Test-Time Degradation Adaptation for Open-Set Image Restoration

Yuanbiao Gou [1] **Haiyu Zhao** [1] **Boyun Li** [1] **Xinyan Xiao** [2] **Xi Peng** [1]

Abstract

In contrast to close-set scenarios that restore images from a predefined set of degradations, open-set image restoration aims to handle the unknown degradations that were unforeseen during the pretraining phase, which is less-touched as far as we know. This work study this challenging problem and reveal its essence as unidentified distribution shifts between the test and training data. Recently, test-time adaptation has emerged as a fundamental method to address this inherent disparities. Inspired by it, we propose a test-time degradation adaptation framework for open-set image restoration, which consists of three components, *i.e.*, i) a pre-trained and degradation-agnostic diffusion model for generating clean images, ii) a test-time degradation adapter adapts the unknown degradations based on the input image during the testing phase, and iii) the adapter-guided image restoration guides the model through the adapter to produce the corresponding clean image. Through experiments on multiple degradations, we show that our method achieves comparable even better performance than those task-specific methods. The code is available at https://github.com/XLearning-SCU/2024-ICML-TAO.

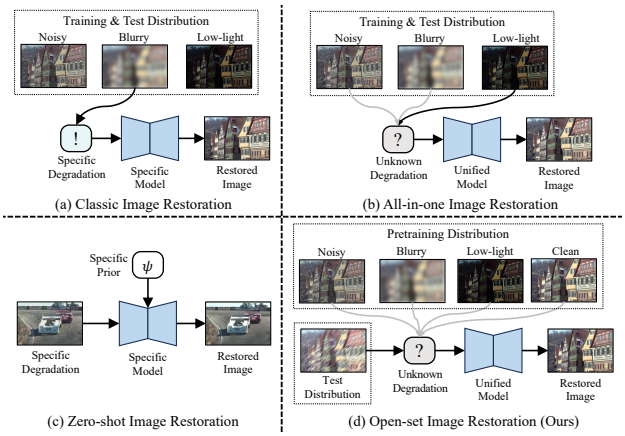

Figure 1. The differences of image restoration (IR) tasks. To be specific, (a) Classic IR works in a close-set scenario where the training and test degradations are the same and known, and customizes a specialized model for each one. (b) All-in-one IR also works in a close-set scenario where the training and test degradations are the same but unknown, and addresses them through a unified model. (c) Zero-shot IR focuses on recovery from single degraded image, which is free from the training degradations, but often requires priors about the test degradations in advance. In contrast, (d) OIR works in an open-set scenario where the test degradations are unknown and different from the pretraining ones. This is analogous to the challenge in natural language processing, which applies the pre-trained large language model to the various downstream tasks not predefined during the pretraining phase.

1. Introduction

In recent years, significant advances have been made in the realm of image restoration, and have demonstrated remarkable capabilities in addressing a multitude of image degradations (Fei et al., 2023; Jiang et al., 2023; Gou et al., 2022; Zhao et al., 2023; Li et al., 2021b). However, a common limitation among most existing methods is their reliance on assumptions rooted in a close-set scenario, *i.e.*, the test degradations closely resemble those encountered during the pretraining phase. This assumption restricts the applicability of these methods to specific types of degradation, hindering their adaptability to a broader range of real-world scenarios where diverse and unforeseen degradations may arise.

To break the above assumption, this work delves into the more challenging and less-touched problem of open-set image restoration (OIR). In contrast to conventional close-set approaches, where models only address the specific degradations encountered during the pretraining phase, OIR poses a formidable challenge by requiring models to handle the unknown degradations absent from the training data. This shift in focus aims to push the boundaries of image restoration, fostering the development of models that exhibit adaptability and resilience in the face of diverse and unexpected degradations. By tackling the OIR problem, this work seeks to pave the way for more robust and versatile solutions applicable to complex degradations in the real-world scenarios.

[1] College of Computer Science, Sichuan University, Chengdu, China [2]Baidu Inc., Beijing, China. Correspondence to: Xi Peng <pengx.gm@gmail.com>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

The essence of OIR lies in addressing the challenges posed by real-world scenarios, where the test data may differ significantly from the training data. This could involve variations in degradation types and levels, lighting conditions and so on that were not explicitly covered in the training data. To tackle this challenge, test-time adaptation (TTA) has recently emerged as an effective methodology to address the inherent disparities between the test and training data (Wang et al., 2021). Specifically, it adapts the pre-trained model during the testing phase based on the specific characteristics of the test data, allowing the model to perform better on a wider range of input scenarios. This is particularly useful in OIR where the test data may vary from the training data, and the pre-trained model should to adapt to unknown and unseen degradations for optimal performance.

Building on this inspiration, we present a **T**est-time degradation **A**daptation framework for **O**pen-set image restoration, dubbed TAO, which harnesses the idea of TTA to provide a robust solution for OIR. Specifically, TAO consists of three components, *i.e.*, a pre-trained image diffusion model (PDM), a test-time degradation adapter (TDA), and the adapter-guided image restoration (AIR). In this framework, PDM is adopted as the foundation model for OIR due to the following considerations. First, PDM captures rich knowledges of generating various high-quality visual scenarios, which could be regarded as a generic pretraining for OIR targeting at producing clean images. Second, PDM is degradation-agnostic and any degradations in the test data could be considered as unforeseen. Note that the test phase of PDM is referred to the process of image generation, *i.e.*, the step-by-step denoising process.

After each denoising step, TDA and AIR are sequentially conducted for adapting to open-set scenarios and guiding image restoration, respectively. Specifically, TDA employs a learnable adapter during the test phase for adapting PDM on the test degraded image. This adapter is devised for domain alignment that aligns the generative domain of PDM to the degraded domain of the test image. In this way, the generated clean image could be translated to the corresponding degraded one, which could be further supervised towards the test degraded image, and in turn updating the generated clean image. AIR is elaborated to conduct this supervised updating for image restoration. In brief, AIR is inspired by the observation that PDM exhibits an intriguing temporal dynamic (Choi et al., 2022) during the step-by-step denoising process. Therefore, AIR dynamically and accordingly adjusts the supervision strategies at different denoising steps for achieving a better performance.

Overall, our contributions could be summarized as follows.

- As far as we know, this could be one of the first work that explicitly studies OIR and discovers its essence, *i.e.*, the unidentified distribution shifts between the test and training data.
- This work reveals that TTA is an effective methodology for OIR, by adapting the pre-trained model on the test data during the test phase, to address the inherent disparities between the test and training data.
- This work presents a test-time degradation adaptation framework for OIR, where two components are devised to adapt open-set scenarios and guide image restoration. Experiments show the effectiveness of this framework.

2. Related Work

In this section, we briefly review recent advances in related topics, which mainly involves test-time adaptation, all-in-one image restoration, and zero-shot image restoration.

Test-time adaptation has shown to be effective at tackling distribution shifts between test and training data, by adapting the pre-trained model on test samples. In the past several years, it has gained increasing attentions and plentiful methods have been proposed. For example, (Wang et al., 2021; Yang et al., 2024) updated the specific model parameters through the test samples by resorting to the unsupervised objectives. Similarly, (Wang et al., 2022a; Gan et al., 2023) aimed to solve the continually changing distribution shifts along the test time. Moreover, (Niu et al., 2023; Zhou et al., 2023) considered more challenging and practical adaptation settings such as single sample, label shifts, and mixed domain shifts. In addition, (Shin et al., 2022; Lee et al., 2023) shifted their focuses on applications beyond image recognition, such as semantic segmentation and pose estimation.

This work focuses on application of solving OIR through TTA methodologies. In brief, we incorporate PDM with an adapter which is updated for adapting to the test degraded image during the test phase. Besides, we also consider the challenging adaptation setting of single test sample.

All-in-one Image Restoration is an emerging issue which has been attracted more and more attentions. Specifically, it aims to address the unknown degradations within a predefined set through one model. For example, (Li et al., 2022) firstly studied this problem and proposed to automatically extract the degradation representation from the unknowingly degraded images for assisting their restoration. (Potlapalli et al., 2023) aimed to encode the degradation information into prompts which are then used to dynamically guide the restoration. (Jiang et al., 2023) focused on using the generated text prompts to guide the latent diffusion model for restoring images with unknown degradations. Although remarkable progress has been achieved, they usually need to be trained on a predefined set of degradations, and recover images from the degradations that are unknown but within this set. In other words, they could only be applied to the

close-set scenarios.

In contrast, this work is devoted to address OIR challenge, wherein the test degradations are unknown during the test phase, meanwhile unforeseen during the pretraining phase. Note that OIR is established on the pre-trained model for image restoration, and adapts the model to the downstream tasks (*i.e.*, unknown and unforeseen degradations) during the test phase. This work empirically regards the diffusion model (Ho et al., 2020) as a generic pretraining task for various image restoration tasks.

Zero-shot Image Restoration focuses on recovering the clean image from a single degraded one, without relying on paired clean-degraded images for training. Classic methods are usually devised for specific tasks. For example, (Shocher et al., 2018) performed super resolution by downsampling the low-resolution image to construct the lower-low training pairs. (Li et al., 2021a) trained dehazing networks by decoupling and coupling the haze image through the physical model of haze formation. In recent, some works focused on methods with more general purpose. For instance, (Kawar et al., 2022; Wang et al., 2022c) applied the decomposition approaches on the predefined linear degradations, and used the diffusion model to address multiple tasks of image restoration. Similarly, (Fei et al., 2023) introduced a conditional guidance for eliminating linear and blind degradations through the predefined and optimizable linear degradation operators, respectively.

This work is different from the above methods on both the problem and solution. On the problem, zero-shot focuses on the restoration from single degraded sample, while open-set focuses on dealing with the unknown degradations that were unforeseen during the pretraining phase, where the degraded sample does not have to be a single one. On the solution, the above methods usually exploit additional degradation priors to assist restoration from single degraded image, while our method only introduces a degradation-agnostic adapter for all unknown and unforeseen degradations.

3. The Proposed Method

In this section, we first briefly review the preliminary knowledges, and then progressively illustrate the proposed TAO and its two core components of TDA and AIR.

3.1. Preliminary

Open-set image restoration is devoted to handle the unknown degradations that were also unforeseen during the pretraining phase. Mathematically, for a given model M pretrained on the distribution D_{train}, OIR requires to restore a test degraded image $y \sim D_{test}$ so that $L(M(y), x)$ is minimal, wherein D_{test} is unknown and different from D_{train}, L is a function to measure the discrepancy, and x is the clean version of y. Formally, OIR essentially poses a problem of the unidentified distribution shifts between the test and training data. For OIR, two fundamental principles should be followed. (i) Unknown: OIR is agnostic to degradations during the test phase, and any prior knowledge about specific degradations is prohibitive. ii) Unforeseen: OIR is independent of a specific pretraining task, and only requires the test degradations were not presented for pretraining.

Test-time adaptation aims to improve the model's performance on new, unseen data by adjusting its parameters based on the test samples. Mathematically, for a given model M pretrained on the distribution D_{train}, TTA optimizes some parameters θ from M on the test samples $y \sim D_{test}$ through the objective function $\min_\theta L(y;\theta)$, wherein D_{test} is different from D_{train}. For TTA, two fundamental principles should be followed. (i) TTA starts with a pre-trained model which has learned general knowledges for the target task. (ii) TTA works on the test data where the training pairs are unavailable. In this work, we adopted the diffusion model as the pre-trained model, the generated and degraded images as the training pairs to implement TTA for OIR task.

The Diffusion Model. Given a data distribution $x_0 \sim q(x_0)$ and a noise distribution $x_T \sim \mathcal{N}(0, I)$, diffusion model (Ho et al., 2020) defines a T-timestep diffusion process, which corrupts x_0 to x_T by sequentially adding the random noise $\epsilon \sim \mathcal{N}(0, I)$, and a T-timestep denoising process, which recovers x_T to x_0 by progressively eliminating the noise ϵ. Importantly, there is an elegant property in the diffusion process, *i.e.*,

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \tag{1}$$

where $t \in \{0, ..., T\}$, $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, $\alpha_i = 1 - \beta_i$ and β_i is the variance of the ith timestep. Based on this formula, the diffusion model learns a denoiser for the denoising process via the following objective, *i.e.*,

$$\nabla_\theta \parallel \epsilon - \mathcal{Z}_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t) \parallel, \tag{2}$$

where $\mathcal{Z}_\theta$ is the denoising network, and t is randomly sampled during the training phase. For a well-trained diffusion model, denoising process progressively yields x_{t-1} from x_t in terms of $\epsilon_t = \mathcal{Z}_\theta(x_t, t)$ until $t = 0$. According to Bayes Theorem, the x_{t-1} could be sampled through the following process, *i.e.*,

$$\begin{aligned} q(x_{t-1}|x_t) &= \mathcal{N}(x_{t-1}; \mu(x_t, \hat{x}_0), \tilde{\beta}_t), \\ \mu(x_t, \hat{x}_0) &= \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t \hat{x}_0 + \sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})x_t}{1 - \bar{\alpha}_t}, \\ \hat{x}_0 &= \frac{x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_t}{\sqrt{\bar{\alpha}_t}}, \ \tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}\beta_t, \end{aligned} \tag{3}$$

where $\hat{x}_0$ is the estimated output from x_t in terms of Eq.(1). To guide the x_{t-1} towards the desired output x_0, the conditional guidance y could be introduced through the following

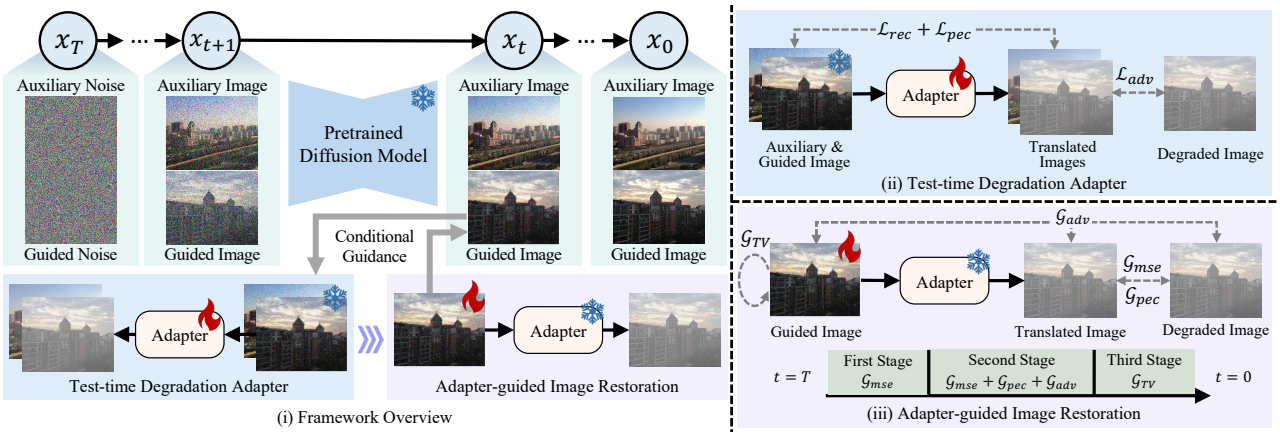

Figure 2. Overview of the proposed method, which exploits (i) a PDM as the generic pre-trained model for OIR. After each denoising step, it first performs (ii) TDA for adapting to the unknown and unseen degradations posed by open-set scenarios, and then conducts (iii) AIR for optimizing the guided image towards the restored clean image. Note: the snowflake icon indicates the image or model is fixed, and the flame icon indicates the image or model will be updated through the gradients.

formula (Dhariwal & Nichol, 2021; Fei et al., 2023), *i.e.*,

$$q(x_{t-1}|x_t, y) \propto N(x_{t-1}; \mu(x_t, \hat{x}_0) + s\mathcal{G}, \tilde{\beta}_t), \quad (4)$$

where s is the scale of guidance, $\mathcal{G} = \nabla_{\hat{x}_0} \log p_\phi(y|\hat{x}_0)$, and $p_\phi(\cdot)$ is a model bridging the gaps between $\hat{x}_0$ and y. For image restoration, the test degraded image is introduced as the y, and the x_0 is the desired clean output.

3.2. Framework Overview

To achieve OIR through PDM, the key lies in implementing $p_\phi(\cdot)$ to produce the appropriate guidances of gradients $\mathcal{G}$, guiding the generated image $\hat{x}_0$ towards the clean image x_0 of the test degraded image y. In this work, we implement $p_\phi(\cdot)$ through TDA and AIR, which are devised for open-set scenarios and image restoration, respectively. To overcome the challenges posed by open-set scenarios, TDA introduces a degradation-agnostic adapter during the test phase (*i.e.*, denoising process of PDM) for adapting PDM to the unknown and unforeseen degradations. This adapter is implemented by a simple neural network ϕ, and optimized once after each denoising step towards aligning the domain of the $\hat{x}_0$ to that of the y. Through the adapter, $\hat{x}_0$ could be translated to the corresponding degraded one $\phi(\hat{x}_0)$, which could be further supervised towards the y, and in turn produce $\mathcal{G}$ to update the $\hat{x}_0$ towards the desired x_0. AIR is elaborated to conduct this supervised process for image restoration. Specifically, AIR is inspired by the observation that a PDM exhibits an intriguing temporal dynamic during the denoising process, *i.e.*, $\hat{x}_0$ are generated from the unrecognizable contents to the perceptually rich contents, and finally the imperceptible details. Therefore, AIR dynamically adjusts the supervision strategies by gradually shifting the focuses from high- to low-level contents during the denoising process.

The framework of TAO is illustrated in Fig.2, where a PDM is fixed to progressively denoising the pure noise x_T to the clean image x_0. For the test degraded image y, TAO first samples two maps of random noise at $t = T$, and thus two images will be generated after each denoising step. We refer to the image updated towards the clean image of the y as the guided image x_t^g, and the other one as the auxiliary image x_t^a. We introduce the x_t^a to prevent TDA from falling into trivial solutions, and assist AIR to guide the x_t^g towards the desired clean image x_0^g. At each step t, the images x_t^g, x_t^a, and y are first fed into TDA for optimizing the adapter ϕ through the objectives for domain alignment, and then sent into AIR for updating x_t^g towards x_0^g, which best matches y undergoing ϕ, through the conditional guidance in Eq.(4).

3.3. Test-time Degradation Adapter

The essence of OIR lies in addressing the unidentified distribution shifts between the test and training data, while TTA recently emerges as an effective methodology to address this inherent disparities. With the motivations, TDA is presented for adapting PDM to the downstream shifted distribution of degradations during the test phase.

TDA employs a four-layer convolution neural network as the adapter ϕ, which is optimized for aligning the domain of the generated images $\hat{x}_0^g$, $\hat{x}_0^a$ to that of the test degraded image y. Note that $\hat{x}_0^g$ and $\hat{x}_0^a$ are the estimated outputs from x_t^g and x_t^a in terms of Eq.(1), respectively. Specifically, we introduce adversarial training which has shown to be effective in aligning the source and target domains, and the translated samples usually capture rich visual characteristics inherent to the target domain. Therefore, TDA aligns $\phi([\hat{x}_0^g, \hat{x}_0^a])$ to y which involves significant visual degradations through

$$\mathcal{L}_{adv} = \log(1 - \mathcal{D}(\phi([\hat{x}_0^g, \hat{x}_0^a])), \quad (5)$$

where $[\cdot]$ denotes the concatenation along the dimension of

batch size, and $\mathcal{D}$ is a discriminator optimized through

$$\mathcal{L}_{dis} = -\log \mathcal{D}([y, y]) - \log(1 - \mathcal{D}(\phi([\hat{x}_0^g, \hat{x}_0^a]))). \quad (6)$$

Meanwhile, TDA adopts a reconstruction loss to maintain the content consistency during the domain alignment, *i.e.*,

$$\mathcal{L}_{rec} = \| [\hat{x}_0^g, \hat{x}_0^a] - \phi([\hat{x}_0^g, \hat{x}_0^a]) \|_2^2 . \quad (7)$$

In addition, the perceptual loss is introduced to remain the semantic consistency during the domain alignment, *i.e.*,

$$\mathcal{L}_{pec} = \| \mathcal{V}([\hat{x}_0^g, \hat{x}_0^a]) - \mathcal{V}(\phi([\hat{x}_0^g, \hat{x}_0^a])) \|_2^2, \quad (8)$$

where $\mathcal{V}(\cdot)$ denotes the feature maps extracted from the pretrained VGG network. Overall, the objective for optimizing the adapter ϕ is

$$\mathcal{L}_\phi = \lambda_1 * \mathcal{L}_{rec} + \lambda_2 * \mathcal{L}_{pec} + \lambda_3 * \mathcal{L}_{adv}, \quad (9)$$

where λ_1, λ_2, and λ_3 are the loss weights. By iterative optimization during the denoising process, ϕ gradually translates $\hat{x}_0^g$, $\hat{x}_0^a$ into the domain of y. Note that $\hat{x}_0^a$ is introduced to prevent ϕ from the trivial solution of identity mapping.

3.4. Adapter-guided Image Restoration

With the adapter ϕ, $\hat{x}_0^g$ could be translated into the degraded domain of y through $\phi(\hat{x}_0^g)$, which could be further supervised towards the y, and in turn produce the $\mathcal{G}$ to update the generated images $\hat{x}_0^g$ to the desired clean output x_0^g. AIR is devised to perform this supervised updating process.

Existing studies (Choi et al., 2022) have show that the denoising process exhibits an intriguing temporal dynamic, *i.e.*, $\hat{x}_0^g$ is generated from the unrecognizable contents to the perceptually rich contents, and finally the imperceptible details. In other words, the contents to be restored are varying as the denoising steps. To take full advantages of PDM, AIR empirically divides the denoising steps into three stages, and guides the restoration through different strategies.

In the first stage where $t \sim T$, the generated images have a lower signal-to-noise ratio, wherein the contents are unrecognizable and only contain some global information such as colors, layouts distributed throughout image. Therefore, AIR applies mse loss on the spatial pixels and color channels to learn global contents, *i.e.*,

$$\mathcal{G} = \gamma_1 * \nabla_{\hat{x}_0} \| y - \phi(\hat{x}_0) \|_2^2, \quad (10)$$

where γ_1 is the loss weight and $\hat{x}_0 \in \{\hat{x}_0^g, \hat{x}_0^a\}$. Note that AIR also use this guidance on the auxiliary image x_0^a so that it involves the similar global properties as y, since it is beneficial to the optimization of adapter on most degradations. In addition, a linear warmup is employed on γ_1 to gradually involve this guidance.

In the second stage where $t \sim T/2$, PDM prefers to generate the perceptually rich contents, and thus AIR focuses more on restoring the perceptual contents from y. To this end, the perceptual loss is adopted, *i.e.*,

$$\mathcal{G}_{pec} = \nabla_{\hat{x}_0^g} \| \mathcal{V}(y) - \mathcal{V}(\phi(\hat{x}_0^g)) \|_2^2 . \quad (11)$$

Meanwhile, an adversarial loss is also introduced to boost the perceptual quality, *i.e.*,

$$\begin{aligned} \mathcal{G}_{adv} &= \nabla_{\hat{x}_0^g} \log(1 - \mathcal{D}(\hat{x}_0^g - y)) \\ \mathcal{L}_{dis} &= -\log(1 - \mathcal{D}(\hat{x}_0^g - y)) \\ &\quad - \log \mathcal{D}([\hat{x}_0^g, \hat{x}_0^a] - \phi([\hat{x}_0^g, \hat{x}_0^a])). \end{aligned} \quad (12)$$

With this loss, $\hat{x}_0^g$ will be further updated toward the clean image whose degraded version matches y, by aligning the domain of content loss caused by degradations. Besides, the mse loss is also employed to remain the global contents, *i.e.*,

$$\mathcal{G}_{mse} = \nabla_{\hat{x}_0^g} \| y - \phi(\hat{x}_0^g) \|_2^2 . \quad (13)$$

Overall, the guidance loss for the second stage is

$$\mathcal{G} = \gamma_2 * \mathcal{G}_{mse} + \gamma_3 * \mathcal{G}_{pec} + \gamma_4 * \mathcal{G}_{adv}, \quad (14)$$

where γ_{2-4} are the loss weights.

In the third stage where $t \sim 0$, the perceptually rich contents are already presented in $\hat{x}_0^g$. AIR is only required to recover the imperceptible details. In other words, $\hat{x}_0^g$ is very close to the natural image, and thus some extra image priors could be introduced to regulate the final output. In consideration of generality, we only introduce a total variation loss, *i.e.*, $\mathcal{G} = \gamma_5 * TVLoss(\hat{x}_0^g)$, where γ_5 is the loss weight. Note that although the third stage mainly affects the imperceptible details, the pixel values still fluctuate and thus the guidance loss in the second stage is used for a better numerical fidelity.

4. Experiments

In this section, we first introduce the experimental settings, and then show quantitative and qualitative results on multiple degradations. Finally, we perform analysis experiments including ablation studies and result visualizations.

4.1. Experimental Settings

To evaluate the effectiveness of TAO on addressing OIR, we conduct experiments on multiple tasks of image restoration with exactly the same settings, except for the loss weights and guidance scale (λ_{1-3}, γ_{1-5}, s), which are finetuned for different types of degradation to obtain the best results. In experiments, we employ an unconditional image diffusion model (Dhariwal & Nichol, 2021) pretrained on ImageNet (Deng et al., 2009) as PDM, and sets the timestep as $T = 1000$, which is further divided into three stages in a

Table 1. Quantitative results of image dehazing on HSTS. Our method outperforms zero-shot methods, and obtains comparable even better results than classic methods through supervised learning. All the compared methods are specifically designed for image dehazing.

Metrics	Classic Learning Methods				Zero-shot Methods						OIR
	DEN	MSN	AOD	CAP	DCP	BCCR	GRM	NLD	DDIP	YOLY	Ours
PSNR ↑	23.96	18.40	19.15	21.69	18.42	15.75	17.41	18.18	18.22	21.24	22.29
SSIM ↑	0.902	0.826	0.860	0.868	0.854	0.783	0.833	0.814	0.832	0.835	0.877

Figure 3. Qualitative results on image dehazing, from which one could observe that existing methods excessively dehazing resulting in darkening and/or artifacting of the images. In contrast, our method obtains clearer results which are closer to the natural ground truths.

heuristic way, *i.e.*, the first stage 999-700, the second stage 700-50, and the third stage 50-0. Since PDM is pre-trained on large-scale natural images for high-quality image generation, any degradation in the test image could be regraded as the unforeseen. Both the adapter and discriminators are four-layer convolutional networks, and optimized once at each denoising step through Adam optimizer with default learning rate of 1e-3. First nine layers of pretrained VGG-16 (Simonyan & Zisserman, 2014) network are employed to extract the semantic feature maps for calculating the perceptual loss. For evaluations, considering the specificity of our methods, *i.e.*, adaptation on single test sample, we mainly compare it with those task-specific zero-shot methods, and some classic learning-based methods. To access their performance, PSNR and SSIM metrics are employed, and all experiments are conducted through PyTorch framework on Ubuntu20.04 with GeForce RTX 3090 GPUs.

4.2. Experimental Results

Here, we evaluate our method on the tasks of image dehazing, low-light image enhancement, and image denoising.

Image Dehazing aims to eliminate the haze and boost the visibility of hazy image, which has spawned many methods in the past decades. In experiments, we introduce the HSTS dataset from RESIDE (Li et al., 2018a) for evaluations. To be specific, RESIDE is a large scale haze image dataset, and HSTS is one of the test subsets, which contains 10 synthetic and 10 real-world hazy images. Before the evaluation, we first center crop the images along the shorter edges, and then resize them to match the image size of the PDM.

For a comprehensive comparison, we compare TAO with 10 representative methods which are specifically designed for image dehazing, and could be roughly divided into two groups, *i.e.*, classic learning-based and zero-shot methods. Specifically, classic learning-based methods are DehazeNet (DEN) (Cai et al., 2016), MSCNN (MSN) (Ren et al., 2016), AOD-Net (AOD) (Li et al., 2017), and CAP (Zhu et al., 2015). Zero-shot methods are DCP (He et al., 2010), BCCR (Meng et al., 2013), GRM (Chen et al., 2016), NLD (Berman et al., 2016), DDIP (Gandelsman et al., 2019) and YOLY (Li et al., 2021b).

The quantitative results are presented in Tab.1, from which one could see that our method achieves comparable even better results than those task-specific methods. To be specific, our method obtains the best performance in zero-shot methods, and outperforms YOLY and DCP with 1.05/0.042 and 3.87/0.023 according to PSNR/SSIM metrics, respectively. Meanwhile, even compared with the classic learning-based methods, our method still outperforms most of them and achieves comparable results with the others. For example, our method obtains 3.89/0.051, 3.14/0.017, and 0.60/0.009 higher PSNR/SSIM values than MSN, AOD and CAP, respectively. Although our method is lower than DEN, the performance gaps are not huge, *e.g.*, 1.67/0.025 lower than DEN in terms of PSNR/SSIM metrics. In addition, we also show and analyze the qualitative results in Fig.3, from which one could observe that our method achieves better fidelity and realness than the compared methods.

Table 2. Quantitative results of low-light image enhancement on LOL dataset. Our method obtains the best SSIM values and the second PSNR values in the zero-shot methods, and only the classic learning-based MBLLEN outperforms our method on both the two metrics.

Metrics	Classic Learning Methods				Zero-shot Methods					OIR
	LNet	RNet	MBLLEN	EGAN	ExCNet	ZDCE	ZDCE+	RRDN	GDP	Ours
PSNR ↑	13.39	17.87	18.03	14.85	17.15	15.33	15.71	11.24	15.75	17.10
SSIM ↑	0.623	0.699	0.809	0.815	0.720	0.763	0.765	0.543	0.665	0.804

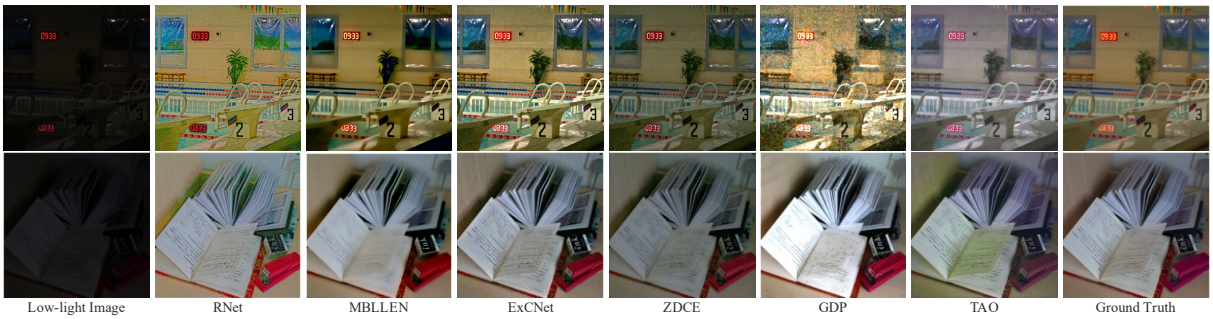

Figure 4. Qualitative results on low-light image enhancement, from which one could see that our results are not as smooth as MBLLEN nor as dark as ZDEC. Although there are slight color biases from ground truths, our method achieves a rational lighting of the dark images.

Low-light Image Enhancement aims at improving the perception of the image captured in the environment with poor illumination. To evaluate our method on this task, we introduce the test subset from LOL (Wei et al., 2018b) dataset. LOL includes 485 training and 15 test image pairs of low- and normal-light, which contain noises produced during the photo capture process. Since the images have a resolution of 400×600, we first center crop them along the shorter edges, and then resize them to be applicable to the PDM.

For a comprehensive comparison, we compare our method with nine low-light image enhancement methods, which are divided into two categories, *i.e.*, classical learning-based and zero-shot methods. To be specific, the classical learning-based methods are LightenNet (LNet) (Li et al., 2018b), Retinex-Net (RNet) (Wei et al., 2018a), MBLLEN (Lv et al., 2018), and EnlightenGAN (EGAN) (Jiang et al., 2021). The zero-shot methods are ExCNet (Zhang et al., 2019), Zero-DCE (ZDEC) (Guo et al., 2020), Zero-DCE++ (ZDEC+) (Li et al., 2021c), RRDNet (RRDN) (Zhu et al., 2020), and GDP (Fei et al., 2023). The above methods except GDP are specially designed for low-light image enhancement, while our method is devised for more general purpose of OIR, and the only difference for different tasks lies in the loss weights and the guidance scale.

The quantitative results are shown in Tab.2. From the Table, one could see that our method achieves comparable even better results than those specific-designed methods. In the zero-shot methods, our method achieves the best and the second performance in terms of SSIM and PSNR metrics, respectively. In PSNR, our method outperforms the other methods except ExCNet with the margins of 1.35-5.86. Although ExCNet achieves a slightly better PSNR value of 0.05, our method achieves the best SSIM values and exceeds it with a margin of 0.084. Besides, even compared with the classic learning-based methods, our method is competitive and obtains better results than most of them. For instance, our method obtains 3.71/0.181 higher PSNR/SSIM value than LNet, 2.25 higher PSNR value than EGAN, and 0.105 higher SSIM value than RNet. Only MBLLEN outperforms our method on both the PSNR and SSIM metrics with the slight margins of 0.93 and 0.005, respectively. In addition, we present and analyze the qualitative results in Fig.4, which demonstrate our method enlightens the dark images reliably and reasonably.

Table 3. Quantitative results of image denoising on Kodak dataset. N.I. denotes the noise images. From the table, one could see that our method is comparable to even better than those zero-shot image denoising methods.

Metrics	N.I.	BM3D	N2S	N2V	DIP	Ours
PSNR ↑	26.35	28.63	28.71	27.17	27.58	28.49
SSIM ↑	0.757	0.795	0.864	0.846	0.808	0.830

Image Denoising aims to remove noises from an noise image. To evaluate our method, we introduce Kodak24 dataset which consists of 24 natural clean images, and is commonly used for testing image denoising methods. Similarly, we center crop and resize the image to match the size of the PDM, and obtain the noise images by adding the Gaussian noises with the noise level of $\sigma = 30$ to the clean images.

Figure 5. Qualitative results on image denoising, from which one could observe that existing methods excessively denoising resulting in smoothing and/or artifacting of the images. In contrast, our method obtains clearer and sharper results which are closer to ground truths.

Table 4. Results on different strategies for domain alignment to unknown and unseen degradations. These strategies were implemented based on BasicSR (Wang et al., 2022b). VGAN-S/B denote the Sigmoid Layer plus BCELoss and BCEWithLogitsLoss, respectively.

Metrics	Haze Image	VGAN_S	VGAN_B	LSGAN	WGAN	WGAN_Softplus	Hinge
PSNR ↑	14.74	22.29	22.06	20.76	13.61	21.47	20.44
SSIM ↑	0.770	0.877	0.870	0.873	0.619	0.864	0.860

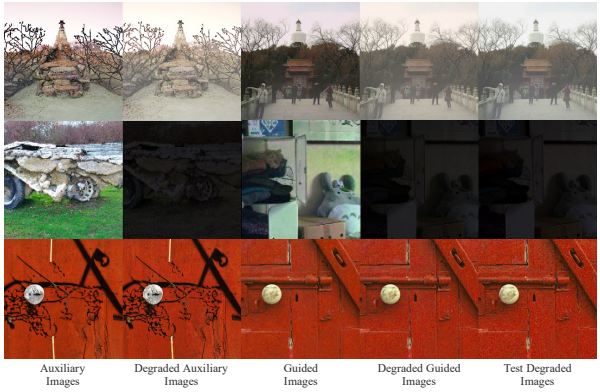

Figure 6. Visualizations from TDA wherein Auxiliary/Guided Images are translated to Degraded Auxiliary/Guided Images by TDA, whose domains align well to that of Test Degraded Images.

We compare our method with four zero-shot image denoising methods, *i.e.*, BM3D (Dabov et al., 2007), N2S (Batson & Royer, 2019), N2V (Krull et al., 2019), and DIP (Ulyanov et al., 2018). The above methods except for DIP are specifically designed for image denoising. The quantitative results are presented in Tab.3. From the table, one could see that our method obtains comparable even better results than those task-specific methods. To be specific, our method obtains 0.035 and 0.022 higher SSIM value than BM3D and DIP, respectively. Meanwhile, our method also outperforms N2V and DIP in PSNR value with a large margin of 1.32 and 0.91, respectively. Although not achieving the best performance, our method enjoys the appealing capacity of addressing OIR through one model. In addition, we present the qualitative results in Fig.5, from which one could see that our method obtains clearer results with better fidelity to ground truths than the compared methods.

4.3. Analysis Experiments

In this section, we conduct analysis experiments w.r.t. our proposed TDA and AIR mainly on image dehazing.

The effectiveness of TDA. Since PDM cannot achieve image dehazing without TDA, we demonstrate its effectiveness by i) observing whether it aligns the domain of generated images to that of the degraded image after the optimization during the denoising process, and ii) comparing with different strategies of adversarial training for domain alignment. The visual results are shown in Fig.6, from which one could observe that TDA exhibits a significant capability of domain alignment to the unknown and unforeseen degradations. For example, the haze in the top row is uneven distributed in the test degraded image, and the same effects exist in the degraded auxiliary/guided image. The degree of darkness in the degraded auxiliary/guided image is similar to that in the test degraded image in the second row. In addition, we conduct experiments on different strategies of adversarial training implemented by BasicSR (Wang et al., 2022b), and present the results in Tab. 4. From the table, one could see that, except for WGAN, all strategies achieve the significant performance gains over the haze images. Namely, our adapter generalizes well to different strategies for domain alignment to the unknown and unseen degradations.

The effectiveness of AIR. Here, we first demonstrate the rationality of AIR in Fig.7 which shows a typical denoising process of PDM. From the figure, one could see the signifi-

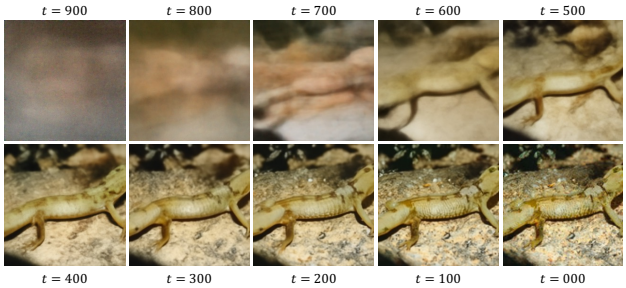

Figure 7. The temporal dynamics of PDM during the denoising process, which motivates our AIR to guide the image restoration through different strategies at different denoising steps.

Table 5. Ablation studies on guidance strategies, where U.G. denotes using the same guidance loss of mean squared error throughout the denoising process. -F.S., -S.S., and -T.S. denote to remove the First Stage, Second Stage, and Third Stage, receptively. The results demonstrate the indispensable roles of the three stages for image dehazing.

Metrics	U.G.	-F.S.	-S.S.	-T.S.	AIR
PSNR ↑	20.49	21.43	21.06	21.37	22.29
SSIM ↑	0.869	0.873	0.865	0.871	0.877

cant temporal dynamics, *i.e.*, the images are generated from the unrecognizable contents to the perceptually rich contents, and finally the imperceptible details. The dynamics also imply the varying of the contents should be focused and restored in the test degraded image, and thus applying different guidance strategies could facilitate image restoration. We present the ablation results in Tab. 5, which show the effectiveness of the three stages on facilitating image dehazing. In addition, we also conduct analysis experiments on stage divisions as shown in Tab. 6, from which one could see that both early and late dividing steps result in the inferior performances. In other words, a proper stage division consistent with the temporal dynamics could improve the performance of image restoration.

Table 6. Analysis experiments on divisions of the three stages. The left is the dividing step between the First and Second stages, and the right is the dividing step between the Second and Third stages. The results demonstrate a proper stage division could significantly improve the restoration.

Metrics	800	700	600	70	50	30
PSNR ↑	21.89	22.29	21.11	21.76	22.29	22.01
SSIM ↑	0.861	0.877	0.868	0.876	0.877	0.872

5. Conclusion

This paper explicitly explores the challenges posed by open-set scenarios, and formally defines the problem of open-set image restoration. To solve this problem, this work reveals its essence from the perspective of distribution shifts, and discovers the methodology of test-time adaptation is adept at addressing this inherent disparities. Motivated by this, we presented a test-time degradation adaptation framework for open-set image restoration, which is ingenious in the following ways. First, it considers a pre-trained image diffusion model as the general pretraining model for solving various tasks of image restoration. Second, it introduces an adapter optimized during the test phase for adapting the pre-trained model to the unknown and unseen test degradations. Third, it dynamically adjusts the guidance strategies following the denoising process to obtain better restoration results. Through experiments on multiple degradations, we demonstrate the effectiveness of our designs.

Impact Statement

This paper considers a novel problem in Image Restoration that restores the clean images from the degraded ones under an open-set scenario. Although there are potential societal consequences of this work, none which we feel serious and must be specifically highlighted here. For example, the most serious cases may be the recovery of images that have been intentionally damaged by someone else, or do not match the facts affecting the decisions made on that.

Acknowledgment

The authors would like to thank the anonymous reviewers and area chair for their valuable comments and constructive suggestions to improve the quality of this paper. This work was supported in part by NSFC under Grant 62176171, U21B2040; in part by the Fundamental Research Funds for the Central Universities under Grant CJ202303; and in part by Sichuan Science and Technology Planning Project under Grant 24NSFTD0130.

References

Batson, J. and Royer, L. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pp. 524–533, 2019.

Berman, D., Avidan, S., et al. Non-local image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1674–1682, 2016.

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

Chen, C., Do, M. N., and Wang, J. Robust image and video dehazing with visual artifact suppression via gradi-

ent residual minimization. In *European Conference on Computer Vision*, pp. 576–591, 2016.

Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., and Yoon, S. Perception prioritized training of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11472–11481, 2022.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8): 2080–2095, 2007.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., and Dai, B. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9935–9946, 2023.

Gan, Y., Bai, Y., Lou, Y., Ma, X., Zhang, R., Shi, N., and Luo, L. Decorate the newcomers: Visual domain prompt for continual test time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7595–7603, 2023.

Gandelsman, Y., Shocher, A., and Irani, M. "double-dip": unsupervised image decomposition via coupled deep-image-priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11026–11035, 2019.

Gou, Y., Hu, P., Lv, J., Zhou, J. T., and Peng, X. Multi-scale adaptive network for single image denoising. *Advances in Neural Information Processing Systems*, 35:14099–14112, 2022.

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1780–1789, 2020.

He, K., Sun, J., and Tang, X. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., and Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

Jiang, Y., Zhang, Z., Xue, T., and Gu, J. Autodir: Automatic all-in-one image restoration with latent diffusion. *arXiv preprint arXiv:2310.10123*, 2023.

Kawar, B., Elad, M., Ermon, S., and Song, J. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.

Krull, A., Buchholz, T.-O., and Jug, F. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2129–2137, 2019.

Lee, T., Tremblay, J., Blukis, V., Wen, B., Lee, B.-U., Shin, I., Birchfield, S., Kweon, I. S., and Yoon, K.-J. Ttacope: Test-time adaptation for category-level object pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21285–21295, 2023.

Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4770–4778, 2017.

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

Li, B., Gou, Y., Gu, S., Liu, J. Z., Zhou, J. T., and Peng, X. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129:1754–1767, 2021a.

Li, B., Gou, Y., Gu, S., Liu, J. Z., Zhou, J. T., and Peng, X. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129:1754–1767, 2021b.

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

Li, C., Guo, J., Porikli, F., and Pang, Y. Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104:15–22, 2018b.

Li, C., Guo, C., and Loy, C. C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4225–4238, 2021c.

Lv, F., Lu, F., Wu, J., and Lim, C. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, pp. 4, 2018.

Meng, G., Wang, Y., Duan, J., Xiang, S., and Pan, C. Efficient image dehazing with boundary constraint and contextual regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 617–624, 2013.

Niu, S., Wu, J., Zhang, Y., Wen, Z., Chen, Y., Zhao, P., and Tan, M. Towards stable test-time adaptation in dynamic wild world. *arXiv preprint arXiv:2302.12400*, 2023.

Potlapalli, V., Zamir, S. W., Khan, S., and Khan, F. S. Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090*, 2023.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., and Yang, M.-H. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, pp. 154–169, 2016.

Shin, I., Tsai, Y.-H., Zhuang, B., Schulter, S., Liu, B., Garg, S., Kweon, I. S., and Yoon, K.-J. Mm-tta: multi-modal test-time adaptation for 3d semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16928–16937, 2022.

Shocher, A., Cohen, N., and Irani, M. Zero-shot super-resolution using deep internal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3118–3126, 2018.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. Deep image prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., and Darrell, T. Tent: Fully test-time adaptation by entropy minimization. In *Proceedings of the International Conference on Learning Representations*, 2021.

Wang, Q., Fink, O., Van Gool, L., and Dai, D. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7201–7211, 2022a.

Wang, X., Xie, L., Yu, K., Chan, K. C., Loy, C. C., and Dong, C. BasicSR: Open source image and video restoration toolbox. https://github.com/XPixelGroup/BasicSR, 2022b.

Wang, Y., Yu, J., and Zhang, J. Zero-shot image restoration using denoising diffusion null-space model. *Proceedings of the International Conference on Learning Representations*, 2022c.

Wei, C., Wang, W., Yang, W., and Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018a.

Wei, C., Wang, W., Yang, W., and Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018b.

Yang, M., Li, Y., Changqing, Z., Hu, P., and Peng, X. Test-time adaption against multi-modal reliability bias. In *Proceedings of the International Conference on Learning Representations*, May 2024.

Zhang, L., Zhang, L., Liu, X., Shen, Y., Zhang, S., and Zhao, S. Zero-shot restoration of back-lit images using deep internal learning. In *Proceedings of the 27th ACM international conference on multimedia*, pp. 1623–1631, 2019.

Zhao, H., Gou, Y., Li, B., Peng, D., Lv, J., and Peng, X. Comprehensive and delicate: An efficient transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14122–14132, 2023.

Zhou, Z., Guo, L.-Z., Jia, L.-H., Zhang, D., and Li, Y.-F. Ods: test-time adaptation in the presence of open-world data shift. In *International Conference on Machine Learning*, pp. 42574–42588. PMLR, 2023.

Zhu, A., Zhang, L., Shen, Y., Ma, Y., Zhao, S., and Zhou, Y. Zero-shot restoration of underexposed images via robust retinex decomposition. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2020.

Zhu, Q., Mai, J., and Shao, L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015.