Controllable Image Restoration for Under-Display Camera in Smartphones

Kinam Kwon* Eunhee Kang* Sangwon Lee Su-Jin Lee Hyong-Euk Lee
ByungIn Yoo Jae-Joon Han
Samsung Advanced Institute of Technology (SAIT), South Korea

Abstract

Under-display camera (UDC) technology is essential for full-screen display in smartphones and is achieved by removing the concept of drilling holes on display. However, this causes inevitable image degradation in the form of spatially variant blur and noise because of the opaque display in front of the camera. To address spatially variant blur and noise in UDC images, we propose a novel controllable image restoration algorithm utilizing pixel-wise UDC-specific kernel representation and a noise estimator. The kernel representation is derived from an elaborate optical model that reflects the effect of both normal and oblique light incidence. Also, noise-adaptive learning is introduced to control noise levels, which can be utilized to provide optimal results depending on the user preferences. The experiments showed that the proposed method achieved superior quantitative performance as well as higher perceptual quality on both a real-world dataset and a monitor-based aligned dataset compared to conventional image restoration algorithms.

1. Introduction

Recently, an under-display camera (UDC) is in the spotlight as a new design form factor for smartphones. UDC is mounted below the display, which enables a full-screen display for better user experience without punch holes or notches. However, inevitable image degradation is accompanied by UDC, and thus, to realize this attractive technology, the corresponding technical breakthrough against the degraded camera imaging performance is required.

One of the major limitations of UDC system is low light transmission of the display. Relatively low signal-to-noise ratio (SNR) and additional noise due to the display cause severe image quality degradation especially in smartphone cameras, whereas other cameras powered by large sensor pixels and a large lens adequately compensate for the signal reduction [36]. In addition, various forms of optical diffraction and interference are generated by the display pattern which consists of transparent and opaque areas; therefore, a complicated blurry image is captured by a particular diffraction blur kernel [10]. Furthermore, the color tone is distorted by the wavelength-dependent light transmittance of the display, which is applied to each color filter array.

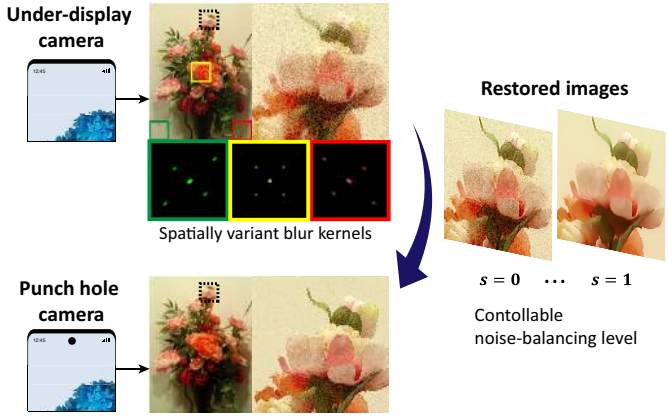


Figure 1: Controllable image restoration algorithm to enhance the quality of UDC images degraded by spatially variant blur and noise.

Meanwhile, image sensor data are normally reconstructed into standard RGB images by an image signal processor (ISP) which consists of the sequential process of demosaicing, white balancing, color space transform, sharpening, etc. The ISP is implemented on-chip and tuned to a specific sensor. Recently, several attempts have been made to replace ISP with neural networks [13, 17, 22]. However, their preliminary results under laboratory experimental settings may be incompatible with existing functions such as scalability, scene-specific detail enhancement, multi-image fusion, and high dynamic range.

Therefore, solving the UDC problem while maintaining compatibility with the existing functions has become an important problem. A simple solution is to handle the degradation caused only by a UDC in the linear RGB sensor data, while maintaining the conventional ISP. The linear RGB domain has the advantage of preserving the physical properties of the UDC system because it does not go through nonlin-

*These two authors contributed equally.

ear processes in the ISP [2, 4]. In addition, color tone distortion can be solved by white-balance correction and tone mapping in the ISP, which can reduce the network burden by focusing on deblurring and denoising. Therefore, we attempted to solve the UDC problem in the linear RGB domain to cope with various imaging conditions in real-world scenarios.

In this study, the UDC problem is defined as image restoration from complicated diffraction blur and high noise, which are described in Fig. 1. In order to address spatially variant blur and noise, the proposed architecture is designed to be controllable pixel-by-pixel with respect to the level of blur and noise. Additionally, in terms of practical usage, a controllable noise-balancing level is proposed because user preferences vary depending on the light conditions.

The training dataset is important for learning-based approach to solve the image restoration problem, but it is difficult to obtain actual pairs of degraded and ground truth images. Although [1, 36] provide an actual paired dataset, it is restricted to artificial settings such as only indoor scenes or monitor-displayed images under fixed lighting, focus, and exposure time. This dataset cannot cover the wide range of light brightnesses and multiple light sources in real-world scenarios. To address the gap between synthetic and actual data, we propose elaborate optical modeling by reflecting the effects of oblique light incidence additionally. Based on the modeling results, realistic training data is synthesized and controllable restoration architectures are trained to address spatially variant degradation.

The contributions of this work are summarized as follows. (1) A realistic optical model of the UDC system is presented by considering the effects of both normal and oblique light incidence. (2) A new pixel-wise controllable architecture is provided to address spatially variant diffraction blur and noise of UDC images from a smartphone in an ISP-compatible manner. (3) Noise-adaptive learning is introduced to control noise levels depending on the user preference and various imaging conditions.

2. Related Works

2.1. Image restoration for UDC

UDC image restoration was firstly addressed in [36], which included the handling of various types of image degradation such as blur, noise, and color shift. The authors proposed data synthesis by optical modeling and collected actual paired data using a monitor-camera imaging system. They presented comprehensive results about the display structures, training data, loss functions, and network architectures. A UDC image restoration challenge [35] was held based on the collected monitor-captured UDC images. The participants applied various well-known techniques such as skip or dense connections, attention layers, utilization of explicit information such as shade maps, and decomposition of images. They all showed feasible performance under the given conditions of the monitor-camera imaging system.

UDC image degradation can differ depending on the display structure and camera specifications [19, 24, 30, 34]. In real-world scenarios such as UDC imaging in smartphones, UDC image restoration must be performed in collaboration with the 3A conditions (auto exposure, auto white-balance and auto focus) and image reconstruction pipelines (from linear RGB to standard RGB). In this study, the UDC image restoration problem is newly defined for smartphones. The proposed method is applied not only to monitor-captured images, but also to actual images captured in various real-world scenarios.

2.2. Controllable network

Deep learning has shown remarkable performance in various image restoration tasks such as denoising [14, 16, 31, 32], deblurring [6, 21, 23, 25], and super-resolution [5, 18, 28, 33]. However, a deterministic result is output by feeding an input into the network. Moreover, they have no option to reflect the user preferences or to adapt to condition variations.

Pioneering works [11, 12, 27] have proposed controllable networks that can produce various results based on the given control parameters to restore single or multiple types of degradations adaptively, such as noise, blur, and compression artifacts. Their networks were trained on synthetic data that were generated by controlling degradation, and the performance was demonstrated using synthetic data under similar experimental settings. Perceptually or quantitatively better results can be achieved by controlling the parameters, even compared to the results for the parameters corresponding to the input degradation levels.

In UDC image restoration, controllable networks are essential to address spatially variant blur and noise. In addition, controlling the balance between deblurring and denoising is also important for better subjective evaluation. In this study, we parameterize the blur and noise induced by UDC, and adaptively relieve actual UDC degradation based on pixel-wise control parameter maps instead of the image-wise control parameter vector in [12]. We also propose a noise-adaptive training method to control the level of denoising, which enables the optimal results to be output for the given user preferences according to the conditions.

3. Method

3.1. Background

The light transmission rate (LTR) through a typical mobile display (P-OLED) is less than 3% [36]. A UDC in-

evitably utilizes more gains (analog and digital gains) to achieve acceptable exposure values, which results in increasing noise levels. This issue is critical for the imaging sensors of smartphones because they are too small to achieve acceptable SNR. To overcome these problems, some display pixels are replaced with transparent window areas [30]. The transparent window areas are regularly arranged to preserve reasonable visibility of the display. Light transmission is allowed only through the transparent window areas by accumulating an additional layer outside the transparent window area. This feature helps simplify the degradation patterns and predict the degradation kernels more accurately. Fig. 2 shows schematic diagrams of the transparent window areas in the camera region. This approach enables LTR improvement up to about 20%.

A P-OLED display attenuates the LTR depending on the wavelength of light, which results in color shift [36]. In other words, blue components (short wavelengths) are attenuated more than red components (long wavelengths). The unbalanced light transmission can be compensated by white balancing and tone mapping in conventional ISP pipelines.

3.2. Problem formulation

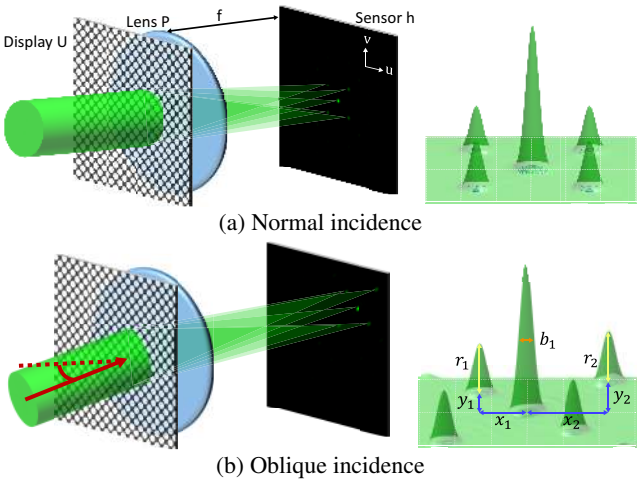
We define the UDC image restoration problem in smartphones as deblurring and denoising, to resolve diffraction induced by the transparent window areas and to suppress noise induced by compensating for the decreased LTR, respectively. We assume that the UDC images and non-UDC images are captured with the same exposure value under an auto-exposure algorithm. The image restoration is processed in the linear raw-RGB domain because the noise distribution and shape of the blur kernel induced by a UDC are significantly changed in the standard-RGB domain [2, 4]. In summary, the UDC image degradation can be represented by

$$y = Ax + n, \tag{1}$$

where y, x, A, and n are a UDC image, a ground-truth image, a spatially variant blurring operator, and noise, respectively. Detailed explanations are provided in the following sections.

3.3. Optical modeling

The degradation model of the UDC panel is explained in detail in [36]. We used the design of transparent window pattern U as in Fig. 2a, instead of a microscope image of the display panel [36]. By multiplying this pattern times the circular aperture P from the entrance pupil distance of the camera and computing the Fourier transform $F = \mathcal{F}\{UP\}$, the point spread function (PSF) of the UDC panel can be derived. Explicitly, the PSF intensity h at the focal plane of the lens is proportional to the Fraunhofer pattern of the



(a) Normal incidence

(b) Oblique incidence

Figure 2: Optical modeling including (a) normal incidence and (b) oblique incidence of light to reflect the spatially variant properties.

incident field [26, 36]:

$$h(u,v) \propto \left| F\left(\frac{u}{\lambda f}, \frac{v}{\lambda f}\right)\right|^2, \tag{2}$$

where (u, v) is the sensor coordinate, λ is the wavelength, and f is the effective focal length of the lens.

Although the PSF in Eq. (2) is valid for the center of the image, where the light is normally incident on the panel (Fig. 2a), it does not work for the corners of the image, where the light is obliquely incident on the panel (Fig. 2b). Owing to the periodic windows pattern, the UDC panel acts as a planar diffraction grating and the obliquely incident light induces conical diffraction [10] and distorted PSFs. In order to reflect the spatially variant properties in our degradation model, we generalized (2) by considering the oblique-incidence cases.

This problem has been addressed in previous studies. For instance, [10] discussed the diffraction with obliquely incident angles in terms of the direction cosines of the incident and diffracted angles, but it was limited to specific types of gratings such as grooves. Although diffraction was analyzed in [20] using arbitrarily oriented gratings under an arbitrary angle of incidence, a complicated three-dimensional vector coupled-wave analysis was utilized, which is difficult to implement. Instead, we simplified the generalization problem of obtaining spatially variant PSF information valid for all image pixel coordinates, using the orthogonal projection. For each sensor pixel p, the orthogonal projection Q_p onto the plane perpendicular to the incident ray to p is considered. Thereafter, the PSF corresponding to sensor pixel p can be approximated from the Fourier transform of the orthogonal projection of the UDC panel pattern

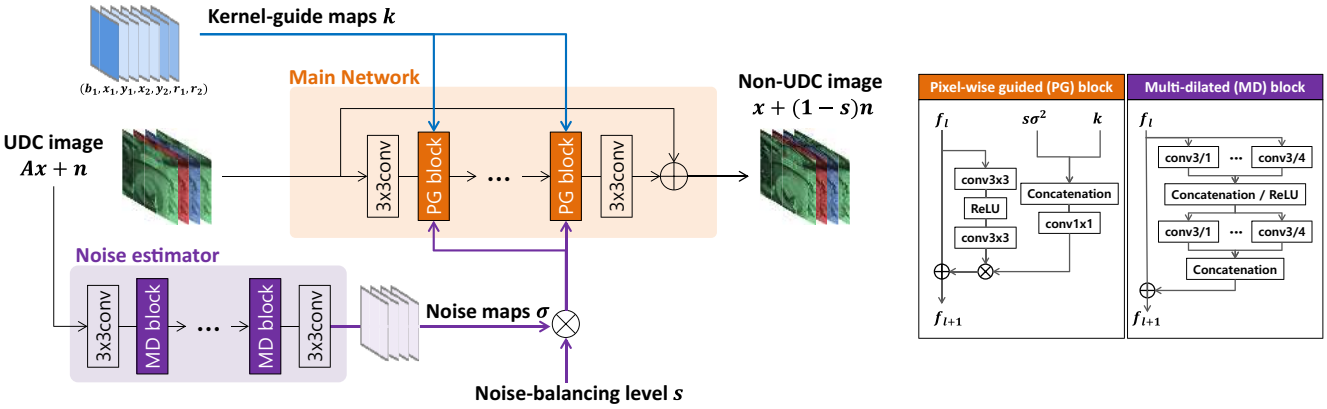


Figure 3: Proposed framework, which consists of a main network for UDC image restoration and auxiliary components to control the main network. Degradation induced by UDC is resolved pixel by pixel based on given kernel-guide maps and noise maps from the noise estimator.

$F_p = \mathcal{F}\{Q_p(UP)\}$:

$$h_p(u, v) \propto \left| F_p\left(\frac{u}{\lambda f}, \frac{v}{\lambda f}\right)\right|^2. \tag{3}$$

3.4. Overall framework

The main network enhances UDC images by utilizing controllable parameters such as explicit kernel-guide maps, estimated noise maps, and an explicit noise-balancing level as shown in Fig. 3. The noise estimator and main network are based on EDSR [18]. The residual block of EDSR is replaced with a multi-dilated (MD) block in the noise estimator. The residual block uses n-channel 3×3 convolutions with a dilation rate of 1, but the MD block uses four $n/4$-channel 3×3 convolutions with four different dilation rates of 1, 2, 3, and 4. It enables the receptive field to be enlarged efficiently in relatively shallow networks and noise levels to be estimated properly from complex texture patterns. The main network uses pixel-wise guided (PG) blocks, which modulate feature maps by utilizing control parameters that represent the kernel and noise level in each position. The control parameters are mapped to the number of channels in each block by using a 1×1 convolution layer. CResMD [12] showed that this feature modulation is better than direct concatenation to input or AdaFM [11].

3.4.1 Controllable kernel representation

The calculated kernels h_p in (3) are represented as seven semantic parameters ($k = (b_1, x_1, y_1, x_2, y_2, r_1, r_2)$) as shown in Fig. 2b. Because the kernel is symmetric with respect to the origin, it is sufficient to use half of the kernels to represent all of the kernels. Each kernel is parameterized as the width b_1 of the main lobe, coordinates (x_1, y_1, x_2, y_2) and the intensity ratios (r_1, r_2) between the peaks of the main lobe and first grating lobe. The width is given by $1/(2\rho)$, where ρ is the pixel pitch size. For example, when the pixel pitch size is 1 μm, the width of the kernel is represented as 0.5.

From the representation parameter k, the original kernels can be reconstructed as the main lobe and four first grating lobes by ignoring high-order grating lobes and small ripples. The root-mean-square error between the reconstructed kernels and optics-based calculated kernels is 0.29% and 0.32% in the center and corners of the image, respectively. Hence, the kernel representation can significantly reduce the number of kernel parameters from 33×33 (kernel size including second grating lobes) to 7, with acceptable error rates.

In the training phase, training data are synthesized by the seven kernel parameters with certain ranges to reflect the actual degradation levels and for robustness against potential variations, as described in 4.1.1. In the inference phase, the pre-calculated kernel-guide maps are utilized. The kernel-guide maps are calculated from the hardware specifications, such as those of the transparent window areas, lens array, and imaging sensor.

3.4.2 Noise estimator

Noise in the linear raw-RGB domain is generated by various sources, and it includes signal-dependent (owing to photon noise and color-dependent transmission rate), and spatially variant characteristics (owing to the lens shading correction). The noise is simplified by using the heteroscedastic Gaussian model. The noise estimator outputs the standard deviation of the Gaussian model to represent the noise from various sources as parameter σ.

In the training phase, the noise n is generated by utilizing the Gaussian distribution $\mathcal{N}$ of a signal-dependent term α_1 and a signal-independent term α_2 as follows:

$$n \sim \mathcal{N}(0, \sigma^2) = \mathcal{N}(0, \alpha_1 Ax + \alpha_2). \tag{4}$$

Spatially variant noise is efficiently relieved by considering explicit known noise level maps [32]. ATDNet [14] utilizes estimated noise level maps and thus retains details in the texture regions while suppressing the noise. The noise changes with the imaging objects or imaging conditions, and the noise level maps cannot be appropriately calculated from only metadata such as the gain, luminance, and aperture size. Therefore, the noise maps are estimated from UDC images, and utilized to enable proper restoration of the UDC images by the main network based on the noise level.

3.4.3 Noise-adaptive training

The noise estimator is trained to minimize the following loss function:

$$\mathcal{L}_{ne} = \|\sigma - N(Ax + n)\|_2 \quad (5)$$
$$= \|\sigma - \hat{\sigma}\|_2, \quad (6)$$

where N is the noise estimator that utilizes the UDC image $(Ax + n)$ as input. The main network is trained by the following loss function:

$$\mathcal{L}_m = \|x + (1 - s)n – M(Ax + n, s\hat{\sigma}, k)\|_1, \quad (7)$$

where M is main network that utilizes the UDC image $(Ax + n)$, estimated noise level maps $\hat{\sigma}$, and kernel-guide maps k as input. The estimated noise maps are multiplied by a random noise-balancing level $s \in [0, 1]$, and the main network reduces the noise level proportional to s while deblurring the UDC images. This training method allows the main network to take full advantage of the estimated noise map and control the noise level efficiently.

4. Experiments

4.1. Data

Synthetic training data were generated from the modeling 3.3. Aligned and real-world datasets were utilized for quantitative and qualitative evaluations. The aligned dataset was collected by utilizing a monitor-camera imaging system [35], and its details are provided in the supplementary material.

4.1.1 Synthetic training data

Synthetic data were generated on the fly in the training phase based on DIV2k data [3]. Original patches with dimensions of 512×512 were randomly cropped from the DIV2k data, and normalized to be within the range [0,1]. The patches were unprocessed by gamma decompression [4]. Degraded patches were generated by filtering the UDC kernels and injecting noise. The UDC kernels were randomly generated by the representation parameters k, described in Section 3.4.1, as follows. (1) The values of the width b_1 were generated to follow a uniform distribution in the range of [0.25, 1]. (2) The values of the coordinates (x_1, y_1, x_2, y_2) were within the range of [4, 14], and were scaled by 100 to balance with other parameters. (3) The intensity ratios (r_1, r_2) were within the range of [0.01, 0.015]. The signal-dependent ($\alpha_1 \in$[0,0.001]) and signal-independent ($\alpha_2 \in$[0,0.001]) noise were injected channel-wise into the images filtered by the UDC kernels.The distributions of the parameters were empirically chosen based on the calculated kernels and the measured noise. The pairs of degraded and original patches were subsampled to follow the Bayer pattern, reshaped to four-channel GRBG images, and cropped to dimensions of 160×160 for training. Each pair corresponded to seven-channel kernel-guide maps and four-channel noise level maps.

4.1.2 Real-world dataset

Real-world images were captured by smartphones (Samsung Galaxy S20 Plus) under on-device 3A conditions. The images were 10-bit Bayer-pattern 3648×2736 linear-RGB images. The images were labeled as UDC images and non-UDC images, depending on whether there was a panel in front of the camera when the relevant image was captured. We collected paired data of UDC and non-UDC images in real-world scenarios such as indoor/outdoor, day/night, moving/stationary objects, and low/high lighting conditions. The real-world paired data were not aligned pixel-wise, and some of the pairs had time and/or perspective gaps. They were utilized as test sets for modulation transfer function (MTF) analysis and qualitative evaluation.

4.2. Implementation details

In our implementation, the noise estimator consisted of 6 MD blocks with 32 channels, and the main network consisted of 12 PG blocks with 32 channels. Each network had approximately 113.4k and 228.9k learnable weights, respectively. The whole framework was trained end to end using the synthetic data by minimizing $\mathcal{L}_m + 100\mathcal{L}_{ne}$. We trained the model using the ADAM optimizer [15] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We set the minibatch size to 1. The learning rate was initialized to 0.0002 and multiplied by 0.97 every 8k updates of the weights. We trained several models with differently initialized weights [8] by performing 800k weight updates, and we chose the model with the best validation performance, which was marked as “Ours”. In the experimental results, the noise-balancing level of Ours was chosen based on the ISO value to achieve the best mean opinion score (MOS), which is described in 4.5. For visualization, all linear-RGB images were processed by in-house ISP pipelines designed for Samsung Galaxy smartphones.

4.3. Ablation study

We performed several ablation studies to demonstrate the effectiveness of each component as follows.

Kernel-guide maps. The skewed blur kernels in image corners can be easily observed in the region with high contrast, as shown in Fig. 4. The restoration performance of the model without utilizing kernel-guide maps (w/o KG) deteriorates in the skewed blur of the corner regions. In addition, the degradations in corner regions were not resolved at all when the kernel-guide parameters representing the blur in the center regions were given for the proposed method, as shown in "incorrectKG" in Fig. 4. However, Ours resolves various blurs when the correct pixel-wise kernel-guide maps are given.

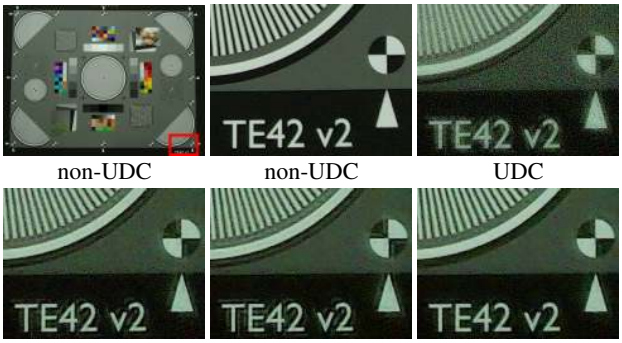


Figure 4: Ablation study of kernel-guide maps.

Noise estimator. The model without utilizing the noise estimator (w/o NE) over-smoothed the fine details, as shown in Fig. 5. The noise estimator was trained to estimate the noise maps properly, enabling the distinction of complex patterns and noise. The main network can reduce the noise, while maintaining delicate structures based on the estimated noise maps. Empirically, the receptive fields of the noise estimator influence the performance of the texture regions, which is why we utilized MD blocks for the noise estimator. We provide an example of this in the supplementary material.

Noise-adaptive training. When noise-adaptive training is not utilized, the effect of the noise-balancing level is slight as shown in Fig. 6. The restored images from the model without noise-adaptive training (w/o NT) are almost the same, although the estimated noise maps were scaled by different noise-balancing levels. In addition, the noise distributions in the UDC image and restored image ($s = 0$) are completely different. On the other hand, the proposed model reduces the noise from the UDC image by a given noise-balancing level. The restored image ($s = 0$) does not change the noise distribution of the UDC image while resolving blur. Noise-adaptive training helps the main network utilize the estimated noise maps fully because the main network cannot perform noise reduction by a certain quantity without utilizing the noise level maps.

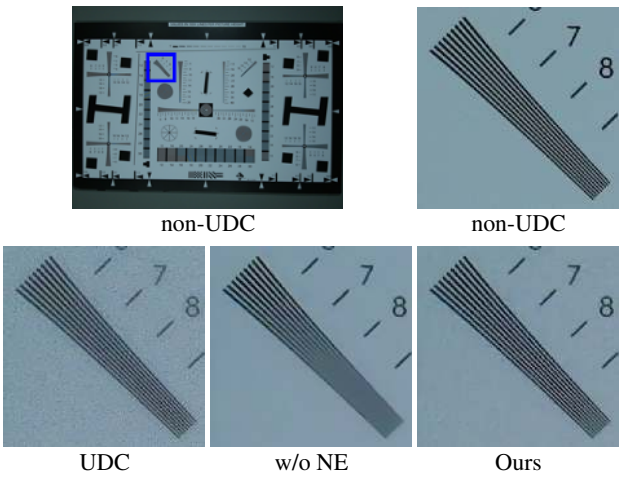


Figure 5: Ablation study of noise estimator.

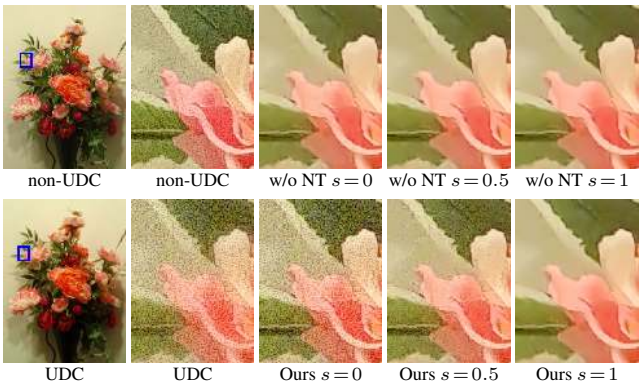


Figure 6: Ablation study of noise-adaptive training.

4.4. Performance evaluation

The proposed method was compared with two classical methods and two learning-based methods as follows. (1) WF [9]: The Wiener filter was used for deconvolution with a fixed kernel from the normal light incidence. (2) CGLS [7]: The conjugate gradient algorithm for least squares problems was used for the spatially variant deconvolution. (3) Real-mon: The network was trained using the aligned monitor-captured dataset. The monitor-camera imaging system was introduced in [36]. (4) Syn-nor: The network was trained using synthetic data that were generated by only considering normal incidence. The data synthesis method corresponds to [36]. The networks of Real-mon and Syn-nor were the same as the main network in the proposed framework. Note that Real-mon and Syn-nor are designed to compare to [36] because their conditions were completely different from ours, for example, the blur

Table 1: Performance comparison between several methods on DIV2k monitor-captured images (PSNR(dB) and SSIM) and TE42v2 chart image (MTFs (cycles/pixel)). The best and second best performances are indicated in red and blue, respectively.

Method	PSNR	SSIM	MTF25	MTF50
UDC	36.46	0.9218	0.3614	0.2017
WF	37.65	0.9460	0.2674	0.1761
CGLS	38.15	0.9511	0.2638	0.1726
Real-mon	38.24	0.9804	0.1623	0.1384
Syn-nor	38.52	0.9676	0.3482	0.3014
Ours	38.68	0.9715	0.3797	0.3024

by transparent windows in the display, size of the target imaging sensors, and domain of the network output (linear RGB or standard RGB) were different.

Quantitative evaluation. The PSNR and SSIM[29] values in Table 1 were calculated using 40 aligned monitor-captured images. The learning-based methods outperformed the classical methods, and the performances of both the classical and learning-based methods improved when the spatially variant blur was considered. Ours and Real-mon showed the best performance in terms of PSNR and SSIM, respectively. Ours showed results comparable to those of Real-mon, although actual monitor-captured data were not utilized for training.

MTF was measured to quantify image sharpness in TE42v2 chart images by using Image Engineering iQ-Analyzer v6.2.2.1. The spatial frequencies with relative contrast values of 25%, and 50% are labeled as MTF25, and MTF50 in Table 1, respectively. The restored images are displayed in Fig. 7. Ours improved MTFs, in other words, the sharpness at all frequencies, but the other methods showed partial improvements at specific frequencies or no improvements in terms of MTFs. MTF50, which is well known to be correlated with perceptual sharpness, dramatically increased by about 50% in Ours.

Qualitative evaluation. A monitor-captured image, and three actual images taken in various imaging conditions were restored using several methods, and are displayed as shown in Fig. 7. The deblurring performances of the two classical methods were marginal, and they suffered from boosted noise in low-light conditions. WF and Syn-nor cannot address skewed blurs in corner regions as shown in chart images of Fig. 7. Syn-nor and Ours showed quite similar performance in center regions with low noise levels. Syn-nor had suboptimal results except in the center regions, and it frequently caused over-smoothing in regions with high noise levels or complex texture because it does not utilize kernel-guide maps and noise estimator. Real-mon was tuned to only monitor-captured data, and thus it frequently suffered from halo artifacts and had residual noise in actual scenarios. Moreover, Real-mon partially caused blur in restored actual images and even monitor-captured images. Ours produced superior deblurring effects in complex texture or character regions. In addition, Ours retained fine details without causing over-smoothing or boosted noise.

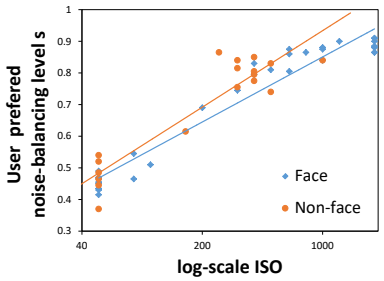


Figure 8: Scatter plots of noise-balancing level s versus ISO. Each sample point represents averaged s value for one test image.

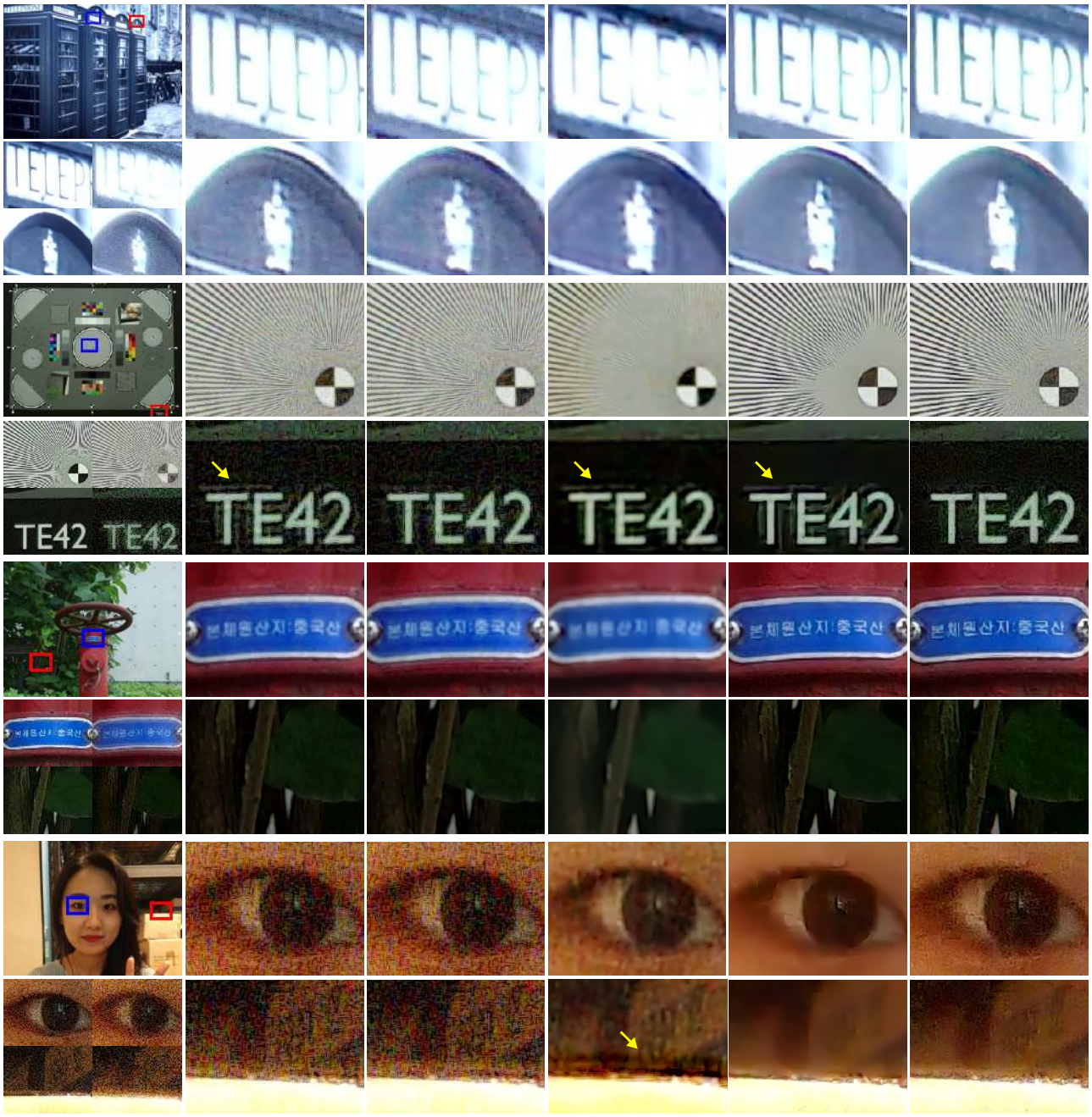
4.5. MOS test on the noise-balancing level

We performed a MOS test to analyze perceptual quality according to noise-balancing levels. Specifically, we asked 25 raters to select the best image among five restored images that were adjusted by changing the noise-balancing level s. The raters compared 50 sets of anonymized restored images including selfies and general images taken under various light conditions. All of the raters performed subjective evaluation under the same environment with a full HD monitor and were given instructions before the experiments. In the instruction, we recommend that the raters select the best image, but if the decision was difficult, they could select multiple images that seemed better than the others.

We analyzed the results depending on the ISO, which is proportional to the gains in the image. The raters preferred high noise-balancing levels for the noisy images that were taken under low-light conditions and low noise-balancing levels for the images that were taken under sufficient illumination. In Fig. 8, the results are divided into face and non-face categories because UDC is designed for front camera, which is typically used to take selfies. According to this experiment, people are more tolerant to noise in face images than in non-face images, and they reported that the residual noise was more natural and realistic.

5. Conclusion

We propose a novel controllable image restoration framework for UDC in smartphones. An optical modeling is elaborated to represent the degradation caused by UDC accurately. In designing the proposed software architecture, a noise estimator is adopted to maintain the fine



non-UDC UDC WF CGLS Real-mon Syn-nor Ours

Figure 7: Restoration results obtained using several methods on (first row) a monitor-captured image from DIV2k, (second row) a TE42v2 chart image, (third row) an outdoor image taken in daylight, and (fourth row) a selfie image taken under low-light conditions. See more samples with large sizes in the supplementary material.

details while deblurring and denoising. And then, kernel-guide maps and estimated noise maps are utilized in the manner of feature modulation for the main restoration network to address spatially variant blur and noise. Moreover, practical computing issues such as ISP compatibility and use-case constraints also have been considered. The noise-adaptive training is adopted to control desired noise levels, which can reflect user preferences depending on various real-world imaging conditions. Finally, we would like to remark that the proposed method quantitatively and qualitatively outperforms the compared methods in various actual scenarios. We believe that the proposed method provides the core technical foundation for enabling a new form factor, UDC, in smartphones.

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S. Brown. NTIRE 2020 Challenge on Real Image Denoising: Dataset, Methods and Results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2020. 2

[2] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3165–3173, 2019. 2, 3

[3] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 126–135, 2017. 5

[4] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11036–11045, 2019. 2, 3, 5

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015. 2

[6] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[7] Silvia Gazzola, Per Christian Hansen, and James G Nagy. IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems. *Numerical Algorithms*, 81(3):773–811, 2019. 6

[8] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010. 5

[9] Rafael C. González and Richard E. Woods. *Digital Image Processing, 3rd Edition*. Pearson Education, 2008. 6

[10] James E Harvey and Richard N Pfisterer. Understanding diffraction grating behavior: including conical diffraction and Rayleigh anomalies from transmission gratings. *Optical Engineering*, 58(8):087105, 2019. 1, 3

[11] Jingwen He, Chao Dong, and Yu Qiao. Modulating image restoration with continual levels via adaptive feature modification layers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11056–11064, 2019. 2, 4

[12] Jingwen He, Chao Dong, and Yu Qiao. Multi-dimension modulation for image restoration with dynamic controllable residual learning. *arXiv preprint arXiv:1912.05293*, 2019. 2, 4

[13] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera ISP with a single deep learning model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2275–2285, 2020. 1

[14] Yoonsik Kim, Jae Woong Soh, and Nam Ik Cho. Adaptively tuning a convolutional neural network by gate process for image denoising. *IEEE Access*, 7:63447–63456, 2019. 2, 5

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[16] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In Jennifer Dy and Andreas Krause, editors, *Proceedings of Machine Learning Research*, volume 80, pages 2965–2974, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR. 2

[17] Zhetong Liang, Jianrui Cai, Zisheng Cao, and Lei Zhang. Cameranet: A two-stage framework for effective camera ISP learning. *arXiv preprint arXiv:1908.01481*, 2019. 1

[18] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 136–144, 2017. 2, 4

[19] Sehoon Lim, Yuqian Zhou, Neil Emerton, Tim Large, and Steven Bathiche. 74-1: Image restoration for display-integrated camera. In *SID Symposium Digest of Technical Papers*, volume 51, pages 1102–1105. Wiley Online Library, 2020. 2

[20] M G Moharam and Thomas K Gaylord. Three-dimensional vector coupled-wave analysis of planar-grating diffraction. *JOSA*, 73(9):1105–1112, 1983. 3

[21] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3883–3891, 2017. 2

[22] Eli Schwartz, Raja Giryes, and Alex M Bronstein. DeepISP: Toward learning an end-to-end image processing pipeline. *IEEE Transactions on Image Processing*, 28(2):912–923, 2018. 1

[23] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 237–246, 2017. 2

[24] Quan Tang, He Jiang, Xindong Mei, Shaojun Hou, Guanghui Liu, and Zhifu Li. 28-2: Study of the image blur through FFS LCD panel caused by diffraction for camera under panel. In *SID Symposium Digest of Technical Papers*, volume 51, pages 406–409. Wiley Online Library, 2020. 2

[25] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8174–8182, 2018. 2

[26] David Voelz. Computational Fourier optics: A MATLAB tutorial. Society of Photo-Optical Instrumentation Engineers, 2011. 3

[27] Wei Wang, Ruiming Guo, Yapeng Tian, and Wenming Yang. CFSNnet: Toward a controllable feature space for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4140–4149, 2019. 2

[28] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018. 2

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7

[30] Zhibin Wang, Yilu Chang, Qi Wang, Yingjie Zhang, Jacky Qiu, and Michael Helander. 55-1: Invited paper: Self-assembled cathode patterning in AMOLED for under-display camera. In *SID Symposium Digest of Technical Papers*, volume 51, pages 811–814. Wiley Online Library, 2020. 2, 3

[31] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2

[32] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 2, 5

[33] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *The European Conference on Computer Vision (ECCV)*, 2018. 2

[34] Zhenhua Zhang. Image deblurring of camera under display by deep learning. In *SID Symposium Digest of Technical Papers*, volume 51, pages 43–46. Wiley Online Library, 2020. 2

[35] Yuqian Zhou, Michael Kwan, Kyle Tolentino, Neil Emerton, Sehoon Lim, Tim Large, Lijiang Fu, Zhihong Pan, Baopu Li, Qirui Yang, et al. UDC 2020 challenge on image restoration of under-display camera: Methods and results. *arXiv preprint arXiv:2008.07742*, 2020. 2, 5

[36] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. *arXiv preprint arXiv:2003.04857*, 2020. 1, 2, 3, 6