# SIDL: A Real-World Dataset for Restoring Smartphone Images with Dirty Lenses

**Sooyoung Choi*, Sungyong Park*, Heewon Kim**

Soongsil University
{csy010921, ejqdl}@soongsil.ac.kr, hwkim@ssu.ac.kr

## Abstract

Smartphone cameras are ubiquitous in daily life, yet their performance can be severely impacted by dirty lenses, leading to degraded image quality. This issue is often overlooked in image restoration research, which assumes ideal or controlled lens conditions. To address this gap, we introduced SIDL (Smartphone Images with Dirty Lenses), a novel dataset designed to restore images captured through contaminated smartphone lenses. SIDL contains diverse real-world images taken under various lighting conditions and environments. These images feature a wide range of lens contaminants, including water drops, fingerprints, and dust. Each contaminated image is paired with a clean reference image, enabling supervised learning approaches for restoration tasks. To evaluate the challenge posed by SIDL, various state-of-the-art restoration models were trained and compared on this dataset. Their performances achieved some level of restoration but did not adequately address the diverse and realistic nature of the lens contaminants in SIDL. This challenge highlights the need for more robust and adaptable image restoration techniques for restoring images with dirty lenses. Project website: https://sidl-benchmark.github.io

## Introduction

Smartphone cameras have become an essential tool in everyday life. They capture moments and memories with ease and convenience at home, at a party, at a pool, or on the beach. However, the quality of images produced by these cameras can be significantly compromised by a common yet often overlooked issue: dirty lenses. Smartphone lenses frequently contain contaminants such as water drops, fingerprints, and dust, which can cause blurred, hazy, or distorted images. Despite its prevalence, this problem has received limited attention in image restoration, where research typically focuses on hardware limitations or environmental conditions.

The existing works in image restoration have significantly addressed challenges such as denoising, deblurring, and super-resolution. However, these approaches often assume that the images are captured under optimal camera lens conditions, which do not fully reflect the complexities of real-world scenarios. Lens contamination introduces unique artifacts that differ from those typically addressed in traditional image restoration tasks, requiring specialized solutions. The contaminants in lenses scatter light or distort images. Figure 1 visualizes the characteristics of image degradation through the contaminants, which differ in type.

To bridge this gap, we proposed SIDL (Smartphone Images with Dirty Lenses), a dataset for learning and evaluating the restoration of images captured by contaminated smartphone camera lenses. SIDL offers a diverse collection of real-world images taken in various lighting conditions and environments *(e.g.* low-light indoor, normal-light indoor, outdoor day, and outdoor night), with common lens contaminants *(e.g.* water drop, fingerprint, dust, scratch, and the mixture of contaminant types.) for 300 scenes, resulting *1,588* degraded images. The dataset pairs each degraded image with a clean image, making it available for supervised learning and reference-based image quality assessment.

SIDL's challenge is its realistic and varied contamination patterns, which present significant obstacles to existing restoration models. To assess the efficacy of current techniques in this context, we trained and tested several state-of-the-art restoration models , AirNet (Li et al. 2022), NAFNet (Chen et al. 2022), Restormer (Zamir et al. 2022), FFTformer (Kong et al. 2023), DiffUIR (Zheng et al. 2024), AutoDIR (Jiang et al. 2023), and MambaIR (Guo et al. 2024), on the SIDL dataset. As a result, these models achieved some level of restoration. However, they struggled to handle the diverse and complex contaminants in SIDL effectively. These results underscore the need to develop more robust and adaptable restoration techniques to address the challenges posed by lens contamination in the real world.

This paper details the construction of the SIDL dataset, provides a thorough evaluation of existing restoration models on this dataset, and presents potential avenues for future research in this underexplored area of image restoration.

## Related Work

### Image Restoration Datasets

Previous research has introduced numerous datasets to address various types of image degradation and environmental conditions. For traditional image restoration tasks, datasets have been developed for deblurring (Nah, Kim, and Lee 2017; Shen et al. 2019; Rim et al. 2020), super-

*These authors contributed equally.

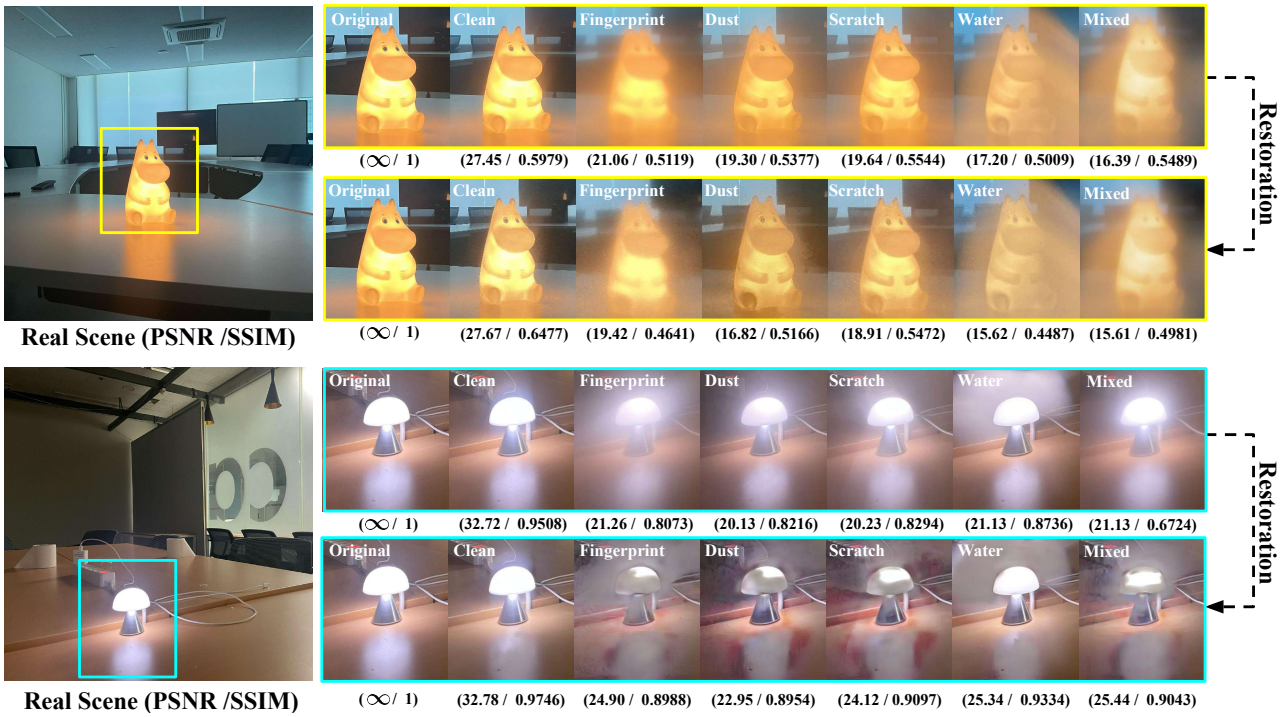



Figure 1: Overview of the proposed **S**martphone **I**mage with **D**irty **L**enses **(SIDL)** dataset. SIDL contains the original images and five realistic physical contaminant types: fingerprint, dust, scratch, water drop, and mixed types. It also includes a control type of clean images without contaminants in our data acquisition setting. The unsatisfactory restoration results from MambaIR (Guo et al. 2024) (top) and DiffUIR (Xia et al. 2023) (bottom) trained on SIDL highlight its challenges. While all contaminants present the overall blurring effect, each type exhibits unique characteristics. *Fingerprint* leads to an overall defocus and blurring effect, *Dust* creates darker regions on the image, *Scratch* causes sharp and multidirectional light scattering, *Water* results in circular light scattering, and *Mixed* type combines these distinct characteristics. These degradation similarities and dissimilarities make the SIDL restoration more challenging.

resolution (Zhou et al. 2020; Zhang et al. 2021; Agustsson and Timofte 2017; Timofte et al. 2017), and denoising (Plotz and Roth 2017; Abdelhamed, Lin, and Brown 2018). These datasets cover a range of degradation types and restoration challenges. However, real-world scenarios often involve more complex environmental factors. Several datasets have been created for specific environmental conditions, such as fog and haze (Ancuti et al. 2018b, 2019, 2018a), rain (Wang et al. 2019; Li et al. 2019), and low-light (Wei et al. 2018; Fu et al. 2023). These datasets facilitate the development of robust algorithms for real-world environmental conditions. Recently, few works have focused on the visual artifacts caused by lens contamination. Wang et al. (Wang et al. 2023) proposed a dataset for a single degradation type (Scratch), and "Let's see clearly" (Li et al. 2021) proposed a synthetic dataset with explicit shapes of stains on the lens. However, the previous dirty lens datasets are limited in scope. Our dataset includes multiple real-world degradation types from dirty lenses. Table 1 detailed the dataset characteristics.

## Image Restoration Methods

Traditional image restoration approaches can be categorized into model-based methods (Buades, Coll, and Morel 2005; Protter et al. 2008; He, Sun, and Tang 2010) and optimization-based methods (Dabov et al. 2007; Gu et al. 2014; Farsiu et al. 2004). Following recent breakthroughs in deep neural networks, various approaches have emerged, addressing a wide range of restoration tasks such as super-resolution (Dong et al. 2014; Lim et al. 2017), denoising (Zhang et al. 2017), deblurring (Nah, Kim, and Lee 2017; Jiang et al. 2020; Zhang et al. 2019; Park et al. 2020), dehazing (Cai et al. 2016; Ren et al. 2016) and derain (Fu et al. 2017; Li et al. 2018; Ren et al. 2019; Yang et al. 2020). However, these methods often rely on specific degradation assumptions and may lack the generalization ability to handle diverse degradations. To address this limitation, recent approaches like NAFNet (Chen et al. 2022) with non-linear activation functions, and transformer-based models (Zamir et al. 2022; Kong et al. 2023) have emerged as versatile solutions. Further advancing this trend, researchers developed multi-degradation or all-in-one models (Chen et al. 2021; Li, Tan, and Cheong 2020) capable of addressing various restoration problems simultaneously. More recently, diffusion-based methods (Zheng et al. 2024; Jiang et al. 2023; Kawar et al. 2022; Xia et al. 2023) and linear state space models like MambaIR (Guo et al. 2024) have emerged, showing promising results in diverse image restoration tasks. Here, we evaluated and analyzed the state-

| Type | Dataset | Resolution | Device | Real / Synthetic | Multiple Distortions | RAW Data Available | Ref images | Distort images |
|---|---|---|---|---|---|---|---|---|
| **Deblurring** | GOPRO | $1280 \times 720$ | Digital camera | Synthetic | ✗ | ✗ | 3,214 | 3,214 |
| | HIDE | $1280 \times 720$ | Digital camera | Synthetic | ✗ | ✗ | 8,422 | 8,422 |
| | RealBlur | $680 \times 773$ | Digital camera | Synthetic | ✗ | ✗ | 232 | 4,738 |
| **Denoising** | SIDD | $5312 \times 2988$ | Smartphone | Real | ✗ | ✓ | 200 | 30,000 |
| | DND | $7360 \times 4912$ | Smartphone | Real | ✗ | ✓ | 50 | 50 |
| **Environmental conditions** | O-HAZE | $5456 \times 3632$ | Digital camera | Real | ✗ | ✓ | 45 | 45 |
| | Dense-Haze | $5456 \times 3632$ | Digital camera | Real | ✗ | ✓ | 33 | 33 |
| | MPID | $1920 \times 990$ | Digital camera | Both | ✓ | ✗ | 4,543 | 4,543 |
| | LOL | $400 \times 600$ | Digital camera | Real | ✗ | ✗ | 500 | 500 |
| **Dirty lens** | Wang et al. | $1920 \times 1080$ | Digital camera | Real | ✗ | ✗ | 1,251 | 1,251 |
| | Let's see clearly | $384 \times 384$ | Digital camera | Synthetic | ✓ | ✗ | 18,000 | 18,000 |
| | **SIDL (Ours)** | $4032 \times 3024$ | Smartphone | Real | ✓ | ✓ | 300 | 1,588 |

Table 1: Comparison with image restoration datasets. Prior dirty lens works captured real single degradation (Wang et al. 2023) or generated synthetic data (Let's see clearly), while our SIDL dataset presents real multiple degradations, including RAW data.

of-the-art models on the SIDL dataset to highlight the challenges in restoring images with dirty lenses.

# The SIDL Dataset

## Image Acquisition Setup

To acquire the SIDL dataset, we used an iPhone 12 Pro with a fixed 26mm focal length camera. Figure 2 visualizes a custom smartphone camera setup to capture images with various lens degradations for a scene. We used a tripod and a Bluetooth remote controller to minimize camera shakes during image capture. Our setup's key component is a custom 'Film holder,' created using `Maya` 3D modeling software, as shown in Figure 2(a). The film holder holds the 'Dirty film' close to the smartphone camera lens. It includes a small external frame to insert and secure the 'Dirty film.' This frame includes a thin groove that helps maintain a close distance between the film and the lens while preventing movement during image capture.

## Dirty Film Setup

We used various materials to simulate common contaminants in smartphone camera lenses, creating diverse and realistic dirty lens effects. We applied these materials to thin PVC films with a refractive index and reflectance similar to glass.

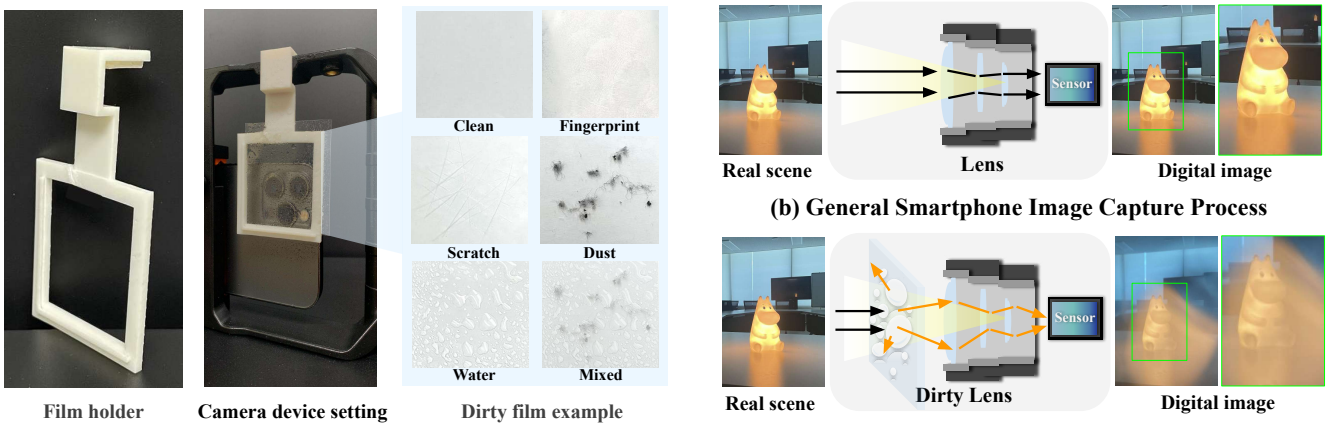
Each type of films replicates specific environmental conditions or common causes of lens degradation. The following descriptions detail the process of creating each type of dirty films:

- **Original**: We captured images using the clean camera lens without any added filters or simulated effects.
- **Clean**: We captured images with clean PVC films without any applied materials or simulated degradation.
- **Dust**: We simulated dust on a camera lens by sprinkling tiny threads and sand onto a film. We secured the particles between two films to replicate various forms of dust *(e.g.* fibers, dirt, sand) and their static cling.
- **Fingerprint**: Fingerprints are a common type of dirt found on smartphone camera lenses. We applied hand cream to a finger and left fingerprints on a film to recreate this effect.
- **Water**: Smartphone cameras often encounter water droplets in humid environments *(e.g.*, rain, swimming pools, seaside). We used a spray bottle to simulate this to create water droplets on a film.
- **Scratch**: Scratches on camera lenses are typical, often caused by dropping the device. We used a knife to create scratch marks on a film to simulate this.
- **Mixed**: Multiple dirty films are combined to simulate various contaminants on a camera lens.

Reusing dirty films to capture multiple scenes may lead to learning-based methods to memorize specific dirty patterns. To prevent this, we created new dirty films for each scene.

## Image Acquisition Process

The SIDL dataset was captured in a controlled environment that minimized the influence of external factors. We carefully designed this setup to ensure consistency and reproducibility across all images. To clearly show the effect of lens degradations, the scenes (or objects) chosen for imaging were static and unchanged throughout the process. This ensured that any changes in image quality were caused only by the lens degradations. Under these controlled conditions, we selected diverse scenes representing various photographic challenges (*e.g.,* indoor and outdoor environments, day and night, various lighting conditions, and subjects with different textures and colors).




(a) Film holder & Camera setting

(b) General Smartphone Image Capture Process

(c) Smartphone Image with Dirty Lens Capture Process

Figure 2: Custom smartphone camera setup and image capture processes for dirty lenses. (a) The 3D-printed 'Film holder' is attached to a smartphone for capturing images with PVC films called 'dirty films,' which contain physical damage to generate realistic distortions. (b) The typical direction of light toward the smartphone camera for a scene generates a clean image. (c) The dirty film in front of the smartphone camera distorts the light and results in a degraded image.

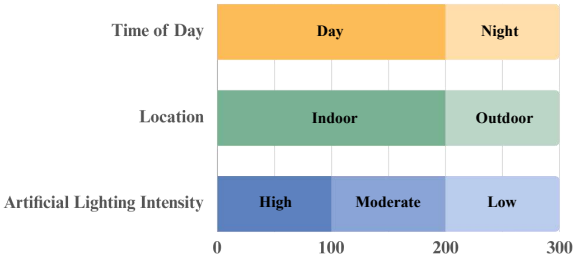



Figure 3: The SIDL statistics. The SIDL dataset consists of 300 scenes, covering various times of day, locations, and artificial lighting intensities. Each category is carefully structured to provide a diverse and balanced composition, ensuring thorough coverage of different environmental and lighting conditions. The data separation for the training, validation, and test sets is 240, 20, and 40 scenes.

The number of scenes for each category in the SIDL dataset is shown in Figure 3. We captured clean reference images for each scene without any lens degradation, which we denote as 'Original'. After capturing the Original images, we applied different lens degradations by changing dirty films in our custom film holder for each scene. We created new dirty films for each scene to prevent learning-based methods from memorizing specific dirty patterns. All images were captured in ProRAW format to preserve the original sensor data with the iPhone's automatic camera adjustments turned off.

## Preprocessing

Following previous image restoration works (Kong et al. 2023; Zheng et al. 2024; Guo et al. 2024) that operate in the sRGB domain, we utilized an open-source ISP (`Rawpy`) to convert the captured ProRAW images to the sRGB space. The ISP includes black-and-white level adjustment, white balancing, gamma correction, color space conversion, and bit-depth quantization, where ProRAW images have already been demosaiced. The image resolution of our dataset is $4032 \times 3024$ pixels, which require long data loading time. For efficient model training, the full-resolution images were divided into $512 \times 512$ patches, a size commonly used in image restoration model training.

## Dataset Characteristics

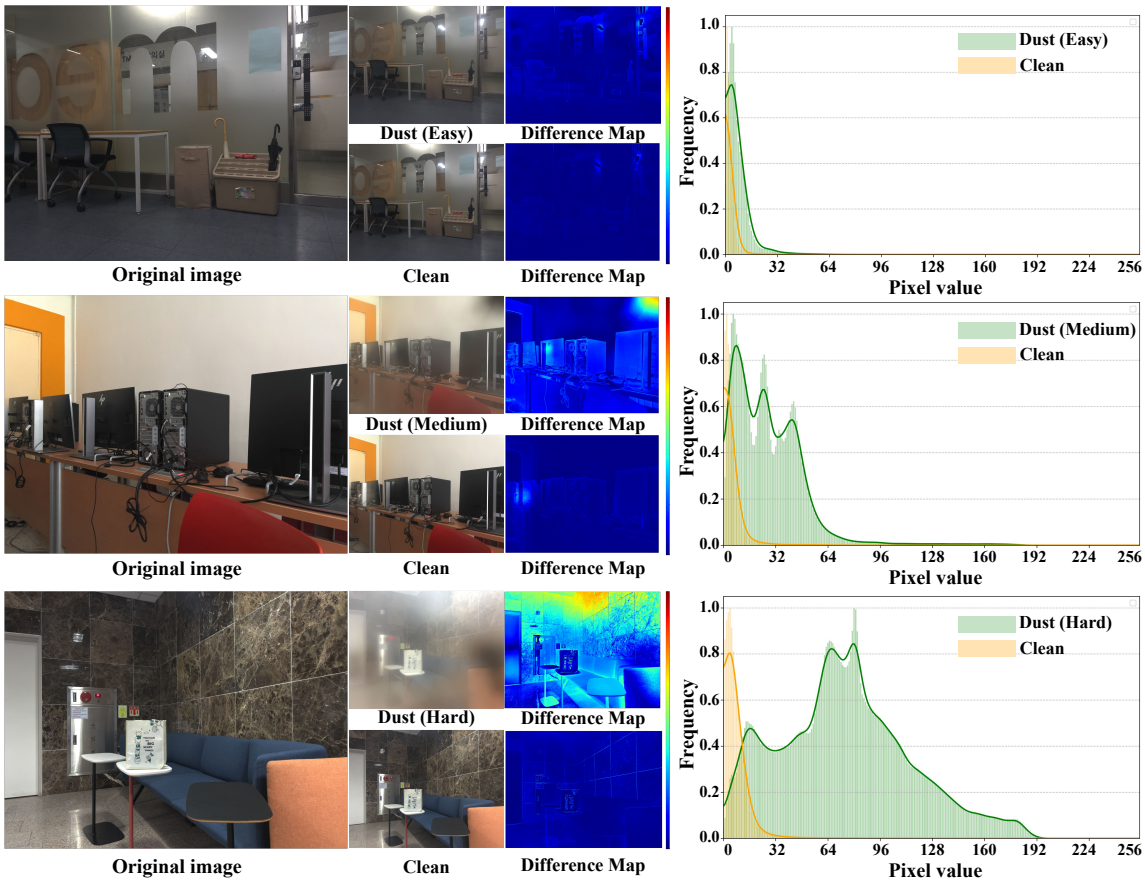
The SIDL dataset, captured using various dirty films, demonstrates a wide range of image degradations that are similar to real-world scenarios. Each type of degradation introduces unique characteristics to the captured images, affecting their visual quality.

**Original** Captured directly using only a camera lens without any film, the 'Original' images present clean contents and sharp patterns without damage or degradation.

**Clean** 'Clean' images, captured using an uncontaminated film in the holder, are a reference for comparing the effects of different lens degradations. In low light conditions, 'Clean' and 'Original' (no film) images may appear similar. However, the film often increases light refraction under intense lighting and causes visible artifacts.

**Dust** Dust particles on the film produced various visual artifacts in the captured images. These particles either obscured parts of the subject or caused color distortions, depending on their size and location. Additionally, they scattered and reflected light, degrading image quality and reducing sharpness.

**Fingerprint** Fingerprints on the film surface caused light scattering, reducing image sharpness. The intensity of the fingerprint affected the degree of blurriness due to the




**(a) Visualization of Difference Maps for Various Degradation Levels** **(b) Frequency Histogram of Pixel Differences**

Figure 4: Visual comparison of the 'Original,' 'Dust,' and 'Clean' images at different degradation levels (easy, medium, hard). (a) The difference maps illustrate the spatial distribution and severity of the degradation effects by highlighting the pixel-wise discrepancies between the original and degraded images. (b) Frequency histograms of pixel differences between the original and degraded images for each degradation level. The histograms provide insights into the distribution and magnitude of the degradation impact. As the degradation level increases from easy to hard, the histograms exhibit a wider spread and higher frequencies of larger pixel differences, indicating more severe and diverse degradation effects.

amount of oil in the print. More distinct fingerprints resulted in more blurred images, with the blurring effect typically spreading evenly across the entire image.

**Water** Water causes different types and degrees of damage to the image depending on the amount and location of water droplets on the film surface. Light scattering doesn't simply result in blurring but appears as if colors spread to the surrounding areas, and circular shapes may also appear. Additionally, as light passes through water, severe refraction and reflection occur, causing more significant degradation than other degradation types.

**Scratch** Scratch exhibits different degrees of light scattering or blurring depending on its direction and severity. Unlike Fingerprints, which appear generally uniform across the entire image, Scratch is characterized by localized or severe effects in the degraded areas.

**Mixed** As a special case, mixed degradation was collected for several scenes to explore more complex lens damage patterns. These images were obtained by overlapping two different types of dirty films, resulting in more severe degradation than using a single film.

## Benchmark

The SIDL dataset is structured to provide a comprehensive benchmark for image restoration tasks. For each degradation type, the data is divided into 240 scenes for training, 20 for validation, and 40 for testing. The mixed degradation subset includes 51, 15, and 20 scenes for training, validation, and testing, respectively. Our data split is performed at the image level to prevent overlap between training, validation, and test data. We used PSNR (Peak Signal-to-Noise Ratio) to measure the difference between degraded (or restored) images and their corresponding reference ('Original'). Based on these PSNR values of degraded images, our dataset is divided into three difficulty levels: **Easy**, **Medium**, and **Hard**.

| Method | Network | | Clean PSNR/SSIM | Dust PSNR/SSIM | Fingerprint PSNR/SSIM | Water PSNR/SSIM | Scratch PSNR/SSIM | Mixed PSNR/SSIM | Average PSNR/SSIM |
|---|---|---|---|---|---|---|---|---|---|
| CNN | AirNet† | Easy | 27.10 / 0.9329 | 23.40 / 0.8669 | 25.10 / 0.8708 | 24.85 / 0.8658 | 25.84 / 0.8917 | 25.30 / 0.8422 | 25.26 / 0.8784 |
| | | Medium | 26.13 / 0.8899 | 19.69 / 0.7577 | 22.53 / 0.8191 | 20.36 / 0.7656 | 21.24 / 0.8051 | 16.79 / 0.7542 | 21.12 / 0.7986 |
| | | Hard | 23.79 / 0.8527 | 16.18 / 0.7350 | 15.83 / 0.6432 | 16.64 / 0.7449 | 18.13 / 0.7975 | 14.48 / 0.7047 | 17.50 / 0.7463 |
| | NAFNet | Easy | 36.37 / 0.9682 | 23.19 / 0.8197 | 28.12 / 0.8987 | 25.21 / 0.8609 | 27.55 / 0.8829 | 22.90 / 0.7569 | 27.22 / 0.8646 |
| | | Medium | 32.25 / 0.9312 | 21.60 / 0.7367 | 25.45 / 0.8470 | 22.53 / 0.7834 | 25.16 / 0.8282 | 20.16 / 0.7579 | 24.53 / 0.8141 |
| | | Hard | 28.56 / 0.9093 | 19.34 / 0.6985 | 17.59 / 0.6614 | 19.61 / 0.7256 | 20.24 / 0.7734 | 17.75 / 0.7411 | 20.51 / 0.7516 |
| Transformer | Restormer | Easy | 38.12 / 0.9786 | 25.65 / 0.9083 | 28.77 / 0.9186 | 26.25 / 0.8984 | 27.00 / 0.9230 | 22.62 / 0.8435 | 28.06 / 0.9117 |
| | | Medium | 34.07 / 0.9368 | 23.39 / 0.8235 | 26.08 / 0.8701 | 24.11 / 0.8400 | 26.32 / 0.8912 | 22.41 / 0.8345 | 26.06 / 0.8660 |
| | | Hard | 30.49 / 0.9338 | 21.41 / 0.8349 | 19.38 / 0.7708 | 20.85 / 0.8417 | 21.57 / 0.8528 | 18.86 / 0.8054 | 22.09 / 0.8399 |
| | FFTformer | Easy | 34.58 / 0.9621 | 22.79 / 0.8742 | 28.34 / 0.9053 | 25.09 / 0.8925 | 22.59 / 0.8665 | 21.80 / 0.6909 | 25.86 / 0.8653 |
| | | Medium | 31.45 / 0.9137 | 21.23 / 0.7786 | 25.22 / 0.8536 | 22.81 / 0.8162 | 21.38 / 0.8315 | 20.64 / 0.7974 | 23.78 / 0.8318 |
| | | Hard | 28.29 / 0.9024 | 19.22 / 0.7935 | 16.55 / 0.6808 | 18.21 / 0.7727 | 18.45 / 0.7891 | 18.06 / 0.7886 | 19.80 / 0.7879 |
| Diffusion | DiffUIR† | Easy | 34.51 / 0.9785 | 26.96 / 0.9276 | 28.82 / 0.9331 | 27.42 / 0.9220 | 30.14 / 0.9375 | 28.22 / 0.9021 | 29.34 / 0.9335 |
| | | Medium | 33.25 / 0.9375 | 22.11 / 0.8325 | 26.16 / 0.8856 | 24.65 / 0.8647 | 27.07 / 0.9105 | 20.32 / 0.8368 | 25.36 / 0.8786 |
| | | Hard | 29.47 / 0.9366 | 20.38 / 0.8602 | 18.93 / 0.7763 | 19.91 / 0.8500 | 21.27 / 0.8599 | 18.97 / 0.8409 | 21.71 / 0.8533 |
| Mamba | MambaIR | Easy | 36.96 / 0.9807 | 24.92 / 0.9125 | 29.26 / 0.9235 | 26.93 / 0.9488 | 29.34 / 0.9322 | 25.30 / 0.8819 | 28.78 / 0.9299 |
| | | Medium | 34.62 / 0.9498 | 23.48 / 0.8298 | 26.62 / 0.8758 | 24.23 / 0.8489 | 27.29 / 0.8955 | 22.18 / 0.8365 | 26.40 / 0.8728 |
| | | Hard | 31.37 / 0.9481 | 21.87 / 0.8388 | 19.27 / 0.7750 | 21.02 / 0.8480 | 21.98 / 0.8615 | 19.53 / 0.8249 | 22.51 / 0.8494 |

Table 2: Quantitative comparison of restoration models on SIDL dataset. For each degradation type and level, the best score is highlighted in red, and the second-best score is indicated with a blue. † denotes **All-in-One** methods.

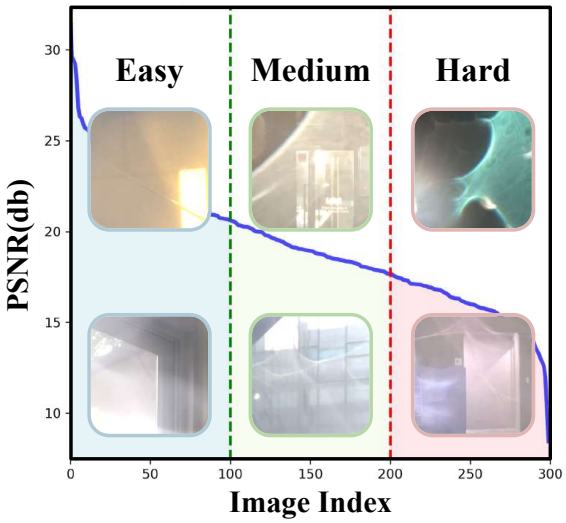



Figure 5: Examples of degradation level classification. The SIDL dataset is divided into 'Easy,' 'Medium,' and 'Hard' categories based on the severity of degradations through PSNR between the reference and degraded images. This example presented the case of water droplets.

The training, validation, and test sets incorporate a balanced representation of various difficulty levels. This categorization enables a more detailed evaluation of restoration algorithms across different degradation levels. Figure 5 illustrates this categorization for the 'Water' degradation type.

# Experiments

## Experimental Setup

**Training Details** To evaluate the SIDL dataset, we selected various state-of-the-art architectures, including CNN-based (AirNet (Li et al. 2022)), NAFNet (Chen et al. 2022)), transformer-based (Restormer (Zamir et al. 2022), FFTformer (Kong et al. 2023)), diffusion-based (DiffUIR (Zheng et al. 2024)), and Mamba-based (MambaIR (Guo et al. 2024)). We employed different training strategies depending on the network architecture. AirNet and DiffUIR are all-in-one methods that were trained simultaneously with all types of degradations as input. In contrast, the other networks were trained in a task-specific manner, focusing on individual degradation types. Each network was trained using an NVIDIA RTX 4090 GPU.

## Restoration Results of the SotA Methods

The quantitative and qualitative results on SIDL datasets are presented in Table 2 and Figure 6. Our experimental results reveal that DiffUIR consistently outperforms other methods across all degradation types and levels. The probabilistic approach seems particularly effective in handling the diverse and challenging scenarios presented in the SIDL dataset. However, the qualitative results are unsatisfactory and present a severe degradation, requiring breakthroughs in the follow-up studies.

## Ablation Studies

**Real Dirty Lens Image Restoration** We conducted experiments with actual smartphone lens degradations to verify that our SIDL dataset accurately reflects real-world dirty

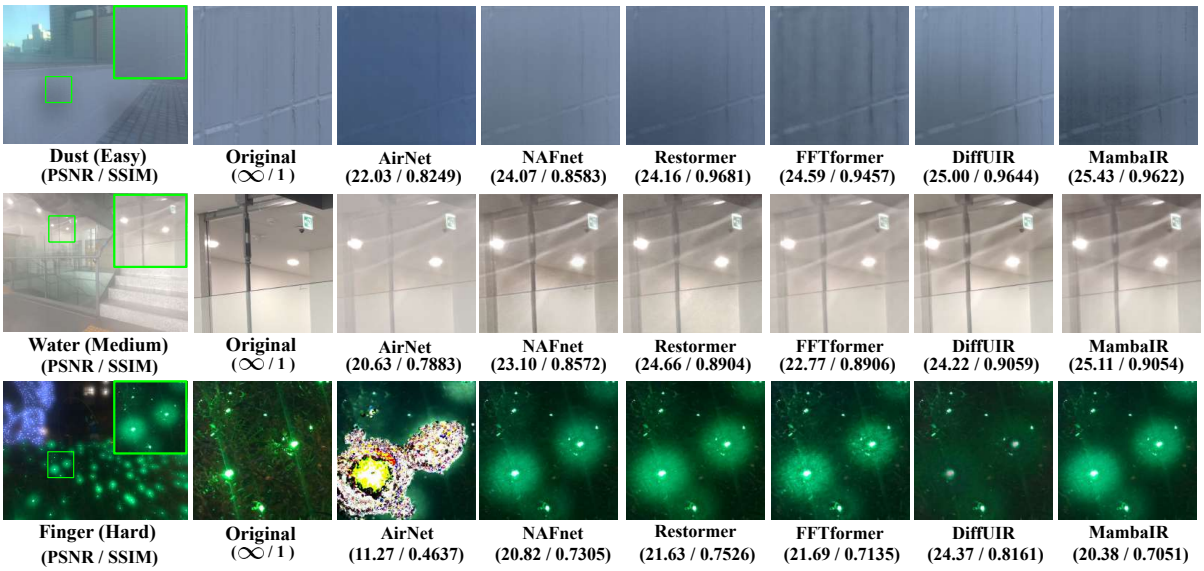



Figure 6: Qualitative comparisons of restoration results on SIDL datasets.

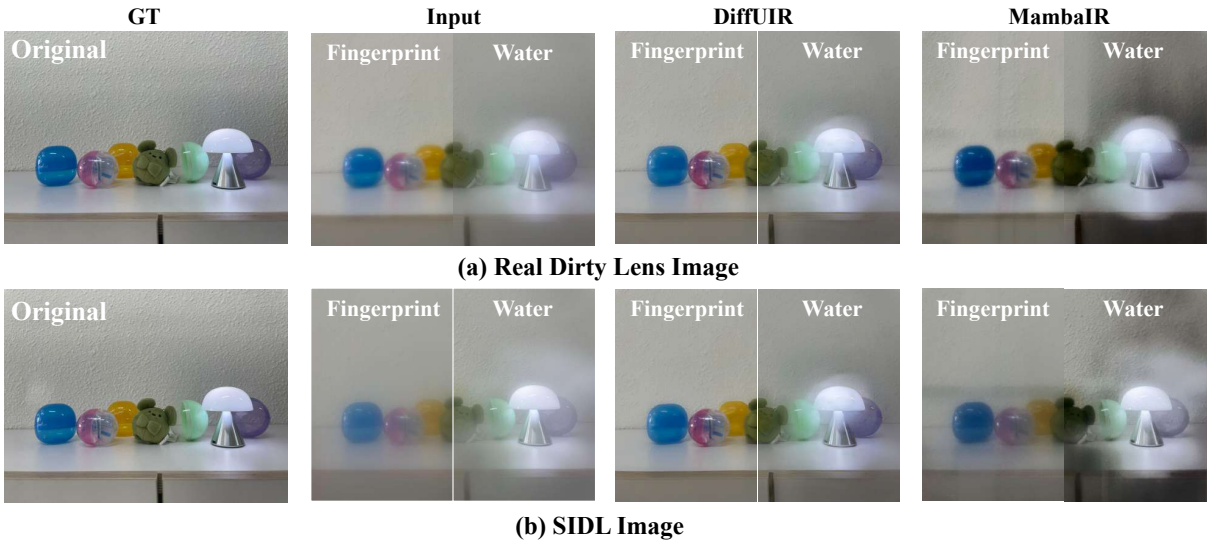



Figure 7: Real dirty lens images *v.s.* SIDL images. (a) Images with camera lenses damaged by fingerprints and water droplets, along with their restoration results, and (b) SIDL images and their restoration results. All real images and their restoration results exhibit similar characteristics to those of SIDL images.

lens characteristics. We focused on fingerprints and water droplets, as these are common and easily reproducible types of dirty lenses in the real world. We applied fingerprints and water droplets directly to smartphone camera lenses, replicating the method used to create our 'Dirty Film' in SIDL. As shown in Figure 7, the 'Real Dirty Lens' degradation patterns closely resemble those in SIDL, exhibiting similar characteristics for each degradation type. Furthermore, we applied image restoration models to the real dirty lens images and our SIDL images, resulting in similar restoration performances. This qualitative comparison demonstrates that SIDL successfully replicates the visual characteristics of real dirty lens images captured from smartphone cameras.

**Evaluation of Pretrained Networks on SIDL** To demonstrate the uniqueness of SIDL degradation patterns, Table 3 compares the quantitative results of pretrained models without fine-tuning and the same models trained on the SIDL training set for the cases of Water and Dust. The results show that pretrained models consistently underperform compared to trained models on the SIDL dataset. This performance gap underscores the unique and challenging nature of SIDL, where the pretrained DiffUIR is an all-in-one

| Method | Type | Pretrained | Trained | Difference |
|---|---|---|---|---|
| DiffUIR | Dust | 20.38 / 0.8130 | 23.15 / 0.8734 | +2.77 / 0.0604 |
| | Water | 21.88 / 0.8408 | 23.99 / 0.8789 | +2.11 / 0.0381 |
| FFTformer | Dust | 19.74 / 0.7926 | 21.08 / 0.8154 | +1.34 / 0.0228 |
| | Water | 21.15 / 0.8075 | 22.04 / 0.8271 | +0.89 / 0.0196 |
| MambaIR | Dust | 19.33 / 0.7461 | 23.42 / 0.8604 | +4.09 / 0.1143 |
| | Water | 21.21 / 0.7572 | 24.06 / 0.8819 | +2.85 / 0.1247 |

Table 3: Performance comparison of Pretrained *vs.* Trained models on the SIDL test set.

| Method / Train Set | Easy | Medium | Hard |
|---|---|---|---|
| NAFNet / **Wang et al.** | 25.55 / 0.8239 | 21.61 / 0.7235 | 18.16 / 0.7598 |
| NAFNet / **SIDL** (scratch) | **27.55 / 0.8829** | **25.16 / 0.8282** | **20.24 / 0.7734** |

Table 4: Comparison of NAFNet models trained on different dirty lens datasets and evaluated on the SIDL test set.

| | Scenes for Training | | | | |
|---|---|---|---|---|---|
| PSNR/SSIM | 10 | 50 | 100 | 200 | 240 |
| Restormer | 20.18 / 0.75 | 21.66 / 0.79 | 22.11 / 0.79 | 23.55 / 0.85 | **23.73 / 0.86** |

Table 5: Analysis with numbers of training scenes.

restoration model. These degradation patterns differ from those in previous research on image restoration, necessitating task-specific adaptation through fine-tuning to improve performance on our benchmark.

**Comparison with Previous Dirty Lens Datasets** Previous dirty lens datasets in Table 1 are limited in scope. Wang et al. (Wang et al. 2023) proposed a dataset only containing scratch degradation, while the "Let's see clearly" (Li et al. 2021) dataset used synthetic data that appears less realistic. In contrast, SIDL has multiple real-world degradation types from the dirty lens. Table 4 compares the performance between models trained on each dataset and evaluated on the SIDL test set. Note that the "Let's see clearly" dataset (Li et al. 2021) is not publicly available. The model trained on Wang et al.'s dataset shows significantly lower performance on the SIDL scratch test set, demonstrating that SIDL provides more diverse lens contamination patterns.

**Analysis of Dataset Scale** Table 5 shows the model performance when trained with different subsets of our dataset (10, 50, 100, 200, 240 scenes) with 15 test scenes. As the number of training scenes increases, the restoration performance improves significantly up to 100 scenes, after which the improvement plateaus. This suggests that the current dataset size is sufficient for evaluating restoration models.

**Comparison with Unsupervised Approach** We compare DiffUIR trained on SIDL with the pretrained AutoDIR (Jiang et al. 2023), an unsupervised text-guided diffusion model for all-in-one image restoration. We used $512 \times 512$ patches to evaluate both models on the SIDL test set (water) images. During inference, AutoDIR first analyzes the input image to identify degradations from its predefined artifact list (*e.g.,* blur, haze, underexposure, etc), then automatically generates a corresponding prompt in the form of "A photo needs [artifact] reduction". As shown in Table 6, both models achieve similar restoration performance across different difficulty levels. However, DiffUIR processes each image in 4.15 seconds with 3 timesteps for testing while AutoDIR requires approximately 90 seconds ($\sim 22\times$ slower).

| PSNR/SSIM | Easy | Medium | Hard |
|---|---|---|---|
| DiffUIR | **27.42** / 0.9220 | 24.65 / 0.8647 | **19.91 / 0.8500** |
| AutoDIR | 26.83 / **0.9231** | **24.67 / 0.8650** | 19.19 / 0.8247 |

Table 6: Comparison with unsupervised learning methods.

## Limitation

Our SIDL dataset, while comprehensive, has some limitations. We focused on common types of lens degradation, but real-world scenarios may present an even wider range of imperfections. The complexity of lens contamination in actual use might exceed our simulations. Future work could expand this dataset to include more varied degradation types, further improving its representation of real-world conditions.

## Conclusion

This paper addresses the need for a high-quality image dataset for image restoration research in smartphone cameras. We propose a public dataset called 'Smartphone Image with Dirty Lens (SIDL),' consisting of 1,588 degraded images and their corresponding original images. Instead of image synthesis, we capture realistic degradations by directly applying physical damage to PVC films. The SIDL dataset includes five types of degradation in smartphone cameras and mixed degradations in which these types appear simultaneously. Experiments demonstrate that images in SIDL closely resemble those where damage is directly applied to the camera lens without films. Benchmarking results of existing methods show that SIDL presents a highly challenging task. This dataset will facilitate diverse follow-up studies, significantly contributing to restoring image degradations commonly encountered in everyday life.


## Acknowledgements

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW) (2024 -0-00071), the Graduate School of Metaverse Convergence support program(IITP-2025-RS-2024-00430997), and the Convergence security core talent training business support program(IITP-2025-RS-2024-00426853) supervised by the IITP(Institute of Information & communications Technology Planning & evaluation). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT)(RS-2024-00456589). We thank Corca AI for sponsoring the compute used in this work.

## References


Abdelhamed, A.; Lin, S.; and Brown, M. S. 2018. A high-quality denoising dataset for smartphone cameras. In *CVPR*, 1692–1700.

Agustsson, E.; and Timofte, R. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *CVPRW*.

Ancuti, C.; Ancuti, C. O.; Timofte, R.; and De Vleeschouwer, C. 2018a. I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images. In *ACIVS*, 620–631. Springer.

Ancuti, C. O.; Ancuti, C.; Sbert, M.; and Timofte, R. 2019. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *ICIP*, 1014–1018.

Ancuti, C. O.; Ancuti, C.; Timofte, R.; and De Vleeschouwer, C. 2018b. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPRW*, 754–762.

Buades, A.; Coll, B.; and Morel, J.-M. 2005. A non-local algorithm for image denoising. In *CVPR*, 60–65.

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

Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; and Gao, W. 2021. Pre-trained image processing transformer. In *CVPR*, 12299–12310.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022. Simple baselines for image restoration. In *ECCV*, 17–33.

Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *TIP*, 16(8): 2080–2095.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Learning a deep convolutional network for image super-resolution. In *ECCV*, 184–199.

Farsiu, S.; Robinson, M. D.; Elad, M.; and Milanfar, P. 2004. Fast and robust multiframe super resolution. *TIP*, 13(10): 1327–1344.

Fu, H.; Zheng, W.; Wang, X.; Wang, J.; Zhang, H.; and Ma, H. 2023. Dancing in the dark: A benchmark towards general low-light video enhancement. In *ICCV*, 12877–12886.

Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017. Removing rain from single images via a deep detail network. In *CVPR*, 3855–3863.

Gu, S.; Zhang, L.; Zuo, W.; and Feng, X. 2014. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, 2862–2869.

Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2024. MambaIR: A Simple Baseline for Image Restoration with State-Space Model. In *ECCV*.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *TPAMI*, 33(12): 2341–2353.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 8346–8355.

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

Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising diffusion restoration models. *NeurIPS*, 35: 23593–23606.

Kong, L.; Dong, J.; Ge, J.; Li, M.; and Pan, J. 2023. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, 5886–5895.

Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; and Peng, X. 2022. All-in-one image restoration for unknown corruption. In *CVPR*, 17452–17462.

Li, R.; Tan, R. T.; and Cheong, L.-F. 2020. All in one bad weather removal using architectural search. In *CVPR*, 3175–3185.

Li, S.; Araujo, I. B.; Ren, W.; Wang, Z.; Tokuda, E. K.; Junior, R. H.; Cesar-Junior, R.; Zhang, J.; Guo, X.; and Cao, X. 2019. Single image deraining: A comprehensive benchmark analysis. In *CVPR*, 3838–3847.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, 254–269.

Li, X.; Zhang, B.; Liao, J.; and Sander, P. V. 2021. Let's see clearly: Contaminant artifact removal for moving cameras. In *ICCV*, 2011–2020.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Mu Lee, K. 2017. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 136–144.

Nah, S.; Kim, T. H.; and Lee, K. M. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 3883–3891.

Park, D.; Kang, D. U.; Kim, J.; and Chun, S. Y. 2020. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *ECCV*, 327–343.

Plotz, T.; and Roth, S. 2017. Benchmarking denoising algorithms with real photographs. In *CVPR*, 1586–1595.

Protter, M.; Elad, M.; Takeda, H.; and Milanfar, P. 2008. Generalizing the nonlocal-means to super-resolution reconstruction. *TIP*, 18(1): 36–51.

Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; and Meng, D. 2019. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, 3937–3946.

Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; and Yang, M.-H. 2016. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, 154–169.

Rim, J.; Lee, H.; Won, J.; and Cho, S. 2020. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, 184–201.

Shen, Z.; Wang, W.; Lu, X.; Shen, J.; Ling, H.; Xu, T.; and Shao, L. 2019. Human-aware motion deblurring. In *ICCV*, 5572–5581.

Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; and Zhang, L. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In *CVPRW*.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *CVPR*, 12270–12279.

Wang, Y.; Wan, R.; Yang, W.; Wen, B.; Chau, L.-p.; and Kot, A. C. 2023. Removing Image Artifacts From Scratched Lens Protectors. *arXiv*.

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

Xia, B.; Zhang, Y.; Wang, S.; Wang, Y.; Wu, X.; Tian, Y.; Yang, W.; and Van Gool, L. 2023. Diffir: Efficient diffusion model for image restoration. In *ICCV*, 13095–13105.

Yang, W.; Tan, R. T.; Wang, S.; Fang, Y.; and Liu, J. 2020. Single image deraining: From model-based to data-driven and beyond. *TPAMI*, 43(11): 4059–4077.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 5728–5739.

Zhang, H.; Dai, Y.; Li, H.; and Koniusz, P. 2019. Deep stacked hierarchical multi-patch network for image deblurring. In *CVPR*, 5978–5986.

Zhang, K.; Li, D.; Luo, W.; Ren, W.; Stenger, B.; Liu, W.; Li, H.; and Yang, M.-H. 2021. Benchmarking ultra-high-definition image super-resolution. In *ICCV*, 14769–14778.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *TIP*, 26(7): 3142–3155.

Zheng, D.; Wu, X.-M.; Yang, S.; Zhang, J.; Hu, J.-F.; and Zheng, W.-S. 2024. Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model. In *CVPR*, 25445–25455.

Zhou, R.; El Helou, M.; Sage, D.; Laroche, T.; Seitz, A.; and Süsstrunk, S. 2020. W2S: microscopy data with joint denoising and super-resolution for widefield to SIM mapping. In *ECCV*, 474–491.