OneRestore: A Universal Restoration Framework for Composite Degradation

Yu Guo[1,2,†], Yuan Gao[1,†], Yuxu Lu[3], Huilin Zhu[1,2], Ryan Wen Liu[1(✉)], and Shengfeng He[2(✉)]

[1] Wuhan University of Technology, Wuhan, China
wenliu@whut.edu.cn
[2] Singapore Management University, Singapore
shengfenghe@smu.edu.sg
[3] The Hong Kong Polytechnic University, Hong Kong, China
† Equal Contribution
https://github.com/gy65896/OneRestore

1 Overview

This supplementary material provides more details on model configuration and experimental results, which can be listed as follows:

- We offer additional configuration details about the Text/Visual Embedder and the OneRestore model.
- We elaborate on model training and inference procedures.
- We present more experimental results to verify the effectiveness and controllability of the proposed method.
- We discuss limitations and outline future research directions.

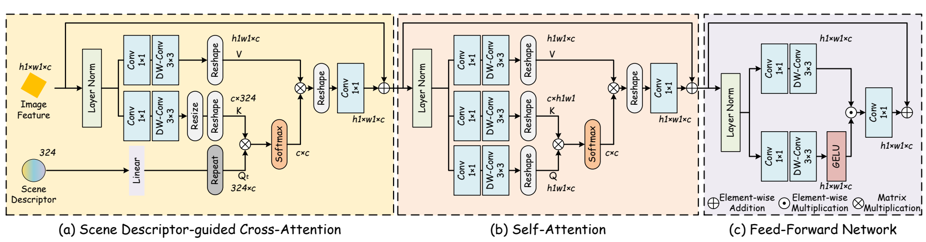


Fig. 1: Architecture of proposed Scene Descriptor-guided Transformer Block (SDTB).

2 Network Details

2.1 Transformer Block

Fig. 1 displays the details of the Scene Descriptor-guided Transformer Block (SDTB), which includes three parts: Scene Descriptor-guided Cross-Attention

Table 1: Configurations of network architecture. n, c, and $head$ denote the numbers of texts, channels, and attention heads, respectively. l is the length of the initial text embedding. k and s represent the kernel size and stride, respectively. bn and $relu$ denote the batch normalization and ReLU activation function, respectively.

Models	Layers	Configurations	Output Size
Text Embedder	Input1	Scene Description Text	1 string
	GloVe [26]	$n = 12, l = 300$	300
	MLP1	$c = 324, relu$	324
Visual Embedder	Input2	RGB Image	$h \times w \times 3$
	Resize	Uniform Size	$224 \times 224 \times 3$
	ResNet	ResNet-18	$7 \times 7 \times 512$
	Conv1	$c = 1024, k = 1, bn, relu$	$7 \times 7 \times 1024$
	Dropout	$rate = 0.35$	$7 \times 7 \times 1024$
	AvgPool	Global	$1 \times 1 \times 1024$
	Linear1	$c = 324$	324
OneRestore	Input3	RGB Image	$h \times w \times 3$
	Input4	Scene Descriptor	324
	Conv2	$c = 32, k = 1$	$h \times w \times 32$
	SDTB1	$c = 32, head = 8$	$h \times w \times 32$
	Down1	Maxpool $k = 3, s = 2$, Conv $c = 64, k = 1$	$\frac{h}{2} \times \frac{w}{2} \times 64$
	SDTB2	$c = 64, head = 8$	$\frac{h}{2} \times \frac{w}{2} \times 64$
	Down2	Maxpool $k = 3, s = 2$, Conv $c = 128, k = 1$	$\frac{h}{4} \times \frac{w}{4} \times 128$
	SDTB3	$c = 128, head = 8$	$\frac{h}{4} \times \frac{w}{4} \times 128$
	Down3	Maxpool $k = 3, s = 2$, Conv $c = 256, k = 1$	$\frac{h}{8} \times \frac{w}{8} \times 256$
	SDTB4	$c = 256, head = 8$	$\frac{h}{8} \times \frac{w}{8} \times 256$
	SDTB5	$c = 256, head = 8$	$\frac{h}{8} \times \frac{w}{8} \times 256$
	SDTB6	$c = 256, head = 8$	$\frac{h}{8} \times \frac{w}{8} \times 256$
	Addition1	Down3 + SDTB6	$\frac{h}{8} \times \frac{w}{8} \times 256$
	SDTB7	$c = 256, head = 8$	$\frac{h}{8} \times \frac{w}{8} \times 256$
	Up1	Bilinear Interpolation, Conv $c = 128, k = 1$	$\frac{h}{4} \times \frac{w}{4} \times 128$
	Addition2	Up1 + SDTB3	$\frac{h}{4} \times \frac{w}{4} \times 128$
	SDTB8	$c = 128, head = 8$	$\frac{h}{4} \times \frac{w}{4} \times 128$
	Up2	Bilinear Interpolation, Conv $c = 64, k = 1$	$\frac{h}{2} \times \frac{w}{2} \times 64$
	Addition3	Up2 + SDTB2	$\frac{h}{2} \times \frac{w}{2} \times 64$
	SDTB9	$c = 64, head = 8$	$\frac{h}{2} \times \frac{w}{2} \times 64$
	Up3	Bilinear Interpolation, Conv $c = 32, k = 1$	$h \times w \times 32$
	Addition4	Up3 + SDTB1	$h \times w \times 32$
	SDTB10	$c = 32, head = 8$	$h \times w \times 32$
	Conv3	$c = 3, k = 1$	$h \times w \times 3$
	Addition5	Conv3 + Input3	$h \times w \times 3$

(SDCA), Self-Attention (SA), and Feed-Forward Network (FFN). Specifically, the SDCA considers the image feature and degraded scene descriptor as input. The values $\mathbf{V}$ and keys $\mathbf{K}$ are generated by processing image features using two modules with the same structure but different weights, concatenating 1×1 Convolution (Conv) with 3×3 DepthWise Convolution (DW-Conv). As for the scene descriptor, we use a linear layer to produce the scene description query $\mathbf{Q}_t$. Then, we adjust the image size to make the number of tokens in $\mathbf{K}$ to be consistent with $\mathbf{Q}_t$. The SA and FFN adopt modules proposed by Restormer [44]. By incorporating SDCA, our model can facilitate restored orientation control by changing different embeddings.

Algorithm 1 Text/Visual Embedder Training

Input: Visual image and scene description text label pairs (I_v, y), all texts $\mathcal{S}_t$.

1: **repeat**

2: $e_t = Emb_t(\mathcal{S}_t)$ ▷ Get text embeddings.

3: $e_v = Emb_v(I_v)$ ▷ Get visual embeddings.

4: Perform the gradient descent step by

$$\mathcal{L}_{cross} = -\frac{1}{N_v}\sum_{i=1}^{N_v}\sum_{j=1}^{N_t} y_{ij}\log\left(S(e_{v_i}, e_{t_j})\right).$$

▷ N_v and N_t are the numbers of visual and text embeddings, respectively. y_i is the truth label of i-th sample. When the truth label is j, $y_{ij} = 1$, otherwise $y_{ij} = 0$.

5: **until** converged

6: **return** Emb_t, Emb_v

Algorithm 2 OneRestore Model Training

Input: Clear, input degraded, and other degraded image pairs $(J, I, \{I_o\})$, scene description text of input degraded image $\mathcal{S}'_t \in \mathcal{S}_t$, text embedder Emb_t.

1: **repeat**

2: $e'_t = Emb_t(\mathcal{S}'_t)$ ▷ Get the scene descriptor with the frozen text embedder.

3: $\hat{J} = OneRestore(I, e'_t)$ ▷ Get the restored image.

4: Perform the gradient descent step by

$$\mathcal{L} = \alpha_1\mathcal{L}_1^s(J,\hat{J}) + \alpha_2\mathcal{L}_{\mathrm{M}}(J,\hat{J}) + \alpha_3\mathcal{L}_{\mathrm{c}}(J,\hat{J},I,\{I_o\}).$$

5: **until** converged

6: **return** $OneRestore$

2.2 Detailed Architecture

All network configurations and output sizes of each layer are displayed in Table 1. Specifically, the MLP layer used in the text embedder consists of a linear operator and a ReLU activation function. The visual embedder introduces a dropout layer to prevent overfitting. Our OneRestore employs three max-pooling operations to downsample images and utilizes the Scene Descriptor-guided Transformer Block (SDTB) to extract multi-scale image features. Finally, bilinear interpolation upsampling and skip connections are employed to fuse features of different scales and levels.

3 Model Training and Inference

The model training includes two steps: text/visual embedder training and OneRestore model training. Meanwhile, the model inference involves a manual mode based on text embedding and an automatic mode controlled by visual attributes.

3.1 Step1: Text/Visual Embedder Training.

The pseudo-code is shown in Alg. 1. Initially, we set RGB image and scene description text label pairs as (I_v, y) and all texts as $\mathcal{S}_t$, respectively. Subsequently,

Algorithm 3 Model Inference

Input: Input degraded image I, scene description text of input degraded image $\mathcal{S}_t' \in \mathcal{S}_t$ (non-essential).

1: **if** exist $\mathcal{S}_t'$ **then** ▷ Manual restoration.
2: $e_t' = Emb_t(\mathcal{S}_t')$
3: **else** ▷ Automatic restoration.
4: $e_v' = Emb_v(I)$
5: Calculate the cosine similarity between e_v' and each text embedding by
$$cos(e_v', e_t) = \gamma \cdot \frac{e_v' \cdot e_t^\top}{\|e_v'\|\|e_t\|}.$$
6: Select the text embedding with the highest similarity as the estimated scene descriptor e_t'.
7: **end if**
8: $\hat{J} = OneRestore(I, e_t')$
9: **return** $\hat{J}$

the text embedder Emb_t is employed to generate 12 scene description embeddings e_t (line 2 in Alg. 1), and the visual embedder Emb_v is utilized to produce visual embeddings e_v from I_v (line 3 in Alg. 1). Finally, we calculate the cosine similarity between visual embeddings and text embeddings and perform gradient descent via the cross-entropy loss (line 4 in Alg. 1).

3.2 Step2: OneRestore Model Training.

The pseudo-code is shown in Alg. 2. We consider the clear, input degraded, and other degraded image pairs $(J, I, \{I_o\})$, scene description text of the input degraded image $\mathcal{S}_t' \in \mathcal{S}_t$, and pre-trained text embedder Emb_t as inputs. Emb_t is first used to generate scene descriptors e_t' (line 2 in Alg. 2). Input degraded images I and scene descriptors e_t' are jointly fed into our OneRestore to generate restored results (line 3 in Alg. 2). For the model converges, the total loss $\mathcal{L}$ is employed for gradient descent (line 3 in Alg. 2).

3.3 Model Inference.

As shown in Alg. 3, our model inference has two modes: manual and automatic. The necessary input is the degraded image I, and the optional input is the scene description text $\mathcal{S}_t' \in \mathcal{S}_t$. When the manual mode is selected, scene descriptors e_t' will be generated directly from $\mathcal{S}_t'$ using Emb_t (line 2 in Alg. 3). When using the automatic mode, Emb_v will first generate visual embedding e_v' (line 4 in Alg. 3). Then, the distance between the visual embedding e_v' and all text embeddings e_t is calculated by cosine similarity, and the top-1 text embedding is selected as the OneRestore input (line 5-6 in Alg. 3). The restored result will be generated by processing I and e_t' through our OneRestore model (line 8 in Alg. 3).

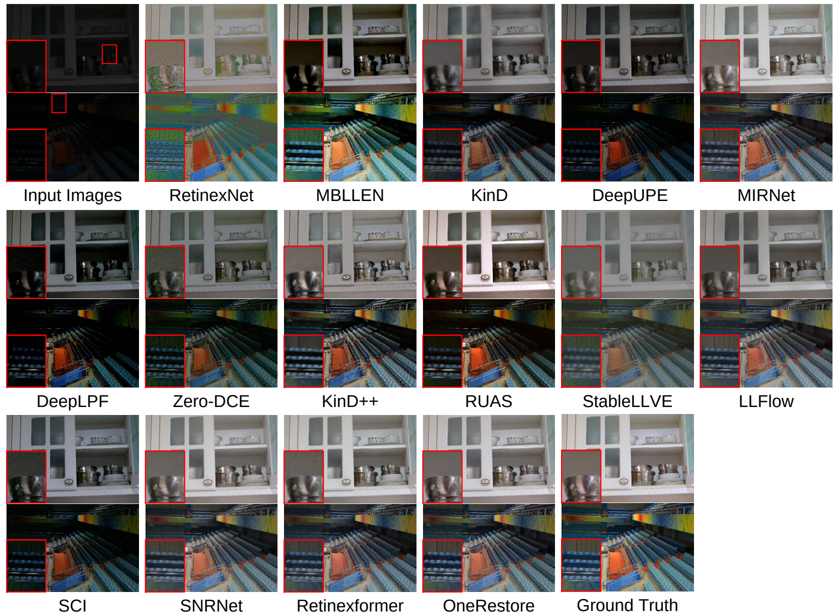


Fig. 2: Comparison of low-light enhancement on LOL dataset [37].

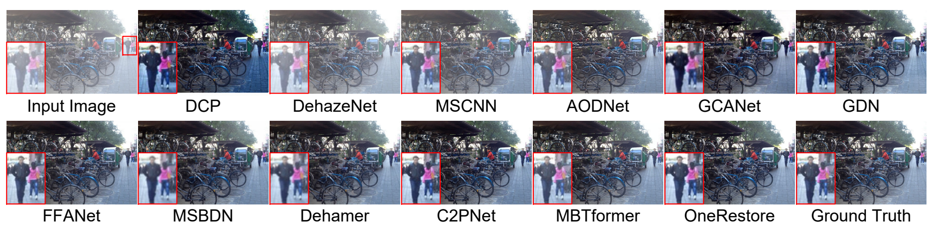


Fig. 3: Comparison of image dehazing on RESIDE dataset [15].

4 Experiments Results

In this section, we conduct a comparative analysis of our OneRestore against state-of-the-art (SOTA) models on benchmark datasets of related tasks, aiming to validate OneRestore's generalization ability. Subsequently, extensive experiments of image restoration, encompassing both synthetic and real-world scenarios, are presented to further illustrate the method's efficacy. Lastly, an analysis focusing on the model controllability is performed.

4.1 Comparison on Classic Benchmarks

We train and test our OneRestore on classic benchmark datasets for low-light enhancement (LOL [37]), dehazing (RESIDE-OTS [15]), draining (Rain1200 [48]),

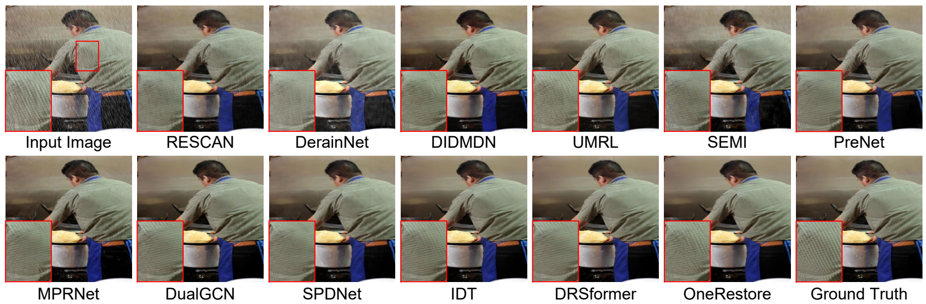


Fig. 4: Comparison of image dehazing on Rain1200 dataset [48].

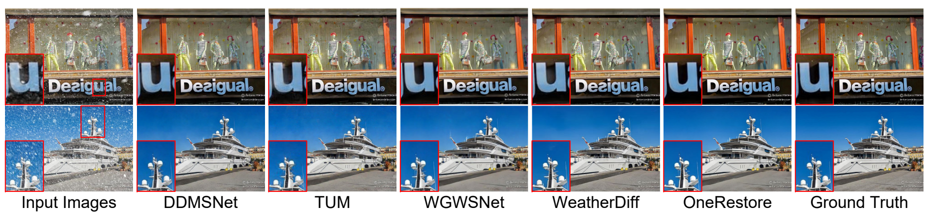


Fig. 5: Comparison of image desnowing on Snow100k-L dataset [21].

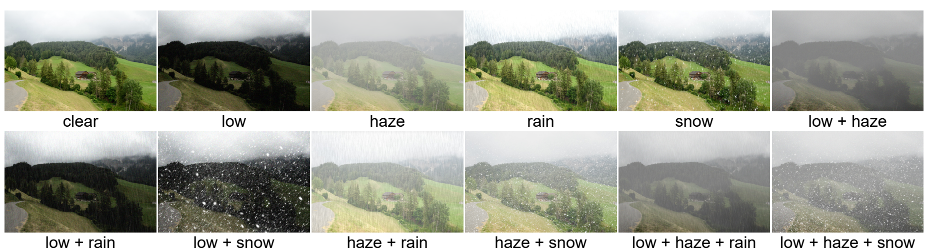


Fig. 6: Visualization of one test sample on CDD-11 dataset.

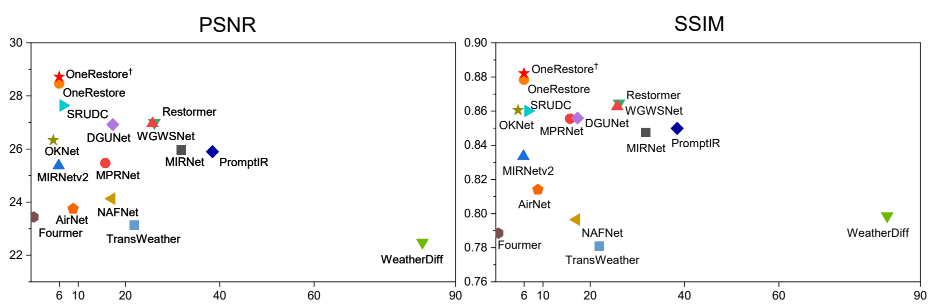


Fig. 7: Comparison of quantitative results and parameter quantities on CDD-11 dataset. OneRestore† indicates the manual mode, while OneRestore denotes the automatic mode.

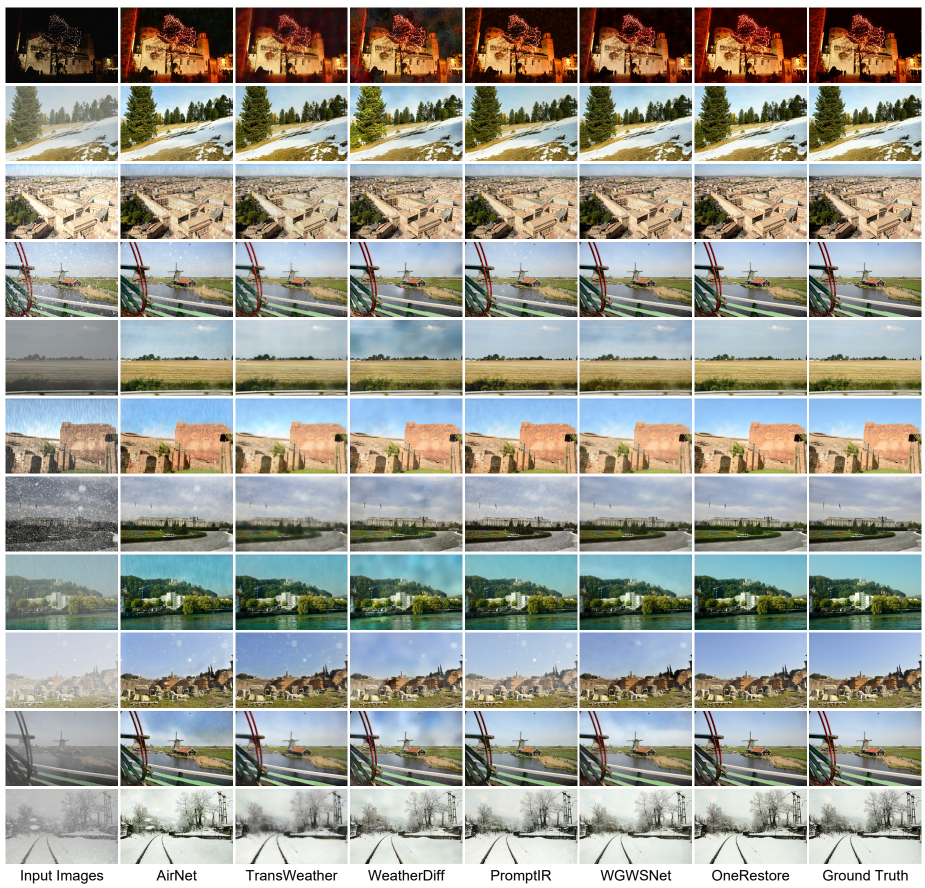


Fig. 8: Comparison of image restoration on *low*, *haze*, *rain*, *snow*, *low+haze*, *low+rain*, *low+snow*, *haze+rain*, *haze+snow*, *low+haze+rain*, and *low+haze+snow* synthetic samples.

Fig. 9: Comparison of image restoration on *low*, *haze*, *rain*, *snow*, *low+haze*, *low+rain*, *low+snow*, *haze+rain*, *haze+snow*, *low+haze+rain*, and *low+haze+snow* samples in real-world scenarios.

Table 2: Comparison of quantitative results on four benchmarks. Red, green, and blue indicate the best, second-best, and third-best results, respectively.

Methods	Venue & Year	PSNR ↑	SSIM ↑
Input		7.77	0.1914
RetinexNet [37]	BMVC2018	17.12	0.5920
MBLLEN [22]	BMVC2018	17.86	0.7247
KinD [51]	ACMMM2019	17.71	0.7734
MIRNet [45]	ECCV2020	24.10	0.8421
Zero-DCE [11]	CVPR2020	14.86	0.5624
KinD++ [50]	IJCV2021	17.75	0.7581
RUAS [19]	CVPR2021	16.40	0.5034
StableLLVE [47]	CVPR2021	17.36	0.7373
LLFlow [36]	AAAI2022	19.34	0.8388
SNRNet [40]	CVPR2022	24.61	0.8401
Retinexformer [2]	ICCV2023	25.15	0.8434
OneRestore		24.25	0.8564

(a) Enhancement results on the LOL dataset [37].

Methods	Venue & Year	PSNR ↑	SSIM ↑
Input		15.92	0.8139
DCP [12]	TPAMI2010	14.67	0.7825
DehazeNet [1]	TIP2016	20.95	0.8841
MSCNN [30]	ECCV2016	20.00	0.8779
AODNet [14]	ICCV2017	20.36	0.8945
GCANet [3]	WACV2019	22.60	0.8985
GDN [20]	ICCV2019	30.77	0.9808
FFANet [27]	AAAI2020	32.13	0.9792
MSBDN [7]	CVPR2020	30.23	0.9458
DeHamer [10]	CVPR2022	30.70	0.9457
C2PNet [52]	CVPR2023	34.05	0.9857
MB-TFormer [28]	ICCV2023	37.94	0.9899
OneRestore		35.58	0.9814

(b) Dehazing results on the RESIDE-OTS dataset [15].

Methods	Venue & Year	PSNR ↑	SSIM ↑
Input		22.16	0.6869
RESCAN [17]	ECCV2016	28.83	0.8430
DerainNet [8]	TIP2017	21.93	0.7814
DID-MDN [48]	CVPR2018	27.99	0.8627
UMRL [42]	CVPR2019	28.62	0.8706
SEMI [38]	CVPR2019	24.39	0.7622
PreNet [29]	CVPR2019	29.79	0.8811
MPRNet [46]	CVPR2021	31.32	0.8907
DualGCN [9]	AAAI2021	32.09	0.9181
SPDNet [43]	ICCV2021	32.84	0.9138
IDT [39]	TPAMI2022	33.13	0.9238
DRSformer [6]	CVPR2023	33.59	0.9274
OneRestore		32.89	0.9182

(c) Deraining results on the Rain1200 dataset [48].

Methods	Venue & Year	PSNR ↑	SSIM ↑
Input		18.68	0.6470
DerainNet [8]	TIP2017	19.18	0.7495
DehazeNet [1]	TIP2016	22.62	0.7975
DeepLab [4]	TPAMI2017	21.29	0.7747
RESCAN [17]	ECCV2018	26.08	0.8108
SPANet [35]	CVPR2019	23.70	0.7930
AIO [16]	CVPR2020	28.33	0.8820
DDMSNet [49]	TIP2021	28.85	0.8772
TransWeather [31]	CVPR2022	29.31	0.8879
TUM [5]	CVPR2022	26.90	0.8321
WGWSNet [53]	CVPR2023	28.94	0.8758
WeatherDiff [25]	TPAMI2023	30.09	0.9041
OneRestore		30.24	0.8947

(d) Desnowing results on the Snow100k dataset [21].

and desnowing (Snow100k [21]), and conduct a comprehensive comparison of our method with SOTA methods in each task. The quantitative results and visual performance of each method are reported in Table 2 and Figs. 2-5, respectively. Given that our OneRestore is designed to address composite degradations, it is inevitable that it cannot surpass the current optimal SOTA methods for all benchmarks. Nonetheless, the comparable performance of the recovery results unequivocally demonstrates the model's robust feature extraction capabilities. It is noteworthy that our method boasts a relatively smaller number of parameters (only 5.98M) and faster inference speed (0.0115s for processing a 1080×720 image on our PC), further reinforcing its practical significance.

4.2 More Results on CDD-11 Dataset

We conduct experiments on the proposed Composite Degradation Dataset (CDD-11), where an example of a clear image and its corresponding 11 types of degraded samples are shown in Fig. 6. Quantitative comparisons of all methods on our CDD-11 dataset are shown in Fig. 7, demonstrating that our OneRestore can achieve a balance between quantitative results and parameter quantities.

Furthermore, Fig. 8 displays more restoration cases on the 11 types of degraded images from the CDD-11 dataset. It is clear that current SOTA methods are limited in their ability to handle all types of image degradation and can produce unstable results. In contrast, our method is designed to be versatile

Table 3: Comparison of quantitative results on four real benchmarks. The best results are in **bold**, and the second-best are with underline.

Methods	Venue & Year	NIQE ↓	PIQE ↓
RUAS [19]	IJCV2021	7.77	19.64
SCI [23]	CVPR2021	3.97	**16.35**
SNRNet [40]	CVPR2022	4.49	20.71
OneRestore		**3.93**	16.69

(a) Enahncement results on the NPE dataset [34].

Methods	Venue & Year	NIQE ↓	PIQE ↓
MSBDN [7]	CVPR2020	4.77	26.14
DeHamer [10]	CVPR2022	5.34	31.93
C2PNet [52]	CVPR2023	5.03	25.09
OneRestore		**4.58**	**23.91**

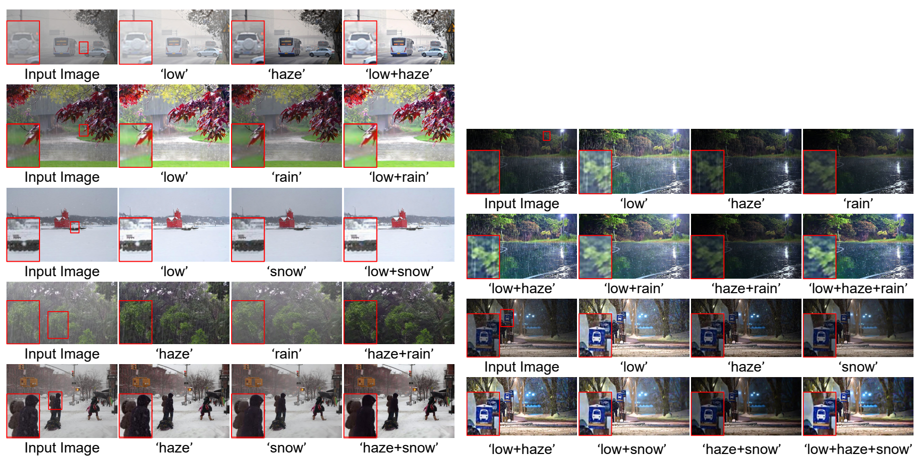
(b) Dehazing results on the RTTS dataset [15].

Methods	Venue & Year	NIQE ↓	PIQE ↓
MPRNet [46]	CVPR2021	3.55	20.78
DualGCN [9]	AAAI2021	**3.27**	15.15
MFDNet [33]	TIP2023	3.37	18.86
OneRestore		3.32	**14.46**

(c) Deraining results on the RS dataset [41].

Methods	Venue & Year	NIQE ↓	PIQE ↓
DRT [18]	CVPRW2022	3.93	12.00
TUM [5]	CVPR2022	3.13	9.14
UMWT [13]	ECCVW2022	3.34	12.29
OneRestore		**3.00**	**9.12**

(d) Desnowing results on the Snow100k-R dataset [21].



(a) Double degeneration coexistence examples (b) Triple degenerate coexistence examples

Fig. 10: Comparison of image restoration on real-world scenarios by using different text descriptors.

and can effectively handle a wide range of degradation scenarios. By incorporating scene descriptors to identify degradation situations, our method can create high-quality images with rich details.

4.3 More Results on Real-World Dataset

To assess the robustness of our model, which was trained using the CDD-11 dataset, against complex degradation scenarios, we conducted extensive real-world image restoration experiments. Our quantitative analysis on four distinct real-world benchmarks utilized two no-reference quality assessment metrics: the Natural Image Quality Evaluator (NIQE) [24] and the Perception-based Image Quality Evaluator (PIQE) [32]. It was noted that these benchmarks were specifically chosen to represent the four degradation challenges we aimed to address: the Naturalness Preserved Enhancement dataset (NPE) [34] for low-light enhance-

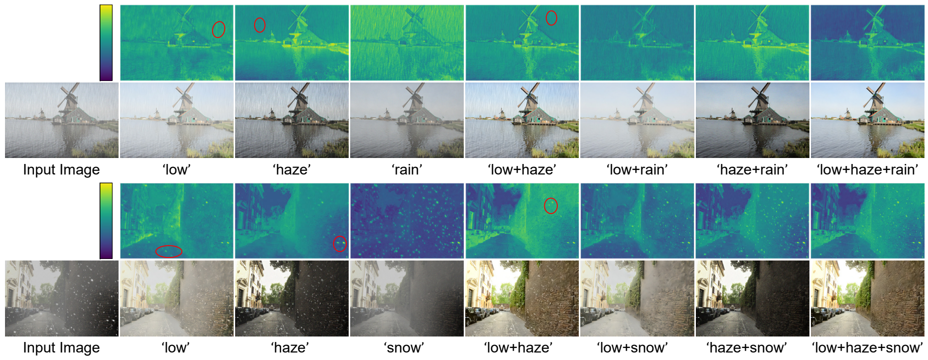


Fig. 11: Comparison of the SDTB10 output feature maps and restoration results based on different texts on *low+haze+rain* and *low+haze+snow* synthetic samples. Due to the intricate interplay of different degradation factors, our model tends to slightly eliminate features not explicitly mentioned in the description text. As indicated in the red oval regions, this phenomenon is often observed in the context of rain and snow degradation. Nevertheless, our model already exhibits impressive controllability.

ment, the RESIDE Real-world Task-driven Testing Set (RTTS) [15] for image dehazing, Yang's dataset (RS) [41] for image deraining, and the real Snow100k dataset (Snow100k-R) [21] for image desnowing. As shown in Table 3, our quantitative assessment comparison with SOTA methods substantiates the exceptional performance of our model in restoring these degradation datasets. Moreover, we show additional real-world image restoration cases in Fig. 9. It is evident that models trained on the CDD-11 dataset exhibit remarkable efficacy in real-world scenarios, thereby validating the proposed synthesis strategy as an effective simulation tool for real composite degradation scenarios. Notably, both WGWSNet and our method, which introduces degraded descriptions, demonstrate superior visual effects. Our OneRestore establishes a more stringent lower-bound constraint by applying the composite degradation restoration loss, enabling its restoration results to closely approximate clear scenarios.

4.4 More Results of Different Scene Descriptors

To verify the controllability of our model, we conduct a comparison experiment on seven distinct types of images employing varying text embeddings. The resulting visual outcomes are shown in Fig. 10. By introducing scene description texts, we can direct the model's attention to different types of degradation factors, thereby achieving controllable restoration. Furthermore, the input of multiple words in conjunction enables our OneRestore to achieve better visual performance in complex composite degradation scenarios. Fig. 11 further illustrates the impact of using different scene description texts on the feature maps output by SDTB10 and restoration results under *low+haze+rain* and *low+haze+snow* synthetic samples. The introduction of scene description texts can serve as prior

Fig. 12: Failure cases of image restoration.

information, enabling the model to establish the distinct representation space for correspondence degradation. This characteristic positions our OneRestore as a superior solution in addressing complex, composite degradations.

5 Limitation and Future Work

Extensive experiments have been conducted to verify the proposed method's performance on both synthetic and real composite degradation scenarios. Impressive results have demonstrated that the proposed unified imaging model effectively simulates real-world complex degradation. Moreover, the introduction of degraded scene descriptors and the proposed composite degradation restoration loss can aid the model in adjusting to complex situations. However, our approach still has failure recovery cases, as illustrated in Fig. 12. Although some degradation factors are well suppressed or eliminated, color abnormalities and distortions still exist in the restoration results. Based on the observation of failure results, we elaborate on the limitations and future research directions of this work in more detail, which can be introduced as follows:

– Disparities between synthetic and real-world data may constrain the image restoration ability of our approach in some real degradation scenarios, such as high-density corruption scenarios.
– More degradation types are not considered in this work, such as raindrops, moiré, shadows, etc. However, the proposed universal framework provides a solid solution, making it possible to cover a wider range of degradation types by incorporating additional scene descriptors for directional restoration.
– While the proposed approach provides restoration control over the degradation type, achieving restoration intensity control for each type remains a direction for future research.
– Ensuring model robustness while reducing computational overhead is also one of our future considerations.

References

1. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE TIP **25**(11), 5187–5198 (2016)

2. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: ICCV. pp. 12504–12513 (2023)
3. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: WACV. pp. 1375–1383 (2019)
4. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE TPAMI **40**(4), 834–848 (2017)
5. Chen, W.T., Huang, Z.K., Tsai, C.C., Yang, H.H., Ding, J.J., Kuo, S.Y.: Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In: CVPR. pp. 17653–17662 (2022)
6. Chen, X., Li, H., Li, M., Pan, J.: Learning a sparse transformer network for effective image deraining. In: CVPR. pp. 5896–5905 (2023)
7. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: CVPR. pp. 2157–2167 (2020)
8. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: A deep network architecture for single-image rain removal. IEEE TIP **26**(6), 2944–2956 (2017)
9. Fu, X., Qi, Q., Zha, Z.J., Zhu, Y., Ding, X.: Rain streak removal via dual graph convolutional network. In: AAAI. vol. 35, pp. 1352–1360 (2021)
10. Guo, C.L., Yan, Q., Anwar, S., Cong, R., Ren, W., Li, C.: Image dehazing transformer with transmission-aware 3d position embedding. In: CVPR. pp. 5812–5820 (2022)
11. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: CVPR. pp. 1780–1789 (2020)
12. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE TPAMI **33**(12), 2341–2353 (2010)
13. Kulkarni, A., Phutke, S.S., Murala, S.: Unified transformer network for multi-weather image restoration. In: ECCV. pp. 344–360 (2022)
14. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: ICCV. pp. 4770–4778 (2017)
15. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE TIP **28**(1), 492–505 (2018)
16. Li, R., Tan, R.T., Cheong, L.F.: All in one bad weather removal using architectural search. In: CVPR. pp. 3175–3185 (2020)
17. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: ECCV. pp. 254–269 (2018)
18. Liang, Y., Anwar, S., Liu, Y.: Drt: A lightweight single image deraining recursive transformer. In: Proc. IEEE CVPR. pp. 589–598 (2022)
19. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: CVPR. pp. 10561–10570 (2021)
20. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: ICCV. pp. 7314–7323 (2019)
21. Liu, Y.F., Jaw, D.W., Huang, S.C., Hwang, J.N.: Desnownet: Context-aware deep network for snow removal. IEEE TIP **27**(6), 3064–3073 (2018)
22. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: Low-light image/video enhancement using cnns. In: BMVC. vol. 220, p. 4 (2018)

23. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: CVPR. pp. 5637–5646 (2022)
24. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE SPL **20**(3), 209–212 (2012)
25. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE TPAMI **45**(8), 10346–10357 (2023)
26. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: EMNLP. pp. 1532–1543 (2014)
27. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: AAAI. vol. 34, pp. 11908–11915 (2020)
28. Qiu, Y., Zhang, K., Wang, C., Luo, W., Li, H., Jin, Z.: Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In: ICCV. pp. 12802–12813 (2023)
29. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: A better and simpler baseline. In: CVPR. pp. 3937–3946 (2019)
30. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: ECCV. pp. 154–169 (2016)
31. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: CVPR. pp. 2353–2363 (2022)
32. Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: NCC. pp. 1–6 (2015)
33. Wang, Q., Jiang, K., Wang, Z., Ren, W., Zhang, J., Lin, C.W.: Multi-scale fusion and decomposition network for single image deraining. IEEE TIP **33**, 191–204 (2023)
34. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE TIP **22**(9), 3538–3548 (2013)
35. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: CVPR. pp. 12270–12279 (2019)
36. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.: Low-light image enhancement with normalizing flow. In: AAAI. vol. 36, pp. 2604–2612 (2022)
37. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: BMVC. pp. 1–12 (2018)
38. Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y.: Semi-supervised transfer learning for image rain removal. In: CVPR. pp. 3877–3886 (2019)
39. Xiao, J., Fu, X., Liu, A., Wu, F., Zha, Z.J.: Image de-raining transformer. IEEE TPAMI **45**(11), 12978–12995 (2022)
40. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: CVPR. pp. 17714–17724 (2022)
41. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: CVPR. pp. 1357–1366 (2017)
42. Yasarla, R., Patel, V.M.: Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In: CVPR. pp. 8405–8414 (2019)
43. Yi, Q., Li, J., Dai, Q., Fang, F., Zhang, G., Zeng, T.: Structure-preserving deraining with residue channel prior guidance. In: ICCV. pp. 4238–4247 (2021)
44. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR. pp. 5728–5739 (2022)

45. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: ECCV. pp. 492–511 (2020)
46. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR. pp. 14821–14831 (2021)
47. Zhang, F., Li, Y., You, S., Fu, Y.: Learning temporal consistency for low light video enhancement from single images. In: CVPR. pp. 4967–4976 (2021)
48. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: CVPR. pp. 695–704 (2018)
49. Zhang, K., Li, R., Yu, Y., Luo, W., Li, C.: Deep dense multi-scale network for snow removal using semantic and depth priors. IEEE TIP **30**, 7419–7431 (2021)
50. Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. IJCV **129**, 1013–1037 (2021)
51. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: ACM MM. pp. 1632–1640 (2019)
52. Zheng, Y., Zhan, J., He, S., Dong, J., Du, Y.: Curricular contrastive regularization for physics-aware single image dehazing. In: CVPR. pp. 5785–5794 (2023)
53. Zhu, Y., Wang, T., Fu, X., Yang, X., Guo, X., Dai, J., Qiao, Y., Hu, X.: Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In: CVPR. pp. 21747–21758 (2023)