Decouple to Reconstruct: High Quality UHD Restoration via Active Feature Disentanglement and Reversible Fusion

Supplementary Material

This supplementary document is organized as follows:
Sec. 1 presents additional visual results.
Sec. 2 presents the results under six degradation experimental settings.
Sec. 3 provides the construction of the UHD dataset and details of the experimental setup.

1. More visual comparison results.

We present additional visual results for low-light image enhancement, image dehazing, image deblurring, and moiré pattern removal in Figures 1 to 4. As can be observed, our method achieves minimal degradation artifacts while maintaining the consistency of background information in the images.

2. The results under six degradation experimental settings

We further design six types of degraded UHD all-in-one experiments, including low-light enhancement, image deblurring, image dehazing, image denoising, image deraining, and image desnowing. The experimental results are shown in Tab. 2. Our method significantly outperforms both traditional all-in-one approaches and UHD restoration methods. By balancing efficiency and performance, we validate the effectiveness of our method.

3. Experimental Details

3.1. Datasets

The various UHD degradation scenarios in this paper are based on UHD-LL [7], UHD-blur [4], UHD-haze [19], UHD-rain [2], and UHD-snow [12]. For UHD denoising, 4k images from [17] are used as the background.The distributions of the training and testing sets for all datasets are shown in Tab. 1.

3.2. Implementation Details

The number of encoder and decoder layers is set to 3, the number of modules in the latent image restoration network is set to 6, and the number of Glow modules in CIMF-Net is set to 3.

For the first stage, we train Clean-VAE on the image reconstruction task. The initial learning rate is set to 5×10^{-4}, gradually reduced to 1×10^{-7} using cosine annealing. The batch size is set to 16, and the images are randomly cropped to 256×256.

In the second stage, we train CD²-VAE based on paired degraded-clean image inputs for feature disentanglement training. On one hand, the degraded latent extracted from the input is combined with the clean latent extracted by the Clean-VAE encoder and input into the degraded decoder for the reconstruction of the degraded image. On the other hand, the disentangled background features are input into Clean-VAE's decoder for the reconstruction of the clean image. The initial learning rate is set to 5×10^{-4}, gradually reduced to 1×10^{-7} using cosine annealing. The batch size is set to 12, and the images are randomly cropped to 256×256.

For the third stage, we train D²R-UHDNet on the image restoration task, keeping the parameters of CD²-VAE frozen. We fine-tune the parameters of the LaReNet and CIMF-Net. The initial learning rate is set to 4×10^{-4}, gradually reduced to 1×10^{-7} using cosine annealing. The batch size is set to 6, and the images are randomly cropped to 512×512.

3.3. Training procedure

In the first phase, Clean-VAE is trained for the image reconstruction task using clean images. The clean input image is denoted as I_h, and the corresponding reconstructed image is represented as I_{r1}.

The loss function for a standard Variational Autoencoder (VAE) comprises two key terms: the reconstruction loss and the KL divergence loss. The reconstruction loss quantifies the difference between the decoder's output and the original input, ensuring the output remains consistent with the input image [15]. The KL divergence loss regularizes the latent space by encouraging the posterior distribution to align with the prior distribution. This regularization improves the model's ability to generate consistent and continuous reconstructions from similar inputs [20].

We adopt this approach, where the reconstruction loss and KL divergence loss are defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{rec}_1} &= \frac{1}{N}\sum_{i=1}^{N} \|I_{\text{r1}}^{(i)} - I_{\text{h}}^{(i)}\|_1, \\ \mathcal{L}_{\text{KL}} &= D_{\text{KL}}(q(z|I)\|p(z)), \end{aligned} \tag{1}$$

where $q(z|I_h)$ represents the approximate posterior distribution of the latent variable z given the input image I_h, and $p(z)$ is the prior distribution, usually chosen as a standard Gaussian distribution $\mathcal{N}(0, I)$. The KL divergence D_{KL} measures the discrepancy between the posterior distribution $q(z|I_h)$ and the prior distribution $p(z)$.

Additionally, we enforce frequency domain consistency in the reconstruction results using the FFT loss, which is

Table 1. Dataset details and corresponding tasks.

Dataset	Training samples	Testing samples	Task
UHD-Snow	2,000	200	Desnowing
UHD-Blur	1,964	300	Deblurring
UHD-Rain	2,000	500	Deraining
UHD-LL	2,000	115	LLIE
UHD-Haze	2,290	231	Dehazing
UHD-Noise	2,000	500	Denoising

Table 2. *Comparison to state-of-the-art on three degradations.* PSNR (dB, ↑), SSIM (↑), LPIPS (↓) and FS represents full-size 4K image inference. FLOPs are computed for an input size of 256×256. **Best** and **second best** performances are highlighted.

Method	FS	FLOPs	Params.	*Low Light*		*Deblurring*		*Dehazing*		*Denoising*		*Deraining*		*Desnowing*		Average		
				UHD-LL		UHD-blur		UHD-haze		UHDN$_{\sigma=50}$		UHD-rain		UHD-snow				
AIRNet [6]	✗	301G	9M	22.68	.887	23.52	.876	18.24	.846	22.38	.876	26.35	.876	27.38	.924	23.43	.874	.1861
IDR [16]	✗	88G	15.3M	24.33	.915	25.64	.788	18.68	.879	29.64	.906	28.82	.906	30.48	.945	26.27	.890	.1912
PromptIR [9]	✗	158G	33M	23.3	.911	26.48	**.805**	20.14	.901	24.88	.835	28.89	.897	30.78	.966	25.74	.886	.2155
CAPTNet [5]	✗	25G	24.3M	24.97	**.921**	**26.32**	.796	**20.32**	.903	21.64	.569	**29.34**	**.908**	**32.21**	**.974**	25.80	.845	.2861
NDR-Restore [14]	✗	196G	36.9M	25.12	.885	25.64	.791	19.21	.896	31.44	.915	29.24	.897	28.41	.948	26.51	.889	.3108
Gridformer [13]	✗	367G	34M	23.92	.898	25.68	.782	18.87	.889	32.86	.915	29.37	.904	28.24	.942	26.49	.895	.2321
DiffUIR-L [18]	✗	10G	36.2M	22.64	.902	25.08	.785	18.62	.889	**33.25**	**.928**	27.89	.886	27.36	.945	25.81	.889	**.1844**
Histoformer [10]	✗	91G	16.6M	25.73	.915	26.55	.796	18.73	.897	33.05	.924	27.96	.884	27.56	.971	26.59	.898	.1855
adaIR [3]	✗	147G	28.7M	23.84	.918	26.86	.803	19.34	**.910**	32.46	.923	28.18	.901	27.72	.953	26.40	**.901**	.2492
HAIR [1]	✗	41G	29M	**25.22**	.897	24.77	.799	18.75	.883	32.50	.915	28.76	.893	27.89	.968	26.31	.892	.2607
UHDformer [11]	✓	3.0G	0.33M	22.87	.891	24.68	.792	20.02	.888	27.23	.892	28.32	.953	28.24	.882	25.23	.883	.2012
UHDDIP [12]	✓	2.2G	0.81M	24.56	.887	24.26	.794	19.68	.872	28.12	.889	28.78	.942	28.07	.893	25.58	.880	.2278
DreamUHD [8]	✓	4.1G	1.46M	25.12	.901	25.82	.796	20.21	.908	29.08	.901	30.42	.950	32.12	.914	**27.13**	.895	.1998
Ours	✓	4G	1.0M	**26.14**	**.916**	**26.87**	**.799**	**20.38**	**.911**	**29.64**	**.912**	**32.28**	**.968**	**33.32**	**.929**	**28.11**	**.906**	**.1842**

defined as:

$$\mathcal{L}_{\text{FFT}_1} = \frac{1}{N}\sum_{i=1}^{N} \|\text{FFT}(I_{\text{rec}}^{(i)}) - \text{FFT}(I_{\text{h}}^{(i)})\|_1. \tag{2}$$

In the second phase, the input degraded image is denoted as I_d, which corresponds to the clean image I_{gt}. After feature disentanglement learning with CD²-VAE, the degraded image $I_{d_{rec}}$ and the clean image $I_{gt_{rec}}$ are reconstructed. The same reconstruction loss and frequency loss from the first phase are applied, and they are expressed as follows:

$$\begin{aligned} \mathcal{L}_{\text{rec}_2} &= \frac{1}{N}\sum_{i=1}^{N} \|I_{\text{d}_{\text{rec}}}^{(i)} - I_{\text{d}}^{(i)}\|_1 + \|I_{\text{gt}_{\text{rec}}}^{(i)} - I_{\text{gt}}^{(i)}\|_1 \\ \mathcal{L}_{\text{FFT}_2} &= \frac{1}{N}\sum_{i=1}^{N} \|\text{FFT}(I_{\text{d}_{\text{rec}}}^{(i)}) - \text{FFT}(I_{\text{d}}^{(i)})\|_1 \\ &\quad + \|\text{FFT}(I_{\text{gt}_{\text{rec}}}^{(i)}) - \text{FFT}(I_{\text{gt}}^{(i)})\|_1 \end{aligned} \tag{3}$$

In the third phase, the parameters of CD²-VAE are frozen, and D²R-UHDNet takes over the image restoration task. The input consists only of the degraded image I_d, and the output of this restoration process is the restored image, denoted as I_r. The loss function is constructed by comparing the restored image I_r with the clean ground truth image I_{gt}, as follows:

$$\begin{aligned} \mathcal{L}_{\text{rec}} &= \frac{1}{N}\sum_{i=1}^{N} \|I_r^{(i)} - I_{gt}^{(i)}\|_1, \\ \mathcal{L}_{\text{FFT}} &= \frac{1}{N}\sum_{i=1}^{N} \|\text{FFT}(I_r^{(i)}) - \text{FFT}(I_{gt}^{(i)})\|_1. \end{aligned} \tag{4}$$

References

[1] Jin Cao, Yi Cao, Li Pang, Deyu Meng, and Xiangyong Cao. Hair: Hypernetworks-based all-in-one image restoration, 2024. 2

[2] Hongming Chen, Xiang Chen, Chen Wu, Zhuoran Zheng, Jinshan Pan, and Xianping Fu. Towards ultra-high-definition image deraining: A benchmark and an efficient method, 2024. 1

[3] Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation, 2024. 2

[4] Senyou Deng, Wenqi Ren, Yanyang Yan, Tao Wang, Fenglong Song, and Xiaochun Cao. Multi-scale separable network for ultra-high-definition video deblurring. In *the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14030–14039, 2021. 1

[5] Hu Gao, Jing Yang, Ning Wang, Jingfan Yang, Ying Zhang, and Depeng Dang. Prompt-based all-in-one image restoration

Figure 1. Additional visual results for LLIE.

Figure 2. Additional visual results for image dehazing.

using cnns and transformer. *arXiv preprint arXiv:2309.03063*, 2023. 2

[6] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In *IEEE Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, 2022. 2

Figure 3. Additional visual results for image deblurring.

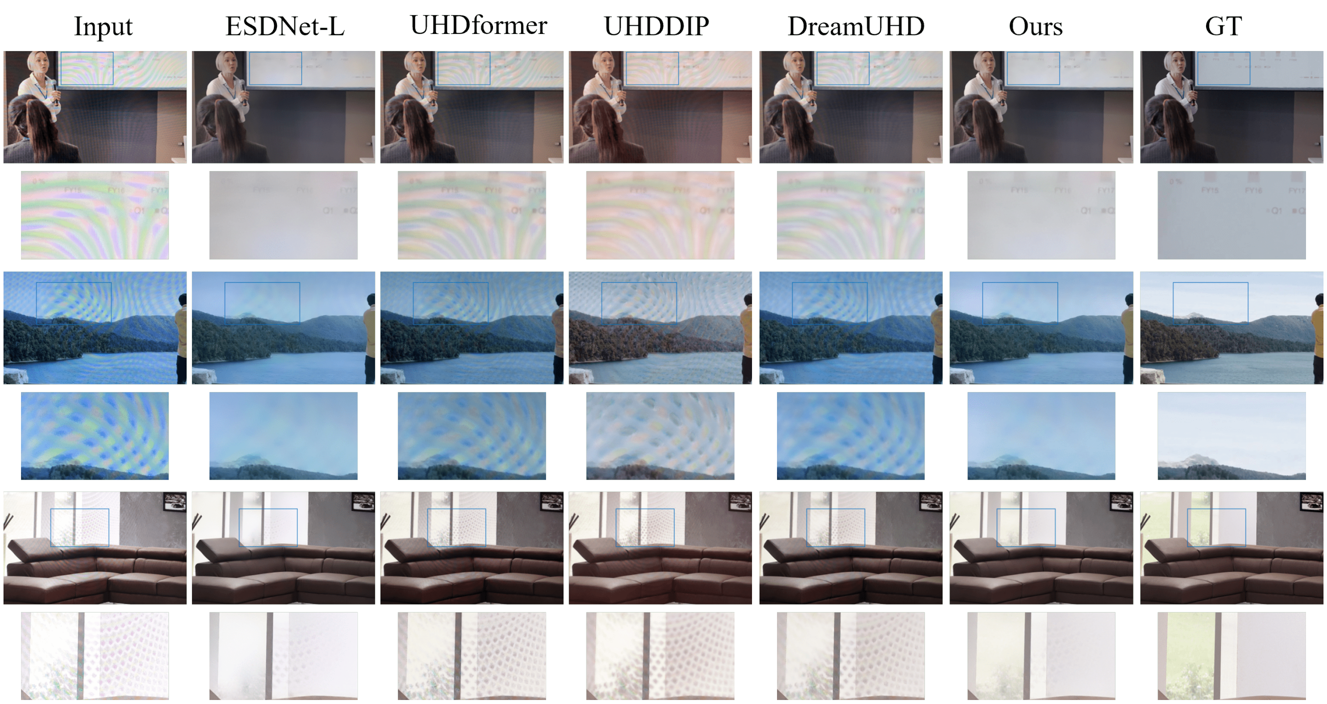

Figure 4. Additional visual results for image demoiring.

[7] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. In *ICLR*, 2023. 1

[8] Yidi Liu, Dong Li, Jie Xiao, Yuanfei Bao, Senyan Xu, and Xueyang Fu. Dreamuhd: Frequency enhanced variational

autoencoder for ultra-high-definition image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5712–5720, 2025. 2

[9] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Khan. Promptir: Prompting for all-in-one image restoration. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 2

[10] Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, pages 111–129. Springer, 2025. 2

[11] Cong Wang, Jinshan Pan, Wei Wang, Gang Fu, Siyuan Liang, Mengzhu Wang, Xiao-Ming Wu, and Jun Liu. Correlation matching transformation transformers for uhd image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5336–5344, 2024. 2

[12] Liyan Wang, Cong Wang, Jinshan Pan, Weixiang Zhou, Xiaoran Sun, Wei Wang, and Zhixun Su. Ultra-high-definition restoration: New benchmarks and a dual interaction prior-driven solution, 2024. 1, 2

[13] Tao Wang, Kaihao Zhang, Ziqian Shao, Wenhan Luo, Bjorn Stenger, Tong Lu, Tae-Kyun Kim, Wei Liu, and Hongdong Li. Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. *International Journal of Computer Vision*, pages 1–23, 2024. 2

[14] Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration. *IEEE Transactions on Image Processing*, 2024. 2

[15] Ronald Yu. A tutorial on vaes: From bayes' rule to lossless compression, 2020. 1

[16] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5825–5835, 2023. 2

[17] Kaihao Zhang, Dongxu Li, Wenhan Luo, Wenqi Ren, Bjorn Stenger, Wei Liu, Hongdong Li, and Yang Ming-Hsuan. Benchmarking ultra-high-definition image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 1

[18] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25445–25455, 2024. 2

[19] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16180–16189, 2021. 1

[20] Lei Zhou, Chunlei Cai, Yue Gao, Sanbao Su, and Junmin Wu. Variational autoencoder for low bit-rate image compression. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 2617–2620, 2018. 1