CVPR #9427

UniRestore: Unified Perceptual and Task-Oriented Image Restoration Model Using Diffusion Prior

Anonymous CVPR submission

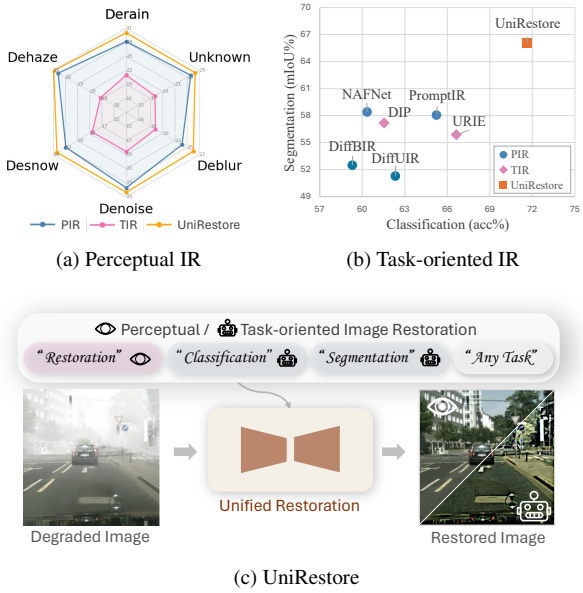
Paper ID 9427

Abstract

Image restoration aims to recover content from inputs degraded by various factors, such as adverse weather, blur, and noise. Perceptual Image Restoration (PIR) methods improve visual quality but often do not support downstream tasks effectively. On the other hand, Task-oriented Image Restoration (TIR) methods focus on enhancing image utility for high-level vision tasks, sometimes compromising visual quality. This paper introduces UniRestore, a unified image restoration model that bridges the gap between PIR and TIR by using a diffusion prior. The diffusion prior is designed to generate images that align with human visual quality preferences, but these images are often unsuitable for TIR scenarios. To solve this limitation, UniRestore utilizes encoder features from an autoencoder to adapt the diffusion prior to specific tasks. We propose a Complementary Feature Restoration Module (CFRM) to reconstruct degraded encoder features and a Task Feature Adapter (TFA) module to facilitate adaptive feature fusion in the decoder. This design allows UniRestore to optimize images for both human perception and downstream task requirements, addressing discrepancies between visual quality and functional needs. Integrating these modules also enhances UniRestore's adaptability and efficiency across diverse tasks. Extensive experiments demonstrate the superior performance of UniRestore in both PIR and TIR scenarios.

1. Introduction

Image restoration [44, 57] aims to restore content degraded by various factors, such as adverse weather, blur, and noise. These factors often reduce image perceptual visibility [29, 62] and negatively impact the performance of high-level vision applications, such as object detection [68, 76] and semantic segmentation [45, 51]. Various restoration methods have been developed and studied over the past decades to address this wide range of image restoration challenges. These include methods focused on improving perceptual quality [46, 74] as well as those designed to enhance the performance of downstream tasks [54, 71].



(a) Perceptual IR (b) Task-oriented IR

(c) UniRestore

Figure 1. **Illustration of UniRestore's capabilities:** (a) PIR - Comparison with existing methods (*e.g.*, URIE [54], PromptIR [44]) under adverse conditions. (b) TIR - Demonstrating UniRestore's robustness for downstream tasks such as classification and segmentation. (c) Unified Restoration - UniRestore's versatility in addressing PIR and TIR.

Perceptual image restoration (PIR) algorithms focus on improving visual clarity and fidelity of images by removing or reducing visible artifacts that affect their aesthetic quality. These methods focus on dealing with noise reduction [8, 74], low-light enhancement [72, 78], deblurring [40, 46], dehazing [6, 74], deraining [59, 69], and desnowing [7, 37, 61]. While these algorithms can enhance the visual quality of images, they do not always improve performance in downstream tasks [4, 5, 21]. This is because the factors contributing to visual quality often differ from those determining recognition quality [43, 54, 71].

On the other hand, task-oriented image restoration (TIR) [54, 71] is specifically designed to optimize images for applications that rely on computer vision, such as object detection [68, 76], classification [53, 54], and autonomous driving systems [45, 51]. This approach aligns the restoration process with the specific requirements of neural network models used in these applications, ensuring that the restored images are suitable for effective machine interpretation. Although these methods improve the performance of downstream tasks, they often produce results that are less visually appealing [36, 76].

Based on the analysis above, existing image restoration algorithms often face a trade-off, highlighting the challenge of balancing technical functionality with aesthetic quality. This dual functionality is crucial because real-world applications often require restoration processes to be specifically adapted to different scenarios. Thus, developing a model that can simultaneously enhance perceptual quality and improve performance for downstream tasks is essential. Such a unified image restoration framework can effectively perform across diverse settings, reducing system redundancy and boosting operational efficiency, as shown in Figure 1.

In this paper, we propose a unified image restoration paradigm, UniRestore, which simultaneously improves the performance of downstream tasks and the human perceptual quality of degraded images. UniRestore leverages a diffusion prior [47] as the backbone, recognized for its generative capabilities in producing high-quality images. However, these images are typically optimized for human aesthetics, which may not align with downstream task requirements. UniRestore addresses this limitation by adapting the diffusion prior to meet both perceptual and functional needs, enabling the model to effectively bridge the gap between visual quality and downstream task performance.

To bridge this gap, we exploit the encoder features from the autoencoder within the diffusion model as complementary elements to tailor the diffusion prior to specific tasks. We introduce a Complementary Feature Restoration Module designed to reconstruct degraded features in the encoder. Subsequently, we propose a Task Feature Adapter, which harmonizes the diffusion features with the restored features within the decoder for various downstream tasks. Given the diversity of downstream tasks and the frequent necessity to adapt these tasks within existing models, the TFA module offers extendability to accommodate new tasks. Extensive experiments validate UniRestore's effectiveness, demonstrating enhancements in both PIR and TIR, with the potential for expansion to additional downstream tasks.

The contributions of this work are:

- We introduce UniRestore, a unified image restoration model that addresses both perceptual image restoration and task-oriented image restoration within a single framework. Experimental results show that UniRestore surpasses existing methods in both visual quality and downstream task performance.
- We propose two components for UniRestore: the CFRM and the TFA. These modules work together to adaptively complement the diffusion prior, enabling simultaneous restoration across diverse tasks.

2. Related Work

2.1. Perceptual Image Restoration

Perceptual image restoration aims to enhance the visual quality of images as perceived by humans, and it can be categorized into *single* degradation and *multiple* degradation restoration. Early work in single degradation restoration, such as SRCNN [13], focused on specific degradations to improve image quality, leading to significant advancements in super-resolution [25, 32, 79], denoising [8, 74], dehazing [6, 74], deraining [59, 69], low-light enhancement [72, 78], and deblurring [40, 46]. To tackle multiple degradations simultaneously, methods like MPRNet [74] and NAFNet [2] introduced unified solutions. Recently, transformer-based approaches, such as SwinIR [30] and Restormer [73], have gained traction for their versatility. Additionally, holistic approaches like All-in-One [29], TransWeather [57], and PromptIR [44] have focused on improving visual quality across a wide range of conditions while providing enhanced adaptability and performance. While these methods excel in enhancing visual quality, they do not always guarantee improved performance in downstream vision tasks.

2.2. Task-oriented Image Restoration

Task-oriented image restoration aims to enhance downstream task performance, as studies [4, 26, 71] show that image degradation significantly impairs high-level task accuracy. DDP [64] aligns feature representations between low- and high-quality images to improve classification accuracy. SFDUnet [70] employs self-feature distillation and uncertainty modeling to extract high-quality-like features from degraded images, enhancing recognition in challenging regions. URIE [54] integrates image enhancement and recognition tasks in an end-to-end framework to mitigate degradation effects. DIP [36] adapts image processing dynamically based on degradation factors for better recognition outcomes. VDR-IR [71] unifies semantic representations of diverse degraded images to recover intrinsic semantics effectively. While these methods enhance downstream task performance, they may result in images that are less visually pleasing.

2.3. Diffusion Model

Diffusion Models (DMs) [18] leverage a parameterized Markov chain to optimize the lower variational bound on

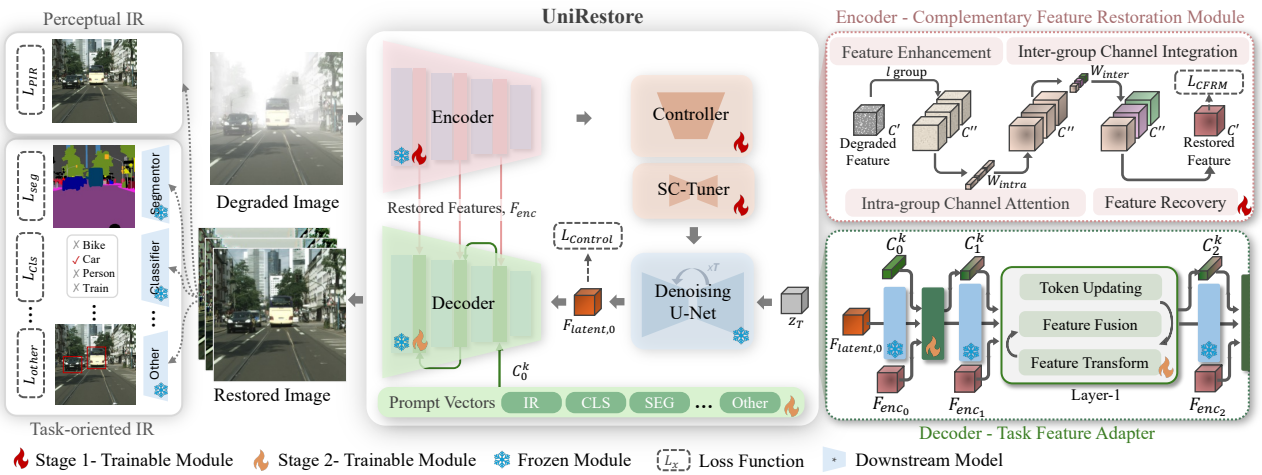


Figure 2. **Overview of UniRestore.** UniRestore augments the diffusion model by integrating a CFRM and a TFA. The training process is divided into two stages: In the first stage, CFRM, Controller, and SC-Tuner are trained to restore clear encoder and latent features. In the second stage, the TFA is trained to adapt the restored encoder features and latent features for various downstream tasks, using task-specific prompts at the decoder to control the output restoration.

the likelihood function, achieving state-of-the-art results in sample quality[23, 47, 55, 56] and various applications [52, 67, 81]. ControlNet [77] further refines this architecture by optimizing it for conditional image generation tasks. It incorporates conditions such as segmentation maps or textual prompts directly into the diffusion process through a specialized encoder, which modifies the latent representation of the input, enabling ControlNet to generate images that closely align with specified conditions [75, 80].

Recently, DMs have become integral to image restoration, including tasks such as super-resolution [28, 48], inpainting [10, 38, 47], and degradation restoration [42, 66, 83]. StableSR [60] effectively leverages diffusion priors for real-world super-resolution, resulting in superior reconstruction quality. DiffUIR [82] utilizes a specialized hourglass architecture to map degraded inputs to high-quality outputs, enhancing both global and local features effectively. DiffBIR [35] follows a two-stage restoration strategy, first addressing specific degradations and then refining the image quality through a diffusion generation model. Despite these advances, adopting pre-trained diffusion models for both PIR and TIR remains an open challenge.

3. Proposed Method

3.1. Architecture of UniRestore

UniRestore is built upon Stable Diffusion [47], leveraging its diffusion prior known for generating high-quality images. However, these images are optimized for human perception, which may not align with the requirements of machine vision tasks. To bridge this gap, we introduce two components: the Complementary Feature Restoration Module (CFRM) and the Task Feature Adapter (TFA). These modules adapt the diffusion prior to address diverse objectives, ensuring suitability for PIR and TIR tasks.

As illustrated in Figure 2, the input degraded image is processed through a modified encoder of VAE enhanced with the CFRM. The latent features produced by the final layer of this encoder are then fed into the Controller [77], equipped with an SC-Tuner [23]. The SC-Tuner, an enhanced module within the Controller architecture, integrates control signals efficiently with the Denoising U-Net. Subsequently, the noisy latent features are denoised to produce clear latent features, which are then passed to the decoder of the VAE augmented with the TFA. The restored features from the CFRM are input into the TFA, which adapts these features, enabling the decoder to generate outputs optimized for specific tasks.

3.2. Complementary Feature Restoration Module

The objective of the CFRM is to restore and enhance features within the encoder, thereby providing complementary inputs to the decoder. As shown in Figure 3, the CFRM is integrated into the output of each encoder layer and consists of four steps:

Feature Enhancement: The enhancement begins with a NAFBlock [9], followed by a convolutional layer and group normalization [65]. Input features have dimensions $(C^{'}, H, W)$, where $C^{'}$ denotes the number of channels, H the height, and W the width. These features are expanded fourfold to dimensions $(4C^{'}, H, W)$ and subsequently divided into l groups, resulting in $(C^{''}, H, W)$, where $C^{''} = 4C^{'}/l$.

Intra-group Channel Attention: In this stage, group con-

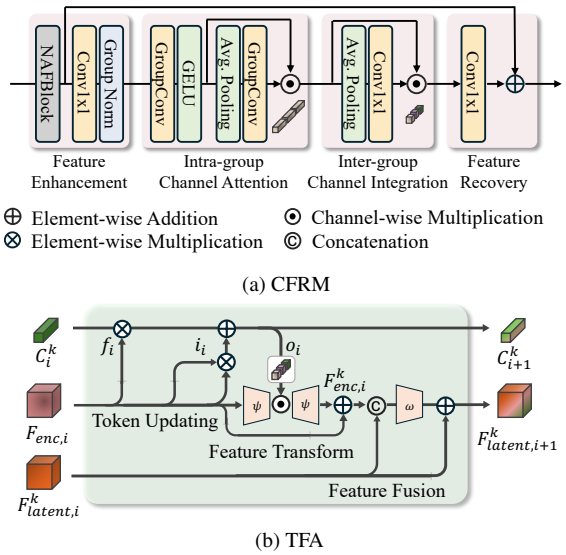


Figure 3. **Schematic diagrams of (a) the Complementary Feature Restoration Module and (b) the Task Feature Adapter.**

volution is employed to model learning from diverse degradation types. Initially, features are processed through a group convolution [24] and Gate Linear Unit (GELU) [17], followed by average pooling. Subsequently, a subsequent group convolution operation calculates the intra-group channel weights W^i_{intra}, each with dimensions $(C'', 1, 1)$, for groups indexed from 0 to l-1. This module sharpens the model's emphasis on pivotal intra-group features.

Inter-group Channel Integration: This module enhances contextual awareness by applying average pooling and a convolutional layer, yielding the inter-group channel weights W_{inter} with dimension $(l, 1, 1)$.

Feature Recovery: In the final stage, a convolution layer merges the refined group features. These are then combined with the enhanced features via a skip connection.

The output feature of the CFRM serves as a complementary feature to the subsequent TFA.

3.3. Task Feature Adapter

The TFA leverages the restored features from the CFRM to adapt the original diffusion features for various objectives. The core idea is to integrate the CFRM output features at each layer with the corresponding output from the decoder at the same scale, enabling feature fusion for the targeted purpose. To achieve this, a straightforward approach involves designing a distinct feature adapter module for each task, which is then individually optimized using the relevant objectives and datasets. However, given the wide range of TIR tasks, this approach requires extensive model parameters and faces scalability challenges.

To address this limitation, we draw inspiration from prompt tuning [22] and LSTM [19] and propose an efficient architecture that reuses TFA, relying only on a lightweight, learnable prompt vector to adapt to different tasks effectively. As shown in Figure 3, for each task, we initialize a lightweight learnable prompt vector C^k_0, where k represents the task index. This vector controls the weights of the CFRM features $F_{\text{enc},i}$ in each decoder layer i, dynamically combining them with the decoder's output features $F^k_{\text{latent},i}$. This prompt is updated within layer i and passed to the next layer $i+1$ as the input prompt for TFA. The procedure can be formulated as:

$$\begin{aligned} f_i &= \sigma\left(\vartheta\left(F_{\text{enc},i}, \theta_{f,i}\right)\right) \\ i_i &= \sigma\left(\vartheta\left(F_{\text{enc},i}, \theta_{i,i}\right)\right) \\ C^k_{i+1} &= f_i \otimes C^k_i + i_i \otimes \tanh\left(\vartheta\left(F_{\text{enc},i}, \theta_{c,i}\right)\right) \\ o_i &= \tanh(\xi(C^k_{i+1}, \theta_{o,i})) \\ F^k_{\text{enc},i} &= \psi(F_{\text{enc},i}, o_i, \theta_{t,i}) + F_{\text{enc},i} \\ F^k_{\text{latent},i+1} &= \omega((F^k_{\text{enc},i}, F^k_{\text{latent},i}), \theta_{l,i}) + F^k_{\text{latent},i} \end{aligned} \tag{1}$$

where σ and $\tanh$ represent the softmax and $\tanh$ activate function, $\otimes$ denotes the element-wise multiplication. $\vartheta\left(\cdot, \theta_{x,i}\right)$ is the prompt updating project function, involving instance normalization, convolution operation, GELU, and global average pooling. $\xi(\cdot, \theta_{o,i}))$ and $\omega(\cdot, \theta_{l,i}))$ are simple projection layers. $\psi(\cdot, c, \theta_{o,i})$ is a tuner-operator [23] with a channel attention based on condition c. Through this process, the restored features adapt to the diffusion features at each scale through different prompts, ultimately producing images suitable for downstream tasks.

3.4. Training Pipeline

Our training pipeline consists of three stages:

Stage 1 primarily focuses on adapting stable diffusion for the image restoration context. In this stage, we utilize the PIR dataset to train the CFRM, Controller [77], and SC-Tuner [23]. The loss function, $\mathcal{L}_{\text{CFRM}}$, is designed to enable the CFRM to learn the restoration of degraded features to their clear states. Specifically, clear latent features f^{Clear}_i are extracted from the i^{th} layer of the vanilla encoder using a clean image. Similarly, restored latent features f^{Restored}_i are derived by inputting a degraded image into the encoder integrated with the CFRM. The loss function is defined as:

$$\mathcal{L}_{\text{CFRM}} = \sum_{i=1}^{M} \lambda_i (f^{\text{Clear}}_i - f^{\text{Restored}}_i), \tag{2}$$

where M denotes the number of layers in the encoder, and λ_i represents the scaling weight for the i^{th} layer.

Additionally, the Controller and SC-Tuner are trained using the loss function $\mathcal{L}_{\text{Control}}$. This function is designed to align the clear latent features z_0 of the clear image with the reconstructed latent features $\hat{z}^t_0$ at any given sampled step t. The loss function can be expressed as:

$$\mathcal{L}_{\text{Control}} = \|z_0 - \hat{z}^t_0\|^2_2. \tag{3}$$

The total loss at Stage 1 is:

$$\mathcal{L}_{\text{Stage 1}} = \mathcal{L}_{\text{CFRM}} + \mathcal{L}_{\text{Control}}. \tag{4}$$

During this stage, TFA is not integrated into the decoder.

Stage 2 aims to optimize TFA to adapt the diffusion prior to different objectives. Therefore, CFRM, Controller, and SC-Tuner do not undergo parameter updates during this stage.

We optimize the network using objectives specific to each task, defined as:

$$\mathcal{L}_{\text{Stage 2}} = \sum_{i=1}^{N} \beta^i_{\text{Task}} \mathcal{L}^i_{\text{Task}}, \tag{5}$$

where N represents the number of tasks, and β^i_{Task} are the weighting coefficients that adjust the importance of each task-specific loss $\mathcal{L}^i_{\text{Task}}$ on the overall multi-task learning objective.

In this paper, we aim to address both PIR and TIR tasks simultaneously, selecting semantic segmentation and image classification as representative TIR tasks. For semantic segmentation, we employ cross-entropy loss using segmentation labels, while for classification, we also use cross-entropy loss but with class labels. For the PIR task, Mean Squared Error is applied to compare the reconstructed images against their corresponding ground truths.

The overall loss for Stage 2 of UniRestores formulated as:

$$\mathcal{L}_{\text{Stage 2}} = \beta_{\text{PIR}}\mathcal{L}_{\text{PIR}} + \beta_{\text{Seg}}\mathcal{L}_{\text{Seg}} + \beta_{\text{Cls}}\mathcal{L}_{\text{Cls}} \tag{6}$$

These tasks may originate from different datasets, and we distribute images from these tasks across each batch. Losses are calculated based on the input and its corresponding task before the model parameters are updated.

Introducing Additional Tasks. After training UniRestore with the two-stage process, adding more tasks in TIR requires only the introduction of a new task-specific prompt, which can then be optimized with the corresponding objective and training data. This process does not require data or loss functions from the original tasks, as only the new task-specific prompt is updated.

4. Implementation Details

To evaluate the effectiveness of the proposed UniRestore, experiments are conducted in PIR and TIR. Within TIR, image classification and semantic segmentation are chosen as downstream tasks. Detailed descriptions of the implementation details are provided in the Supplementary Material.

4.1. Training Dataset

We reference the dataset configurations from previous PIR [66] and TIR [54] studies. Specifically, we use a blend of the DIV2K [1], FlickrK [31], and OST [63] datasets for PIR tasks. For image classification, we randomly select 80,000 images from the training set of ImageNet [12], and for semantic segmentation, we use the training set from the Cityscapes datasets [11]. These datasets are synthesized with 15 types of degradation, including blur, noise, adverse weather conditions, etc., following the procedures outlined in [16] to create our training set.

4.2. Evaluation Dataset

For PIR evaluation, UniRestore is evaluated on the test set of DIV2K [1] using the same degradation synthesis procedure as in training. To further assess the robustness of UniRestore on unseen data, we utilize multiple benchmarks with synthetic degradations. These benchmarks include various tasks such as image denoising, which uses a composite dataset labeled 'Noise'—comprising Urban100 [20], BSD68 [39], CBSD68 [39], Kodak [39], McMaster [39], and Set12 [39]. Additionally, we utilize Rain100L [69] for deraining, RESIDE [27] for dehazing, UHDSnow [61] for desnowing, and GoPro [40] for deblurring.

In the TIR context, UniRestore is evaluated as follows: For classification, we sample 20,000 images from the test set ImageNet and utilize the entire CUB dataset [58] for validation of unseen data, applying the same degradation synthesis method as during training. For semantic segmentation, UniRestore is assessed on the test set of Cityscapes using identical degradation synthesis. Additionally, UniRestore is tested on the FoggyCityscapes dataset [49], specifically the subsets Fog1, Fog2, and Fog3, with results reported as the average across these three subsets. Further evaluations include the unseen ACDC dataset [50].

4.3. Evaluation Protocol

For the evaluation of PIR, we utilize multiple metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). For image classification, we measure performance using accuracy (ACC), and for semantic segmentation, we use the mean Intersection over Union (mIoU).

5. Experiments

5.1. Baselines

To evaluate the effectiveness of UniRestore, we compare it against multiple TIR methods, including DIP [36] and URIE [54], as well as PIR methods such as NAFNet [2], and PromptIR [44]. Furthermore, we include comparisons with diffusion-based approaches including DiffBIR [35] and DiffUIR [82]. We report results across two settings: First, models are trained only for their intended purpose (*i.e.*, PIR or TIR) using the corresponding dataset from our training set. Second, for a fair comparison and following the training pipeline of UniRestore, all baseline models are initially

Methods	*Seen Dataset* DIV2K [1]		*Unseen Datasets* Rain100L [69]		RESIDE [27]		UHDSnow [61]		Noise [20, 39]		GoPro [40]		*Average*	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DIP [36]	18.47	0.5810	22.65	0.7884	21.30	0.7819	19.03	0.8089	15.41	0.2494	23.08	0.8041	17.13	0.5734
DIP* [36]	18.62	0.5516	23.16	0.8097	19.83	0.7586	16.77	0.7830	14.51	0.2328	21.05	0.7624	16.28	0.5569
URIE [54]	17.72	0.5202	20.97	0.7293	18.30	0.7449	18.11	0.7626	18.57	0.5180	19.21	0.5683	18.81	0.6406
URIE* [54]	17.98	0.5967	19.97	0.6993	20.37	0.7694	16.18	0.7526	17.41	0.3624	18.57	0.4624	18.41	0.6071
NAFNet [2]	22.23	0.7905	24.57	0.8178	25.13	0.8632	20.71	0.8672	23.22	0.6951	22.18	0.8042	23.01	0.8063
NAFNet* [2]	19.81	0.7005	20.51	0.7314	21.24	0.8178	18.39	0.7958	20.38	0.6019	19.79	0.7293	20.02	0.7295
PromptIR [44]	23.90	0.8321	28.17	0.9034	27.26	0.8957	22.10	0.8877	23.72	0.7269	23.93	0.8221	24.85	0.8447
PromptIR* [44]	21.94	0.7421	24.76	0.8134	24.16	0.8317	19.13	0.8265	19.68	0.6283	20.18	0.7657	21.64	0.7680
DiffBIR [35]	22.76	0.8053	27.25	0.8695	26.97	0.8770	20.84	0.8785	23.67	0.7661	23.49	0.8076	24.16	0.8340
DiffBIR* [35]	18.32	0.6847	23.48	0.8143	23.13	0.8068	18.29	0.8167	21.59	0.6419	20.13	0.7413	20.82	0.7510
DiffUIR [82]	23.79	0.8397	28.25	0.9154	27.12	0.8820	20.74	0.8753	24.27	0.7481	23.93	0.8241	24.68	0.8474
DiffUIR* [82]	21.47	0.7742	25.44	0.8276	23.58	0.8174	18.62	0.8318	22.76	0.6691	21.71	0.7649	22.26	0.7808
UniRestore	**24.32**	**0.8434**	**30.02**	**0.9237**	**27.91**	**0.9043**	**23.44**	**0.8943**	**24.37**	**0.7811**	**25.94**	**0.8541**	**26.00**	**0.8668**

Table 1. **Performance comparison of existing methods on one seen and five unseen PIR datasets.**

Figure 4. **Qualitative analysis of perceptual image restoration:** A visual comparison on unseen datasets highlighting the performance improvements of the UniRestore over existing methods.

trained on the PIR training set and then fine-tuned on multiple downstream tasks (*i.e.*, PIR and TIR) using the loss function in (6), indicated by the suffix "*".

5.2. Perceptual Image Restoration

The results presented in Table 1 show that UniRestore achieves the best overall performance on the seen dataset. Additionally, it highlights UniRestore's generalizability across several unseen datasets, especially in high-resolution scenarios (UHDSnow) and dynamic scenes (GoPro). Moreover, TIR methods generally show limited performance in PIR task, as their learning objectives are optimized for specific downstream tasks. Additionally, in scenarios involving multiple downstream tasks, both PIR and TIR methods exhibit limited performance in due to the absence of a mechanism to learn different objectives simultaneously. We also present a qualitative comparison in Figure 4, where UniRestore reconstructs more details and delivers better visual quality.

5.3. Task-oriented Image Restoration

Image Classification. During training, UniRestore employs ResNet-50 [15] as the recognition model. For evaluation, both ResNet-50 [15] and ViT-B [14] serve as recognition backbones on the restored images. All recognition models are pre-trained on the training sets of their corresponding classification datasets without the degradation synthesis process.

Table 2 demonstrates UniRestore's effectiveness in enhancing image classification performance compared to ex-

Inputs	Seen Dataset ImageNet [12]		Unseen Dataset CUB [58]	
	ResNet-50 [15] ↑	ViT-B [14] ↑	ResNet-50 [15] ↑	ViT-B [14] ↑
LQ	*51.75*	*67.65*	*33.69*	*44.83*
DIP [36]	61.55	72.05	47.91	54.10
DIP* [36]	59.80	70.35	45.99	52.48
URIE [54]	66.65	73.95	49.64	57.24
URIE* [54]	65.20	72.15	46.89	54.93
NAFNet [2]	60.35	70.80	46.47	53.82
NAFNet* [2]	57.65	68.25	43.17	51.88
PromptIR [44]	65.25	73.90	49.52	58.04
PromptIR* [44]	64.05	73.00	48.52	57.39
DiffBIR [35]	59.30	68.05	41.68	52.38
DiffBIR* [35]	57.55	66.85	40.65	51.34
DiffUIR [82]	62.35	72.10	46.75	57.28
DiffUIR* [82]	61.15	71.60	45.44	56.31
UniRestore	**71.65**	**77.05**	**53.70**	**60.79**
HQ	*72.80*	*78.70*	*58.22*	*64.41*

Table 2. **Performance comparison of existing methods on seen and unseen datasets for image classification.** Results are reported in terms of accuracy.

Inputs	Seen Dataset Cityscapes [49]		FoggyCityscapes [49]	Unseen Dataset ACDC [50]
	DeepLabv3+ [3]	RefineNet-lw [41]	RefineNet-lw [41]	RefineNet-lw [41]
LQ	*40.36*	*40.75*	*65.20*	*28.30*
DIP [36]	57.17	57.67	67.81	38.19
DIP* [36]	51.81	50.35	67.16	32.98
URIE [54]	55.88	51.45	65.93	37.90
URIE* [54]	50.56	48.23	65.93	32.71
NAFNet [2]	58.41	58.19	66.06	37.59
NAFNet* [2]	51.91	53.29	65.40	36.03
PromptIR [44]	58.05	57.54	66.76	37.86
PromptIR* [44]	54.67	52.25	63.44	35.51
DiffBIR [35]	52.49	53.68	66.29	36.28
DiffBIR* [35]	48.90	48.56	63.26	33.12
DiffUIR [82]	51.28	51.46	66.24	35.78
DiffUIR* [82]	47.92	45.01	62.82	34.83
UniRestore	**66.05**	**65.73**	**70.77**	**39.27**
HQ	*75.64*	*75.66*	*75.66*	-

Table 3. **Performance comparison of existing methods on seen and unseen datasets for semantic segmentation.** Results are reported in terms of mIoU.

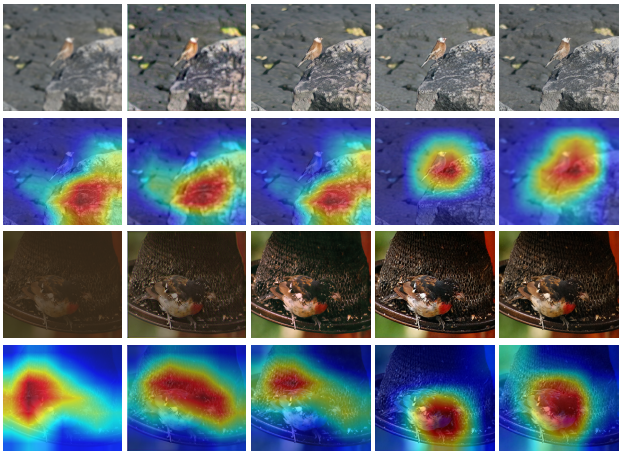


Figure 5. **Qualitative analysis of image classification.** The first and third rows display the input images, while the second and fourth rows show their corresponding activation maps.

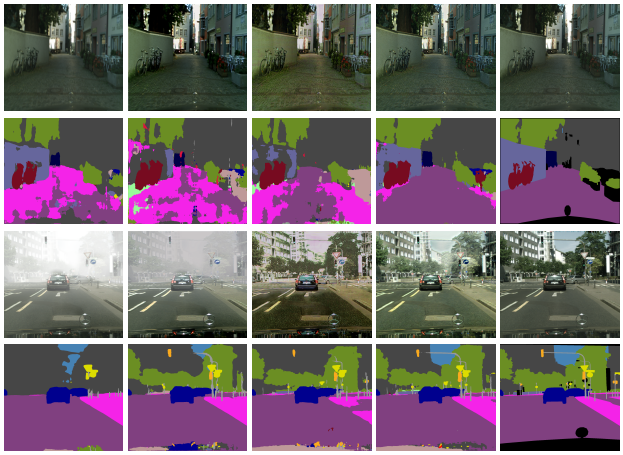


Figure 6. **Qualitative analysis of semantic segmentation.** The first and third rows present the input images, while the second and fourth rows display the corresponding segmentation results.

isting image restoration models, achieving accuracy comparable to that obtained using high-quality ground truth inputs (*i.e.*, HQ). Moreover, PIR methods show limited performance in classification tasks because they are optimized for human perception, which does not guarantee recognition accuracy. However, when these methods are trained on multiple downstream tasks, classification performance decreases. This decline may be due to these methods' potential inability to effectively handle multiple downstream tasks simultaneously. As detailed in Table 2, on the CUB dataset [58], UniRestore enhances classification accuracy by 20.01% for ResNet-50 [15] and by 15.96% for ViT-B [14] in scenarios involving unseen images. Figure 5 visually demonstrates that when images restored by UniRestore are used as inputs for recognition, their activation maps align more closely with those from high-quality ground truth images used as inputs for recognition.

Semantic Segmentation. In the training stage, we adopt DeepLabv3+ [3] as the segmentation model, while for evaluation, we employ both DeepLabv3+ [3] and RefineNet-lw [41]. Both models are pretrained on the training set of Cityscapes dataset [11]. As shown in Table 3, UniRestore achieves decent performance in semantic segmentation on both seen datasets and the unseen dataset. These results underscore UniRestore's ability to restore fine-grained details crucial for semantic segmentation tasks, demonstrating that the TFA modules integrate diffusion features with restored features to produce restored images with high-quality feature representations. Similar to the results observed in classification, PIR methods show limited performance in semantic segmentation compared to TIR methods, with performance decreasing in multiple downstream task scenar-

Methods	PIR PSNR ↑	Cls ACC ↑	Seg mIoU ↑
Baseline	19.35	57.65	46.76
UniRestore w/o CFRM	21.43	63.10	55.48
UniRestore w/o TFA	22.16	64.25	58.13
UniRestore	**24.32**	**71.65**	**66.05**

Table 4. **Effectiveness of CFRM and TFA in UniRestore.**

Methods	# of Tuned Parameters	PIR PSNR ↑	Cls ACC ↑	Seg mIoU ↑
Multiple Adapters	65.17M	23.06	68.95	64.64
Multiple TFAs	63.03M	**25.48**	71.20	65.78
UniRestore-SP	21.01M	23.91	70.05	64.99
UniRestore	21.03M	24.32	**71.65**	**66.05**

Table 5. **Comparative analysis of different TFA variants.**

Method	LQ	DIP [36]	PromptIR [44]	**UniRestore**
mAP ↑	45.63	54.29	50.61	**58.06**

Table 6. **Performance of Extendability Tested on Object Detection Using the RTTS [27] Dataset.**

ios. A qualitative comparison in Figure 6 shows that images restored by UniRestore enable segmentation models to generate more accurate object boundaries in the segmented results.

5.4. Ablation Study

To verify the effectiveness of the proposed modules, we conduct an ablation study. All experiments are evaluated on the DIV2K [1] test set for PIR, the ImageNet [12] test set for image classification, and the Cityscapes [49] test set for semantic segmentation, with degradation synthesis applied in the above tasks. These three sets are the same as we used in the evaluation dataset.

Effectiveness of Proposed Modules. We have established several configurations for our experiments: (i) Baseline: training the controller with a pre-trained Stable Diffusion model; (ii) UniRestore w/o CFRM: using the vanilla encoder features without any restoration; (iii) UniRestore w/o TFA: employing only the latent features from the denoising U-Net without adapting the encoder features; (iv) UniRestore: incorporating all modules. Table 4 demonstrates that both CFRM and TFA significantly enhance performance across PIR and TIR scenarios.

Investigation of TFA. To further investigate the effectiveness of the TFA module in multi-task scenarios, we conducted experiments with four variants: (i) Multiple Adapters: concatenates the output of the denoising U-Net with the restored features from CFRM and processes them through the same number of convolutional blocks as in TFA; (ii) Multiple TFAs: optimizes each task with its own TFA; (iii) UniRestore-SP: employs a single TFA with a single prompt for all tasks; (iv) UniRestore: utilizes one TFA with specific prompts for each task.

The results are shown in Table 5. The Multi-TFA outperforms Multi-Adapter, indicating the importance of dynamically fusing features by utilizing an updated prompt from the previous layer. Although UniRestore-SP requires the fewest parameters to be tuned, its performance is inferior to that of UniRestore, highlighting the significance of having a specific prompt for each task. UniRestore delivers performance comparable to Multi-TFA in PIR tasks and better performance in TIR tasks. This may be attributed to the single TFA block's ability to update using different objectives simultaneously, unlike in Multi-TFA, where each TFA is updated with only one objective. This also suggests that knowledge from various TIR tasks can potentially benefit other TIR tasks. Additionally, the number of parameters needing updates does not significantly increase with the number of tasks, indicating that the proposed TFA structure effectively balances scalability and performance.

Extendability Evaluation. To validate the extensibility of UniRestore, we incorporate an additional downstream task—object detection based on the model trained for PIR, image classification, and semantic segmentation. Specifically, we use a RetinaNet [33] pre-trained on the COCO [34] as the backbone. We randomly select 69,242 images from the COCO training set and synthetic the degradation as our training set. As outlined in Section 3.4, we utilize the current model configuration and update only with a new learnable prompt, optimizing it using the object detection loss. We then evaluate the object detection performance of UniRestore on the RTTS [27] dataset in comparison with other methods optimized concurrently for PIR, image classification, semantic segmentation, and object detection. As shown in Table 6, UniRestore achieves promising results in object detection. Additionally, compared to existing methods that require retraining models on complete task datasets, UniRestore only needs fine-tuning of a prompt with new downstream data and optimizing with its specific objective. This highlights UniRestore's potential for extensibility to other downstream tasks using our designed TFA module.

6. Conclusion

This paper introduces UniRestore, an approach capable of addressing PIR and TIR simultaneously. Building on diffusion models, we propose adapting diffusion features for diverse applications. To achieve this, we introduce a complementary feature restoration module that restores features within the encoder and a task feature adapter that dynamically and efficiently combines these restored features with diffusion features for downstream tasks. Experimental results validate the effectiveness and extendability of UniRestore, demonstrating its ability to alleviate the trade-offs associated with existing PIR and TIR methods.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, 2017. 5, 6, 8

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 2, 5, 6, 7

[3] Liang-Chieh Chen. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 7

[4] Wei-Ting Chen, I-Hsiang Chen, Chih-Yuan Yeh, Hao-Hsiang Yang, Hua-En Chang, Jian-Jiun Ding, and Sy-Yen Kuo. Rvsl: Robust vehicle similarity learning in real hazy scenes based on semi-supervised learning. In *ECCV*, 2022. 1, 2

[5] Wei-Ting Chen, I-Hsiang Chen, Chih-Yuan Yeh, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Sjdl-vehicle: Semi-supervised joint defogging learning for foggy vehicle re-identification. In *AAAI*, 2022. 1

[6] Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen Kuo. Pms-net: Robust haze removal based on patch map for single images. In *CVPR*, 2019. 1, 2

[7] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *ICCV*, 2021. 1

[8] Hamadi Chihaoui and Paolo Favaro. When self-supervised pre-training meets single image denoising. In *ICIP*, 2024. 1, 2

[9] Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *CVPRW*, 2022. 3

[10] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *CVPR*, 2022. 3

[11] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016. 5, 7

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 5, 7, 8

[13] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 2

[14] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 6, 7

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *ICPR*, 2016. 6, 7

[16] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 5

[17] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. 4

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NIPS*, 2020. 2

[19] S Hochreiter. Long short-term memory. *NC*, 1997. 4

[20] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015. 5, 6

[21] Zhi-Kai Huang, Wei-Ting Chen, Yuan-Chun Chiang, Sy-Yen Kuo, and Ming-Hsuan Yang. Counting crowds in bad weather. In *ICCV*, 2023. 1

[22] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *ECCV*, 2022. 4

[23] Zeyinzi Jiang, Chaojie Mao, Yulin Pan, Zhen Han, and Jingfeng Zhang. Scedit: Efficient and controllable image diffusion generation via skip connection editing. In *CVPR*, 2024. 3, 4

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *NIPS*, 2012. 4

[25] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 2

[26] Sohyun Lee, Taeyoung Son, and Suha Kwak. Fifo: Learning fog-invariant features for foggy scene segmentation. In *CVPR*, 2022. 2

[27] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *TIP*, 2018. 5, 6, 8

[28] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single image super-resolution with diffusion probabilistic models. *IJON*, 2022. 3

[29] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *CVPR*, 2020. 1, 2

[30] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, 2021. 2

[31] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017. 5

[32] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017. 2

[33] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 8

[34] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 8

[35] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 3, 5, 6, 7

[36] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke

Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. In *AAAI*, 2022. 2, 5, 6, 7, 8

[37] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *TIP*, 2018. 1

[38] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022. 3

[39] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001. 5, 6

[40] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 1, 2, 5, 6

[41] Vladimir Nekrasov, Chunhua Shen, and Ian Reid. Lightweight refinenet for real-time semantic segmentation. *arXiv preprint arXiv:1810.03272*, 2018. 7

[42] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *PAMI*, 2023. 3

[43] Yanting Pei, Yaping Huang, Qi Zou, Yuhang Lu, and Song Wang. Does haze removal help cnn-based image classification? In *ECCV*, 2018. 1

[44] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In *NIPS*, 2024. 1, 2, 5, 6, 7, 8

[45] Chenghao Qian, Mahdi Rezaei, Saeed Anwar, Wenjing Li, Tanveer Hussain, Mohsen Azarmi, and Wei Wang. Allweatherernet: Unified image enhancement for autonomous driving under adverse weather and lowlight-conditions. *arXiv preprint arXiv:2409.02045*, 2024. 1, 2

[46] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *ICCV*, 2023. 1, 2

[47] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 2, 3

[48] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *PAMI*, 2022. 3

[49] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *IJCV*, 2018. 5, 7, 8

[50] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *CVPR*, 2021. 5, 7

[51] Vani Suthamathi Saravanarajan, Rung-Ching Chen, Cheng-Hsiung Hsieh, and Long-Sheng Chen. Improving semantic segmentation under hazy weather for autonomous vehicles using explainable artificial intelligence and adaptive dehazing approach. *IEEE Access*, 2023. 1, 2

[52] Ruizhi Shao, Zerong Zheng, Hongwen Zhang, Jingxiang Sun, and Yebin Liu. Diffustereo: High quality human reconstruction via diffusion-based stereo using sparse cameras. In *ECCV*, 2022. 3

[53] Vivek Sharma, Ali Diba, Davy Neven, Michael S Brown, Luc Van Gool, and Rainer Stiefelhagen. Classification-driven dynamic image enhancement. In *CVPR*, 2018. 2

[54] Taeyoung Son, Juwon Kang, Namyup Kim, Sunghyun Cho, and Suha Kwak. Urie: Universal image enhancement for visual recognition in the wild. In *ECCV*, 2020. 1, 2, 5, 6, 7

[55] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2020. 3

[56] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2020. 3

[57] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, 2022. 1, 2

[58] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 5, 7

[59] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *CVPR*, 2020. 1, 2

[60] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *IJCV*, 2024. 3

[61] 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. *arXiv preprint arXiv:2406.13607*, 2024. 1, 5, 6

[62] Pei Wang, Hongzhan Huang, Xiaotong Luo, and Yanyun Qu. Data-free learning for lightweight multi-weather image restoration. In *ISCAS*, 2024. 1

[63] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, 2018. 5

[64] Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, and Zhiwei Xiong. Deep degradation prior for low-quality image classification. In *CVPR*, 2020. 2

[65] Yuxin Wu and Kaiming He. Group normalization. In *ECCV*, 2018. 3

[66] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV*, 2023. 3, 5

[67] Jiarui Xu, Sifei Liu, Arash Vahdat, Wonmin Byeon, Xiaolong Wang, and Shalini De Mello. Open-vocabulary panoptic segmentation with text-to-image diffusion models. In *CVPR*, 2023. 3

[68] Xin Xu, Shiqin Wang, Zheng Wang, Xiaolong Zhang, and Ruimin Hu. Exploring image enhancement for salient object detection in low light images. *TMM*, 2021. 1, 2

[69] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *CVPR*, 2017. 1, 2, 5, 6

[70] Zhou Yang, Weisheng Dong, Xin Li, Jinjian Wu, Leida Li, and Guangming Shi. Self-feature distillation with uncertainty modeling for degraded image recognition. In *ECCV*, 2022. 2

[71] Zizheng Yang, Jie Huang, Jiahao Chang, Man Zhou, Hu Yu, Jinghao Zhang, and Feng Zhao. Visual recognition-driven

image restoration for multiple degradation with intrinsic semantics recovery. In *CVPR*, 2023. 1, 2

[72] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *CVPR*, 2023. 1, 2

[73] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 2

[74] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 1, 2

[75] Juntao Zhang, Yuehuai Liu, Yu-Wing Tai, and Chi-Keung Tang. C3net: Compound conditioned controlnet for multimodal content generation. In *CVPR*, 2024. 3

[76] Kaiwen Zhang, Xuefeng Yan, Yongzhen Wang, and Junchen Qi. Adaptive dehazing yolo for object detection. In *ICANN*, 2023. 1, 2

[77] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023. 3, 4

[78] Yi Zhang, Xiaoyu Shi, Dasong Li, Xiaogang Wang, Jian Wang, and Hongsheng Li. A unified conditional framework for diffusion-based image restoration. In *NIPS*, 2023. 1, 2

[79] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, 2018. 2

[80] Shihao Zhao, Dongdong Chen, Yen-Chun Chen, Jianmin Bao, Shaozhe Hao, Lu Yuan, and Kwan-Yee K Wong. Uni-controlnet: All-in-one control to text-to-image diffusion models. *NIPS*, 36, 2024. 3

[81] Dian Zheng, Xiao-Ming Wu, Zuhao Liu, Jingke Meng, and Wei-shi Zheng. Diffuvolume: Diffusion model for volume based stereo matching. *arXiv preprint arXiv:2308.15989*, 2023. 3

[82] 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 *CVPR*, 2024. 3, 5, 6, 7

[83] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *CVPR*, 2023. 3