Ingredient-oriented Multi-Degradation Learning for Image Restoration Supplementary Material

Jinghao Zhang*, Jie Huang*, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, Feng Zhao†
University of Science and Technology of China
{jhaozhang,hj0117,mdyao,yzz6000,yuhu520,manman}@mail.ustc.edu.cn, fzhao956@ustc.edu.cn

This supplementary document is organized as follows:
Sec. 1 presents the discussion about the clean image priors.
Sec. 2 presents architecture of the supervised degradation attention module.
Sec. 3 presents more implementation details.
Sec. 4 presents the latent visualization.
Sec. 5 presents more visual comparison results.
Sec. 6 presents the broader impacts.

1. Discussion about the clean image priors

In our proposed ingredients-oriented degradation reformulation framework (IDR), the correlation among diverse image restoration tasks are explored for task-scalable learning. Beyond that, it is worth noting that the natural opposition [2] between the corrupted images and clean images provides another avenue, which are *two roads lead to Rome*.

Prior work [1] has already dedicated to the development of this line, and advocated to learn normal image priors for handling various image restoration tasks. However, several problems have been exposed: a) it need auxiliary network to pre-generate pseudo clean images at inference time to provide normal image priors; b) requires fine-tuning on downstream tasks; c) inflexibility to arbitrary image sizes as fixed number of prior queries. Therefore, it remains a challenge to obtain generalizable and convenient normal image priors in context of various degradations.

Table 1. Comparison experiments on the clean image priors.

Methods	Rain100L	SOTS	BSD68	GoPro	LOL
TAPE [1]	29.67	22.16	30.18	24.47	18.97
clean	35.31	25.64	30.81	26.97	20.93
w/o. clean	35.63	25.24	31.60	27.87	21.34

To this end, we attempt to integrate the clean priors into the IDR framework, via simply include the clean images during training, and establish the corresponding clean prior hub $\mathcal{T}_{clean} \in \mathbb{R}^{N \times C_d}$ as other degradations in the meta-prior learning module. Unfortunately, the overall degradation removal performance benefits poor little from the participation of clean images, as shown in Table 1. We present the t-SNE statistics of the learned clean prior embeddings and degraded prior embeddings in the main body of the paper, where the apparent distinctions are observed. Therefore, how to collaborate the **correlation among diverse degradations** and the **opposition between the normal clean images and degraded images** deserves further investigation.

2. Supervised degradation attention module

We present the architecture of the supervised degradation attention module (SDAM) in Fig. 1 as its similar structure inspired by [3]. Their key differences are summarized as follows. The intention of the SDAM is to intensify the latent degradation while dilute the background content for content-agnostic prior learning, which is completely different from [3]. The supervision of the SDAM lies in the degraded images instead of clean images. The position of the SDAM is mainly concentrated in the encoder stage of the network to achieve better degradation perception, rather than the decoder stage in [3]. Note that we stop the gradients on the input degraded features to prevent the inferior impact on the backbone network, while only the intensified degraded features are needed.

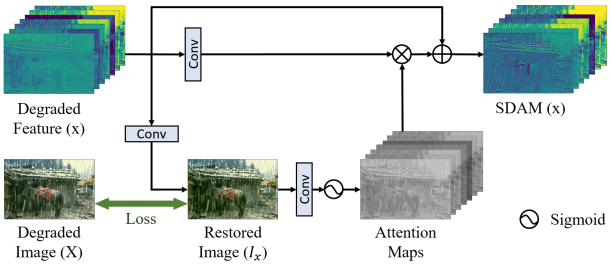

Figure 1. Supervised degradation attention module.

3. More implementation details

We implement our IDR with a trimmed restormer backbone to verify the model scalability under the constrained

*Both authors contributed equally to this research.
†Corresponding author.

capacity bottleneck and to be consistent with previous all-in-one fashion methods in terms of model size. Specifically, from stage-1 to stage-4, the number of transformer blocks are [2,3,3,4], attention heads are [1,2,4,8], and the number of channels are [32,64,128,256]. The refinement stage contains 2 blocks. The capability of the prior hub N is empirically set as 24, as shown in Table 3.

We embed two meta-prior learning modules (MPL) at the encoder stage and one at the decoder stage. Table 2 presents the ablation experiments on the distribution of the MPL. One can see that the complete decoder assignment (*i.e.* 3D) leads to the worst performance, as inconspicuous degradations contribute weakly to the knowledge collection, compared to the preceding encoder stage.

Table 2. Ablation experiments on the distribution of the MPL.

Methods	Rain100L	SOTS	BSD68	GoPro	LOL
3E	35.15	25.81	31.41	27.23	21.70
2E+1D	35.63	25.24	31.60	27.87	21.34
3D	32.88	23.29	31.13	26.63	21.61

Table 3. Ablation experiments on the capability of the prior hub.

Capability N	Rain100L	SOTS	BSD68	GoPro	LOL
12	35.51	24.60	31.49	28.00	21.18
24	35.63	25.24	31.60	27.87	21.34
36	35.48	24.76	31.61	27.99	21.61
48	35.69	24.56	31.62	28.01	21.25

Table 4. Ablation experiments on the components design.

Methods	Rain100L	SOTS	BSD68	GoPro	LOL
a	34.73	25.26	31.37	27.42	21.51
b	34.46	25.33	31.41	27.24	21.19
c	35.01	24.98	31.45	27.48	21.01
d	35.27	25.13	31.52	27.45	20.99
e	35.51	25.28	31.48	27.69	21.21
f	35.63	25.24	31.60	27.87	21.34

Additionally, in Table 4, we provide a detailed supplement to Table 4 of the main paper with metrics on each task.

4. Latent visualization

We visualize the latent representations inside the meta-prior learning module on diverse image degradations in Fig. 2. The latent representations include the input degraded feature x, prior-oriented degradation representation h_p, and the output pseudo clean feature x'.

For x with different degradation types, it typically suffer from severe corruption, such as intensive rain streaks or noise, unclear information affected by low-light and hazy conditions, and blurry scenes.

For h_p with different degradation types, it is constructed via the aggregation of the learned prior embeddings $\mathcal{P}$. It can be observed that the extracted prior-oriented degradation representations contain much more degradation-related

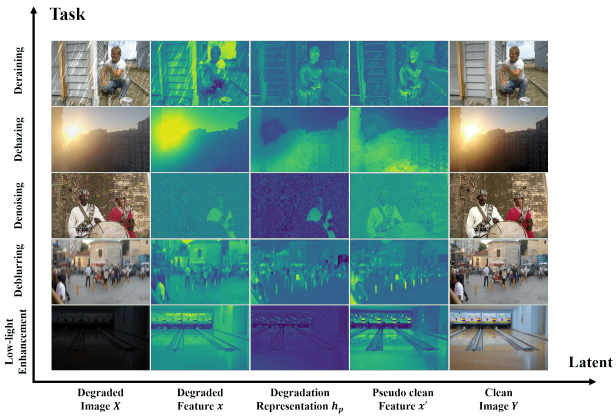

Figure 2. Latent visualization in the meta-prior learning module. Please zoom in for details.

information with diluted background content, compared to x. For example, the rain streaks and the noise are much more decoupled from the backgrounds, compared with their entangled state in x. Additionally, the unclear regions of the low-light and hazy conditions are well recognized, such as the discrete bowling balls and the underlying main body of the building. While the h_p for blur reflects more potential directionality compared to x, *e.g.* up and down.

For x' with different degradation types, it alleviates the corruption on the feature to some extent. For example, the rain streaks and the noise are significantly reduced. The contrast of the low-light and hazy features is improved. The motion effect of the blurry features is suppressed.

5. Additional visual comparison results

In this section, we provide more visual comparison results on the foregoing five image restoration tasks in Figs. 3 to 7, and an additional unknown task (*i.e.* under-display camera image restoration) in Fig. 8, demonstrating the scalability and generalization ability of our method.

6. Broader impacts

Our work provides a novel perspective to explore the correlation among diverse image restoration tasks to realize the intrinsic degradation ingredients, and exhibits favorable generalization ability and scalability. In a broader vision, it potentially release the redundant model deployments in real world scenarios, and avoid the model switching when faced with complex environments. We note that it sincerely benefits a lot of applications with limited resources, such as mobile photography and 24/7 surveillance. However, the privacy of our method may raise potential concerns when used improperly. For example, some important occlusions in the original images may be removed, resulting in the disclosure of private information. Therefore, how to ensure the user-agnostic security of our method needs further research.

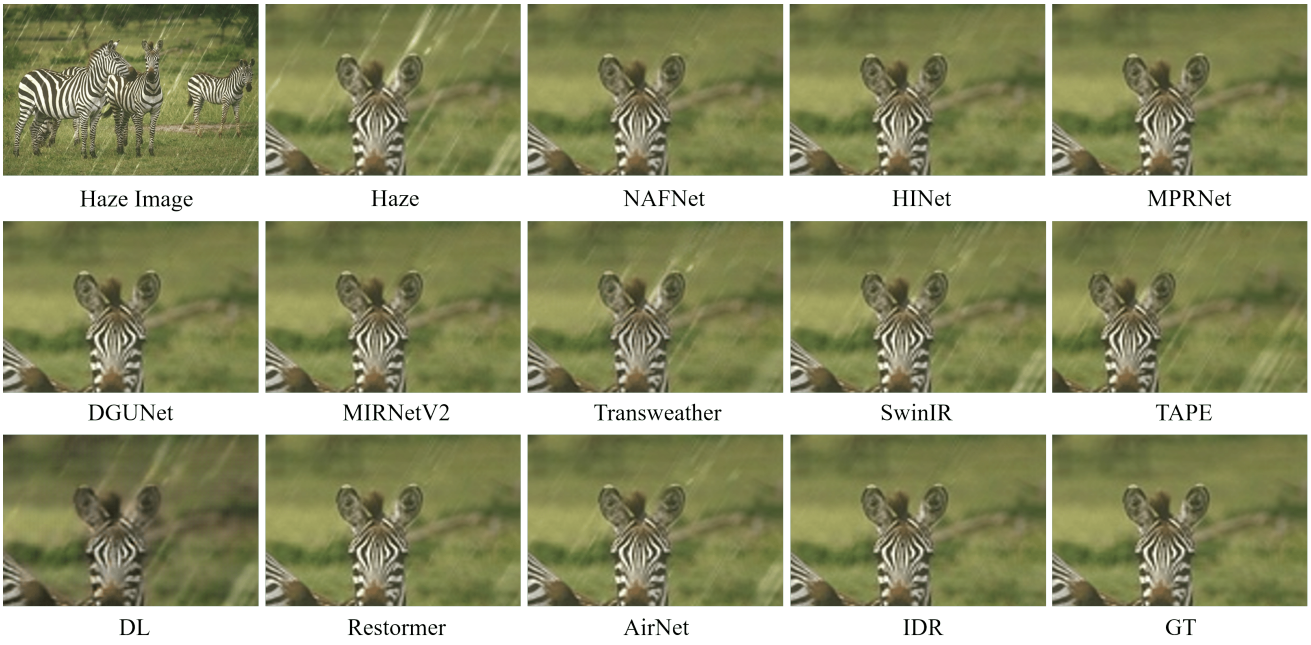

Figure 3. Visual comparison with state-of-the-art methods on Rain100L dataset. Please zoom in for details.

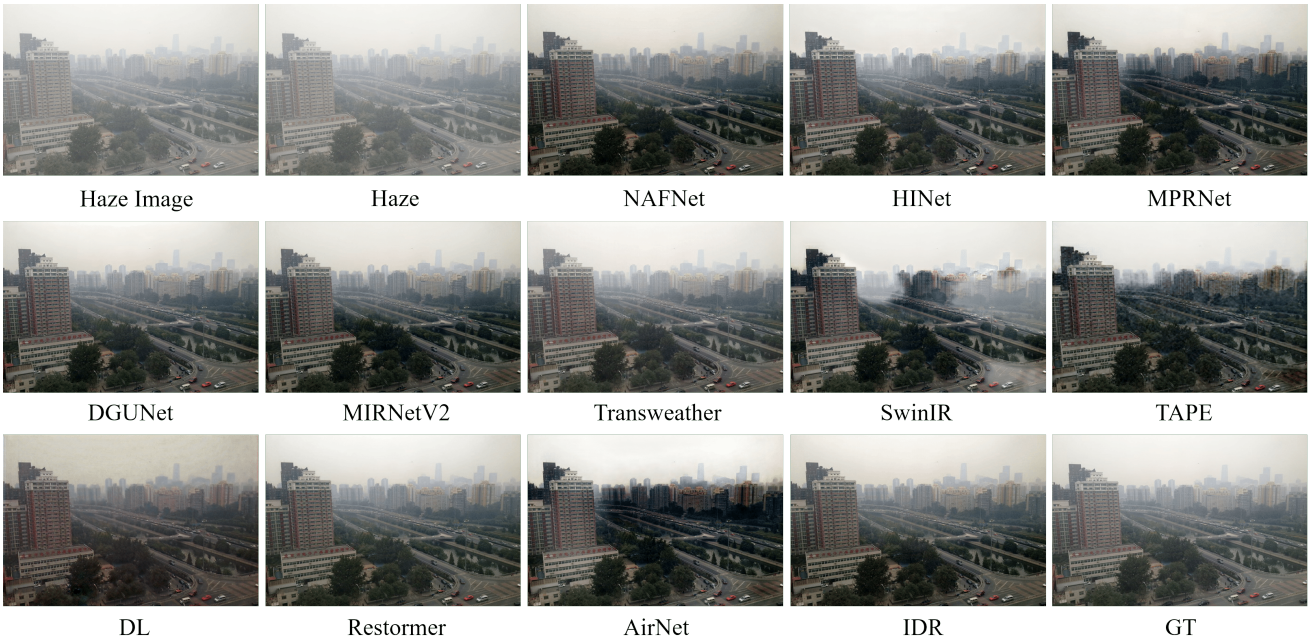

Figure 4. Visual comparison with state-of-the-art methods on SOTS dataset. Please zoom in for details.

References

[1] Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. Tape: Task-agnostic prior embedding for image restoration. In *In Proceedings of the European conference on computer vision*, pages 447–464. Springer Nature Switzerland, 2022. 1

[2] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1

[3] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 1

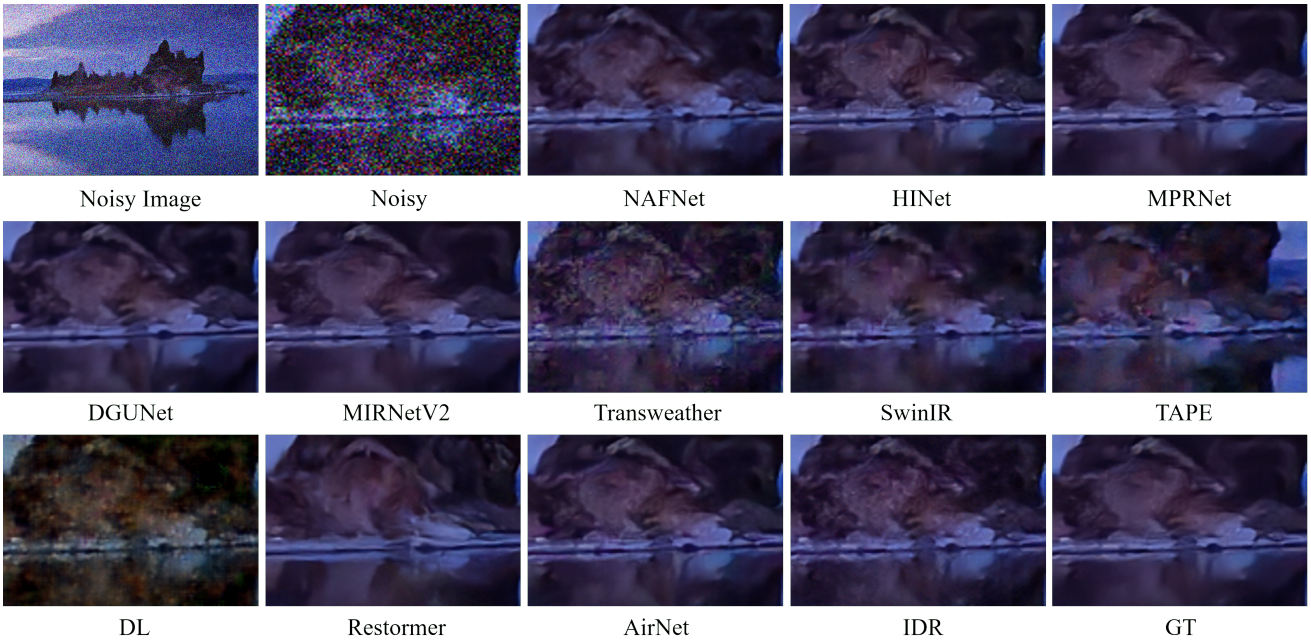

Figure 5. Visual comparison with state-of-the-art methods on BSD68 dataset. Please zoom in for details.

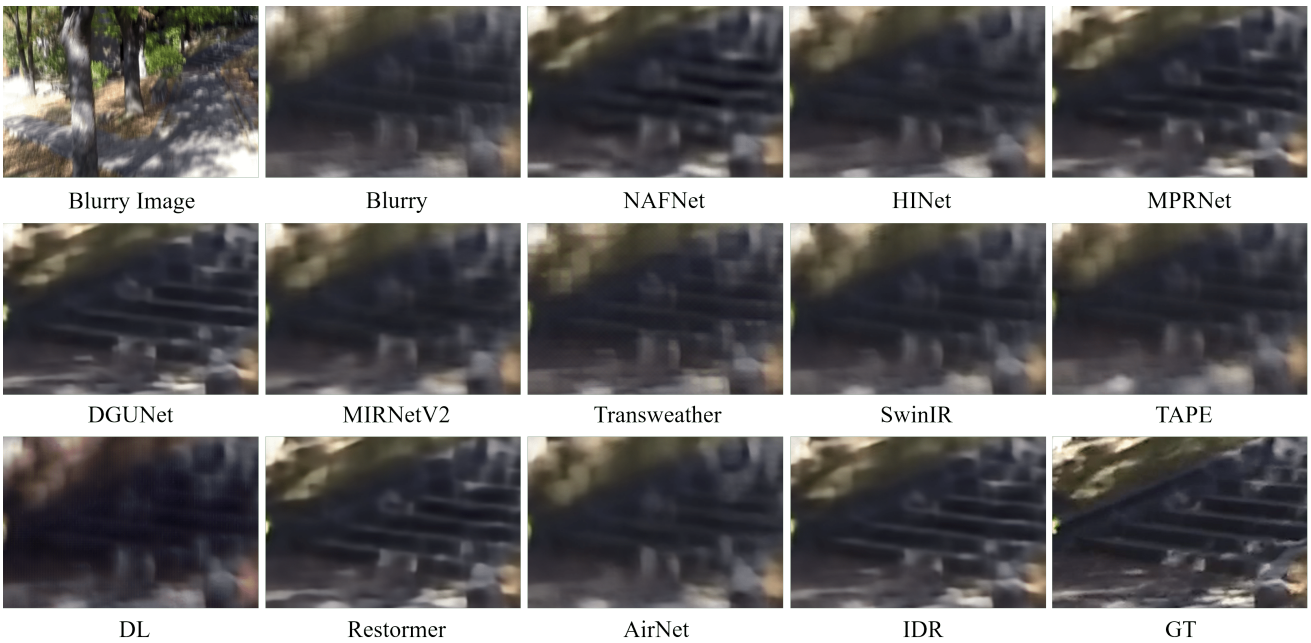

Figure 6. Visual comparison with state-of-the-art methods on GoPro dataset. Please zoom in for details.

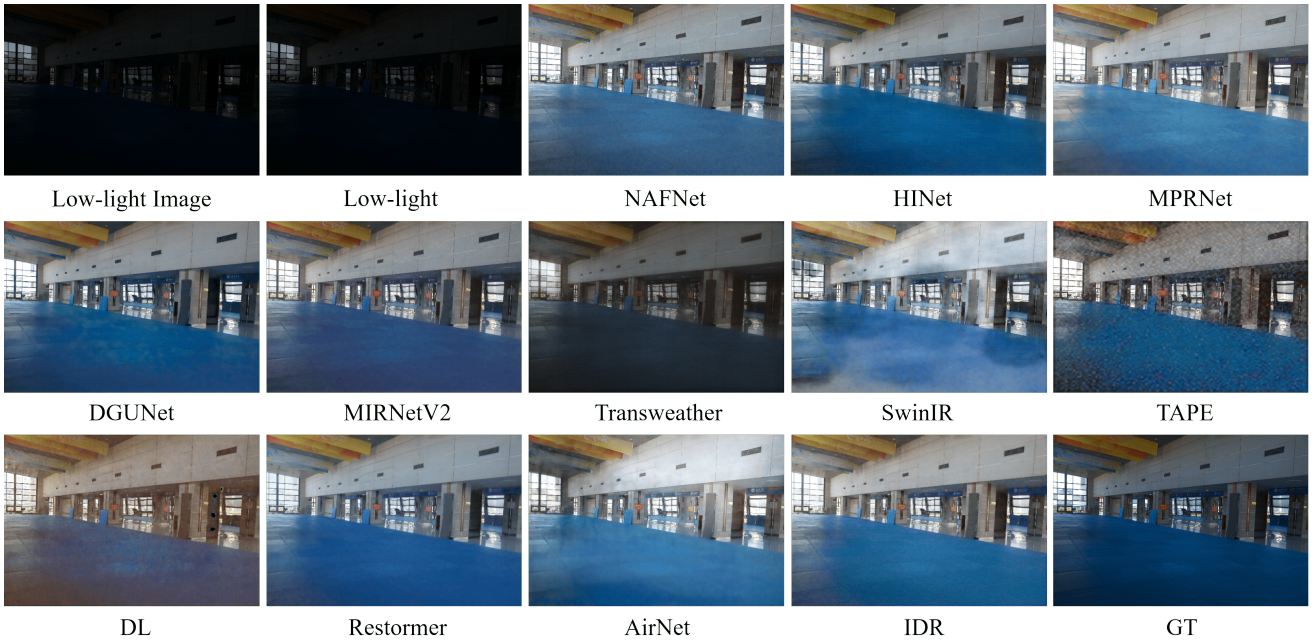

Figure 7. Visual comparison with state-of-the-art methods on LOL dataset. Please zoom in for details.

Figure 8. Visual comparison with state-of-the-art methods on UDC (TOLED) dataset. Please zoom in for details.