# Panel-Specific Degradation Representation for Raw Under-Display Camera Image Restoration

Youngjin Oh[1], Keuntek Lee[1], Jooyoung Lee[3], Dae-Hyun Lee[3], and Nam Ik Cho[1,2]

[1] Department of ECE & INMC, Seoul National University, Seoul, Korea
{yjymoh0211,leekt000,nicho}@snu.ac.kr
[2] IPAI, Seoul National University, Seoul, Korea
[3] SK hynix Inc., Korea
{jooyoung2.lee,daehyun1.lee}@sk.com

**Abstract.** Under-display camera (UDC) image restoration aims to restore images distorted by the OLED display panel covering the frontal camera on a smartphone. Previous deep learning-based UDC restoration methods focused on restoring the image within the RGB domain with the collection of real or synthetic RGB datasets. However, UDC images in these datasets exhibit domain differences from real commercial smartphone UDC images while inherently constraining the problem and solution within the RGB domain. To address this issue, we collect well-aligned sensor-level real UDC images using panels from two commercial smartphones equipped with UDC. We also propose a new UDC restoration method to exploit the disparities between degradations caused by different panels, considering that UDC degradations are specific to the type of OLED panel. For this purpose, we train an encoder with an unsupervised learning scheme using triplet loss that aims to extract the inherent degradations caused by different panels from degraded UDC images as implicit representations. The learned panel-specific degradation representations are then provided as priors to our restoration network based on an efficient Transformer network. Extensive experiments show that our proposed method achieves state-of-the-art performance on our real raw image dataset and generalizes well to previous datasets. Our dataset and code is available at https://github.com/OBAKSA/DREUDC.

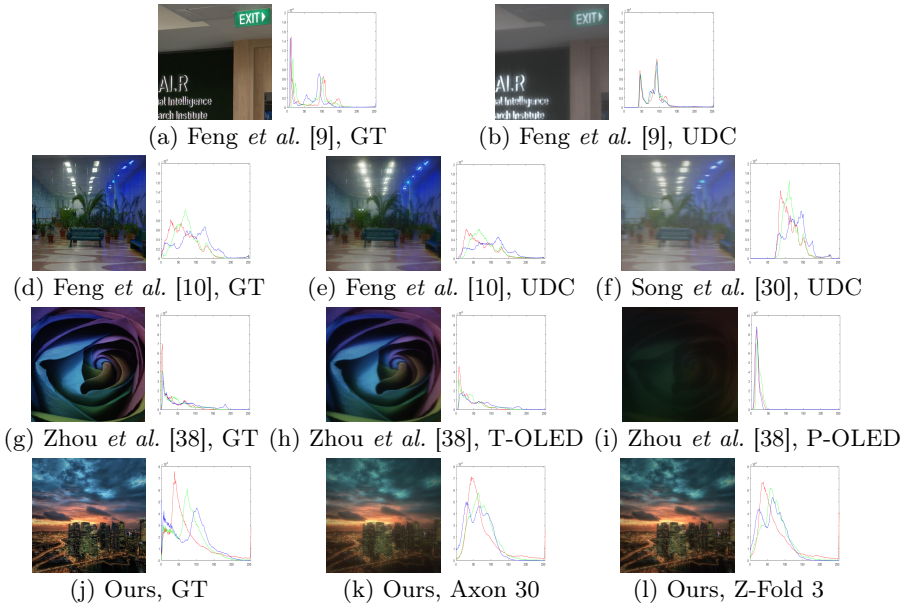
**Keywords:** Under-Display Camera · Image Restoration · Representation Learning

## 1 Introduction

To hide the camera hole in the frontal side of smartphones, under-display camera (UDC) systems have been developed, enabled by recent advancements in imaging and display panel technologies. With the UDC technology, smartphone users can enjoy a full-screen display without any camera holes or notches on the front side of their device. However, since the camera is hidden under the panel, the light that reaches the camera is altered by the complex layers of the panel, causing

**Table 1:** Comparison of previous datasets [9, 10, 30, 38] and our commercial dataset.

| Dataset | Zhou *et al.* [38] | Feng *et al.* [10] | Feng *et al.* [9] | Song *et al.* [30] | Ours |
|---|---|---|---|---|---|
| Format | RGB | RGB | RGB | RGB | RGB/RAW |
| UDC Modeling | Real (prototype) | Synthetic | Real | Synthetic | Real |
| Aligned GT | ✓ | ✓ | ✗ | ✓ | ✓ |
| Multiple Display | ✓ | ✗ | ✗ | ✗ | ✓ |



(a) Feng *et al.* [9], GT (b) Feng *et al.* [9], UDC

(d) Feng *et al.* [10], GT (e) Feng *et al.* [10], UDC (f) Song *et al.* [30], UDC

(g) Zhou *et al.* [38], GT (h) Zhou *et al.* [38], T-OLED (i) Zhou *et al.* [38], P-OLED

(j) Ours, GT (k) Ours, Axon 30 (l) Ours, Z-Fold 3

**Fig. 1:** Samples and their RGB channel histograms from previous UDC datasets [9, 10, 30, 38] and our commercial dataset.

severe degradation in the resulting image. Degradations include blur, diffraction, noise, low-lightness, color shift, haziness, *etc.*, and their spatially-variant nature makes the UDC image restoration very challenging [10, 17, 38].

Most UDC image restoration methods used deep neural networks [9–11, 15, 17, 23, 24, 30, 31, 37, 38], providing promising results on public benchmark datasets [9, 10, 30, 38] gathered to facilitate research on UDC image restoration. However, these existing datasets either use prototype display panels to inadequately represent the degradations observed in real UDC images caused by commercial panels [38], lack realistic degradations due to incomplete UDC modeling (*e.g.*, noise and color shift) in the synthesis pipeline [10,30], or mishandles aligning image pairs with occlusion and parallax due to the mechanism of the capturing imaging system [9]. In addition, the potential for enhanced image restoration performance through sensor-level restoration [1, 19, 25, 35] cannot be exploited, because previous datasets predominantly restrict experiments to RGB domain as raw data is unreleased. Tab. 1 shows comparison of the datasets.

In this work, we propose a new UDC dataset that is collected in the raw domain. The proposed dataset contains pairs of real UDC images and clean images that are well-aligned using a monitor-based imaging system with two different panels from commercial UDC smartphones, ZTE Axon 30 5G and Samsung Galaxy Z-Fold 3. In our experiments, we find that training with raw data enhances the performance of UDC image restoration. This raw-domain approach with real data also has the advantage of enabling the use of end-to-end neural ISPs that are being installed in recent smartphones [26], allowing a practical implementation on devices. Additionally, as shown in Fig. 1, real UDC data captured with commercial smartphone panels display noticeable distinct degradations visually and statistically when compared to previous datasets [10,30,38].

Furthermore, we develop a two-stage restoration method for UDC images. First, we introduce a framework for learning the implicit representation of degradations in the UDC images. We notice that degradation in a patch of a UDC image is similar to that of other patches of the same UDC image while being distinct from degradations of the same scene taken with a different panel. With this observation, we develop a novel unsupervised learning scheme involving triplet loss [28] to train our encoder. The learning scheme aims to leverage the information of the data collected with different types of panels, ultimately condensing the degradation information into an implicit representation vector that reflects the panel-specific degradation moderately. We also propose an efficient Transformer-based restoration network that incorporates the learned panel-specific degradation representations as a prior by generating channel-wise calibration coefficients for attention. We discover that our strategy yields improved results for UDC image restoration. To the best of our knowledge, this is the first case in UDC image restoration where a dataset collected with one panel is utilized to enhance the performance on a dataset collected with another panel.

To summarize, our contributions are as follows:

- We introduce a new real UDC dataset, which is collected with commercial smartphone panels. Our dataset is provided in raw format, which can improve the performance of the UDC restoration task.
- We propose an unsupervised learning approach that allows for implicit representations of degradations caused by various display panels in the embedding space. The panel-specific degradation representations are leveraged as priors in the restoration network, which is an efficient Transformer and brings notable improvement.
- Extensive experiments demonstrate that our restoration network embedded with learned representations achieves state-of-the-art performance on the proposed dataset and is consistently applicable to previous datasets.

## 2 Related Works

### 2.1 UDC Image Restoration

Since UDC image restoration is a relatively new area, research is still ongoing to provide users with a satisfying experience when UDC is applied in real situa-

tions. Research on UDC image restoration was sparked by the pioneering work of [38]. They defined a UDC imaging system and released the first UDC dataset for researchers to experiment with. Synthetic datasets were also made available by simulating the camera pipeline using a pre-calculated point spread function (PSF), which adequately represents the diffraction [10], with further improvements on scattering effects in [30] to account for the haziness of UDC images. More recently, a real UDC dataset generated by pseudo-aligning geometrically misaligned image pairs for a single UDC panel [9] was presented. Although the proposed datasets have some drawbacks, several works [9–11,15,17,23,24,30,31, 37,38] have investigated UDC image restoration using deep neural networks and have shown promising results.

Among these deep-learning-based methods, several of them use priors to improve performance. For instance, in [10], PCA is employed on a PSF of the UDC display panel to stretch it as a 'kernel code' and condition the network with it as a prior. In [30], the network is divided into a scattering branch and an image branch, and the former is trained to estimate the scattering parameters, which is prior information that can be used. However, these methods require precise priors; otherwise, the performance is degraded. Our approach differs from previous methods because we do not need strict parameters or ground truth settings. Instead, we aim to learn an acceptable representation of the degradation induced by panel-specific properties in the embedding space rather than a rigorous estimation of the degradation. We only need images taken with different panels because our method is unsupervised, and learn the implicit priors by leveraging the discrepancies of degradation caused by different panels.

### 2.2 Image Restoration with Raw Data

Most of the previous research on image restoration relies on RGB images as training data as they are easier to collect. However, some recent studies have shown that using raw sensor data can lead to improved performance. Specifically, [1, 25] are the works on image denoising where they collected DND and SIDD datasets that are available as raw and RGB images, where denoising with raw images usually yields higher PSNR. CycleISP [35] demonstrated that noise is more complex in RGB images and showed that noise modeling is more realistic in the sensor domain. This is because compared to the noise samples in the sensor pixels, which are nearly uncorrelated to each other, the noise samples become highly correlated in RGB images as they pass through the camera ISP pipeline, which involves processes like demosaicing, gamma correction, and white balancing [18,22]. DeepISP [29] proposed an end-to-end camera ISP modeled with deep neural networks, suggesting that a deep network-based camera ISP is promising. Furthermore, [4] solved the issue of low-light enhancement problem by operating in the sensor domain, and [19] presented a dataset for raw image deblurring and proved that operating on raw sensor data achieves better performance than RGB-based methods. These researches suggest that it is easier to restore raw data than RGB images with deep neural networks, as sensor data contains the linear representation of incoming light intensity of the scene. This ensures that

the scene's authenticity is maintained more faithfully in raw data than in heavily processed RGB pixel values that are nonlinear with respect to light intensity.

## 3 Proposed Method

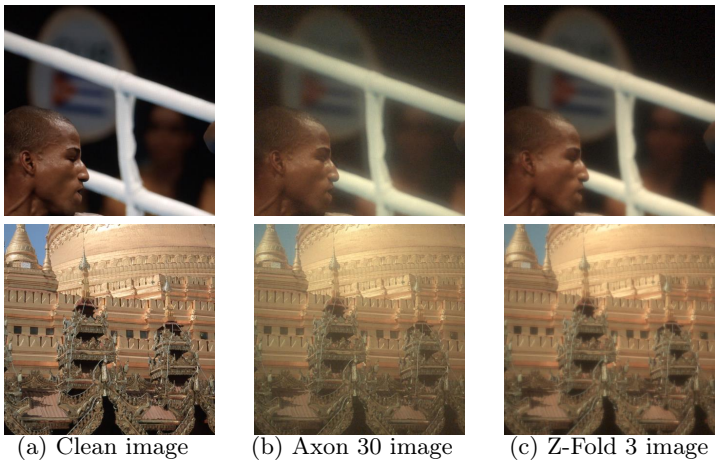
### 3.1 Sensor-level Real UDC Dataset

A raw dataset is necessary to achieve end-to-end image restoration using sensor images. However, the available UDC datasets are limited to the RGB domain. Moreover, public UDC datasets do not show the types of degradation that are present in images taken with commercially available smartphones equipped with UDC, limiting their assessment of practical scenarios. Therefore, we capture real raw UDC data by using OLED panels obtained from commercial UDC smartphones, a smartphone camera, and a monitor-based imaging system [38].

**UDC data acquisition process** Capturing a raw UDC image requires that we model the UDC imaging system as similarly as possible. Consequently, our data-capturing process involves rendering an image onto a high-resolution monitor, followed by capturing the displayed image using a smartphone camera sensor located closely behind a UDC panel. The panels used to capture the data are from two commercial UDC smartphones: ZTE Axon 30 5G and Samsung Galaxy Z-Fold 3. They introduce low-light/color shift, scattering, diffraction, and blur. As noise is usually generated from the sensor, we chose Samsung Galaxy S23 Ultra as our camera device to capture the authentic noise of smartphone camera sensors. Raw images are collected using the internal software of the smartphone that automatically saves the minimally processed raw data. The preference for monitor-displayed images [17, 38] over real natural scenes is driven by the need to control additional factors beyond the panels' presence. Specifically, we can enforce the target scene to be static as dynamic scenes are prone to motion blur, and therefore lead to misaligned data pairs. We can also eliminate interference from external ambient sources of light other than the monitor's emitted light.

The monitor, UDC panel, and camera sensor are aligned in sequence, and the principal axis of the camera is aligned perpendicular to the plane of the monitor and the panel. The camera is mounted on a sturdy tripod, and the panel is situated close to the sensor using a firm holding device. The displayed images are then captured remotely with/without the smartphone panels to ensure the alignment of clean/UDC image pairs during capture while keeping the camera settings the same. We provide more details in the Supplementary Material.

**Our UDC dataset** The dataset is captured using 400 images from the DIV2K dataset [2], which is comprised of natural scene images from various themes, encompassing diverse objects captured under different lighting conditions. This makes our dataset 1,200 images in total, as there are three different image sets (Axon 30, Z-Fold 3, ground truth) in raw format. The images are stored in 16 bits with a Bayer pattern of G-B-R-G. For the ground truth images, we capture



(a) Clean image (b) Axon 30 image (c) Z-Fold 3 image

**Fig. 2:** Cropped examples from our captured dataset. All images are visualized with the same procedure that includes normalization, demosaicing, channel gain adjustment, and gamma correction. Due to the difference in smartphone display panels, noise and haze are more apparent in Axon 30 UDC images, while the Z-Fold 3 UDC images show blurrier results. The diffraction induces flare artifacts when the image intensity becomes too strong that the sensor saturates.

bursts of a scene without the display panel and average them. To check for misalignments that can be induced by optical image stabilization in the smartphone hardware [1], we utilize registration and feature matching algorithms [3, 12].

Samples from our dataset are displayed in Fig. 2, with each UDC image taken with Axon 30 and Z-Fold 3 display panels exhibiting panel-specific degradations. As such, our dataset can be used to evaluate if a restoration method generally performs well in the restoration of various UDC systems. Also, our real commercial UDC data show different behaviors from previous datasets, as shown in Fig. 1. Specifically, low-light, color shift, blur, and noise are observed, and an object is seen repeated faintly near the main object due to diffraction. We also detect a phenomenon similar to haze due to the scattering effects of the panel [30]. Synthetic datasets that are generated using UDC modeling [10, 30] do not take noise into account. However, including noise in UDC images is vital as the level of noise captured with frontal smartphone camera sensors, which are relatively inferior to DSLR camera sensors [4, 26], is amplified when normalizing sensor data. This amplification is related to the attenuation caused by the panel which reduces light intensity. More samples are available in the Supplementary.

### 3.2 Proposed UDC Restoration Framework

In this section, we propose our restoration framework called Degradation Representation Embedded Transformer for UDC image restoration (DREUDC), which consists of a degradation representation encoder and a restoration network. We

first train the encoder with our proposed unsupervised scheme using triplet loss to implicitly learn the degradations of a display panel. Then, we utilize the representations generated from the encoder by integrating them in the restoration network as a prior. It is worth noting that our approach's novelty lies in how we take of advantage of UDC data from one display panel to restore a UDC image from another panel rather than in the restoration network's architecture itself.

**Motivation** Representation learning has been demonstrated to be effective in blind super-resolution, as demonstrated in [32]. Inspired by this, we explore the use of representation learning to enhance the restoration of UDC images by extracting useful prior information about the panels. Since each UDC system has a different panel, it yields panel-specific degradations. For example, the UDC images presented in Fig. 2, which are captured using two different panels of UDC smartphones, show differences in the degree of degradations. This implies that the degradations present in an image captured using one panel are similar to each other and that they differ significantly from those of the image with the same content but are captured using another panel. Using this information, our approach intends to utilize these differences and train an encoder to encode them in an unsupervised manner into the embedding space.

**Unsupervised representation encoder training** Our proposed unsupervised training scheme is visualized in Fig. 3. As mentioned above, the objective of training the panel-specific degradation representation encoder is to learn the unique properties of the panels and produce implicit representations of the degradations to be used as priors in UDC image restoration. To achieve this, the encoder should learn the degradations using attributes related to a specific panel, such as the level of noise, blur, and color shift. On the other hand, the encoder should be discouraged from learning degradations with features in an image that are panel-agnostic, such as textures and contents of a scene.

We believe that this condition is analogous to the situation of tasks where triplet loss [28] is exploited to learn favorable representations. Triplet loss is widely applied to various tasks of computer vision such as facial recognition [28], person re-identification [13,34], and domain adaptation [8] for its ability to learn useful embeddings by leveraging the distances between embeddings of similar instances and dissimilar instances. For instance, in person re-identification, networks are guided to identify among different people using features that are unique to a specific person, not with features that are shared among people, by learning suitable representations using triplet loss. In our case, the features unique to a specific person would correspond to properties unique to a specific panel, and the features that are shared among people would correspond to panel-agnostic features. Therefore, we opt to apply triplet loss in the training of our encoder to capture the panel-specific degradation representations.

Nevertheless, triplet loss is only effective when the network is provided with valid triplets [28, 33]. To collect meaningful triplets, we impose constraints on the training pairs that will be used to train our encoder. We choose a pair of

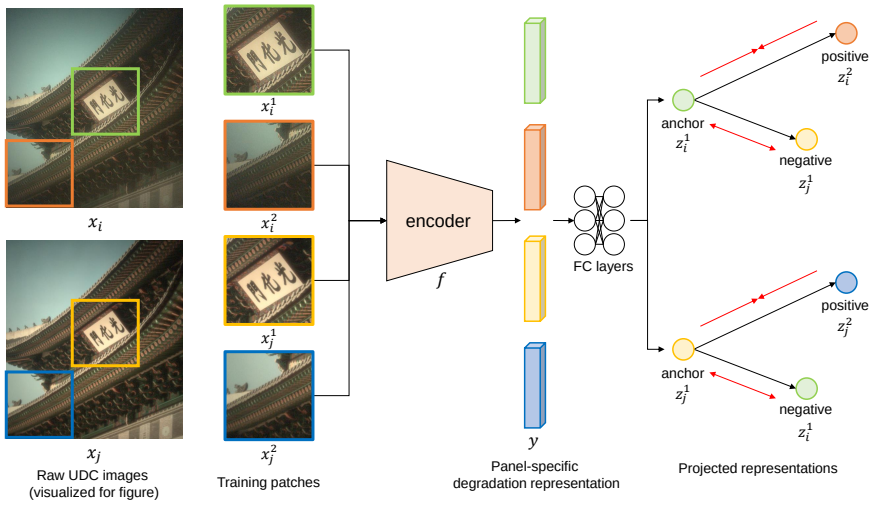



**Fig. 3:** Our proposed unsupervised panel-specific degradation representation learning scheme. By using triplet loss, representations of the same panel-specific degradations with different image content are pulled closer together, while representations of different panel-specific degradations with the same image content are pushed away.

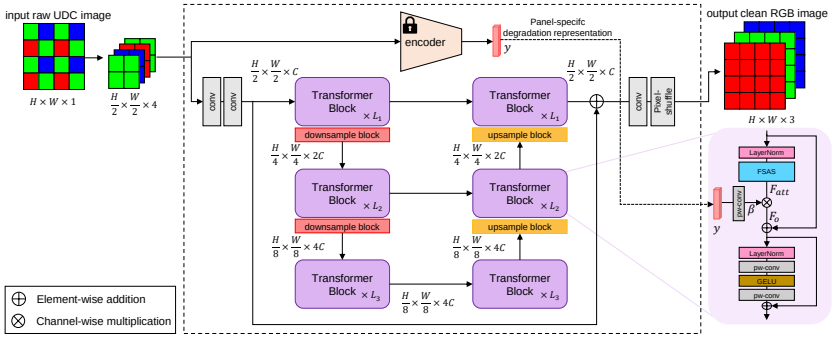
raw images $x_i$ and $x_j$ from our dataset that contains the same content but have been taken with different display panels. Then, we randomly crop two pairs of patches $(x_i^1, x_i^2)$ and $(x_j^1, x_j^2)$ from each image, where $x_i^1$ and $x_j^1$, and $x_i^2$ and $x_j^2$, respectively, are cropped from the same location of $x_i$ and $x_j$. We choose patch $x_i^1$ as the anchor, and the positive and negative samples are decided to be $x_i^2$ and $x_j^1$. The same process is repeated with $x_j^1$ as the anchor.

With this triplet selection, the encoder is guided to learn the intrinsic properties of the degradations caused by the display panel by pulling representation of images that show similar degradation while disregarding the features that are unrelated to degradations by pushing the representation of the same contents. The encoder $f$ is trained with triplet loss using the selected triplets to obtain the panel-specific degradation representation $y \in \mathbb{R}^{256}$. The representations are further projected to a projection $z$ with two fully connected layers, following [5,6]. The triplet loss for training the encoder is as follows:

$$L = max(||z_i^1 - z_i^2||_2^2 - ||z_i^1 - z_j^1||_2^2 + \alpha, 0) \\ + max(||z_j^1 - z_j^2||_2^2 - ||z_j^1 - z_i^1||_2^2 + \alpha, 0), \quad (1)$$

where $\alpha$ is the margin of the triplet loss. The margin is set as $\alpha = 1$. Following [32], the encoder is a simple network with six convolutional layers.

**Representation embedded restoration network** For our restoration network, we design it to be as simple and efficient as possible to emphasize the ef-




(a) Our proposed framework, DREUDC (b) DREblock

**Fig. 4:** Our proposed DREUDC. The framework consists of an encoder that produces an implicit representation of the panel-specific degradations of a UDC image and a restoration network that exploits the representations via the proposed DREblocks.

fectiveness of our contribution of how we leverage the difference in data collected with various panels to enhance the restoration performance rather than focusing on the impact of complex restoration network architectures. Therefore, we adopt a U-shaped [27] Transformer with Frequency domain-based Self-Attention Solver (FSAS) [16] as our restoration architecture. In previous works on UDC image restoration [9, 10, 17, 31], researchers have highlighted the importance of a wide receptive field. As FSAS utilizes the convolutional theorem of Fourier transform to model global contexts in the frequency domain efficiently [16], we choose it as a component of the building block for our restoration network.

Our restoration network is based on stacking repeated Transformer blocks, which we name DREblock (Degradation Representation Embedded Transformer block, Fig. 4(b)). The proposed DREblock efficiently integrates the degradation representation $y$ of the input UDC image as a condition of our restoration network by calibrating the features in the block in a channel-wise manner, *i.e.*,

$$F_o = F_{att} \otimes \beta, \tag{2}$$

where $F_{att}$ is the feature maps of the estimated attention of FSAS, $F_o$ is the feature after the conditioning, $\otimes$ denotes channel-wise multiplication, and $\beta$ is the channel conditioning coefficients that are produced by reshaping the panel-specific degradation representation $y$ using a single layer of $1\times1$ convolution. The features $F_o$ are then processed by the point-wise feed-forward network.

The entire framework DREUDC is illustrated in Fig. 4. In training, the encoder is first trained with Eq. (1). Then, the encoder is frozen while the restoration network is trained with $\mathcal{L}_1$ loss. During inference, the encoder takes the input raw UDC image and outputs a panel-specific degradation representation vector that contains prior information on UDC degradation. The representation is seamlessly embedded into the restoration network through DREblocks, finally producing a clean restored RGB image.

# 4 Experiment and Analysis

## 4.1 Datasets

We evaluate the restoration performance using our captured data with display panels of ZTE Axon 30 5G and Samsung Galaxy Z-Fold 3. The raw images are normalized to have values within the range of [0,1]. To train and evaluate RGB-based methods, we use a simplified ISP algorithm written with MATLAB to visualize the normalized raw images to RGB images. The algorithm contains demosaicing, channel-wise gain control, and gamma correction in sequence. Note that the visualized images are not sRGB images, as actual camera ISPs in practice involve additional steps to render the image in sRGB domain. Please check the Supplementary for more discussion on the usage of simplified ISP. For both display panels, we split the dataset into 300 images for training and 100 images for testing. The training set is cropped into patches of 512×512. We also use previous UDC datasets [38], which consists of 300 images of both T-OLED and P-OLED scenes, to verify that our method is applicable to other real datasets.

## 4.2 Training and Implementation Details

We use NVIDIA RTX 3090 GPUs in our experiments. To train the encoder, we use Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) for $1 \times 10^5$ iterations with a batch size of 64, and the loss is given as Eq. (1). The restoration networks are trained with $\mathcal{L}_1$ loss to minimize the distance between the restored image of the network and the ground truth. The initial number of channels $C$ of our restoration network is 32, and the number of blocks in each layer $L_1$, $L_2$, and $L_3$ are [5,5,6]. We use Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) for $2 \times 10^5$ iterations with a batch size of 2. For both networks, the initial learning rate is set as $2 \times 10^{-4}$ and gradually decayed to $1 \times 10^{-6}$ with cosine annealing [20]. This setting is applied to all experiments on our dataset, and every previous method reported in this paper is also reproduced by training on our dataset with the same setting for fairness.

## 4.3 Raw UDC Image Restoration

First, we experiment to validate that using raw images can improve the UDC image restoration process. For this purpose, we employ a baseline neural network called DE-UNET [38] and use our captured data. We compare the performance of the network under three different settings: RAW-to-RAW, RAW-to-RGB, and RGB-to-RGB. We evaluate the performance of the network using the PSNR and SSIM metrics. To assess the performance of the RAW-to-RAW setting, we convert the output restored raw images to RGB using the same ISP process mentioned in Sec. 4.1 and measure them in the RGB domain. The results are presented in Tab. 2, showing that using raw sensor images as input (RAW-to-RAW, RAW-to-RGB) yields superior results compared to only using RGB images. Among the two settings that use raw images as input, the latter shows better performance overall, so we choose the method of converting and restoring

**Table 2:** Comparison of different input-to-output settings of DE-UNET [38] on our captured datasets. The best results are bold-faced.

| Method | Axon 30 (raw) | | Z-Fold 3 (raw) | | Axon 30 (rgb) | | Z-Fold 3 (rgb) | |
|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| DE-UNET(RGB-to-RGB) | - | - | - | - | 23.33 | 0.9109 | 26.90 | 0.9390 |
| DE-UNET(RAW-to-RAW) | 27.36 | 0.9335 | 29.66 | 0.9637 | **24.21** | 0.9170 | 27.18 | 0.9388 |
| DE-UNET(RAW-to-RGB) | - | - | - | - | 24.20 | **0.9196** | **27.29** | **0.9419** |

**Table 3:** Results on our captured datasets for ZTE Axon 30 5G and Samsung Galaxy Z-Fold 3 smartphone display panels on various methods of UDC image restoration. The best results are bold-faced.

| Input-to-output | Method | Params | MACs | Axon 30 (rgb) | | Z-Fold 3 (rgb) | |
|---|---|---|---|---|---|---|---|
| | | | | PSNR | SSIM | PSNR | SSIM |
| RGB-to-RGB | DE-UNET [38] | 8.93M | 273.84G | 23.33 | 0.9109 | 26.90 | 0.9390 |
| | DAGF [31] | 1.09M | 73.84G | 24.49 | 0.9150 | 28.42 | 0.9404 |
| | DISCNET [10] | 3.80M | 596.95G | 23.11 | 0.9063 | 26.67 | 0.9361 |
| | BNUDC [15] | 4.57M | 993.38G | 25.85 | 0.9283 | 28.54 | 0.9458 |
| | DWFormer [36] | 1.45M | 216.77G | 21.52 | 0.8923 | 26.36 | 0.9338 |
| | SRUDC [30] | 6.70M | 371.32G | 22.67 | 0.8998 | 27.34 | 0.9336 |
| | PPM-UNet [9] | 4.23M | 409.29G | 18.21 | 0.8502 | 23.85 | 0.9188 |
| | AWNet [7] | 47.00M | 1.53T | 20.27 | 0.8839 | 25.60 | 0.9316 |
| RAW-to-RGB | DE-UNET [38] | 8.94M | 69.29G | 24.20 | 0.9196 | 27.29 | 0.9419 |
| | PyNet [14] | 47.56M | 1.79T | 22.58 | 0.9069 | 27.35 | 0.9377 |
| | AWNet [7] | 49.07M | 392.25G | 20.98 | 0.8885 | 27.25 | 0.9369 |
| | DREUDC | 6.44M | 200.07G | **26.66** | **0.9368** | **29.04** | **0.9511** |

raw UDC images to a clean RGB image (RAW-to-RGB). This is also the setting of most neural ISPs currently being used in smartphones [26] since a successful network is able to model the image processing pipeline as well.

### 4.4 Evaluation on Our Dataset

To provide experimental results on our dataset and demonstrate the performance of DREUDC on UDC image restoration, we train numerous previous methods using our dataset for comparison. The quality of the restoration performance is evaluated with PSNR and SSIM as performance metrics. Additionally, we assess the efficiency of the methods by comparing the number of parameters and MACs. To calculate the MACs, we use a dummy image with a resolution of $1024 \times 1024$.

Tab. 3 shows experimental results for raw and RGB image restoration using our dataset. Because there are no neural networks on UDC image restoration that take raw sensor images as input except [38] as previous datasets were provided only in the RGB domain, we choose [7] and [14] for further comparison, which are RAW-to-RGB image conversion models. We observe that DREUDC achieves state-of-the-art performance on both Axon 30 and Z-fold 3 datasets,


**Fig. 5:** Visual comparison of restoration examples for our dataset. Zoom in for a better comparison between the methods. The top two rows are results for Axon 30 images, and the bottom two rows are for Z-Fold 3 images. Methods that take raw input and produce RGB images are annotated with the asterisk *. Our method is capable of removing UDC degradations more clearly than previous UDC image restoration networks.

surpassing BNUDC [15] by 0.81dB and 0.50dB, demonstrating that our method can successfully restore various UDC degradations. With the exception of a lightweight model [31] and the baseline model using raw data input [38], our method is the most efficient when it comes to computation while showing superior performance. Qualitative results are illustrated in Fig. 5, and more results including restoration of scenes that are not displayed on a monitor are available in the Supplementary Material.

### 4.5 Analysis on Panel-Specific Degradation Representation

In this subsection, we analyze the effectiveness of DREUDC and panel-specific degradation representations learned with our proposed method. The strength of DREUDC is reported in Tab. 4 with an ablation study. We observe that our strategy of embedding an implicit representation of degraded UDC images is indeed effective by comparing (a) and (c) of Tab. 4, where Model 1 is our framework with only the restoration network without channel calibration and representation learning, while DREUDC is our entire framework trained with the protocol explained in Sec. 3.2.

However, with only this comparison, we cannot rule out the possibility that the rise in performance is solely because of increased parameters. Therefore, we further compare (a) and (c) with (b), where Model 2 is our framework trained in

**Table 4:** Results on ablation of our method.

| Method | Channel calibration | Representation learning | Params | MACs | Axon 30 | | Z-Fold 3 | |
|---|---|---|---|---|---|---|---|---|
| | | | | | PNSR | SSIM | PNSR | SSIM |
| (a) Model 1 | ✗ | ✗ | 4.52M | 160.57G | 26.09 | 0.9338 | 28.66 | 0.9491 |
| (b) Model 2 | ✓ | ✗ | 6.44M | 200.07G | 26.38 | 0.9348 | 28.71 | 0.9499 |
| (c) DREUDC | ✓ | ✓ | 6.44M | 200.07G | 26.66 | 0.9368 | 29.04 | 0.9511 |

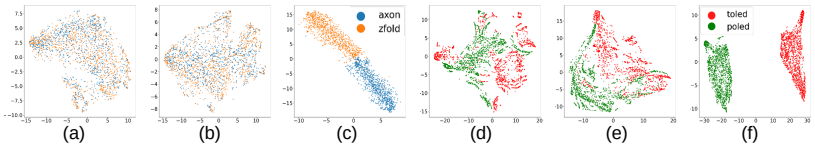



**Fig. 6:** T-SNE [21] visualization of representation generated from the encoder. (a) and (b) show the representations generated from encoders of Model 2 in Tab. 4 for Axon 30 and Z-fold 3 images, respectively, which are trained without representation learning. (d) and (e) show the representations generated from encoders of Model 2 in Tab. 5 for T-OLED and P-OLED images, respectively, which are trained without representation learning. (c) and (f) illustrate the representations that are generated with our unsupervised triplet learning on our dataset and Zhou's dataset [38].

an end-to-end manner by training the encoder and the restoration network jointly with $\mathcal{L}_1$ loss without our proposed representation learning. By comparing Model 1 and Model 2, we see that channel calibration with representations learned from a single dataset without triplet training increases the performance marginally. On the contrary, when given an appropriate representation learned from two datasets by leveraging the difference in degradations with our proposed method, we notice that the restoration model benefits substantially from it. Moreover, it is noteworthy that our proposed embedding scheme has a negligible effect on the computational increase (MACs) of the restoration network, as most of the increase is due to the encoder.

The validity of our panel-specific degradation representation learning scheme is visualized in Fig. 6. Without representation learning (Fig. 6(a) and (b)), the encoders are unable to distinguish different degradations. Conversely, with representation learning (Fig. 6(c)), the encoder identifies the degradation and discriminates them in clusters, offering advantageous priors for the restoration network.

This finding suggests an approach to train a restoration network by making good use of datasets other than the one gathered for a specific UDC system. Typically, a specific dataset is used to train a restoration network tailored for a particular UDC system. However, this dataset would be sub-optimal or ineffective when future UDC smartphones have different panels. Our method offers a way to utilize other datasets collected with different panels, mediating this issue.

**Table 5:** Results of our method and ablation on T-OLED/P-OLED datasets [38].

| Method | Channel calibration | Representation learning | Params | MACs | T-OLED | | P-OLED | |
|---|---|---|---|---|---|---|---|---|
| | | | | | PSNR | SSIM | PSNR | SSIM |
| DE-UNET [38] | - | - | 8.94M | 69.29G | 36.70 | 0.9745 | 28.58 | 0.9344 |
| Model 1 | ✗ | ✗ | 4.52M | 160.57G | 37.37 | 0.9774 | 30.85 | 0.9510 |
| Model 2 | ✓ | ✗ | 6.44M | 200.07G | 37.46 | 0.9777 | 31.01 | 0.9515 |
| DREUDC | ✓ | ✓ | 6.44M | 200.07G | 37.55 | 0.9778 | 31.17 | 0.9521 |

**Application to previous datasets** Our restoration method using panel-specific degradation representations is also applied to public UDC datasets to show that it generalizes well to other datasets. Our method requires a real dataset comprised of degraded images taken with more than a single panel from the same capturing environment, so we choose Zhou's P-OLED/T-OLED datasets [38]. DREUDC is designed to take raw images as input, so we make a pseudo-raw version of the dataset by mosaicing the RGB images. This pseudo-raw version of the RGB images that used to train the methods, including a new encoder, is created by reversing the in-camera pipeline mentioned in [38]. We only reversed the demosaicing step as it was the only mentioned process; hence we call it 'pseudo-raw' as the images have not been fully reversed to the original high-bit raw data. We choose DE-UNET [38] to compare with our results because it is a model that can be trained with RAW-to-RGB setting. Additional results on RGB image restoration are provided in the Supplementary Material.

Tab. 5 shows the results on T-OLED/P-OLED datasets. Our method outperforms DE-UNET on both panels with slightly fewer parameters. Most importantly, our strategy of representation embedding is also effective on [38] when we compare Model 1 to DREUDC, showing an increase of 0.18dB and 0.32dB in the restoration of T-OLED and P-OLED images. We also prove that the boost in performance is not only due to increased parameters but also because we enhance the restoration network with effective representations of the degradations learned using our training scheme by comparing Model 2 and DREUDC. Visualization of the representations are provided in Fig. 6(d), (e), and (f). Overall, the results imply that our method is also effective on other UDC datasets.

## 5 Conclusion

We have collected a new UDC dataset using commercial UDC-equipped smartphones and demonstrated that UDC image restoration using raw sensor images achieves better performance than using RGB images. We have also proposed a UDC image restoration framework that generates panel-specific degradation representations in the embedding space with a novel unsupervised representation learning scheme. Our restoration network efficiently accommodates the learned representations as a prior with our proposed Transformer blocks, improving the performance. Experiments show that our method achieves state-of-the-art performance on our dataset and is generalizable to previous datasets.

## Acknowledgment

This paper was result of the research project supported by SK hynix Inc., including the support of camera equipments and lab environment for collecting the dataset. It was also partially supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-01062), and in part by the BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University in 2024.

## References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1692–1700 (2018)
2. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 126–135 (2017)
3. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. pp. 404–417. Springer (2006)
4. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3291–3300 (2018)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. pp. 1597–1607. PMLR (2020)
6. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297 (2020)
7. Dai, L., Liu, X., Li, C., Chen, J.: Awnet: Attentive wavelet network for image isp. In: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. pp. 185–201. Springer (2020)
8. Deng, W., Zheng, L., Sun, Y., Jiao, J.: Rethinking triplet loss for domain adaptation. IEEE Transactions on Circuits and Systems for Video Technology **31**(1), 29–37 (2020)
9. Feng, R., Li, C., Chen, H., Li, S., Gu, J., Loy, C.C.: Generating aligned pseudo-supervision from non-aligned data for image restoration in under-display camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5013–5022 (2023)
10. Feng, R., Li, C., Chen, H., Li, S., Loy, C.C., Gu, J.: 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. pp. 662–671 (2021)
11. Feng, R., Li, C., Zhou, S., Sun, W., Zhu, Q., Jiang, J., Yang, Q., Loy, C.C., Gu, J., Zhu, Y., et al.: Mipi 2022 challenge on under-display camera image restoration: Methods and results. In: European Conference on Computer Vision. pp. 60–77. Springer (2022)
12. Guizar-Sicairos, M., Thurman, S.T., Fienup, J.R.: Efficient subpixel image registration algorithms. Optics letters **33**(2), 156–158 (2008)

13. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)
14. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera isp with a single deep learning model. arXiv preprint arXiv:2002.05509 (2020)
15. Koh, J., Lee, J., Yoon, S.: Bnudc: a two-branched deep neural network for restoring images from under-display cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1950–1959 (2022)
16. Kong, L., Dong, J., Ge, J., Li, M., Pan, J.: Efficient frequency domain-based transformers for high-quality image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5886–5895 (2023)
17. Kwon, K., Kang, E., Lee, S., Lee, S.J., Lee, H.E., Yoo, B., Han, J.J.: Controllable image restoration for under-display camera in smartphones. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2073–2082 (2021)
18. Lee, W., Son, S., Lee, K.M.: Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17725–17734 (2022)
19. Liang, C.H., Chen, Y.A., Liu, Y.C., Hsu, W.H.: Raw image deblurring. IEEE Transactions on Multimedia **24**, 61–72 (2020)
20. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
21. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
22. Nam, S., Hwang, Y., Matsushita, Y., Kim, S.J.: A holistic approach to cross-channel image noise modeling and its application to image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1683–1691 (2016)
23. Oh, Y., Park, G.Y., Cho, N.I.: Restoration of high-frequency components in under display camera images. In: 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). pp. 1040–1046. IEEE (2022)
24. Oh, Y., Park, G.Y., Chung, H., Cho, S., Cho, N.I.: Residual dilated u-net with spatially adaptive normalization for the restoration of under display camera images. In: 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). pp. 151–157. IEEE (2021)
25. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1586–1595 (2017)
26. Punnappurath, A., Abuolaim, A., Abdelhamed, A., Levinshtein, A., Brown, M.S.: Day-to-night image synthesis for training nighttime neural isps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10769–10778 (2022)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
28. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 815–823 (2015)

29. Schwartz, E., Giryes, R., Bronstein, A.M.: Deepisp: Toward learning an end-to-end image processing pipeline. IEEE Transactions on Image Processing **28**(2), 912–923 (2018)
30. Song, B., Chen, X., Xu, S., Zhou, J.: Under-display camera image restoration with scattering effect. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12580–12589 (2023)
31. Sundar, V., Hegde, S., Kothandaraman, D., Mitra, K.: Deep atrous guided filter for image restoration in under display cameras. In: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16. pp. 379–397. Springer (2020)
32. Wang, L., Wang, Y., Dong, X., Xu, Q., Yang, J., An, W., Guo, Y.: Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10581–10590 (2021)
33. Yu, B., Liu, T., Gong, M., Ding, C., Tao, D.: Correcting the triplet selection bias for triplet loss. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 71–87 (2018)
34. Yuan, Y., Chen, W., Yang, Y., Wang, Z.: In defense of the triplet loss again: Learning robust person re-identification with fast approximated triplet loss and label distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 354–355 (2020)
35. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Cycleisp: Real image restoration via improved data synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2696–2705 (2020)
36. Zhou, Y., Song, Y., Du, X.: Modular degradation simulation and restoration for under-display camera. In: Proceedings of the Asian Conference on Computer Vision. pp. 265–282 (2022)
37. Zhou, Y., Kwan, M., Tolentino, K., Emerton, N., Lim, S., Large, T., Fu, L., Pan, Z., Li, B., Yang, Q., et al.: Udc 2020 challenge on image restoration of under-display camera: Methods and results. In: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16. pp. 337–351. Springer (2020)
38. Zhou, Y., Ren, D., Emerton, N., Lim, S., Large, T.: Image restoration for under-display camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9179–9188 (2021)