Supplemental Material

Anonymous Author(s)
Affiliation
Address
email

1 Additional Ablation Studies

We conduct further ablation studies to illustrate the effectiveness of various design choices of the PromptIR framework. We examine various key design choices like the usage of prompt tokens and plugging in prompt blocks only on the decoder branch of the network.

1.1 Contrastive learning-based Degradation Encoder embedding v/s Prompt Tokens

To strengthen the design rationale for incorporating prompts instead of following recent methods [3] that use embeddings learned through contrastive training, we replace the generated prompt from our PGM module with embeddings extracted from the Contrastive- learning based Degradation Encoder of the AirNet [3] model. We observed that the use of contrastive embeddings resulted in significantly weaker performance compared to prompt tokens. Moreover, achieving good performance with contrastive embeddings requires a custom-designed restoration network, whereas our Prompt Blocks can be seamlessly integrated as plug-and-play modules into any restoration network.

Table 1: Comparisons under all-in-one setting: between the usage of degradation embedding extracted from the Contrastive-learning Based Degradation Encoder (CBDE) of the Airnet [3] Model and the usage of prompt tokens in the PromptIR framework.

Method	Dehazing on SOTS [4]	Deraining on Rain100L [1]	Denoising on BSD68 dataset [5]) $\sigma = 15$	$\sigma = 25$	$\sigma = 50$	Average
CBDE+PromptIR	23.92/0.881	32.03/0.0.972	32.96/0.910	30.36/0.860	26.93/0.732	29.24/0.875
PromptIR (Ours)	**30.58/0.974**	**36.37/0.972**	**33.98/0.933**	**31.31/0.888**	**28.06/0.799**	**32.06/0.913**

1.2 Prompt Blocks on both Encoder branch and Decoder branch

We study the importance of decoder-only prompting by evaluating the usage of prompt blocks on both the encoder and decoder branches. We show that it is important the prompt block is only used on the decoder side.

Table 2: Comparisons under the all-in-one setting: between the usage of the Prompt-block on both the encoder branch and encoder branch with using the prompt block only on the decoder branch.

Method	Dehazing on SOTS [4]	Deraining on Rain100L [1]	Denoising on BSD68 dataset [5]) $\sigma = 15$	$\sigma = 25$	$\sigma = 50$	Average
Enc+Dec+PromptIR	28.52/0.927	35.43/0.965	33.59/0.927	30.85/0.878	27.35/0.732	31.14/0.885
PromptIR (Ours)	**30.58/0.974**	**36.37/0.972**	**33.98/0.933**	**31.31/0.888**	**28.06/0.799**	**32.06/0.913**

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

2 Transformer Block in PromptIR Framework

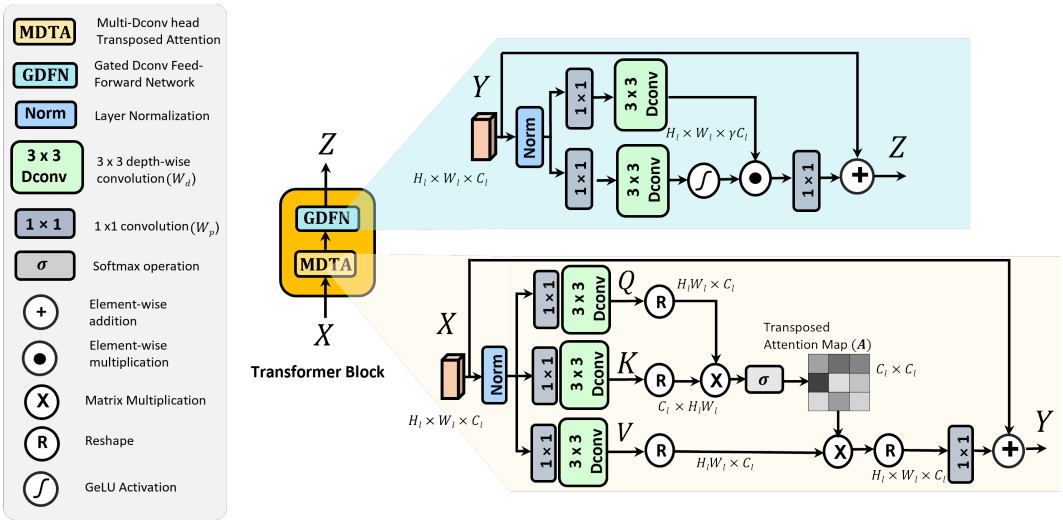

Figure 1: Overview of the Transformer block used in the PromptIR framework. The Transformer block is composed of two sub modules,the Multi Dconv head transposed attention module(MDTA) and the Gated Dconv feed-forward network(GDFN).

As mentioned in section 3.1.2 of the main manuscript,we present the block diagram1 of the transformer block and further elaborate on the details of the transformer block employed in the PromptIR framework.The tranformer block follows the design and hyper-parameters outlined in [6]

To begin, the input features $\mathbf{X} \in \mathbb{R}^{H_l \times W_l \times C_l}$ are passed through the MDTA module. In this module, the features are initially normalized using Layer normalization. Subsequently, a combination of 1×1 convolutions followed by 3×3 depth-wise convolutions are applied to project the features into Query ($\mathbf{Q}$), Key ($\mathbf{K}$), and Value ($\mathbf{V}$) tensors. An essential characteristic of the MDTA module is its computation of attention across the channel dimensions, rather than the spatial dimensions. This effectively reduces the computational overhead. To achieve this channel-wise attention calculation, the Q and K projections are reshaped from $H_l \times W_l \times C_l$ to $H_l W_l \times C_l$ and $C_l \times H_l W_l$ respectively, before calculating dot-product, hence the resulting transposed attention map with the shape of $C_l \times C_l$. Bias-free convolutions are utilized within this submodule. Furthermore, attention is computed in a multi-head manner in parallel.

After MDTA Module the features are processed through the GDFN module. In the GDFN module the input features are expanded by a factor γ using 1×1 convolution and they are then passed through 3×3 convolutions. These operations are performed through two parallel paths and output of the paths are activated using GeLU non-linearity. This is then combined with the output of the other path using element-wise product.

3 Qualitative results:

We present more qualitative results from single-task models to further elucidate the effectiveness of prompt-block even when under the single-task setting.

3.1 Dehazing

Figure 2: **Image deraining comparisons** under single task setting on images from the SOTS dataset [4]. Our method effectively removes haze to produce visually better images.

3.2 Deraining

Figure 3: **Image deraining comparisons** under single task setting on images from the Rain100L dataset [1]. Our method effectively removes rain streaks to generate rain-free images.

3.3 Denoising

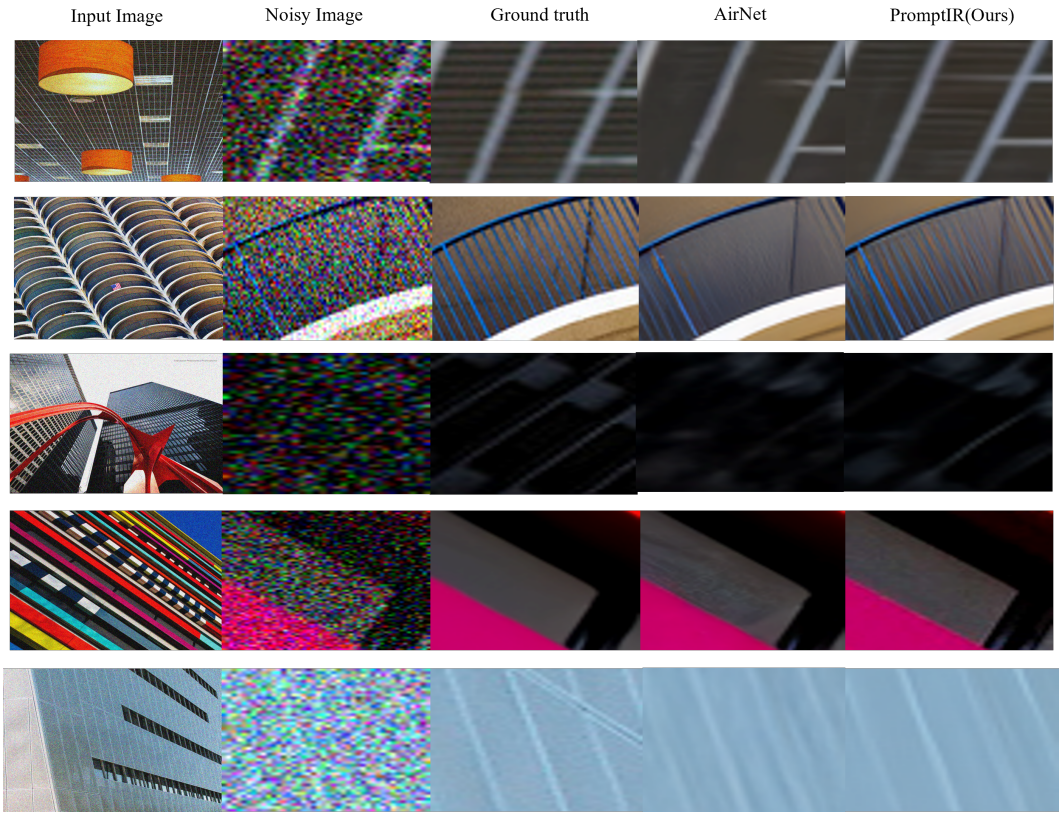

Figure 4: **Image deraining comparisons** under single task setting on images from the URBAN100 dataset [2] with $\sigma = 50$. Our method produces visually better images as compared to previous methods. We show selected patches from the images.

References

[1] Fan, Q., D. Chen, L. Yuan, G. Hua, N. Yu, and B. Chen (2019). A general decoupled learning framework for parameterized image operators. *IEEE transactions on pattern analysis and machine intelligence 43*(1), 33–47.

[2] Huang, J.-B., A. Singh, and N. Ahuja (2015). Single image super-resolution from transformed self-exemplars. In *CVPR*.

[3] Li, B., X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng (2022, June). All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17452–17462.

[4] Li, B., W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang (2018). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing 28*(1), 492–505.

[5] Martin, D., C. Fowlkes, D. Tal, and J. Malik (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*.

[6] Zamir, S. W., A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang (2022). Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*.