

Under-Display Camera Image Restoration with Scattering Effect

Binbin Song¹, Xiangyu Chen^{1,2}, Shuning Xu¹, and Jiantao Zhou^{1†}

¹State Key Laboratory of Internet of Things for Smart City

Department of Computer and Information Science, University of Macau

²Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

{yby97426, jtzhou}@umac.mo, {chxy95, rebeccaxu0418}@gmail.com

Abstract

The under-display camera (UDC) provides consumers with a full-screen visual experience without any obstruction due to notches or punched holes. However, the semi-transparent nature of the display inevitably introduces the severe degradation into UDC images. In this work, we address the UDC image restoration problem with the specific consideration of the scattering effect caused by the display. We explicitly model the scattering effect by treating the display as a piece of homogeneous scattering medium. With the physical model of the scattering effect, we improve the image formation pipeline for the image synthesis to construct a realistic UDC dataset with ground truths. To suppress the scattering effect for the eventual UDC image recovery, a two-branch restoration network is designed. More specifically, the scattering branch leverages global modeling capabilities of the channel-wise self-attention to estimate parameters of the scattering effect from degraded images. While the image branch exploits the local representation advantage of CNN to recover clear scenes, implicitly guided by the scattering branch. Extensive experiments are conducted on both real-world and synthesized data, demonstrating the superiority of the proposed method over the state-of-the-art UDC restoration techniques. The source code and dataset are available at <https://github.com/NamecantbeNULL/SRUDC>.

1. Introduction

The adoption of under-display cameras (UDCs) has gained the popularity as a solution to avoid the inconvenience of notches and punched holes on the screen. This design meets the growing demand for full-screen hand-held devices and improves consumers' visual experience. However, the semi-transparent nature of the display covered on

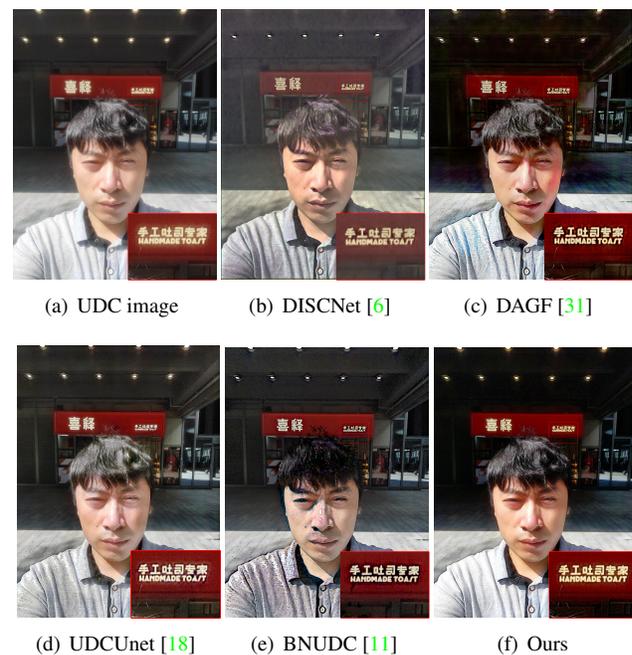


Figure 1. An example of the UDC image restoration on a real-scene image. (a) Degraded UDC image, (b-e) Outputs of state-of-the-art methods w/o considering the scattering effect, (f) Our result, capable of removing the haziness and contrast distortion, while restoring fine details. (Zoom in for a better view.)

the camera inevitably introduces various types of the degradation, e.g., contrast distortion, blurring, haziness, noise, and diffraction artifacts around light sources, into the captured UDC images. To mitigate these issues, rapidly evolving UDC image restoration methods [42, 6, 14, 11, 8, 21, 43, 12, 4, 31, 7, 5] in recent years are dedicated to recovering degraded images. Most of these methods focus on the degradation of high-frequency components, e.g., blurring, noise, and diffraction artifacts, while simply modeling the deterioration of low-frequency components as the pixel intensity attenuation. Whereas, the scattering effect of the display, which can result in the degradation such as con-

[†]Corresponding author.

trast distortion and haziness, has been neglected in previous UDC restoration methods. Such a negligence may significantly limit the restoration performance and generalization of existing methods on real-scene images.

Essentially, the scattering effect is a common phenomenon in UDC imaging systems [26]. When incident light enters the display, it travels through multiple stacked layers, e.g., the cathode, the substrate, etc., before reaching the camera sensor. During this process, collisions with small particles within these layers alter the propagation direction of some photons. Consequently, the total energy of the incident radiance is suppressed and part of the light is converted into scattered light, which generates contrast distortion and haziness. These types of degradation due to the scattering effect not only affect the visual quality of the captured UDC images, but also may cause the high-level vision systems to fail when UDC images are fed as the input. Surprisingly, as far as we know, the scattering effect has never been explicitly addressed in the UDC image restoration task. Due to the scattering effect, on one hand, there is a gap between the real-world and synthetic UDC images generated by the existing data synthesis pipeline. On the other hand, the restoration methods necessitate special designs for the scattering-induced degradation in UDC images. These two reasons motivate us to take into account the scattering effect for better restoring the captured UDC images.

In this work, we model the UDC scattering effect by decomposing the irradiance received by the camera into transmitted and scattered components and calculating them separately. Based on the proposed UDC scattering model, we enhance the existing image formation pipeline (IFP) to generate realistic UDC images with paired ground truths. We show that compared to the existing IFP, our enhanced method produces synthetic UDC images that are closer to real-captured ones. To restore UDC images with the scattering effect, we propose a two-branch deep network consisting of a scattering branch and an image branch. The scattering branch estimates the parameters of the scattering effect from degraded images, while the image branch recovers clear scenes guided by the scattering branch. In the scattering branch, we propose transposed self-attention blocks (TSABs) to leverage the global modeling capability [37, 32] of the channel-wise self-attention (CSA). In the image branch, we use Convolutional Neural Network (CNN) as the backbone to exploit its local representation advantage [29, 3]. To effectively combine the information from both branches, we devise a feature fusion module, which adaptively modulates the deep features with the global information from the scattering branch. Fig. 1 presents an example of UDC image restoration on a real-scene image.

In summary, our contributions are as follows:

- We propose a new UDC imaging model by explicitly considering the scattering effect, which causes unex-

pected haziness and contrast distortion. The image formation pipeline is then enhanced, generating more realistic UDC synthetic images.

- Guided by the proposed UDC imaging model with the scattering effect, we specially design a dual-branch deep framework, which leverages global modeling capabilities of CSA and the local representation advantage of CNN, to restore the UDC image.
- Extensive experiments demonstrate significant performance gains on both real-world and synthetic data. Compared with state-of-the-art methods, the visual quality of resulting images is dramatically improved.

2. Related Works

In this section, we first briefly introduce the works related to modeling the degradation process of UDC images. Then the recently developed UDC image restoration methods are reviewed.

2.1. UDC imaging modeling

Many previous works [42, 6, 11, 13, 14, 23] have investigated the degradation of the UDC imaging. Zhou *et al.*[42] first formulate the UDC imaging process with the classical convolution model as:

$$\mathbf{I} = \gamma(\mathbf{B} * \mathbf{k}) + \mathbf{n}, \quad (1)$$

where \mathbf{B} denotes the clean scene image, \mathbf{I} denotes the degraded image. Here, \mathbf{k} is the point spread function (PSF), \mathbf{n} is the zero-mean noise, γ is the intensity scaling factor, and $*$ represents the convolution operator. However, directly using this model in the limited dynamic range suppresses the diffraction artifacts around the saturated areas. To get closer to the real-world data, Feng *et al.*[6] propose to migrate the convolution into the RAW domain with high dynamic range and measure the actual PSF of UDC in a smartphone for the data synthesis. Kwon *et al.*[14] further consider the spatial variability of the PSF due to the incident angle of the light. To simulate complicated forms of the degradation with low spatial frequency, Koh *et al.* [11] exploit a 3D transformation to substitute the single scaling factor γ .

Although the aforementioned works have improved the UDC degradation model, the convolution model with the single γ cannot generate the widely-observed contrast distortion and haziness caused by the UDC scattering effect. This phenomenon motivates us to specifically model the light scattering process in the UDC imaging.

2.2. UDC restoration methods

The pioneering UDC image restoration method based on U-Net is proposed in [42], as a blind deconvolution solution. To deal with the different PSFs in saturated regions, Feng *et al.*[6] design a Dynamic Skip Connection Network

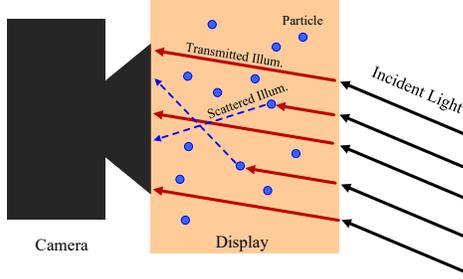


Figure 2. The schematic diagram of the UDC scattering model. The irradiance received by the camera can be decomposed into the transmitted illumination and scattered illumination.

(DISCNet) to estimate the latent clean image. Koh *et al.* [11] propose a branched network for UDC image restoration (BNUDC), which simultaneously removes the high-frequency noise and the low-frequency degradation. To save memory for the restoration of high-resolution images, a Deep Atrous Guided Filter (DAGF) network is utilized by Sundar *et al.* [31]. Luo *et al.* [21] discover distinct statistical features of UDC and ground-truth images in HSV space. To fit the estimated images adaptively in H and S channels, they propose a cascaded curve estimation network for the enhancement of UDC images.

While there have been some successful attempts on the UDC image restoration, existing methods have largely neglected the treatment of the scattering effect, which could severely affect the restoration performance on real-world UDC images.

3. Enhancing IFP with UDC Scattering Model

In this section, we first model the scattering phenomenon in UDC imaging systems by decomposing the irradiance received by the camera into transmitted and scattered components (Section 3.1). Subsequently, we propose an enhanced approach for the UDC image formation pipeline (IFP) utilizing the scattering model for synthesizing UDC images (Section 3.2). We demonstrate that our proposed method yields UDC images, which exhibit higher similarity to real-captured UDC images, compared to the existing UDC IFP.

3.1. Modeling UDC Scattering Effect

In this study, we assume the incident radiance striking the display comes from a single light source, so as to facilitate the modeling of the UDC scattering effect in a simplified context. By exploiting the radiance accumulation, our modeling approach can be easily extended to the circumstance with multiple sources. In reality, the incident light can have a variety of sources, including sunlight, ground-reflected light, scene-reflected light, etc. Also, when the light source is considerably distant from the UDC, the incident light rays can be assumed as parallel light beams [27].

When the incident parallel light travels through the dis-



Figure 3. The comparison of synthetic images generated by our enhanced IFP and Feng *et al.* [6]. Our enhanced IFP generates haziness and contrast distortion, which can also be observed in real-world captured UDC images. PSNR and SSIM are calculated by using the captured UDC images as the reference.

play, a fraction of photons collides with tiny particles inside the display, leading to the scattering phenomenon [26]. The unaffected photons continue to pass through the display and eventually reach the camera sensor with attenuated amplitudes. A schematic illustration of the UDC scattering phenomenon is presented in Figure 2. In our analysis, we treat the display in the UDC imaging system as a piece of homogeneous scattering medium with a constant scattering coefficient β . The UDC scattering effect can then be expressed as:

$$\mathbf{L}(\mathbf{x}) = \mathbf{L}_t(\mathbf{x}) + \mathbf{L}_s(\mathbf{x}), \quad (2)$$

where \mathbf{L} denotes the camera-captured light intensity, \mathbf{L}_t is the transmitted illumination component and \mathbf{L}_s is the scattered illumination component. Here, \mathbf{x} represents the pixel location. Inspired by the scattering model used in image de-haze [22, 15] and under-water image enhancement [10, 16], we can write \mathbf{L}_t and \mathbf{L}_s in an explicit form:

$$\begin{aligned} \mathbf{L}_t(\mathbf{x}) &= \mathbf{L}_b(\mathbf{x}) e^{-\beta d(\mathbf{x})}, \\ \mathbf{L}_s(\mathbf{x}) &= m(1 - e^{-\beta d(\mathbf{x})}), \end{aligned} \quad (3)$$

where $d(\mathbf{x})$ represents the routing path distance of the light ray within the display, and m is the proportionality constant related to the incident radiance \mathbf{L}_b (also known as the light intensity of the clean background image). For the light rays with the same incident angle, $d(\mathbf{x})$ approximates a fixed value for all the pixel positions.

As will be discussed below, this UDC scattering model will be used to enhance the UDC IFP.

3.2. Enhanced UDC IFP

Based on the proposed UDC scattering model, we now enhance the existing IFP to generate more realistic UDC images with paired ground truths for training the subsequent

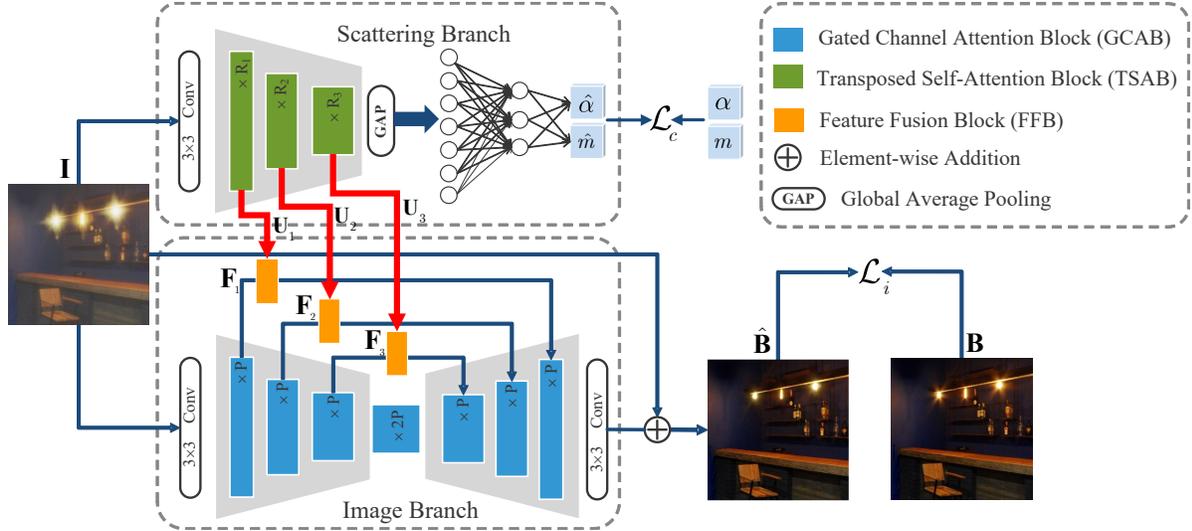


Figure 4. The overview of our proposed SRUDC for the UDC image restoration. The whole framework consists of two branches: the scattering branch and the image branch. The scattering branch estimates the parameters of the scattering effect from degraded images, while the image branch recovers clear scenes guided by the scattering branch.

restoration networks. Specifically, we integrate the diffraction blur related to the PSF \mathbf{k} and zero-mean noise \mathbf{n} into (2). Our enhanced UDC IFP can then be defined as:

$$\mathbf{I}(\mathbf{x}) = TC \left[(\mathbf{L}_t(\mathbf{x}) + \mathbf{L}_s(\mathbf{x})) * \mathbf{k} + \mathbf{n} \right], \quad (4)$$

where \mathbf{x} is the pixel position, \mathbf{I} is the degraded UDC image, and $TC(\cdot)$ denotes the tone mapping and clipping functions. Here, the purpose of introducing $TC(\cdot)$ is to convert the image intensities into the low-dynamic-range space after generating the anisotropic diffraction blur in the high-dynamic-range space. Substituting (3) into (4) yields:

$$\mathbf{I}(\mathbf{x}) = TC \left\{ [\mathbf{L}_b(\mathbf{x}) e^{-\beta d(\mathbf{x})} + m(1 - e^{-\beta d(\mathbf{x})})] * \mathbf{k} + \mathbf{n} \right\}. \quad (5)$$

where $\mathbf{L}_b(\mathbf{x})$ is the light intensity of the background \mathbf{B} . As $d(\mathbf{x})$ approximates a constant for all \mathbf{x} s and both \mathbf{k} and \mathbf{n} are assumed to be independent with \mathbf{x} , the above UDC scattering model can be seen as a global transformation. By introducing $\alpha = e^{-\beta d}$ and omitting \mathbf{x} , the enhanced IFP can be further simplified as:

$$\mathbf{I} = TC \left\{ [\alpha \mathbf{L}_b + m(1 - \alpha)] * \mathbf{k} + \mathbf{n} \right\}. \quad (6)$$

Regarding the parameter m , which is the proportionality constant related to the incident radiance \mathbf{L}_b , we set it as the mean of the gray-scale \mathbf{B} to maintain the average brightness. Such a setting is reasonable since the under-display photography generally could not influence the distribution of the value (V) channel in HSV space [21]. Also, α is a constant empirically sampled in the interval [0.6, 0.9].

To justify the rationality of our enhanced IFP, we compare it with the existing data synthesizing method [6] on

a small real dataset [21]. As demonstrated in Fig. 3, our generated UDC images exhibit a closer visual resemblance to the real-captured images; also with significantly higher PSNR and SSIM values. To further illustrate the superiority of our enhanced IFP, we conduct a statistical analysis of the entire synthetic dataset in the HSV color space. We find that the distributions of the hue (H) and saturation (S) channels in our synthetic dataset are similar to those of the real-captured dataset. The histogram representations of statistical properties can be found in the supplementary material.

4. Scattering Removal Network for UDC Image Restoration

To specifically address the problem of the scattering effect in the UDC image restoration, we propose a Scattering Removal network for the UDC image restoration (SRUDC). Based on our enhanced UDC IFP, we notice that accurate estimates of α and m are rather beneficial to the UDC image recovery. To this end, we purposely design a scattering branch in SRUDC that estimates these parameters to guide the restoration process of the background image in the image branch. The overview of the framework is shown in Fig. 4. Given the degraded UDC image \mathbf{I} , our target is to restore the clean background image \mathbf{B} . Our SRUDC framework consists of two branches, namely the scattering branch and the image branch. Concretely, the scattering branch estimates α and m from \mathbf{I} (Section 4.1), while the image branch restores \mathbf{B} with the guidance of the scattering branch (Section 4.2). To further enhance the feature representation of the image branch, we design a new feature fusion block (FFB) that effectively combines the global information from

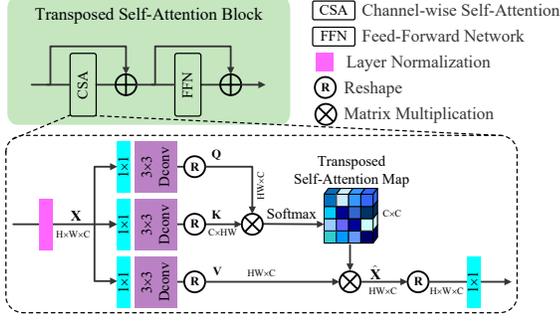


Figure 5. The structure of TSAB.

the scattering branch (Section 4.3).

4.1. Scattering Branch

The scattering branch is designed to estimate α and m , which numerically represent the scattering effect in the UDC imaging system. Recall from (6) that these two coefficients determine the global transformation of the clean image due to the UDC scattering effect. As a result, the accurate estimation of α and m requires the network to maintain the capability of modeling spatially invariant global information. To this end, inspired by the recent success of channel-wise self-attention [37, 35, 41], we develop the Transposed Self-Attention Block (TSAB), which captures long-range dependencies in the channel dimension. Furthermore, we employ global average pooling (GAP) to extract the useful global statistics of the features, and then calculate the conditional vector of $\hat{\alpha}$ and \hat{m} through a two-layer multi-layer perceptron (MLP).

The detailed structure of the scattering branch is demonstrated in the upper part of Fig. 4. Given a degraded UDC image \mathbf{I} , we apply a 3×3 convolution layer to extract the shallow features. The extracted features then pass through a three-level encoder built upon several TSABs to generate the set of deep features $\mathbf{U}_o (o = 1, 2, 3)$, where o is the level index. The numbers of TSABs in each level are denoted as $\{R_1, R_2, R_3\}$. We downscale features twice to gradually reduce the spatial size and increase the number of channels. Eventually, the estimated $\hat{\alpha}$ and \hat{m} are obtained by:

$$(\hat{\alpha}, \hat{m}) = W_2 \odot g(W_1 \odot GAP(\mathbf{U}_3)), \quad (7)$$

where $GAP(\cdot)$ means the global average pooling layer, W_1 and W_2 are parameters of the two-layer MLP, $g(\cdot)$ denotes a Sigmoid function, and \odot represents the Hadamard product. Also, the features \mathbf{U}_o s are further transformed to guide the reconstruction of the clean scenes in the image branch.

Regarding the TSAB, which is an essential building block in the scattering branch, its architecture is shown in Fig. 5. Given the input features, to capture the global information through long-distance dependencies in the channel dimension and significantly reduce the storage consumption, we compute CSA for the modulation of the feature

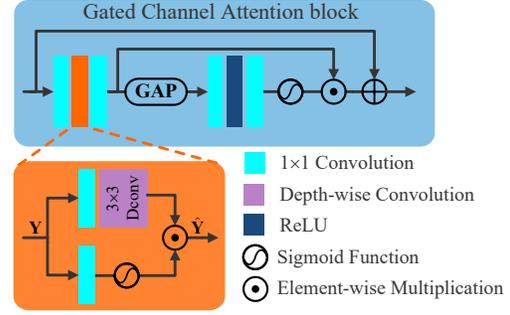


Figure 6. The structure of the proposed GCAB.

residue. From a layer normalized tensor $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, we use the combinations of 1×1 convolution and 3×3 depth-wise convolution followed by the reshaping to generate $\{\mathbf{Q}, \mathbf{K}, \mathbf{V}\} \in \mathbb{R}^{HW \times C}$. The CSA modulation is then defined as:

$$\hat{\mathbf{X}} = \text{softmax}(\mathbf{Q}\mathbf{K}^T)\mathbf{V}, \quad (8)$$

where $\hat{\mathbf{X}}$ means channel-wise modulated features. Then we reshape $\hat{\mathbf{X}}$ into a dimension $\mathbb{R}^{H \times W \times C}$ for further processing. After the modulation by the CSA, the feed-forward network is added in TSAB to introduce the nonlinear activation, further enhancing the capability of CSA.

4.2. Image Branch

With the guidance of \mathbf{U}_o s from the scattering branch, the image branch is proposed to generate the estimated background $\hat{\mathbf{B}}$. We leverage CNN as the backbone to exploit its local representation capability. As shown in the bottom part of Fig. 4, \mathbf{I} is first transformed into the feature space through a 3×3 convolution layer. Then, a symmetric U-shape architecture reconstructs the background scenes in the feature space. The encoder and the decoder consist of three levels, built based on our proposed Gated Channel Attention Block (GCAB). To assist the recovery process, the features from the encoder, denoted as $\mathbf{F}_e (e = 1, 2, 3)$, are modulated by \mathbf{U}_o and then concatenated with the decoder features via skip connections [25].

To better reconstruct the spatially variant UDC image content, we design GCAB as the basic building block of the encoder and the decoder in the image branch. The structure of GCAB is depicted in Fig. 6. GCAB is based on the widely used residual channel attention block [40, 38]. To enhance the spatial representation capability of the feature maps, we introduce the gating mechanism [1, 30], which adaptively activates the feature maps. As shown in the orange block in Fig. 6, given an input tensor \mathbf{Y} , the gating operator is defined as:

$$\text{Gating}(\mathbf{Y}) = W_{dw} W_{pw}^1 \mathbf{Y} \odot g(W_{pw}^2 \mathbf{Y}), \quad (9)$$

where W_{pw}^1 and W_{pw}^2 respectively represent the weight matrices of two 1×1 convolution layers, and W_{dw} is the parameter of the 3×3 depth-wise convolution. By integrating the gate mechanism and the channel attention, we can

adaptively modulate the feature in both the spatial and the channel dimensions for the restoration of UDC images.

4.3. Feature Fusion Block

To implicitly guide the estimation of $\hat{\mathbf{B}}$ in the decoder of the image branch, we recalibrate \mathbf{F}_e with the scattering effect information embedded in \mathbf{U}_o , via Feature Fusion Blocks (FFBs). Since \mathbf{F}_e and \mathbf{U}_o respectively encode the global and local information of the input, directly concatenating or element-wisely adding them may cause ambiguity. To solve this challenge, we devise the FFB to adaptively modulate \mathbf{F}_e . According to the formulation of the UDC scattering model (see (6)), we propose to utilize the affine transformation in FFB. Inspired by spatial feature transformation (SFT) [33, 2], we compute the coefficient and offset of the affine transformation from \mathbf{U}_o . The illustration of the design of FFB is put in the supplement. Given \mathbf{F}_e and \mathbf{U}_o with $e = o$, the fusion weights $\{\mathbf{v}, \mathbf{w}\}$ can be obtained through the combination of two 1×1 convolution layers and a leaky ReLU function. Then, the output feature \mathbf{F}_o can be generated by:

$$\mathbf{F}_o = \mathbf{v} \odot \mathbf{F}_e + \mathbf{w}. \quad (10)$$

4.4. Training Objectives

We apply the supervision on the estimated $\{\hat{\alpha}, \hat{m}\}$ and the restored $\hat{\mathbf{B}}$. Our framework is trained in an end-to-end manner, with the following objective function:

$$\mathcal{L} = \omega_c \mathcal{L}_c + \omega_i \mathcal{L}_i, \quad (11)$$

where \mathcal{L}_c and \mathcal{L}_i respectively represent training losses for the scattering branch and the image branch. Here, the weighting parameters ω_c and ω_i that balance the relative importance of the scattering and image branches are empirically set as $\omega_c = 0.1$ and $\omega_i = 1$. For the scattering branch, we use mean square error (MSE) as the criterion, while for the image branch, the training loss is defined in a l_1 form. Specifically, we have:

$$\begin{aligned} \mathcal{L}_c &= \mathcal{L}_{MSE}(\hat{\alpha}, \alpha) + \mathcal{L}_{MSE}(\hat{m}, m), \\ \mathcal{L}_i &= \mathcal{L}_{l_1}(\hat{\mathbf{B}}, \mathbf{B}). \end{aligned} \quad (12)$$

5. Experimental Results

We now describe some details regarding the implementation of SRUDC. For the numbers of TSABs in the scattering branch, we set $\{R_1 = 2, R_2 = 3, R_3 = 4\}$, while for the number of GCABs in the image branch, we set $P = 16$. The proposed SRUDC is implemented with Pytorch [24] framework and trained for 800 epochs on four NVIDIA V100 GPUs. During the training process, the images are randomly cropped into 256×256 patches, and the mini-batch size is set to 64. The AdamW optimizer [20] is used to update learnable parameters. We use the cosine annealing

Table 1. Quantitative comparison on the synthetic dataset. The best results are in **bold** and the second best results are underlined.

Method	PSNR	SSIM	LPIPS	Params (M)	MACs (G)
DISCNet [6]	25.54	0.9246	0.1508	3.442	272.33
DAGF [31]	27.12	0.9230	0.1024	1.117	45.89
DWFormer [43]	28.85	0.9392	0.1366	1.447	131.84
UDCUNet [18]	31.33	0.9759	0.0459	<u>1.406</u>	402.16
BNUDC [11]	34.91	0.9777	<u>0.0331</u>	4.575	317.38
SRUDC-light	<u>35.56</u>	<u>0.9804</u>	0.0365	3.105	<u>106.79</u>
SRUDC-full	36.02	0.9833	0.0322	8.734	240.36

strategy [19] to adjust the learning rate, which is initialized to be 4×10^{-4} and gradually decreases to 4×10^{-6} .

5.1. Training and Testing Datasets

The proposed SRUDC is trained with a synthetic dataset. The degraded image \mathbf{I} is obtained by using (6). For the PSF \mathbf{k} in (6), we collect 10 kernels in total, including 9 measured ones with ZTE Axon 20 phone [6] and a real-world one with Transparent-OLED [36]. With these PSF kernels, we could generate 20160 training image pairs from 2016 image patches, which are cropped with the size 800×800 from the HDR images in HDRI Haven dataset [9].

Regarding testing datasets, we exploit the combination of synthetic and real-world data for the evaluation. Similar to the synthetic training dataset, we select 360 cropped image patches from HDRI Haven dataset to generate the synthetic testing set. The real testing set is formed by three parts, including UDC-ZTE [6] with 30 high-resolution UDC images taken by ZTE Axon 20 phone, UDC-UAV [6] with 162 images captured from the aerial video, and UDCIE [21] with 12 images taken in the wild environment. Hence, in total, we have 3600 synthetic and 204 real-world testing images for evaluating our method and the existing competitors.

5.2. Quantitative results

The quantitative results tested on the synthetic data are reported in Table 1. We provide two versions of SRUDC: a full version and a lightweight version denoted as SRUDC-full and SRUDC-light, respectively. The numbers of the feature channel for the three levels of the encoder-decoder in SRUDC-full and SRUDC-light are respectively set as $\{24, 48, 96\}$ and $\{16, 32, 64\}$. We compare the restoration performance of SRUDC with that of five state-of-the-art UDC image restoration methods, including DISCNet [6], DAGF [31], DWFormer [43], UDCUNet [18], and BNUDC [11]. For a fair comparison, we also retrain these models on our training data with the default parameters set as suggested, and choose the better results between pretrained and retrained models for the comparison. This is certainly advantageous to the competing methods. We adopt PSNR, SSIM [34], and LPIPS [39] as evaluation metrics. The numbers of parameters and multiply-accumulation operations (MACs) are used to compare the model complexity. As

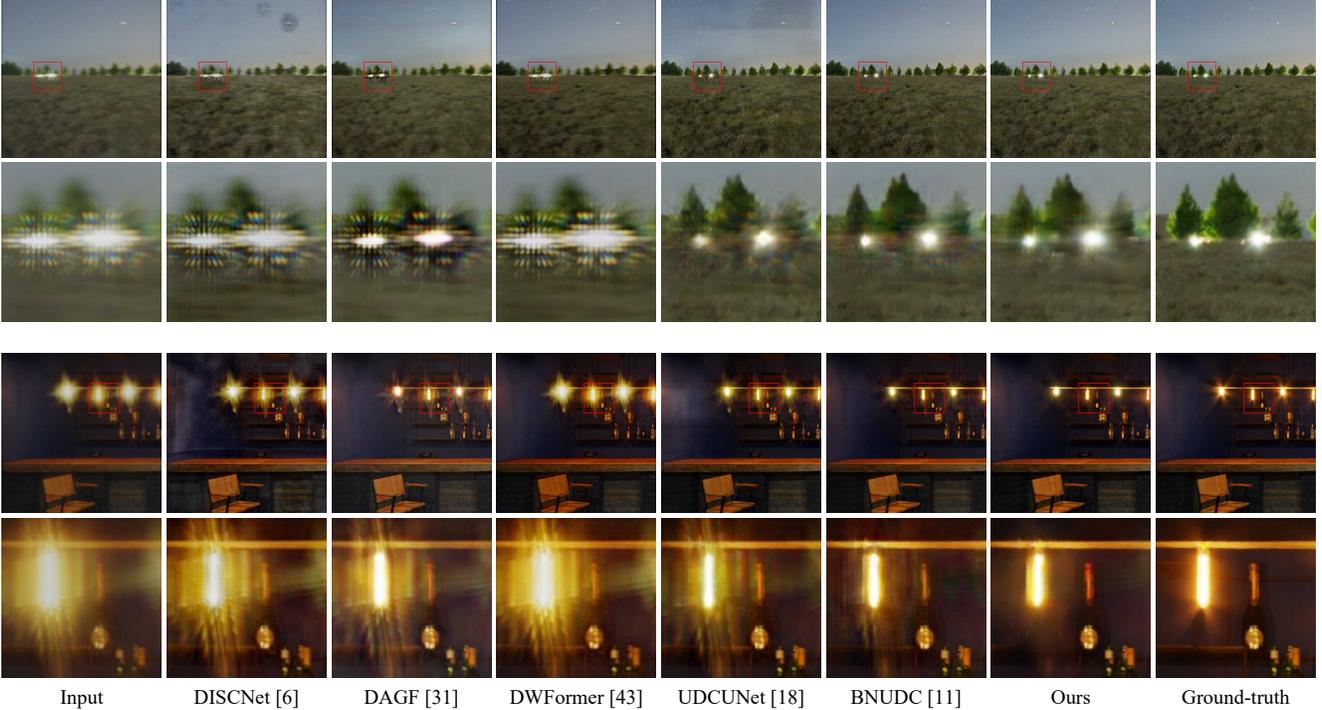


Figure 7. Visual comparison on synthetic images. Our method removes the scattering-induced haziness and contrast distortion spread over the entire image. Meanwhile, the diffraction artifacts around the light sources are greatly suppressed.

tabulated in Table 1, SRUDC-full outperforms the second-best method BNUDC by a large margin, i.e., 1.11dB in PSNR, 0.0056 in SSIM, and 0.0009 in LPIPS. For the model complexity, SRUDC-light achieves the second-lowest computation overhead with 106.79G MACs and a comparable number of parameters with existing SOTA methods. Meanwhile, SRUDC-light still maintains a performance gain of 0.65dB in PSNR over BNUDC.

5.3. Qualitative Results

We now compare the qualitative results of our method and the competing approaches on both synthetic and real-world UDC images. Fig. 7 presents the results on two synthetic examples. It can be seen that our method is capable of removing the scattering-induced haziness and contrast distortion spread over the entire image, and greatly suppressing the diffraction artifacts around the light sources. We also give the visual comparison of different methods on real-world datasets in Fig. 8. As can be observed, our SRUDC achieves the best perceptual quality, especially in terms of the reconstruction of the textual regions, e.g., repeated fences and building structures. In addition, SRUDC avoids color shifts as UDCUNet. More qualitative comparisons can be found in the supplementary file.

5.4. Ablation study

Scattering Branch: To demonstrate the benefits of the scattering branch, we build the following ablation experiments.

Table 2. Impact of the scattering branch.

Model	PSNR	SSIM	LPIPS
w/o the supervisions of m and α	35.13	0.9817	0.0347
w/o the entire scattering branch	34.66	0.979	0.0394
w/ CNN scattering branch	35.02	0.9798	0.0336
SRUDC-full	36.02	0.9833	0.0322

We first evaluate the importance of including the supervision of m and α . As shown in Table 2, switching off the supervision of m and α leads to a significant performance drop of 0.89dB, compared with SRUDC-full. We further try to justify the effect of the whole scattering branch, by deleting it from SRUDC. In this case, SRUDC degenerates to a variant of Unet. As can be seen from the second row of Table 2, the whole scattering branch brings a significant performance gain of around 1.36dB PSNR. In addition, we substitute TSABs in the scattering branch by CNN structure GCABs to show the effectiveness of the CSA. It can be observed that our proposed TSABs could introduce a significant performance improvement of around 1dB PSNR gain, due to the strong capability of modeling global information.

Image branch: We evaluate the components in the proposed GCAB in the image branch. We sequentially remove the gating operation and channel attention and retrain the counterpart models. As shown in Table 3, eliminating the gating operation and the channel attention respectively causes 0.38 dB and 0.7 dB PSNR drops.

Feature Fusion Block: To show the benefit of the FFB, we

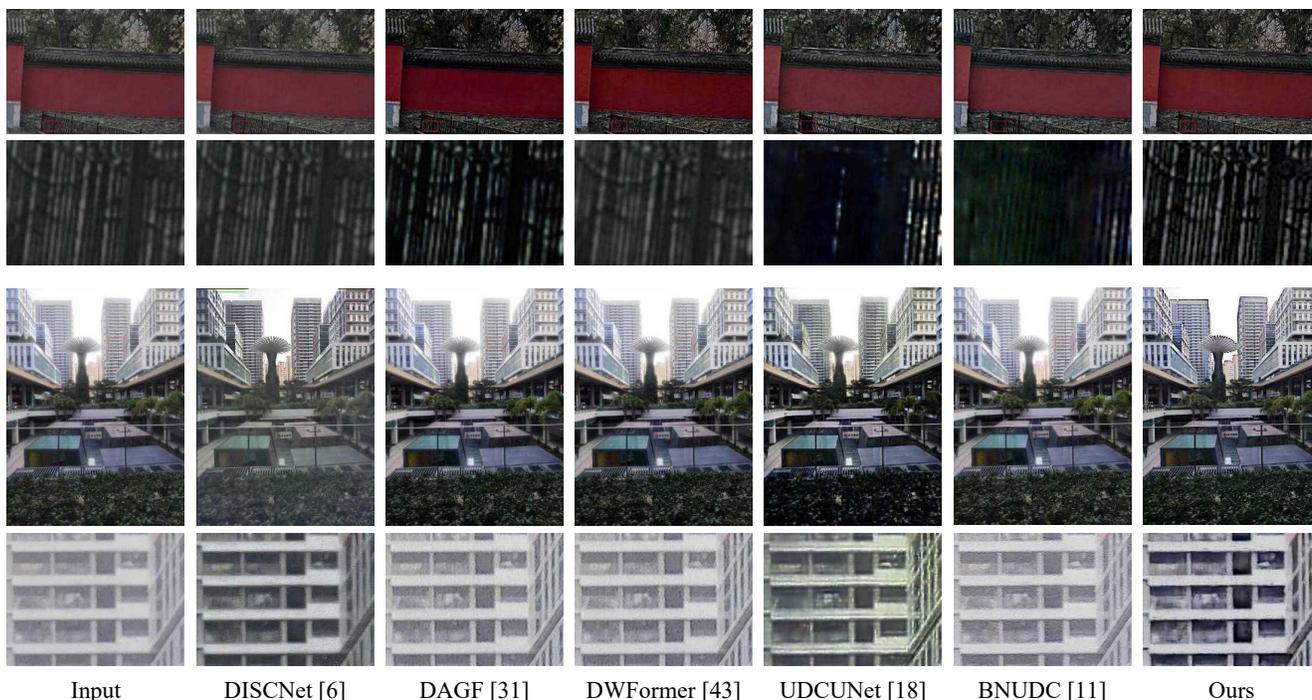


Figure 8. Visual comparison on real-world images. Our method achieves the best perceptual quality for the reconstruction of the textual regions, e.g., repeated fences and building structures.

Table 3. Effect of the components in the image branch.

Model	PSNR	SSIM	LPIPS
w/o gating operation	35.64	0.9824	0.0356
w/o channel attention	35.32	0.9806	0.0417
GCAB	36.02	0.9833	0.0322

Table 4. Effect of the feature fusion block.

Model	PSNR	SSIM	LPIPS
Concatenation	35.84	0.9813	0.0325
SKfusion [17]	35.42	0.9824	0.0335
SCAM [28]	33.59	0.9707	0.0479
Our FFB	36.02	0.9833	0.0322

build an ablation experiment on this module. Specifically, we substitute FFB with the concatenation, SKfusion [17] and SCAM [28]. The results summarized in Table 4 verify the superiority of SFT with around 0.18dB PSNR gains.

5.5. Effectiveness of Enhanced IFP

To present the importance of considering the scattering effect in the UDC image restoration, we compare the pretrained and retrained model of the state-of-the-art UDC restoration method BNUDC. The pretrained model is trained with the data generated from the existing IFP, Feng *et al.* [6], and the retrained model is trained with data synthesized by our enhanced IFP. The visual results tested on the real-world UDC images are shown in Fig. 9. From the comparison, our synthetic data effectively improves the visual quality of the restored results from BNUDC.



Figure 9. Comparison of BNUDC pretrained and retrained with synthetic data from different IFPs. (a) Real-world input UDC images. (b) Results of BNUDC pretrained with data generated by Feng *et al.* [6]. (c) Results of BNUDC retrained with data from our enhanced IFP. Our synthetic data effectively improves the visual quality of the restored results from BNUDC.

6. Conclusion

In this study, we propose to address the UDC image restoration problem with the consideration of the scattering effect caused by the display. With the physical UDC

scattering model, we improve the IFP for the realistic UDC image synthesis. To recover the clean background scenes, we specially design a dual-branch deep network, where the scattering branch estimates the parameters of the scattering effect from degraded images, while the image branch restores the background image. The devised feature fusion block provides guidance to the UDC image restoration with the global information from the scattering branch. Our model significantly outperforms the state-of-the-art methods on both synthetic and real-world UDC images.

Acknowledgement. This work was supported in part by Macau Science and Technology Development Fund under SKLIOTSC-2021-2023, 0072/2020/AMJ, and 0022/2022/A1; in part by Research Committee at University of Macau under MYRG2020-00101-FST and MYRG2022-00152-FST; in part by Natural Science Foundation of China under 61971476; and in part by Alibaba Group through Alibaba Innovative Research Program.

References

- [1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Eur. Conf. Comput. Vis.*, pages 17–33, 2022. [5](#)
- [2] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 354–363, 2021. [6](#)
- [3] Xiangyu Chen, Zhengwen Zhang, Jimmy S Ren, Lynhoo Tian, Yu Qiao, and Chao Dong. A new journey from sdrtv to hdrtv. In *Int. Conf. Comput. Vis.*, pages 4500–4509, 2021. [2](#)
- [4] Florin Conde, Marcos V. and Vasluianu, Sabari Nathan, and Radu Timofte. Real-time under-display cameras image restoration and hdr on mobile devices. In *Eur. Conf. Comput. Vis. Worksh.*, pages 747–762, 2022. [1](#)
- [5] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Jinwei Gu, and Chen Change Loy. Generating aligned pseudo-supervision from non-aligned data for image restoration in under-display camera. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5013–5022, 2023. [1](#)
- [6] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 662–671, 2021. [1](#), [2](#), [3](#), [4](#), [6](#), [8](#)
- [7] Ruicheng Feng, Chongyi Li, Shangchen Zhou, Wenxiu Sun, Qingpeng Zhu, Jun Jiang, Qingyu Yang, Chen Change Loy, Jinwei Gu, Yurui Zhu, et al. Mipi 2022 challenge on under-display camera image restoration: Methods and results. In *Eur. Conf. Comput. Vis. Worksh.*, pages 60–77, 2022. [1](#)
- [8] KeMing Gao, Meng Chang, Kunjun Jiang, Yaxu Wang, Zhihai Xu, Huajun Feng, Qi Li, Zengxin Hu, and YueTing Chen. Image restoration for real-world under-display imaging. *Opt. Express*, 29(23):37820–37834, 2021. [1](#)
- [9] HDRI. Hdri dataset. <https://polyhaven.com/hdri>. [6](#)
- [10] Jules S Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, 1990. [3](#)
- [11] Jaihyun Koh, Jangho Lee, and Sungroh Yoon. Bnucd: A two-branched deep neural network for restoring images from under-display cameras. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1950–1959, 2022. [1](#), [2](#), [3](#), [6](#)
- [12] Melvin Kuriakose, P S Hrishikesh, Densen Puthussery, and C. V. Jiji. Transform domain pyramidal dilated convolution networks for restoration of under-display camera images. In *Eur. Conf. Comput. Vis. Worksh.*, pages 364–378, 2020. [1](#)
- [13] Hyeok-Jun Kwon, Chang-Mo Yang, Min-Cheol Kim, Choon-Woo Kim, Ji-Young Ahn, and Pu-Reum Kim. Simulation of blur in transmitted image through transparent plastic for transparent oleds. *Journal of Display Technology*, 12(8):851–858, 2016. [2](#)
- [14] Kinam Kwon, Eunhee Kang, Sangwon Lee, Su-Jin Lee, Hyong-Euk Lee, ByungIn Yoo, and Jae-Joon Han. Controllable image restoration for under-display camera in smartphones. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2073–2082, 2021. [1](#), [2](#)
- [15] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.*, 28(1):492–505, 2018. [3](#)
- [16] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.*, 30:4985–5000, 2021. [3](#)
- [17] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 510–519, 2019. [8](#)
- [18] Xina Liu, Jinfan Hu, Xiangyu Chen, and Chao Dong. Udcunet: Under-display camera image restoration via u-shape dynamic network. In *Eur. Conf. Comput. Vis. Worksh.*, pages 113–129, 2022. [1](#), [6](#)
- [19] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. [6](#)
- [20] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. [6](#)
- [21] Jun Luo, Wenqi Ren, Tao Wang, Chongyi Li, and Xiaochun Cao. Under-display camera image enhancement via cascaded curve estimation. *IEEE Transactions on Image Processing*, 31:4856–4868, 2022. [1](#), [3](#), [4](#), [6](#)
- [22] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *Int. J. Comput. Vis.*, 48(3):233, 2002. [3](#)
- [23] Yongmin Park, Dongwoo Kang, Sungjin Kim, Taeseong Han, JangJin Yoo, and Moojong Lim. Paper no 5.3: simulation of visual quality for transparent oled display. In *SID Symposium Digest of Technical Papers*, volume 44, pages 156–159. Wiley Online Library, 2013. [2](#)
- [24] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. [6](#)

- [25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, pages 234–241, 2015. 5
- [26] Amin Salehi, Xiangyu Fu, Dong-Hun Shin, and Franky So. Recent advances in oled optical design. *Advanced Functional Materials*, 29(15):1808803, 2019. 2, 3
- [27] Heiko H Schuett, Franziska Baier, and Roland W Fleming. Perception of light source distance from shading patterns. *Journal of Vision*, 16(3):9–9, 2016. 3
- [28] Binbin Song, Jiantao Zhou, Xiangyu Chen, and Shile Zhang. Real-scene reflection removal with raw-rgb image pairs. *IEEE Trans. Circuits Syst. Video Technol.*, 99(99):1–1, 2023. 8
- [29] Binbin Song, Jiantao Zhou, and Haiwei Wu. Multi-stage curvature-guided network for progressive single image reflection removal. *IEEE Trans. Circuits Syst. Video Technol.*, 32(10):6515–6529, 2022. 2
- [30] Yuda Song, Yang Zhou, Hui Qian, and Xin Du. Rethinking performance gains in image dehazing networks. *arXiv preprint arXiv:2209.11448*, 2022. 5
- [31] Varun Sundar, Sumanth Hegde, Divya Kothandaraman, and Kaushik Mitra. Deep atrous guided filter for image restoration in under display cameras. In *Eur. Conf. Comput. Vis. Worksh.*, pages 379–397, 2020. 1, 3, 6
- [32] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5769–5780, 2022. 2
- [33] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. 6
- [34] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. 6
- [35] Jiaying Yan, Hong Zhao, Penghui Bu, and YuSheng Jin. Channel-wise attention-based network for self-supervised monocular depth estimation. In *2021 International Conference on 3D vision (3DV)*, pages 464–473, 2021. 5
- [36] Anqi Yang and Aswin C. Sankaranarayanan. Designing display pixel layouts for under-panel cameras. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(7):2245–2256, 2021. 6
- [37] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5728–5739, 2022. 2, 5
- [38] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14821–14831, 2021. 5
- [39] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 586–595, 2018. 6
- [40] Yulun Zhang, Kungpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Eur. Conf. Comput. Vis.*, pages pp. 286–301, 2018. 5
- [41] Wang Zhixue, Zhang Yu, Luo Lin, and Wang Nan. Csa-cdgan: channel self-attention-based generative adversarial network for change detection of remote sensing images. *Neural Comput. Appl.*, 34(24):21999–22013, 2022. 5
- [42] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9179–9188, 2021. 1, 2
- [43] Yang Zhou, Yuda Song, and Xin Du. Modular degradation simulation and restoration for under-display camera. In *Proceedings of the Asian Conference on Computer Vision*, pages 265–282, 2022. 1, 6