TASKIR: TASK-DRIVEN IMAGE RESTORATION VIA DEGRADATION ADAPTATION AND TASK FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Task-driven image restoration aims to improve both image quality and downstream task performance. However, existing methods predominantly focus on single degradation type and struggle to handle the diverse degradations encountered in real-world scenarios. Different degradations impose distinct restoration demands, and insufficient restoration may leave residual degradations and artifacts that impair object boundaries and semantic cues, thereby compromising downstream task performance. To address these challenges, we propose TaskIR, a two-stage task-driven unified image restoration framework that integrates degradation-adaptive restoration with task feedback refinement. In Stage I, a Degradation Representation Module (DRM) extracts degradation representations, enabling a Degradation-Guided Transformer Block (DGTB) to dynamically modulate feature transformations for adaptive restoration. In Stage II, a Task-to-Restoration Feedback Generation module (TRFG) transforms heterogeneous task features into restoration feedback by modeling task-representation discrepancies associated with the current restoration. Subsequently, a Selective Task Feedback Refinement module (STFR) assesses feedback relevance and selectively refines intermediate restoration features to mitigate interference with well-restored content. Extensive experiments demonstrate that TaskIR achieves competitive restoration quality and downstream task performance across diverse degradations and tasks.

1 INTRODUCTION

Image restoration aims to reconstruct high-quality images from degraded observations and provide reliable inputs for subsequent visual analysis. Images captured in real-world scenarios are often affected by various degradations, such as noise (Abdelhamed et al., 2018), blur (Ren et al., 2023), haze (Chen et al., 2019), and low illumination (Wei et al., 2018). These degradations not only deteriorate visual quality but may also corrupt task-relevant information, such as object textures, structures, and boundaries, thereby degrading the performance of downstream vision tasks, including image classification (Huang et al., 2017), semantic segmentation (Zhang et al., 2022), and object detection (Liu et al., 2016). Consequently, task-driven image restoration (Kim et al., 2025; Lee et al., 2026; Zhang et al., 2026) has attracted increasing attention, aiming to improve restoration quality while preserving or enhancing information relevant to downstream tasks. Existing task-driven image restoration methods typically guide the restoration process using task losses, task features, or prediction results and have achieved promising performance in specific scenarios such as super-resolution (Haris et al., 2021; Kim et al., 2024) and dehazing (Wan et al., 2025; Zhang et al., 2026). However, most of these methods are designed for a single degradation, with their network architectures and task-guidance mechanisms often tailored to specific degradation conditions, making them difficult to directly adapt to the diverse and complex degradations encountered in real-world scenarios. In fact, different degradations exhibit distinct corruption patterns and restoration requirements, and their effects on downstream task representations and prediction performance also vary. Therefore, extending task-driven image restoration from single-degradation settings to a unified multi-degradation scenario represents an important direction toward practical deployment in complex real-world environments.

In recent years, all-in-one image restoration (Potlapalli et al., 2023; Jiang et al., 2025; Zamfir et al., 2025; Tang et al., 2026) has attracted increasing attention by handling multiple degradations within a single model and enhancing multi-degradation restoration through mechanisms such as degradation

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

representation, prompt learning, and dynamic experts. However, different degradations have distinct restoration requirements, making it difficult for shared feature transformations to achieve sufficient degradation adaptation, while improved restoration quality does not necessarily lead to better downstream task performance (Chen et al., 2025). On the other hand, downstream task features emphasize high-level semantic and discriminative information, which differs from the texture and structural information required for image restoration, and directly using such features for restoration may introduce interference. Moreover, different task feedback signals contribute differently to restoration, and indiscriminately using them to refine restoration features may disrupt already restored content (Lee et al., 2026). Therefore, adapting to the restoration requirements of diverse degradations while effectively utilizing heterogeneous task feedback to improve downstream task performance without compromising restoration quality remains a core challenge for task-driven image restoration.

To address the above challenges, we propose TaskIR, a two-stage task-driven all-in-one image restoration framework. In Stage I, we design a Degradation Representation Module (DRM) to jointly model global degradation characteristics and spatial degradation responses and employ a Degradation-Guided Transformer Block (DGTB) to dynamically modulate feature transformations for degradation-adaptive restoration. In Stage II, we further incorporate downstream task feedback to improve the suitability of restoration results for downstream tasks. Specifically, Task-to-Restoration Feedback Generation (TRFG) adaptively transforms heterogeneous task features and generates targeted task feedback based on discrepancies in the task representations of the current restoration results. Subsequently, Selective Task Feedback Refinement (STFR) assesses the contributions of different feedback signals to the current restoration state and selectively refines restoration features to minimize interference with already restored content. By combining degradation-adaptive restoration with task feedback refinement, TaskIR balances restoration quality and downstream task performance under diverse degradation conditions.

The main contributions of this work are summarized as follows:

- We propose TaskIR, a task-driven unified image restoration framework that combines degradation-adaptive restoration and task feedback refinement to balance restoration quality and downstream task performance under various degradation conditions.
- We develop DGTB that dynamically modulates feature transformations in a shared restoration network using degradation representations, enabling adaptive restoration across diverse degradations.
- We introduce TRFG and STFR to transform heterogeneous task features into targeted feedback and selectively refine restoration features, improving downstream task performance while minimizing interference with previously restored content.
- Extensive experiments across diverse image degradations and downstream vision tasks demonstrate that TaskIR effectively improves downstream task performance while maintaining high restoration quality.

2 Related Work

All-in-One Image Restoration. All-in-One Image Restoration (AiOIR) (Jiang et al., 2025; Tu et al., 2026; Cui et al., 2026; Lin et al., 2026b) aims to handle multiple degradations, such as noise, blur, haze, and compression artifacts, within a single model, providing more flexible multi-degradation restoration than task-specific methods (Cai et al., 2023; He et al., 2026). Existing studies mainly improve the adaptability of unified models through degradation representation and dynamic restoration mechanisms. For example, AirNet (Li et al., 2022) learns degradation representations via contrastive learning to guide restoration; PromptIR (Potlapalli et al., 2023) employs learnable prompts to modulate the restoration process under different degradation conditions; and MoCE-IR (Zamfir et al., 2025) introduces a complexity-adaptive mixture-of-experts mechanism for dynamic assignment across degradation tasks. More recently, MIRAGE (Ren et al., 2026), BioIR (Cui et al., 2026), and BaryIR (Tang et al., 2026) further improve multi-degradation restoration performance and generalization from the perspectives of feature decomposition, bio-inspired visual modeling, and degradation-content decoupling, respectively. Despite these advances, existing AiOIR methods still mainly optimize pixel fidelity and visual quality, with limited attention to the suitability of restored images for downstream tasks.

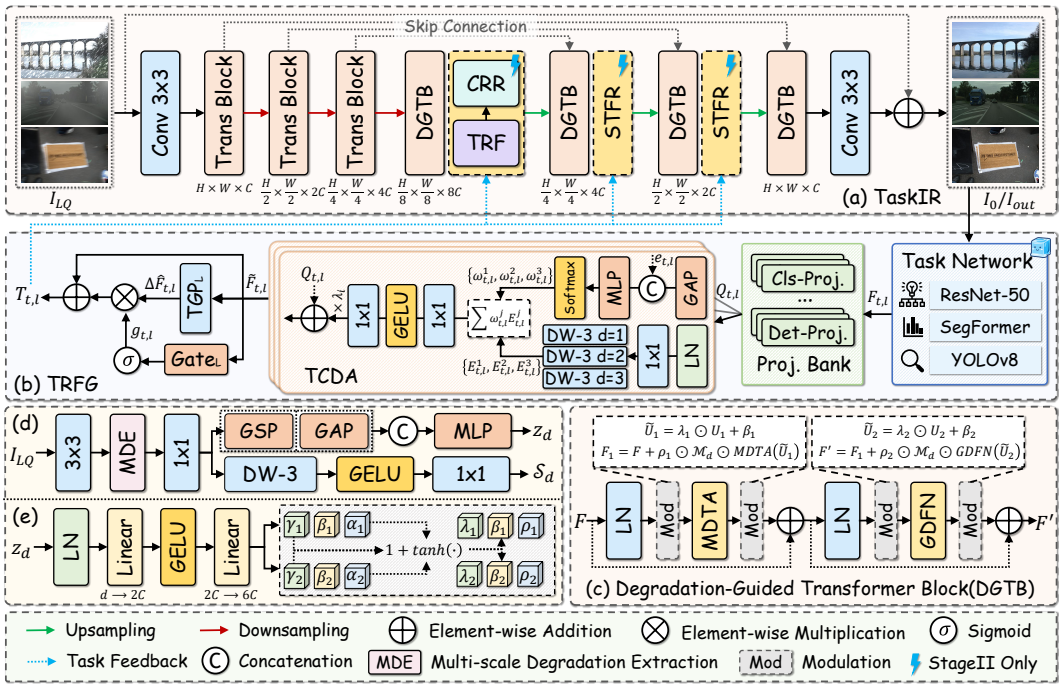

Figure 1: Overview of the proposed TaskIR framework and its key components: (a) Overall architecture of TaskIR; (b) Task-to-restoration feedback generation (TRFG); (c) Degradation-guided transformer block (DGTB); (d) Degradation representation module (DRM); and (e) Degradation-conditioned parameter generator (DPG).

Task-Driven Image Restoration. Task-Driven Image Restoration (TDIR) (Yang et al., 2023; Kim et al., 2024; Lin et al., 2026a; Zhang et al., 2026) introduces downstream visual task constraints to make restored images more suitable for machine vision analysis. Existing studies (Chen et al., 2025) have shown that image degradations can substantially affect the performance of tasks such as classification, semantic segmentation, and object detection, while optimizing only pixel-level or perceptual quality is insufficient for practical visual applications. To address this issue, SR4IR (Kim et al., 2024) and TDSR (Haris et al., 2021) employ task-related losses to improve the recognition and detection performance of super-resolved images, respectively; MS-FODN (Wan et al., 2025) jointly optimizes object detection and dehazing to improve detection under hazy conditions; ADeT-Net (Zhang et al., 2026) combines task feedback with textual instructions to perform dynamic dehazing for different downstream requirements; and DTIUIE (Lin et al., 2026a) introduces task-driven constraints to improve the suitability of underwater image enhancement for recognition tasks. However, most existing methods are designed for a single degradation and are therefore difficult to balance restoration requirements and task performance under multiple degradations. Moreover, downstream task representations do not fully align with the requirements of image restoration, and indiscriminate refinement using task information may introduce irrelevant information and interfere with already restored content.

3 METHOD

As shown in Fig. 1, given a degraded image I_{LQ}, TaskIR adopts a two-stage task-driven all-in-one image restoration framework to jointly improve restoration quality and downstream task performance under diverse degradations. In Stage I, the DRM extracts a global degradation representation z_d and a spatial degradation response S_d. The DGTB then dynamically modulates the feature transformations of the shared restoration network to produce an initial restored image I_0. In Stage II, TRFG transforms heterogeneous downstream task features into task feedback and predicts the task-representation discrepancy of the current restored image. Based on this feedback, STFR dynamically integrates multiple task feedback signals according to their relevance to the current restoration state and selectively refines intermediate restoration features, yielding the final restored image I_{out}.

3.1 STAGE I: DEGRADATION-GUIDED RESTORATION

Degradation Representation Module. Different degradations exhibit distinct local spatial patterns and also induce changes in feature statistics. To characterize both types of information, we design the DRM, which generates a global degradation representation z_d and a spatial degradation response S_d. Specifically, given a degraded image I_{LQ}, a 3×3 convolution is first applied to obtain shallow features $X \in \mathbb{R}^{B\times C\times H\times W}$. Subsequently, Multi-scale Degradation Extraction (MDE) employs parallel depth-wise convolution branches with different receptive fields to extract multi-scale degradation patterns, followed by feature fusion and a residual connection to obtain the degradation feature X_d, as illustrated in Fig. 1(d). To characterize degradation-induced changes in global statistics, we apply GAP and Global Standard Deviation Pooling (GSP) to X_d to capture its mean and dispersion, respectively. The resulting statistics are concatenated and mapped into the global degradation representation $z_d \in \mathbb{R}^{B\times D}$ through an MLP.

In addition to image-level degradation statistics, the degradation distribution may also vary across spatial locations. We further employ a spatial mapper $P_s(\cdot)$ to predict the spatial degradation response from X_d. For the k-th restoration stage, S_d is resized according to the resolution of the current features and then mapped through a Sigmoid function to obtain the corresponding spatial degradation map:

$$S_d = P_s(X_d), \qquad M_d^{(k)} = 2\sigma\left(\mathrm{Resize}_k(S_d)\right). \tag{1}$$

where $\sigma(\cdot)$ denotes the Sigmoid function, and $M_d^{(k)}$ denotes the spatial degradation map whose resolution matches the features at the k-th restoration stage.

Degradation-Guided Transformer Block. After obtaining the global degradation representation z_d and the spatial degradation map $M_d^{(k)}$, we design the DGTB, as illustrated in Fig. 1(c). DGTB explicitly incorporates degradation information into the MDTA and GDFN sublayers of the Transformer, enabling the shared restoration network to adjust channel responses, residual update strengths, and feature update magnitudes at different spatial locations according to the current degradation state. For the k-th restoration stage, a Degradation-conditioned Parameter Generator (DPG) first generates two sets of degradation modulation parameters from z_d. As shown in Fig. 1(e), DPG employs normalization and two mapping layers to generate channel scales $\lambda_i^{(k)}$, channel biases $\beta_i^{(k)}$, and residual scales $\rho_i^{(k)}$ for the MDTA and GDFN branches, respectively, where $i \in \{1, 2\}$.

Given the input feature F of DGTB, the first set of degradation parameters modulates the normalized features of the MDTA branch, followed by feature updates using the residual scale and the spatial degradation map of the current stage:

$$F_1 = F + \rho_1^{(k)} \odot M_d^{(k)} \odot \mathrm{MDTA}\left(\lambda_1^{(k)} \odot \mathrm{LN}(F) + \beta_1^{(k)}\right). \tag{2}$$

The second set of degradation parameters takes the updated feature F_1 as input to modulate the GDFN branch, yielding the DGTB output:

$$F_2 = F_1 + \rho_2^{(k)} \odot M_d^{(k)} \odot \mathrm{GDFN}\left(\lambda_2^{(k)} \odot \mathrm{LN}(F_1) + \beta_2^{(k)}\right). \tag{3}$$

Here, $\lambda_i^{(k)}$ and $\beta_i^{(k)}$ adjust the channel scales and biases of the input features to MDTA or GDFN, while $\rho_i^{(k)}$ controls the update strength of the corresponding residual branch. The spatial degradation map $M_d^{(k)}$ further modulates the update magnitudes at different spatial locations according to the spatial distribution of degradations. We employ DGTBs in the latent and decoder stages, allowing degradation information to continuously guide feature restoration and ultimately producing the Stage I restoration result I_0.

3.2 STAGE II: TASK-DRIVEN FEEDBACK MODELING

Building upon the initial restoration results generated in Stage I, Stage II introduces downstream task feedback to further improve task performance. Specifically, we design Task-to-Restoration Feedback Generation (TRFG) and Selective Task Feedback Refinement (STFR). TRFG transforms features from different downstream tasks into task feedback, while STFR reduces interference with already restored content through feedback fusion and selective feature refinement.

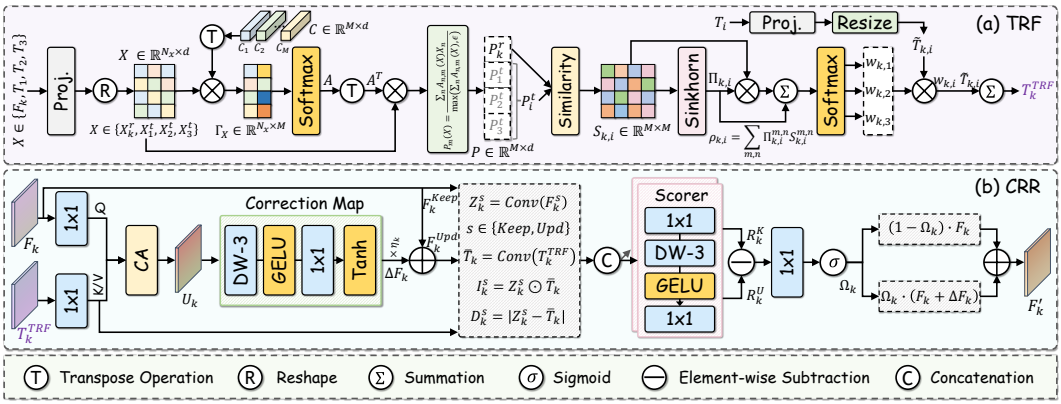

Figure 2: Architecture of the proposed selective task feedback refinement (STFR) module. (a) Task-related feedback (TRF); (b) Conditional restoration refinement (CRR).

3.2.1 Task-to-Restoration Feedback Generation

Features from different downstream tasks vary in channel dimensions, spatial scales, and representation forms. Moreover, these features are primarily designed for task prediction, making them difficult to directly apply to image restoration. To address this issue, we design TRFG to generate task feedback through adaptive feature transformation and task representation discrepancy prediction.

Given the feature $F_{t,l}$ extracted at position l for task t, a task- and position-specific Projection Bank first performs channel mapping and spatial alignment to obtain $Q_{t,l} = P_{t,l}(F_{t,l})$, where $P_{t,l}(\cdot)$ denotes the corresponding independent projection mapping. Subsequently, the Task-Conditioned Dynamic Adapter (TCDA) employs three depth-wise convolution branches with dilation rates $r_j \in \{1, 2, 3\}$ to extract multi-scale information (Yu & Koltun, 2015), as illustrated in Fig. 1(b). TCDA predicts channel-wise weights for each branch based on the global statistics of $Q_{t,l}$ and the learnable condition embedding $e_{t,l}$ associated with task t and position l:

$$\omega_{t,l}^{j} = \mathrm{Softmax}_j\left(\mathrm{MLP}\left([\mathrm{GAP}(Q_{t,l}); e_{t,l}]\right)\right). \tag{4}$$

Denoting the output of the j-th branch as $E_{t,l}^{j}$, the adapted task features are obtained through dynamic weighted aggregation, lightweight feature mapping, and a residual connection:

$$\hat{F}_{t,l} = Q_{t,l} + \xi_l \mathcal{A}\left(\sum_{j=1}^{3} \omega_{t,l}^{j} \odot E_{t,l}^{j}\right), \tag{5}$$

where $\mathcal{A}(\cdot)$ denotes the lightweight feature mapping, and ξ_l is a learnable residual scale.

On this basis, the Task Gap Predictor (TGP) predicts the task representation discrepancy of the current restoration result through $\Delta\hat{F}_{t,l} = \tanh(\Phi_l(\hat{F}_{t,l}))$, where $\Phi_l(\cdot)$ denotes the discrepancy predictor corresponding to position l. During training, an exponential moving average (EMA) teacher (Tarvainen & Valpola, 2017) is employed to construct supervision targets based on the task representation discrepancy between the initial restored image I_0 and the reference image I_{GT}. During inference, the reference image is not required. Subsequently, a lightweight gate modulates the contribution of the predicted discrepancy to generate the final task feedback:

$$T_{t,l} = \hat{F}_{t,l} + \sigma\left(\Psi_l(\hat{F}_{t,l})\right) \odot \Delta\hat{F}_{t,l}, \tag{6}$$

where $\Psi_l(\cdot)$ denotes the gating mapping.

3.2.2 Selective Task Feedback Refinement

The multiple task feedback signals generated by TRFG differ in their relevance to the current restoration state, and directly applying them to restoration features may interfere with already restored content. We therefore design STFR, as shown in Fig. 2, in which Task-Related Feedback (TRF)

estimates the relative contributions of different feedback signals, while Conditional Restoration Refinement (CRR) further controls their actual updates to the restoration features.

Task-Related Feedback. Given the restoration feature F_k at the k-th restoration position and three task feedback signals $\{T_i\}_{i=1}^3$, TRF first projects them into a unified d-dimensional space and flattens them to obtain X_k^r and $\{X_i^t\}_{i=1}^3$. To model the overall relationship between restoration features and task feedback, we introduce a shared prototype bank $C = \{C_m\}_{m=1}^M \in \mathbb{R}^{M\times d}$ containing M learnable prototypes, which is shared across all restoration positions and feedback paths. For any projected feature X, soft assignments are computed based on its cosine similarity to the shared prototypes:

$$A_{n,m}(X) = \mathrm{Softmax}_m\left(\frac{\langle \bar{X}_n, \bar{C}_m\rangle}{\tau_p}\right), \tag{7}$$

where $\bar{X}_n$ and $\bar{C}_m$ denote the L2-normalized feature and prototype, respectively. $\mathrm{Softmax}_m$ denotes normalization along the prototype dimension, and τ_p denotes the prototype assignment temperature. Subsequently, the m-th prototype representation is obtained through weighted aggregation:

$$P_m(X) = \frac{\sum_n A_{n,m}(X)X_n}{\max\left(\sum_n A_{n,m}(X), \epsilon\right)}. \tag{8}$$

This yields the restoration feature prototypes P_k^r and the prototypes P_i^t for the i-th task feedback signal. We then compute the cosine similarity matrix $S_{k,i} \in \mathbb{R}^{M\times M}$ between the two sets of prototypes and approximate the entropy-regularized transport plan $\Pi_{k,i}$ through log-domain Sinkhorn (Cuturi, 2013) iterations. The relevance score between the i-th task feedback signal and the current restoration state, along with its contribution weight, is defined as:

$$\rho_{k,i} = \sum_{m=1}^{M}\sum_{n=1}^{M} \Pi_{k,i}^{m,n} S_{k,i}^{m,n}, \qquad w_{k,i} = \mathrm{Softmax}_i\left(\frac{\rho_{k,i}}{\tau_r}\right). \tag{9}$$

where $\mathrm{Softmax}_i$ denotes normalization across the three task feedback signals, and τ_r is the path weighting temperature. Each task feedback signal is then projected and resized to match the spatial resolution of the current restoration position, yielding $\widetilde{T}_{k,i}$. The aligned feedback signals are fused according to their corresponding contribution weights:

$$T_k^{\mathrm{TRF}} = \sum_{i=1}^{3} w_{k,i}\widetilde{T}_{k,i}. \tag{10}$$

Thus, TRF dynamically determines the relative contributions of different task feedback signals based on the prototype matching relationships between the current restoration state and each feedback signal, providing fused feedback for subsequent feature refinement.

Conditional Restoration Refinement. Given T_k^{TRF}, CRR further controls the extent to which task feedback modifies the current restoration features. First, channel-wise cross-attention is employed to establish interactions between restoration features and task feedback, where the query is projected from F_k, while the key and value are projected from T_k^{TRF}, yielding a candidate refinement:

$$\Delta F_k = \eta_k \tanh\left(\mathcal{H}_k\left(\mathrm{CA}_k(F_k, T_k^{\mathrm{TRF}})\right)\right), \tag{11}$$

where $\mathcal{H}_k(\cdot)$ consists of a 3×3 depth-wise convolution, GELU, and a 1×1 convolution, and η_k is a learnable refinement scale. Based on ΔF_k, we construct a keep state $F_k^K = F_k$ and an update state $F_k^U = F_k + \Delta F_k$. The two candidate states and the fused feedback are mapped into a unified feature space, yielding Z_k^s and $\bar{T}_k$, respectively. A shared Scorer then computes responses based on the conditional relationship between each candidate state and the fused feedback:

$$R_k^s = \mathcal{S}_k\left([Z_k^s; \bar{T}_k; Z_k^s \odot \bar{T}_k; |Z_k^s - \bar{T}_k|]\right), \quad s \in \{K, U\}. \tag{12}$$

An update gate is then generated based on the response difference between the update and keep states:

$$\Omega_k = \sigma\left(\mathcal{G}_k(R_k^U - R_k^K)\right), \tag{13}$$

where $\mathcal{G}_k(\cdot)$ denotes a learnable gating mapping at the k-th restoration position, implemented using a 1×1 convolution. Finally, the refined restoration features are obtained by selectively fusing the keep and update states:

$$F_k' = (1-\Omega_k)\odot F_k^K + \Omega_k \odot F_k^U = F_k + \Omega_k \odot \Delta F_k. \tag{14}$$

Table 1: Overall restoration performance across eight degradation types. The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	ImageNet-1K		Cityscapes		VOC2012	
		PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑
AirNet (Li et al., 2022)	CVPR'22	25.20	0.7785	30.81	0.9005	25.30	0.7792
PromptIR (Potlapalli et al., 2023)	NeurIPS'23	27.37	0.8280	33.43	0.9290	27.32	0.8268
NDR-Restore (Yao et al., 2024)	TIP'24	27.26	0.8261	33.67	0.9318	27.19	0.8243
AdaIR (Cui et al., 2025)	ICLR'25	27.70	0.8330	33.70	0.9321	27.63	0.8314
DFPIR (Tian et al., 2025)	CVPR'25	28.23	0.8420	<u>34.81</u>	0.9334	<u>28.17</u>	0.8408
MoCE-IR (Zamfir et al., 2025)	CVPR'25	24.23	0.7070	28.97	0.7977	24.20	0.7071
VLU-Net (Zeng et al., 2025)	CVPR'25	27.75	0.8359	34.01	0.9341	27.67	0.8342
BaryIR (Tang et al., 2026)	TPAMI'26	<u>28.26</u>	<u>0.8451</u>	34.76	0.9384	28.17	<u>0.8434</u>
C2SSM (Wu et al., 2026)	CVPR'26	27.32	0.8295	33.60	0.9345	27.29	0.8286
MIRAGE (Ren et al., 2026)	ICLR'26	27.83	0.8407	34.24	<u>0.9397</u>	27.80	0.8398
TaskIR (Ours)	–	**28.31**	**0.8491**	**35.31**	**0.9481**	**28.28**	**0.8484**

Table 2: Overall downstream task performance (%). The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	ImageNet-1K		Cityscapes		VOC2012	
		Top-1 ↑	Top-5 ↑	mIoU ↑	Dice ↑	mAP_{50} ↑	$mAP_{50:95}$ ↑
AirNet (Li et al., 2022)	CVPR'22	68.31	87.30	70.95	82.08	76.90	58.35
PromptIR (Potlapalli et al., 2023)	NeurIPS'23	72.26	89.78	74.12	84.36	81.92	63.32
NDR-Restore (Yao et al., 2024)	TIP'24	72.15	90.06	74.08	84.33	81.65	63.13
AdaIR (Cui et al., 2025)	ICLR'25	72.10	90.36	74.12	84.35	82.05	63.44
DFPIR (Tian et al., 2025)	CVPR'25	72.13	90.41	74.05	84.30	<u>82.44</u>	63.14
MoCE-IR (Zamfir et al., 2025)	CVPR'25	65.95	86.50	66.39	78.39	72.11	54.01
VLU-Net (Zeng et al., 2025)	CVPR'25	72.63	<u>90.56</u>	74.19	84.40	82.06	63.65
BaryIR (Tang et al., 2026)	TPAMI'26	<u>72.95</u>	90.30	74.17	<u>84.40</u>	82.40	**64.15**
C2SSM (Wu et al., 2026)	CVPR'26	71.49	89.86	73.89	84.19	80.98	62.46
MIRAGE (Ren et al., 2026)	ICLR'26	72.25	90.43	<u>74.63</u>	84.71	81.95	63.45
TaskIR (Ours)	–	**73.00**	**90.65**	**74.69**	**84.77**	**82.46**	<u>63.98</u>

By comparing the responses of the two candidate states under the same task feedback, CRR adaptively controls the extent of feature refinement, reducing interference with already restored content.

4 EXPERIMENTS

We evaluate TaskIR on ImageNet-1K (Deng et al., 2009), Cityscapes (Cordts et al., 2016), and PASCAL VOC2012 (Everingham et al., 2010) across eight degradation types in terms of both image restoration quality and downstream task performance. We further conduct comprehensive comparisons and ablation studies to validate the effectiveness of the proposed framework.

4.1 IMPLEMENTATION DETAILS

TaskIR adopts a two-stage training strategy. In Stage I, the degradation-guided restoration network is trained using a pixel-wise L_1 reconstruction loss and a degradation classification loss. In Stage II, the pretrained Stage-I weights are loaded, while the restoration backbone and downstream task networks are frozen, and only TRFG and STFR are optimized. All experiments are conducted on four NVIDIA A100 GPUs with a batch size of 8. We use the AdamW (Loshchilov & Hutter, 2017) optimizer with an initial learning rate of 1.0×10^{-4}. Stage I and Stage II are trained for 200K and 100K iterations, respectively.

4.2 COMPARISONS WITH THE STATE-OF-THE-ART METHODS

Image Restoration Performance. Tab. 1 reports the overall image restoration performance of different methods after two-stage training. TaskIR achieves competitive PSNR and SSIM in the overall evaluation across eight degradation types, demonstrating that it maintains high restoration quality even after incorporating downstream task feedback. Fig. 3 further presents a qualitative comparison of image restoration results from different methods. Compared with other methods, TaskIR more

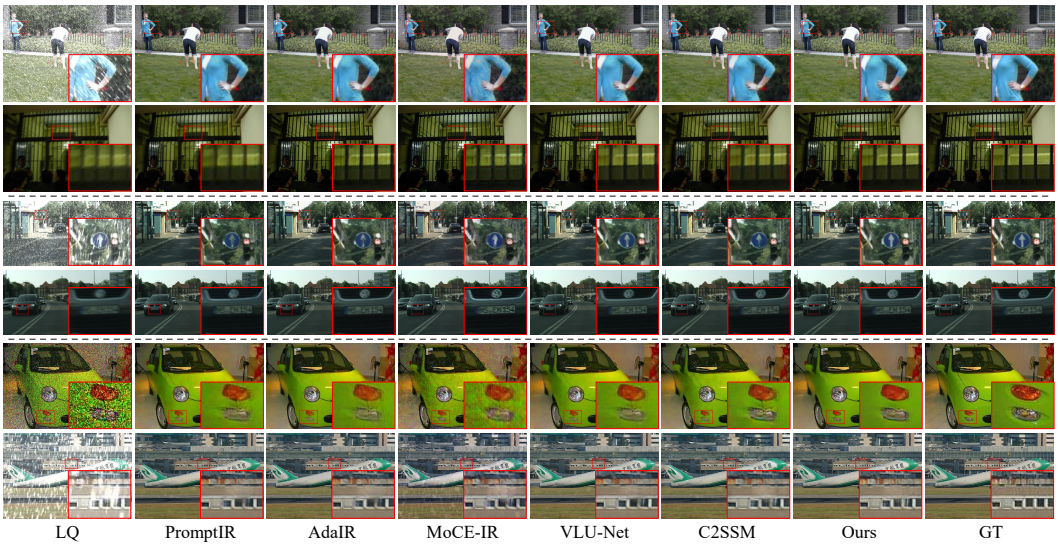

Figure 3: Qualitative comparison of image restoration results under diverse degradations on ImageNet-1K (Deng et al., 2009), Cityscapes (Cordts et al., 2016), and PASCAL VOC2012 (Everingham et al., 2010). Zoom in for a better view.

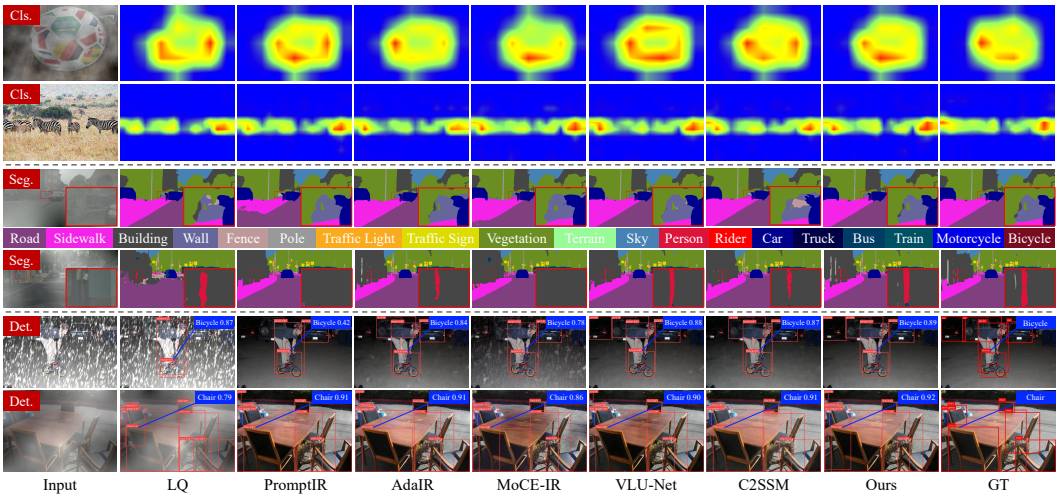

Figure 4: Qualitative comparison of downstream task results under diverse degradations, including classification (Cls.), segmentation (Seg.), and detection (Det.). Zoom in for a better view.

effectively removes degradations and restores texture and structural details. For example, in the car headlight region, some competing methods produce blurred edges and distorted contours, whereas TaskIR better recovers the sharp boundaries and original shape of the headlight, yielding a visual result closer to the ground truth.

Downstream Task Performance. We further evaluate the performance of the restored images on image classification, semantic segmentation, and object detection. As shown in Tab. 2, compared with other restoration methods trained with downstream task supervision, TaskIR achieves competitive performance in terms of Top-1/Top-5 accuracy (Russakovsky et al., 2015), mIoU/Dice (Everingham et al., 2010; Dice, 1945), and mAP (Lin et al., 2014), demonstrating that its restoration results can better support diverse downstream vision tasks. Fig. 4 presents qualitative comparisons of downstream task results obtained with different methods. Compared with other methods, TaskIR produces classification predictions that are more consistent with the ground truth, segmentation results that better preserve object boundaries and local structures, and detection boxes that more closely align with target locations. For example, in building segmentation, some competing methods exhibit local

Table 3: Ablation study on the feedback modules. We report both restoration quality and downstream task performance across three datasets.

TRFG	STFR		ImageNet-1K		Cityscapes		VOC2012	
	TRF	CRR	Restoration	Classification	Restoration	Segmentation	Restoration	Detection
✓	✗	✗	28.29/0.8489	72.95/90.59	35.28/0.9479	74.52/84.64	28.26/0.8481	82.24/63.83
✓	✓	✗	28.28/0.8489	72.98/90.58	35.28/0.9480	74.65/84.74	28.25/0.8482	82.29/63.87
✓	✗	✓	28.30/0.8487	72.98/90.64	35.30/0.9480	74.67/84.75	28.27/0.8484	82.37/63.92
✗	✓	✓	28.31/0.8490	72.99/90.62	35.29/0.9481	74.64/84.73	28.28/0.8483	82.44/63.91
✓	✓	✓	**28.31/0.8491**	**73.00/90.65**	**35.31/0.9481**	**74.69/84.77**	**28.28/0.8484**	**82.46/63.98**

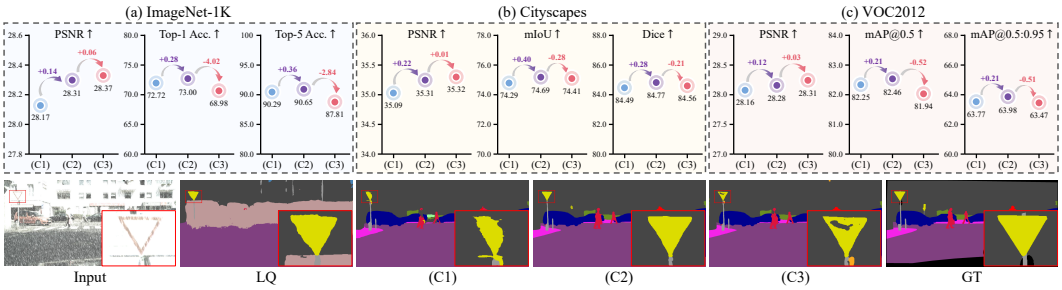

Figure 5: Ablation study on the two-stage design. Qualitative segmentation results on Cityscapes are shown for further comparison.

misclassifications and discontinuous semantic regions, whereas TaskIR more completely preserves building regions and their boundaries, yielding better segmentation results.

4.3 Ablation Study

To evaluate the effectiveness of each component in TaskIR, we conduct ablation studies under the same training settings and report both image restoration quality and downstream task performance.

Effectiveness of the Two-Stage Design. We first conduct ablation studies on Stage I and Stage II to evaluate the effects of the spatial degradation map M_d and task feedback, respectively. As shown in Fig. 5, C1 uses only the global degradation representation z_d while retaining task feedback, C2 represents the full TaskIR model, and C3 removes task feedback and directly uses the initial restored images for task prediction. Compared with C1, C2 improves restoration quality by incorporating M_d. Compared with C3, C2 improves downstream task performance while maintaining comparable restoration quality.

Effectiveness of Feedback Modules. We further conduct ablation studies on the feedback modules in Stage II. As shown in Tab. 3, retaining only TRFG results in relatively low downstream task performance. Incorporating either TRF or CRR improves overall task performance, indicating that feedback fusion and selective feature refinement both contribute to more effective utilization of task feedback. Compared with configurations in which any individual module is removed, the full TaskIR model achieves better results across all downstream task metrics while maintaining comparable restoration quality.

5 Conclusion

In this paper, we propose TaskIR, a task-driven all-in-one image restoration framework that jointly models degradation adaptation and task feedback through a two-stage design. In Stage I, degradation representations are used to dynamically modulate Transformer feature transformations, enabling adaptive restoration across diverse degradations. In Stage II, downstream task feedback is introduced to selectively refine restoration features, enhancing their suitability for downstream tasks while reducing interference with already restored content. Extensive experiments demonstrate that TaskIR achieves competitive performance across diverse degradations and downstream vision tasks, validating the effectiveness of the proposed framework.

REFERENCES

Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1692–1700. IEEE, 2018.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12504–12513, 2023.

I Chen, Wei-Ting Chen, Yu-Wei Liu, Yuan-Chun Chiang, Sy-Yen Kuo, Ming-Hsuan Yang, et al. Unirestore: Unified perceptual and task-oriented image restoration model using diffusion prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17969–17979, 2025.

Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen Kuo. Pms-net: Robust haze removal based on patch map for single images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11673–11681. IEEE, 2019.

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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In *International Conference on Learning Representations*, volume 2025, pp. 101306–101327, 2025.

Yuning Cui, Wenqi Ren, and Alois Knoll. Bio-inspired image restoration. *Advances in Neural Information Processing Systems*, 38:80452–80481, 2026.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, 1945.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2): 303–338, 2010.

Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Task-driven super resolution: Object detection in low-resolution images. In *International conference on neural information processing*, pp. 387–395. Springer, 2021.

Jin-Ting He, Fu-Jen Tsai, Yan-Tsung Peng, Min-Hung Chen, Chia-Wen Lin, and Yen-Yu Lin. Hndiff: Haze-noise diffusion for image dehazing. *arXiv preprint arXiv:2608.10995*, 2026.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

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

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 2261–2269. Ieee, 2017.

Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolo. *Zenodo*, 2023.

Jaeha Kim, Junghun Oh, and Kyoung Mu Lee. Beyond image super-resolution for image recognition with task-driven perceptual loss. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2651–2661. IEEE, 2024.

Jaeha Kim, Junghun Oh, and Kyoung Mu Lee. Exploiting diffusion prior for task-driven image restoration. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10151–10161. IEEE, 2025.

Hongjae Lee, Sojung Kang, Jaeseong Yu, and Seung-Won Jung. Tasktok: Delving into task tokens for task-driven image restoration. *arXiv preprint arXiv:2606.26615*, 2026.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17431–17441. IEEE, 2022.

Bosen Lin, Feng Gao, Yanwei Yu, Junyu Dong, and Qian Du. Downstream task inspired underwater image enhancement: a perception-aware study from dataset construction to network design. *IEEE Transactions on Image Processing*, 2026a.

Jingbo Lin, Zhilu Zhang, Wenbo Li, Renjing Pei, Hang Xu, Hongzhi Zhang, and Wangmeng Zuo. Unirestorer: Universal image restoration via adaptively estimating image degradation at proper granularity. In *International Conference on Learning Representations*, volume 2026, pp. 2369–2409, 2026b.

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 *European conference on computer vision*, pp. 740–755. Springer, 2014.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer, 2016.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in neural information processing systems*, 36:71275–71293, 2023.

Bin Ren, Yawei Li, Xu Zheng, Yuqian Fu, Danda Pani Paudel, Hong Liu, Ming-Hsuan Yang, Luc Van Gool, and Nicu Sebe. Efficient degradation-agnostic image restoration via channel-wise functional decomposition and manifold regularization. In *The Fourteenth International Conference on Learning Representations(ICLR)*, 2026.

Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10687–10699. IEEE, 2023.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

Xiaole Tang, Xiaoyi He, Jiayi Xu, Xiang Gu, and Jian Sun. Learning continuous wasserstein barycenter space for generalized all-in-one image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.

Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, and Chao Ren. Degradation-aware feature perturbation for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28165–28175, 2025.

Yanjie Tu, Qingsen Yan, Axi Niu, and Jiacong Tang. Tpgdiff: Hierarchical triple-prior guided diffusion for image restoration. *arXiv preprint arXiv:2601.20306*, 2026.

Yu Wan, Jie Li, Liupeng Lin, Qiangqiang Yuan, and Huanfeng Shen. Collaboration of dehazing and object detection tasks: A multitask learning framework for foggy image. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

Chen Wu, Ling Wang, Zhuoran Zheng, Yuning Cui, Zhixiong Yang, Xiangyu Chen, Yue Zhang, Weidong Jiang, and Jingyuan Xia. Scan clusters, not pixels: A cluster-centric paradigm for efficient ultra-high-definition image restoration. *arXiv preprint arXiv:2602.21917*, 2026.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

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 *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14059–14070. IEEE, 2023.

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.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12753–12763, 2025.

Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. Vision-language gradient descent-driven all-in-one deep unfolding networks. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 7524–7533, 2025.

Wenqiang Zhang, Zilong Huang, Guozhong Luo, Tao Chen, Xinggang Wang, Wenyu Liu, Gang Yu, and Chunhua Shen. Topformer: Token pyramid transformer for mobile semantic segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12073–12083. IEEE, 2022.

Yafei Zhang, Shuaitian Song, Huafeng Li, Shujuan Wang, and Yu Liu. Adaptive dynamic dehazing via instruction-driven and task-feedback closed-loop optimization for diverse downstream task adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12888–12896, 2026.

A Overview of Supplementary

In this supplementary material, we provide additional details and experimental results to complement the main paper. We first describe the experimental protocols, including datasets, evaluation metrics, downstream architectures, optimization objectives, and the comparison protocol. We then present additional quantitative results on image restoration, downstream tasks, and computational complexity. Next, we provide more qualitative comparisons for both image restoration and downstream task results. Finally, we describe the use of Large Language Models (LLMs) in preparing the manuscript. The supplementary material is organized as follows:

1. Experimental Protocols (Sec. B)
   (a) Datasets (Sec. B.1)
   (b) Evaluation Metrics (Sec. B.2)
   (c) Downstream Architectures (Sec. B.3)
   (d) Optimization Objectives (Sec. B.4)
   (e) Comparison Protocol (Sec. B.5)
2. Additional Quantitative Results (Sec. C)
   (a) Image Restoration Performance (Sec. C.1)
   (b) Downstream Task Performance (Sec. C.2)
   (c) Computational Complexity (Sec. C.3)
3. Additional Visual Results (Sec. D)
   (a) Image Restoration Results (Sec. D.1)
   (b) Downstream Task Results (Sec. D.2)
4. LLM Usage (Sec. E)

B Experimental Protocols

B.1 Datasets

We evaluate TaskIR on three representative high-level vision tasks, including image classification, semantic segmentation, and object detection, using the ImageNet-1K (Deng et al., 2009), Cityscapes (Cordts et al., 2016), and PASCAL VOC2012 (Everingham et al., 2010) datasets, respectively. For image classification, we randomly select 4,000 images from ImageNet-1K for training and 1,000 images for validation. For semantic segmentation and object detection, we use Cityscapes and PASCAL VOC2012, respectively. Cityscapes contains 2,975 training images and 500 validation images, while PASCAL VOC2012 contains 5,717 training images and 5,823 validation images.

To construct a unified multi-degradation experimental setting, we follow the degradation generation protocol of (Hendrycks & Dietterich, 2019; Chen et al., 2025) and apply eight types of synthetic degradations to each clean image, including Gaussian noise, Shot noise, Defocus blur, Motion blur, Snow, Fog, Brightness, and JPEG compression. Thus, eight images with different degradation types are generated from each original image. After degradation augmentation, ImageNet-1K contains 32,000/8,000 training/validation images, Cityscapes contains 23,800/4,000 images, and PASCAL VOC2012 contains 45,736/46,584 images. All three datasets use the same degradation types and generation protocol, providing a unified multi-degradation evaluation setting across different downstream tasks and enabling a systematic evaluation of the model's restoration capability and task applicability under different degradation conditions. Visual examples of different degradation types are shown in Fig. 6.

B.2 Evaluation Metrics

We comprehensively evaluate TaskIR in terms of image restoration quality and downstream task performance. For image restoration quality, we adopt Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) (Wang et al., 2004) in the RGB space to measure pixel fidelity and structural similarity between restored images and their corresponding high-quality images. For



Figure 6: Visualization of representative degraded samples from ImageNet-1K, Cityscapes, and PASCAL VOC2012 under eight degradation types.

downstream task performance, we report Top-1 and Top-5 accuracy (Russakovsky et al., 2015) for image classification on ImageNet-1K, Dice score (Dice, 1945) and mean Intersection-over-Union (mIoU) (Everingham et al., 2010) for semantic segmentation on Cityscapes, and the mAP_{50} and $\text{mAP}_{50:95}$ scores (Lin et al., 2014) for object detection on PASCAL VOC2012. These metrics are jointly used to evaluate the model's visual restoration capability and its effectiveness in supporting high-level vision tasks under different degradation conditions.

B.3 Downstream Architectures

We adopt representative networks for three downstream vision tasks, including ResNet-50 (He et al., 2016) for image classification, SegFormer (Xie et al., 2021) for semantic segmentation, and YOLOv8 (Jocher et al., 2023) for object detection. To obtain task representations at different semantic levels and spatial scales, we extract multi-scale features from each network and feed them into TRFG to generate task feedback for image restoration. During Stage II training, all three downstream task networks remain frozen and are used only for task feature extraction and supervision to ensure the stability of feedback signals.

B.4 Optimization Objectives

We adopt a two-stage training strategy to optimize TaskIR. In Stage I, the degradation-guided restoration network is trained independently to learn unified multi-degradation restoration capabilities. In Stage II, we load the Stage-I pretrained model and freeze the Stage-I restoration backbone, the independent Stage-I reference network, and the downstream task networks. Only TRFG and STFR are optimized, allowing task feedback to further improve downstream task performance while limiting its impact on the basic restoration quality.

In Stage I, the training objective consists of an image reconstruction loss and a degradation classification loss:

$$\mathcal{L}_{\text{I}} = \beta_{\text{rec}}\mathcal{L}_{\text{rec}}^{\text{I}} + \beta_{\text{deg}}\mathcal{L}_{\text{deg}}, \tag{15}$$

where the Stage-I reconstruction loss is defined as $\mathcal{L}_{\text{rec}}^{\text{I}} = \frac{1}{N}\|I_0 - I_{\text{GT}}\|_1$. Here, I_0 denotes the Stage-I restoration result, I_{GT} denotes the corresponding high-quality reference image, and N denotes the total number of pixel and channel elements involved in the loss computation.

In Stage II, the overall optimization objective is defined as:

$$\mathcal{L}_{\mathrm{II}} = \beta_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \mathcal{L}_{\mathrm{task}} + \beta_{\mathrm{gap}}\mathcal{L}_{\mathrm{gap}} + \beta_{\mathrm{preserve}}\mathcal{L}_{\mathrm{preserve}}, \tag{16}$$

where the Stage-II reconstruction loss is $\mathcal{L}_{\mathrm{rec}} = \frac{1}{N}\|I_{\mathrm{out}} - I_{\mathrm{GT}}\|_1$, and I_{out} denotes the final restoration result after task feedback optimization. During training, Stage II adopts a task-balanced scheduling strategy. Each iteration processes only the batch corresponding to the current task t and employs the corresponding task loss:

$$\mathcal{L}_{\mathrm{task}} = \beta_t \mathcal{L}_t, \tag{17}$$

where $\mathcal{L}_t$ denotes the training loss for task t, and β_t denotes the corresponding task-balancing weight. To supervise TRFG in accurately predicting the task representation differences between the Stage-I restoration results and the high-quality reference images, we first extract multi-level task features from I_0 and I_{GT} using the frozen downstream task networks. We then construct the target task representation differences $\Delta^*_{t,l}$ through the EMA teacher of the TRFG front-end. Subsequently, we jointly constrain the magnitude and directional consistency of the predicted differences:

$$\begin{aligned}\mathcal{L}_{\mathrm{gap}} = \frac{1}{L}\sum_{l=1}^{L}\Big[&\mathrm{SmoothL1}\left(\hat{\Delta}_{t,l}, \Delta^*_{t,l}\right) \\ &+ \lambda_{\mathrm{dir}}\left(1 - \cos\left(\hat{\Delta}_{t,l}, \Delta^*_{t,l}\right)\right)\Big],\end{aligned} \tag{18}$$

where $\hat{\Delta}_{t,l}$ and $\Delta^*_{t,l}$ denote the predicted and target task representation differences at the l-th level, respectively, and λ_{dir} denotes the weight of the directional consistency term.

Furthermore, to suppress per-sample degradation in restoration quality caused by task feedback, we define the Stage-II and Stage-I reconstruction errors as

$$e^{(i)}_{\mathrm{II}} = \mathcal{E}\left(I^{(i)}_{\mathrm{out}}, I^{(i)}_{\mathrm{GT}}\right), \qquad e^{(i)}_{\mathrm{I}} = \mathcal{E}\left(I^{(i)}_{0}, I^{(i)}_{\mathrm{GT}}\right). \tag{19}$$

where $\mathcal{E}(\cdot,\cdot)$ denotes the per-sample mean L_1 error, and the Stage-I reconstruction error is computed with gradient propagation stopped. The corresponding restoration preservation loss is defined as:

$$\mathcal{L}_{\mathrm{preserve}} = \frac{1}{B}\sum_{i=1}^{B}\left[e^{(i)}_{\mathrm{II}} - e^{(i)}_{\mathrm{I}}\right]_+, \tag{20}$$

where B denotes the batch size, and $[\cdot]_+$ denotes the ReLU operation. This loss imposes a penalty only when the Stage-II reconstruction error of a sample exceeds its Stage-I reconstruction error, thereby limiting the adverse impact of task feedback optimization on the basic restoration quality.

B.5 Comparison Protocol

To ensure fair comparisons, we retrain all comparison methods using a unified two-stage training procedure. In Stage I, each method undergoes basic image restoration training. In Stage II, all methods are trained with the same downstream task networks and task losses for task-driven restoration. All methods use identical data splits, degradation settings, and evaluation protocols to minimize the influence of differences in training configurations and supervision on experimental results, thereby enabling a fair comparison of image restoration quality and downstream task performance.

C Additional Quantitative Results

This section provides detailed comparisons of TaskIR with other restoration methods in terms of image restoration quality and downstream task performance under different degradation conditions. Note that non-additive metrics, including mIoU, Dice, and mAP, are computed over the entire validation set. Therefore, their overall results do not necessarily equal the arithmetic mean of the results across individual degradation types.

Table 4: Image restoration quality on ImageNet-1K under eight degradations. Each entry reports PSNR (dB)/SSIM. The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	ImageNet-1K							
		Gaussian	Shot	Defocus	Motion	Snow	Fog	Brightness	JPEG
AirNet	CVPR'22	27.26/0.75	27.14/0.75	24.55/0.67	23.58/0.71	24.43/0.81	21.41/0.86	25.97/0.91	27.27/0.78
PromptIR	NeurIPS'23	27.84/0.76	27.69/0.77	26.76/0.76	27.68/0.83	26.68/0.85	24.37/0.92	30.28/0.94	27.63/0.79
NDR-Restore	TIP'24	27.81/0.77	27.63/0.78	26.25/0.74	27.77/0.83	26.55/0.84	24.66/0.92	29.94/0.94	27.48/0.78
AdaIR	ICLR'25	27.91/0.77	27.74/0.77	27.25/0.77	27.71/0.83	26.98/0.86	25.32/0.93	31.02/0.94	27.68/0.79
DFPIR	CVPR'25	27.98/0.77	27.83/0.77	**27.84**/0.79	**29.57/0.87**	**27.64**/0.87	25.54/0.93	31.64/0.94	27.80/0.79
MoCE-IR	CVPR'25	25.05/0.59	24.96/0.60	23.01/0.57	23.79/0.70	23.59/0.72	19.78/0.83	27.02/0.89	26.66/0.75
VLU-Net	CVPR'25	27.99/0.77	27.86/0.78	27.21/0.78	27.89/0.84	27.16/0.86	24.99/0.93	31.26/0.94	27.64/0.79
BaryIR	TPAMI'26	28.12/0.78	28.01/0.78	27.62/0.79	28.95/0.86	27.53/0.87	25.67/0.93	**32.40**/0.95	27.79/0.79
C2SSM	CVPR'26	27.88/0.78	27.76/0.78	26.95/0.76	26.90/0.81	26.56/0.85	24.42/0.92	30.45/0.94	27.64/0.79
MIRAGE	ICLR'26	28.22/0.79	28.09/0.79	27.26/0.77	27.27/0.83	27.43/0.87	24.65/0.93	31.92/**0.95**	27.79/0.79
Ours	–	**28.32/0.79**	**28.19/0.80**	27.64/**0.79**	29.13/0.86	27.54/**0.87**	**26.39/0.94**	31.43/0.94	**27.85/0.80**

Table 5: Image restoration quality on Cityscapes under eight degradations. Each entry reports PSNR (dB)/SSIM. The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	Cityscapes							
		Gaussian	Shot	Defocus	Motion	Snow	Fog	Brightness	JPEG
AirNet	CVPR'22	32.56/0.86	32.98/0.86	32.81/0.90	30.02/0.88	30.01/0.90	22.39/0.92	31.74/0.97	33.98/0.92
PromptIR	NeurIPS'23	33.36/0.87	33.64/0.88	35.86/0.94	36.78/0.96	31.89/0.92	26.40/0.96	34.92/0.98	34.63/0.92
NDR-Restore	TIP'24	33.41/0.88	33.71/0.89	35.23/0.93	36.59/0.96	31.61/0.92	27.86/0.96	36.54/0.98	34.39/0.92
AdaIR	ICLR'25	33.47/0.87	33.71/0.88	36.66/0.95	36.99/0.96	32.23/0.93	27.01/0.96	34.86/0.99	34.69/0.93
DFPIR	CVPR'25	33.39/0.87	33.69/0.87	37.08/0.95	38.63/0.97	32.88/0.92	29.01/0.97	38.89/0.99	34.93/0.92
MoCE-IR	CVPR'25	28.51/0.60	28.74/0.63	29.96/0.83	30.61/0.89	27.20/0.72	21.98/0.89	32.01/0.96	32.74/0.87
VLU-Net	CVPR'25	33.23/0.88	33.73/0.89	36.45/0.94	36.26/0.96	32.44/0.93	27.13/0.96	38.40/0.99	34.46/0.92
BaryIR	TPAMI'26	33.73/0.88	34.03/0.89	36.54/0.94	38.15/0.97	32.85/0.93	27.96/0.97	**40.00**/0.99	34.85/0.93
C2SSM	CVPR'26	33.49/0.88	33.78/0.89	35.98/0.94	36.58/0.96	31.75/0.92	27.63/0.97	34.96/0.99	34.65/0.92
MIRAGE	ICLR'26	33.86/0.89	34.09/0.90	36.41/0.94	36.78/0.96	32.62/0.94	26.58/0.97	38.79/0.99	34.80/0.93
Ours	–	**34.39/0.91**	**34.65/0.91**	**37.48/0.95**	**38.81/0.97**	**33.15/0.94**	**30.14/0.98**	38.70/**0.99**	**35.13/0.93**

C.1 Image Restoration Performance

To further analyze the restoration capabilities of different methods under specific degradations, we report PSNR and SSIM for eight degradation types on ImageNet-1K, Cityscapes, and PASCAL VOC2012, as shown in Tabs. 4–6. Compared with the overall results presented in the main text, the degradation-specific metrics provide a more detailed assessment of each method's ability to adapt to different degradations. By comparing restoration performance across degradation types, we further examine the differences among methods under noise, blur, and other degradation conditions.

C.2 Downstream Task Performance

To evaluate the impact of different degradations on downstream vision tasks, we report detailed image classification, semantic segmentation, and object detection performance under eight degradation types.

Image Classification. As shown in Tab. 7, we report Top-1 and Top-5 accuracy for eight degradation types on ImageNet-1K to compare the effects of different restoration methods on image classification performance and assess their ability to support category discrimination under various degradation conditions.

Semantic Segmentation. As shown in Table 8, we report the mIoU and Dice scores for each degradation type on Cityscapes. The per-degradation results enable further analysis of the impact of different restoration methods on pixel-level semantic prediction performance under complex degradation conditions.

Object Detection. As shown in Table 9, we report mAP@0.5 and mAP@0.5:0.95 on PASCAL VOC2012 under eight degradation conditions. These results are used to evaluate the impact of different restoration methods on object localization and classification performance under different degradation conditions.

Table 6: Image restoration quality on VOC2012 under eight degradations. Each entry reports PSNR (dB)/SSIM. The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	VOC2012							
		Gaussian	Shot	Defocus	Motion	Snow	Fog	Brightness	JPEG
AirNet	CVPR'22	27.17/0.74	27.09/0.74	24.38/0.67	23.74/0.71	24.70/0.82	21.83/0.87	26.41/0.91	27.07/0.77
PromptIR	NeurIPS'23	27.72/0.76	27.64/0.77	26.49/0.76	27.55/0.83	26.83/0.85	24.54/0.92	30.38/0.94	27.40/0.78
NDR-Restore	TIP'24	27.68/0.77	27.57/0.77	25.93/0.74	27.63/0.83	26.66/0.85	24.82/0.93	29.99/0.94	27.25/0.78
AdaIR	ICLR'25	27.78/0.76	27.69/0.77	26.94/0.77	27.52/0.82	27.13/0.86	25.50/0.93	31.06/0.94	27.45/0.79
DFPIR	CVPR'25	27.86/0.76	27.77/0.77	**27.62**/0.79	**29.42/0.87**	**27.79**/0.87	25.62/0.93	31.74/0.95	27.56/0.79
MoCE-IR	CVPR'25	25.01/0.58	24.97/0.60	22.76/0.57	23.80/0.70	23.69/0.73	19.82/0.83	27.08/0.90	26.49/0.75
VLU-Net	CVPR'25	27.89/0.77	27.81/0.78	26.90/0.77	27.67/0.83	27.28/0.86	25.12/0.93	31.29/0.95	27.41/0.78
BaryIR	TPAMI'26	27.99/0.77	27.95/0.78	27.39/0.79	28.75/0.86	27.64/0.87	25.68/0.94	**32.39**/0.95	27.56/0.79
C2SSM	CVPR'26	27.78/0.77	27.71/0.78	26.67/0.76	26.90/0.81	26.73/0.85	24.54/0.93	30.58/0.95	27.41/0.78
MIRAGE	ICLR'26	28.09/0.78	28.03/0.79	26.96/0.77	27.22/0.83	27.64/**0.88**	24.83/0.93	32.10/**0.96**	27.55/0.79
Ours	–	**28.24/0.79**	**28.19/0.80**	27.41/**0.79**	28.94/0.86	27.68/0.87	**26.58/0.94**	31.54/0.95	**27.70/0.79**

Table 7: Classification accuracy under eight degradations on ImageNet-1K. Each entry reports Top-1/Top-5 accuracy (%). The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	ImageNet-1K							
		Gaussian	Shot	Defocus	Motion	Snow	Fog	Brightness	JPEG
AirNet	CVPR'22	63.50/83.40	63.10/84.30	61.90/82.80	65.00/87.10	70.00/88.30	76.50/91.60	77.10/91.80	69.40/89.10
PromptIR	NeurIPS'23	66.50/84.90	67.70/86.70	69.70/90.20	74.80/91.80	74.00/90.50	77.60/92.50	77.40/92.50	70.40/89.10
NDR-Restore	TIP'24	67.40/87.40	68.30/87.20	68.80/89.10	74.70/92.10	73.60/90.40	77.40/92.20	77.30/92.20	69.70/**89.90**
AdaIR	ICLR'25	67.00/87.10	67.50/87.40	70.30/90.00	73.70/92.00	73.90/91.00	**77.90**/92.10	77.10/92.50	69.40/89.90
DFPIR	CVPR'25	**69.10**/88.70	68.10/88.70	**71.70**/89.70	75.50/92.10	74.50/90.60	77.70/91.80	77.60/92.40	70.80/89.30
MoCE-IR	CVPR'25	64.30/85.50	63.60/85.70	54.70/77.70	65.40/87.30	63.30/84.90	73.30/91.40	75.00/91.60	68.00/87.90
VLU-Net	CVPR'25	67.40/87.60	69.10/87.90	69.50/**91.40**	75.20/92.30	74.30/90.90	77.70/92.00	77.40/92.10	70.40/89.10
BaryIR	TPAMI'26	68.10/88.10	68.30/88.00	71.30/89.90	75.30/91.90	74.60/90.70	77.40/92.40	**77.90**/92.20	70.70/89.20
C2SSM	CVPR'26	65.10/87.40	68.00/86.70	69.30/89.30	72.40/91.10	72.30/90.50	77.40/92.30	77.70/92.00	69.70/88.60
MIRAGE	ICLR'26	68.40/87.90	68.50/88.90	67.80/88.10	72.30/91.40	**75.10/91.90**	77.70/**92.60**	77.30/**92.70**	**70.90**/89.90
Ours	–	69.00/**89.00**	**70.10/89.00**	71.30/90.30	**76.30/92.40**	73.40/90.90	77.70/92.20	76.80/92.50	69.40/88.90

C.3 Computational Complexity

We further evaluate the computational complexity of TaskIR in terms of the number of parameters (Params.) and floating-point operations (FLOPs). All methods are evaluated using an input resolution of 256×256 under consistent complexity measurement settings. As shown in Tab. 10, we report the complexity of Stage I and the full inference paths for classification, segmentation, and detection, and compare them with other restoration methods. The full TaskIR includes the corresponding downstream feature extractor and task feedback modules, whereas only the restoration networks are considered for competing methods. The results show that Stage I achieves competitive restoration performance with moderate parameter counts and computational costs. Although incorporating task feedback increases the computational complexity of the full TaskIR, it further improves downstream task performance.

D Additional Visual Results

To further compare the performance of different restoration methods under diverse degradations, we provide additional qualitative results for image restoration and three downstream vision tasks. These visualizations illustrate the differences in degradation removal, texture and structural detail preservation, and downstream predictions across different methods. Specifically, we present restoration results alongside classification, semantic segmentation, and object detection results to provide a more comprehensive evaluation of restoration quality and task performance.

D.1 Image Restoration Results

As shown in Figs. 7–10, we present image restoration results under different degradation conditions on ImageNet-1K, Cityscapes, and PASCAL VOC2012. By comparing the overall restoration results and enlarged local regions, we can further observe the differences among methods in degradation removal, texture recovery, and structural detail preservation.

Table 8: Semantic segmentation performance under eight degradations on Cityscapes. Each entry reports mIoU/Dice (%). The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	Cityscapes							
		Gaussian	Shot	Defocus	Motion	Snow	Fog	Brightness	JPEG
AirNet	CVPR'22	63.14/75.87	66.65/78.75	71.80/82.73	72.85/83.41	71.42/82.48	73.70/83.90	80.20/88.52	68.47/80.07
PromptIR	NeurIPS'23	64.37/76.78	67.96/79.71	77.23/86.54	79.14/87.82	75.07/85.09	79.69/88.18	80.61/88.80	69.96/81.24
NDR-Restore	TIP'24	65.32/77.61	68.78/80.33	76.53/86.04	78.89/87.66	74.16/84.41	79.23/87.82	80.42/88.67	70.03/81.25
AdaIR	ICLR'25	64.57/76.95	68.08/79.72	77.45/86.68	78.81/87.59	75.23/85.22	79.40/87.95	80.66/88.83	69.81/81.09
DFPIR	CVPR'25	64.58/76.95	67.11/79.01	77.69/86.85	79.55/88.09	74.09/84.31	79.20/87.84	80.82/88.93	**70.27/81.44**
MoCE-IR	CVPR'25	59.75/72.94	63.76/76.33	64.90/77.07	70.77/81.75	57.71/70.75	71.62/82.28	79.24/87.88	65.41/77.52
VLU-Net	CVPR'25	64.46/76.76	68.54/80.15	77.49/86.72	79.27/87.90	75.57/85.41	78.99/87.72	80.57/88.77	69.58/80.93
BaryIR	TPAMI'26	63.95/76.41	67.50/79.32	77.36/86.62	79.63/**88.23**	75.41/85.34	79.71/88.19	80.84/88.95	69.95/81.21
C2SSM	CVPR'26	63.98/76.52	69.05/80.55	76.69/86.14	78.00/87.04	74.41/84.59	79.69/88.14	80.87/88.96	69.21/80.62
MIRAGE	ICLR'26	**65.39/77.63**	**69.95/81.22**	77.02/86.38	79.01/87.74	75.75/85.57	79.59/88.06	80.95/89.02	70.21/81.33
Ours	–	64.82/77.17	69.33/80.71	**78.24/87.21**	**79.64**/88.17	**76.43/86.03**	**80.41/88.65**	**81.01/89.07**	68.79/80.29

Table 9: Object detection performance under eight degradations on VOC2012. Each entry reports $mAP_{50}/mAP_{50:95}$ (%). The best and second-best results are shown in bold and underlined, respectively.

Methods	Venue	VOC2012							
		Gaussian	Shot	Defocus	Motion	Snow	Fog	Brightness	JPEG
AirNet	CVPR'22	74.21/55.65	74.71/56.12	66.57/48.92	70.20/51.06	79.72/60.63	85.71/67.52	86.31/68.27	78.38/59.56
PromptIR	NeurIPS'23	77.67/58.66	79.40/60.38	78.83/60.51	82.86/64.22	83.46/64.80	86.86/68.93	86.94/69.10	79.59/60.86
NDR-Restore	TIP'24	77.52/58.50	78.84/60.14	77.48/59.23	83.27/65.00	82.91/64.18	86.77/68.72	86.76/68.98	79.97/61.27
AdaIR	ICLR'25	77.77/58.42	79.13/60.27	79.87/61.67	82.84/64.07	83.66/65.06	86.90/69.00	86.75/68.97	79.79/60.99
DFPIR	CVPR'25	77.96/58.86	79.19/60.14	80.66/62.58	**84.90/66.90**	84.13/65.43	86.87/68.95	86.89/69.10	**80.28/61.63**
MoCE-IR	CVPR'25	63.63/46.16	67.17/49.37	56.71/40.44	72.32/52.98	72.51/53.63	84.90/66.52	86.04/68.10	74.71/56.00
VLU-Net	CVPR'25	77.41/58.74	79.10/60.37	79.73/61.56	83.07/64.75	83.97/65.46	86.87/69.00	86.88/69.12	79.76/61.17
BaryIR	TPAMI'26	78.29/59.26	79.80/60.93	80.42/62.35	84.04/66.05	**84.30/65.68**	86.89/69.03	86.91/69.19	80.03/61.37
C2SSM	CVPR'26	76.64/57.75	78.94/60.06	77.50/59.45	79.96/61.55	82.56/63.61	86.77/68.89	86.93/69.10	79.04/60.16
MIRAGE	ICLR'26	**78.47/59.43**	80.15/61.08	79.02/60.90	80.25/62.12	84.26/65.65	86.91/69.02	**87.10/69.26**	79.99/61.24
Ours	–	78.18/59.32	**80.32/61.36**	**80.83/62.59**	84.06/66.05	83.99/65.26	**86.92/69.04**	86.87/68.98	78.88/60.28

D.2 Downstream Task Results

As shown in Figs. 11–14, we present visualization results for image classification, semantic segmentation, and object detection under eight degradation types on ImageNet-1K, Cityscapes, and PASCAL VOC2012, respectively. By comparing the classification predictions, segmentation results, and detection bounding boxes produced by different restoration methods against the ground truth, we can visually examine the impact of image restoration on category recognition, semantic region prediction, and object localization. These visualizations further illustrate the differences in downstream task performance among restoration methods under diverse degradation conditions.

E LLM USAGE

We used OpenAI's GPT-based Large Language Models (LLMs) to polish the language of the manuscript and improve its clarity and readability.

Table 10: Computational complexity comparison. Ours (I) denotes Stage I, while Ours (Cls), Ours (Seg), and Ours (Det) denote the full inference paths for classification, segmentation, and detection, respectively.

Methods	AirNet	PromptIR	NDR-Restore	AdaIR	DFPIR	MoCE-IR	VLU-Net	BaryIR	C2SSM	MIRAGE	Ours (I)	Ours (Cls)	Ours (Seg)	Ours (Det)
Venue	CVPR'22	NeurIPS'23	TIP'24	ICLR'25	CVPR'25	CVPR'25	CVPR'25	TPAMI'26	CVPR'26	ICLR'26	–	–	–	–
Params. (M)	8.930	35.592	36.962	28.785	182.382	25.354	188.212	53.765	2.842	6.207	28.595	84.813	143.528	61.870
FLOPs (G)	416.272	175.637	156.589	165.148	171.520	96.960	193.782	216.168	6.772	22.460	109.869	233.471	243.604	230.676



Figure 7: Qualitative comparison of image restoration results under gaussian noise and shot noise. Zoom in for a better view.

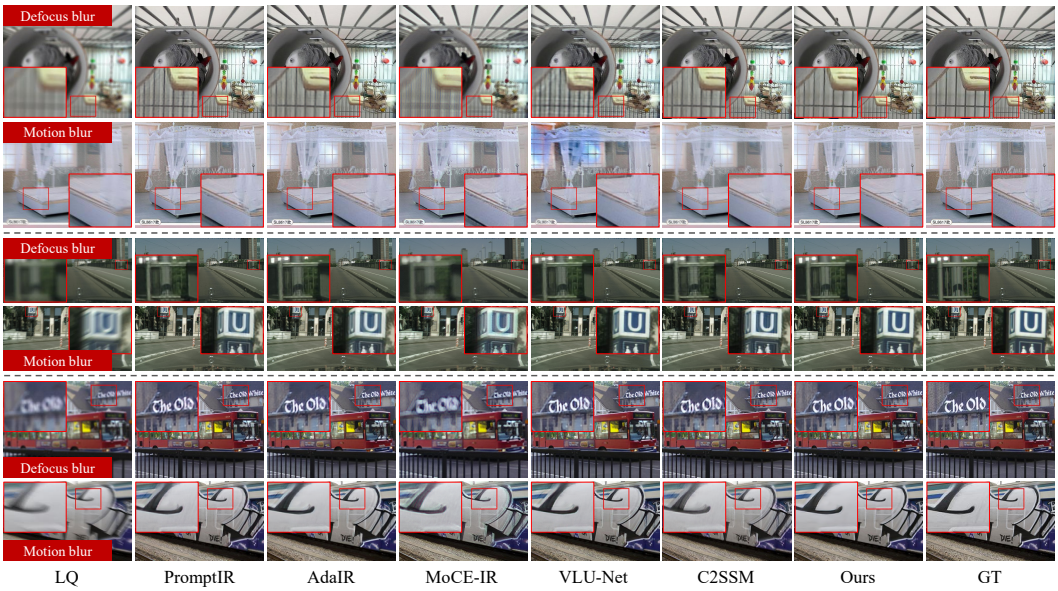

Figure 8: Qualitative comparison of image restoration results under defocus blur and motion blur. Zoom in for a better view.



Figure 9: Qualitative comparison of image restoration results under snow and fog. Zoom in for a better view.



Figure 10: Qualitative comparison of image restoration results under brightness degradation and JPEG compression. Zoom in for a better view.

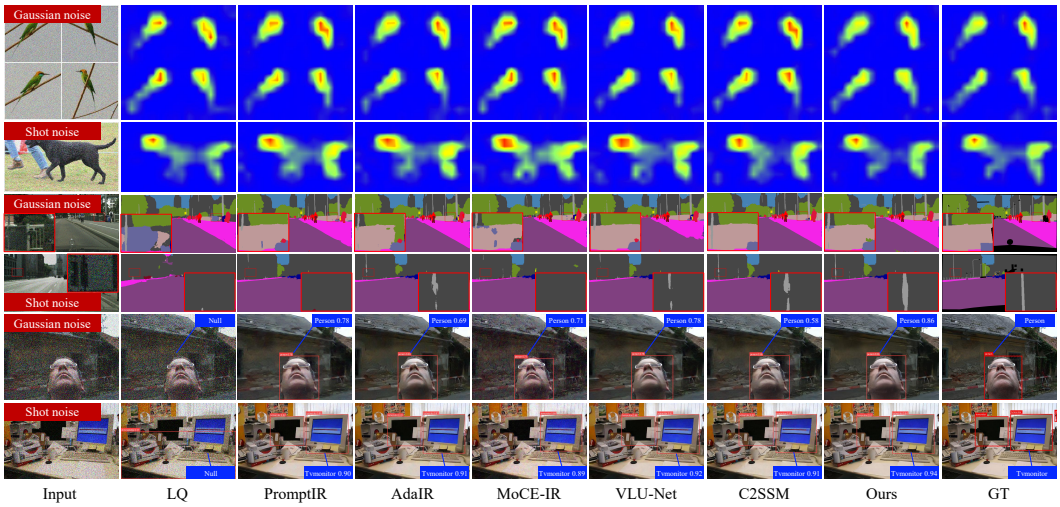

Figure 11: Qualitative comparison of downstream task performance under gaussian noise and shot noise. Zoom in for a better view.

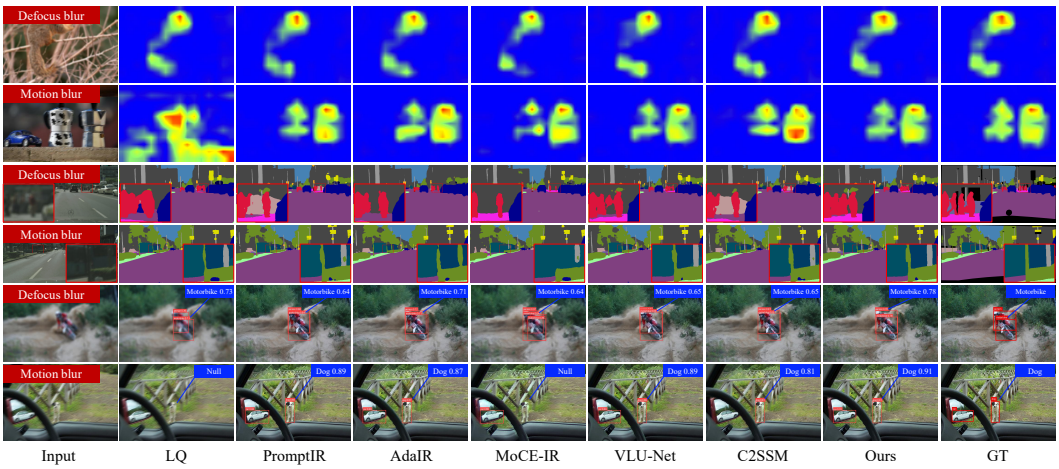

Figure 12: Qualitative comparison of downstream task performance under defocus blur and motion blur. Zoom in for a better view.

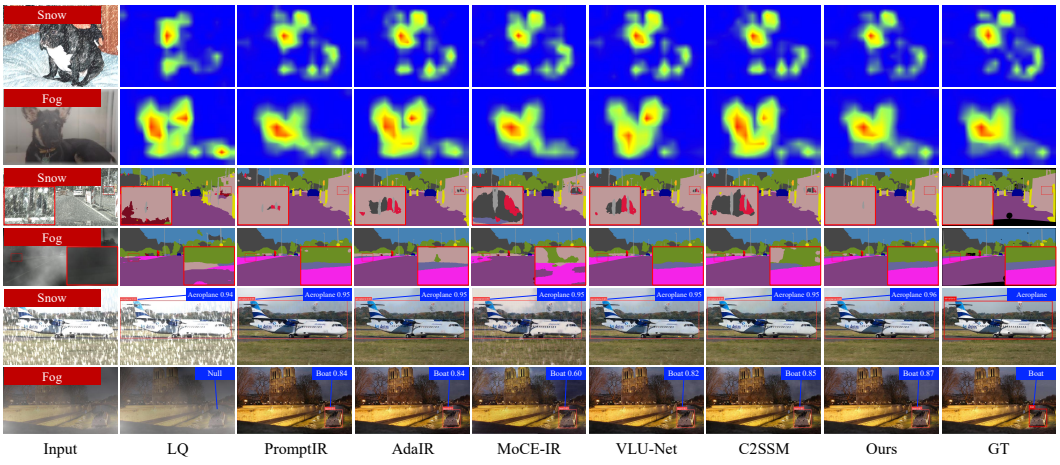

Figure 13: Qualitative comparison of downstream task performance under snow and fog. Zoom in for a better view.

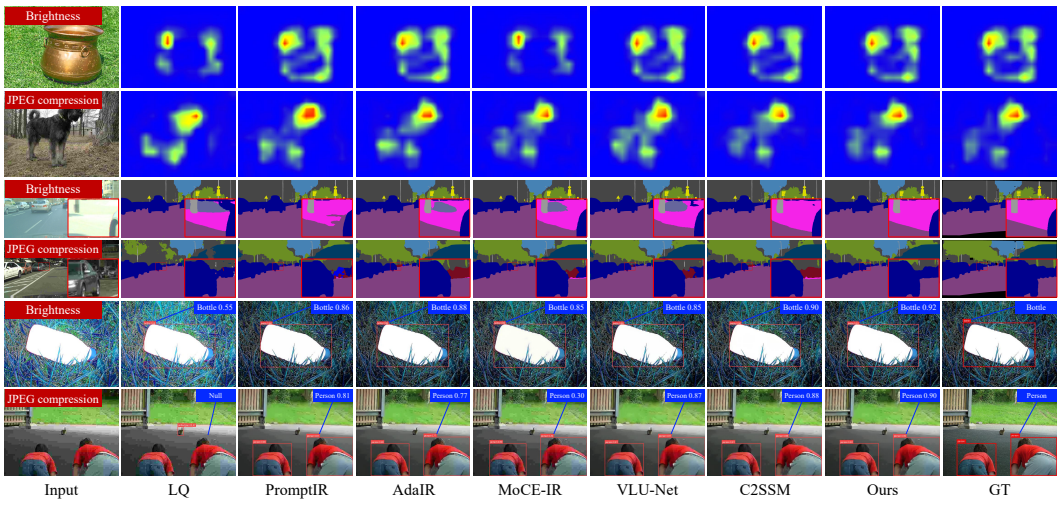

Figure 14: Qualitative comparison of downstream task performance under brightness degradation and JPEG compression. Zoom in for a better view.