Extreme-Quality Computational Imaging via Degradation Framework

Shiqi Chen, Huajun Feng, Keming Gao, Zhihai Xu, Yueting Chen
State Key Laboratory of Modern Optical Instrumentation, Zhejiang University
Hangzhou, Zhejiang, China
{chenshiqi, fenghj, gaokeming, xuzh, chenyt}@zju.edu.cn

Abstract

To meet the space limitation of optical elements, free-form surfaces or high-order aspherical lenses are adopted in mobile cameras to compress volume. However, the application of free-form surfaces also introduces the problem of image quality mutation. Existing model-based deconvolution methods are inefficient in dealing with the degradation that shows a wide range of spatial variants over regions. And the deep learning techniques in low-level and physics-based vision suffer from a lack of accurate data. To address this issue, we develop a degradation framework to estimate the spatially variant point spread functions (PSFs) of mobile cameras. When input extreme-quality digital images, the proposed framework generates degraded images sharing a common domain with real-world photographs. Supplied with the synthetic image pairs, we design a Field-Of-View shared kernel prediction network (FOV-KPN) to perform spatial-adaptive reconstruction on real degraded photos. Extensive experiments demonstrate that the proposed approach achieves extreme-quality computational imaging and outperforms the state-of-the-art methods. Furthermore, we illustrate that our technique can be integrated into existing postprocessing systems, resulting in significantly improved visual quality.

1. Introduction

Combined with powerful hardware and manual-made image signal processing (ISP) systems, mobile cameras have achieved great success. Recently, smartphones have become the primary source of photographs, and the pursuit of better imaging results has increased even more. However, to meet the space limitation of lenses, high-order aspherical lenses or free-form surfaces are applied in mobile cameras. Most of them aim to compress the volume of wide-angle lenses [17], but in the meanwhile, they also introduce the problem of image quality mutation (as shown in Fig. 2).

Recent cameras have shifted some of these correction tasks from lens design to ISP systems to correct this abrupt degradation. However, different steps of the traditional ISP are independent of each other, where errors accumulate and magnify in the following steps. So for the postprocessing pipeline, how to accurately estimate the abrupt degradation and correct it are the keys to reconstruction.

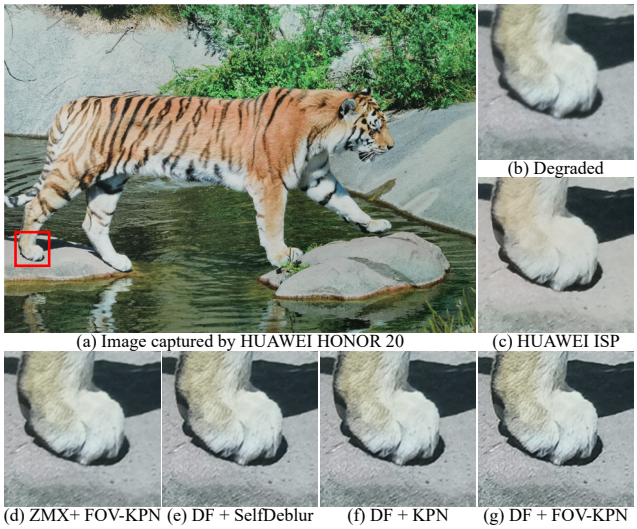

Figure 1. The reconstructions of a real-world image (a) captured by HUAWEI HONOR 20. (b) is the primary degradation, (c) is the output of HUAWEI ISP. Results of the proposed FOV-KPN model trained using image pairs simulated by (d) $Zemax^{®}$ (ZMX). (e), (f), and (g) are the restorations of SelfDeblur [39], KPN [31], and FOV-KPN model (Ours), respectively, all of them trained on the data generated by our degradation framework (DF).

One approach is to estimate the camera intrinsic degradation kernel through iterative optimization [33]. However, for the PSFs that show a wide range of spatial variation over regions, optimization methods have difficulty converging. The other way is reconstructing the degraded image with deep learning method [20], which depends heavily on the accuracy of data pairs. Moreover, because the degradation of diverse cameras is different, it is urgent to develop feasible and convenient approaches to generate the training data designed for each camera. In [37], the researchers

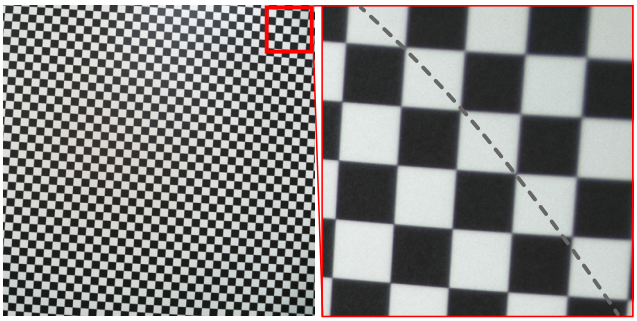

Figure 2. Image quality mutation of the mobile camera. One can see that in the magnified region (right), the two sides separated by the dotted line have a significant difference in blurring.

put forward an indoor method of automatic image capturing. However, sophisticated operations, including registration, color calibration, distortion correction, are required to post-process the captured images. These operations modify the degradation of raw data. Therefore, the challenges of degradation correction lie in two aspects. **On the one hand**, we need targeted methods to simulate accurate degraded data of diverse cameras. **On the other hand**, there is not exist a spatial-adaptive reconstruction technique that can integrate into existing post-processing systems.

In this work, we propose a robust degradation framework for estimating spatially variant kernels from real photographs. And the framework generates accurate imaging results of diverse cameras. As shown in Fig. 1(d) and 1(g), accuracy of simulation has a great impact on restoration. In the meanwhile, based on the kernel prediction network (KPN) [7], we propose a Field-Of-View (FOV) shared network architecture aiming at performing spatial-adaptive recovery on degraded images. Different from KPN, the predicted kernel output results in a dilated convolution manner. The dilated way increases the influence scope with a relatively lower computational overhead. The rest of the network is divided into the FOV encoder and the Deformable Decoder. They aim to enhance the spatial adaptability of the model. Fig. 1(e)$\sim$1(g) show the reconstructions of the state-of-the-art methods and FOV-KPN. Our approach generates more vivid results than the compared methods.

The contributions of our work are as follows:

- A semi-supervised degradation framework is proposed to estimate spatially variant kernels from real photographs and generate realistic imaging results of a specific camera.
- We propose a spatial-adaptive network architecture to correct spatially variant degradation, and it can integrate into existing post-processing systems.

We conduct extensive experiments to analyze the performance of the proposed degradation framework and network architecture. Moreover, we prove that the proposed solution can replace the traditional ISP in practical scenarios, resulting in significantly improved image quality.

2. Related Work

Degradation Estimation. To calculate the degradation kernel of imaging procedure, scholars have put forward many constructive proposals, which can be divided into two categories. One is the optimization-based deconvolution method [36, 34, 35], which has a great performance in dealing with spatially consistent degradation. However, when applied to real images, these methods have difficulty in converging under low signal to noise ratio (SNR) condition [33, 21]. The other category is to estimate PSFs by *deep linear network*. It is first proposed by [3, 22], where an *internel-gan* is used to predict super-resolution kernel. Due to its unsupervised training procedure, the predicted kernels are unstable in application [29]. Therefore, compared to the deconvolution methods, the deep learning approaches can converge in extreme cases, but they need stronger constraints to predict the degradation kernel accurately.

ISP Pipeline. Because of the differences between human eyes sensitivity and sensor response, ISP systems post-process the raw data received by hardware to generate photographs adapted to human vision [41]. However, the traditional ISP systems in modern imaging devices are step-by-step operations [27, 5, 6, 26, 18], where errors of each step will accumulate and amplify in the following discrete steps. Due to the popularity of the end-to-end deep learning method, many researchers focus on substituting the traditional ISP systems with a trained deep learning model [2, 38]. In the previous works [19, 10, 37], data acquisition is an essential issue, where a lot of shooting, registration, and color correction is needed to generate proper data pairs for training. Moreover, when migrates to a new device, the data set needs to be collected again [7]. The work of recollection is time-consuming and requires a lot of manual costs. Therefore, it is urgent to design a targeted method for simulating the degradation of the imaging procedure.

Image Reconstruction. Single image reconstruction has been the focus of a significant body of researches in computer vision, image processing [53, 50, 49]. Most classic reconstruction works suppose that the degradation can be modeled by convolution [16]. In this way, they propose model-based deconvolution methods to restore the data generated by degraded kernel [47, 36]. However, because these approaches need to adjust parameters according to the actual situation, they can not be applied to real-time scenes. Recently, deep learning is becoming the most popular technology in image reconstruction [9], and many approaches combine the optimization schemes and CNNs [51, 12].

Nevertheless, the methods mentioned above are all pro-

posed to deal with globally consistent blur. Considering that the degradation in natural images is spatially variant [42, 24], we note that the KPN architecture is customized for solving this problem. KPN is first proposed in [31], and it has been proved to have better stability than direct prediction [44]. Nevertheless, KPN is not effective in dealing with severe degradation because computational overhead will increase if the predicted kernel grows larger [13]. Our work proposes an efficient way to integrate the spatial information and recover the spatially variant degradation adaptively.

3. Natural Images Degradation

In this section, we focus on the proposed degradation framework. First, we analyze the degradation of natural images in Section 3.1. Second, the proposed degradation framework is detailed in Section 3.2. And in Section 3.3, we illustrate the optical geometric priors, aiming at carrying out an accurate estimation for different FOVs.

3.1. Realistic Degradation in Imaging Pipeline

The pipeline of the image formation model is divided into two stages. The first stage is to record light intensity information via optics and a sensor. The second stage is to post-process the measurements, known as the ISP systems.

In the first stage, the degradation of light energy can be formulated as the following equation:

$$I_e(x,y) = \int R_e(\lambda) \cdot [p(x,y,d,\lambda) * i_e(x,y)] d\lambda + n(x,y), \tag{1}$$

Here $I_e(x,y)$ and $i_e(x,y)$ are the degradated measurement and the latent image, respectively. We emphasize that both $I_e(x,y)$ and $i_e(x,y)$ represent the energy intensity received by sensor. $R_e(\lambda)$ is the response distribution of sensor. $n(x,y)$ can be approximated as a single heteroscedastic Gaussian, we refer readers to [14] for more details. And the degraded PSF $p(x,y,d,\lambda)$ can be expressed as:

$$p(x,y,d,\lambda) = p_o(x,y,d,\lambda) * p_c(\lambda), \tag{2}$$

$p_o(x,y,d,\lambda)$ is the PSF of optical lenses. It varies with the spatial coordinates (x,y) of sensor plane, the shooting distance d, and the wavelength λ. $p_c(\lambda)$ is the PSF of sensor crosstalk, which just depends on the wavelength λ of input signal (well described in [45]).

In the second stage, ISP systems are conducted to the energy measurements $I_e(x,y)$ (the detail of each step in ISP is in [4]). We note that the steps cascaded down the ISP are independent of each other, where a slight error of each operation will magnify in the following steps. The solution to this problem can be obtained if only the data pairs are constructed end-to-end from degraded domain to latent domain. In this way, the error accumulated in ISP systems can eliminate with a deep learning model. Therefore, we firstly construct an authentic degradation framework based on the image formation pipeline. The detail of the framework will illustrate in the following section.

3.2. Degradation Framework

As shown in Fig. 3, the proposed degradation framework is mainly divided into two stages.

Stage I The first stage is to construct data pairs for degradation estimation. Assuming the blur and the noise level of real checkers are similar in some neighborhoods, we pick out the degraded label Y in the checkers taken by a specific camera. Then the structure Y_s can be easily obtained by edge detection algorithms. The pixel value of edges in the degraded input X will be interpolated in coloring to prevent discontinuous solutions at the boundary of data pairs. In this way, we preprocess Y to acquire the training pairs $\{x,y\}$. Furthermore, we transmit the normalized FOV and the ISP parameters (including white balance, CCM, and gamma value) to the second stage.

Stage II As analysed in Section 3.1, only the degradation in energy domain can be modeled as convolution. Therefore, we incorporate the ISP pipeline into the training procedure of degradation transfer. By using the ISP parameters transmitted from stage I, we convert extreme-quality images to raw-like data for training (the pipelines of unprocess and process are in Fig. 4). We must emphasize is that, in training, the ISP parameters follow the data of real checkers, but in the test, these parameters vary with a random color temperature. And the λ_{read} and the λ_{shot} in noise injection are determined by the camera measurements for a particular exposure (see [4] for more details).

Inspired by KernelGAN [3], we adopt *deep linear model* to perform the mapping of $\{x_l, x_l^d\}$ in energy domain. However, we discovered that the unsupervised way used in [3] is unstable in implementation. Hence, a novel semi-supervised manner is applied to train the degradation transfer, which significantly enhances the training stability. Unfortunately, because the degradation shows a wide range of spatial variation over regions, the PSFs of different FOVs cannot be accurately estimated by the degradation transfer with fixed hyperparameters. Therefore, it is significant to carry out specific training for different FOVs. To achieve this goal, we introduce the optical geometric priors for degradation transfers. They will illustrate in the following section.

3.3. Optical Geometric Priors

We design the optical geometric priors to carry out an accurate estimation. We devide them into two aspects: (1) constraint of PSFs size and shape, (2) symmetry constraint.

Shape constraint For each lens in modern camera, Strehl Ratio [30] indicates the energy diffusion of it. Given the

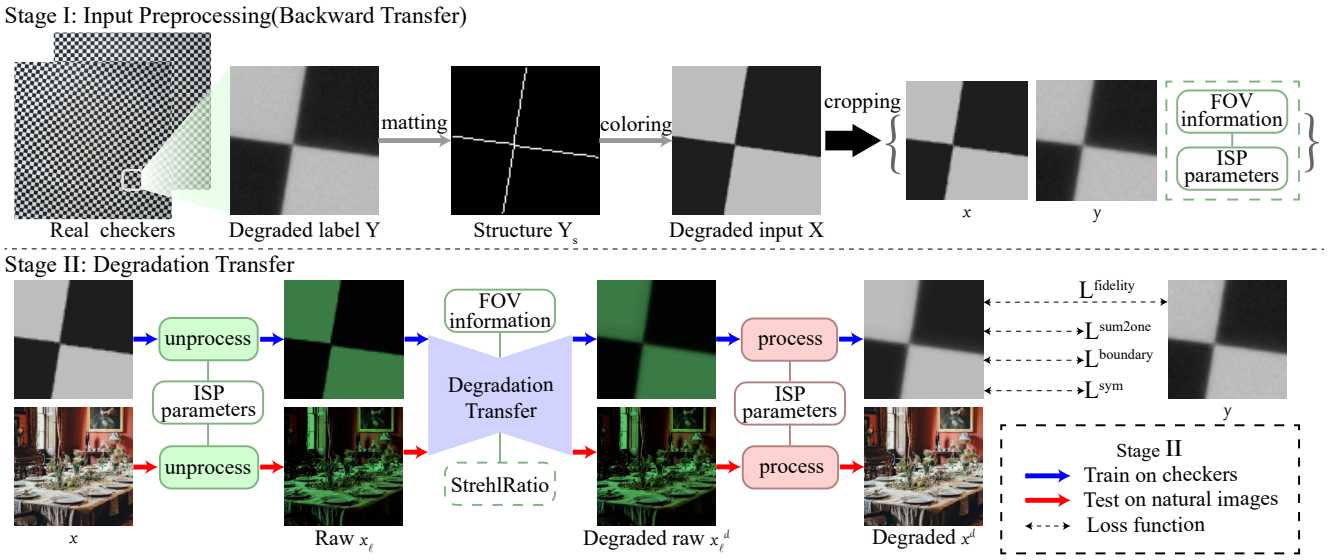

Figure 3. **Overview of the degradation framework.** In stage I, backward transfer constructs the data pairs for the training procedure of degradation transfer. And in stage II, each degradation transfer is tailored to one FOV, aiming at estimating the spatially variant PSFs of different FOVs and generating realistic imaging results.

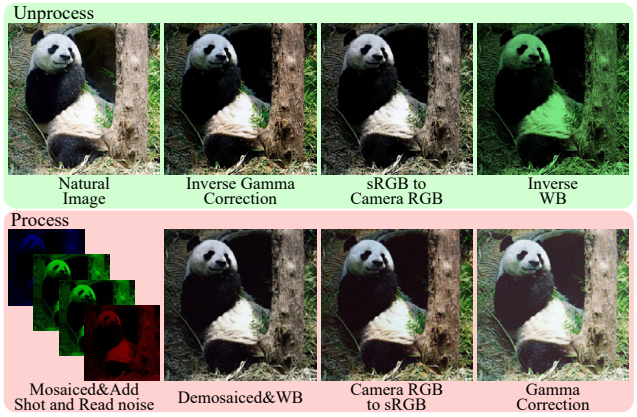

Figure 4. **Unprocess and Process Pipeline.** See Section 3.2 and [4] for details.

sensor pixel size, lenses airy disk, and Strehl Ratio, it is easy to derive the approximate PSF sizes of different FOVs. In this way, we construct a relationship between the Strehl Ratio and the receptive field of degradation transfer. To be more specific, the number of convolution layers in *deep linear model* will change according to the PSF size (detailed in the **supplementary file**). Moreover, the whole framework needs to meet following constraints:

$$\mathcal{L} = \alpha\mathcal{L}_{fidelity} + \beta\mathcal{L}_{sum2one} + \gamma\mathcal{L}_{boundary}, \quad (3)$$

here α, β, γ are the weight coefficients of different loss functions. $\mathcal{L}_{fidelity} = ||x^d - y||_1$ supervises the fidelity of the degraded result x^d. $\mathcal{L}_{sum2one} = |1 - \sum k_{i,j}|$ and $\mathcal{L}_{boundary} = \sum_{i,j} |k_{i,j} \cdot m_{i,j}|$ are the same in [3] (i, j is the 2D Cartesian coordinate of PSF k). γ is used to penalize the non-zero values close to the boundaries of k, so γ will be lower by the decrease of Strehl Ratio. In this way, specific training strategy according to the shape of PSF is designed.

Symmetry constraint For each off-axis pixel, there is a corresponding pixel that is central symmetric to it [48]. So the local PSF k and its symmetric one k_{sym} can be constrained with a general rotating constraint:

$$\mathcal{L}_{sym} = \sum_{i,j}(k(i,j) - \frac{1}{2}(k(i,j) + k_{sym}(-i,-j)))^2, \quad (4)$$

Here (i, j) is the 2D Cartesian coordinate of PSF. In this way, we force each PSF to be similar to its central symmetry counterpart. The overall constraints as follows:

$$\mathcal{L} = \alpha\mathcal{L}_{fidelity} + \beta\mathcal{L}_{sum2one} + \gamma\mathcal{L}_{boundary} + \delta\mathcal{L}_{sym}, \quad (5)$$

In summary, we design the optical geometric priors to carry out specific training for different FOVs: one prior is to constrain the size of degradation the other prior is to constrain the central symmetry of deterioration. Our priors directly derive from the optical parameters of a camera. The training procedures of the degradation transfer illustrate in Algorithm 1, where the subscript sym indicate the form of central symmetric. After training the degradation transfers $\mathcal{D}$ of different FOVs, the framework generates realistic imaging results as follows. Each FOV patch of a digital image $\mathbf{I}_{GT}$ is input into the corresponding $\mathcal{D}$, and all the outputs join together to get $\mathbf{I}_{DE}$. Although trained on the real checkerboards taken by a camera, the framework can generalize the

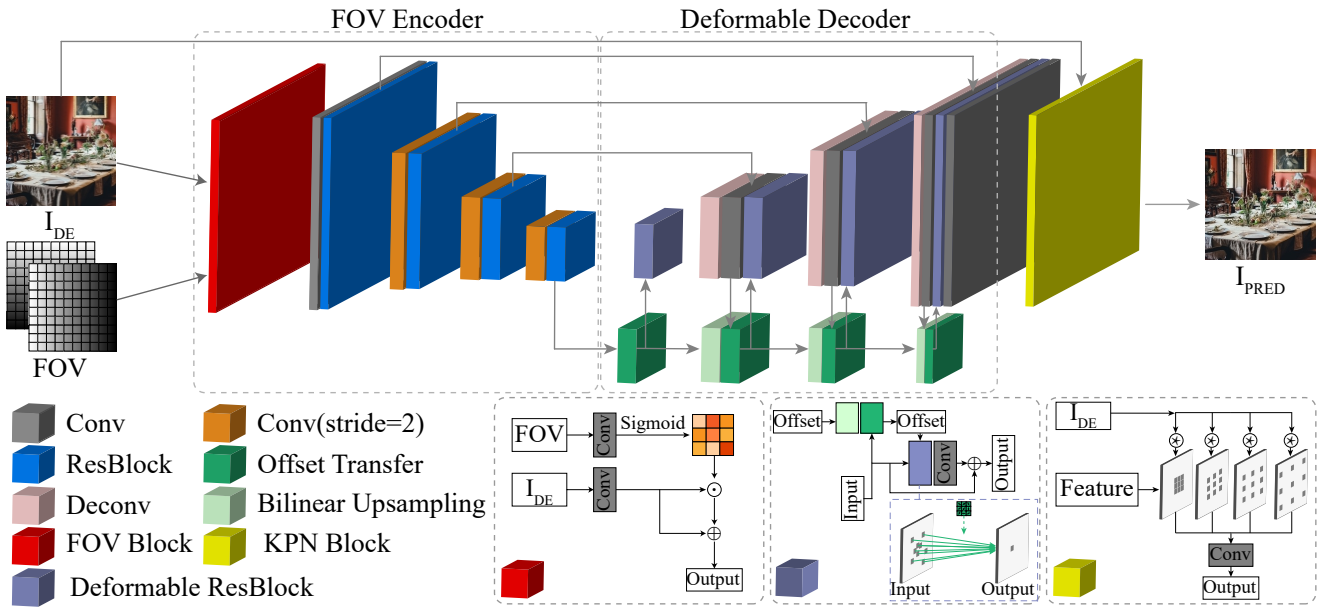

Figure 5. **Network Architecture** The UNet-based network architecture is shown on the top, and the layer configurations illustrate with different colored blocks (bottom left). The FOV Block, Deformable ResBlock, and KPN Block are detailed in the bottom right corner.

Algorithm 1: Training of Degradation Transfer

Input: data-pair $\{x, y\}$ and $\{x, y\}_{sym}$
Output: trained Degradation Transfer $\mathcal{D}$ and $\mathcal{D}_{sym}$

Initialize Degradation Transfer $\mathcal{D}$ and $\mathcal{D}_{sym}$ according to the corresponding Strehl Ratio;
for *t = 1 to T* **do**
  x^d = process($\mathcal{D}$(unprocess(x)));
  x^d_{sym} = process($\mathcal{D}_{sym}$(unprocess(x_{sym})));
  Calculate k and k_{sym} following [3];
  Calculate $\mathcal{L}(x^d, y, k, k_{sym})$ using Eq.5;
  Calculate $\mathcal{L}_{sym}(x^d_{sym}, y_{sym}, k_{sym}, k)$ using Eq.5;
  Compute the gradient w.r.t. $\mathcal{D}$ and $\mathcal{D}_{sym}$;
  Update the weight of $\mathcal{D}$ and $\mathcal{D}_{sym}$;
end
Output the trained $\mathcal{D}$ and $\mathcal{D}_{sym}$;

estimated deterioration to natural images. It is due to the powerful generalization capability of the *deep linear model* (detailed in [3]). In this way, we prepare the data pairs $\{\mathbf{I}_{DE}, \mathbf{I}_{GT}\} \in \{x_d, x\}$ for extreme-quality computational imaging, where the mapping from $\mathbf{I}_{GT}$ to $\mathbf{I}_{DE}$ covers all authentic deteriorations in image formation.

4. Image Reconstruction Model

An end-to-end reconstruction model is implemented to realize extreme-quality computational imaging. It takes degraded sRGB images $\mathbf{I}_{DE}$ as input and output reconstructed images in the same domain [1]. The architecture of the proposed model is shown in Fig. 5, which can be divided into three parts: FOV Encoder (Section 4.1), Deformable Decoder (Section 4.2), and KPN Block (Section 4.3).

4.1. FOV Encoder

Due to the strong correlation between degradation and spatial information, we propose the FOV Attention Block to integrate spatial features. It calculates spatial attention masks from pixel coordinate matrices [11], then the image features are modulated by the masking through element-wise multiplication and addition (as shown in the bottom right of Fig. 5). Moreover, ResBlocks [15] are applied in each scale to enhance the expression ability.

4.2. Deformable Decoder

Owing to the irregular shapes and sizes of spatially variant PSFs, conventional convolution layers are not efficient in solving these degradations. Because it only takes features from fixed positions and output results. To deal with singularly shaped PSFs, we introduce deformable convolution to calculate information more flexibly. Following the works proposed by [46, 8], Deformable ResBlocks are employed in each scale of the decoder (as shown in the bottom right of Fig. 5).

4.3. KPN Block

In [44], scholars have proved that kernel prediction is more stable than direct prediction and can handle spatially

variant degradation. However, the computation overhead of KPN will increase heavily when the kernel size grows larger. Therefore, we adopt a dilated convolution manner to output the restoration [28] (as shown in the bottom right of Fig. 5). This operation enlarges the influence scope of the predicted kernel with a relatively lower amount of computation.

Because $\{\mathbf{I}_{DE}, \mathbf{I}_{GT}\}$ are pixel-aligned, extreme-quality results can be obtained only with pixel-level fidelity loss. Moreover, we note that perceptual metric [23] will generate authentic results when dealing with natural images. Therefore, the overall combination of the loss function is as follows:

$$L_{total} = \lambda_{mse} \cdot L_{mse} + \lambda_{per} \cdot L_{per}, \tag{6}$$

Where λ_{mse} and λ_{per} are set as 0.1 and 1 empirically.

5. Experiments

To evaluate the proposed technique, we carry out a comprehensive set of experiments aiming at answering the following three questions:

- What are the advantages of our degradation framework over the model-based kernel estimation approach?
- Is the proposed solution superior to the state-of-the-art methods when applied to solve degradation?
- How good is the extreme-quality computational imaging compared to the built-in ISP system?

5.1. Advantages of Degradation Framework

The performance of our degradation framework is analyzed in two ways: *Degradation estimation accuracy* and *Reconstruction performance*. To evaluate the accuracy of the degradation estimation, the model-based deconvolution method [33] and the calibration approach [21] are chosen for comparing. We apply the real estimated kernels of different methods to checkerboards, and $imatest^{®}$ is used to calculate MTF from the degraded checker. As shown in Fig. 6, we plot the MTF50 *vs.* FOV curves, where the size of circles indicates the SNR of the corresponding FOV. The proposed framework accurately simulates the mutational degradation of different FOVs, and it is robust to noise. Yet, the accuracy of other approaches will be significantly affected by the SNR of the estimated region.

To evaluate the reconstruction performance, we calculate the theoretical PSFs of the experimental camera by $Zemax^{®}$ and applied these PSFs to checkerboards by partition convolution. We use the method in [33] and our framework to estimate PSFs on the corrupted checkerboards. The results show in Fig. 7. SSIMs between the estimated kernel and the ground-truth, are noted in the bottom right of each PSF. The visual comparison of PSFs validates the accuracy

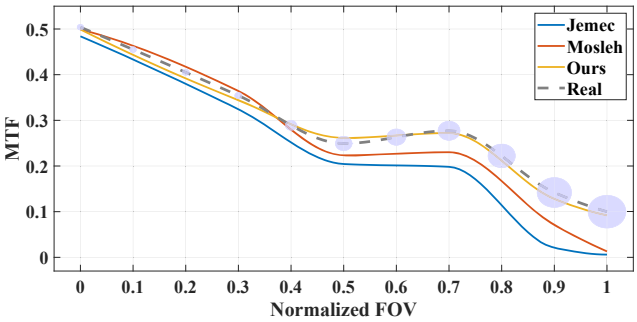

Figure 6. **Degradation MTF comparison.** The MTF50 *vs.* FOV curves of different compared methods are plotted: Jemec [21], Mosleh [33], and Ours. The MTF50 of the experimental camera is calculated by $imatest^{®}$ for reference. And the size of purple circle indicates the SNR of corresponding FOV.

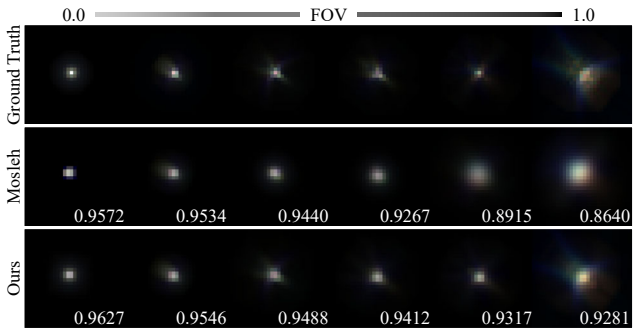

Figure 7. **Kernel estimation and Reconstruction.** The PSFs are arranged from left to right according to the corresponding FOV. We use $Zemax^{®}$ to calculate the ground truth PSFs. And these PSFs are used to corrupt checkerboards. The following PSFs are estimated by [33] and our method when supplied with the corrupted checkerboards. We note the SSIMs of estimations in the bottom right of each PSF.

of the framework, and the SSIMs emphasize the authentic reconstruction performance of our method.

In summary, the proposed degradation framework has two advantages compared with other approaches: (1) adapt to image quality mutation and robust to noise. (2) authentic data generation designed for better deep learning reconstruction. We refer readers to the **supplementary file** for the generalizing capability of our framework.

5.2. Testing on Synthetic Images

To demonstrate the advantages of the proposed method, our technique is compared with many advanced deblur methods, including SRN [43], DeblurGan-V2 (GAN) [25], IRCNN [51], GLRA [40], SelfDeblur [39], KPN [31], and LP-KPN [7]. The training/test dataset consists of 600/50 ground-truth images selected from DIV2K (the degraded images are generated by the proposed Degradation Framework). We retrain all the models with the same data, and

Table 1. **Performance of the proposed model and its ablation study on synthetic data**

Method	PSNR↑		SSIM↑		NIQE↓		LPIPS↓		Times (ms)
SRN	34.27	(17.4%)	0.9602	(19.3%)	2.8917	(31.7%)	0.3386	(63.9%)	102.4
DeblurGAN-v2	32.79	(30.3%)	0.9566	(23.1%)	3.3814	(41.6%)	0.2458	(50.2%)	52.8
IRCNN	34.14	(18.6%)	0.9679	(11.2%)	2.6578	(25.7%)	0.3043	(59.8%)	194.1
GLRA	34.53	(14.9%)	0.9663	(12.8%)	3.2675	(39.5%)	0.3156	(61.2%)	246.6
SelfDeblur	34.96	(10.6%)	0.9709	(8.0%)	3.0735	(35.7%)	0.2707	(54.8%)	264.9
KPN, $k = 19$	35.62	(3.5%)	0.9762	(2.6%)	1.9759	(0.0%)	0.1448	(15.5%)	70.4
LP-KPN	35.58	(3.9%)	0.9764	(2.4%)	2.0929	(5.6%)	0.1410	(13.3%)	53.6
Ours	35.93	(0.0%)	0.9787	(0.0%)	2.0578	(4.0%)	0.1223	(0.0%)	47.4
Ablation study									
Degradation Kernel	33.52	(24.2%)	0.9641	(15.1%)	3.5102	(43.7%)	0.2749	(55.5%)	47.4
Unprocess and Process	35.11	(9.0%)	0.9697	(9.3%)	3.0493	(35.2%)	0.1647	(25.7%)	45.6
Optical Geometric Priors	33.96	(20.3%)	0.9649	(14.3%)	3.3727	(41.4%)	0.2285	(46.5%)	47.4
FOV Encoder	34.69	(13.3%)	0.9684	(10.6%)	3.0684	(35.6%)	0.1622	(24.6%)	36.2
Deformable Decoder	35.34	(6.6%)	0.9769	(1.8%)	2.2928	(13.8%)	0.1245	(1.8%)	44.6
KPN Block	35.26	(7.4%)	0.9743	(4.5%)	2.3059	(14.3%)	0.1301	(6.0%)	45.4

Figure 8. **Real image restoration.** The results of different methods magnify in the right and the positions highlight on the left.

the deconvolution methods are optimized ten times (time comparison on 500×500 color images). The PSNR, SSIM, NIQE [32], and LPIPS [52] indices of these methods are listed in Table 1. One can notice that the combination methods of model-based deconvolution schemes and CNNs (*i.e.* IRCNN, GLRA, SelfDeblur) are overall better than the direct pixel synthesis networks (*i.e.*, SRN, DeblurGan-v2) when tested on synthetic data. However, optical degradation is spatially variant and highly correlate with spatial information. These methods, aiming at solving globally consistent blur, cannot obtain extreme-quality performances in correcting optical degradation. Moreover, we note that KPN and LP-KPN are designed to deal with spatially variant degradation, and they perform better in NIQE. However, directly increasing the kernel size will quadratically aggravate the computational overhead. Utilizing the dilated convolution strategy, our FOV-KPN enlarges the influence scope of the predicted kernel with a relatively small size and low time cost (detailed in the **supplementary file**). And benefiting from the FOV encoder, our method engages the spatial information and obtains better performance than the state-of-the-art methods.

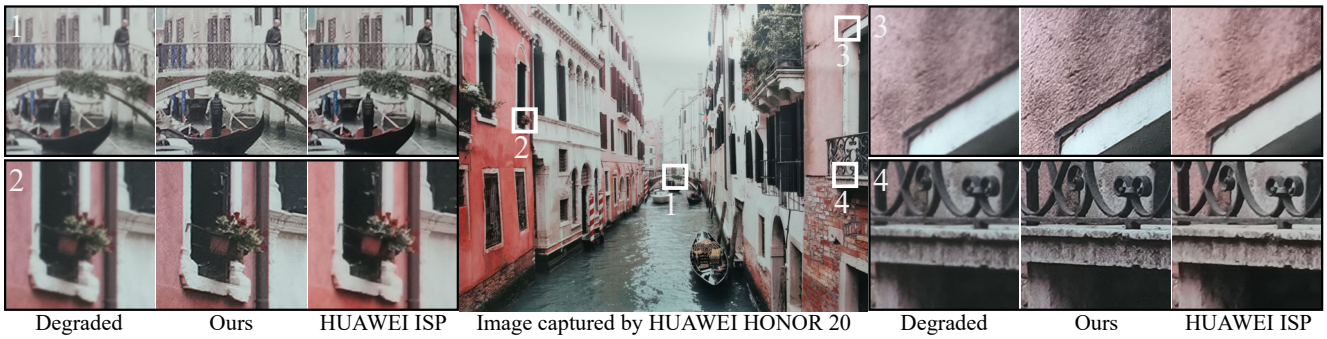

Figure 9. **Postprocessing performance.** Our restoration pipeline compares with the built-in HUAWEI ISP. See Section 5.5 for details.

5.3. Evaluation on Real Photographs

In Fig. 8, we visualize the blur-resolved images. We show two patches of each image, aiming at evaluating the performance of different methods in solving spatially variant blur. As can be seen, the techniques designed for globally consistent degradation tend to generate the compromise of severe blur (edge) and mild degradation (center). The KPN architecture achieves perfect restoration in the center, but it is not effective in dealing with the severe blur in the edge due to the limitation of kernel size. On the contrary, the proposed spatial-adaptive model successfully deals with the deteriorations, resulting in extreme-quality enhancement. More real-world photographs and their quantitative evaluation are shown in the **supplementary file**.

5.4. Ablation Study

In this section, we perform an ablation study on our method to further verify the necessity of each step. The setting of each evaluation is as follows: **Degradation Kernel** indicates that the degradation kernels are replaced by the PSFs calculated by $Zemax^{®}$. **Unprocess and Process** indicates that the unprocess and process operations are removed in the degradation framework. **Optical Geometric Priors** indicates that the receptive field of *deep linear network* is fixed, and symmetry constraint is abandoned. **FOV Encoder**, **Deformable Decoder**, and **KPN Block** indicate that these modules are replaced by convolution layers.

As shown in Table 1, apart from the optical aberrations of lenses, there are manufacturing error, electron diffusion, and aperture restriction that will deteriorate the image quality. Moreover, the unprocess and process operations are crucial because the degradation can be modeled by convolution only in the energy domain. When the optical geometric priors are not applied to the degradation framework, the estimation process is unstable. We conduct a detailed discussion on the optical geometric priors and show it in the **supplementary file**. As for network architecture, the FOV Encoder is an efficient way to utilize spatial information. The Deformable Decoder and the KPN Block effectively deal with spatially variant degradation.

5.5. Applications

Our ultimate goal is to replace the existing hand-made ISP system with the proposed post-processing pipeline (including dark level subtraction, demosaic, white balance, color correction, gamma compression, and reconstruction), achieving extreme-quality computational imaging. Therefore, we further present the comparison between our method and HUAWEI ISP in Fig. 9, both of them perform RAW-to-RGB translation. We apply the same JPEG compression algorithm for the sake of a fair comparison. Owing to the additional sharpening operation, the image post-processed by HUAWEI ISP is excellent in the center of photograph, and our result obtains the same recovery as the built-in ISP. However, sharpening is globally consistent. The advantage of our method is evident when FOV increases. According to the spatial information in off-axis regions, the proposed pipeline recovers the degraded image adaptively, especially in the edge. We refer readers to the **supplementary file** for more comparisons concerning time costs and restoration examples.

6. Conclusion

In this paper, we constructed a degradation framework for estimating the spatially variant PSF of a specific camera, including but not limited to the device that shows image quality mutation. The proposed framework generates authentic imaging results that resemble real-world photographs, where a lot of shooting, registration, and color correction are needless. Then a KPN-based network architecture was designed, aiming at dealing with spatially variant degradation. Our comprehensive experiments validate the advantages of our degradation framework when compared with the model-based deconvolution methods. Moreover, we demonstrate that the proposed network is superior to the state-of-the-art approaches in dealing with spatially varying blur. Furthermore, the proposed method can be embedded in existing ISP systems, resulting in extreme-quality computational imaging. We hope this work could inspire further research in real-world image reconstruction.

References

[1] Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes. Proposal for a standard default color space for the internet—srgb. In *Color and imaging conference*, volume 1996, pages 238–245. Society for Imaging Science and Technology, 1996.

[2] Chaim Baskin, Natan Liss, Yoav Chai, Evgenii Zheltonozhskii, Eli Schwartz, Raja Giryes, Avi Mendelson, and Alexander M Bronstein. Nice: Noise injection and clamping estimation for neural network quantization. *arXiv preprint arXiv:1810.00162*, 2018.

[3] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *Advances in Neural Information Processing Systems*, pages 284–293, 2019.

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

[5] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[6] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.

[7] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019.

[8] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. *arXiv preprint arXiv:2001.10291*, 2020.

[9] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[10] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2018.

[11] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11065–11074, 2019.

[12] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1604–1613, 2019.

[13] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, and Yang Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. *arXiv preprint arXiv:2009.09238*, 2020.

[14] Samuel W Hasinoff. Photon, poisson noise., 2014.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] Felix Heide, Mushfiqur Rouf, Matthias B Hullin, Bjorn Labitzke, Wolfgang Heidrich, and Andreas Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics (TOG)*, 32(5):1–14, 2013.

[17] Jia Hou, Haifeng Li, Zhenrong Zheng, and Xu Liu. Distortion correction for imaging on non-planar surface using freeform lens. *Optics Communications*, 285(6):986–991, 2012.

[18] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017.

[19] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3277–3285, 2017.

[20] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 536–537, 2020.

[21] Jurij Jemec, Franjo Pernuš, Boštjan Likar, and Miran Bürmen. 2d sub-pixel point spread function measurement using a virtual point-like source. *International journal of computer vision*, 121(3):391–402, 2017.

[22] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 466–467, 2020.

[23] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[24] Rudolf Kingslake. *Applied Optics and Optical Engineering V6*, volume 6. Elsevier, 2012.

[25] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019.

[26] Ngai Ming Kwok, HY Shi, Quang Phuc Ha, Gu Fang, SY Chen, and Xiuping Jia. Simultaneous image color correction and enhancement using particle swarm optimization. *Engineering Applications of Artificial Intelligence*, 26(10):2356–2371, 2013.

[27] Xin Li, Bahadir Gunturk, and Lei Zhang. Image demosaicing: A systematic survey. In *Visual Communications and Image Processing 2008*, volume 6822, page 68221J. International Society for Optics and Photonics, 2008.

[28] Yuhong Li, Xiaofan Zhang, and Deming Chen. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1091–1100, 2018.

[29] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3408–3416. IEEE, 2019.

[30] Virendra N Mahajan. Strehl ratio for primary aberrations: some analytical results for circular and annular pupils. *JOSA*, 72(9):1258–1266, 1982.

[31] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2502–2510, 2018.

[32] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[33] Ali Mosleh, Paul Green, Emmanuel Onzon, Isabelle Begin, and JM Pierre Langlois. Camera intrinsic blur kernel estimation: A reliable framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4961–4968, 2015.

[34] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. l_0-regularized intensity and gradient prior for deblurring text images and beyond. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):342–355, 2016.

[35] Jinshan Pan, Wenqi Ren, Zhe Hu, and Ming-Hsuan Yang. Learning to deblur images with exemplars. *IEEE transactions on pattern analysis and machine intelligence*, 41(6):1412–1425, 2018.

[36] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.

[37] Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph.*, 38(6):219–1, 2019.

[38] Sivalogeswaran Ratnasingam. Deep camera: A fully convolutional neural network for image signal processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[39] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3341–3350, 2020.

[40] Wenqi Ren, Jiawei Zhang, Lin Ma, Jinshan Pan, Xiaochun Cao, Wangmeng Zuo, Wei Liu, and Ming-Hsuan Yang. Deep non-blind deconvolution via generalized low-rank approximation. *Advances in neural information processing systems*, 31:297–307, 2018.

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

[42] Warren J Smith. *Modern optical engineering: the design of optical systems*. McGraw-Hill Education, 2008.

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

[44] Thijs Vogels, Fabrice Rousselle, Brian McWilliams, Gerhard Röthlin, Alex Harvill, David Adler, Mark Meyer, and Jan Novák. Denoising with kernel prediction and asymmetric loss functions. *ACM Transactions on Graphics (TOG)*, 37(4):1–15, 2018.

[45] Ching-Chun Wang and Charles G Sodini. A crosstalk study on cmos active pixel sensor arrays for color imager applications. In *Proc. IEEE Workshop on CCDs and Advanced Image Sensors*, 2001.

[46] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[47] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *European conference on computer vision*, pages 157–170. Springer, 2010.

[48] Tao Yue, Jinli Suo, Jue Wang, Xun Cao, and Qionghai Dai. Blind optical aberration correction by exploring geometric and visual priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1684–1692, 2015.

[49] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2529, 2018.

[50] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

[51] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017.

[52] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[53] Wangmeng Zuo, Dongwei Ren, David Zhang, Shuhang Gu, and Lei Zhang. Learning iteration-wise generalized shrinkage–thresholding operators for blind deconvolution. *IEEE Transactions on Image Processing*, 25(4):1751–1764, 2016.