GenDeg: Diffusion-Based Degradation Synthesis for Generalizable All-in-One Image Restoration

Sudarshan Rajagopalan
Johns Hopkins University
sambasa2@jhu.edu

Nithin Gopalakrishnan Nair
Johns Hopkins University
ngopala2@jhu.edu

Jay N. Paranjape
Johns Hopkins University
jparanj1@jhu.edu

Vishal M. Patel
Johns Hopkins University
vpatel36@jhu.edu

Abstract

Deep learning–based models for All-In-One image Restoration (AIOR) have achieved significant advancements in recent years. However, their practical applicability is limited by poor generalization to samples outside the training distribution. This limitation arises primarily from insufficient diversity in degradation variations and scenes within existing datasets, resulting in inadequate representations of real-world scenarios. Additionally, capturing large-scale real-world paired data for degradations such as haze, low-light, and raindrops is often cumbersome and sometimes infeasible. In this paper, we leverage the generative capabilities of latent diffusion models to synthesize high-quality degraded images from their clean counterparts. Specifically, we introduce GenDeg, a degradation and intensity-aware conditional diffusion model, capable of producing diverse degradation patterns on clean images. Using GenDeg, we synthesize over 550k samples across six degradation types: haze, rain, snow, motion blur, low-light, and raindrops. These generated samples are integrated with existing datasets to form the GenDS dataset, comprising over 750k samples. Our experiments reveal that image restoration models trained on GenDS dataset exhibit significant improvements in out-of-distribution performance as compared to when trained solely on existing datasets. Furthermore, we provide comprehensive analyses on implications of diffusion model-based synthetic degradations for AIOR. The code will be made publicly available.

1. Introduction

Image restoration is a well-studied computer vision problem that aims to reverse the effects of image corruptions or artifacts. It is important for numerous applications, including autonomous driving, imaging and surveillance. Early approaches focused on handling specific degradations such as haze [16, 51], rain [19, 43], snow [11, 52], blur [33, 38] etc. More recent methods such as Restormer [50], MPRNet [49] and SwinIR [26] proposed architectures capable of addressing any single restoration task. However, these approaches are limited to addressing one type of degradation at a time, making them inefficient for scenarios involving multiple types of corruptions.

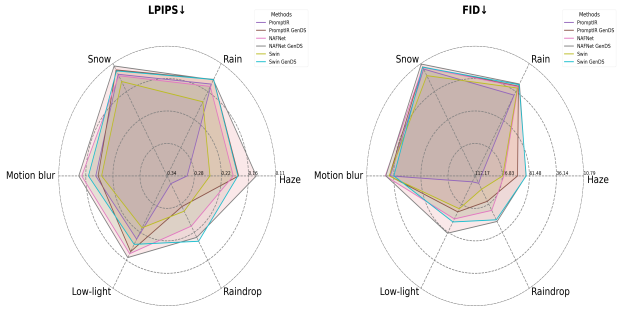

Figure 1. Out-of-distribution performance of image restoration models when trained solely using existing datasets and our proposed GenDS dataset. Significant improvements can be observed across all degradations. Metric values reduce outward.

All-In-One Restoration (AIOR) methods overcome this limitation by employing a single model capable of handling multiple types of degradations. Recent approaches include PromptIR [34], DA-CLIP [31], DiffUIR [55], DiffPlugin [28], InstructIR [13] and AutoDIR [18]. Most AIOR methods are trained using a single dataset for each restoration task such as RESIDE [23] for dehazing, Snow100k [29] for desnowing, Rain13K [49] for deraining and GoPro [33] dataset for motion deblurring. Although these approaches perform well on degradations from these dataset distributions, they often exhibit poor generalization when con-

Project Page: https://sudraj2002.github.io/gendegpage/

fronted with new scenes or out-of-distribution (OoD) degradation patterns, which is very common in real-world scenarios. Recent studies have discussed the problem of generalization in image restoration models in great depth [17, 22]. We hypothesize that the limited generalization of these models is mainly due to two reasons:

1. **Lack of large datasets with real degradations under diverse scenes.** In this paper, we consider the degradations haze, snow, rain, raindrop, motion blur and low-light. Fig. 2 shows the number of synthetic and real images for each degradation from its existing publicly available datasets, along with the number of unique scenes. To the best of our knowledge, we have included most of the existing datasets. Firstly, the figure shows that existing restoration datasets are significantly smaller than those used to train generalizable models for other low-level vision tasks, such as SAM [21] for segmentation and Depth-Anything [47] for depth estimation ($> 1.5M$ samples). This limited dataset size hinders the ability of models to generalize well to diverse real-world scenarios. Secondly, the figure illustrates that degradations such as haze, raindrop, low-light and snow have very few real images compared to synthetic ones. This scarcity is because of challenges in capturing real images under these conditions. For instance, haze is an atmospheric phenomenon which is difficult to simulate in real scenarios. Conversely, motion blur and rain have a decent number of real-world examples as they can be generated from existing videos [15, 25, 33, 38, 43]. Thirdly, degradations such as haze, raindrop and low-light have very limited scene diversity which can further limit the generalization of models. Finally, it can be observed that the number of samples across different degradations is highly imbalanced.
2. **Lack of variety in degradation patterns within datasets.** Previously, we analyzed the distribution of samples for each degradation. Examining individual datasets, especially synthetic ones, reveals that they contain degradations generated using only a particular model. For instance, the images in the RESIDE [23] dataset are generated by the atmospheric haze model [7] with specific parameters. Consequently, training a network on such a dataset can tailor it to work only for the degradation patterns of that dataset, limiting its generalization capability for real-world dehazing.

Due to the above reasons, existing AIOR methods overfit to specific training distributions, thus, limiting their ability to generalize to real-world degradations. To overcome this limitation, we aim to develop robust all-in-one restoration models capable of generalizing to OoD restoration. We define OoD samples as those that are from a test set, whose training set was not utilized for training. Achieving this requires a large number of degraded images with diverse

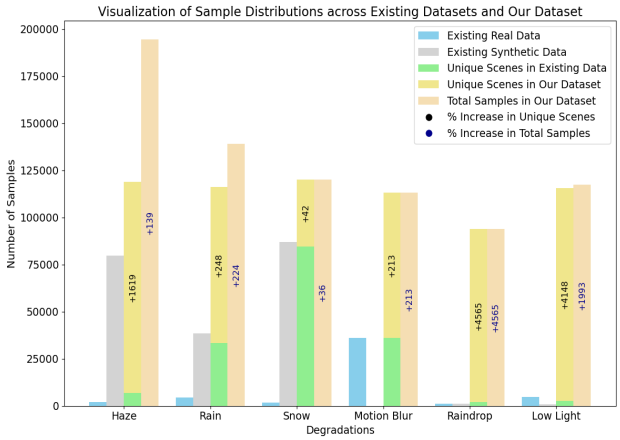

Figure 2. Analysis of real and synthetic image restoration datasets for various degradations. Existing datasets are small and less diverse, especially for haze, low-light, and raindrop. Our diffusion-generated synthetic data substantially increases the number of samples as well as scene diversity.

degradations. Since collecting real-world data for all the degradations is infeasible as already mentioned, we propose a novel degradation generation network capable of producing diverse degradation patterns for each degradation.

Latent diffusion models (LDMs) [37] have demonstrated immense potential in generating diverse high quality images. We propose to harness the generative capability of LDMs for synthesizing diverse degradations. Specifically, we train GenDeg, a diffusion model based on Stable Diffusion that conditions on text prompts, clean images and the levels of degradation to generate diverse degraded images under different degradations. We train GenDeg by combining multiple existing datasets for each degradation type to ensure that it does not heavily rely on a specific degradation pattern or physical model. Thus, it can produce both synthetic and realistic degradations, thereby enriching the diversity of degradation patterns in the generated data. Furthermore, GenDeg offers fine-grained control over the intensity and spatial variations of generated degradations. We achieve this by conditioning GenDeg on the mean (μ) and standard deviation (σ) of the degradation map during training. Using GenDeg, we generate over 550k degraded images from roughly 120k clean images. We augment existing restoration datasets with our generated images to create a dataset, GenDS, with over 750k paired images under haze, rain, snow, motion blur, low-light and raindrop degradations. GenDS provides a significant boost in scene diversity and number of samples as seen from Fig. 2. Training models with our GenDS dataset shows substantial improvements in OoD performance as seen from Fig. 1. In Fig. 3 we illustrate that the diverse degradations in the GenDS dataset help bridge the domain gap between existing and OoD datasets (see Sec. 4.2 for more details). Additionally,

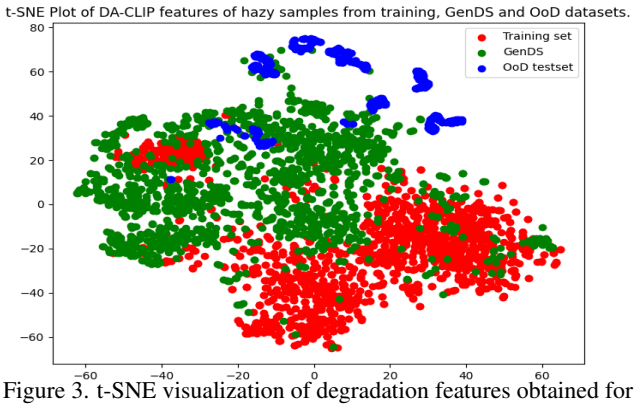

Figure 3. t-SNE visualization of degradation features obtained for hazy samples from existing training data, GenDS dataset and OoD test sets. The features were obtained using DA-CLIP [31].

our dataset consists of the same clean images under different degradations that, to the best of our knowledge, is the first such dataset.

Finally, we train three models on the GenDS dataset, namely NAFNet [10], PromptIR [34] and a Swin Transformer-based model that we propose. Our experiments demonstrate that these models achieve significant improvements in their generalization performance when trained on our large-scale dataset.

In summary, our contributions are as follows:

1. We propose a novel diffusion model-based degradation generation framework, GenDeg, which is capable of producing diverse degradations on any clean image.
2. Using GenDeg, we synthesize over 550k degraded images which when combined with existing datasets forms the comprehensive GenDS dataset comprising approximately 750k samples across highly diverse scenes. Furthermore, each image in GenDS has multiple degraded versions, making it, to the best of our knowledge, the first restoration dataset of its kind.
3. Finally, we train restoration models on the GenDS dataset and demonstrate that incorporating our synthetic data significantly improves the out-of-distribution restoration capabilities of these networks.

2. Related Works

In this section, we discuss relevant works on all-in-one image restoration and diffusion models for synthetic data. Related works on diffusion models are given in supplementary.

2.1. All-in-one image restoration

All-in-one restoration (AIOR) methods employ a single model to address multiple corruptions. Early approaches include All-in-one [24], which employed neural architecture search to select optimal encoders for weather tasks, and Transweather [42], which unified multiple encoders for efficient multi-weather restoration. Airnet and [12] used contrastive loss to learn well-separated degradation representations. PromptIR [34] utilized learnable prompt embeddings to handle multiple degradations. Recent approaches have leveraged the potential of diffusion models for AIOR. DA-CLIP [31] used degradation information from CLIP to guide diffusion-based image restoration. Diff-Plugin [28] leveraged multiple task plugins to guide a latent diffusion model for restoration. DiffUIR [55] proposed selective hourglass mapping to create task-specific distributions with high image quality. AutoDIR [18] developed an automatic approach using vision-language models for degradation detection and restoration. Additionally, InstructIR [13] utilized text guidance as instructions for AIOR. Despite these advancements, no existing work (to the best of our knowledge) has explored using diffusion models to generate degradations. Our approach enables the creation of large datasets with realistic degradations to train generalizable image restoration models.

2.2. Diffusion models for synthetic data

Recent research has focused on leveraging the potential of latent diffusion models for generating synthetic data. [3, 4, 39, 40, 48] demonstrated that diffusion-generated images improve classification and zero-shot classification performance. However, classification tasks do not require preservation of intricate details in the generated images. Some approaches [32, 45] explored the use of diffusion-generated data for pixel-level semantic segmentation task and demonstrated promising directions. Further, [41, 54] showed that augmenting real data with diffusion-generated samples enhances aerial segmentation performance. Nonetheless, these approaches primarily generate only segmentation masks which lack detailed scene content. In contrast, we propose to generate high quality degraded images for image restoration tasks, for which ensuring precise scene consistency is crucial, as discrepancies can degrade restoration performance. Our approach effectively addresses these challenges, leading to significant improvements in the generalization of image restoration models trained with our generated data.

3. Proposed Method

In this section, we detail GenDeg, our diffusion-based method for generating large-scale synthetic data for image restoration. We also discuss the process of data generation, curation and training of restoration models.

3.1. Diffusion based degradation generation

Our goal is to leverage the generative priors of pre-trained diffusion models to produce diverse degradations on clean images while preserving the underlying scene semantics. We consider the synthesis of six common degradations,

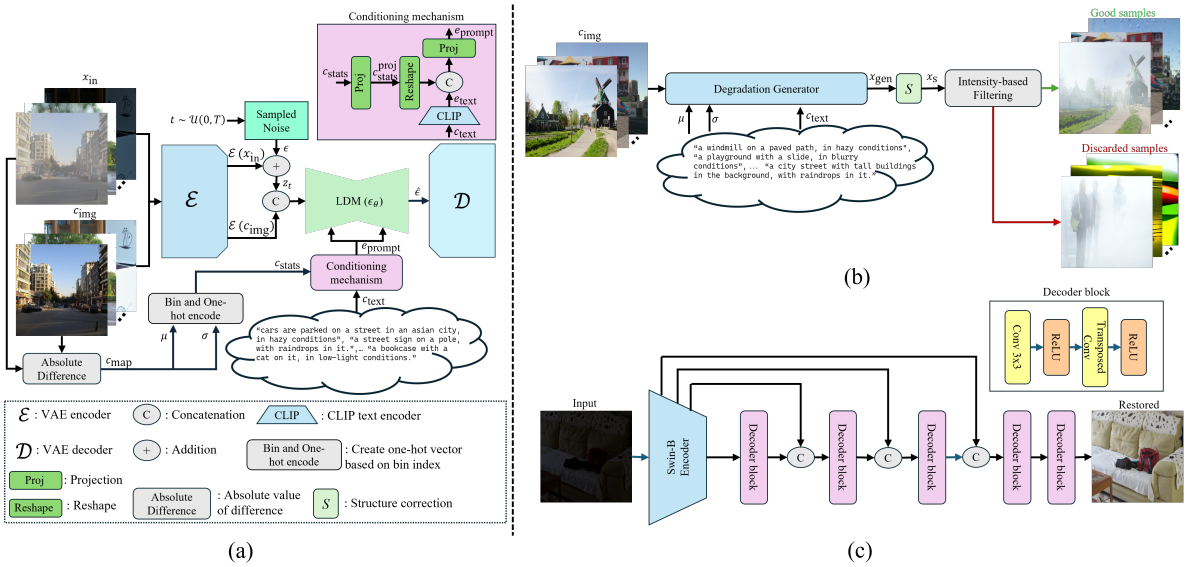

Figure 4. (a) Illustrates the training stage of the GenDeg model where it is trained to condition on the clean image, text prompt and mean intensity (μ) and variation (σ) of the degradation pattern. (b) Shows the inference stage where the model generates a degraded image based on these conditions; and (c) Depicts the architecture of the Swin-transformer-based restoration network.

namely, haze, rain, snow, low-light, motion blur and raindrops. To achieve this, we require a diffusion model that conditions on an input clean image (that needs to be degraded) and a prompt specifying the desired degradation. One popular approach that aligns with our objectives is the InstructPix2pix [5] model. It is a text-based image editing framework that leverages the latent diffusion model (LDM), Stable Diffusion [37], to generate edited images that are consistent with the input image. LDM operates in the latent space of a pre-trained variational auto-encoder [20] whose encoder and decoder are denoted by $\mathcal{E}$ and $\mathcal{D}$, respectively. Given an image x_{in}, image condition c_{img} and text condition c_{text}, the diffusion model (ϵ_θ) minimizes the following objective during training

$$L = \mathbb{E}_{\mathcal{E}(x_{\text{in}}),\, \mathcal{E}(c_{\text{img}}),\, c_{\text{text}},\, \epsilon\sim\mathcal{N}(0,1),\, t} \left[\|\epsilon - \epsilon_\theta(z_t, t, \mathcal{E}(c_{\text{img}}), c_{\text{text}})\|_2^2 \right], \tag{1}$$

where z_t is the noised version of $\mathcal{E}(x_{\text{in}})$ at timestep t of the forward diffusion process and ϵ is the added noise.

In our adaptation, x_{in} represents the degraded image from existing paired restoration datasets, c_{img} is the corresponding clean image, and c_{text} is a text-prompt conveying information about the degradation to be produced. For training the diffusion model, we combine multiple synthetic and real paired image restoration datasets (see supplementary for dataset details). This approach ensures that the diffusion model learns to produce diverse degradation patterns not specific to any single dataset.

The text condition c_{text} includes both high-level scene information along with the degradation specifics. To generate these text descriptions, we process the clean images through the BLIP-2 image captioning model to obtain scene descriptions. We then append degradation-specific phrases such as ", in hazy conditions." to these descriptions, forming the final text prompt. Incorporating scene descriptions provides initial guidance to Stable Diffusion during training, helping it generate an image related to c_{img}. A few examples of the text data are shown in Fig. 4.

While this method produces degraded images effectively, we observed that using only the degradation type in the prompt causes the diffusion model to default to extreme degradation patterns during inference. For instance, the generated haze is excessively thick, or the rain is unrealistically heavy or minimal (see supplementary for examples). Such degradations could negatively impact the performance of a restoration model trained on this data, as the patterns differ significantly from typical real-world scenarios. To overcome this limitation, we introduce an additional conditioning on the level of the degradation, quantified by the mean intensity (μ) and standard deviation (σ) of the degradation map c_{map} defined as $c_{\text{map}} = |x_{\text{in}} - c_{\text{img}}|$. μ represents mean intensity of degradation in the degraded image while σ indicates its spatial distribution across the image. We fuse the conditioning information in the form of μ and σ with the CLIP [36] embedding, e_{text}, of the prompt, c_{text}, as follows. First, we compute the range, $[a, b]$ of μ and σ for each degradation type from all their respective datasets. We divide this range into 128 bins and obtain a one-hot encoding for the bins corresponding to particular μ and σ values calculated from c_{map} during training. An additional bin is included for null-prompt conditioning [5], resulting in vectors of length 129.

We then concatenate the one-hot vectors for μ and σ to

obtain the vector $c_{\text{stats}} \in \mathbb{R}^{2\times 129}$. c_{stats} is then projected to $c^{\text{proj}}_{\text{stats}} \in \mathbb{R}^{2\times 77}$ using a learnable transformation. $c^{\text{proj}}_{\text{stats}}$ is then transposed to $\mathbb{R}^{77\times 2}$ and concatenated with $e_{\text{text}} \in \mathbb{R}^{77\times 768}$ to obtain a vector of size $\mathbb{R}^{77\times 770}$. We project this vector back to the CLIP text embedding dimension and obtain $e_{\text{prompt}} \in \mathbb{R}^{77\times 768}$ to be fed as conditioning to Stable Diffusion. All projection layers are learned during training. This conditioning mechanism ensures that the diffusion model is aware of the degradation level to be added to the clean image, resulting in generated images, x_{gen}, with diverse and realistic degradation patterns. The effect of varying μ and σ is studied in Sec. 4.2. Fig. 4 (a) summarizes the above steps.

Finally, we need to tackle the challenge of aligning the generated degraded images (x_{gen}) precisely with the input clean images (c_{img}). The VAE encoding and decoding process in latent diffusion models causes the loss of fine details in the image [9, 18]. To mitigate this issue, we draw inspiration from AutoDIR [18], which introduced a Structure Correction Module (SCM) to reverse VAE-induced distortions. In our framework, the SCM, denoted by S, corrects x_{gen} as follows:

$$x_S = x_{\text{gen}} + S([x_{\text{gen}}, c_{\text{img}}]) \tag{2}$$

The goal of S is to undo the corruptions causes by the LDM and VAE without affecting the generated degradation. We train S after the degradation generator has been trained (with the generator's parameters kept frozen) using a one-step reverse diffusion process:

$$z_{\text{gen}} = \frac{(z_t - \sqrt{1-\bar{\alpha}_t}\cdot\epsilon)}{\sqrt{\bar{\alpha}_t}} \tag{3}$$

Here $\bar{\alpha}_t$ is the cumulative product of the noise schedule up to timestep t, and z_t is the noisy latent at timestep t. We then obtain $x_{\text{gen}} = \mathcal{D}(z_{\text{gen}})$. The loss function for training S is given by

$$L_S = \sqrt{\bar{\alpha}_{t-1}}\cdot\sqrt{1-\bar{\alpha}_t}\cdot\|x_{\text{in}} - x_S\|_2^2 \tag{4}$$

The term $\sqrt{\bar{\alpha}_{t-1}}\cdot\sqrt{1-\bar{\alpha}_t}$ weights the performance of S for each timestep, recognizing that structure correction is easier near the initial timesteps ($t \approx 0$) and quite challenging near the final timesteps ($t \approx T$). This weighting reduces the influence of these extreme cases during training.

We found that S works well for degradations that possess smooth characteristics such as haze, raindrops and motion blur. However, for degradations such as rain and snow, S tends to blur out fine details in the generated image such as rain streaks and snowflakes. Similarly, for low-light conditions, the SCM can produce blurry outputs due to the low pixel intensities. Thus, for rain, snow and low-light degradations, we omit the usage of S. Instead, we pass the clean image through the VAE encoder and decoder to obtain a slightly altered version $\hat{c}_{\text{img}}$ that is better aligned with the generated image than the original clean image (c_{img}). Visual results showcasing the effect of S can be found in the supplementary.

3.2. Dataset creation

With GenDeg, we can now synthesize diverse degradations on any clean image. We generate these degradations using unique clean images taken from the training datasets of GenDeg. Since we use a large number of training datasets, we obtain approximately 120k distinct scenes. For each clean image, we produce the degradations that were not present in its original training set, resulting in five degradations per image. This strategy supplements existing restoration datasets with our synthetic data, thereby enhancing the potential for generalization capabilities of restoration models when trained on them.

To generate a particular degradation, we randomly select a dataset associated with that degradation type. We sample μ_{gen} from the histogram of μ values in the selected dataset, which is created by the same binning strategy used during training. Subsequently, we sample σ_{gen} from a similar histogram of σ values obtained from images belonging to the sampled μ_{gen} bin. This process ensures that the value of σ_{gen} is meaningfully correlated with the chosen μ_{gen}, resulting in realistic degradation patterns. To further enhance diversity, for every 1 in 20 images, we select a random value of σ_{gen} (within acceptable limits) for a chosen μ_{gen}. The clean image is then degraded using the chosen μ_{gen} and σ_{gen} values. After generation, we filter the images based on the mean value of the generated degradation map to discard poor quality images (see Fig. 4 (b)). The thresholds for filtering are empirically determined for each degradation type (see supplementary for details). In total, after filtering, we create approximately 550k degraded images which are combined with samples from existing datasets to obtain the GenDS dataset.

3.3. Training image restoration models

Transformer based architectures have demonstrated enormous potential in learning generalizable image features. However, the usage of pre-trained transformers for image restoration has been limited. We hypothesize that transformer encoders pre-trained on large datasets such as ImageNet [14] can serve as effective feature encoders for improving generalization in restoration tasks. Hence, we choose a pre-trained Swin transformer encoder [46] as a strong initialization for extracting generalizable features from degraded images. We specifically choose the Swin Transformer [30] over the standard Vision Transformer (ViT) as it provides hierarchical features at multiple resolutions, which is crucial for preserving fine details in restored images. To reconstruct the restored image from the features extracted by the Swin Transformer, we employ a

Table 1. Quantitative comparisons of NAFNet [10], PromptIR [34], and Swin-transformer models using LPIPS and FID metrics (lower is better), trained with and without our GenDS dataset. Performance is evaluated on OoD test sets. The table also includes the performance of existing state-of-the-art (SOTA) approaches. Training with the GenDS dataset significantly enhances OoD performance. (R) indicates real images and (S) indicates synthetic images.

Method	**REVIDE [53]**	**O-Haze [1]**	**RainDS [35]**	**LHP [15]**	**RSVD [8]**	**GoPro [33]**	**LOLv1 [44]**	**SICE [6]**	**RainDS [35]**
Degradation Type	Haze (R)	Haze (R)	Rain (S)	Rain (R)	Snow (S)	Motion Blur (R)	Low-light (R)	Low-light (R)	Raindrop (R)
DiffUIR	0.268/58.5	0.334/147.2	0.088/31.2	0.187/26.5	0.176/26.1	0.144/25.2	0.148/65.1	0.442/102.9	-
Diff-Plugin	0.281/72.9	0.377/164.7	0.194/45.0	0.178/30.2	0.207/22.5	0.217/32.8	0.195/70.5	0.233/69.2	-
InstructIR	0.313/65.4	0.341/154.5	0.117/29.2	0.139/21.2	-	0.146/21.1	0.132/57.3	0.234/65.2	-
AutoDIR	0.247/57.9	0.315/144.1	0.105/30.6	0.181/27.0	-	0.157/22.2	0.116/43.7	0.249/74.0	0.157/52.4
PromptIR	0.262/62.0	0.333/150.9	0.111/49.3	0.186/29.3	0.128/15.8	0.186/32.9	0.258/111.8	0.391/99.3	0.208/106.8
PromptIR GenDS	0.212/56.0	0.160/89.0	0.096/34.4	0.182/28.1	0.119/13.9	0.191/31.9	0.178/87.9	0.375/90.7	0.182/79.8
Swin	0.242/62.9	0.254/109.9	0.182/38.8	0.189/27.1	0.143/21.7	0.198/31.7	0.241/112.0	0.293/88.3	0.232/82.6
Swin GenDS	0.209/54.3	0.165/74.6	0.116/35.8	0.162/24.1	0.121/14.1	0.170/36.2	0.167/73.1	0.241/72.1	0.197/70.7
NAFNet	0.211/71.3	0.183/99.2	0.107/34.4	0.200/29.3	0.131/14.3	0.155/28.2	0.167/78.8	0.304/83.5	0.178/73.4
NAFNet GenDS	0.151/52.5	0.143/76.7	0.100/31.5	0.180/27.1	0.110/11.3	0.149/28.7	0.147/63.7	0.278/78.5	0.170/60.5

lightweight convolutional decoder. This decoder aggregates information from different hierarchical levels of the encoder to produce a high-quality image. The overall architecture is depicted in Fig. 4 (c). The usage of 3×3 convolutions in the decoder helps to overcome a major limitation of patch border artifacts [27] that occur when using transformer models for image restoration. The effect is more exacerbated when using vision transformers due to its large patch size. In addition to training the Swin Transformer-based architecture described above, we also train two other restoration networks: NAFNet [10] and PromptIR [34], on the combined dataset.

4. Experiments

In this section, we provide detailed results and analysis of our method. Implementation details and dataset details can be found in the supplementary.

4.1. Results

To understand the impact of the GenDS dataset, we initially trained NAFNet [10], PromptIR [34] and the Swin model (Sec. 3.3) exclusively on existing restoration datasets, without incorporating any of our synthetic data. We then evaluated their performance on both within-distribution and out-of-distribution (OoD) test sets. Subsequently, we retrained the same models using the entire GenDS dataset and evaluated the performance. Additionally, we compared the performance of these models against state-of-the-art (SOTA) AIOR models, namely, DiffUIR [55], Diff-Plugin [28], InstructIR [13] and AutoDIR [18].

Quantitative comparisons. Due to space constraints, we present quantitative comparisons using only the LPIPS and FID metrics (following [28]). Table 1 presents these scores for OoD test sets across all six degradations. We observe that PromptIR, NAFNet and the Swin model exhibit significant improvements in OoD performance when trained on the GenDS dataset. Motion blur performance remains nearly the same even after training with the GenDS dataset. Since, motion blur already contains sufficient real data with diverse scenes (see Fig. 2), introducing more synthetic data does not necessarily improve performance on real OoD samples. This observation highlights the importance of diverse high-quality data for generalizable image restoration.

Furthermore, the results indicate that the synthetic data generated by GenDeg aids in bridging the domain gap with OoD samples. In certain instances, SOTA methods (upper portion of Table 1) outperform our models on specific OoD datasets (e.g., AutoDIR raindrop removal on RainDS). However, it is important to note that our models serve as simple baselines compared to the more complex SOTA architectures, and that SOTA methods do not consistently perform well for OoD datasets across degradations. Furthermore, SOTA architectures could further enhance their OoD performance by training with our GenDS dataset, as evidenced by the improvements observed in PromptIR, NAFNet, and the Swin model.

Table 2 presents the mean within distribution performance for each degradation. Interestingly, the performance remains almost identical even after training with our GenDS dataset. Moreover, the within-distribution performance shows substantial improvements for haze, low-light and raindrop degradations. This improvement is likely due to the GenDS dataset effectively addressing the limited scene diversity present in existing datasets for these specific degradations.

Detailed quantitative results (including PSNR and

Table 2. Quantitative comparisons of mean LPIPS and FID scores (lower is better) across within distribution datasets for each degradation. Comparisons are shown for PromptIR [34], NAFNet [10] and Swin-transformer models trained with and without our GenDS dataset, along with SOTA models.

Method	Haze	Rain	Snow	Motion Blur	Raindrop	Low-light
DiffUIR	0.329/141.46	0.175/53.53	0.305/23.67	0.182/42.89	-	0.551/260.26
DiffPlugin	0.351/154.12	0.205/47.10	0.227/26.86	0.218/50.57	-	0.464/180.67
InstructIR	0.355/158.81	0.144/35.03	-	0.148/31.86	-	0.402/157.74
AutoDIR	0.306/136.27	0.139/38.12	-	0.161/33.34	0.195/68.09	0.420/155.14
PromptIR	0.309/141.05	0.097/32.61	0.100/18.34	0.163/35.79	0.189/84.48	0.421/189.87
PromptIR GenDS	0.210/112.54	0.080/27.95	0.091/16.19	0.171/34.93	0.188/74.65	0.354/168.59
NAFNet	0.190/118.22	0.074/21.84	0.067/8.20	0.136/28.72	0.085/39.91	0.349/172.36
NAFNet GenDS	0.171/104.43	0.077/22.13	0.069/8.56	0.136/29.31	0.069/29.92	0.316/148.65
Swin	0.244/121.92	0.092/24.50	0.080/10.80	0.194/40.69	0.097/47.09	0.420/187.65
Swin GenDS	0.182/105.242	0.090/24.48	0.083/11.44	0.189/42.17	0.092/39.63	0.368/166.13

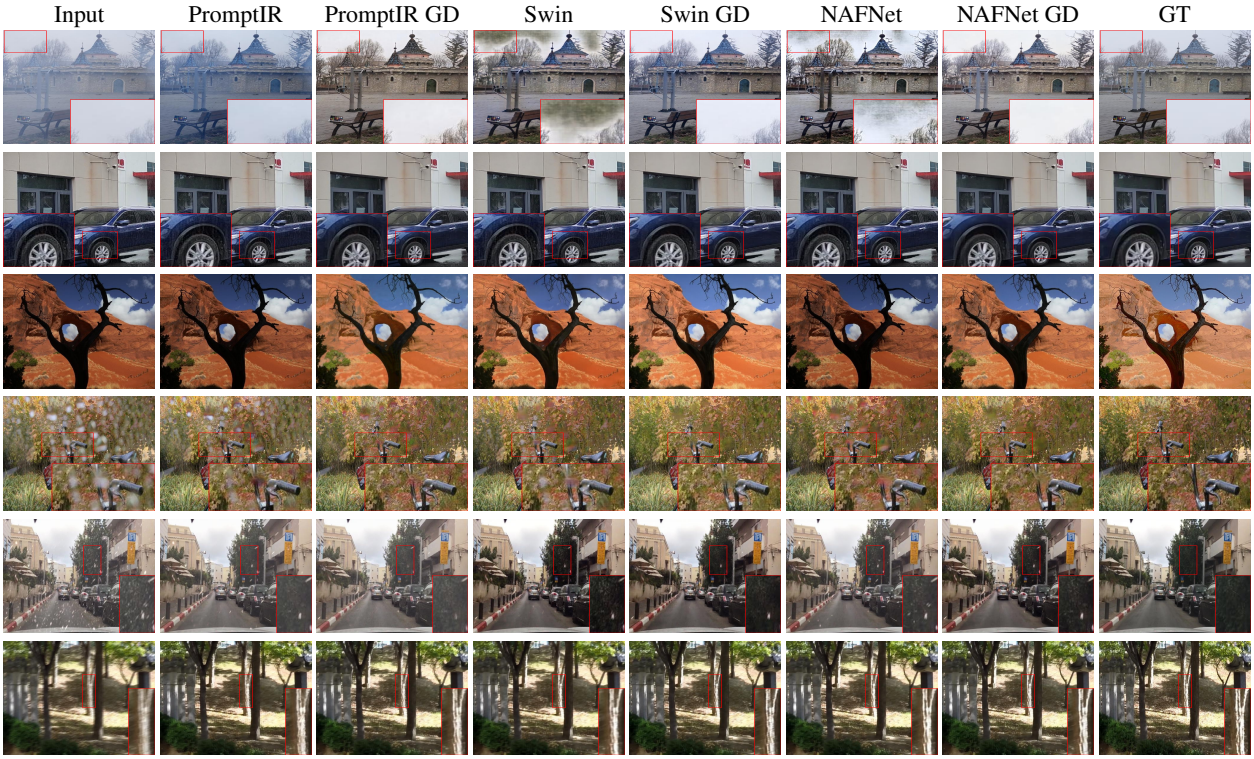

Figure 5. Qualitative comparisons of image restoration models trained with and without our GenDS dataset. The suffix GD represents training with the GenDS dataset. Zoomed-in patches are provided for viewing fine details.

SSIM) are available in the supplementary material, which we highly recommend readers consult.

Qualitative comparisons. We provide qualitative results from one OoD test set for each degradation in Fig. 5. Qualitative comparisons with SOTA methods are in the supplementary. The models trained with our GenDS dataset consistently achieve the best restoration results. Notably, the enhanced images often contain richer colors than the ground truth (see first row), which can cause bad PSNR and SSIM scores. Thus, LPIPS and FID scores are more reliable metrics for testing the OoD performance.

These results demonstrate that the synthetic data generated by GenDeg effectively bridges the domain gap and enhances the generalization capabilities of the restoration models.

4.2. Analysis

In this section, we use our generated synthetic data to conduct various insightful analyses.

Synthetic Data Scaling. We analyze the impact of progressively adding synthetic data generated by our GenDeg framework to existing real data on out-of-distribution

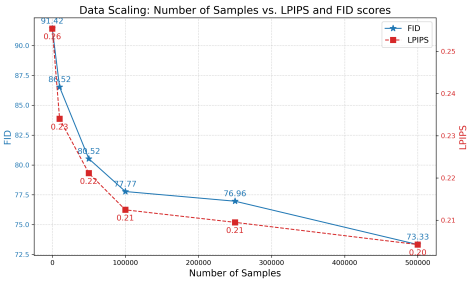

Figure 6. Effect of scaling number of synthetic samples augmented with real data on OoD performance (LPIPS and FID).

(OoD) performance, utilizing the Swin Transformer model. Fig. 6 illustrates the variation in mean OoD LPIPS and FID scores with increasing synthetic data. There is substantial OoD performance improvements with the addition of up to 100k synthetic samples, after which the performance improvement is marginal. Due to limited computational resources, we were unable to scale the synthetic dataset beyond 500 ksamples.

Generated degradation diversity. Our results demonstrate that the degradations generated by GenDeg significantly aid in improving OoD performance of restoration models. This improvement is primarily due to the enhanced scene diversity (as shown in Fig. 2) in our dataset and the variety of degradation patterns produced by GenDeg. To illustrate the variety in degradation patterns, we utilize degradation-aware CLIP (DA-CLIP [31]), a robust CLIP model trained to extract degradation-specific features from images. Fig. 3 presents a t-SNE visualization of the DA-CLIP embeddings obtained from hazy samples in existing training datasets, our GenDS dataset, and the OoD test sets. The visualization reveals a substantial gap in degradation features between the training dataset and OoD testsets. This indicates that the degradations patterns in these datasets are different, thereby, posing generalization challenges. However, our GenDS dataset bridges this gap by introducing numerous samples that resemble those in the OoD test sets, thereby enhancing generalization. Note that GenDeg was *never* trained on the OoD test sets. Furthermore, the t-SNE plot showcases the diversity of degradation patterns produced by our model, as evidenced by our samples spanning a wide area.

Domain gap with real data. We examine the domain gap between existing datasets and synthetic samples generated by GenDeg by training the Swin model exclusively on GenDeg synthesized data. For this experiment, we provide mean of the within distribution and OoD performance across all degradations. Table 3 provides the mean LPIPS and FID scores for both within distribution and OoD performance across all degradations. The results indicate that training solely with GenDeg data significantly reduces within distribution performance compared to using existing datasets for training. This decline is expected due to the domain gap caused by factors such as alignment discrepancies between diffusion-generated samples and corresponding clean samples. However, utilizing both existing and GenDeg synthesized data, i.e., GenDS dataset, enhances performance as the restoration model benefits from exposure to diverse degradation patterns and scenes while maintaining performance on existing data.

Table 3. LPIPS/FID scores for analyzing the performance difference between training on solely existing data, solely GenDeg data, and both real and GenDeg data (GenDS data).

Setting	Existing data	GenDeg data	GenDS data
Within distribution	0.1879/72.10	0.2358/81.08	0.1669/64.84
OoD	0.211/59.36	0.207/61.71	0.1694/47.99

Figure 7. Effect of varying μ and σ in our GenDeg framework for the degradation of haze.

For OoD performance, the model trained on existing datasets alone, experiences a notable decrease from its within distribution performance. In contrast, models trained exclusively on GenDeg data show improved OoD performance demonstrating that the diversity in scenes and degradation patterns enhances generalization. Nevertheless, the best performance is achieved when incorporating both existing and GenDeg data, i.e., GenDS dataset.

Effect of μ and σ. GenDeg allows us to control the intensity and variations in the generated degradation patterns, thereby enhancing degradation diversity. We demonstrate the effect of varying μ and σ on two images for the synthesized degradation of haze. Fig. 7 illustrates the same by first fixing the value of σ and varying μ and then fixing the value of μ and varying σ. When μ is increased, the intensity of haze in the image expectedly increases. σ controls the variation of haze in the image. For $\sigma = 0.15$, the non-homogenous haze (similar to NH-Haze [2]) is spread throughout the image. As σ increases, the spread of haze becomes more localized with higher intensity as seen from the figure.

5. Conclusions

In this paper, we addressed the important problem of generalization in All-In-One Restoration (AIOR) models. Toward this aim, we introduced GenDeg, a novel diffusion model–based framework for synthesizing diverse degradation patterns on clean images, offering fine-grained control over degradation characteristics. Utilizing GenDeg, we generated over 550k degraded samples encompassing a wide range of scenes and degradations. Training AIOR models with both existing and GenDeg data yielded significant improvements in out-of-distribution performance. Our work suggests a promising research direction for addressing generalization challenges in AIOR, aiding in the development of more robust restoration models.

6. Acknowledgments

This work is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DOI/IBC) contract number 140D0423C0076. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI/IBC, or the U.S. Government.

References

[1] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 754–762, 2018. 6

[2] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 444–445, 2020. 8

[3] Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J Fleet. Synthetic data from diffusion models improves imagenet classification. *arXiv preprint arXiv:2304.08466*, 2023. 3

[4] Hritik Bansal and Aditya Grover. Leaving reality to imagination: Robust classification via generated datasets. *arXiv preprint arXiv:2302.02503*, 2023. 3

[5] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023. 4

[6] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 6

[7] A. Cantor. Optics of the atmosphere–scattering by molecules and particles. *IEEE Journal of Quantum Electronics*, 14(9): 698–699, 1978. 2

[8] Haoyu Chen, Jingjing Ren, Jinjin Gu, Hongtao Wu, Xuequan Lu, Haoming Cai, and Lei Zhu. Snow removal in video: A new dataset and a novel method. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13165–13176. IEEE, 2023. 6

[9] Jingye Chen, Yupan Huang, Tengchao Lv, Lei Cui, Qifeng Chen, and Furu Wei. Textdiffuser: Diffusion models as text painters. *Advances in Neural Information Processing Systems*, 36, 2024. 5

[10] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 3, 6, 7

[11] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. Berlin, Heidelberg, 2020. Springer-Verlag. 1

[12] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. 2022. 3

[13] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2025. 1, 3, 6

[14] 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*, pages 248–255. Ieee, 2009. 5

[15] Yun Guo, Xueyao Xiao, Yi Chang, Shumin Deng, and Luxin Yan. From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12097–12107, 2023. 2, 6

[16] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, 2009. 1

[17] Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *arXiv preprint arXiv:2410.15067*, 2024. 2

[18] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. *arXiv preprint arXiv:2310.10123*, 2023. 1, 3, 5, 6

[19] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4): 1742–1755, 2012. 1

[20] Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 4

[21] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 2

[22] Xiangtao Kong, Jinjin Gu, Yihao Liu, Wenlong Zhang, Xiangyu Chen, Yu Qiao, and Chao Dong. A preliminary exploration towards general image restoration. *arXiv preprint arXiv:2408.15143*, 2024. 2

[23] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 1, 2

[24] Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3172–3182, 2020. 3

[25] Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*, 2022. 2

[26] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1833–1844, 2021. 1

[27] Yihao Liu, Jingwen He, Jinjin Gu, Xiangtao Kong, Yu Qiao, and Chao Dong. Degae: A new pretraining paradigm for low-level vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23292–23303, 2023. 6

[28] Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson WH Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4197–4208, 2024. 1, 3, 6

[29] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073, 2018. 1

[30] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 5

[31] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023. 1, 3, 8

[32] Chaofan Ma, Yuhuan Yang, Chen Ju, Fei Zhang, Jinxiang Liu, Yu Wang, Ya Zhang, and Yanfeng Wang. Diffusionseg: Adapting diffusion towards unsupervised object discovery. *arXiv preprint arXiv:2303.09813*, 2023. 3

[33] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 1, 2, 6

[34] 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, 2024. 1, 3, 6, 7

[35] Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing raindrops and rain streaks in one go. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9147–9156, 2021. 6

[36] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 4

[37] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 4

[38] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5572–5581, 2019. 1, 2

[39] Jordan Shipard, Arnold Wiliem, Kien Nguyen Thanh, Wei Xiang, and Clinton Fookes. Diversity is definitely needed: Improving model-agnostic zero-shot classification via stable diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 769–778, 2023. 3

[40] Jordan Shipard, Arnold Wiliem, Kien Nguyen Thanh, Wei Xiang, and Clinton Fookes. Boosting zero-shot classification with synthetic data diversity via stable diffusion. *arXiv preprint arXiv:2302.03298*, 3(5), 2023. 3

[41] Aysim Toker, Marvin Eisenberger, Daniel Cremers, and Laura Leal-Taixé. Satsynth: Augmenting image-mask pairs through diffusion models for aerial semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27695–27705, 2024. 3

[42] J. Jose Valanarasu, R. Yasarla, and V. M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2343–2353, 2022. 3

[43] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12262–12271, 2019. 1, 2

[44] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6

[45] Weijia Wu, Yuzhong Zhao, Mike Zheng Shou, Hong Zhou, and Chunhua Shen. Diffumask: Synthesizing images with pixel-level annotations for semantic segmentation using diffusion models. In *Proceedings of the IEEE/CVF Interna-*

tional Conference on Computer Vision, pages 1206–1217, 2023. 3

[46] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9653–9663, 2022. 5

[47] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10371–10381, 2024. 2

[48] Jianhao Yuan, Francesco Pinto, Adam Davies, Aarushi Gupta, and Philip Torr. Not just pretty pictures: Text-to-image generators enable interpretable interventions for robust representations. *arXiv preprint arXiv:2212.11237*, 3, 2022. 3

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

[50] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 1

[51] He Zhang, Vishwanath Sindagi, and Vishal M. Patel. Joint transmission map estimation and dehazing using deep networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7):1975–1986, 2020. 1

[52] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30:7419–7431, 2021. 1

[53] Xinyi Zhang, Hang Dong, Jinshan Pan, Chao Zhu, Ying Tai, Chengjie Wang, Jilin Li, Feiyue Huang, and Fei Wang. Learning to restore hazy video: A new real-world dataset and a new method. In *CVPR*, pages 9239–9248, 2021. 6

[54] Chenbo Zhao, Yoshiki Ogawa, Shenglong Chen, Zhehui Yang, and Yoshihide Sekimoto. Label freedom: Stable diffusion for remote sensing image semantic segmentation data generation. In *2023 IEEE International Conference on Big Data (BigData)*, pages 1022–1030. IEEE, 2023. 3

[55] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 1, 3, 6