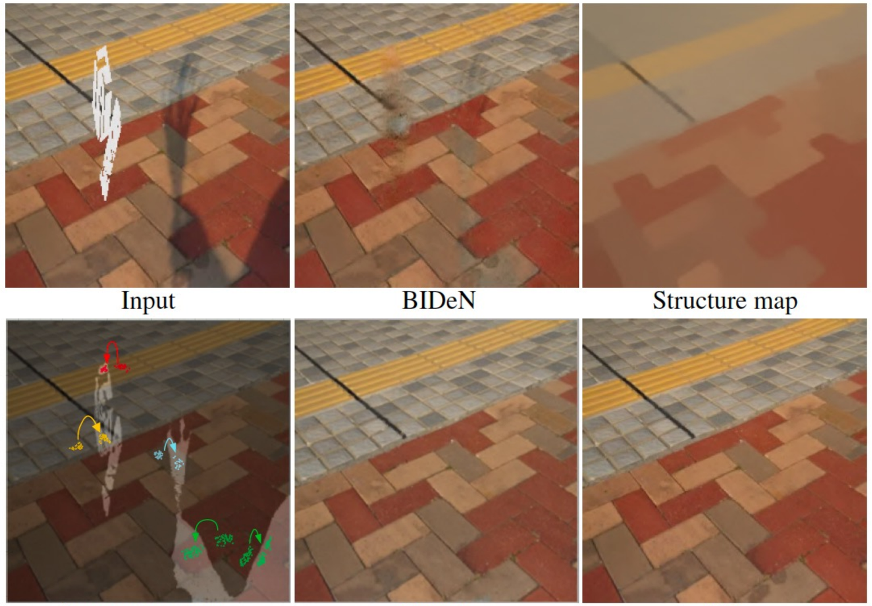


Figure 7. Visualizations of the outputs during pretraining and fine-tuning. To visualize the appearance flow fields, we plot part of the sample points of typical missing regions. The arrows show the direction of the appearance flow. Please zoom in to see the details.

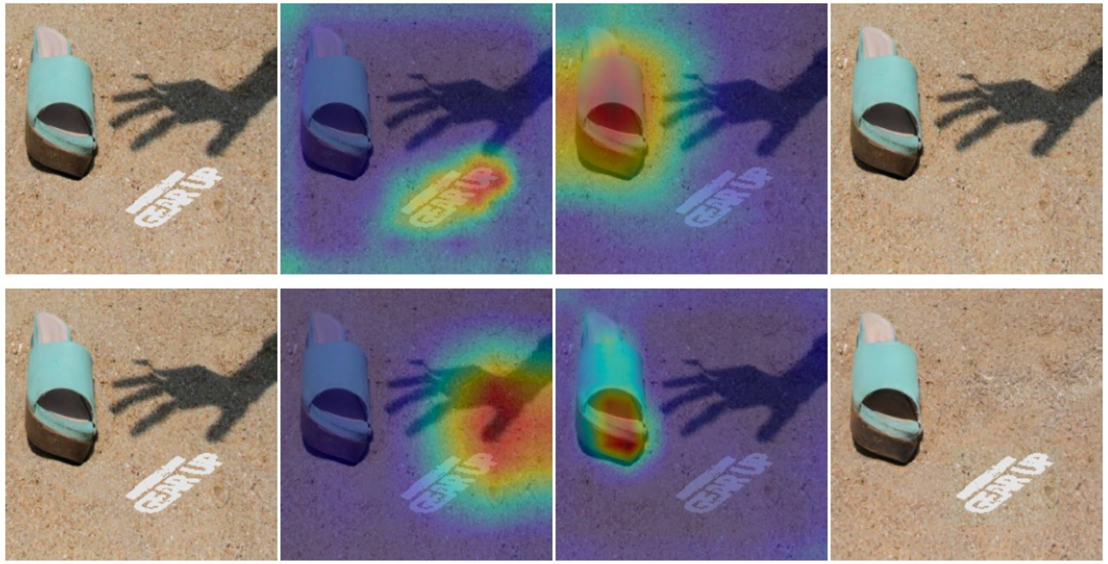


Figure 8. Attention visualizations. Mask attention represents the feature activation map in E_A, while the context attention comes from E_B . The top row shows the outputs for watermark removal and the bottom row shows shadow removal results.

Conclusion

- In this paper, we propose a new context-aware pretraining paradigm (CP) for the BID task. Different from previous methods, we shed light on the possibilities of self-supervised pretraining to remove multiple general noises in one go.
- During pretraining, the CPNet model is designed with two entangled encoders serving different image processing tasks, i.e., mixed image separation and masked image reconstruction, for joint context-aware learning.
- Experiments on seven representative restoration tasks and three BID tasks demonstrate that CPNet consistently facilitates state-of-the-art performance in terms of both image restoration quality and efficiency.

Paper

Code