Annular Computational Imaging: Capture Clear Panoramic Images Through Simple Lens

Qi Jiang, Hao Shi, Graduate Student Member, IEEE, Lei Sun, Shaohua Gao, Kailun Yang, and Kaiwei Wang

Abstract—Panoramic Annular Lens (PAL) composed of few lenses has great potential in panoramic surrounding sensing tasks for mobile and wearable devices because of its tiny size and large Field of View (FoV). However, the image quality of tiny-volume PAL confines to optical limit due to the lack of lenses for aberration correction. In this paper, we propose an Annular Computational Imaging (ACI) framework to break the optical limit of light-weight PAL design. To facilitate learning-based image restoration, we introduce a wave-based simulation pipeline for panoramic imaging and tackle the synthetic-to-real gap through multiple data distributions. The proposed pipeline can be easily adapted to any PAL with design parameters and is suitable for loose-tolerance designs. Furthermore, we design the Physics Informed Image Restoration Network (PIRNet) considering the physical priors of panoramic imaging and single-pass physics-informed engine. At the dataset level, we create the DIVPano dataset and the extensive experiments on it illustrate that our proposed network sets the new state of the art in the panoramic image restoration under spatially-variant degradation. In addition, the evaluation of the proposed ACI on a simple PAL with only 3 spherical lenses reveals the delicate balance between high-quality panoramic imaging and compact design. To the best of our knowledge, we are the first to explore Computational Imaging (CI) in PAL.

Index Terms—Computational imaging, optical aberration, image restoration, panoramic annular lens.

I. INTRODUCTION

MAGES with very large Field of View (FoV) are attracting more attention on account of more abundant information they provide about the surrounding environment for some Computer Vision (CV) tasks [1]. A considerable amount of research is carried out on semantic segmentation [2], [3], [4], object detection [5], [6], visual SLAM [7], [8], [9] and other CV tasks [10], [11], [12] under images with 360° FoV, namely panoramic CV tasks. Mainstream camera models (e.g. pinhole camera, fisheye camera) are incapable of panoramic imaging with a FoV of 360°. As a feasible solution, these cameras are equipped with multi-view image mosaic algorithms [5], [6], [7] to make 360° panoramic imaging possible. However, the problems introduced by this sub-optimal choice (e.g. high latency between devices, inaccurate multi-view alignment and heavy computational burden) are unbearable for its application in mobile and wearable devices or some extreme environments.

Panoramic Annular Lens (PAL) is a compact solution to panoramic imaging with 360° FoV but a late starter [13], [14]. Due to the catadioptric structure of the PAL head, a PAL can achieve 360° panoramic imaging, where incident rays of 360° FoV are collected and bent to a beam with small angle. To correct aberrations in different FoVs, complex relay lens groups are applied so that traditional PAL systems normally consist of more than 8 lenses [15], [16], [17]. Simplifying the number of lenses is becoming a hot research topic. Some researchers leverage Q-type surfaces in high-quality panoramic systems design [18]. However, the special surface design for reduction in the number of lenses brings high cost and difficulties in manufacture, not to mention the still numerous lenses after reduction. Designing PAL with fewer spherical lenses is theoretically possible. But without the lenses for aberration correction, the image will degrade seriously, which limits its application. A valuable question emerges: Can we capture clear panoramic images through simple lens?

In the last decade, the great success of convolutional neural network (CNN) promotes the advance of Computational Imaging (CI) methods, in which a main group of efforts attempt to restore the aberration images from simple, imperfect optical lens [19], [20], [21], or compact imaging systems with large FoV [22], [23], [24]. Apart from network design, the key to applying CNN is to prepare datasets for network training, which is often realized by live shooting [22], [25] and optical simulation based on raytracing [21], random zernike coefficients [19], optical design software [26], or PSFs calibration [23], [24]. However, the imaging simulation of PAL, where the 360° panorama is distributed around the center of the image, is seldom explored in previous CI approaches [23], [24], [27], [28], [29].

In this paper, to maintain both tiny-volume lens system and high-quality 360° panoramic imaging, we pioneer the Annular
Computational Imaging (ACI) framework that can be applied to correct the aberration of any compact PAL design. A simulation pipeline is proposed for panoramic imaging with PAL which generates data pairs for network training based on the aberration distribution in unfolded PAL images and the spatially-variant sampling process during unfolding. The trained CNN can correct optical aberration of given PAL and enables simple lens to capture clear panoramic images.

On the other hand, deep learning based aberration correction [19], [21] often regards the problem as image restoration which discards the physical model of imaging. To integrate physical priors into restoration, we adopt the idea of single-pass physics-informed engine [30] and propose the Physics Informed Image Restoration Network (PI²RNet). An approximate recovery is made first to consider the distribution of point spread functions (PSFs) and convolution process of imaging, namely physics-informed engine. The followed fine restoration is also improved by degradation characteristics from the physics-informed engine, specially designed convolution kernels and global context components.

Last but not least, although synthetic datasets [21], [26] are easier to be mass produced and transferred to different lenses compared to manual shooting [22], [25], the synthetic-to-real gap is non-negligible considering that the manufactured lenses often deviate from the theoretical results. In our ACI framework, optical simulation is considered as data augmentation, where the random distributions of PSFs contribute to addressing errors introduced by manufacture and assembly, facilitating loose-tolerance requirements of our framework for lens design. The framework is evaluated on real-world images from a PAL composed of three spherical lenses but with aberration. Finally, we successfully capture clear panoramic images through simple lens, as shown in Fig. 1.

The main contributions of our work are in the following:

- We propose an Annular Computational Imaging (ACI) framework based on the physical model of panoramic imaging, improving the imaging quality of PAL while keeping its compact structure.
- The Physics Informed Image Restoration Network (PI²RNet) is designed to restore the degraded image, considering the physics-informed engine and spatial distribution of PSFs.
- We create the DIVPano dataset through wave-based simulation pipeline, in which random distributions of PSFs are adopted to bridge the synthetic-to-real gap.
- Finally, we test the qualitative effects of ACI on a real PAL composed of only three spherical lenses.

II. RELATED WORK

A. Aberration Correction Through Computation

Simple optical systems, which are inherently flawed because of the compact design, produces sub-optimal images with aberrations. In these cases, computational aberration correction is of great significance, and among these methods, the deconvolution approaches such as the Richardson-Lucy algorithm [31], [32] and Wiener filter [33] are classical approaches. Since the degradation caused by optical aberration is usually spatially-variant, i.e. the linear shift variant (LSV), the traditional hypothesis that blur kernel is invariant is no longer applicable. Trussell and Hunt [34], [35] first propose the idea of patch-wise restoration to address the problem. More work [36], [37] is then done to alleviate fringe and ringing effects of patch-wise restoration.

Another option is to calibrate the non-uniform PSFs of the LSV system. Early works [27], [38], [39] usually revolve around calibration process based on parameters of given lens and physical aberration model. Additionally, the idea of low-rank decomposition is applied to describe the PSFs of LSV system, which approximates the spatially varying PSFs as weighted sum of shift-invariant kernel [23], [24], [40], [41]. The calibrated PSFs are then used in image restoration.

As data-driven deep learning methods are extensively applied in image restoration [42], [43], [44], more aberration correction works are equipped with powerful CNN. These approaches often utilize networks with encoder-decoder structure [45] combined with efficient blocks for complex spatial distribution of aberrations [19], [21], [22], [25], [46], obtaining more efficient and robust results than traditional methods. Other networks for image restoration with spatially diverse degradation are also fit for aberration correction which fuse multi-scale information [43], [47], and use deformable convolution [48], [49] or dynamic convolution [50]. Nevertheless, these methods are usually evaluated on datasets and rarely applied to image restoration of real lenses.
Although computational imaging applied to perspective images captured by pinhole cameras is well developed, the solution to PAL with convenient design is hardly explored. At the same time, physical information corresponding to the imaging model is proved to be a vital factor in aberration-induced degradation [23], [24]. In our ACI framework, to solve the problems of image recovery under complex degradation, the network is designed based on the specific distribution of degradation in unfolded PAL images.

### B. Imaging Simulation Under Optical Aberration

Data-hungry supervised learning methods require hundreds of degraded-sharp image pairs for training. Although manual acquisition methods [22], [25] which capture the same scene with high-end cameras and simple lens can produce real degraded-sharp image pairs, the complex alignment and shooting process make this method restrict to specific case and difficult to transfer to other different lenses.

Consequently, imaging simulation [19], [21], [46] becomes a preferable choice by generating synthetic aberration datasets. The key point is to calculate the PSFs under optical aberration, which are often achieved by calibration, raytracing or wave-based methods. PSFs calibration is often assisted by low-rank decomposition [23], [24], [51] to improve the sparsely calibrated PSFs in LSV system. Raytracing is a direct method for PSF calculation. Shih et al. [52] calculate the PSFs of given lens based on the design parameters by combining raytracing with calibration, while Sun et al. [46] and Chen et al. [21] produce synthetic data pairs through raytracing-based imaging simulation. In contrast, wave-based methods [19], [20], [53] are more flexible and can generate multiple distributions of PSFs, whose core part is to estimate the phase function on pupil plane. Zernike polynomials model is widely adopted to represent the phase under aberration [54], especially when the influence of aberration is complex [55], [56].

Yet, due to the special catadioptric imaging mechanism of PAL, no approaches are explored for its imaging simulation. In addition, most of aforementioned pipelines rely on the lens design while ignoring the errors introduced by manufacture and assembly, making them highly depend on the low tolerances of the lens design. Hence, our simulation is set up for PAL in particular and takes into account the gap between the synthetic data and real panoramic imaging through wave-based PSF model.

### C. Panoramic Computer Vision Tasks

Panoramic images for vision tasks are often obtained by image mosaic or panoramic cameras. Multi-view image mosaic is widely adopted for panoramic imaging with 360° FoV. Patil et al. [5] and Deng et al. [6] perform panoramic object detection task with multiple perspective and fisheye lenses respectively. Lin et al. [7] propose panoramic visual odometry for a multi-camera system with two fisheye lenses, while Seok et al. design Rovo [57] and Omnislam [8] for wide-baseline multi-camera systems in subsequent years.

Recently, computer vision tasks with 360° FoV panoramic images from PAL show advantages. Yang et al. propose panoramic annular semantic segmentation (PASS) [2] which is valuable for surrounding sensing [3] based on PAL. Chen et al. [9] design the PALVO which applies PAL to visual odometry by modifying the camera model and specially designed initialization process. Fang et al. [10] employ PAL for visual place recognition, leveraging panoramas for omnidirectional perception with FoV up to 360°. Shi et al. [12] explore 360° optical flow estimation based on the cyclicity of panoramic images.

The robust and efficient performance of the panoramic vision algorithm depends on high-quality images captured by panoramic lens, while the cost and portability of the equipment is crucial if applied in mass-produced mobile terminals. In comparison to multi-camera systems, PAL is a more compact panoramic solution capable of imaging with FoV of 360°. Nevertheless, current designs [16], [18] require additional lenses to optimize image quality, otherwise the produced images are with severe aberrations. We propose the ACI framework to recover the images for low-cost PAL designs based on computational methods and make a sweet pot between the trade-off of the high-quality imaging and light-weight lens design. We hope that PAL design adopting ACI framework can maintain both tiny volume and high imaging quality, benefiting the panoramic computer vision community.

### III. OVERVIEW

For PAL, compact design with fewer lenses means sacrifice for image quality. Tiny volume and high image quality are contradictory in optical design. The goal of our work is to correct the image degradation of simple designed PAL through computational methods, enabling 360° FoV panoramic lens design with more compact structure, higher image quality, and lower manufacturing costs. The captured degraded annular image is unfolded and then fed to the CNN to recover a panoramic image without degradation (as shown in Fig. 2). Since perspective high quality images are publicly available, the PAL image is processed in unfolded type because it has similar distribution with public datasets. The annular flow chart in Fig. 2 shows the proposed ACI framework. The degraded image is synthesized from high-quality perspective images through asymmetrical sampling and optical simulation. Then we train the PI²RNet to reconstruct the image from simulated degradation, making it fit the ground truth well. Due to the similar degradation of synthesized dataset and real captured unfolded panorama, the PI²RNet can produce clear unfolded panoramic images. Detailed descriptions of the simulation pipeline and PI²RNet are given in Sections IV and V respectively.

### IV. SIMULATION FOR PANORAMIC ANNULAR IMAGING

For panoramic image restoration, a fatal problem is the lack of datasets, especially for the PAL annular image. However, the unfolded PAL image has similar data distributions with perspective images which is publicly available. It is reasonable to approximate the perspective images as unfolded PAL images and simulate corresponding degradation on them after considering the uneven sampling induced by unfolding. By this means, we can enjoy the benefits of large amount of the existing datasets.
Fig. 2. The proposed ACI framework. The “annular” closes the loop when the restored image approximate the high-quality sharp image. The proposed P^2RNet is trained on synthetic degraded dataset through optical simulation and restore the degraded image captured by a PAL with compact design. We map the annular image into unfolded one for the similar data distribution of perspective image by which our network is trained with. In this way, the restored image from P^2RNet enables light-weight PAL design with high imaging quality. For more detailed description please refer to Section III.

Fig. 3. Simulation pipeline for panoramic annular imaging. A perspective clear image on object plane is downsampled and upsampled asymmetrically to simulate the PAL unfolding process. The object plane image then degrades to aberration image based on aberration distribution on unfolded PAL image plane. The image signal processing (ISP) contains mosaiced, adding noise, demosaiced, white balance (WB), color correction matrix (CCM), and gamma compression while the Invert ISP is composed of gamma decompression, invert color correction matrix (Invert CCM), and invert white balance (Invert WB). The patch-wise convolution convolves the raw image after invert ISP with spatially variant PSFs from Zernike Model considering the Zernike coefficients, illumination distribution and spot diagram from the given PAL.

The simulation pipeline for panoramic annular imaging is depicted in Fig. 3, which maps clear perspective images to images with degradation characteristics of the given PAL. Our process follows the principle of real imaging equipment, i.e., the modulation of optical system is on the raw image [58] while the sensor presents the image we are familiar with through image signal processing (ISP) systems. It can be expressed as:

\[ A = \Gamma(R \otimes K + N). \]
The key point of the simulation is to compute the PSF $K$ of the optical system. At last, aberration image $A$ is produced by ISP $\Gamma(\cdot)$ after the convolution $\otimes$ between the raw image $R$ and PSF with the addition of sensor’s shot and read noise $N$.

The rest of this section is organized as follows: We first discuss the calculation of phase function on pupil plane and focal intensity on image plane in a general way. Then, the solutions to the particularities of PAL are depicted. Finally, we review the whole pipeline which completes the simulation for panoramic annular imaging.

A. Phase Function on Pupil Plane

Without loss of generality, we select general imaging system with circular pupil as an example, and firstly only discuss PSF calculation under a certain fixed FoV $\theta_0$ and wavelength $\lambda_0$. Assuming that a point light source is on the object plane, the intensity distribution on the image plane after propagation is the PSF.

According to Fourier optics, the pupil function on pupil plane affected by aberration needs to be described first, which counts for PSF calculation. In this case, it’s unnecessary to discuss the specific types of aberrations, because the effect of them is all to deviate the wavefront at the pupil from the ideal sphere. The wave aberration $W(x, y)$ is applied to represent the optical path difference between the two surfaces, so the phase function $\Phi(x, y)$ on the pupil plane is expressed as:

$$\Phi(x, y) = k_0 W(x, y),$$  \hspace{2cm} (2)

where $k_0 = \frac{2\pi}{\lambda}$ denotes the wave number.

Zernike polynomials are adopted as a mathematical description of optical wavefronts propagating through circular pupils [59]. $\Phi(x, y)$ can therefore be described by Zernike circle polynomials $W(\rho, \theta)$ in polar coordinates as:

$$W(\rho, \theta) = \sum_{m,n} C_n^m Z_n^m(\rho, \theta),$$  \hspace{2cm} (3)

where $\rho$ denotes Zernike coefficients and $Z$ refers to polynomials. The combination of different $m$ and $n$ represents different orders. For the specific expressions of each item please refer to [60]. The above polar expression of $(\rho, \theta)$ can be rewritten in rectangular coordinates of $(x, y)$ by coordinate transformation $(x = \rho \cos \theta, y = \rho \sin \theta)$.

With the phase function on the pupil, pupil function $\mathcal{P}(x, y)$ can be written by:

$$\mathcal{P}(x, y) = P(x, y) e^{iW(x, y)},$$  \hspace{2cm} (4)

where $P(x, y)$ is circ function in our case.

B. Focal Intensity on Image Plane

In such imaging system, the light field propagating from the back surface to the image plane satisfies Fresnel diffraction theory and the amplitude can be calculated by scalar diffraction integral [61]. Before integral, we distinguish the coordinates of the image plane and the pupil plane. Let the coordinate of the image plane be $(x, y)$ and that of the pupil plane be $(x', y')$.

Therefore, the amplitude of the image plane $E(x, y)$ can be expressed as:

$$E(x, y) = \frac{E_0}{\lambda_0 d} \int \int \mathcal{P}(x', y') e^{-\frac{i 2\pi}{\lambda_0 d} (x'x + y'y)} dx' dy',$$  \hspace{2cm} (5)

where $E_0$ is a constant amplitude related to illumination, $d$ denotes the distance from the pupil plane to the image plane and $\lambda_0$ is the selected wavelength mentioned above.

The above discussion is carried out at a certain FoV $\theta_0$ and wavelength $\lambda_0$. In fact, the calculated $E$ on image plane is a function of FoV $\theta$ and wavelength $\lambda$. More generally, we write (5) as:

$$E(x, y, \theta, \lambda) = \frac{E_0}{\lambda d} \int \int \mathcal{P}(x', y', \theta, \lambda) e^{-\frac{i 2\pi}{\lambda d} (x'x + y'y)} dx' dy'.$$  \hspace{2cm} (6)

The dependence of pupil function $\mathcal{P}(x', y', \theta, \lambda)$ on FoV and wavelength is reflected in wave aberration $\mathcal{W}(x', y', \theta, \lambda)$ which has diverse distributions under different FoVs and wavelengths.

Finally, the focal intensity on the image plane is:

$$I(x, y, \theta, \lambda) = |E(x, y, \theta, \lambda)|^2.$$  \hspace{2cm} (7)

Under our hypothetical point light source condition, the above intensity is the PSF distribution $K(x, y, \theta, \lambda)$ required.

C. Adaptation to Unfolded PAL Images

In (2)–(7) provide the calculation approach for PSFs in general imaging system. In this section, the adaptation to PAL will be discussed. Since most of the available high-quality images are perspective, our simulation for panoramic annular imaging is aimed at the unfolded PAL image plane. Following previous works [62], we adopt the unit sphere model to describe the projection relationship for the catadioptric system. A unit sphere which takes the camera as its origin is included with the Taylor polynomial to fit the parametric model of PAL. We calibrate the coefficients of the Taylor polynomial for PAL through the omnicalib toolbox [63]. Then, we apply the camera model to project the original image onto the unit sphere and use the EquiRectangular Projection (ERP) to produce an equirrectangular panorama unfolded image.

For an annular PAL image, the patches of the inner circle with fewer pixels will be upsampled during unfolding, resulting in the degradation of image quality. This process can be expressed as upsampling with different scale factors which should be considered in simulation. The annular image is unfolded based on a certain FoV $\theta_0$, and the width of the unfolded image is the circumference of the circle at $\theta_0$. For a panoramic camera whose vertical field of view contains 90°, we take $\theta_0$ as 90°. For cameras that not included, we take the $\theta_0$ angle closest to the equator on the unit sphere [63]. The scale factor $S_\theta$ of any FoV $\theta$ is calculated by:

$$S_\theta = \frac{2\pi r_{\theta_0}}{2\pi r_{\theta}},$$  \hspace{2cm} (8)

where $r_{\theta_0}$ is the radius of $\theta_0$ and $r_\theta$ is the radius of $\theta$. We divide a perspective image into patches longitudinally and downsample each patch with corresponding scale factor $S_\theta$. After upsampling
these patches back to their original resolutions, the degradation of unfolding is simulated on the perspective image.

Moreover, we consider the distribution of PSFs and the geometric deformation introduced by projection in the unfolding process. As shown in Fig. 4, PSFs annularly distribute with FoVs along the radial direction on the original PAL image plane. On the unfolded image plane, PSFs distribute along the transverse stripes instead. Therefore, each row of pixels in the image corresponds to one PSF under its FoV. In addition, the geometric deformation of the PSFs in the ERP format is linearly related to the introduced scale factor $S_\theta$ of each FoV. The square PSF kernels on annular image plane in Fig. 4(a) is distorted to a trapezoidal one on unfolded image plane in Fig. 4(b) based on distribution of $S_\theta$ with respect to FoVs ($S_\theta$ varies in vertical but keeps fixed in horizontal). In this way, based on the kernel size on annular image computed from spot diagram, we can obtain the mapped size of PSF kernels on unfolded image plane, adapting the PSFs calculation to unfolded PAL images.

D. Simulation Pipeline

For the simulation pipeline described in (1), we need a more rigorous expression. The calculated PSF $K(x, y, \theta, \lambda)$ on unfolded PAL image plane is a function of FoV $\theta$ and wavelength $\lambda$. Under a certain FoV, (1) can be expressed as the combined action of all wavelengths within corresponding band:

$$
A(x, y, \theta) = \Gamma \left( \int R(x, y, \theta) \otimes K(x, y, \theta, \lambda) d\lambda \right), \tag{9}
$$

where the process of adding noise is included in the ISP $\Gamma(\cdot)$. $A(x, y, \theta)$ is the image patch of degraded image under FoV $\theta$ and $R(x, y, \theta)$ is the corresponding patch on raw image. We sample and divide the raw image into patches according to the FoV distributions of unfolded PAL images described in Section IV-C. The image patch is convolved with PSF under $\theta$ and then merged to form the whole degraded image.

The Zernike coefficients of $W(x', y', \theta, \lambda)$ can be calculated by Zemax during lens design. This process requires to sample different FoVs and wavelengths within the imaging range. In addition, the illumination distribution and the spot diagram of the lens are also considered to determine $E_0$ and the size of PSF kernels respectively.

V. PHYSICS INFORMED IMAGE RESTORATION NETWORK

Based on the physical model in Section IV, we first adapt the single-pass physics-informed engine [30] (Section V-A) to our work. In Section V-B, we introduce the whole network architecture composed of double U-nets named Physics-informed Network and Restoration Network. The special convolution kernel designed for spatially varying PSFs and the bridge to inform the Restoration Network the physical information from Physics-informed Network are depicted in Sections V-C and V-D respectively.

A. Single-Pass Physics-Informed Engine

In CI problem where the reconstructed image $\hat{R}$ is predicted from input $A$, Barbastathis et al. [30] adopt the idea of physics-informed learning, which helps solve inverse problems efficiently through data and physical model, to incorporate physical priors into deep learning, and propose the single-pass physics-informed engine demonstrated in Fig. 5(a). Consistency with [30], we use the operator $H$ to represent the forward model from $\hat{R}$ to $A$. $H^*$ is the approximate inverse operator of $H$ so
that the engine can be modeled as:

$$\hat{R} = f_{RN}(H^*A; \Theta),$$  \hspace{1cm} (10)

where $f_{RN}(\cdot)$ denotes the Restoration Network and $\hat{R}_0 = H^*A$ in Fig. 5 is a single approximant of the restored image. $\Theta$ are learnable parameters for Restoration Network. In our case, the forward operator $H$ is the process described in (9), in which the information of PSFs: $K(x,y,\theta,\lambda)$ is available from simulation pipeline.

However, solutions to the approximate inverse operator $H^*$ are no longer convincing in our case because of the difference between simulated PSFs and the real ones. Consequently, we design the novel Physics-informed Network $f_{H^*}(\cdot)$ as a substitute for operator $H^*$ and adapt the engine to our work. As shown in Fig. 5(b), the proposed Physics-informed Network makes an approximate restoration $\hat{R}_0$. The delicate design of Physics-informed Network has potential for better integration of physical priors.

### B. Network Architecture

We propose the Physics Informed Image Restoration Network to address the complex degradation introduced by spatially-variant PSFs and uneven upsampling. As shown in the Fig. 6, our model consists of two major parts: Physics-informed Network $f_{H^*}(\cdot)$ and the Restoration Network $f_{RN}(\cdot)$.

The degraded image and physical information are firstly fed into Physics-informed Network to make an approximate restoration. As Fig. 7 shows, our physical information consists of two parts: The distributions of PSF kernel sizes (height and width), and scale factors that indicates the degradation information of different spatial regions of the unfolded image captured by particular PAL. To enable $f_{H^*}(\cdot)$ to extract degradation characteristics from physical features and transfer them from $f_{H^*}(\cdot)$ to $f_{RN}(\cdot)$, we propose the Physics-informed Bridge (PiB) in the bottleneck of two encoder-decoder network.

The encoder and decoder of Restoration Network is composed of HINBlocks and ResBlocks respectively adopted from HINet [65] which proposes half-instance normalization for image restoration. The Down operation is a $4 \times 4$ convolution with stride 2 and the Up operation is a $2 \times 2$ transpose convolution with stride 2. Consider that the PSFs which generate aberration are spatially-variant in both content and shape, we apply Dynamic Deformable Convolution (D$^2$Conv) to feature maps between encoder and decoder based on encoded feature map from $f_{H^*}(\cdot)$. In addition, to maintain the global context information and enlarge the receptive filed, Context Block (CB) [64] is introduced which also constructs multi-scale information and makes network robust to noise.

### C. Dynamic Deformable Convolution

For PAL designs with very few lenses, the aberration distribution varies greatly with FoVs, causing the shape and content of PSFs spatially-variant. The traditional convolution operation overlooks this variation and applies spatially-fixed convolution kernel to pixels across the whole image. Since the image degradation is relevant to the spatial distribution of PSFs based on (9), we should redesign our convolution kernel to tackle the challenge of spatially-variant image degradation.

Deformable Convolution (DConv) [48] proposes a novel convolution format, in which the kernel shape changes with the position of pixels, while the content of the kernel is fixed for all locations across the feature map. Coincidentally, the Filter Adaptive Convolution (FAC) layer [50] is able to generate different kernel content for pixels with different spatial coordinates. To handle the spatially-variant degradation of images from PAL, we combine kernel-shape-variant convolution (i.e. DConv) with...
Fig. 8. The proposed Dynamic Deformable Convolution (D²Conv). The feature map containing physical information is chuncked into two branches to make a Dynamic Deformable Kernel Prediction (D²KP), where the dynamic branch produces an adaptive filter for every pixel and the deformable branch predicts the offset for it. Then, the input image (or feature) is convolved with the kernel.

kernel-content-variant convolution (i.e. FAC) and propose our Dynamic Deformable Convolution.

As is illustrated from Fig. 8, the feature map is firstly chuncked into two branches to make a Dynamic Deformable Kernel Prediction (D²KP). The dynamic branch produces an adaptive filter for every pixel and the deformable branch predicts the offset of each coordinate to decide the filter shape for it. Then, the input image (or feature maps) is convolved with the filter. When the feature map contains physical information, the filter will be adaptive to each pixel in terms of shape and content based on the degradation model.

D. Physics-Informed Bridge

A simple U-net to represent \( H^* \) may cause the loss of physical information and the coarse restoration process is not connected with the convolution model in degradation, bringing poor performance (depicted in Section VI-B). Inspired by the convolution model of image degradation, Kernel Prediction (KP) [66] is often applied to predict the dynamic kernel weights and biases for image restoration, but restricted to fixed kernel shape in different pixels. Thus, we replace the KP with our D²KP to better model the inverse process of panoramic imaging. Consequently, we build a Physics-informed Bridge (illustrated in Fig. 9) to equip

\[ f_{H^*} \] with physical significance and pass on modeled physical feature to \( f_{RN} \).

As Fig. 9 shows, we make D²KP first with the decoded feature map from \( f_{H^*} \). In PiB, the input image is applied D²Conv with the kernel for an approximate restoration supervised by the ground truth with L1Loss to ensure that \( f_{H^*} \) is trained to reconstruct feature of complex distributions of PSFs. Then, we fuse the physical features with the feature maps from input image and produce the output feature for \( f_{RN} \). In this way, the physical model of image restoration in single-pass physics-informed engine is replaced with our Physics-informed Network and its physical priors are transferred to the Restoration Network.

VI. EXPERIMENTS

A. Setup

1) Optical System: Our optical system for experiments on ACI is a PAL system consisting of three spherical lenses which is thin and light. Our system applies the prototype of a PAL head of one lens and a relay lens group composed of several lenses. Considering the crucial catadioptric structure of the PAL head for 360° panorama, we can only reduce the number of lenses of the relay lens group for more compact system, i.e. 2 lenses in our case. The FoV of the system is 360° × (30°-100°). Reasonable focal power distribution of lenses according to the power distribution principle can greatly simplify the number of lenses, promoting its application under volume and space constraints. The light path diagram of selected PAL is shown in Fig. 10(a). The blue rays are rays path of the 30° half FoV. Green rays are rays with a maximum half FoV of 100°. The number of lenses of this PAL is reduced to less than 30% of traditional one, leading a relatively deficient imaging quality with aberration.

Fig. 10(b) shows the adopted PAL system, the maximum optical lens diameter of our device is comparable to the size of a coin, enabling panoramic CV tasks for space-constrained scenarios. The MV-SUA133GC camera with resolution of 1024 × 1280 and pixel size of 4 μm is applied as the sensor considering the comparable target plane size with our PAL. Due to the reduction of the number of lenses, the optical aberration of the large FoV is hard to correct through optical design, resulting in the degradation of the captured image shown in Fig. 10(c). The proposed ACI framework will alleviate the drawback of relatively low imaging quality while keeping the tiny size.

2) Data Preparation: We generate degraded images through applying simulation pipeline on DIV2K [67]. The illumination
distribution, spot diagram and Zernike coefficients (keep the first 37 polynomials as a common practice \cite{59,60}) under each FoV and wavelength are calculated by Zemax. For each wavelength, the FoVs are sampled in the range of \([30^\circ, 100^\circ]\) with step of 0.7\(^\circ\) for 101 related PSFs. We produce 31 groups of such PSFs (PSFs from 101 FoVs in each group) from wavelengths of 400 nm to 700 nm in 10 nm intervals. Then we translate them to red, green, and blue channels according to the wave response of the sensor, respectively. Considering that the panorama is unfolded to 288 \(	imes\) 1504, we crop and downsample the images of DIV2K into 512 \(	imes\) 1024 and produce degraded images through simulation pipeline (depicted in Section IV) based on aforementioned 101 PSFs in RGB channels.

To improve the generalization ability of our model to complex degradation distribution, we randomly adjust each Zernike coefficient at all FoVs within \(\pm 25\%\) to produce image pairs with diverse aberration distributions. For each FoV \(\theta\), a random number of the uniform distribution \(r_{i,\theta}\) within \([-0.25, 0.25]\) is generated, and coefficient \(Z_{i,\theta}\) is then replaced by \((1 + r_{i,\theta})Z_{i,\theta}\).

By this means, we propose the DIVPano dataset composed of image pairs under different aberration distributions from random adjustment. To validate the generalization ability of image restoration approaches, the aberration distributions in validation set is different from those in training set. We use Strehl ratio \cite{68}, which can measure the energy diffusion of PSFs, to describe the aberration distributions under different FoVs. As is shown in Fig. 11, centering around a standard distribution, training set consists of 3 deviated PSFs distributions while validation set is composed of other 3 different ones. The ground truth for training set and validation set are divided in the same way as DIV2K while the testing images are captured by the chosen PAL.

3) Training Details: The DIVPano dataset and the physical information introduced in Section V-B are used to train our PI\(^2\)RNet. The approximate restoration and final restored image are both supervised by ground truth \(G\) with L1Loss:

\[
L = \lambda ||\hat{R}_0 - G|| + ||\hat{R} - G||,
\]

where \(\lambda\) is the loss weight and empirically set to 0.5. We also conduct experiments on the choice of \(\lambda\) in Section VI-B.

The network is trained with Adam optimizer on a single NVIDIA GeForce RTX 3090. We also use rotation and flip as data argumentation. We set the initial learning rate to \(2 \times 10^{-4}\) and apply cosine annealing strategy \cite{69} with a minimum learning rate of \(1 \times 10^{-7}\). Finally, we train our model on random crops of 256 \(	imes\) 256 image patch with a batch size of 8 for 1.4 \(	imes\) 10\(^5\) iterations.

B. Ablation Study

We conduct sufficient ablation studies on the components of physics-informed engine (Table I), comparison of different design in the D\(^2\) Conv module (Table II), different context components (Table III), and different data distributions (Table IV) of our DIVPano dataset. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) \cite{70} are employed as our evaluation metrics for all the experiments.

Ablations on single-pass physics-informed engine are shown in Table I. Our baseline (row 1) is a cascade structure of two
U-nets which has considerable parameters with our network. We first concatenate the physical information with our input image into the baseline (row 2). Physical information can inform network the degradation distributions and improves PSNR by 0.67 dB and SSIM by 0.71%. Then we adopt the single-pass engine to make an approximate restoration but without our PiB. The drop in both metrics (i.e., 0.9 dB in PSNR and 4.6% in SSIM, row 3 and 4) indicates that replacing the physical model of $H^*$ with a single U-net is ineffective. However, PiB is able to guide the former network to extract features of dynamic deformable kernels and inform the latter network the physical feature, contributing to an improvement of 0.3 dB in PSNR and 0.34% in SSIM (row 3 and 6). Even replacing the $D^2$ KP in PiB with KP, we can achieve better results compared to that without PiB (i.e., 1.15 dB in PSNR and 4.89% in SSIM, row 4 and 5) or single-pass engine (0.25 dB in PSNR and 0.29% in SSIM, row 3 and 5).

Ablations on $D^2$Conv Module are shown in Table II. The baseline is the network with single-pass engine and $D^2$ KP PiB evaluated above. We replace the $D^2$Conv in the bottleneck of Restoration Network with DConv [48] and FAC [50] respectively to evaluate the effectiveness of different convolution kernels. The network with $D^2$Conv stands out with improving PSNR by 0.14 dB and SSIM by 0.13% compared to baseline (row 1 and 4). $D^2$Conv considers both kernel-shape-variant convolution and kernel-content-variant convolution, which is consistent with the degradation process.

Ablations on global context are shown in Table III. We also take the best model in row 6 of Table I as baseline and perform the ablation study. Except for the mentioned $D^2$Conv, CB is applied to enlarge the receptive field and helps keep the global context of features at the bottleneck of Restoration Network. Context components bring considerable improvement in metrics (i.e., 0.15 dB in PSNR and 0.13% dB in SSIM, row 1 and 4) and make the deep features sharper and more even in global region (shown in Fig. 12).

Ablations on random distributions of aberration for training set are shown in Table IV and the corresponding results on a real captured unfolded PAL image are shown in Fig. 13. We train our Pi²RNet and SRN [43] (a classical network for image restoration) on training sets with standard PSFs distribution (blue curves in Fig. 11) and random distributions (other 3 curves around the standard one in Fig. 11). Random data distributions can improve the generalization ability of the model and enhance the performance under the synthetic-to-real gap. The quantitative results on validation set with other 3 random distributions illustrate the effectiveness of our dataset. We note that our method requires the random distributions for training a robust $f_{H^*}$ to learn degradation characteristics. Consequently, it performs more poorly than SRN when training with single standard distribution. In addition, the random distributions contribute to the better performance in our testing images. Compared with the degraded image in Fig. 13(a), the image in Fig. 13(b) restored by Pi²RNet trained on standard distribution is slightly sharper but with more fake texture. Because of the single data distribution, the network overfits to the standard aberration characteristics and performs badly in real captured PAL images. Obviously, training with random data distributions brings a superior restoration, as shown in Fig. 13(c).

Ablations on the loss weight $\lambda$ of the approximate restoration are shown in Table V. When the $\lambda$ is set to 0, the modeling of degradation characteristics in $f_{H^*}$ becomes implicit, which is ineffective to the task and leads to the worse performance (28.71 dB to 28.43 in PSNR and 89.70 % to 89.52 % in SSIM). As a supervision for the learning of degradation characteristics, the loss function of the approximate restoration is essential which improves the PSNR by 0.22~0.28 and SSIM by 0.08~0.18%. The optimal value for $\lambda$ is 0.5 with advantages of 0.05 dB in PSNR and 0.08% in SSIM.

### C. Comparisons With State-of-The-Art Methods

We compare our proposed Pi²RNet with state-of-the-art image restoration methods on both our DIvPano dataset and real unfolded PAL images. SRN [43] and DeepRFT [71] are selected to represent the multi-scale methods while DeepRFT processes images in frequency domain. KPN [72] is a classical network adopting KP which considers the spatially-variant degradation. In addition, we also choose latest NAFNet [73] and HiNet [65] because of their excellent performances in various image restoration tasks. We replicate these methods and retrain them on DIvPano with the default setting according to the corresponding papers. We keep batch size and training iterations the same as ours for all the methods for fair comparison.

Table VI shows the quantitative results (PSNR and SSIM) of selected methods on our validation set. The parameters and runtime for the inference of one unfolded panoramic image with resolution of $288 \times 1504$ tested on one single NVIDIA RTX 3090 are also illustrated. As mentioned in Section VI-A2, the image degradation of our dataset varies greatly with FoVs, and
to model the unknown aberration of the real lens, the aberration distribution is different between training set and validation set, placing a huge challenge for all the methods. Consequently, our PI\(^2\)RNet which considers physical prior of panoramic imaging is superior to all the competitive methods. The physics-informed engine guides the network with the degradation distributions and proposed D\(^2\)Conv makes the recovery process conform to the degradation model. Our Physics-informed Network can learn robust characteristics of degradation based on physical information and degraded image from random distributions. In other words, the different degradation characteristics are first modeled and extracted by \(f_{HR}\). As long as the degradation distributions of validation images approximates the theoretical design, the proposed \(H^*\) promotes the network performance by strengthen the generalization of the model. All the proposed components and physical information help our method outperform the state-of-the-art methods by 0.11~2.38 dB in PSNR and 0.34~5.49% in SSIM on validation set of DIVPano, with slight increase in parameters and sacrifice of speed.

We capture testing images with selected panoramic optical system to evaluate our model on real panoramic images. An unfolded PAL image is shown at the top of sub figure in Fig. 14 where the lower left part is the original degraded one and the upper right part is the image restored by our work. We select two image patches in each scene from different FoVs, and show the original image patches and restoration image patches from our method and other 3 best methods in Fig. 14. With physical priors integrated, PI\(^2\)RNet is a superior solution to the spatially-variant degradation of unfolded panorama. In the edge FoV regions (red boxes) with severer aberrations, our work has better aberration correction effectiveness in terms of higher contrast and sharper edges, while it can also produce convincing results with fewer artifacts and more realistic details in the regions with slight aberrations (yellow boxes).

The Physics-informed Network helps the network adapt to different degradation distributions without overfitting to synthetic data and prevents the restored image from fake textures, compared with the results from HINet. Comprehensively, based on the physical model in the image restoration, our proposed PI\(^2\)RNet dominates other state-of-the-art methods in restoration of unfolded PAL images.

### D. MTF Analysis

The qualitative results of restored PAL images reveal the effectiveness of our ACI framework. Rigorously, we evaluate the

| Method | PSNR(dB) | SSIM(%) | Parameters(M) | Runtime(s) |
|--------|----------|---------|---------------|------------|
| SRN [43] | 28.26 | 88.58 | 10.25 | 0.16 |
| DeepRFT [71] | 27.81 | 84.21 | **09.60** | 0.24 |
| KPN [72] | 26.33 | 85.80 | 27.65 | **0.10** |
| NAFNet [73] | 28.60 | 89.36 | 17.11 | 0.14 |
| HINet [65] | 28.24 | 89.19 | 88.70 | 0.19 |
| Ours | **28.71** | **89.70** | 14.48 | **0.21** |
Fig. 14. Qualitative results on real unfolded PAL image compared with state-of-the-art methods. According to Table VI, we compare our PI$^2$RNet with other 3 most competitive methods, i.e. SRN, NAFNet and HINet. We unfold the image captured by selected PAL and shows a partial FoV of it at the top of each sub-figure. The upper right part of this image is replaced with restoration result of our work. We highlight regions of details under different FoVs by chromatic boxes.

Fig. 15. MTF analysis of unfolded PAL image. Average MTF curves of real unfolded PAL image and restored image by ACI are plotted along with the diffraction limit. The detected edges and corresponding image patches are shown on the top and lower left respectively. We also show the results of MTF50 of degraded image and restored image in the figure.

In this paper, we explore the trade-off of the high-quality imaging and compact PAL design by proposing the ACI framework. As a flexible high-level framework, our ACI can be deployed to any light-weight PAL system for higher imaging quality in volume-critical scenes. We expect our work paves...
the way for the subsequent advancement of practical panoramic vision tasks.

Specifically, we first obtain degraded PAL unfolded images through a wave-based simulation pipeline. For the target PAL lens, we create the DIVPano dataset and impose various aberration modulation to enhance the robustness of the model. We then propose PF-RNet to recover degraded images, which utilizes the physical priors of image degradation to assist the recovery process. Both quantitative and qualitative experiments show that our method outperforms current state-of-the-art methods in the PAL image restoration task.

Our framework, with the random distribution of aberrations, is proved to be robust to small vibration in the optical system, e.g., the error of assembly and manufacture. However, the simulation and training process of ACI should be re-conducted for new hardware configuration, e.g., a new prototype PAL with different types or numbers of lenses, whose aberrations distribution varies greatly from that of applied system. This is also a limitation of most CI pipelines for aberration correction of optical lens.

In the future, we will focus on panoramic imaging with higher quality in terms of resolution, noise and dynamic range. Further researches on color correction and exposure of imaging sensor are also necessary for impressive imaging results. We are particularly interested in the trade-off between sharpness and noise level of restored PAL images. Moreover, we will investigate to establish a pre-training model library for improving the generalization ability of our framework to different designs of PAL. In addition, an adaptation of ACI framework to annular PAL images will be explored by shifting the data domain, which is meaningful to some panoramic CV tasks such as panoramic SLAM.

ACKNOWLEDGMENT

The authors would like to thank Dr. Shiqi Chen from Zhejiang University for discussion on imaging simulation.

REFERENCES

[1] S. Gao, K. Yang, H. Shi, K. Wang, and J. Bai, “Review on panoramic imaging and its applications in scene understanding,” IEEE Trans. Instrum. Meas., vol. 71, pp. 1–34, 2022, Art. no. 5026034.
[2] K. Yang, X. Hu, L. M. Bergasa, E. Romera, and K. Wang, “PASS: Panoramic annular semantic segmentation,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 10, pp. 4171–4185, Oct. 2020.
[3] K. Yang et al., “Can we PASS beyond the field of view? Panoramic annular semantic segmentation for real-world surrounding perception,” in Proc. IEEE Intell. Veh. Symp., 2019, pp. 446–453.
[4] L. Sun et al., “Aerial-PASS: Panoramic annular scene segmentation in drone videos,” in Proc. Eur. Conf. Mobile Robots, 2021, pp. 1–6.
[5] R. Patil, P. E. Rybski, T. Kanade, and M. M. Veloso, “People detection and tracking in high resolution panoramic video mosaic,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., vol. 2, 2004, pp. 1323–1328.
[6] F. Deng, X. Zha, and J. Ren, “Object detection on panoramic images based on deep learning,” in Proc. 3rd Int. Conf. Control, Automat. Robot., 2017, pp. 375–380.
[7] M. Lin, Q. Cao, and H. Zhang, “PVO: Panoramic visual odometry,” in Proc. 3rd Int. Conf. Adv. Robot. Mechatronics, 2018, pp. 491–496.
[8] C. Won, H. Seok, Z. Cui, M. Pollefeys, and J. Lim, “OmniSLAM: Omnidirectional localization and dense mapping for wide-baseline multi-camera systems,” in Proc. IEEE Int. Conf. Robot. Automat., 2020, pp. 559–566.
[9] H. Chen et al., “PALVO: Visual odometry based on panoramic annular lens,” Opt. Exp., vol. 27, no. 17, pp. 24481–24497, 2019.
[10] Y. Fang, K. Yang, R. Cheng, L. Sun, and K. Wang, “A panoramic localizer based on coarse-to-fine descriptors for navigation assistance,” Sensors, vol. 20, no. 15, 2020, Art. no. 4177.
[11] H. Jiang, Z. Sheng, S. Zhu, Z. Dong, and R. Huang, “UniFuse: Unidirectional fusion for 360 panorama depth estimation,” IEEE Robot. Automat. Lett., vol. 6, no. 2, pp. 1519–1526, Apr. 2021.
[12] H. Shi et al., “PanoFlow: Learning 360 optical flow for surrounding temporal understanding,” 2022, arXiv:2202.13388.
[13] P. Grguss, “Panoramic imaging block for three-dimensional space,” U.S. Patent 4,566,763A, Jan. 28, 1986.
[14] J. Powell, “Panoramic lens,” Appl. Opt., vol. 33, no. 31, pp. 7356–7361, 1994.
[15] S. Niu, J. Bai, X. Hou, and G. Yang, “Improvement in design of panoramic annular lens using cemented lenses,” Proc. SPIE, vol. 6834, pp. 515–522, 2007.
[16] D. Cheng, C. Gong, C. Xu, and Y. Wang, “Design of an ultrawide angle catadioptric lens with an annularly scribed aspherical surface,” Opt. Exp., vol. 24, no. 3, pp. 3266–3277, 2016.
[17] K. Zhang, X. Zhong, L. Zhang, and T. Zhang, “Design of a panoramic annular lens with ultrawide angle and small blind area,” Appl. Opt., vol. 59, no. 19, pp. 5737–5744, 2020.
[18] S. Gao, E. A. Tsyganok, and X. Xu, “Design of a compact dual-channel panoramic annular lens with a large aperture and high resolution,” Appl. Opt., vol. 60, no. 11, pp. 3094–3102, 2021.
[19] L. Hu, S. Hu, W. Gong, and K. Si, “Image enhancement for fluorescence microscopy based on deep learning with prior knowledge of aberration,” Opt. Lett., vol. 46, no. 9, pp. 2055–2058, 2021.
[20] Y. Peng, Q. Fu, F. Heide, and W. Heidrich, “The diffractive achromat full spectrum computational imaging with diffractive optics,” ACM Trans. Graph., vol. 35, no. 4, pp. 1–11, 2016.
[21] S. Chen, H. Feng, D. Pan, Z. Xu, Q. Li, and Y. Chen, “Optical aberrations correction in postprocessing using imaging simulation,” ACM Trans. Graph., vol. 40, no. 5, pp. 1–15, 2021.
[22] Y. Peng, Q. Sun, X. Dun, G. Wetzstein, W. Heidrich, and F. Heide, “Learned large-field-of-view imaging with thin-plate optics,” ACM Trans. Graph., vol. 38, no. 6, pp. 1–14, 2019.
[23] X. Zhe, J. Wang, Y. Guo, and L. Tian, “Deep-learning-aided computational miniature microscope,” Optica, vol. 9, no. 9, pp. 1009–1021, 2022.
[24] K. Yamy, K. Monakhova, R. W. Shuai, and L. Weller, “Deep learning for fast spatially varying deconvolution,” Optica, vol. 9, no. 1, pp. 96–99, 2022.
[25] A. Ignatov, L. V. Gool, and R. Timofte, “Replacing mobile camera isp with a single deep learning model,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 536–537.
[26] X. Li, J. Suo, W. Zhang, X. Yuan, and Q. Dai, “Universal and flexible optical aberration correction using deep-prior based deconvolution,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 2613–2621.
[27] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, “Non-stationary correction of optical aberrations,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2011, pp. 659–666.
[28] T. Yue, J. Suo, J. Wang, X. Cao, and Q. Dai, “Blind optical aberration correction by exploring geometric and visual priors,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1684–1692.
[29] F. Heide, Q. Fu, Y. Peng, and W. Heidrich, “Encoded diffractive optics for full-spectrum computational imaging,” Sci. Rep., vol. 6, no. 1, pp. 1–10, 2016.
[30] G. Barbastathis, A. Ozcan, and G. Sito, “On the use of deep learning for computational imaging,” Optica, vol. 6, no. 8, pp. 921–943, 2019.
[31] W. H. Richardson, “Bayesian-based iterative method of image restoration,” J. Opt. Soc. Amer., vol. 62, no. 1, pp. 55–59, 1972.
[32] L. B. Lucy, “An iterative technique for the rectification of observed distributions,” Astronomical J., vol. 79, pp. 745–754, 1974.
[33] N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications, vol. 113, MIT Press, 1964.
[34] H. J. Trussell and B. R. Hunt, “Image restoration of space variant blurs by sectioned methods,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 1978, vol. 3, pp. 196–198.
[35] H. Trussell and B. Hunt, “Sectioned methods for image restoration,” IEEE Trans. Acoust., Speech, Signal Process., vol. 26, no. 2, pp. 157–164, Apr. 1978.
[36] J. Kim, A. Tsai, M. Cetin, and A. S. Willsky, “A curve evolution-based variational approach to simultaneous image restoration and segmentation,” in Proc. Int. Conf. Image Process., 2002, vol. 1, pp. 109–112.
[37] L. Bar, N. Sochen, and N. Kiryati, “Restoration of images with piecewise space-variant blur,” in Proc. Int. Conf. Scale Space Variational Methods Comput. Vis., 2007, pp. 533–544.

[38] N. Joshi, R. Szelszki, and D. J. Kriegman, “PSF estimation using sharp edges prediction,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2008, pp. 1–8.

E. Kee, S. Paris, S. Chen, and J. Wang, “Modeling and removing spatially-varying optical blur,” in Proc. IEEE Int. Conf. Comput. Photo., 2011, pp. 1–8.

[40] E. Maalouf, B. Colicchio, and A. Dieterlen, “Fluorescence microscopy three-dimensional depth variant point spread function interpolation using Zernike moments,” J. Opt. Soc. Amer. A, vol. 28, no. 9, pp. 1864–1870, 2011.

[42] J. Sun, W. Cao, Z. Xu, and J. Ponce, “Learning a convolutional neural network for non-uniform motion blur removal,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 769–777.

[43] X. Yao, H. Gao, X. Shen, J. Wang, and J. Jia, “Scale-recurrent network for deep image deblurring,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8174–8182.

[44] H. Zhang, Y. Dai, H. Li, and P. Koniusz, “Deep stacked hierarchical multi-patch network for image deblurring,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 5978–5986.

[45] C. A. Metzler, H. Ikkoma, Y. Peng, and G. Wetzstein, “Deep optics for single-shot high-dynamic-range imaging,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 1375–1385.

[46] Q. Sun, C. Wang, Q. Fu, X. Dun, and W. Heidrich, “End-to-end complex lens design with differentiable ray tracing,” ACM Trans. Graph., vol. 40, no. 4, pp. 1–13, 2021.

[47] R. Xu, Z. Xiao, J. Huang, Y. Zhang, and Z. Xiong, “EDPN: Enhanced deep pyramid network for blurry image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2021, pp. 414–423.

[48] J. Dai et al., “Deformable convolutional networks,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2017, pp. 764–773.

[49] X. Zhu, H. Hu, S. Lin, and J. Dai, “Deformable ConvNets V2: More deformable, better results,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 9308–9316.

[50] S. Zhou, J. Zhang, J. Pan, H. Xie, W. Zuo, and J. Ren, “Spatio-temporal filter adaptive network for video deblurring,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 2482–2491.

[51] K. Yanny et al., “Miniscop3D: Optimized single-shot miniature 3D fluorescence microscopy,” Light: Sci. Appl., vol. 9, no. 1, pp. 1–13, 2020.

[52] Y. Shih, B. Guenter, and N. Joshi, “Image enhancement using calibrated lens simulations,” in Proc. Eur. Conf. Comput. Vis., 2012, pp. 42–56.

[53] V. Sitzmann et al., “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging,” ACM Trans. Graph., vol. 37, no. 4, pp. 1–13, 2018.

[54] X. Dun, H. Ikkoma, G. Wetzstein, Z. Wang, X. Cheng, and Y. Peng, “Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging,” Optica, vol. 7, no. 8, pp. 913–922, 2020.

[55] X. Hao, C. Kuang, T. Wang, and X. Liu, “Effects of polarization on the de-excitation dark focal spot in STED microscopy,” J. Opt., vol. 12, no. 11, Art. no. 115707.

[56] X. H. H. Antonello, E. S. Allgeyer, J. Bewersdorff, and M. J. Booth, “Aberations in 4Pb microscopy,” Opt. Exp., vol. 25, no. 13, pp. 14049–14058, 2017.

[57] H. Seok and J. Lim, “ROVO: Robust omnidirectional visual odometry for wide-baseline wide-FOV camera systems,” in Proc. IEEE Int. Conf. Robot. Automat., 2019, pp. 6344–6350.

[58] T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet, and J. T. Barron, “Unprocessing images for learned raw denoising,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 11036–11045.

[59] V. Lakshminarayanan and A. Fleck, “Zernike polynomials: A guide,” J. Modern Opt., vol. 58, no. 7, pp. 545–561, 2011.

[60] V. N. Mahajan, “Zernike circle polynomials and optical aberrations of systems with circular pupils,” Appl. Opt., vol. 33, no. 34, pp. 8121–8124, 1994.

[61] E. Huggins, “Introduction to Fourier optics,” Phys. Teacher, vol. 45, no. 6, pp. 364–368, 2007.

[62] D. Scarabuzza, A. Martinelli, and R. Siegwart, “A flexible technique for accurate omnidirectional camera calibration and structure from motion,” in Proc. IEEE Int. Conf. Comput. Vis. Syst., 2006, pp. 45–45.
Shaohua Gao received the B.S. and M.S. degrees in measurement technology and instrument from the Changchun University of Science and Technology, Changchun, China, in 2018 and 2021, respectively, and the M.S. degree in optical engineering from the Saint Petersburg State University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, in 2020. He is currently working toward the Ph.D. degree in electronic information from the State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China. His research interests include optical design and computer vision.

Kailun Yang received the B.S. degree in measurement technology and instrument from the Beijing Institute of Technology, Beijing, China, and the dual degree in economics from Peking University, Beijing, in 2014, and the Ph.D. degree in information sensing and instrumentation from the State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China, in 2019. He performed the Ph.D. Internship with the RobeSafe Lab, University of Alcalá, Alcalá Henares, Spain, from 2017 to 2018. He is currently a Postdoctoral Researcher with the Computer Vision for Human-Computer Interaction Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany. For more information, visit his website: https://yangkailun.com.

Kaiwei Wang received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2001 and 2005, respectively. He is currently a Full Professor with the State Key Laboratory of Modern Optical Instrumentation, and the Deputy Director of the National Optical Instrument Engineering Research Center, Zhejiang University, Hangzhou, China. In October 2005, he started his Postdoctoral research with the Center of Precision Technologies, Huddersfield University, Huddersfield, U.K., funded by the Royal Society International Visiting Postdoctoral Fellowship and the British Engineering Physics Council. In February 2009, he joined Zhejiang University, and has been mainly researching on intelligent optical sensing technology and visual assisting technology for the visually impaired. Up to date, he owns 80 patents and has authored or coauthored more than 150 refereed research papers. For more information, visit his Website: http://wangkaiwei.org/.