

Optical Aberrations Correction in Postprocessing Using Imaging Simulation

SHIQI CHEN, HUAJUN FENG, DEXIN PAN, ZHIHAI XU, QI LI, and YUETING CHEN,
Zhejiang University, China



Fig. 1. We present a practical method for optical aberrations correction. We calculate the spatially variant PSFs (shown on the left in 10 \times resampling) of a specific lens by the proposed optical PSF model (Section 4). The PSFs are used for imaging simulation to generate training data for the proposed postprocessing chain (Section 5). Our approach significantly improves image quality (shown on the right) and is convenient for deploying to new devices.

As the popularity of mobile photography continues to grow, considerable effort is being invested in the reconstruction of degraded images. Due to the spatial variation in optical aberrations, which cannot be avoided during the lens design process, recent commercial cameras have shifted some of these correction tasks from optical design to postprocessing systems. However, without engaging with the optical parameters, these systems only achieve limited correction for aberrations.

In this work, we propose a practical method for recovering the degradation caused by optical aberrations. Specifically, we establish an imaging simulation system based on our proposed optical point spread function model. Given the optical parameters of the camera, it generates the imaging results of these specific devices. To perform the restoration, we design a spatial-adaptive network model on synthetic data pairs generated by the imaging simulation system, eliminating the overhead of capturing training data by a large amount of shooting and registration.

Moreover, we comprehensively evaluate the proposed method in simulations and experimentally with a customized digital-single-lens-reflex camera lens and HUAWEI HONOR 20, respectively. The experiments

demonstrate that our solution successfully removes spatially variant blur and color dispersion. When compared with the state-of-the-art deblur methods, the proposed approach achieves better results with a lower computational overhead. Moreover, the reconstruction technique does not introduce artificial texture and is convenient to transfer to current commercial cameras. Project Page: <https://github.com/TanGeeGo/ImagingSimulation>.

CCS Concepts: • Computing methodologies → Reconstruction;

Additional Key Words and Phrases: Optical aberrations, imaging simulation, deep-learning networks, image reconstruction

ACM Reference format:

Shiqi Chen, Huajun Feng, Dexin Pan, Zhihai Xu, Qi Li, and Yueting Chen. 2021. Optical Aberrations Correction in Postprocessing Using Imaging Simulation. *ACM Trans. Graph.* 40, 5, Article 192 (September 2021), 15 pages. <https://doi.org/10.1145/3474088>

1 INTRODUCTION

As an extension of human vision, modern digital imaging techniques have greatly enriched our methods of interacting with the world. Coupled with complex optical systems, powerful digital sensors, and heavy **image signal processing (ISP)** algorithms, today's digital cameras can handle many extreme shooting scenarios, such as low-light [Chen et al. 2018], long-distance [Zhang et al. 2019], and high-dynamic-range scenarios [Hasinoff et al. 2016; Yan et al. 2019]. However, as sophisticated as today's digital imaging systems are, it is difficult to totally avoid the degradation caused by optical aberration. To alleviate the aberrations while satisfying the constraints of spacing, mobile imaging devices (smartphones,

This work was supported by National Natural Science Foundation of China (NSFC) under Award No. 61975175.

Authors' addresses: S. Chen, H. Feng, D. Pan, Z. Xu, Q. Li, and Y. Chen, Zhejiang University, China; emails: {chenshiqi, fenghj, pdx, xuzhh, liqi, chenyt}@zju.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Association for Computing Machinery.

0730-0301/2021/09-ART192 \$15.00

<https://doi.org/10.1145/3474088>

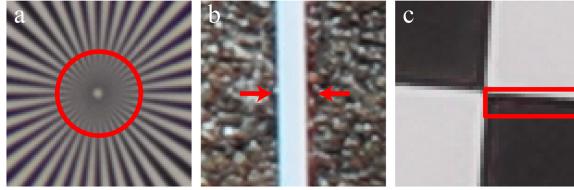


Fig. 2. Typical optical aberration degradation appearing on photos. From a to c: blurring(Canon EOS 80D + EF 50mm f/1.8 STM), chromatism (Canon EOS 80D + EF 50 mm f/1.2 L USM), and “ringing effect” (HUAWEI HONOR 20). We mark typical degradation with red indicators so that readers can spot them quickly.

digital-single-lens-reflex (DSLR) cameras, etc.) adopt many different approaches. For example, the professional large aperture lens replaces ordinary glass with fluorite, and the optical stack of HUAWEI HONOR 20 has more than eight extended aspherical elements, whose aspheric coefficients are up to order 16 [Zhenggang 2016].

In addition to the optical lens module, ISP algorithms also make a large difference in optical aberration correction. However, as shown in Figure 2(a) and Figure 2(b), the blur and chromatic dispersion vary greatly when the optical system changes. Moreover, the current classical ISP systems still cannot handle these issues completely, therefore introducing the “ringing effect” (see Figure 2(c)). These defects are very common in high-contrast photos from many recent flagship devices.

Over the past decade, methods based on the deep learning model have been evolving rapidly and have the potential to replace conventional ISP systems [Ignatov et al. 2020]. The classical postprocessing system is a step-by-step process. The optical aberrations of lens and the error of noise diagnosis will accumulate and be amplified in the subsequent **white balance (WB)**, **color correction matrix (CCM)**, gamma operations, and so on [Heide et al. 2014]. Nevertheless, end-to-end deep learning methods do not suffer from this problem [Abdelhamed et al. 2018]. Although deep learning networks allow us to restore latent images from degraded images, the learned models always achieve unsatisfactory results when evaluating real data [Cai et al. 2020]. Therefore, generating targeted datasets that follow the distribution of real scenes is the principal part of the deep learning method [Ji et al. 2020]. In recent years, many scholars have gained insight into this field. Their proposed approaches are characterized by a considerable amount of shooting and alignment [Ignatov et al. 2017]. However, a large quantity of training data is required to prevent overfitting, which conversely exacerbates the workload [Peng et al. 2019]. Furthermore, because each imaging system exhibits entirely different characteristics, adapting a **Convolutional Neural Network (CNN)**-based algorithm to a new camera may require capturing many new data pairs [Ignatov and Timofte 2019]. Therefore, it is imperative to design a brand new method for dataset generation that combines optical parameters and is easy to transfer to other cameras.

In this work, we propose a practical method for recovering the degradation caused by optical aberrations. To achieve this, we establish an imaging simulation system based on raytracing and coherent superposition [Paul 2014]. Given the lens data and parameters of the sensor, it works out the imaging result of a definite

object at a fixed distance. With datasets established by the imaging simulation system, we design a postprocessing chain, which combines **field-of-view (FOV)** information and optical parameters to perform adaptive restoration on the degraded data [Heide et al. 2014]. The proposed method outperforms many state-of-the-art deblurring methods on synthetic data pairs, as does the evaluation. When evaluated on real images, our proposed technique achieves better visual results (see in Figure 1) and requires lower computational overhead. Moreover, our approach is convenient for transferring to other commercial lenses, and there are no limitations on the optical design.

Our main technical contributions are as follows:

- An optical point spread function model is proposed for accurate **Point Spread Function (PSF)** calculation in a simulation manner.
- Based on the ISP algorithms, an imaging simulation system is established. Given the camera’s optical parameters, it generates targeted data pairs following the distribution of lens optical aberrations.
- A novel spatial-adaptive CNN architecture is proposed, which introduces FOV input, deformable ResBlock, and context block to adapt to spatially variant degradation.

This article proceeds as follows. In Section 2, we review the related works. In Section 3, we present an overview of our methods. In Section 4, we detail the proposed optical PSF model and the composition of the imaging simulation system. In Section 5 and Section 6, we present a brand new postprocessing pipeline. In Section 7, we compare our approach’s improved performance with the state-of-the-art methods and perform an ablation study on each aspect of our approach. In Section 8, we assess the proposed approach and demonstrate its potential applications.

2 RELATED WORK

Optical Aberration. According to Fermat’s principle, rays of light traverse the path of stationary optical path length with respect to variations in the path [Westphal et al. 2002]. Therefore, when the incident angle of rays changes, so does the optical length [Smith 2005]. Generally, the distinction of optical length manifests as weird blur and pixel displacement, which becomes more severe with increasing numerical aperture and FOV [Smith 2008]. While limited by the requirements of physical space, optical parameters, and machining tolerance, it is difficult for classical lens design to suppress optical aberrations [Sliusarev 1984].

Point Spread Function Model. A large body of work in this field has proposed calibrating and measuring PSFs from a single image [Ahi 2017; Hirsch and Scholkopf 2015; Jemec et al. 2017; Shih et al. 2012]. These methods often estimate optical blur functions or chromatic aberrations either blindly or through a calibration process. Moreover, recent works propose developing a machine learning framework to estimate the underlying physical PSF model from the given camera lens [Herbel et al. 2018]. However, these approaches have several limitations in operation, such as the noise level and the field interval of calibration or estimation [Heide et al. 2013]. The closest related works to us are those approaches that compute PSFs with lens prescriptions [Shih et al. 2012]. Given the focal

length, aperture, focusing distance, and white balance information, these methods simulate the image plane PSFs of each virtual object point.

Unfortunately, the PSFs calculated by their method are different from the measured PSFs (after ISP). The reasons for mismatching are summarized as manufacturing tolerances and fabrication errors during lens production [Shih et al. 2012]. However, we hold different views from them. We believe that the PSFs calculated by lens prescription represent the diffusion of energy, which are essentially different from the measured PSFs after ISP [Paulin-Henriksson et al. 2008]. And the complex amplitude of each ray must be considered while calculating the PSFs. Therefore, coherent superposition must be conducted after raytracing. To make a comprehensive comparison to prove the accuracy of the computed energy PSFs, we apply the calculated PSFs to the checkerboard chart in the energy domain and then transform the degraded image to a visualized image. The similarity of the real images and the synthetic results illustrates that it is a practical method for simulating PSFs with the advantages of lens prescriptions if the process of imaging is properly modeled.

Image Reconstruction. The optical aberration correction of a real image can be regarded as a problem of image reconstruction, where the goals are to increase sharpness and restore detailed features of images. Traditional deconvolution methods [Heide et al. 2013; Pan et al. 2016; Sun et al. 2017] use various natural image priors for iterative and mutual optimization, which unfortunately are not robust and have poor computational efficiency when handling large spatially variant blur kernels. The point spread functions of large FOV lenses vary from 11 to 53 pixels, which presents a great challenge to existing deconvolution methods, necessitating a custom image reconstruction approach. In recent years, a large number of efficient solutions have been proposed to use deep learning models for image processing tasks [Agustsson and Timofte 2017; Nah et al. 2017; Schuler et al. 2015; Vedaldi and Fulkerson 2010], starting from the simplest CNN approaches [Dong et al. 2015; Kim et al. 2016; Shi et al. 2016; Sun et al. 2015] to deep residual and attention models [Ratnasingam 2019; Wang et al. 2019; Zhang et al. 2019] and complex **generative adversarial networks (GANs)** [Ignatov et al. 2020; Ledig et al. 2017; Sajjadi et al. 2017; Wang et al. 2018]. More recently, Peng et al. [2019] proposed a learned generative reconstruction model, a single deep Fresnel lens design tailored to this model, obtaining pleasant results.

All of these approaches have in common that they require either accurate PSF estimation [Heide et al. 2013] or large training data that have been manually acquired [Ignatov et al. 2020]. Moreover, when the lens design changes, painful acquisition is still needed to obtain targeted image pairs, which include shooting, registration, and color correction [Peng et al. 2019]. In contrast, our proposed imaging simulation system generates a large number of training corpora with encoded optical information. Moreover, our approach is practical and easy to transfer to other lens designs. Our network model combines the optical parameters and FOV information, allowing us to efficiently address the large scene-dependent blur and color shift of commercial lens design.

Mobile Image Signal Processing. There exist many classical approaches for various image signal processing subtasks, such as image demosaicing [Dubois 2006; Hirakawa and Parks 2005; Li

et al. 2008], denoising [Buades et al. 2005; Chang et al. 2020; Condat 2010; Foi et al. 2008], white balancing [Gijsenij et al. 2011; Schwartzburg et al. 2014; Van De Weijer et al. 2007], and color correction [Kwok et al. 2013]. Only a few works have explored the applicability of combining optical parameters with imaging signal processing systems. Recently, some deep learning methods were proposed to map low-quality smartphone RGB/RAW photos to superior-quality images obtained with a high-end reflex camera [Ignatov et al. 2017]. The collected datasets were later used in many subsequent works [Ignatov et al. 2018, 2020; Mei et al. 2019], which achieved significantly improved results. Additionally, in Agustsson and Timofte [2017] and Ignatov and Timofte [2019], the authors examined the possibility of running image enhancement models directly on smartphones and proposed several efficient solutions for this task.

It is worth mentioning that all these datasets' construction methods require a considerable amount of shooting and registration. Moreover, owing to the different perspectives of data pairs, the resulting image enhancement model will introduce a small displacement to the input data, which is different from lens distortion correction. Additionally, these methods map the degradation images shot by a poor imaging system to the photos shot by a better imaging system; in other words, the upper bound of recovery depends on the quality of the ground truth. In contrast, the quality of the ground truth is also important but not critical to our method, because we encode the optical aberration degradation in the PSFs instead of the image pairs. And our approach generates data in a simulation manner, which liberates us from many duplicate acquisitions.

3 OVERVIEW

The goal of our work is to propose a practical method for recovering the degradation caused by optical aberrations. To achieve this, we engage optical parameters with a brand new postprocessing chain (as shown in Figure 3(b)). The two core ideas behind the proposed method are as follows: First, to eliminate the blurring, displacement, and chromatic aberrations caused by lens design, we design a spatially adaptive network architecture and insert it into the postprocessing chain. Second, for the sake of portability, we design an imaging simulation system for dataset generation, which simulates the imaging results of a specific camera and generates a large training corpus for deep learning methods.

Therefore, we rely on the proposed optical PSF model to establish the imaging simulation system. Different from the traditional raytracing or **fast Fourier transform (FFT)** methods to calculate PSF, we regard each ray as a secondary light source. The complex amplitude of each ray is superposed on the image plane; therefore, the PSF is the intensity of complex amplitude after superposition. In this way, the proposed method determines the degenerative consequences of object points. Figure 8 shows the spatially variant point spread functions of the customized DSLR camera lens and HUAWEI HONOR 20, which are computed following our optical PSF model. After calculating the PSFs of the lens, we establish an imaging simulation system based on the ISP pipeline of the digital camera. The simulation pipeline is detailed in Figure 5. To the best of our knowledge, we are the first to generate datasets of optical

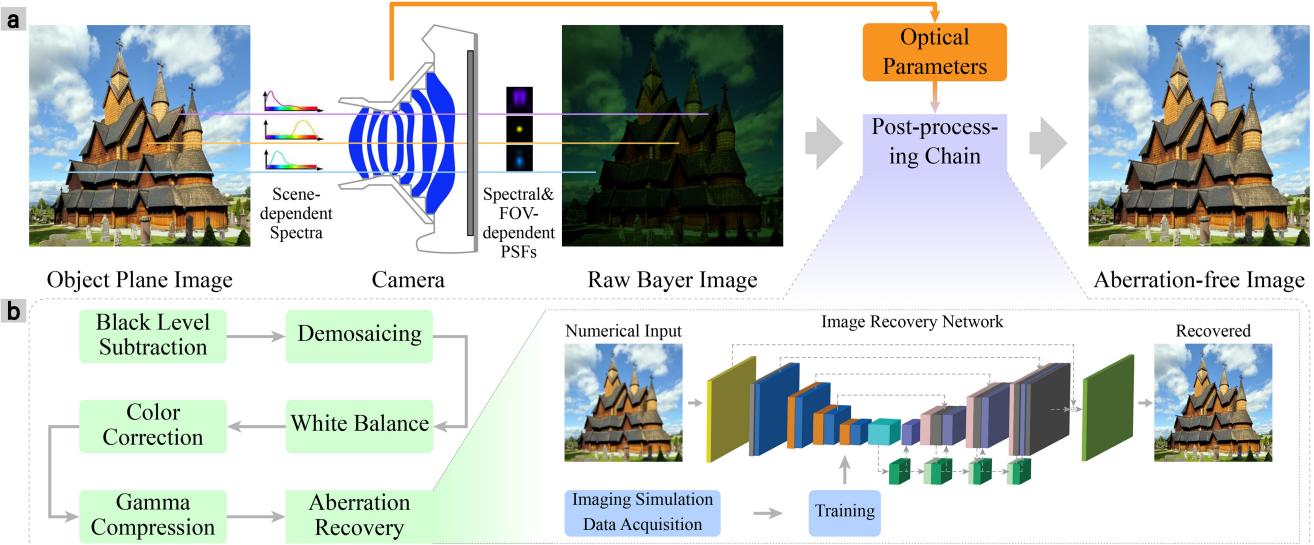


Fig. 3. Overview. The image formation model and the postprocessing chain are sketched in (a). In the image formation process, the camera lens encodes spectrally and field-of-view dependent PSFs in the resulting raw Bayer image. In the postprocessing chain, the optical parameters of the lens are applied to decode the spatially variant blurring and chromatic aberrations. The proposed postprocessing chain is detailed in (b). It jointly combines synthetic data acquisition, spatial-adaptive network architecture and optical parameters to recover latent images.

aberrations in a simulation manner, which do not need a great deal of shooting, registration, and color correction. We emphasize that our simulation method is easy to transfer to other lenses and has no limitations on optical designs.

Because of the spatially variant characteristics of optical aberrations, we design a brand new spatial-adaptive network architecture that takes the degraded image and FOV information as the input to directly reconstruct the aberration-free image. Together with the datasets generated by the imaging simulation system and the newly proposed postprocessing chain (Figure 3(b)), we can address the spectral and FOV-dependent blurring, color shift, and sharpness loss caused by lens aberrations.

4 IMAGING SIMULATION SYSTEM

In this section, we start by introducing the proposed optical PSF model, which consists of raytracing and coherent superposition. Second, we illustrate the relative illumination distribution and sensor spectral response calculation, which are used to connect the separate PSFs of different wavelengths to the accurate energy PSFs. Third, based on the ISP systems of modern digital cameras, we design an imaging simulation pipeline (Figure 5), which is applied to the input images, and generate the results degraded by PSFs depending on the spectral distribution and FOV information.

4.1 Optical PSF Model

For photographic objective systems with large FOVs and apertures, it is inaccurate to derive the PSF analytic solution by Gaussian paraxial approximation [Born and Wolf 2013] or FFT [Manual 2009] (we recommend readers to supplementary materials for relevant information). In other words, for various commercial optical designs, sequential raytracing and coherent superposition are the

only universal methods to obtain precise point spread functions on the imaging plane.

In the image formation process, every point of the object can be regarded as a coherent light source S . For a particular light source S , a grid of rays is launched through the optical system, which forms the wavefront Σ at the exit pupil. According to the Huygens principle, every point on a wavefront is the source of spherical wavelets, and the secondary wavelets emanating from different points mutually interfere [Baker and Copson 2003; Paul 2014]. In other words, the diffraction intensity at any point on the image plane is the complex sum of all these wavelets, squared. Therefore, in our proposed optical PSF model, the PSF on the image grid is computed in two stages, the first stage uses the sequential raytracing method to compute the wavefront distribution at the exit pupil plane, and the second stage uses coherent superposition to output the PSF calculation.

4.1.1 Raytracing. The skew ray emitted by a coherent monochromatic light source S is a perfectly general ray, which can be defined by the starting point $\mathbf{O} = (x_0, y_0, z_0)$ of S , and by its normalized direction vector $\mathbf{D} = (X, Y, Z)$. We can express the ray $\mathbf{P} = (x, y, z)$ in parametric form, where t is the ray marching distance:

$$\mathbf{P} = \mathbf{O} + t\mathbf{D}, t \geq 0. \quad (1)$$

After defining a general ray, raytracing can be divided into two steps: The first step is to calculate the intersection point of light and the surface. The surface equation generally used in optical design can be formulated as follows:

$$z = f(x, y) = \underbrace{\frac{c\rho^2}{1 + \sqrt{1 - c^2\rho^2}}}_{\text{sphere part}} + \underbrace{A_2\rho^2 + A_4\rho^4 + \cdots + A_j\rho^j}_{\text{deformation terms}}, \quad (2)$$

where z is the longitudinal coordinate of a point on the surface, and $\rho = \sqrt{x^2 + y^2}$ is the distance from the point to z axis. c is the curvature of the spherical part. The subsequent terms represent deformations to the spherical part, with A_2, A_4 , and so on, as the constant of the second, fourth, and so on, power deformation terms.

We build the simultaneous equations of Equations (1) and (2) for solving the ray marching distance from the starting point \mathbf{O} to the next ray-surface intersection point. However, their solution cannot be directly determined by the higher-order equations. Therefore, our raytracing is accomplished by a series of approximations. In each approximation, we add a small increment to the ray marching distance and compute $\mathbf{P} = (x, y, z)$ with Equation (1) to obtain the coordinates of ray. Additionally, we substitute x and y into Equation (2) to acquire the z_{ref} , which is the longitudinal coordinate of the surface. The iteration is repeated until the disparity between z and z_{ref} is negligible. In this way, the intersection of light and the surface is computed.

The second step is to calculate the refraction of light on the medium surface. Following Snell's law, the refracted angle θ_t can be computed by:

$$\sin^2 \theta_t = \left(\frac{\eta_1}{\eta_2} \right)^2 (1 - \cos^2 \langle \mathbf{n}, \mathbf{D} \rangle), \quad (3)$$

Here, \mathbf{n} is the normal unit vector of the surface equation, and η_1 and η_2 are the refractive indexes on both sides of the surface. $\cos \langle \cdot, \cdot \rangle$ is the operation of computing the cosine value of two vectors. In this way, the refracted direction vector \mathbf{D}' can be calculated by

$$\mathbf{D}' = \frac{\eta_1}{\eta_2} \mathbf{D} + \left(\frac{\eta_1}{\eta_2} \cos \langle \mathbf{n}, \mathbf{D} \rangle - \sqrt{1 - \sin^2 \theta_t} \right) \mathbf{n}, \quad (4)$$

In the implementation, we perform sequential raytracing for every sample position (x', y') on the pupil plane. The direction vector $D_{x'y'}$ and the total optical path length $l_{x'y'}$ are recorded. We note that every object point can be regarded as a coherent light source S , so the diffusion of energy is the superposition result of the coherent wavelet (as demonstrated in Section 4.1). In the following section, we introduce the details of coherent superposition.

4.1.2 Coherent Superposition. After tracing a grid of rays that is uniformly sampled on the pupil, we obtain the wavefront distribution at the exit pupil plane. Figure 4 shows the wavefront distribution Σ of converging light at the virtual exit pupil plane. We note that each ray of this wavefront distribution represents a spherical wavelet with a particular amplitude and phase, so the complex amplitude of Huygens' wavelet in the pupil plane can be represented by

$$E(l_{x'y'}) = a_0 \frac{e^{ikl_{x'y'}}}{l_{x'y'}}, \quad (5)$$

where $l_{x'y'}$ is the length of total optical path from (x_0, y_0, z_0) to (x', y', z') , where x' and y' are the spatial coordinates of the pupil plane. a_0 is the amplitude of the spherical wave at the unit distance from the source. $k = 2\pi/\lambda$ is the wave number of the light.

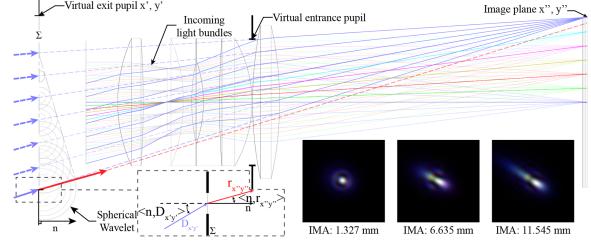


Fig. 4. Coherent superposition. Each position uniformly sampled on wavefront distribution Σ is a spherical wavelet, which coherently superposes on the image plane. In other words, the spherical wavelets interfere with each other when reaching the image plane. In the lower-right corner, we show the different fields' PSFs of customized DSLR camera lenses computed by raytracing and coherent superposition.

The complex amplitude propagating to the image plane can be formulated as follows:

$$E(l_{x'y'}, r_{x''y''}, K) = a_0 \frac{e^{ikl_{x'y'}}}{l_{x'y'}} \frac{e^{ikr_{x''y''}}}{r_{x''y''}} K(D_{x'y'}, r_{x''y''}, n), \quad (6)$$

where the $r_{x''y''} = (x'' - x', y'' - y', z'' - z')$, and $K(D_{x'y'}, r_{x''y''}, n)$ is the obliquity factor of the spherical wavelet, which is defined as follows:

$$K(D_{x'y'}, r_{x''y''}, n) = \frac{1}{2} [\cos \langle \mathbf{n}, \mathbf{r}_{x''y''} \rangle - \cos \langle \mathbf{n}, \mathbf{D}_{x'y'} \rangle], \quad (7)$$

where \mathbf{n} is the normal unit vector of the exit pupil plane and $\cos \langle \cdot, \cdot \rangle$ is the operation of computing the cosine value of the two vectors. The relationships of \mathbf{n} , $D_{x'y'}$, and $\mathbf{r}_{x''y''}$ are magnified in Figure 4.

Therefore, the complex amplitude $\tilde{E}_{x''y''}$ at the point (x'', y'') on the image surface is the complex sum of all the wavefront distributions propagated from the exit pupil plane, which can be formulated as the following equation:

$$\tilde{E}_{x''y''} = \sum_{y'} \sum_{x'} E(l_{x'y'}, r_{x''y''}, K(D_{x'y'}, r_{x''y''}, n)), \quad (8)$$

The diffraction intensity $I_{x''y''}$ can be expressed by:

$$I_{x''y''} = \tilde{E}_{x''y''} * \tilde{E}_{x''y''}^*. \quad (9)$$

Here the corresponding elements of the complex amplitude matrix $\tilde{E}_{x''y''}$ and $\tilde{E}_{x''y''}^*$ are multiplied. $\tilde{E}_{x''y''}^*$ is the complex conjugate of $\tilde{E}_{x''y''}$, and $I_{x''y''}$ is the resulting PSF matrix. Algorithm 1 outlines all the steps of the optical PSF calculation. To keep the article reasonably concise, a brief description of the raytracing is provided in Algorithm 1. We refer the readers to the work by Smith [2008] for an in-depth discussion about raytracing. Moreover, computing the optical PSF of the whole image plane is meaningless, because the complex amplitudes of many sampled points are almost zero. Therefore, as shown in Algorithm 1, we first trace the main ray to acquire the imaging center (x_c'', y_c'', z_c'') on the image plane and then perform the optical PSF calculation within the sampling range $x_r'' \times y_r''$.

Altogether, we illustrate each step of the optical PSF model in Section 4.1.1 and Section 4.1.2. Different from traditional raytracing or FFT [Heckbert 1995] to calculate PSF, we regard each light arriving at the pupil plane as a secondary wavelet. Because they

ALGORITHM 1: Optical PSF model

Input: coordinates of object point (x_0, y_0, z_0) , coordinates of virtual exit pupil z' , coordinates of image plane z'' , sample range on virtual exit pupil $x'_r \times y'_r$, sample interval of virtual exit pupil τ' , sample range on image plane $x''_r \times y''_r$, sample interval of sampling range at image plane τ'' , normal unit vector of virtual exit pupil \mathbf{n}

Output: PSF matrix $I_{x''y''}$

- 1 Generate an $\text{ceil}(\frac{x''}{\tau''}) \times \text{ceil}(\frac{y''}{\tau''})$ PSF matrix $I_{x''y''}$;
- 2 Generate an $\text{ceil}(\frac{x''}{\tau''}) \times \text{ceil}(\frac{y''}{\tau''})$ complex amplitude matrix $\tilde{E}_{x''y''}$;
- 3 Compute the intersection (x''_c, y''_c, z'') of the main ray and the image plane by raytracing;
- 4 Initialize $\tilde{E}_{sum} = 0 + 0i$;
- 5 **for** $x'' = x''_c - \frac{x''}{2}$ to $x''_c + \frac{x''}{2}$, step: τ'' **do**
- 6 **for** $y'' = y''_c - \frac{y''}{2}$ to $y''_c + \frac{y''}{2}$, step: τ'' **do**
- 7 **for** $x' = -\frac{x'_r}{2}$ to $\frac{x'_r}{2}$, step: τ' **do**
- 8 **for** $y' = -\frac{y'_r}{2}$ to $\frac{y'_r}{2}$, step: τ' **do**
- 9 Compute optical path length $l_{x'y'}$ and direction cosines $D_{x'y'}$ by ray tracing from point (x_0, y_0, z_0) to point (x', y', z') ;
 $r_{x''y''} = (x'' - x', y'' - y', z'' - z')$;
calculate $K(D_{x'y'}, r_{x''y''}, \mathbf{n})$ using Equation (7);
calculate $E(l_{x'y'}, r_{x''y''}, K)$ using Equation (6);
 $\tilde{E}_{sum} += E(l_{x'y'}, r_{x''y''}, K)$;
- 10 **end**
- 11 **end**
- 12 $\tilde{E}_{x''y''} = \tilde{E}_{sum}$;
- 13 $\tilde{E}_{sum} = 0 + 0i$;
- 14 **end**
- 15 **end**
- 16 **end**
- 17 **end**
- 18 **end**
- 19 **end**
- 20 $I_{x''y''} = \tilde{E}_{x''y''} \cdot * \tilde{E}_{x''y''}^*$;
- 21 **return** $I_{x''y''}$;

are emitted by a coherent object point S , the wavelets will interfere with each other when the wavefront propagates to the image plane. Therefore, the PSF of a coherent object point S is the diffraction result of the pupil plane, where the complex amplitude of each sampled wavelet must be considered. In the lower-right corner of Figure 4, we show the different FOVs' energy PSFs computed by our method. However, because of the ISP systems in digital cameras, the computed optical PSF, which represents the energy diffusion on the imaging plane, cannot be directly used for imaging simulation. Therefore, in the following section, we illustrate how to integrate optical PSF into the imaging simulation system.

4.2 Wavelength Response and Relative Illumination

Each pixel in a conventional camera sensor is covered by a single red, green, or blue color filter [Brooks et al. 2019]. These color filters have different responses to light waves of different

wavelengths, which can be formulated as $C_{wav}(\lambda)$. Moreover, in the imaging formation process, the off-axis field will suffer a loss of illumination, which is caused by lens shading and can be formulated as $C_{ill}(fov, \lambda)$. Therefore, for a precise simulation, the PSFs of different wavelengths and FOVs are multiplied by the specific distribution of wavelength responses and lens shading parameters. The influence of the wave response and relative illumination can be easily modeled by a coefficient:

$$C_e(fov, \lambda) = C_{ill}(fov, \lambda) \cdot C_{wav}(\lambda). \quad (10)$$

Here fov is the normalized FOV, and λ is the wavelength of object point S .

4.3 Imaging Simulation Pipeline

Owing to the differences between the dynamic range of raw sensor data and the sensitivity of the humans eye, imaging signal processing systems are typically used to convert the raw sensor energy data to photographs [Hasinoff et al. 2016]. In our imaging simulation pipeline, we transform the natural object image L to the energy domain data L_e with the energy domain transformation, which is shown in Figure 5. Specifically, for a general natural image, we first perform gamma decompression with the standard gamma curve, while fixing the input to the gamma curve with $\epsilon = 10^{-8}$ to prevent numerical instability. We refer the readers to the work by Brooks et al. [2019] for a more in-depth discussion of gamma decompression. After gamma decomposition, we convert the sRGB image to camera RGB data with the inverse of the fixed 3×3 CCM. The CCM is determined by the camera sensor. Then, the WB of the image is inverted by dividing the white balance value at random color temperature. The WB values of different color temperatures are obtained by calibration. In this way, we obtain the energy domain image L_e .

Based on the optical PSFs $I_{x''y''}$ calculated in Section 4.1, we degrade the energy domain image L_e with partitioned convolution. Partitioned convolution can be divided into three parts: First, we crop the image into uniform small pieces, then these pieces are convoluted with the PSFs of the corresponding FOVs, and finally the degraded pieces are spliced together. It is worth mentioning that the optical PSF calculated in Algorithm 1 is the PSF of one point (x_0, y_0, z_0) . To precisely simulate the imaging results, we uniformly sample the object plane and perform the PSF calculation in Algorithm 1 on every sampled point. The overall PSF can be represented by $I(h, w, \lambda)$. Here, h, w are the coordinates on the sensor plane (corresponding to the sampled point on the object plane), and λ is the wavelength of the light source. For a captured image with $n \times m$ resolution, let $J_e, L_e, N_e \in \mathbb{R}^{n \times m}$ be the observed energy image, the underlying sharp image, and the measurement noise image, respectively. The formation of the blurred observation J_e can then be formulated as

$$J_e(h, w) = \int C_e(fov, \lambda) \cdot [I(h, w, \lambda) * L_e(h, w)] d\lambda + N_e(h, w). \quad (11)$$

Here, fov can be calculated by (h, w) and the full FOV of the lens. It is important to note that J_e, L_e are rawlike images that indicate the energy received by the sensor.

After generating the blurred observation J_e with Equation (11), we invert the energy domain transformation by sequentially

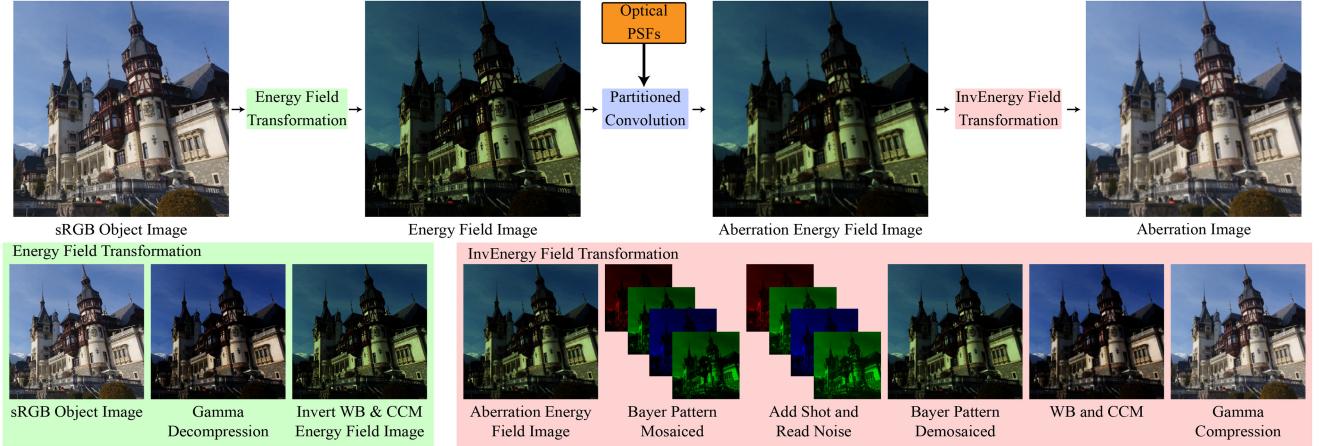


Fig. 5. Simulation pipeline of the imaging simulation system. The input image is converted to the energy domain first. Then, the optical PSFs reunited in Section 4.2 are applied to the energy domain image in a partitioning convolution manner. Finally, the degraded image is transformed into raw Bayer data and converted to aberration images. The details of the energy domain transformation and inverted energy domain transformation are shown at the bottom.

applying the operations in the lower-right corner of Figure 5. To be more specific, because the red, green, and blue color filters are arranged in a Bayer pattern, such as R-G-G-B, we omit two of its three color values according to the Bayer filter pattern of the camera. After acquiring the mosaiced raw energy data, we approximate the shot and read noise together as a single heteroscedastic Gaussian distribution and add it to each channel of the raw image [Brooks et al. 2019]. We refer the readers to Brooks et al. [2019] for details on synthetic noise implementation. Moreover, we sequentially apply the variation in the AHD demosaic algorithm [Lian et al. 2007], white balance, color correction matrix and the inverse of the gamma operator to J_e . The pipeline of the imaging simulation system is shown in Figure 5.

In conclusion, there are two significant differences between our proposed method and the imaging simulation methods that have been used in *Zemax* or *CODEV*. First, our method calculates the superposition of every spherical wavelet, where the phase term of the ray is considered. Therefore, the proposed method can generate more accurate PSFs when compared with the traditional raytracing or FFT methods. Second, we combine the sensor information and the ISP parameters into the imaging simulation system. In this way, our method not only simulates the imaging process from the physical process but also obtains very similar imaging results visually.

5 ABERRATIONS IMAGE RECOVERY

Due to the spatially variant properties of optical aberrations, a spatial-adaptive recovery framework is constructed to restore the degraded image. In this section, we first explain the network architecture and the additional input. Then we illustrate the deformable ResBlock that changes the shapes of convolutional kernels according to the related FOV information. Third, we introduce the context block inserted in the minimum scale of our model. The input used to train our network is the synthetic data that has been processed using the proposed imaging simulation system. In summary, this neural network model combines the optical parameters of specific lens designs, and has spatial adaptiveness to handle aberrations varying with the FOV.

5.1 Recovery Framework

We propose a novel network for the retrieval of the latent image L from corrupted data J . Because we generate the training data by simulation, the input and the ground truth of our network are pixel-to-pixel aligned. This characteristic greatly improves the stability of training. In addition, the aberrations have a high correlation with the FOV, and we compute the x and y coordinates of each pixel and concatenate the FOV image with the aberration inputs (as shown in Figure 6). This operation enables our network to directly perceive the FOV information, which can be regarded as another kind of natural prior. The proposed network is shown in Figure 6. Specifically, we adopt a variant of the UNet architecture [Ronneberger et al. 2015].

In regard to the loss function, the strictly paired data for supervised training allow us to use fidelity loss to learn robust texture details, thereby preventing fake information in the recovered image L . Specifically, the fidelity loss function is

$$L(\theta) = \frac{1}{N} \sum_{n=1}^N \|Net(I_n^{ab}) - I_n^{gt}\|_2^2, \quad (12)$$

where θ denotes the learned parameters in the network, I_n^{ab} are the aberration inputs, and I_n^{gt} are the corresponding ground truths of I_n^{ab} . We note that many loss functions (such as perceptual loss [Johnson et al. 2016] or GAN loss [Ledig et al. 2017]) will generate sharper visual results on natural images. These loss functions also introduce fake texture to the restorations and are not good at suppressing noise. Because the noise level of mobilephone varies greatly while illumination and color temperature are changing. Therefore, we adopt the MSELoss, which is robust to different noise levels.

5.2 Deformable ResBlock

When the field of view increases, the higher-order aberration coefficient deviates from that computed by the Gaussian paraxial model. The different sizes and shapes of off-axis PSFs are caused by these deviations. However, the traditional convolution

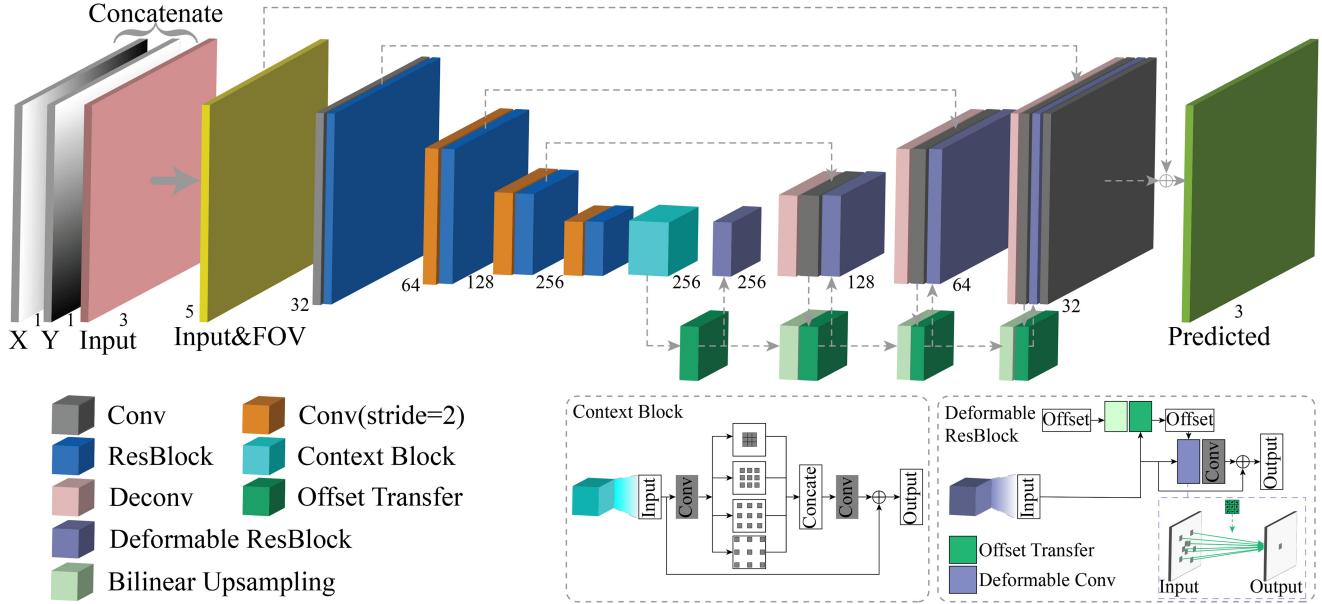


Fig. 6. The proposed neural network architecture. The UNet-based network architecture is shown on the top and the layer configurations of encoder/decoder stages are illustrated with different colored blocks (bottom left). The deformable ResBlock and the context block are detailed in the bottom-right corner. The number of channels for each scale is marked with each block. In particular, we choose MSELoss as the loss function when training, which evaluates the fidelity between the predicted image and the ground truth.

operation strictly adopts the feature of fixed locations around its center, which will introduce irrelevant features into the output and discard the information truly related to PSFs.

The size of optical PSFs ranges from 100 pixels to 1,600 pixels (squared), which means that our network must extract the features based on the FOV information. Therefore, we introduce the deformable convolution in the decoder path [Dai et al. 2017; Zhu et al. 2019]. While different from the method in Chang et al. [2020], we directly concatenate the input and the upsampled offset. After the concatenation, we refine the output offset with a convolution layer. In this way, the shape and the size of the convolution kernel can be changed adaptively according to the field-of-view information.

5.3 Context Block

Multiscale information is important for handling large PSFs. However, multiple downsamplings will increase the overhead of the network and destroy image structure. Therefore, we introduce a context block [Chang et al. 2020] into the minimum scale between encoder and decoder, which increases the receptive field and reconstructs multiscale information without further downsampling. As shown in Figure 6, several dilated convolutions are used to extract features from different receptive fields, which are concatenated later to estimate the output.

6 DATA PREPARATION

How to acquire targeted data pairs that follow the distribution of real scenes is the principle part of the supervised learning method. Our proposed method does not need any shooting or registration operation, which is a practical method for correcting the optical

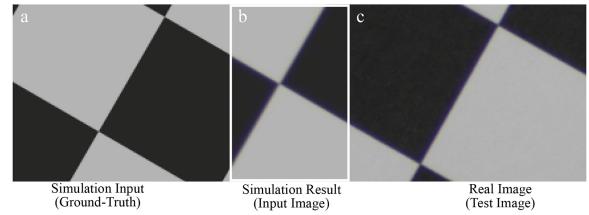


Fig. 7. Illustration of data preparation. In imaging simulation, (a) is the input, (b) is the synthetic result, and (c) is the real image taken by the DSLR camera. In network training, (a) is the ground truth, (b) is the input data and (c) is the image in the test datasets.

aberrations of various commercial lenses. In this section, we illustrate how we generate the data pairs of our networks.

6.1 Simulate Training Data

As shown in Figure 7, the simulation result of our simulation system and the corresponding digital image are the input and ground truth of our network, respectively. In the partitioned convolution operation in Figure 5, we calculate 240,000 (400×600) PSFs for the customized DSLR camera lens and 120,000 (300×400) PSFs for HUAWEI HONOR 20. The number of calculated PSFs is decided by the resolution of the sensor, i.e., the resolution of the sensor in the DSLR camera is $4,000 \times 6,000$, and we compute PSFs for every 10×10 patch. We adopt DIV2K [Martin et al. 2001], which contains 800 images of 2K resolution as the digital images to generate training data pairs.

In regard to the PSF calculation, we set the fixed object distance z_0 for each camera. The shooting distance is 1,750 mm for the DSLR camera, and it is 600 mm for HUAWEI HONOR 20. For the DSLR

camera, we compute 31 PSFs with wavelengths varying from 400 nm to 700 nm with a 10-nm interval. For HUAWEI HONOR 20, we calculate 34 PSFs varying from 400 nm to 730 nm. The wave distribution of PSFs is determined by the upper limit and the bottom limit of the sensor's wave response. Finally, the PSFs $I(h, w, \lambda)$ are reunited by the wave response functions of the sensor and lens shading parameters $C_e(fov, \lambda)$.

6.2 Generate Testing Images

To evaluate the effect of our trained network, we photograph 100 images with a DSLR camera and HUAWEI HONOR 20, respectively. After acquiring the raw data taken by the camera, we sequentially apply the steps in Figure 3(b) to obtain the sRGB image. These sRGB photographs make up the testing datasets. We note that the sharpen operation in classical ISP will affect optical aberration degradation, therefore, we use the self-defined postprocessing pipeline to transform raw data into sRGB images.

6.3 Training Details

In this way, we construct the data pairs for our network. In our network architecture, the channel numbers of each layer are marked at the bottom of the corresponding blocks (as shown in Figure 6). In terms of the hyperparameter of training, our settings are basically the same as in Chang et al. [2020]. In addition, the H, W of our training input is 256×256 , which allows the network to receive a large amount of spatially variant degradation.

7 ANALYSIS

In this section, we first analyze the PSF characteristics of the customized DSLR camera lens and HUAWEI HONOR 20 (*the extrinsic parameters of optical designs are listed in the supplementary file*). Second, comprehensive experiments are conducted to prove the accuracy of the imaging simulation system. Here, the experiments are divided into two parts: one part validates the accuracy of the calculated PSFs, and the other part compares the simulation results with the results generated by Zemax and the photograph taken by the camera. Third, we perform the ablation study of our network architecture. Finally, we compare the proposed method with the state-of-the-art deblurring methods, and the evaluation metrics are **Peak Signal-to-Noise Ratio (PSNR)** and **Structural Similarity (SSIM)**.

7.1 Optical PSF Analysis

The optical PSF distributions of the customized DSLR camera lens and HUAWEI HONOR 20 are sketched on the top of Figure 8, which are centrally symmetric and have a high correlation with the FOV. We note that the optical PSFs of HUAWEI HONOR 20 reveal a wide range of spatial variation over the region, while the PSFs of the customized DSLR camera are relatively consistent. Our network architecture takes advantage of these characteristics. It engages with the FOV information to perform spatial adaptive reconstruction for degraded images. It is worth mentioning that our method directly calculates the resulting PSFs, which is different from the optimization method of PSF estimation [Mosleh et al. 2015].

We plot the Strehl ratio [Mahajan 1982] curves and the measured patches in 0.1, 0.3, 0.7, and 0.96 FOVs of two optical designs,

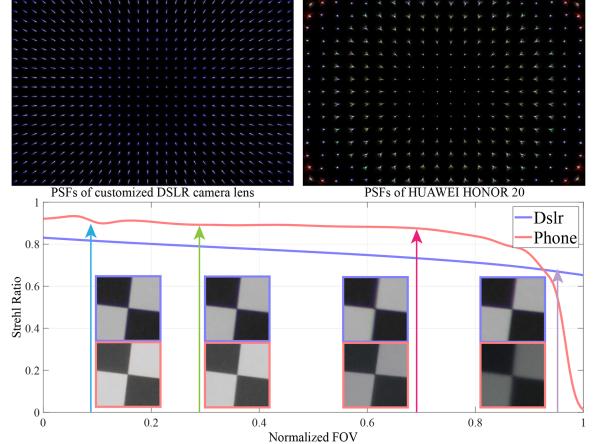


Fig. 8. PSFs characteristics. The patchwise PSF distributions of the customized DSLR camera lens and HUAWEI HONOR 20 are sketched on top. The bottom panel shows the Strehl ratio [Mahajan 1982] curves of the DSLR camera and HUAWEI HONOR 20. The real measurements of different FOVs are plotted by side, where the blue boxes show the DSLR results and the orange boxes show the images taken by HUAWEI HONOR 20. These results demonstrate the spatially variant characteristics of PSFs.

see the bottom of Figure 8. The Strehl ratio can be used to evaluate the energy diffusion in different FOVs, which has some correlation to the degradation of the image. As shown in Figure 8, the Strehl ratio turns lower from the center to the whole field, resulting in more blurred photography. Moreover, unlike the customized DSLR camera lens, the shape and size of PSFs in HUAWEI HONOR 20 abruptly change in regard to the 0.9 FOV, which is common to camera lenses. Thus, the PSF characteristics of different devices are totally different, which means that it is significant to generate targeted datasets for a specific lens.

7.2 Authenticity of Imaging Simulation System

To prove the accuracy of the proposed imaging simulation system, we designed a two-pronged experiment. One prong is to verify the accuracy of the calculated PSFs: First, we calibrate the PSFs of the DSLR camera with the virtual pointlike source [Jemec et al. 2017]. Then, the PSFs of the same camera are estimated by the optimization method used in Mosleh et al. [2015] (we use the chart of the noise pattern). Finally, we calculate the PSFs by Zemax's ray-tracing. The comparison is shown in Figure 9, and we postprocess all the PSFs by the same postprocessing operations for visualization. When the FOV increases, the PSFs estimated by Mosleh et al. [2015] (Figure 9(a)) are becoming more blurry and the PSNRs decrease greatly. We note that from the center to the edge of the image, the relative illumination decreases a lot. Therefore, the edge's results of Mosleh et al. [2015] are not ideal, because the method has difficulty converging when the degraded information is masked by noise. The marginal PSFs calibrated by Jemec et al. [2017] (Figure 9(b)) also suffer from this problem. Different from these methods that need to shoot targets, our method directly calculate the PSFs from the lens prescriptions and will not be affected by noise. By comparing the PSNRs of the proposed method and Zemax, we prove that our results are more accurate and more

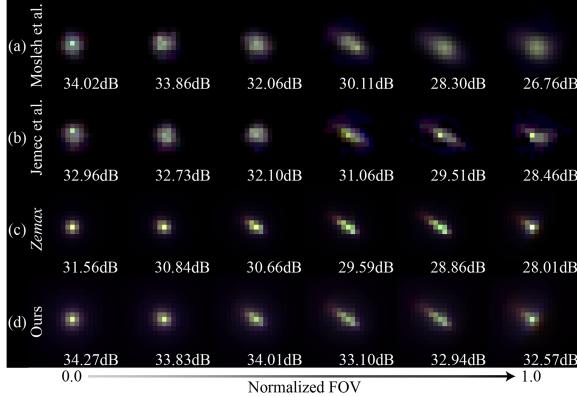


Fig. 9. Authenticity of PSFs. Here, we show the PSFs calculated by different methods, which are arranged from left to right according to the corresponding FOV. (a) shows the PSFs optimized by the method in [Mosleh et al. 2015] (we use the chart of the noise pattern). (b) is the PSFs calibrated by the virtual pointlike source [Jemec et al. 2017]. (c) is the PSFs calculated by *Zemax*®’s raytracing. (d) shows the results of the proposed method. The PSNRs of these PSFs’ imaging simulation (compared to the real image) are noted in the bottom-right corner (we use 100×100 checkerboard for the PSNR calculation).

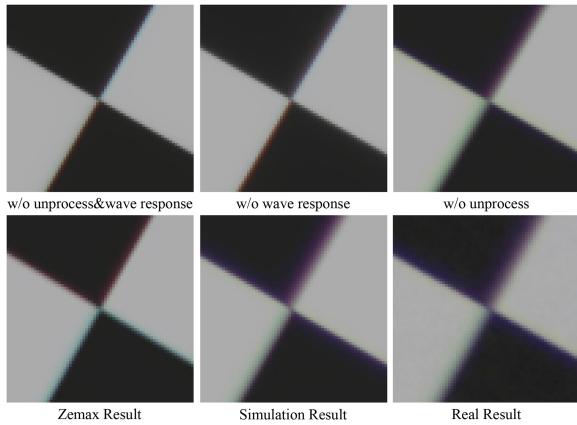


Fig. 10. Accuracy of Imaging Simulation. In the first row, we show the ablation study of the imaging simulation system. In the second row we compare our result with the image generated by *Zemax* and the data taken by the DSLR camera.

similar to the real image. Moreover, these results also validate the necessity of the coherent superposition.

Another method for verifying the authenticity of our method is to compare the simulation results. We shoot the checkerboard chart with the corresponding camera and simulate the imaging results with *Zemax*, where the parameters used are the same as those in our simulation. Moreover, we also perform an ablation study on the proposed imaging simulation system. The visual comparison of the results is shown in Figure 10. Obviously, the imaging results of our method is the closest to the real data. We note that the image generated by *Zemax* has no combination of the sensor’s spectral information and the ISP parameters, accurately.

Table 1. Quantitive Comparison of MTF Area

FOV	Siml MTF Area	Real MTF Area	Ratio
0.0	0.25421	0.24805	1.02898
0.3	0.24284	0.24285	0.99995
0.5	0.22133	0.22122	1.00052
0.7	0.22034	0.21804	1.01055
0.9	0.21156	0.20553	1.02934

We assess the MTF area, which is the area between the MTF curve and the cycle per pixel axis, for comparison. The ratio is the result of the simulation MTF area divided by the real MTF area.

This might be the reason that *Zemax* produces inaccurate results. In addition, we compare the **modulation transfer function (MTF)** [Schroeder 1981] in different FOVs, which shows the accuracy of the proposed method through the whole region of the sensor (as shown in Table 1).

In summary, we prove the authenticity of the proposed imaging simulation system mainly from two aspects. First, we demonstrate that our PSF calculation method is more accurate than the optimization method. When compared with the calibration methods, the proposed method is more practical and will not be affected by noise. The comparison between *Zemax* and the proposed method proves the accuracy of our method and validates the necessity of coherent superposition. Second, we compare the simulation results with the image of *Zemax*®’s simulation and the photograph of the camera, which verifies the authenticity of the imaging simulation system and every component is necessary.

7.3 Ablation Study

We perform the ablation study of our network architecture on the simulation dataset of the DSLR camera lens. The results are shown in Table 2, and the evaluation metric is PSNR.

Ablation on FOV Input. We concatenate the FOV image with the aberrations input image. It is obvious that the FOV image helps the network determine the spatially variant blurring and displacement of optical aberrations.

Ablation on Deformable ResBlock. Without the deformable ResBlock, the network loses the spatial adaptability to handle different PSFs of optical aberrations. If we replace the deformable ResBlock with ResBlock, then the performance of our networks significantly decreases, which indicates its effect.

Ablation on Context Block. The context block helps the network capture larger field information while not damaging the structure of the feature map. The comparison validates its effectiveness.

7.4 Recovery Comparison

Because of the spatial variation over aberrant PSFs, we design a brand new architecture to perform spatially adaptive recovery on the degraded image. In this section, we compare our algorithm with state-of-the-art deblurring methods.

7.4.1 Synthetic Aberrations Image. To evaluate the effectiveness of the proposed approach, we select the IRCNN [Zhang et al. 2017], GLRA [Ren et al. 2018], and Self-Deblur [Ren et al. 2020] as the representatives of the combination of model-based

Table 2. Ablation Study

Field-Of-View Input		✓			✓	✓		✓
Deformable ResBlock			✓		✓		✓	
Context Block				✓		✓	✓	
PSNR	40.14	43.72	40.63	40.38	44.36	44.01	41.23	44.77

The evaluation metric PSNRs are based on the simulation datasets for customized DSLR camera lens.

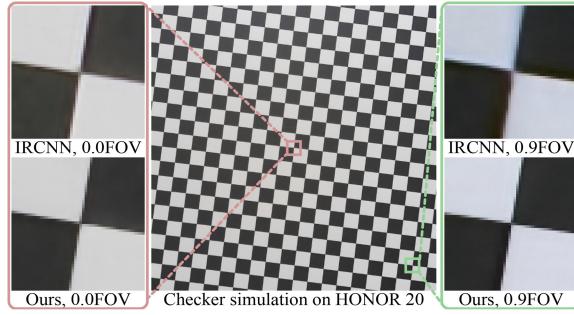


Fig. 11. Spatial adaptability of the proposed network. We compare the different FOV restorations of IRCNN and ours. Obviously, IRCNN is not effective in dealing with the blur varying with the space.

optimization methods and discriminative learning methods. Moreover, we choose some deep learning-based methods, such as SRN [Tao et al. 2018], for comparisons. For a fair comparison, all the state-of-the-art methods employ the default setting provided by the corresponding authors. It is worth emphasizing that all these comparison methods are retrained by the synthetic datasets of the DSLR camera (DSLR) and HUAWEI HONOR 20 (Phone).

To prove the spatial adaptability of the proposed approach, we compare the restoration results with other deblur methods when evaluating synthetic data. As shown in Figure 11, even though the Phone PSFs show a wide range of spatial variation over regions (as shown in Figure 8), the proposed method performs spatial-adaptive recovery on the degraded input of different FOVs. IRCNN, which is designed for globally consistent blur, generates imperfect recovery results at the border of the checker. Because the edge of the FOV is much more blurred than the center of the image, the globally consistent deblur approach will receive a compromise between the center and the edge. Table 3 shows the quantitative comparison. It is obvious that our network architecture outperforms the state-of-the-art methods on synthetic data. We attribute this to the additional FOV input and the deformable ResBlock, which endow the proposed architecture with spatial adaptability. Moreover, we retrain the state-of-the-art methods by uniformly separating all the data into 3×4 patches (overlapping with each other). The quantitative results are shown in Table 3. We note that the model-based optimization methods make a leap in PSNR and SSIM when data are partitioned. For DSLR datasets, the PSFs are relatively consistent, so the IRCNN and Self-Deblur can achieve better quantitative results for patches with relatively consistent blur. For Phone datasets, the PSFs show a wide range of spatial variation over regions, which means that higher restoration cannot be achieved by simply partitioning the data. The smaller size of the patch might improve the restored result, but so would the computation overhead. In summary, our proposed architecture efficiently integrates

Table 3. Quantitative Comparison of Synthetic Aberrations Correction

Dataset	Metric	SRN	GLRA	IRCNN	Self-Deblur	Ours
DSLR	PSNR	43.91	44.13	43.67	44.37	44.77
	SSIM	0.996	0.994	0.992	0.996	0.998
Phone	PSNR	36.28	35.94	36.05	36.22	37.85
	SSIM	0.980	0.978	0.975	0.978	0.986
DSLR	PSNR	44.28	44.34	45.07	45.04	44.96
	SSIM	0.997	0.995	0.997	0.999	0.998
Phone	PSNR	36.98	37.34	37.65	37.72	38.15
	SSIM	0.981	0.982	0.983	0.985	0.988

The compared methods are SRN, GLRA, IRCNN, Self-Deblur, and our proposed network model. We assess PSNR and SSIM to complete the comparisons.

the spatial information and will not occupy many more computing resources.

7.4.2 Real Testing Image. With the testing images taken by the camera (detailed in Section 6.2), we evaluate the proposed technique with state-of-the-art deblurring methods. As shown in Figure 12, we show two patches of each test image: One patch is near the center of the image, and the other patch is on the edge. We note that the PSFs of the DSLR camera are relatively consistent (as shown in Figure 8), so for the mainstream deblurring algorithms, which are used to deal with globally consistent blur, the restoration of the center and edge of the comparison methods are similar. When compared with other techniques, the results of the proposed approach in the central FOV are similar to those of other methods. In the edge of the image, our result of the DSLR camera is slightly better than those of the other methods. However, for the situation in which the PSFs show a wide range of spatial variation over regions, such as mobile phone cameras, the advantages of our approach are evident. In Figure 12, one can see that the proposed method is superior to all other methods for center and edge restoration. The reconstruction effectively removes the blurring caused by optical aberrations, while the spatial-adaptive network architecture successfully handles the spatially variant PSFs and preserves the texture of the original image. We note that some results of GLRA and IRCNN are somewhat comparable to the proposed method, while in general, our approach can obtain better results in different FOVs. It must be emphasized that the additional input is useful for optical aberration correction, because the degradation is highly correlated with the FOV information. Overall, we demonstrate the spatial adaptability of our network architecture. When compared with other methods, the proposed approach is more effective in dealing with the PSFs of spatial variation and therefore is suitable for optical aberration correction.

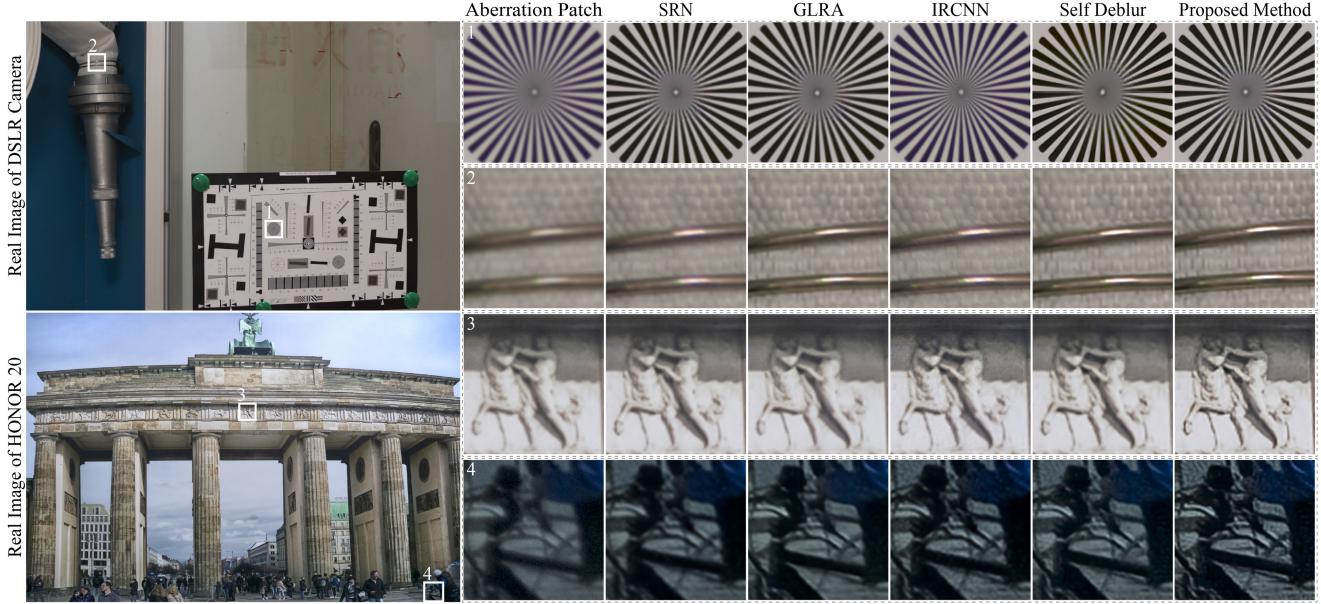


Fig. 12. Real-shoot image restoration comparison. The first row and the second row present the test results of the DSLR camera lens. The third and the fourth row are the results of HUAWEI HONOR 20. In the first column, we highlight the corresponding regions of aberration patches with white boxes.

8 EXPERIMENTAL ASSESSMENT

In this section, we take a comprehensive look at the proposed approach and demonstrate its potential applications. First, we assess how much the proposed method improves MTF, chromatismmm and line-pair resolution. Second, the restored results are compared with the image postprocessed by HUAWEI ISP, to evaluate the improvement in integrating aberration correction into an ISP system.

8.1 MTF and Line-pair per Inch Enhancement

We compare the chromatism and the MTF of degraded image and recovered image (shown in Figure 14). Though only trained with synthetic data, the proposed method can restore perfect knife edge and greatly improve the MTF results. Moreover, the pixel values of the R, G, and B channels of the reconstructed image are closer to each other, which indicates that our method successfully handles the chromatism between different channels. The lower restored **Chromatic Aberration (CA)** curve also proves that the shift between RGB channels is smaller. Moreover, the enhancement of line-pair per inch is shown in Figure 12, where the line-pair resolution is improved in the center of the ISO12233 chart. We note that the restored MTF results approach the diffraction limit of the lens, which is the physical limit resolution of an imaging device.

8.2 Aberrations Correction vs. HUAWEI ISP

The ultimate goal of the proposed work is to integrate the module of aberration correction into an ISP system and to produce nearly perfect images. In today's handcrafted ISPs, sharpening is the core method for alleviating the blurring caused by optical aberrations. Unfortunately, the sharpening operation is globally consistent and cannot solve the spatially variant blur. Therefore, it introduces unsolved blur in the boundary of FOV and "ringing" effect in the center of figure (as shown in Figure 2(c)). In Figure 13, we compare the

results of our proposed pipeline (as shown in Figure 3) and the images generated by HUAWEI ISP (*for more results, please refer to the supplementary materials*). Obviously, our proposed method adapts to the spatially variant degradation and the results are visually superior to HUAWEI's ISP. Together with accurate synthetic data and the customized restoration algorithm, we demonstrate that it is possible to handle optical aberrations in postprocessing. This also encourages us to use a lightweight lens for high-quality imaging.

9 DISCUSSION AND CONCLUSION

In this article, we demonstrated that it is feasible to correct optical aberrations in postprocessing systems. Based on the optical PSF model, the imaging simulation system was developed. Moreover, with synthetic data generated in simulation, we designed the spatial-adaptive network for effective aberration correction.

To be more specific, we constructed an imaging simulation system based on the optical PSF model and the ISP system in the digital camera. Although the lenses in the real world are ever-changing, accurate imaging results can be obtained if crucial parameters are given. In this way, we established the datasets for a specific camera lens. Moreover, to correct the aberrations of optical design, we proposed a spatial-adaptive CNN that combines FOV information to handle aberrations of different sizes and shapes. After that, trained with the simulation datasets, the learned network model was used to map the real degraded images to the restored results. We address spatially variant blurring and chromatism by introducing FOV information, deformable ResBlock, and context block. The proposed method achieves perfect image quality and is more effective in dealing with PSFs varying with space when compared with state-of-the-art deblur methods. Moreover, we successfully corrected the aberrations of the DSLR camera and HUAWEI HONOR 20, which proves that it is a practical method for

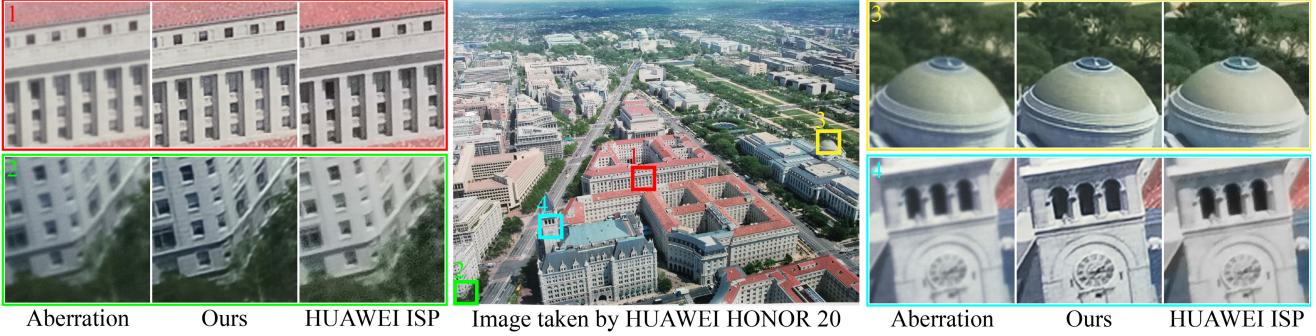


Fig. 13. Aberrations correction vs. HUAWEI ISP. Experimental results of the aberration correction pipeline and the HUAWEI ISP are amplified on both sides of the image. We show different FOVs' patches for validation.

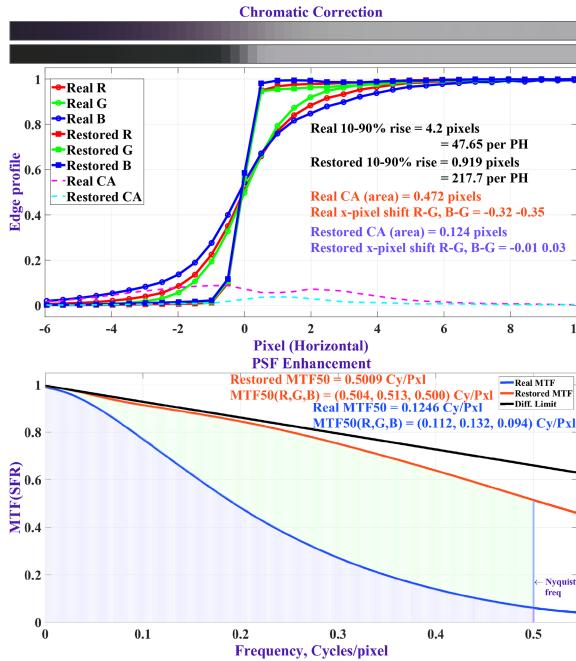


Fig. 14. Chromatic correction and MTF enhancement. The detected edge of the degraded image and the restored edge are plotted on the top. The RGB chromatic aberrations, the average MTF of the real image and reconstruction, and the diffraction limitation MTF are sequentially plotted below. We also show chromatic aberration curves to evaluate the dispersion of RGB channels.

correcting aberrations and easily transport them to other lenses. Finally, we demonstrate that it is feasible to integrate aberration correction with the existing ISP systems, which will bring great positive benefits.

We find that the introduction of optical information and spatial adaptive capability can result in better recovery, but today's devices still can't afford the computational overhead of our post-processing system. The dedicated NPUs and AI chips equipped on many modern imaging devices still limited to scene classification or postprocessing of a image with low resolution. Today's optical design and imaging postprocessing system are divisive and we strongly insist that the optical lens and ISP systems can be seen as

encoder and decoder, respectively. In this way, the lens-ISP joint design is possible for modern imaging systems. Although this work focuses on the postprocessing system of cameras, we envision that it is practical to jointly design optical lenses and ISP systems for aberration correction or many other tasks such as 3D visual reconstruction.

ACKNOWLEDGMENTS

We thank Huawei for their support. Finally, we thank the TOG reviewers for their valuable comments.

REFERENCES

- A. Abdelhamed, S. Lin, and M. S. Brown. 2018. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 694–711.
- Eirikur Agustsson and Radu Timofte. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE Computer Society, 126–135.
- Kiarash Ahi. 2017. Mathematical modeling of THz point spread function and simulation of THz imaging systems. *IEEE Trans. Terahertz Sci. Technol.* 7, 6 (2017), 747–754.
- Bevan B. Baker and Edward Thomas Copson. 2003. *The Mathematical Theory of Huygens' Principle*. Vol. 329. American Mathematical Society, Boston.
- Max Born and Emil Wolf. 2013. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Elsevier, Boston.
- Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T. Barron. 2019. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 11036–11045.
- Antoni Buades, Bartomeu Coll, and J.-M. Morel. 2005. A non-local algorithm for image denoising. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. IEEE, 60–65.
- J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang. 2020. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV'20)*. IEEE Computer Society, 15–463.
- Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. 2020. Spatial-adaptive network for single image denoising. 26, 10 (2020), 171–187. Retrieved from <https://arxiv.org/abs/2001.10291>.
- Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3291–3300.
- Laurent Condat. 2010. A simple, fast and efficient approach to denoising: Joint demosaicking and denoising. In *Proceedings of the 2010 IEEE International Conference on Image Processing*. IEEE, 905–908.
- Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. 2017. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*. IEEE, 764–773.
- Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2015. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 2 (2015), 295–307.

- Eric Dubois. 2006. Filter design for adaptive frequency-domain Bayer demosaicking. In *Proceedings of the 2006 International Conference on Image Processing*. IEEE, 2705–2708.
- Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans. Image Process.* 17, 10 (2008), 1737–1754.
- Arjan Gijssenij, Theo Gevers, and Joost Van De Weijer. 2011. Improving color constancy by photometric edge weighting. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 5 (2011), 918–929.
- Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph.* 35, 6, Article 192 (2016), 1–12. <https://doi.org/10.1145/2980179.2980254>
- Paul Heckbert. 1995. Fourier transforms and the fast Fourier transform (FFT) algorithm. *Comput. Graph.* 2 (1995), 15–463.
- Felix Heide, Mushfiqur Rouf, Matthias B. Hullin, Bjorn Labitzke, Wolfgang Heidrich, and Andreas Kolb. 2013. High-quality computational imaging through simple lenses. *ACM Trans. Graph.* 32, 5, Article 149 (oct 2013), 14 pages. <https://doi.org/10.1145/2516971.2516974>
- Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pajak, Dilipkumar Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, et al. 2014. Flexispis: A flexible camera image processing framework. *ACM Trans.* 33, 6, Article 231 (Nov. 2014), 13 pages. <https://doi.org/10.1145/2661229.2661260>
- Jörg Herbel, Tomasz Kacprzak, Adam Amara, Alexandre Refregier, and Aurelien Lucchi. 2018. Fast point spread function modeling with deep learning. *J. Cosmol. Astropart. Phys.* 2018, 07 (2018), 054. <https://doi.org/10.1088/1475-7516/2018/07/054>
- Keigo Hirakawa and Thomas W. Parks. 2005. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Trans. Image Process.* 14, 3 (2005), 360–369. <https://doi.org/10.1109/TIP.2004.838691>
- Michael Hirsch and Bernhard Scholkopf. 2015. Self-calibration of optical lenses. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 612–620.
- Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 3277–3285.
- Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2018. Wespe: Weakly supervised photo enhancer for digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 691–700.
- Andrey Ignatov and Radu Timofte. 2019. Ntire 2019 challenge on image enhancement: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 0–0.
- Andrey Ignatov, Luc Van Gool, and Radu Timofte. 2020. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 536–537.
- Jurij Jemec, Franjo Pernuš, Boštjan Likar, and Miran Bürmen. 2017. 2D sub-pixel point spread function measurement using a virtual point-like source. *Int. J. Comput. Vis.* 121, 3 (2017), 391–402.
- X. Ji, Cao Y., Tai Y., C. Wang, J. Li, and F. Huang. 2020. Real-world super-resolution via kernel estimation and noise injection. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW’20)*. IEEE, 15–463.
- Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision*. Springer, Berlin, 694–711.
- Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1646–1654.
- Ngai Ming Kwok, H. Y. Shi, Quang Phuc Ha, Gu Fang, S. Y. Chen, and Xiuping Jia. 2013. Simultaneous image color correction and enhancement using particle swarm optimization. *Eng. Appl. Artif. Intell.* 26, 10 (2013), 2356–2371.
- Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 4681–4690.
- Xin Li, Bahadir Gunturk, and Lei Zhang. 2008. Image demosaicing: A systematic survey. In *Proceedings of the Visual Communications and Image Processing 2008*, Vol. 6822. International Society for Optics and Photonics, 68221J.
- Nai-Xiang Lian, Lanlan Chang, Yap-Peng Tan, and Vitali Zagorodnov. 2007. Adaptive filtering for color filter array demosaicking. *IEEE Trans. Image Process.* 16, 10 (2007), 2515–2525.
- Virendra N. Mahajan. 1982. Strehl ratio for primary aberrations: some analytical results for circular and annular pupils. *J. Opt. Soc. Am.* 72, 9 (1982), 1258–1266.
- Zemax Manual. 2009. Optical Design Program User’s Guide.
- David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV’01)*, Vol. 2. IEEE, 416–423.
- Kangfu Mei, Juncheng Li, Jiajie Zhang, Haoyu Wu, Jie Li, and Rui Huang. 2019. Higher-resolution network for image demosaicing and enhancing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW’19)*. IEEE, 3441–3448.
- Ali Mosleh, Paul Green, Emmanuel Onzon, Isabelle Begin, and J. M. Pierre Langlois. 2015. Camera intrinsic blur kernel estimation: A reliable framework. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 4961–4968.
- Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3883–3891.
- Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. 2016. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1628–1636.
- Gunther Paul. 2014. *Huygens’ Principle and Hyperbolic Equations*. Academic Press, Venice.
- S. Paulin-Henriksson, A. Amara, L. Voigt, A. Refregier, and S. L. Bridle. 2008. Point spread function calibration requirements for dark energy from cosmic shear. *Astron. Astrophys.* 484, 1 (2008), 67–77.
- Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. 2019. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph.* 38, 6 (2019), 219–1.
- Sivalingeswaran Ratnasingam. 2019. Deep camera: A fully convolutional neural network for image signal processing. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. IEEE, 0–0.
- Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. 2020. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 3341–3350.
- Wenqi Ren, Jiawei Zhang, Lin Ma, Jinshan Pan, Xiaochun Cao, Wangmeng Zuo, Wei Liu, and Ming-Hsuan Yang. 2018. Deep non-blind deconvolution via generalized low-rank approximation. *Adv. Neural Inf. Process. Syst.* 31 (2018), 297–307.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 234–241.
- Mehdi S. M. Sajjadi, Bernhard Scholkopf, and Michael Hirsch. 2017. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*. Springer, 4491–4500.
- Manfred R. Schroeder. 1981. Modulation transfer functions: Definition and measurement. *Acta Acust.* 49, 3 (1981), 179–182.
- Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. 2015. Learning to deblur. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 7 (2015), 1439–1451.
- Yuliy Schwartzburg, Romain Testuz, Andrea Tagliasacchi, and Mark Pauly. 2014. High-contrast computational caustic design. *ACM Trans. Graph.* 33, 4 (2014), 1–11.
- Wenzhi Shi, José Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Springer, 1874–1883.
- Yichang Shih, Brian Guenter, and Neel Joshi. 2012. Image enhancement using calibrated lens simulations. In *Proceedings of the European Conference on Computer Vision*. Springer, 42–56.
- Georgii Georgievich Sluisarev. 1984. Aberration and optical design theory. *Annal. Henri Lebesgue* 17, 10 (1984), 1737–1754.
- Warren J. Smith. 2005. *Modern Lens Design*. American Mathematical Society, Boston.
- Warren J. Smith. 2008. *Modern Optical Engineering*. Tata McGraw-Hill Education, Boston.
- Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. 2015. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 769–777.
- Tianchang Sun, Yifan Peng, and Wolfgang Heidrich. 2017. Revisiting cross-channel information transfer for chromatic aberration correction. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 3248–3256.
- Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. 2018. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 8174–8182.
- Joost Van De Weijer, Theo Gevers, and Arjan Gijssenij. 2007. Edge-based color constancy. *IEEE Trans. Image Process.* 16, 9 (2007), 2207–2214.
- Andrea Vedaldi and Brian Fulkerson. 2010. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM International Conference on Multimedia*. IEEE, 1469–1472.
- Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. 2019. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 0–0.

- Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV'18)*. IEEE, Venice, 0–0.
- Volker Westphal, Andrew M. Rollins, Sunita Radhakrishnan, and Joseph A. Izatt. 2002. Correction of geometric and refractive image distortions in optical coherence tomography applying Fermat's principle. *Opt. Express* 10, 9 (2002), 397–404.
- Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. 2019. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1751–1760.
- Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. 2017. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3929–3938.
- Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. 2019. Zoom to learn, learn to zoom. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3762–3770.
- L. I. Zhenggang. 2016. Periscope lens and terminal device. US Patent 9,523,847.
- Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. 2019. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 9308–9316.

Received September 2020; revised June 2021; accepted June 2021