Non-blind optical degradation correction via frequency self-adaptive and finetune tactics

TING LIN, SHIQI CHEN,* ⓘ HUAJUN FENG, ZHIHAI XU, QI LI, AND YUETING CHEN ⓘ

State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310000, China
*chenshiqi@zju.edu.cn

Abstract: In mobile photography applications, limited volume constraints the diversity of optical design. In addition to the narrow space, the deviations introduced in mass production cause random bias to the real camera. In consequence, these factors introduce spatially varying aberration and stochastic degradation into the physical formation of an image. Many existing methods obtain excellent performance on one specific device but are not able to quickly adapt to mass production. To address this issue, we propose a frequency self-adaptive model to restore realistic features of the latent image. The restoration is mainly performed in the Fourier domain and two attention mechanisms are introduced to match the feature between Fourier and spatial domain. Our method applies a lightweight network, without requiring modification when the fields of view (FoV) changes. Considering the manufacturing deviations of a specific camera, we first pre-train a simulation-based model, then finetune it with additional manufacturing error, which greatly decreases the time and computational overhead consumption in implementation. Extensive results verify the promising applications of our technique for being integrated with the existing post-processing systems.

1. Introduction

Due to the development of social media and the speedy transmission of information, mobile photography along with user-friendly image signal processing (ISP) has become the most popular way to capture images. However, a mobile device generally stacks 7 to 8 lenses in limited physical space to correct optical aberrations. Moreover, injection molding technology is applied in the manufacturing pipeline of cellphone cameras to cut costs and accelerate production. As a consequence, cramped physical space limits the variety of optical prescription and injection molding introduces more complex deviations, all of which increase the difficulty of tolerance controlling, resulting in diverse image degradation among cellphones with identical lens designing [1].

As discussed above, the degradation introduced by lens consists of intrinsic optical aberrations and manufacturing errors as well as assembly biases, where the former characterize the differences of a lens to focus light across FoVs and the latter comes from the optical component defects introduced in production. Classically, the aberrations are corrected in optical designing and the system variations are compensated by tolerance analysis, but their solutions are very limited by simulation accuracy, structure complexity, and manufacturing capability. Moreover, eliminating optical aberrations completely through a chunky lens stack is still challenging. Therefore, many scholars have taken insight into computational optics to correct the lens degradation by post-processing, and some prior researches indicate that the computational method is quite efficient to alleviate residual aberrations [2–4]. However, owing to the randomicity in manufacturing procedures, the optical system produced on the same pipeline can also deviate from each other. Therefore, generalizing all the optical devices of the same prescription with a concrete model is

neither accurate nor robust, where more investigation should be invested into the manufacturing deviations.

When the aforesaid real lens degradation is abstracted into a theoretical model, point spread functions (PSFs) are widely used to characterize it. PSF is described as the impulse response of a focused optical system and is widely applied in system evaluation and imaging synthesis. In mathematics, image degradation can be expressed as the convolution of the clean image $I_{clear}(x, y)$ and the spatially variant kernels $k(x, y)$, which can be formulated as follows:

$$I_{deg}(x, y) = \iint_D I_{clear}(x, y) \otimes k(x, y) dx dy + n(x, y), \qquad (1)$$

where $\otimes$ is the convolution operation. $I_{deg}(x, y)$ is the measured image, and $n(x, y)$ is the noise introduced in image formation. x and y indicate the pixel position along the two-dimension. The reconstruction of $I_{clear}(x, y)$ is the popular issue where considerable efforts have been invested in it. Recovery methods comprise non-blind and blind deconvolution as well as learning-based method. For the non-blind method, PSF can be obtained through ray tracing or computational approximation [5–7], then utilize PSF as one of the inputs in subsequent deconvolution. Due to the presence of noise, deconvolution is an ill-posed problem and must rely on empirical image priors, where the relationship between fidelity items and image priors requires manual fine-tuning for different conditions. As for blind method [2,8], a large body of work in these fields is performed through alternately optimizing the PSF and the latent image. Nevertheless, due to the PSFs being spatial-variant of a wide range over regions, high computational overhead prevents these methods from implementation in real-time imaging. Over the past decades, the methods based on deep learning outperform traditional deconvolution. Its accuracy exceedingly depends on the data pairs, where target datasets of a specific camera are required for better reconstruction. Moreover, a single network needs a larger model capacity to handle the spatially varying degradation.

In this work, we propose a plug-and-play method to correct the degradation of diverse cameras, including optical aberrations, manufacturing errors, and assembly biases. Drawing from the deconvolution of removing non-uniform blur by combining single PSF restoration linearly, we develop a deep-learning method, whose linear weights are realized via attention mechanism. To be more specific, we introduce the linear combination of Wiener filter [9] in the feature domain and propose an effective PSF calibration method to support the feature-based deconvolution, bridging the gap between the optical degradation and learning-based model. Moreover, since predicting the signal-to-noise ratio (SNR) of Wiener filtering from the complex real scene is arduous and inaccurate estimation will introduce ringing artifacts, we engage the frequency domain feature with the basic functions to construct SNR, denoted as a frequency-related map. To achieve plug-and-play, we propose to pre-train a base model with intrinsic optical aberrations and then finetune it with measured overall degradation, which conveniently adapts to the target real camera and results in significantly improved visual quality. The contributions of our work are as follows:

- A frequency self-adaptive block (FSB) is proposed for performing feature-based Wiener deconvolution and its linear combination is applied to correct non-uniform blur.

- We design a plug-and-play postprocessing pipeline, which is pre-trained with intrinsic optical aberrations and fine-tuned with overall degradation of the target real camera, realizing a fast adaption to manufacturing deviation and significantly improved quality.

2. Related work

PSF Estimation Image restoration algorithm is divided into non-blind and blind methods, the former requires PSF estimation of which accuracy is closely related to the measured image.

A direct method to obtain PSF is calibration, such as obtaining the impulse response of the sensor through an array pinhole and then eliminating the noise in the measured PSF. However, the calibration requires a lot of time consumption and specific devices. Now many research generally uses a single checkerboard or noise image to estimate PSF, including the iterative and learning methods. The former method involves optimization problems and is suitable for spatially invariant blur kernels. The previous work [1] minimized the L2 norm between the simulated and measured PSF by adjusting lens prescription. In [10], a PSF estimation approach was introduced, where the frequency spectrum of the target image is taken into account. This method benefits from a homogeneous spectrum density function (SDF) of Bernoulli noise patterns to introduce a SDF constraint. However, the optimization process is cumbersome and even can not converge, especially for low signal-to-noise ratio observation [10,11]. The learning method can make up for that deconvolution is not effective for estimation. The fully unsupervised internal-GAN was proposed by using a linear generator in down-sampling PSF estimation [12]. But in some real situations, it is unstable and requires strong constraints added to evaluate more reliable kernels.

Aberration Correction Due to aberration is spatial-variant, a straightforward way is to convert the non-uniform blur into several uniform ones. The work [13] introduced the optical symmetry to radially split FOV and warp technique. In deconvolution, they propose sharp-to-blur strategy to improve restoration performance, but this method may fail in severe nonlinear camera response. Unlike [13] and [4] split image along the radius, the other method is splitting the image into rectangular patches. In [14], authors used rectangular patches for spatial-invariant deconvolution, then apply projection network to suppress ringing and blocking artifacts. In non-blind deconvolution, improper PSF estimation may cause artifacts. To solve it, the work in [15] introduced to generate Gabor filters for each deblurred image, with regards to PSF frequency information, then the response is used as an additional regularization scheme.

Optimization-Deconvolution based Restoration Since image deconvolution is an ill-posed problem, adding priors are essential to improve the performance of restored images. The pioneers of deconvolution are the Wiener filter [9] and Richardson-Lucy algorithm [16]. They assumpted that natural images satisfy Gaussian and Poisson distribution, respectively. Nowadays, a more general prior is hyper-Laplacian which is consistent with heavy-tailed distribution. Thus, many scholars have been exploring more accurate methods for estimating it. A Bayesian minimum mean squared error estimates (MMSE) was adopted to explore favorable high-order priors [17]. Cho, Wang and Lee employed expectation-maximization (EM) method to remove saturation regions under the assumption that blur kernel is spatially invariant [18]. The Gaussian mixture model (GMM) was also used as a prior to fit the heavy-tailed distribution of natural image gradient [19], where Zoran and Weiss proposed a patch-wise prior based on it [20]. Roth and Black used a Products-of-Experts framework to model the PSF characteristic and any Bayesian inference that requires image priors can use this model [21]. But these methods need complex optimization and is usually time-consuming.

Deep-learning based Restoration In addition to traditional methods, deep learning is also applied for image restoration. In paper [22], the deconvolution convolutional neural network (CNN) was described with separable 1D kernels to reconstruct image. However, the network requires to finetune with different inverse kernels initialization. The scholars of [23] provided a convolutional submodel to learn denoised image gradients as image priors. The method resolves deblurring into two subproblems, requiring two training processes: denoising and deconvolution, which fits the physical models but is difficult to apply in real-time imaging. Kruse, Rother and Schmidt proposed an extension of iterative Fourier deconvolution, by introducing CNNs to provide more useful regularization and using a simpler boundary adjustment method [24]. A multi-scale blind deconvolution was proposed in [25], whose each scale contains feature extraction, kernel estimation and image estimation modules. But this work is poor-performed in larger PSF and only handles uniform kernels.

Most of the above methods focus on solving spatially uniform PSFs. As for non-uniform blur, they might require to finetune for every PSFs, which requires extensive computational overhead. Challenges in optical degradation correction lie in the trade-off between computational consumption and performance. Based on the similarity of PSFs in the neighborhood, we combine traditional Wiener filter and CNNs for restoration, aiming to output high-quality images without relying on an enormous network.

3. Overview

Our objective is to correct the optical degradation introduced in the designing and manufacturing process of the lens. To achieve this, we engage the post-processing chain with the random degradation caused by an optical system. The two core steps of the proposed method are as follows: first, we apply an effective PSF calibration method and use the blur kernel to bridge the mismatch between the optical system with the post-processing pipeline. Second, to eliminate the stochastic aberration, we design CNN-based post-processing to self-adaptively perform the mapping between the degraded data and the latent RGB image.

Different from the blur that is uniform across the entire FoV, optical degradation is spatially variant due to the FoV-related optical aberration and the random manufacturing deviation. Fortunately, the natural signal is spatially continuous, so the degradation is similar in some neighborhoods of the image. Assuming the blur and the noise level are the same in one patch $I_{degraded}(x, y)$, the model of its degradation can be formulated as:

$$I_{degraded}(x, y) = I_{latent}(x, y) \otimes k(x, y) + n(x, y) \tag{2}$$

here the $I_{latent}(x, y)$ is the latent image of the patch $I_{degraded}(x, y)$. $\otimes$ is the operation of convolution. $k(x, y)$ is the PSF that indicates the energy diffusion caused by optical degradation. $n(x, y)$ is the stochastic noise introduced during image acquisition. A typical method to retrieve the latent image $I_{latent}(x, y)$ is deconvolution, where a Wiener filter is the most classic approach. By introducing Tikhonov regularization, the wiener deconvolution attempts to restore the latent image from the noisy degraded data. This process is as follows:

$$I_{latent}(x, y) = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(k(x, y))}}{|\mathcal{F}(k(x, y))|^2 + 1/SNR} \cdot \mathcal{F}(I_{degraded}(x, y))\right) \tag{3}$$

where $\mathcal{F}$ denotes the discrete Fourier transform. $\overline{(\cdot)}$ is the conjugate operation. SNR is the signal-to-noise ratio of the patch, it varies with spatial frequency. However, the statistical properties (PSF of different FoV and SNR of different frequencies) are difficult to measure when only a degraded patch is given. Therefore, we separate the solution into PSF calibration and self-adaptive correction. The PSF calibration aims to predetermine the PSF of optical degradation (illustrated in Sec. 4.1) and the self-adaptive correction is designed for addressing the problem of statistical properties in different frequencies (detailed in Sec. 4.2).

4. Method

4.1. PSF calibration

Supposing the degradation of the natural image is similar in some neighborhoods, we crop a patch $I_{degraded}(x, y)$ from the checkers taken by a specific camera. Then, the patch $I_{degraded}(x, y)$ is transformed into the latent image $I_{latent}(x, y)$ by the same method as [26]. In this way, we obtain the data pairs $\{I_{latent}(x, y), I_{degraded}(x, y)\}$ for PSF calibration. Furthermore, the normalized FoV (calculated by (x, y)) is prepared for accurate PSF calibration in the energy domain.

In PSF calibration, we adopt the *deep linear model* to transfer the optical degradation from $I_{degraded}(x, y)$ to $I_{latent}(x, y)$ [26]. To be more specific, the *deep linear model* is feed with the perfect

image $I_{latent}(x, y)$, and the output $I^d_{latent}(x, y)$ is supervised by the degraded patch $I_{degraded}(x, y)$. Therefore, the *deep linear model* represents the mapping from the perfect scene to the degraded image received on the sensor. Moreover, the *deep linear model* is also supervised with the PSF characteristics it represents. In one training step, a two-dimensional impulse function is feed into the model besides the $I_{latent}(x, y)$, and the output PSF $k(i, j)$ is supervised by two loss functions to ensure the sum of PSF equals to 1 and penalize the singular values in the edge. The overall loss function of the *deep linear model* is expressed as follows:

$$\mathcal{L} = \alpha ||I^d_{latent} - I_{degraded}||_1 + \beta |1 - \sum_{i,j} k(i,j)| + \gamma \sum_{i,j} |k(i,j) \cdot m(i,j)| \tag{4}$$

here α, β, γ are the trade-off parameters of different loss functions. The first loss is to supervise the output pixel-by-pixel, aiming to ensure the fidelity of the model. The second and the third loss are used to constrain the characteristics of the PSF generated by the model, where (i, j) is the coordinates of the PSF and $m(i, j)$ is a 2D Gaussian mask. We note that the $m(i, j)$ in [26] is a constant matrix, whose weights exponentially grow with distance from the center of k. However, due to the off-axis aberrations, the PSF spreads differently in tangential and sagittal directions. Therefore, we apply the heteroscedastic Gaussian mask to constrain the asymmetric diffusion. With knowing the geometric spot radius in tangential r_{tan} and sagittal r_{sag} of this FoV, the mask is first formulated as follows:

$$m(i, j) = \frac{1}{2\pi\sigma_1\sigma_2} exp[-\frac{1}{2}(\frac{i^2}{\sigma_1^2} - \frac{2ij}{\sigma_1\sigma_2} + \frac{j^2}{\sigma_2^2})], \frac{\sigma_2}{\sigma_1} = \frac{r_{tan}}{r_{sag}} \tag{5}$$

where σ_1 and σ_2 are the variances of the distribution. σ_1 is the same as [26] and σ_2 is calculated by the ratio of geometric spot radius. Then we rotate the mask according to the angle of this FoV (as shown in Fig. 1) and use $m(i, j)$ to penalize the singular values at the edge. After training the *deep linear model* of all (x, y), the PSFs of different FoVs are the output of each model when inputs the impulse function. In this way, we predetermine the PSF k by calibration, where the k covers all the optical degradation in image formation.
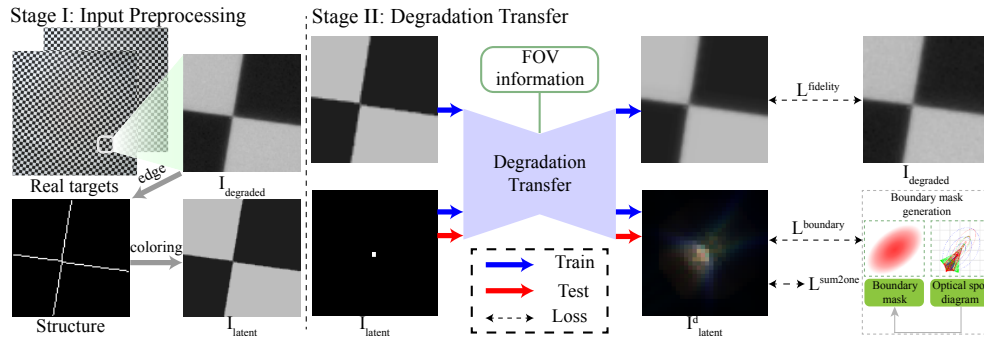

Fig. 1. PSF calibration is devided into input processing (Stage I) and degradation transfer (Stage II). In Stage I, the latent checker is generated by the degraded measurement. In Stage II, the latent image together with the impulse function are feed into *deep linear model*, and the output is supervised by fidelity loss and kernel-related constraints.

4.2. Self-adaptive reconstruction

As mentioned in Section, the statistical properties of input are difficult to measure when only degraded data is given. So for the spatial-variant degradation, the latent image is generally

modeled as the linear combination of a series of deconvolutional results [27]. When the PSFs k_i of the entire field are given, the reconstruction of a latent image can be formulated as follows:

$$I_{latent}(x, y) = \sum_i \mathcal{F}^{-1}(\eta_i \cdot \frac{\overline{\mathcal{F}(k_i)}}{|\mathcal{F}(k_i)|^2 + 1/SNR} \cdot \mathcal{F}(I_{degraded}(x, y))) \tag{6}$$

where η_i is the weight of k_i at different (x, y). Traditionally, the $1/SNR$ is usually fixed or calculated by Gaussian distribution. However, the real distributions of signal and noise are more sophisticated, *e.g.*, the shading introduced by the optical system will let the signal received by the center of the sensor stronger than the edge, so the shot noise and read-out noise dominate at the center and edges, respectively.

Due to the complexity of the noise statistical properties, we explore using the deep learning method to predict $1/SNR$ according to the input data. For convenience, a fast Fourier transform (FFT) map $\epsilon(u, v)$ is used to describe the $1/SNR$ at different spatial frequencies, where (u, v) indicates the coordinates of spatial frequency. The advantage of this representation is that it can self-adaptively suppress or enhance the intensity of different spatial frequencies in the Fourier domain. We express the $\epsilon(u, v)$ as a linear combination of basic functions to represent the $1/SNR$ at different spatial frequencies:

$$\epsilon(u, v) = \sum_j \phi_j \cdot |\mathcal{F}(p_j)|^2 \tag{7}$$

here p_j denotes the gaussian basic functions set consisting of the gaussian kernels of different standard deviations. ϕ_j is the weight of different basic functions. Therefore, when the PSFs are given and the linear representation of $1/SNR$ can be predicted, the accurate reconstruction for spatially-variant degradation can be obtained as follows:

$$I_{latent}(x, y) = \sum_i^m \mathcal{F}^{-1}(\eta_i \cdot \frac{\overline{\mathcal{F}(k_i)}}{|\mathcal{F}(k_i)|^2 + \sum_j^n \phi_{ij} \cdot |\mathcal{F}(p_j)|^2} \cdot \mathcal{F}(I_{degraded}(x, y))) \tag{8}$$

In this way, we realize a self-adaptive reconstruction of non-uniform degradation. When implement in the deep-learning method, a frequency self-adaptive block (FSB) is developed for performing deconvolution in the Fourier domain. We feed the FSB with the image feature and a set of calibrated PSFs and output the image feature after reconstruction. As shown in Fig. 2, the proposed FSB is composed of two main parts: PSF-combining deconvolution and Frequency self-Adaptive modification (framed by red dotted lines). In the first part, the input PSFs k_i are engaged with the image feature as the form of Eq. (8) and the weight η_i of k_i is obtained by performing spatial attention and channel attention on the input image feature g:

$$\{\eta\} = SpatialAttention(g) \cdot ChannelAttention(g), \tag{9}$$

And in the second part, the weight map ϕ_j of different basic functions are predicted by the FFT map of the image feature, which can be formulated as follows:

$$\phi = SoftMax((Conv(AvgPool(\mathcal{F}(g))))), \tag{10}$$

where *AvgPool*, *Conv*, *SoftMax* denotes average pooling, convolution and softmax function. After calculating the weight of PSF and basic functions, the reconstruction feature can be obtained by Eq. (8). Therefore, with the predetermined PSFs (details in Sec. 4.1), the FSB reconstructs the degraded feature in the Fourier domain, suppressing or enhancing the feature self-adaptively.
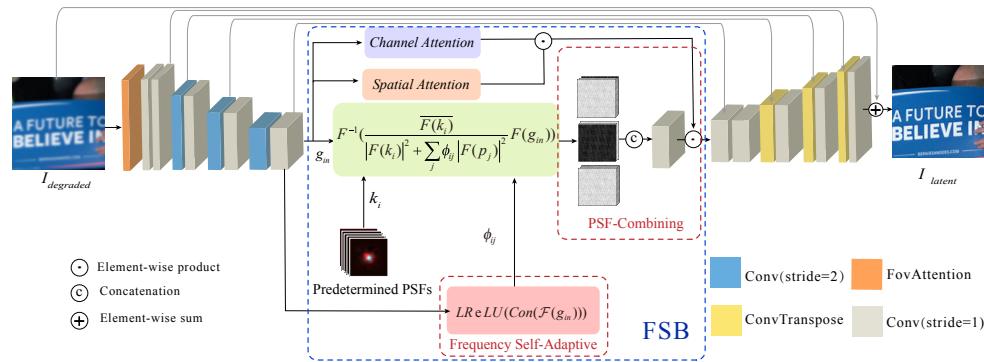
Fig. 2. Optical Degradation Correction Model aims to perform frequency self-adaptive restoration on the input data. It engage with the predetermined PSFs to deconvolution and predict the targeted 1/*SNR* map.

4.3. Details in network

In this subsection, we illustrate some details of the network. As shown in Fig. 2, the network structure is based on U-Net with skip connection and ResBlock is applied for more sophisticated expression. The network model includes an encoder and decoder with four scales for feature and image domain transferring. FSB is located at the bottom of the model framed by a blue dotted line, which is the key element of reconstructing a degraded image. Since the spatially variant PSFs are highly correlated with the FoV of the input image, we adopt the FoV attention module as [26] to let the model perceive the spatial information. In the following, two attention modules in PSF-combining deconvolution and loss function are illustrated. Moreover, for a non-blind reconstruction, containing all the predetermined PSFs (150*200 across the whole FoV) as inputs are redundant, we design a strategy for PSFs reduction.

Channel attention. Due to the concatenation of multiple PSFs restoration, it is worth exploiting the inter-channel relationship of features. Different PSFs deconvolution tends to recover different frequency information. Specifically, small size PSF deconvolution results seem to be smoother, but large size PSF deconvolution results generate more high-frequency components. Thus we design a channel attention module to fuse information of different frequencies. Like [28], global average pooling is presented to squeeze spatial information, then forward to double convolution layers. ReLU and sigmoid function are activate-functions in the middle layer and output layer, details in the left of Fig. 3.
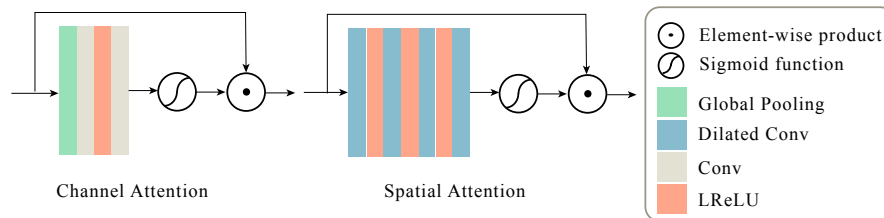

Fig. 3. Channel Attention and **Spatial Attention**.

Spatial attention. Spatial attention is complementary to channel attention. In spatial dimension, textures are diverse with coordinates changing, thus spatial attention is introduced to update a proper value in the position. Following the works proposed by [29], spatial attention module stacks a set of dilated convolution layers, particulars in the Fig. 3. Compare with

the conventional convolution layer, it broadens the receptive field without introducing more parameters.

Loss function. We employ the mean absolute error (MSE) as the fidelity loss and add perceptual loss to induce more high-frequency details.

$$L_{total} = \lambda_{mse} L_{mse} + \lambda_{per} L_{per}, \tag{11}$$

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

Predetermine PSFs. (1) PSFs reduction strategy. Considering that aberrations are locally highly correlated, both size and shape of them are similar. Therefore, inputting all PSFs is not only generates huge redundancy but also may cause overfitting in subsequent restoration. We design a mask to sample them, retaining more PSFs in the edge and fewer PSFs in the center. Together with the symmetry of the optical system, upper-left/lower-left/upper-right/lower-right degradation is quite similar, so a quarter of PSFs almost represents all FoV degradation. Due to the randomness of the selection, the attention mechanism parameters intend to capture more helpful issues in PSFs combination. Furthermore, this strategy also generates a stabler model, which facilitates faster adaptation to new PSFs in subsequent fine-tuning. (2) In order that PSFs and images are at the same scale, PSFs are downsampled four times to realize scale matching.

5. Experiments

5.1. Analysis

We adopt DIV2K [30] dataset which contains 800 images of 2K resolution for evaluation. It is divided into the training set, validation set and test set at 3:1:1 for processing. The procedure we construct image pairs is as follows. First, we calculate 30000 PSFs (we have divided the whole FoV into 150*200, then calculated the PSF by FoV) by Sec. 4.1 for HUAWEI HONOR 20 pro. Then, clean images are convoluted with the obtained PSFs and add 1% and 2% Gaussian noise to generate data pairs.

In network training, the data is cropped by the patch size of 200*200 and the batch size set to 16 for each iteration. We use the Adam optimizer with an initial learning rate of 10^{-4}. The training procedure is terminated after iterating 100 epochs in total and we decrease the learning rate by half after every 10 epochs. We use PyTorch implementation on a single NVIDIA GTX 1080 Ti GPU. In training process, when the batch size sets 16, the corresponding memory is about 5G. By eliminating similar PSFs with the strategy proposed in Sec. 4.3, the number of input PSFs for recovery decreases from 30000 to 375.

In Sec. 5.2, we employ simulation data to calculate objective evaluation value, which highlights the superiority of our method with other reconstruction methods. Experiments are carried out for spatially variant PSFs and spatially invariant PSFs. In Sec. 5.3, we ablate some modules in the original network to demonstrate their necessity. In Sec. 5.4, 5 mobile phones data are utilized to valid the finetune trick, and Sec. 5.5, we compare our method with HUAWEI ISP.

5.2. Qualitative evaluation

Our method is compared with some advanced deblurring algorithms, including DeblurGANv2 [31], DPDNN [32], DWDN [33], SRN [34]. The PSNR, SSIM, and LPIPS [23] value of these methods are listed in Table 1 and Table 2. The PSNR, SSIM and LPIPS are averaged over the whole FoV. DWDN is a non-blind restoration algorithm and others are blind recovery.

We evaluate the effects of uniform and non-uniform blur separately, all methods using the same dataset. For the uniform blur, we randomly pick PSF from the PSF set (the PSF set contains various PSFs calibrated in Sec. 4.1), then convolve sharp images to build uniform-blur image pairs. For the non-uniform blur, we calculate PSFs of a mobile phone through Sec. 4.1 methods, then the clear image is degraded by these PSFs (150*200) in different FOV.

Table 1. Comparison with advanced methods, PSF is spatial-invariant.

Method	PSNR ↑		SSIM ↑		LPIPS ↓		time(s)
Input	29.12		0.9111		0.1671		
SRN	28.1	(59.4 %)	0.9133	(5.4%)	0.1486	(32.6%)	3.0145
DPDNN	31.59	(16.0%)	0.9381	(2.0%)	0.1126	(2.8%)	6.0183
DWDN	32.75	(5.1%)	0.9574	(0.8%)	0.1552	(38.4%)	12.8729
DeblurGANv2	31.5	(16.9%)	0.9226	(4.4%)	0.1586	(41.5%)	1.2412
Ours	33.36	(0.0%)	0.9650	(0%)	0.1104	(0.0%)	1.0543

Table 2. Comparison with advanced methods, PSF is spatial-variant.

Method	PSNR ↑		SSIM ↑		LPIPS ↓		time(s)
Input	27.57		0.8970		0.1849		
SRN	28.40	(76.0%)	0.9151	(5.8%)	0.1595	(115.2%)	3.2490
DPDNN	31.35	(34.4%)	0.9562	(1.6%)	0.1092	(47.3%)	6.6355
DeblurGANv2	28.09	(81.2%)	0.8509	(12.4%)	0.2654	(258.0%)	1.2446
Ours	34.97	(0%)	0.9720	(0%)	0.0741	(0%)	1.0945

Table 3. Performance of the proposed model and its ablation study on synthetic data.

Method	PSNR ↑		SSIM ↑		LPIPS ↓	
Input	27.71		0.9031		0.1685	
whole Network	35.89	(0.0%)	0.9786	(0.0%)	0.0741	(0.0%)
Corresponding PSFs	35.34	(4.3%)	0.9620	(1.7%)	0.0880	(18.7%)
FoV Encoder	34.55	(10.9%)	0.9734	(0.5%)	0.0872	(17.7%)
PSF DownSampling	33.29	(22.7%)	0.9754	(0.3%)	0.0871	(17.5%)
Adaptive ϵ	35.78	(0.8%)	0.9775	(0.1%)	0.0842	(13.6%)
2 FSBs	35.84	(0.4%)	0.9760	(0.3%)	0.0864	(16.6%)

Table 1 shows results with spatial-invariant degradation, our method is better than three blind methods. DeblurGANv2 has a sharper edge but emerges color spots in some areas which affect the overall appearance. In indices, our method is better than the non-blind method DWDN and less time consuming. Additionally, DWDN is relatively smoother compared with other methods, as shown in Fig. 4.

input   label   SRN   DPDNN   DWDN   DeblurGANv2   Ours
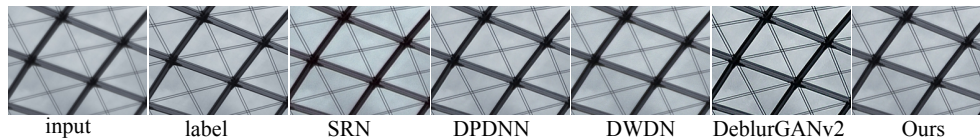
Fig. 4. Spatial-invariant PSFs aberration correction results with 1% additive noise.

For spatial-variant PSFs, more severe degradation occurs when FoV gets large, hence image reconstruction in the edge of the image is tougher. SRN, DeblurGANv2, and DPDNN handle degraded images globally, which lead to compromised restoration and bring unoptimized texture on the edge. Benefitting from the FoV encoder and spatial attention module, our method efficiently engages the spatial information and achieves better performance. Especially on the marginal FoV, the superiority of our method is more obvious as shown in Fig. 5.
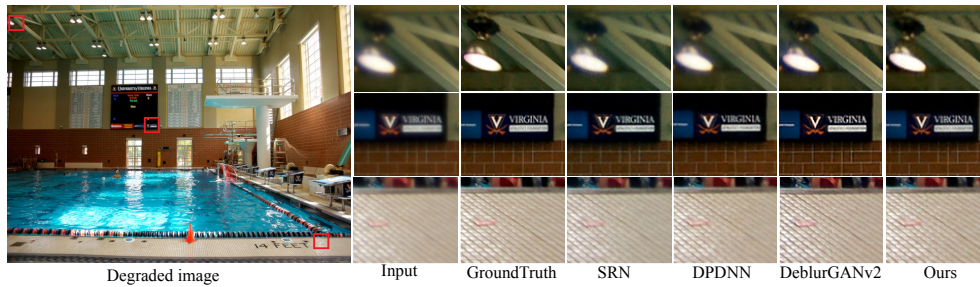
Degraded image   Input   GroundTruth   SRN   DPDNN   DeblurGANv2   Ours

Fig. 5. Spatial-variant PSFs aberration correction results with 1% additive noise.

5.3. Ablation

A comprehensive ablation study is performed to verify that every step in our method is necessary, as shown in Table 3. The dataset of ablation study is degenerated by Gaussian kernels.

Corresponding PSFs: The corresponding PSFs input is conducive to image restoration. The corresponding PSFs of the dataset are asymmetric and we utilize symmetric Gaussian kernels as mismatch input PSFs to validate the work of PSF calibration. **FoV encoder:** PSF is related to the FoV, therefore the coordinates of the pixel are encoded/removed in the input information. **PSF downsampling:** PSF downsampling matches the scale of the image in the network and we input the PSF in the original scale for ablation. **Adaptive ϵ:** To valid Eq. (7), we set ϵ a constant: 0.05 in ablation experiment. **2 FSBs:** To prove the sufficiency of a single FSB, we adapt FSB respectively to the last two scales for processing feature in ablation.

Frequency domain adaptive filter: We validate the basis function fitting data. For a patch, the range of the maximum of ϵ: 0.3~0.5, the range of the minimum: -0.07~-0.04, and the average range is 0.01~0.06. Typical ϵ value in traditional Wiener filter is 0.01, 0.1, etc. Therefore, the ϵ estimated by linear fitting is within a reasonable interval of empirical values. Basic function fitting is feasible and can predict ϵ for different frequencies pertinently.

5.4. Finetune in 5 mobile phones

Image degradation is mainly composed of two aspects: intrinsic optical aberrations and manufacturing error. Even with the identical optical design, the existence of manufacturing error will cause the cameras' PSFs a bit different from each other. If every mobile phone trains the specific model with its own data, it costs too much time and computational overhead. Therefore, we propose to pre-train the base model with optical design degradation, then finetune it with the dataset designed for a specific phone.

Our experiment device is the HUAWEI HONOR20 pro. The configurations of the experiment are illustrated as follows: (a) With the simulation PSFs (only optical design aberration included) to build the dataset for pre-training, and then apply calibrated PSFs (optical design aberration plus manufacturing error) to build the dataset for finetuning. (b) The whole process is trained with the calibration PSFs dataset. Comparing (a) and (b) results in Fig. 6, both objective evaluation and subjective perception of human eyes is not significant. (a) training epochs are less than 25 and (b) training epochs are about 60. Experiments show the finetune method is feasible with fewer iterations.

Moreover, to show the generalization of finetuning processing, we conducted experiments on 5 mobile phones as shown in Table 4. In Table 4, the calculation of PSNR, SSIM, LPIPS is calculated on the simulation dataset. The degradation of simulation dataset is generated by 5 mobile phones PSFs, which are calculated by Sec. 4.1, respectively. Experiments (a) and (b) are recorded as Dire and Finetune and the numbers in the first row indicate the indices of mobile
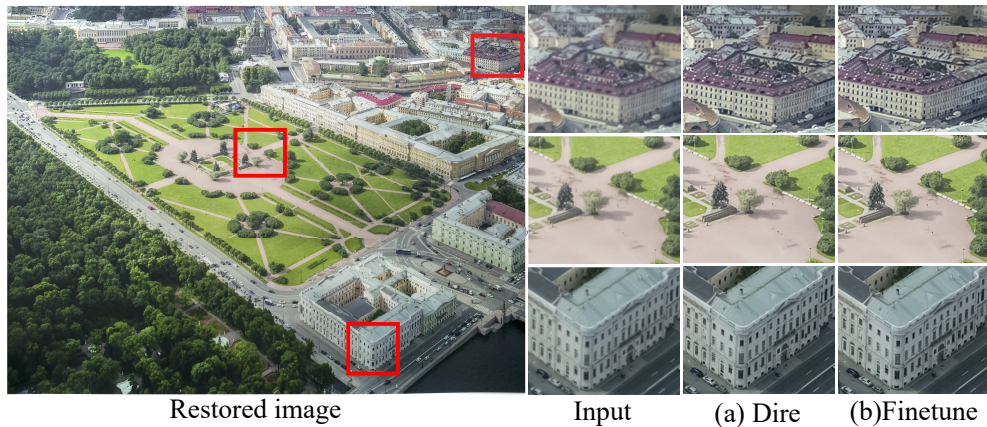
| | Restored image | Input | (a) Dire | (b)Finetune |

Fig. 6. Direct training results compared with finetune results. Test image is real captured image.

phones in Table 4. Therefore, finetuning with a shorter time can achieve closer to or even better results than direct training on a large-scale dataset designed for a specific camera.

Table 4. Finetune evaluations of simulated images on 5 mobile phones.

Evaluation	Experiment	1	2	3	4	5
PSNR	Dire	31.21	31.28	30.58	30.94	31.09
	Finetune	31.22	31.45	30.72	31.15	31.32
SSIM	Dire	0.9369	0.9249	0.9143	0.9207	0.9226
	Finetune	0.9372	0.9262	0.9152	0.9227	0.9249
LPIPS	Dire	0.1888	0.2144	0.2368	0.2211	0.2184
	Finetune	0.1918	0.2156	0.2378	0.2225	0.2187

Besides, to show the performance of our method in real images, we add the evaluation of no-reference metric NIQE on real photographs, as show in Table 5.

Table 5. NIQE metrics of real image test results of 5 mobile phones.

NIQE	1	2	3	4	5
input	5.2590	6.2921	6.3884	6.3241	6.1219
output	4.9036	5.4365	5.1196	5.1788	4.8939

5.5. Application

The goal of our correction is to replace the existing ISP post-processing pipeline and achieve the ultimate image quality improvement. Therefore, we compare our restoration with HUAWEI ISP image. To show the robustness of the proposed algorithm when applied on mobile phone, we evaluate the performance on 5 manufacturing samples. Due to space constraints, the recovery of 2 mobile phones are shown in Fig. 7. Owing to the stochastic deviation introduced in manufacturing, one can see that the degraded images of each samples are different from each other. Because the output of ISP is JPEG format, we post-process our results with JPEG compression for fair evaluation. In the center of outcome, our restorations are comparable to the

built-in performance because Huawei ISP includes additional sharpening. However, as the FoV increase, the advantages of our algorithm show up. Especially in the edge, the textures of our restorations are more realistic and free from artifacts. In addition to the ability for correcting spatially varying degradation, the proposed method also performs adaptive adjustment for each processing samples. When comparing the outcomes on Camera2 and Camera3, we note that the Huawei ISP performs the same processing on sensor measurements, ignoring the manufacturing bias between each samples. On the contrary, the proposed method engages with the predetermined optical degradation to carry out a self-adaptive processing in frequency domain, resulting in better adjustment for different camera.
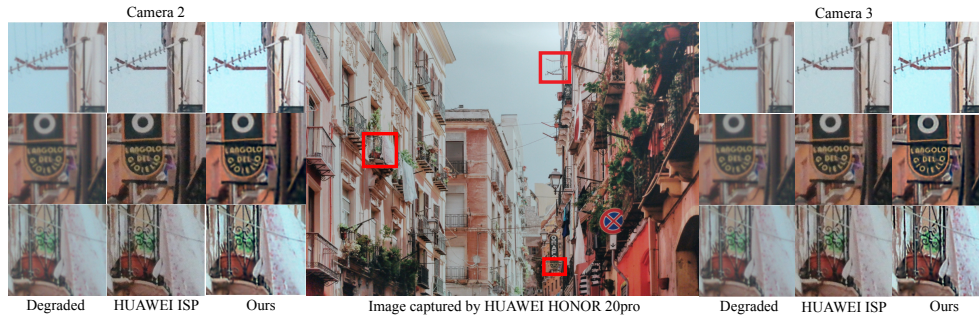

Fig. 7. Our restoration pipeline is compared with the built-in HUAWEI ISP. See Sec. 5.5 for details. Test image is real captured image.

To better quantify the image quality enhancement of the proposed method for different mobile camera samples, we measure the 0.3 FoV spatial frequency response (SFR) on checkerboards for comparing. As shown in Fig. 8, different cameras have various degradation because of deviations, but they are enhanced to the proximate image quality after restoration. Moreover, owing to the proposed fine-tuning tactics, our method has the potential to realize fast adaptation to each mobile cameras in mass production. As illustrated in Sec. 5.4, our training strategy achieve better
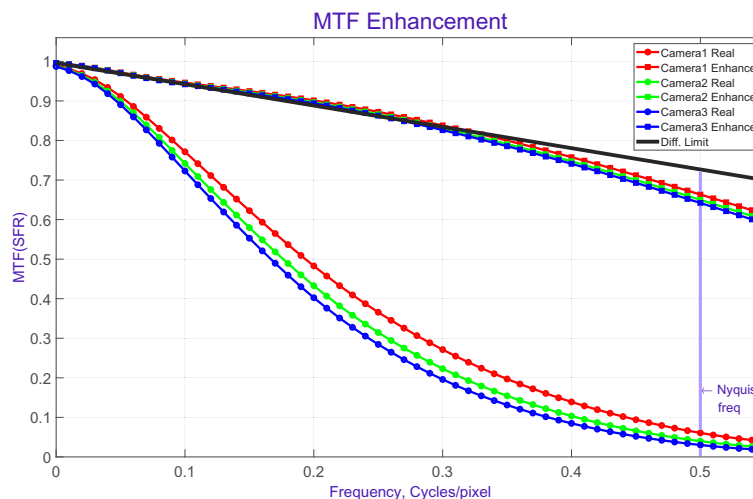

Fig. 8. SFR (MTF) enhancement of different manufacturing samples. "Real", "Enhance" are MTF measured from the degraded checkerboards and the restorations, respectively. "Diff.Limit" is the diffraction limitation of lens prescription.

quantitative indicators with only 40% time consuming, which makes it possible for performing targeted restoration on each mobile camera in mass production.

6. Conclusion

In this paper, a frequency self-adaptive block for non-blind image restoration was proposed. We add the attention mechanism module and the FoV module to extract spatial information and deal with non-uniform optical degradation. The FSB module combined traditional filters with learning mechanisms to restore features in the frequency domain, generating ringing-free and high-frequency details across the whole FoV of the camera. Experiments verified that our method is more efficient and superior to the advanced methods in solving optical degradation. Moreover, it could be quickly fine-tuned for manufacturing deviations, mitigating the difficulty in deployment on mass production. Compared with HUAWEI ISP, the proposed method is capable to replace existing ISPs to achieve image quality enhancement. Benefiting from our finetune strategy, we hope to implement the proposed method in real mobile devices to realize targeted and robust image reconstruction in the future.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. Y. Shih, B. Guenter, and N. Joshi, "Image enhancement using calibrated lens simulations," in *European conference on computer vision*, (Springer, 2012), pp. 42–56.
2. 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. **38**(6), 1–14 (2019).
3. K. Rahbar and K. Faez, "Blind correction of lens aberration using zernike moments," in *2011 18th IEEE International Conference on Image Processing*, (IEEE, 2011), pp. 861–864.
4. C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Blind correction of optical aberrations," in *European Conference on Computer Vision*, (Springer, 2012), pp. 187–200.
5. F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb, "High-quality computational imaging through simple lenses," ACM Trans. Graph. **32**(5), 1–14 (2013).
6. E. Kee, S. Paris, S. Chen, and J. Wang, "Modeling and removing spatially-varying optical blur," in *2011 IEEE international conference on computational photography (ICCP)*, (IEEE, 2011), pp. 1–8.
7. J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 1628–1636.
8. A. Ignatov, L. Van Gool, and R. 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*, (2020), pp. 536–537.
9. N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications Application: With Engineering Applications* (MIT press, 1949).
10. A. Mosleh, P. Green, E. Onzon, I. Begin, and J. Pierre Langlois, "Camera intrinsic blur kernel estimation: A reliable framework," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 4961–4968.
11. J. Jemec, F. Pernuš, B. Likar, and M. Bürmen, "2d sub-pixel point spread function measurement using a virtual point-like source," Int. J..Comput. Vis. **121**(3), 391–402 (2017).
12. S. Bell-Kligler, A. Shocher, and M. Irani, "Blind super-resolution kernel estimation using an internal-gan," Advances in Neural Information Processing Systems **32** (2019).
13. T. Yue, J. Suo, J. Wang, X. Cao, and Q. Dai, "Blind optical aberration correction by exploring geometric and visual priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2015), pp. 1684–1692.
14. X. Li, J. Suo, W. Zhang, X. Yuan, and Q. Dai, "Universal and flexible optical aberration correction using deep-prior based deconvolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), pp. 2613–2621.

15. A. Mosleh, J. Langlois, and P. Green, "Image deconvolution ringing artifact detection and removal via psf frequency analysis," in *European Conference on Computer Vision*, (Springer, 2014), pp. 247–262.

16. W. H. Richardson, "Bayesian-based iterative method of image restoration," J. Opt. Soc. Am. **62**(1), 55–59 (1972).

17. U. Schmidt, K. Schelten, and S. Roth, "Bayesian deblurring with integrated noise estimation," in *CVPR 2011*, (IEEE, 2011), pp. 2625–2632.

18. S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *2011 International Conference on Computer Vision*, (IEEE, 2011), pp. 495–502.

19. R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *ACM SIGGRAPH 2006 Papers*, (2006), pp. 787–794.

20. D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *2011 International Conference on Computer Vision*, (IEEE, 2011), pp. 479–486.

21. S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2 (IEEE, 2005), pp. 860–867.

22. L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," Advances in neural information processing systems **27** (2014).

23. K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), pp. 3929–3938.

24. J. Kruse, C. Rother, and U. Schmidt, "Learning to push the limits of efficient fft-based image deconvolution," in *Proceedings of the IEEE International Conference on Computer Vision*, (2017), pp. 4586–4594.

25. C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," IEEE Trans. Pattern Anal. Mach. Intell. **38**(7), 1439–1451 (2016).

26. S. Chen, H. Feng, K. Gao, Z. Xu, and Y. Chen, "Extreme-quality computational imaging via degradation framework," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), pp. 2632–2641.

27. Y. Bando and T. Nishita, "Towards digital refocusing from a single photograph," in *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, (IEEE, 2007), pp. 363–372.

28. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, (2018), pp. 286–301.

29. H. Son, J. Lee, S. Cho, and S. Lee, "Single image defocus deblurring using kernel-sharing parallel atrous convolutions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), pp. 2642–2650.

30. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2 (IEEE, 2001), pp. 416–423.

31. O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), pp. 8878–8887.

32. W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," IEEE Trans. Pattern Anal. Mach. Intell. **41**(10), 2305–2318 (2019).

33. J. Dong, S. Roth, and B. Schiele, "Deep Wiener deconvolution: Wiener meets deep learning for image deblurring," Advances in Neural Information Processing Systems 33, 1048–1059 (2020).

34. X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 8174–8182.