

# UDC-VIT: A Real-World Video Dataset for Under-Display Cameras

## Supplementary Material

### A. Analyzing the Limitations of Synthetic Datasets

This section provides a comprehensive analysis of the limitations of synthetic datasets (*e.g.*, VidUDC33K [8]). As outlined in Sec. 4 in the main body, the VidUDC33K dataset exhibits several strange scenes. In this section, we analyze three representative anomalies frequently observed in the dataset: *flares occurring in physically improbable scenarios*, *unintended white artifacts*, and *darkened and nearly featureless degraded frames*.

#### Improbable situations and unintended white artifacts.

Liu *et al.* [8] attempt to synthesize flares by convolving the PSF with ground-truth images. However, the desired flares do not manifest as expected. To address this, they apply a scaling procedure to pixel values exceeding a predefined threshold to amplify these intensities, followed by PSF convolution.

This process leads to two notable phenomena. First, *flares appear in physically improbable scenarios*. Their method generates flares with values of white pixels exceeding the threshold regardless of whether these correspond to actual light sources, as shown in Fig. 8(a)-(c) in the main body and Fig. A.1.

Second, *unintended white artifacts* occur. Since pixel values above a predefined threshold are amplified, some regions become saturated, appearing as white, and lose their original colors. For example, areas with clouds in the sky, waterfalls, and white walls become excessively white, as depicted in Fig. 8(f)-(g) in the main body and Fig. A.2. To verify the relationship between the scaling and flare generation, we conducted an experiment shown in the bottom row of Fig. A.2. Without the scaling procedure, flares fail to appear even in frames where they are expected, as illustrated in the bottom row of Fig. A.2(d). Approximately 12% of the videos exhibit these unintended white artifacts caused by the scaling, which makes the data unsuitable for deep-learning training.

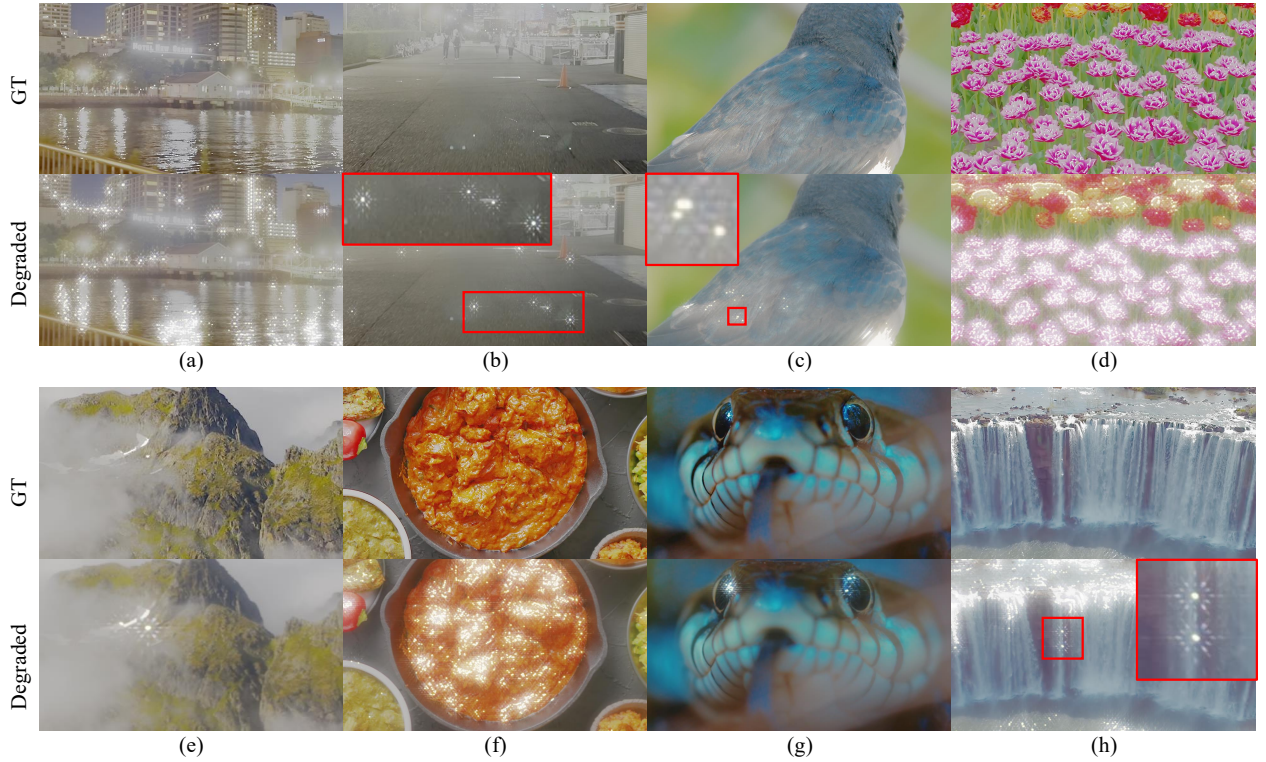


Figure A.1. The visual illustration that showcases improbable flares resulting from excessive scaling in the VidUDC33K dataset [8]. (a) Flare in the river. (b) Flare from the dust on the camera lens. (c) Flare on the bird feathers. (d) Flare on the flower petals. (e) Flare on the mountain peaks. (f) Flare on the food. (g) Flare in the snake eyes. (h) Flare on the waterfalls.

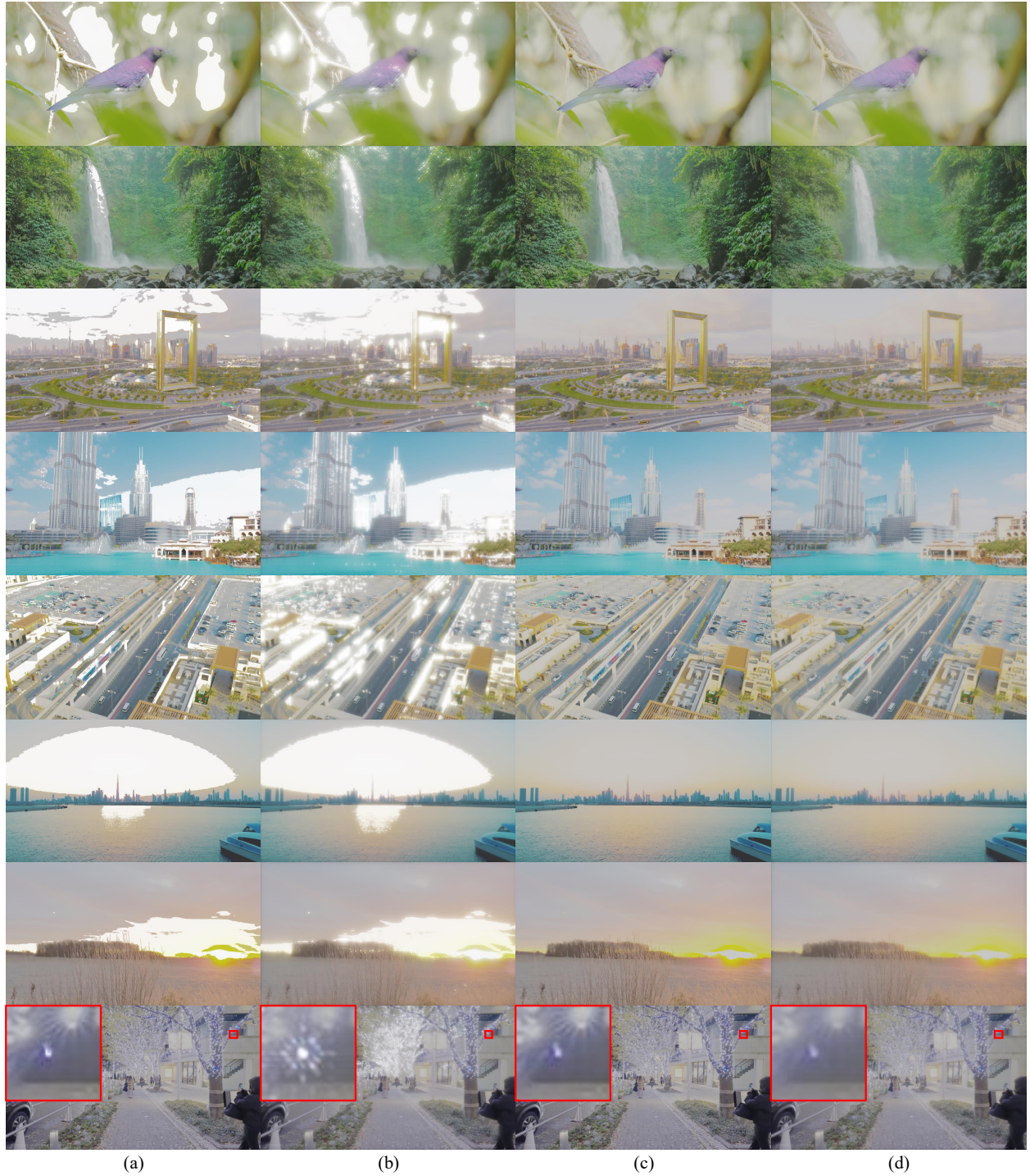


Figure A.2. The visual depiction that shows white artifacts resulting from excessive scaling in the VidUDC33K dataset [8]. The frames processed *without* the scaling procedure do not exhibit these artifacts (see (c) and (d)), in contrast to the frames processed *with* the scaling procedure (see (a) and (b)). Notably, flares are visible only when the scaling procedure is applied (see (b)), while absent without it (see (d)). This suggests that the authors rely on scaling to generate flares, unintentionally producing unrealistic white artifacts. (a) The ground-truth frame *with* scaling procedure. (b) The degraded frame *with* scaling procedure. (c) The ground-truth frame *without* scaling procedure. (d) The degraded frame *without* scaling procedure.



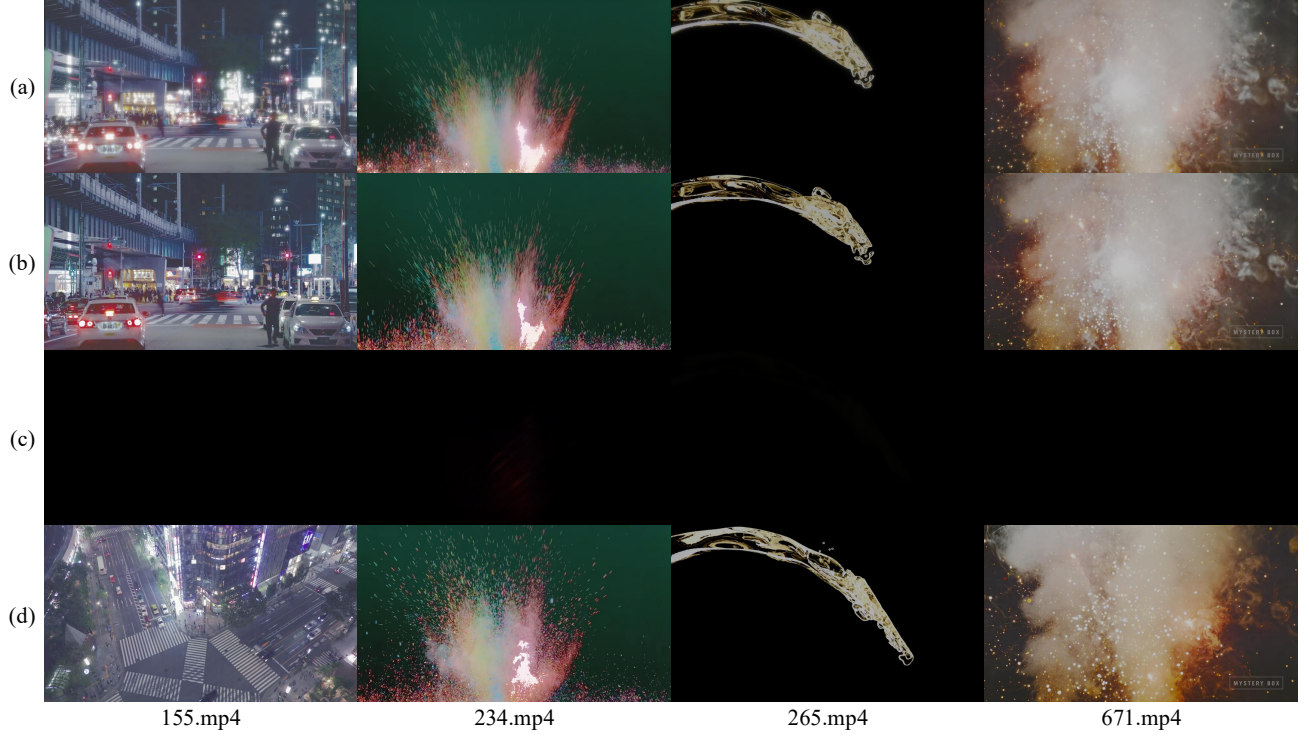


Figure A.3. The visual representation that demonstrates black frames resulting from incorrectly transformed PSFs in the VidUDC33K dataset [8]. (a) The first frame of the degraded video. (b) The first frame of the ground-truth video. (c) The tenth frame of the degraded video. (d) The tenth frame of the ground-truth video.

**The darkened and nearly featureless frames.** Liu *et al.* [8] strive to create temporally variant flares in continuous video sequences. They simulate the dynamic changes of the PSF during motion by computing the inter-frame homography matrix  $H_{t-1 \rightarrow t}$ , formulated as Eq. (1), between consecutive frames.

$$\begin{aligned}
 k_t &= \mathcal{T}(k_{t-1}, H_{t-1 \rightarrow t}) \\
 &= \left| \mathcal{F} \left( H_{t-1 \rightarrow t}^{-1} \left( \mathcal{F}^{-1} \left( \sqrt{k_{t-1}} \right) \right) \right) \right|^2, \quad (1) \\
 H_{t-1 \rightarrow t} &= \mathcal{M}(I_{t-1}^{GT}, I_t^{GT}),
 \end{aligned}$$

where  $\mathcal{T}(\cdot)$  is the transformation function that utilizes  $H_{t-1 \rightarrow t}^{-1}$  to perform a perspective warp on the PSF of the previous frame,  $k_{t-1}$ .  $H_{t-1 \rightarrow t}^{-1}$  denotes the inverse matrix of  $H_{t-1 \rightarrow t}$ .  $\mathcal{F}(\cdot)$  and  $\mathcal{F}^{-1}(\cdot)$  represent the Fourier transform and its inverse, respectively.  $\mathcal{M}(\cdot)$  is the matching component used to calculate the homography matrix between frames.

However, this process occasionally results in PSF values approaching zero, causing the degraded frames to appear entirely black. Specifically, this issue occurs in 4 out of 677 videos. The first frame does not undergo PSF transformation, while subsequent frames do. Therefore, as seen in

Fig. 8(e) in the main body and Fig. A.3(c), the frames after the first one (*e.g.*, the tenth frame) sometimes become black.

## B. Cross-dataset Validation

This section demonstrates cross-dataset validation to address the unique dataset distribution and degradation patterns of UDC datasets, as discussed in Sec. 6. Transfer learning techniques are crucial for addressing varying dataset distributions and degradation patterns in practical UDC restoration. These methods include approaches such as fine-tuning and domain adaptation [3, 6, 7]). In this section, we focus on fine-tuning and evaluate its effectiveness across multiple UDC datasets.

To better illustrate the challenges in cross-dataset generalization and motivate the need for transfer learning, we analyze device-level differences across representative UDC datasets. For example, Samsung Galaxy Z-Fold 5 (UDC-VIT) and ZTE Axon 20 [17] (VidUDC33K [8]) have vastly different pixel designs, as they come from different vendors. Similarly, Samsung Galaxy Z-Fold 3 [12] (UDC-SIT [1]) and Samsung Galaxy Z-Fold 5 [13] (UDC-VIT) share relatively similar designs, but they still exhibit differences.

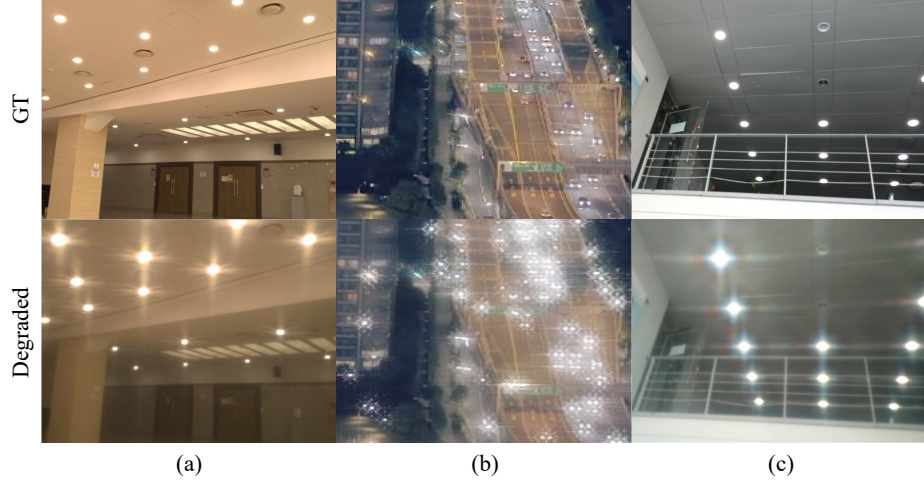


Figure B.1. Comparison of the UDC datasets, showing varied data distribution and degradation patterns. (a) UDC-SIT [1]. (b) VidUDC33K [8]. (c) UDC-VIT.

Table B.1. The design of experiments to verify the effect of fine-tuning and the use of a real-world dataset (*e.g.*, UDC-VIT). The first and the second subscripts beside  $\mathcal{M}$  indicate the training and fine-tuning datasets, respectively. For example,  $\mathcal{M}_{s3}$  refers to the model trained on UDC-SIT without fine-tuning, while  $\mathcal{M}_{s3s5}$  denotes the model trained on UDC-SIT and subsequently fine-tuned on UDC-VIT. Models without subscripts are trained and tested on the same dataset.

Experiments	Model name	Training dataset	Fine-tuning dataset	Test dataset
Exp. 1	$\mathcal{M}_{s3}$	UDC-SIT	-	UDC-VIT
	$\mathcal{M}_{s3s5}$	UDC-SIT	UDC-VIT	UDC-VIT
	$\mathcal{M}$	UDC-VIT	-	UDC-VIT
Exp. 2	$\mathcal{M}_{s5}$	UDC-VIT	-	VidUDC33K
	$\mathcal{M}_{s5z20}$	UDC-VIT	VidUDC33K	VidUDC33K
	$\mathcal{M}$	VidUDC33K	-	VidUDC33K
Exp. 3	$\mathcal{M}_{z20}$	VidUDC33K	-	UDC-VIT
	$\mathcal{M}_{z20s5}$	VidUDC33K	UDC-VIT	UDC-VIT
	$\mathcal{M}$	UDC-VIT	-	UDC-VIT

Fig. B.1(a) and (c) illustrate that the UDC-SIT and UDC-VIT datasets show similar degradation, such as blur, decrease in transmittance, and flare shape. In contrast, Fig. B.1(b) and (c) highlight the stark difference between the VidUDC33K and UDC-VIT datasets. This discrepancy arises from two factors: the variation in pixel design and the synthetic nature of the VidUDC33K dataset, which results in unrealistic degradation patterns.

Fine-tuning models to address variant dataset distributions or degradation patterns is crucial in practical applications. To evaluate the effect of fine-tuning and validate the effectiveness of UDC-VIT, which reflects real-world degradation, we conduct three experiments (Exp. 1-3), as shown in Tab. B.1. The subscripts beside the model name  $\mathcal{M}$  specify the datasets used for training and fine-tuning. For example,  $\mathcal{M}_{s3}$  refers to the model trained on UDC-SIT (Samsung Galaxy Z-Fold 3) without fine-tuning,  $\mathcal{M}_{s5z20}$  is

trained on UDC-VIT (Samsung Galaxy Z-Fold 5) and fine-tuned on VidUDC33K (ZTE Axon 20), while  $\mathcal{M}_{z20s5}$  is trained on VidUDC33K (ZTE Axon 20) and fine-tuned on UDC-VIT (or Samsung Galaxy Z-Fold 5). We use models  $\mathcal{M}$  such as UDC-UNet [9], DISCNet [4], and DDR-Net [8] among six benchmark models in Tab. 3. Fine-tuning is performed for 10% or 20% of the total iterations, with the learning rate set to 10% or 20% of the original value.

**Experiment 1: impact of fine-tuning on UDC-VIT.** This experiment evaluates the impact of fine-tuning on UDC-VIT when tested on UDC-VIT. We compare the performance of models trained on UDC-SIT, with or without fine-tuning on UDC-VIT, and models trained solely on UDC-VIT. Since UDC-SIT is a still image dataset, we use UDC-UNet and DISCNet as models  $\mathcal{M}$ , which are specifically designed for UDC still images. As presented

in Tab. B.2, DISCNet<sub>s3</sub> and UDC-UNet<sub>s3</sub> trained exclusively on UDC-SIT struggle to generalize to UDC-VIT. In contrast, DISCNet<sub>s3s5</sub> and UDC-UNet<sub>s3s5</sub>, which incorporate fine-tuning with UDC-VIT, demonstrate superior restoration performance for UDC-VIT degradations. In particular, increasing the number of fine-tuning iterations further enhances the performance. This is also evident in Fig. B.2. Models without fine-tuning (*e.g.*, DISCNet<sub>s3</sub> and UDC-UNet<sub>s3</sub>) struggle to restore UDC-VIT, as shown in Fig. B.2(b). In contrast, models with fine-tuning (*e.g.*, DISCNet<sub>s3s5</sub> and UDC-UNet<sub>s3s5</sub>), as shown in Fig. B.2(c) and (d), achieve restoration performance com-

parable to those solely trained on UDC-VIT (*e.g.*, DISCNet and UDC-UNet), as depicted in Fig. B.2(e).

These findings lead to the following conclusions. Although the Samsung Galaxy Z-Fold 3 (UDC-SIT) and Samsung Galaxy Z-Fold 5 (UDC-VIT) share similar pixel designs, being from the same vendor, their differences are substantial enough to require fine-tuning. With adequate adaptation, however, these models effectively leverage degradations from other UDC devices, underscoring the potential for cross-device generalization with fine-tuning.

Table B.2. **[Exp. 1]** Restoration performance of DISCNet [4] and UDC-UNet [9] when tested on UDC-VIT. They are trained on UDC-SIT [1], either with or without fine-tuning on UDC-VIT, or solely trained on UDC-VIT. Models without subscripts refer to those solely trained on UDC-VIT, meaning their PSNR, SSIM, and LPIPS values match those in Tab. 3 in the main body. The number of iterations represents the percentage of fine-tuning iterations relative to the total iterations in the original configurations provided by the authors.

Model name	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	Training	Fine-tuning ( $\#Iterations$ )	Test
DISCNet <sub>s3</sub>	16.81	0.7139	0.3293	UDC-SIT	-	UDC-VIT
DISCNet <sub>s3s5</sub>	23.16	0.8281	0.2527	UDC-SIT	UDC-VIT (10%)	UDC-VIT
DISCNet <sub>s3s5</sub>	23.57	0.8331	0.2459	UDC-SIT	UDC-VIT (20%)	UDC-VIT
DISCNet	24.70	0.8403	0.2675	UDC-VIT	-	UDC-VIT
UDC-UNet <sub>s3</sub>	17.21	0.7260	0.3400	UDC-SIT	-	UDC-VIT
UDC-UNet <sub>s3s5</sub>	24.94	0.8709	0.2113	UDC-SIT	UDC-VIT (10%)	UDC-VIT
UDC-UNet <sub>s3s5</sub>	25.41	0.8758	0.2015	UDC-SIT	UDC-VIT (20%)	UDC-VIT
UDC-UNet	28.00	0.8911	0.1779	UDC-VIT	-	UDC-VIT

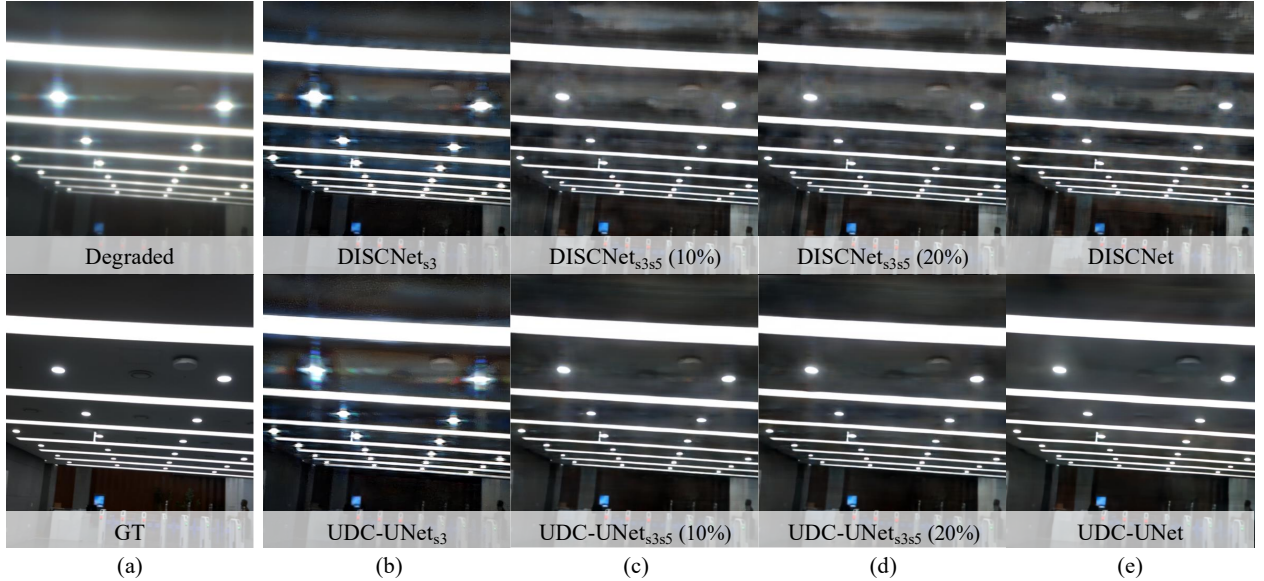


Figure B.2. **[Exp. 1]** Comparison of restoration performance across different models when tested on the UDC-VIT dataset. (a) UDC-degraded (at the first row) and GT (at the second row) frames of the UDC-VIT dataset. Test frames from UDC-VIT by DISCNet (at the first row) and UDC-UNet (at the second row) trained on UDC-SIT (b) without fine-tuning on UDC-VIT, (c) with fine-tuning on UDC-VIT (10%), and (d) with fine-tuning on UDC-VIT (20%). (e) Test frames from UDC-VIT by DISCNet (at the first row) and UDC-UNet (at the second row), solely trained on UDC-VIT.



**Experiment 2: impact of fine-tuning on VidUDC33K.**

This experiment aims to assess the impact of fine-tuning on VidUDC33K when tested on VidUDC33K. It compares the performance of models trained on UDC-VIT, with or without fine-tuning on VidUDC33K, and models trained solely on VidUDC33K. For models  $\mathcal{M}$ , we use UDC-UNet, DISCNet, and DDRNet, which are explicitly designed to address UDC degradations.

As described in Tab. B.3, fine-tuning across datasets from different devices (*e.g.*, Samsung Galaxy Z-Fold 5 and ZTE Axon 20) improves generalization to different data distributions. Interestingly, the fine-tuned models  $\text{DISCNet}_{s5z20}$  and  $\text{UDC-UNet}_{s5z20}$  outperform DISCNet and UDC-UNet, solely trained by VidUDC33K, as shown in Tab. B.3. This performance boost can be attributed to the fact that UDC-VIT exhibits more realistic and severe degra-

Table B.3. **[Exp. 2]** Restoration performance of DISCNet [4], UDC-UNet [9], and DDRNet [8] when tested on VidUDC33K [8]. They are trained on UDC-VIT, either with or without fine-tuning on VidUDC33K, or solely trained on VidUDC33K. Models without subscripts refer to those solely trained on VidUDC33K, meaning their PSNR, SSIM, and LPIPS values match those in Tab. 3 in the main body. The number of iterations represents the percentage of fine-tuning iterations relative to the total iterations in the original configurations provided by the authors.

Model name	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	Training	Fine-tuning (#Iterations)	Test
DISCNet <sub>s5</sub>	18.73	0.7503	0.4159	UDC-VIT	-	VidUDC33K
DISCNet <sub>s5z20</sub>	28.89	0.9129	0.1727	UDC-VIT	VidUDC33K (10%)	VidUDC33K
DISCNet	28.89	0.8405	0.2432	VidUDC33K	-	VidUDC33K
UDC-UNet <sub>s5</sub>	19.84	0.7682	0.3737	UDC-VIT	-	VidUDC33K
UDC-UNet <sub>s5z20</sub>	29.57	0.9139	0.1506	UDC-VIT	VidUDC33K (10%)	VidUDC33K
UDC-UNet	28.37	0.8361	0.2561	VidUDC33K	-	VidUDC33K
DDRNet <sub>s5</sub>	20.10	0.8313	0.3446	UDC-VIT	-	VidUDC33K
DDRNet <sub>s5z20</sub>	29.12	0.8994	0.2073	UDC-VIT	VidUDC33K (10%)	VidUDC33K
DDRNet	31.91	0.9313	0.1306	VidUDC33K	-	VidUDC33K

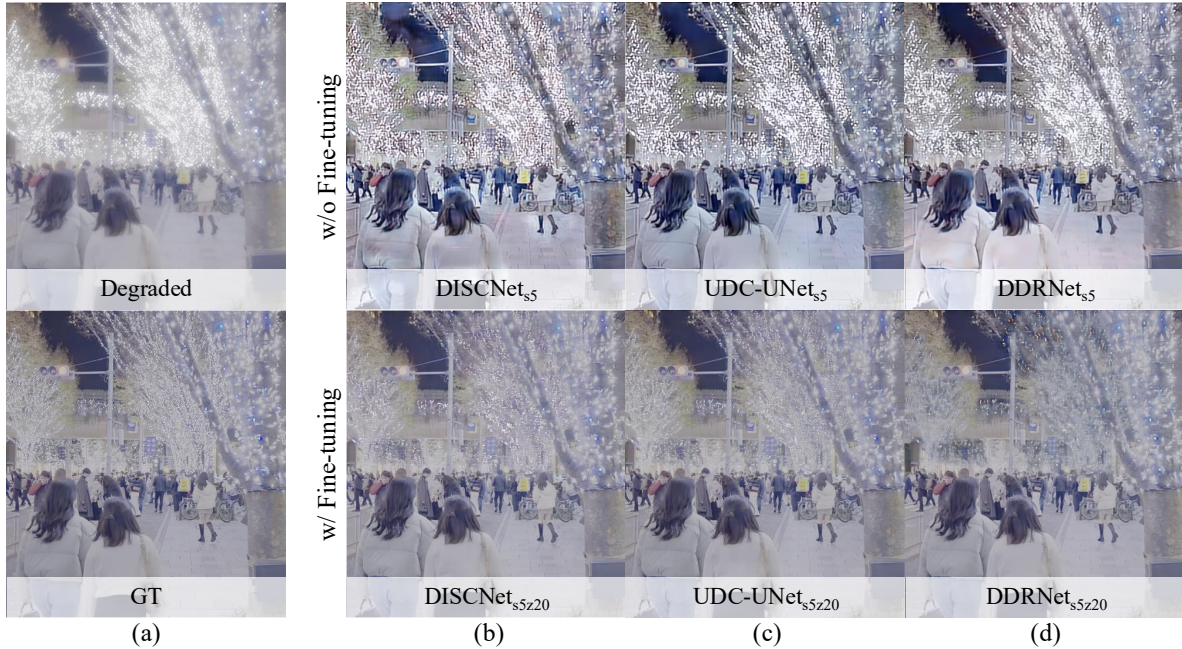


Figure B.3. **[Exp. 2]** Comparison of restoration performance across different models when tested on the VidUDC33K dataset [8]. (a) UDC-degraded (at the first row) and GT (at the second row) frames of the VidUDC33K dataset. The models in the first row are pre-trained on UDC-VIT without fine-tuning on VidUDC33K: (b)  $\text{DISCNet}_{s5}$ . (c)  $\text{UDC-UNet}_{s5}$ . (d)  $\text{DDRNet}_{s5}$ . The models in the second row are pre-trained on UDC-VIT with fine-tuning on VidUDC33K: (b)  $\text{DISCNet}_{s5z20}$ . (c)  $\text{UDC-UNet}_{s5z20}$ . (d)  $\text{DDRNet}_{s5z20}$ , showing improved restoration performance.

dation patterns, such as noise, blur, transmittance decrease, and variant flares, compared to the synthetic VidUDC33K dataset, as discussed in Sec. 4 in the main body. Consequently, models pre-trained on UDC-VIT exhibit improved performance with fine-tuning on VidUDC33K, underscoring the benefits of utilizing real-world images as training datasets.

As described in the first row of Fig. B.3(b), (c), and (d), models trained on UDC-VIT without fine-tuning (*e.g.*, DISCNet<sub>s5</sub>, UDC-UNet<sub>s5</sub>, and DDRNet<sub>s5</sub>) restore blur but fail to address flare artifacts. Interestingly, they show better restoration of transmittance decrease compared to VidUDC33K’s ground truth, probably due to the brighter tone in UDC-VIT’s ground truth compared to VidUDC33K’s. On the other hand, models fine-tuned on VidUDC33K (*e.g.*, DISCNet<sub>s5z20</sub>, UDC-UNet<sub>s5z20</sub>, and DDRNet<sub>s5z20</sub>) effectively restore the complex degradation patterns specific to VidUDC33K, as shown in the second row of Fig. B.3(b), (c), and (d).

**Experiment 3: comparison of UDC-VIT and VidUDC33K.** This experiment evaluates the effect of fine-tuning on UDC-VIT when tested on UDC-VIT. We compare the performance of models trained on VidUDC33K, with or without fine-tuning on UDC-VIT,

and models trained solely on UDC-VIT. For the models  $\mathcal{M}$ , we use DDRNet, the only publicly available pre-trained model trained on VidUDC33K. As shown in Tab. B.4, DDRNet<sub>z20</sub> without fine-tuning on UDC-VIT fails to adequately handle the complex, severe, and real-world degradations present in UDC-VIT. In contrast, DDRNet<sub>z20s5</sub>, fine-tuned on UDC-VIT, demonstrates significant performance improvements over DDRNet<sub>z20</sub>.

However, as illustrated in Fig. B.4, even with fine-tuning, DDRNet<sub>z20s5</sub> still shows limitations in handling specific real-world degradations, such as severe flares. Notably, models pre-trained on the real-world UDC-VIT dataset and fine-tuned on the synthetic VidUDC33K dataset show strong performance in restoring degraded images in VidUDC33K (Experiment 2). In contrast, models pre-trained on the synthetic VidUDC33K dataset and fine-tuned on the real-world UDC-VIT dataset struggle to handle real-world degradation in UDC-VIT effectively (Experiment 3). These findings underscore the importance of pre-training on real-world datasets, such as UDC-VIT, to capture better complex degradations that synthetic data cannot fully replicate.

Table B.4. [Exp. 3] Restoration performance of DDRNet [8] trained on VidUDC33K [8], with or without additional fine-tuning on UDC-VIT. Models without subscripts refer to those trained solely on UDC-VIT, meaning their PSNR, SSIM, and LPIPS values match those in Tab. 3 in the main body. The number of iterations represents the percentage of fine-tuning iterations relative to the total iterations in the original configurations provided by the authors.

Model name	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	Training	Fine-tuning ( $\#Iterations$ )	Test
DDRNet <sub>z20</sub>	11.20	0.5331	0.5609	VidUDC33K	-	UDC-VIT
DDRNet <sub>z20s5</sub>	21.92	0.8302	0.2519	VidUDC33K	UDC-VIT (10%)	UDC-VIT
DDRNet	24.68	0.8539	0.2218	UDC-VIT	-	UDC-VIT

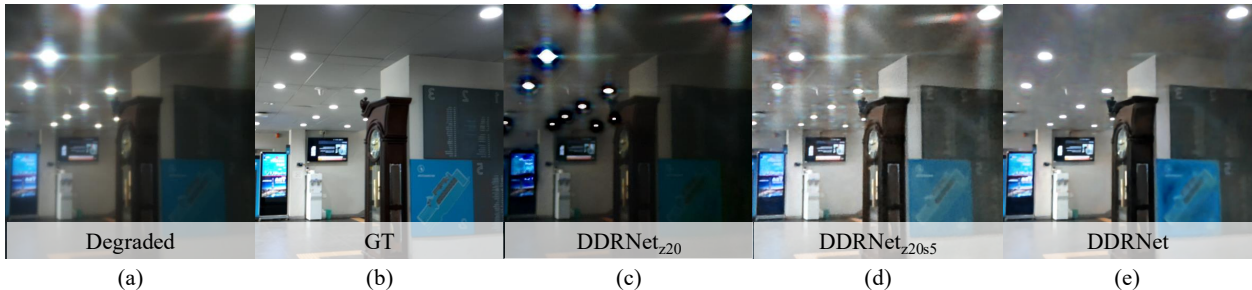


Figure B.4. [Exp. 3] Restoration performance of DDRNet [8] tested on the UDC-VIT dataset. (a) UDC-degraded and (b) GT images from the UDC-VIT dataset. Restored images by (c) DDRNet<sub>z20</sub>, (d) DDRNet<sub>z20s5</sub>, and (e) DDRNet. The model DDRNet<sub>z20</sub> is pre-trained on VidUDC33K without fine-tuning on UDC-VIT, while DDRNet<sub>z20s5</sub> is pre-trained on VidUDC33K with fine-tuning on UDC-VIT, showing improved restoration performance. However, compared to the results in Fig. B.3, the fine-tuned model still struggles to handle the real-world degradations present in the UDC-VIT dataset, since it is pre-trained on the synthetic VidUDC33K dataset.

## C. Effects of Video Capturing Environment

Tab. C.1 provides a comparative evaluation of the restoration performance across different deep-learning models, considering the impact of UDC-VIT dataset annotations. Since a single UDC-VIT video may include multiple annotations (*e.g.*, an indoor scene with flares caused by artificial light), the annotation type listed in a column in Tab. C.1 cannot be interpreted as the sole factor influencing UDC degradation. Nevertheless, it is reasonable to consider the annotation type as a key factor influencing PSNR, SSIM, and LPIPS values.

**Light sources.** As illustrated in Fig. C.1, flares can be categorized into glare, shimmer, and streak [1, 2], marked as red, green, and yellow arrows, respectively. A glare is characterized by intense and robust light, resulting in circular patterns as artifacts. Shimmer entails rapid and nuanced variations in light or color intensity across an image. A streak manifests as a lengthy, slender, and usually irregular line of light or color within an image.

As shown in Tab. C.1, all models encounter difficulties in restoring scenes with flare (Flare - Present - Average) compared to those without flare (Flare - Absent). Within flare-present scenes, the severity of degradation varies based on the light source (*i.e.*, *light source variant flare*) and location (*i.e.*, *spatially variant flare*). For example, in Fig. C.1(a),

sunlight-induced flares are intense, causing all models to struggle to restore obscured objects. In contrast, flares from artificial light, shown in Fig. C.1(b) and (c), are relatively more straightforward to restore than sunlight-induced flares. Flares caused by natural light are often scattered by windows, as illustrated in Fig. C.1(c) and (d), whereas those from artificial light sources with diffusers appear softened, as shown in Fig. C.1(e). All models demonstrate effective restoration of these flare artifacts. However, some models fail to restore reflected light on the human face, as shown in Fig. C.1(f), which may impact face recognition accuracy.

**Shooting location.** In the UDC-VIT dataset, 33.3% of outdoor and 15.4% of indoor scenes contain sunlight-induced flares. The benchmark models face more difficulty restoring outdoor scenes than indoor ones, as shown in Tab. C.1. Sunlight-induced flares in outdoor scenes (Fig. C.1(a)) are more intense than those scattered by windows in indoor scenes (Figures C.1(c) and (d)).

Restoration performance is sometimes more affected by the presence of flares than by the shooting location alone. For example, among two indoor scenes, the models restore the flare-free frame (Fig. C.1(h)) more effectively than the frame with flare (Fig. C.1(b)). Similarly, among two outdoor scenes, the models perform better on the flare-free frame (Fig. C.1(i)) than on the frame with flare (Fig. C.1(a)).

Table C.1. Comparison of restoration performance. Each row’s best and worst scores within each category are bold-faced and underlined, respectively.

Model	Metric	Flare presence and light sources					Shooting location		Human presence		Average
		Present				Absent	Indoor	Outdoor	Present	Absent	
		Natural sunlight	Artificial light	Both	Average						
DISCNet [4]	PSNR $\uparrow$	22.34	24.15	22.37	23.64	26.92	25.43	22.34	26.94	20.33	24.70
	SSIM $\uparrow$	0.7704	0.8451	0.8191	0.8303	0.8611	0.8573	0.7852	0.8795	0.7636	0.8403
	LPIPS $\downarrow$	0.2633	0.2894	0.3250	0.2901	0.2202	0.2608	0.2891	0.2247	0.3512	0.2675
UDC-UNet [9]	PSNR $\uparrow$	23.73	27.76	25.36	26.83	30.46	29.13	24.33	31.34	21.47	28.00
	SSIM $\uparrow$	0.8197	0.8995	0.8857	0.8856	0.9027	0.9092	0.8324	0.9276	0.8198	0.8911
	LPIPS $\downarrow$	0.1878	0.1814	0.2173	0.1871	0.1588	0.1679	0.2107	0.1398	0.2526	0.1779
FastDVDNet [14]	PSNR $\uparrow$	23.13	23.78	21.49	23.38	24.97	24.34	22.45	25.34	21.06	23.89
	SSIM $\uparrow$	0.7902	0.8523	0.8244	0.8392	0.8538	0.8593	0.7940	0.8720	0.7891	0.8439
	LPIPS $\downarrow$	0.2621	0.2772	0.3048	0.2785	0.2403	0.2568	0.2965	0.2364	0.3244	0.2662
EDVR [15]	PSNR $\uparrow$	21.99	23.14	21.57	22.76	25.20	24.07	21.88	25.11	20.51	23.55
	SSIM $\uparrow$	0.7711	0.8422	0.8132	0.8276	0.8445	0.8484	0.7833	0.8612	0.7781	0.8331
	LPIPS $\downarrow$	0.2565	0.2843	0.3039	0.2826	0.2351	0.2605	0.2893	0.2389	0.3227	0.2673
ESTRNN [16]	PSNR $\uparrow$	23.61	25.54	24.08	25.05	26.05	26.07	23.12	26.99	22.22	25.38
	SSIM $\uparrow$	0.7997	0.8818	0.8577	0.8662	0.8637	0.8847	0.8025	0.8938	0.8098	0.8654
	LPIPS $\downarrow$	0.2415	0.2192	0.2640	0.2285	0.2072	0.2086	0.2636	0.1920	0.2794	0.2216
DDRNet [8]	PSNR $\uparrow$	23.24	24.14	23.49	23.92	26.25	25.35	22.49	26.43	21.23	24.68
	SSIM $\uparrow$	0.8015	0.8628	0.8455	0.8512	0.8594	0.8697	0.8025	0.8810	0.8007	0.8539
	LPIPS $\downarrow$	0.2283	0.2267	0.2433	0.2291	0.2066	0.2079	0.2672	0.1936	0.2771	0.2218



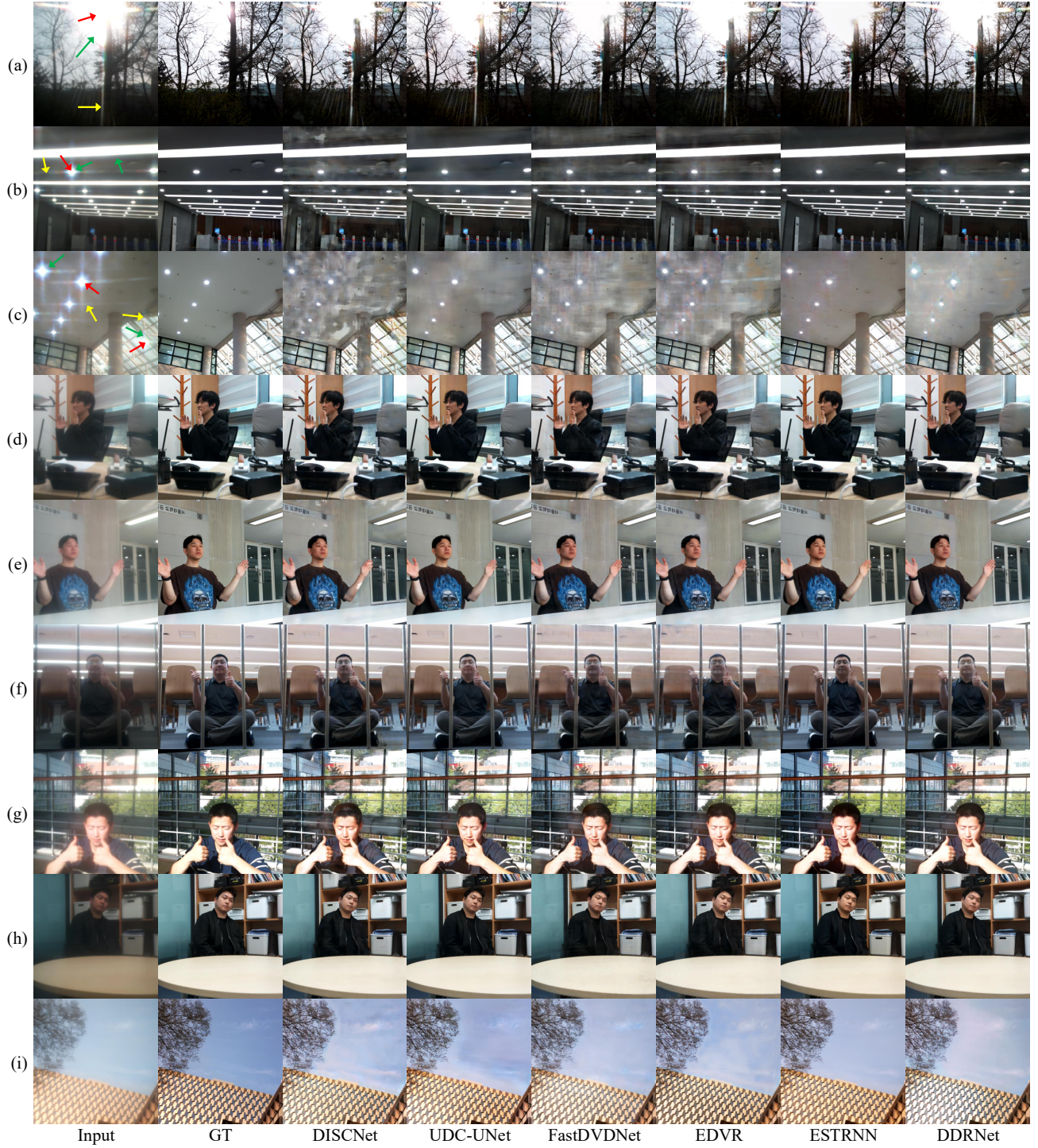


Figure C.1. The visual comparison of the restoration performance regarding different annotations. The red, green, and yellow arrows represent the flares' glare, shimmer, and streak, respectively. (a) Natural sunlight + Human absent + Outdoor. (b) Artificial light + Human absent + Indoor. (c) Both + Human absent + Indoor. (d) Natural sunlight + Hand waving + Indoor. (e) Artificial light + Hand waving + Indoor. (f) Artificial light + Thumbs-up + Indoor. (g) Natural sunlight + Thumbs-up + Indoor. (h) Flare absent + Body-swaying + Indoor. (i) Flare absent + Human absent + Outdoor.

**Human.** The presence of humans alone does not pose a significant challenge to restoration. Instead, the restoration difficulty hinges on how UDC degradations, such as noise, blur, transmittance decrease, and flare, impact humans. For example, in Fig. C.1(d) and (e), despite the presence of flares in the frames, they do not affect humans. However, in Fig. C.1(f), the reflection of fluorescent light on the person’s glasses poses challenges to restoring fine details around the eyes. In Fig. C.1(g) and (h), human faces appear reddish in the input frames compared to the ground-truth frames due to UDC-induced diffraction occurring differently across RGB channels. In addition, the restored facial colors vary among the models. Considering the importance of facial color in image and video-based applications, addressing diverse UDC degradations is essential for reliable human restoration in face recognition and video conferencing.

**Flicker.** The visual comparison presented in this paper is limited to a single frame. While specific frames in Fig. C.1 exhibit successful restoration, temporal inconsistencies, such as flickering, are frequently observed across multiple frames in the video for all models. This flickering may result from varying degradations between consecutive frames, such as decreases in transmittance and spatially and temporally variant flares. To view the restored videos featuring flickers, please visit <https://mcrl.github.io/UDC>.

## D. Details of the UDC-VIT Dataset

In this section, we provide detailed information on the UDC-VIT dataset.

### D.1. Dataset Acquisition

As described in Sec. 3 in the main body of the paper, to construct a real-world UDC video dataset with precise alignment and synchronization, we propose a video-capturing system. We first describe the synchronization strategy between the two camera modules, which includes controller considerations, stable auto-exposure settings, fixed frame rates, and consistent pixel binning to ensure matched resolution. This section then details the alignment algorithm based on the Discrete Fourier Transform (DFT) and its advantages.

**The controller.** When capturing videos, we discard the initial 30 frames because it takes approximately 15 frames for the ground-truth camera and 25 frames for the UDC to achieve focus. The UDC requires more frames for focusing due to its degradation. Furthermore, we use a solid-state drive (SSD) instead of a secure digital (SD) card, since the SD card takes longer to save FHD resolution videos, which disrupts synchronization between the two cameras.

**Exposure setting and synchronization.** Since the presence of the display panel affects illuminance levels, all captures are performed in well-lit environments to ensure stable auto-exposure (AE). We fix the frame rate for temporal synchronization of corresponding frames in GT and degraded videos, while AE adjusts the exposure time. In such lighting conditions, both UDC and non-UDC cameras exhibit similar AE responses in terms of exposure time and gain. This is attributed to the QBC sensor’s high sensitivity, achieved through pixel binning that aggregates light from adjacent pixels. As a result, even with reduced transmittance from the display, the UDC camera achieves adequate brightness without significantly more prolonged exposure. In contrast, in low-light conditions (e.g., nighttime scenes), exposure mismatches can lead not only to synchronization issues but also to poorly rendered dark regions due to noise, underexposure, or loss of detail. Thus, we exclude such cases from the dataset.

**Resolution consistency and image processing.** To ensure consistent resolution between GT and degraded frames, we apply pixel binning consistently for both cameras using `rpicam` utility [5], constraining the output resolution below 16 megapixels (MP) (e.g., FHD). This binning process ensures that the same effective resolution is used across both clean and degraded captures. The output of this binning is then processed by the onboard ISP.

**Alignment.** The alignment algorithm we use involves *shifting*, *rotating*, and *cropping* paired frames with DFT. The detailed algorithm is illustrated in Algorithm D.1. In this algorithm, following Ahn *et al.* [1]’s setting, we use  $\lambda_1 = \lambda_3 = 1$  and  $\lambda_2 = 0$ , and we do not apply rotation. Their experiments show that applying rotation reduces the Percentage of Correct Keypoints (PCK) when varying  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $\theta_{\text{rotation}}$ .

The loss function in Eq. (1) in the main body enables the incorporation of both local (i.e., MSE) and global (i.e., DFT) information across spatial and frequency domains. Using DFT to align the paired frames offers a significant advantage because it can decompose a frame into its constituent spatial frequency components. Fig. D.1(a) and (c) depict paired frames  $\mathcal{G}$  and  $\mathcal{D}$  comprising multiple sinusoidal gratings, indicating a noticeable spatial shift. Fig. D.1(b) and (d) represent the phase and amplitude differences between the paired frames, respectively. Thus, reducing the phase component is critical for effectively aligning the paired frames for the same scene.

### D.2. Dataset Details and Statistics

This section provides detailed information about the UDC-VIT dataset.

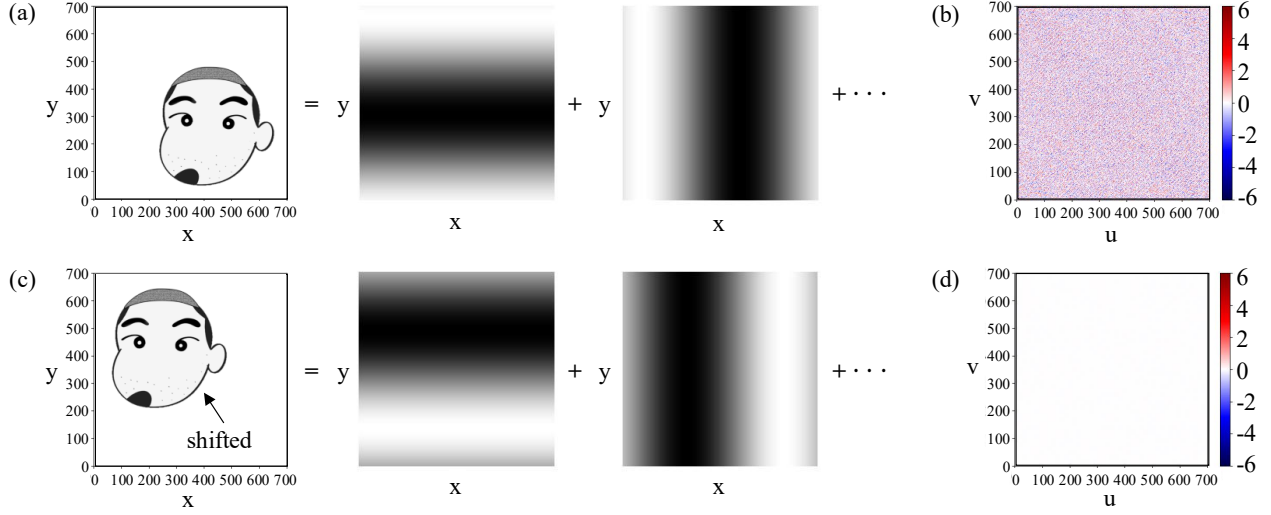


Figure D.1. Frequency analysis based on the conceptual illustration for paired frames involving shifts without degradation. (a) The original frame  $\mathcal{G}$  consists of multiple sinusoidal gratings. The inverse  $DFT$  applied to  $\mathcal{F}_G(u, v)$  produces each sinusoidal grating. (b) The phase difference between  $\mathcal{G}$  and  $\mathcal{D}$ . (c) The spatially shifted frame  $\mathcal{D}$  in the spatial domain comprises multiple sinusoidal gratings, as in (a). (d) The amplitude difference between  $\mathcal{G}$  and  $\mathcal{D}$ , showing no difference.

---

**Algorithm D.1** Alignment of paired images  $I_G$  and  $I_D$  [1].

---

**Require:** Images  $I_G$ ,  $I_D$  of size  $(H, W)$ , hyperparameters  $s, \theta_r, r, \lambda_1, \lambda_2, \lambda_3$

**Ensure:** Aligned images  $\mathcal{G}$ ,  $\mathcal{D}$  of size  $(H^*, W^*)$

Crop  $\mathcal{G}$  from  $I_G$  using center crop

Crop  $\mathcal{D}$  from  $I_D$  to the size of  $\mathcal{G}$

Initialize best loss  $\mathcal{L}_{\text{best}}$  to a large value

Initialize optimal shifts  $s_{\text{opt}_x}$ ,  $s_{\text{opt}_y}$ , and rotation  $\theta_{\text{opt}}$  to 0

**for**  $\theta_{\text{rotation}}$  from  $-\theta_r$  to  $\theta_r$  with step  $r$  **do**

    Apply rotation of  $\theta_{\text{rotation}}$  to  $I_D$  to get  $\mathcal{D}_{\text{rotated}}$

**for**  $x_{\text{shift}}$  from  $-s$  to  $s$  with step 1 **do**

**for**  $y_{\text{shift}}$  from  $-s$  to  $s$  with step 1 **do**

            Calculate crop position  $(p, q)$  relative to the center crop:

$p = x_{\text{center\_crop}} + x_{\text{shift}}$

$q = y_{\text{center\_crop}} + y_{\text{shift}}$

            Crop image  $\mathcal{D}_{\text{tmp}}$  from  $\mathcal{D}_{\text{rotated}}$  at position  $(p, q)$

            Calculate loss  $\mathcal{L}$  using the loss function in **Eq. (1)** between  $\mathcal{D}_{\text{tmp}}$  and  $\mathcal{G}$

**if**  $\mathcal{L} < \mathcal{L}_{\text{best}}$  **then**

                Update  $\mathcal{L}_{\text{best}}$  to  $\mathcal{L}$

                Update  $s_{\text{opt}_x}$  to  $x_{\text{shift}}$

                Update  $s_{\text{opt}_y}$  to  $y_{\text{shift}}$

                Update  $\theta_{\text{opt}}$  to  $\theta_{\text{rotation}}$

**end if**

**end for**

**end for**

**end for**

Apply optimal rotation  $\theta_{\text{opt}}$  to  $I_D$  to get  $\mathcal{D}_{\text{rotated}}$

Calculate crop position  $(p_{\text{opt}}, q_{\text{opt}})$  relative to the center crop:

$p_{\text{opt}} = x_{\text{center\_crop}} + s_{\text{opt}_x}$

$q_{\text{opt}} = y_{\text{center\_crop}} + s_{\text{opt}_y}$

Crop  $\mathcal{D}_{\text{rotated}}$  to acquire an aligned image  $\mathcal{D}$  at position  $(p_{\text{opt}}, q_{\text{opt}})$

---



**Statistics.** From a pool of 647 videos, we have randomly selected 510 for training, 69 for validation, and 68 for the test set. The UDC-VIT dataset is available in PNG format accompanied by a conversion script from PNG to NPY. We have annotated each video pair, providing a detailed overview of the total count and the distribution of different annotation labels. The video pairs are thoughtfully categorized into various conditions, including the presence of flare and light sources, the presence of humans and their motion types, and whether the scene is indoor or outdoor.

**IRB approval.** We have obtained the approval of the Institutional Review Board (IRB) for the UDC-VIT dataset, as our research involves human subjects. This rigorous process ensures the highest standards of research ethics. Using IRB-approved procedures, we enlisted 22 voluntary research participants. As shown in Tab. D.1, the IRB-approved information sheet provides detailed instructions to the participants.

Similarly, it is essential to note that the users of the UDC-VIT dataset are engaged in research involving human subjects. Therefore, users are required to obtain IRB approval by the regulations of their respective countries. When users download the dataset, instructions on IRB approval will be provided, as shown in Fig. D.3.

### D.3. Rigorous Maintenance Plan

This section provides easy accessibility and a rigorous maintenance plan for the UDC-VIT dataset.

**Easy accessibility.** The UDC-VIT dataset will be publicly available on our research group’s website, as shown in Fig. D.3. Users can request access by completing a form on the site; a download link will be sent via email upon submission. Detailed instructions are available at <https://github.com/mcrl/UDC-VIT>. Hosting the dataset on the research group’s website ensures long-term availability, while handling contact and bug reports via email facilitates ongoing maintenance and updates.

**License.** The UDC-VIT dataset is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). Under this license, users of the UDC-VIT dataset can freely utilize, share, and modify this work by adequately attributing the original author, distributing any derived works under the same license, and utilizing it exclusively for non-commercial purposes. It is essential to mention that the UDC-VIT dataset is restricted to UDC research purposes only, as outlined in our IRB documentation. Detailed information on this license can be found on the official Creative Commons website.

### E. Reproducibility

This section provides detailed information on the deep-learning models for reproducibility. The code can be found and downloaded at <https://github.com/mcrl/UDC-VIT>.

The learnable restoration models used for evaluating the UDC-VIT dataset include DISCNet [4], UDC-UNet [9], FastDVDNet [14], EDVR [15], ESTRNN [16], and DDRNet [8]. We use a single-node GPU cluster to train each benchmark model. Each node has eight AMD Instinct MI100 GPUs. While we mainly stick to the original authors’ code and training settings for the models, we introduce some modifications to all models except ESTRNN.

- **DISCNet.** DISCNet is designed to restore UDC still images in high dynamic range (HDR) (e.g., SYNTH [4]). Accordingly, we modify the PyTorch DataLoader to use normalization instead of Reinhard tone mapping [11]. The DataLoader randomly selects one frame per video from the UDC-VIT dataset for each iteration during the training and validation phases.
- **UDC-UNet.** UDC-UNet is designed to restore UDC still images in HDR. The original authors do not conduct normalization or tone mapping in the DataLoader and employ a tone mapping L1 loss function. However, since the UDC-VIT dataset has a low dynamic range (LDR), we modify the DataLoader to normalize the input. The model output is clamped between 0 and 1 before computing the

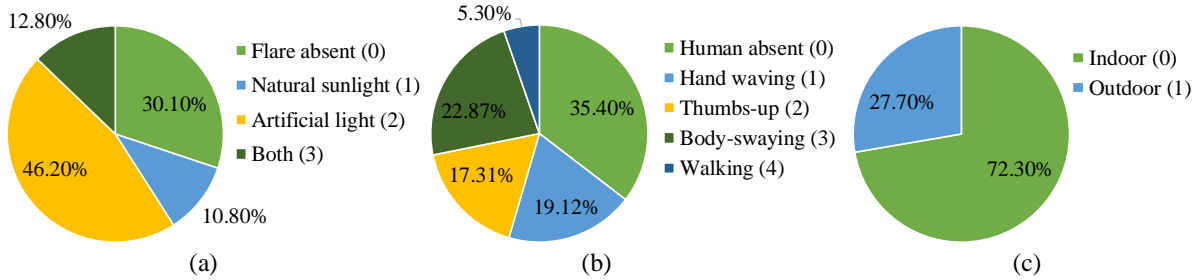


Figure D.2. The dataset distribution. The parenthesis beside a label is the encoding of the label. Note that a video pair can have multiple annotation labels. The distribution of (a) lighting conditions, (b) human presence and their actions, and (c) shooting location.

L1 loss. The DataLoader randomly selects one frame per video from the UDC-VIT dataset for each iteration.

- **FastDVDNet.** FastDVDNet is a video denoising model that utilizes NVIDIA’s Data Loading Library (DALI) [10], processing a noise map and multiple frames as inputs. Instead of DALI, we employ the PyTorch DataLoader tailored to the UDC-VIT dataset in NPY format. We set the noise level to zero. To accommodate the FHD resolution and multiple degradations in the UDC-VIT dataset, we increase the patch size from 64 to 256. Moreover, we extend the training duration of FastDVDNet from the original 95 epochs to 400, allowing the model to reach full convergence.
- **EDVR.** To address the out-of-memory issues of EDVR, which contains 23.6M parameters, we reduce the training patch size from 256 to 192. During inference on the test set, we divide each frame into two patches of size  $3 \times 1060 \times 1060$  and merge the outputs afterward.
- **DDNet.** During inference, the authors of DDNet par-

tion each frame into patches of size  $3 \times 256 \times 256$  and process 50 frames simultaneously. However, this patch-wise inference introduces visible seams at the patch boundaries. To mitigate this artifact, we instead perform inference at full resolution ( $3 \times 1060 \times 1900$ ), using ten consecutive frames at a time.

## F. Discussion on the Responsible Use of the Dataset

This section discusses potential negative societal impacts, corresponding user guidelines, and our responsibilities.

### F.1. Potential Negative Societal Impacts

The UDC-VIT dataset contains facial images and motions of 22 research participants, which raises concerns about potential misuse, such as in deepfake applications. This technology can generate convincingly altered videos, posing risks to individual privacy and societal trust. Deepfakes

Table D.1. Prescribed instructions from IRB-approved participant information sheet. Out of thirty shots per person, videos displaying issues such as being out-of-focus are eliminated from the dataset.

---

#### Q. What procedures will be followed if the participants take part in the study?

---

A. The participants will undergo 30-shot photography using the UDC and regular digital cameras with the following motions:

- 5-second shots of body-swaying  $\times 9$  shots (6 indoors / 3 outdoors)
- 5-second shots of waving hands  $\times 9$  shots (6 indoors / 3 outdoors)
- 5-second shots of giving a thumbs-up  $\times 9$  shots (6 indoors / 3 outdoors)
- 5-second shots of walking indoors/outdoors  $\times 3$  shots

Since the UDC is located beneath the display and operates in low-light environments, it is necessary to capture data in various locations (indoors and outdoors) and lighting conditions (bright and dark) to account for the diverse degradation patterns of the UDC. Additionally, recognizing individuals from different angles is crucial for tasks such as face recognition, especially in financial authentication. This requires deep-learning models capable of restoring a subject’s appearance from multiple perspectives (e.g., front, left, and right), and thus, a dataset with multi-angle captures is essential. The recorded videos will be publicly released as a dataset to support UDC research.

---

#### Q. How long will the study participation last?

---

A. The study will take approximately 30 minutes. While the actual recording will take 2 minutes and 30 seconds (5 seconds  $\times 30$  shots), additional time will be needed for:

- Adjusting the subject’s shooting angle (5 minutes)
- Transitioning between locations (5 minutes)
- Verifying alignment accuracy (5 minutes)
- Conducting necessary adjustments (10 minutes)

---

#### Q. Will compensation be provided for participating in this study?

---

A. As a token of appreciation, participants who complete all 30 video captures will receive a \$40 Starbucks gift card. However, participants who withdraw before completing all 30 shots or request video deletion will not be compensated. If a participant requests video deletion after receiving compensation, they must return the full amount.

---

may infringe upon personal integrity and privacy, potentially leading to social unrest and confusion. Given these potential negative societal impacts, careful consideration is required when using the dataset.

## F.2. User Guidelines


Users of the UDC-VIT dataset are expected to adhere to the following guidelines:

- **Responsible use.** Users must utilize the dataset ethically and responsibly, ensuring that its use does not cause societal harm or infringe on individual privacy.
- **Compliance with legal and ethical standards.** Users must comply with all applicable legal and ethical standards. This includes obtaining approval from an Institutional Review Board (IRB) by the regulations of the respective country and adhering to any restrictions or conditions imposed by the IRB or other regulatory bodies. Any legal violations will be the sole responsibility of the user.
- **Restricted usage.** The UDC-VIT dataset must not be used for harmful applications, such as deep fake technologies or the creation of misinformation and manipulation. Furthermore, the 22 participants agreed to a research scope defined during our IRB review, which is limited to acquiring UDC video data and developing restoration models. Therefore, the dataset must be used exclusively for UDC-related research purposes.

## F.3. Our Responsibility

As the administrators of the UDC-VIT dataset, we acknowledge our responsibility to:

- **Protect participant privacy.** Our primary priority is to maintain the privacy and confidentiality of our research participants. Although the participants provided consent for the public use of their facial appearances and movements within the dataset, we are committed to guiding users toward appropriate research practices and to making efforts to safeguard any additional personal information.
- **Facilitate ethical use.** We provide comprehensive documentation and ethical usage guidelines through the *Datasheets for Datasets* and our GitHub repository at <https://github.com/mcrl/UDC-VIT>. The dataset can be accessed via an automatic email system after users complete the application form on our research group's website. This process also informs users of potential risks and ethical considerations associated with the dataset.
- **Respond to concerns.** We are committed to the responsible stewardship of the UDC-VIT dataset and will respond promptly to any concerns or complaints that may arise. We value user feedback and are prepared to take appropriate actions—such as data correction or updates—to prevent or mitigate harm, should any misuse of the dataset be reported through our research group's website (see Fig. D.3).

 **THUNDER Research Group**

AboutResearch ~People ~Publications ~Software and SystemsContact

## UDC-VIT *Under Display Camera's Videos by Thunder Research Group*

### Overview

Even though an Under-Display Camera (UDC) is an advanced imaging system, the display panel significantly degrades captured images or videos, introducing low transmittance, blur, noise, and flare issues. Tackling such issues is challenging because of the complex degradation of UDCs, including diverse flare patterns. However, no dataset contains videos of real-world UDC degradation. In this paper, we propose a real-world UDC video dataset called UDC-VIT. Unlike existing datasets, UDC-VIT exclusively includes human motions for facial recognition. We propose a video-capturing system to acquire clean and UDC-degraded videos of the same scene simultaneously. Then, we align a pair of captured videos frame by frame, using discrete Fourier transform (DFT). We compare UDC-VIT with six representative UDC still image datasets and two existing UDC video datasets. Using six deep-learning models, we compare UDC-VIT and an existing synthetic UDC video dataset. The results indicate the ineffectiveness of models trained on earlier synthetic UDC video datasets, as they do not reflect the actual characteristics of UDC-degraded videos. We also demonstrate the importance of effective UDC restoration by evaluating face recognition accuracy concerning PSNR, SSIM, and LPIPS scores. UDC-VIT is available at our official GitHub repository.

### Download

If you would like to download the UDC-VIT dataset, kindly complete the provided form. Please note that the UDC-VIT dataset is intended solely for UDC research purposes and can only be utilized by researchers with valid IRB approval. You must comply with legal regulations governing dataset usage in both the Republic of Korea and your nationality, obtaining IRB clearance accordingly. Additionally, it's essential to understand that any misuse or unauthorized distribution of the dataset beyond specified guidelines and the license will result in legal repercussions, for which you are solely responsible. By proceeding, you agree to these terms.

Name	Organization	E-mail	Download
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="button" value="Download"/>

### Contact and Bug Report

E-mail: [udcvit@aces.snu.ac.kr](mailto:udcvit@aces.snu.ac.kr)

Figure D.3. Our research group's official website providing detailed information and a download link for the UDC-VIT dataset.



## References

- [1] Kyusu Ahn, Byeonghyun Ko, HyunGyu Lee, Chanwoo Park, and Jaemin Lee. Udc-sit: A real-world dataset for under-display cameras. *Advances in Neural Information Processing Systems*, 36, 2024. 3, 4, 5, 8, 10, 11
- [2] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. *Advances in Neural Information Processing Systems*, 35:3926–3937, 2022. 8
- [3] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14483–14492, 2020. 3
- [4] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 662–671, 2021. 4, 5, 6, 8, 12
- [5] Raspberry Pi Foundation. Raspberry pi documentation, 2023. Available at [https://www.raspberrypi.com/documentation/computers/camera\\_software.html#building-rpicam-apps-without-building-libcamera](https://www.raspberrypi.com/documentation/computers/camera_software.html#building-rpicam-apps-without-building-libcamera). 10
- [6] Yuanbiao Gou, Haiyu Zhao, Boyun Li, Xinyan Xiao, and Xi Peng. Test-time degradation adaptation for open-set image restoration. In *Proceedings of the 41st International Conference on Machine Learning*, pages 16167–16177. PMLR, 2024. 3
- [7] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3482–3492, 2020. 3
- [8] Chengxu Liu, Xuan Wang, Yuanting Fan, Shuai Li, and Xueming Qian. Decoupling degradations with recurrent network for video restoration in under-display camera. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3558–3566, 2024. 1, 2, 3, 4, 6, 7, 8, 12
- [9] Xina Liu, Jinfan Hu, Xiangyu Chen, and Chao Dong. Udc-unet: Under-display camera image restoration via u-shape dynamic network. In *European Conference on Computer Vision*, pages 113–129. Springer, 2022. 4, 5, 6, 8, 12
- [10] Nvidia. Nvidia’s data loading library (dali) library, 2018. Available at <https://docs.nvidia.com/deeplearning/dali/user-guide/docs/index.html>. 13
- [11] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 267–276, 2002. 12
- [12] Samsung Electronics Co., Ltd. Samsung galaxy z fold 3, 2021. Available at <https://www.samsung.com/sg/business/smartphones/galaxy-z/galaxy-z-fold3-f926-5g-sm-f926bzgdxsp/>. 3
- [13] Samsung Electronics Co., Ltd. Samsung galaxy z fold 5, 2023. Available at <https://www.samsung.com/ae/smartphones/galaxy-z/galaxy-z-fold5-icy-blue-512gb-sm-f946blbgmea/>. 3
- [14] Matias Tassano, Julie Delon, and Thomas Veit. Fastdvdnet: Towards real-time deep video denoising without flow estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1354–1363, 2020. 8, 12
- [15] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 8, 12
- [16] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 191–207. Springer, 2020. 8, 12
- [17] ZTE Corporation. Zte axon 20, 2020. Available at <https://global.ztedevices.com/products/zte-axon-20-5g/>. 3