mathematics

MDPI

Article

Decoupling Induction and Multi-Order Attention Drop-Out Gating Based Joint Motion Deblurring and Image Super-Resolution

Yuezhong Chu, Xuefeng Zhang and Heng Liu *

School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243002, China; yzchu@ahut.edu.cn (Y.C.); zxf_06@ahut.edu.cn (X.Z.)
* Correspondence: hengliu@ahut.edu.cn

Abstract: Resolution decrease and motion blur are two typical image degradation processes that are usually addressed by deep networks, specifically convolutional neural networks (CNNs). However, since real images are usually obtained through multiple degradations, the vast majority of current CNN methods that employ a single degradation process inevitably need to be improved to account for multiple degradation effects. In this work, motivated by degradation decoupling and multiple-order attention drop-out gating, we propose a joint deep recovery model to efficiently address motion blur and resolution reduction simultaneously. Our degradation decoupling style improves the continence and the efficiency of model construction and training. Moreover, the proposed multi-order attention mechanism comprehensively and hierarchically extracts multiple attention features and fuses them properly by drop-out gating. The proposed approach is evaluated using diverse benchmark datasets including natural and synthetic images. The experimental results show that our proposed method can efficiently complete joint motion blur and image super-resolution (SR).

Keywords: motion deblurring; image super-resolution; multi-order attention; gated learning; decoupling

MSC: 37M99

Citation: Chu, Y.; Zhang, X.; Liu, H. Decoupling Induction and Multi-Order Attention Drop-Out Gating Based Joint Motion Deblurring and Image Super-Resolution. *Mathematics* **2022**, *10*, 1837. https://doi.org/10.3390/math10111837

Academic Editors: Jianping Gou, Weihua Ou, Shaoning Zeng, Lan Du and Catalin Stoean

Received: 11 March 2022
Accepted: 24 May 2022
Published: 26 May 2022

1. Introduction

Motion blur and resolution decrease are the two dominant forms of image quality degradation. The former is caused by the relative motion between the camera and the object, while the latter is generally originated by down-sampling. The inverse processes of these degradation forms are individual motion deblurring and image SR—recovering clear images from blurred ones or reconstructing high-resolution (HR) images from low resolution (LR) ones, respectively, which are the practical main means to deal with image quality degradation.

Assuming the original sharp image is $\boldsymbol{x}$, and the blurred image is $\boldsymbol{y}$; if ignoring the effect of the non-linear camera response function (CRF), theoretically the motion blur degradation may be represented as:

$$\boldsymbol{y} = (\boldsymbol{x} * \boldsymbol{h}) + \boldsymbol{n}\,, \tag{1}$$

where $\boldsymbol{h}$ represents the motion blur kernel, $*$ denotes the convolution operation and $\boldsymbol{n}$ usually indicates the random noise. According to Equation (1), obviously, the inverse motion deblurring process is a typical ill-conditioned problem because for one clear image there are possibly many blur images corresponding to it.

Actually, there are two different implementation methods for motion deblurring, namely, blind deblurring or the non-blind method. The usual non-blind method acquires the clear image $\boldsymbol{x}$ based on the estimated blur kernel and the observation $\boldsymbol{y}$. However, the

major difference in blind deblurring is that no kernel estimation is required. Due to the end-to-end mapping and the powerful approximation properties, CNNs are particularly well suited for blind motion deblurring. For example, some recent CNN-based works [1,2] are presented for blind deblurring duties.

Compared with motion blur degradation, in the process of resolution degeneration, in addition to the low pass blur filter $\boldsymbol{k}$ and the noise $\boldsymbol{n}$, there is another degradation down-sampling operator at work. For an HR image $\boldsymbol{x}$, a typical resolution degradation model to acquire the corresponding LR image $\boldsymbol{y}$ is formulated as:

$$\mathbf{y} = (\mathbf{x} * \boldsymbol{k}) \downarrow_{\mathbf{s}} + \mathbf{n}. \tag{2}$$

where $\boldsymbol{k}$ denotes a low pass blur filter, $*$ denotes the convolution operation, $\downarrow_s$ indicates a $s\times$ down-sampling (decimating) operation, and $\boldsymbol{n}$ represents the noise. Obviously, the inverse problem of Equation (2)—image SR—is also a typical ill-posed one as there is usually a non-unique solution.

Motion blur degradation is superficially seen to be a simpler problem than resolution degeneration due to there being no down-sampling operation. However, motion blur is most likely to be non-linear or non-uniform, which is usually more complex than resolution degradation (the blur kernel is generally linear and uniform). This makes it a difficult challenge to directly estimate the blur kernel used for non-blind deblurring.

In recent years, deep learning-based networks, especially CNN-based methods, have been the mainstream of image SR and motion deblurring research, such as [3–5]. Although CNN-based image SR and deblurring methods have reported fairly good results, CNN-based image restoration is far from simple when the resolution and motion blur are reduced simultaneously. In this case, either image SR or motion deblurring does not work well due to degraded convolution or the blur kernels not being equivalent, and the two degradation processes are not complementary with each other when they occur simultaneously.

Recently, there have been some CNN works [6–8] that have addressed simultaneous image SR and motion deblurring. All of these methods explicitly or implicitly adopt a global or local feature coupling structure—a deblurring part and an SR part are involved or intervene with each other, to recover the resolution and the motion details at the same time. Actually, these recovery methods only construct different comprehensive CNN mapping networks from the degraded images to the corresponding sharp and high-resolution ones, but do not fully utilize the characteristics of motion deblurring and image SR to achieve decoupling. Therefore, even if the results of these methods are good, they lack an explanation and have low efficiency.

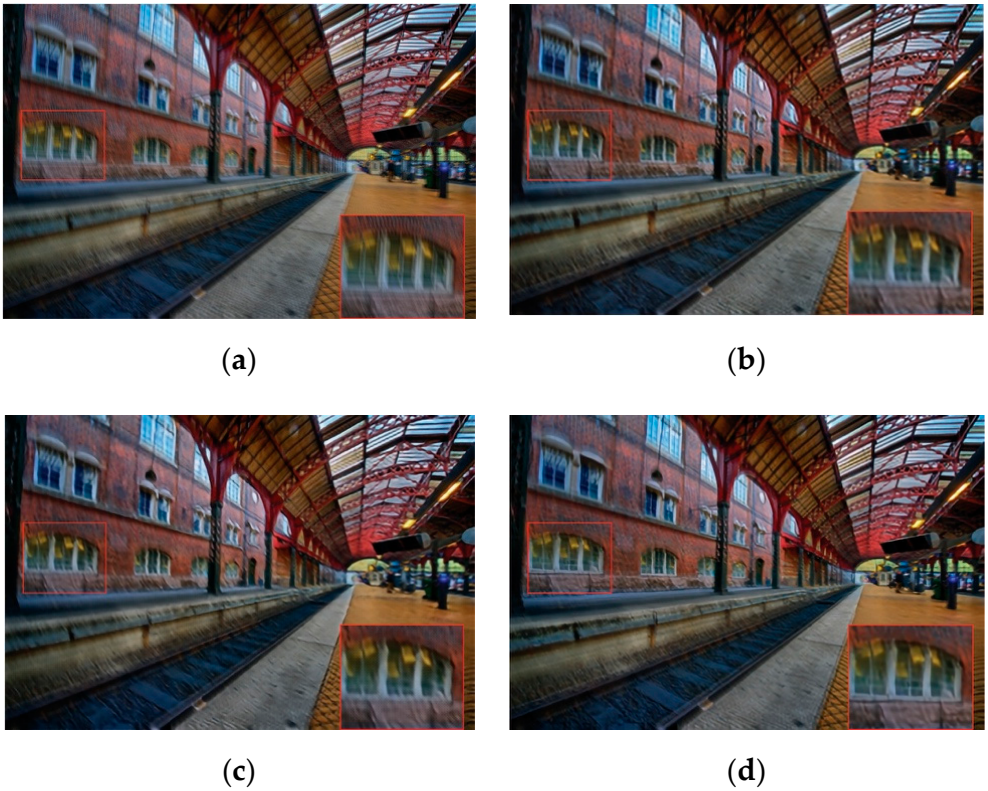
On the other hand, typical deep image recovery models always use the residual connection to convey features. However, due to a lack of ability to mine the feature information across different layers, some complex residual variants are proposed, such as DRRN [9] and RDN [10], etc. Among them, RDN (Residual Dense Network) is representative, which uses not only local dense residual learning but also global residual learning, to extract and adaptively fuse the local and global features from all the observed layers. Since RDN makes full use of multiple hierarchical features, it is very beneficial to construct an image restoration model. In addition to the work on feature learning across different layers, recent attention mechanism-based methods, for example, RCAN [11] and SAN [12], select and enhance useful and important channel feature maps of the same layer through weighting for image or video restoration. In fact, these channel feature attention methods utilize the first- or second-order statistics of the channel maps of certain layers to calculate the dependence of channel features, and then select and weigh the important features. However, single first- or second-order feature statistics cannot make full use of the relationship between different channel feature maps.

In order to overcome the limitations of coupling recovery and single-order attention feature weighting, in this work, we first analyze the compound multiple degradation model of motion blur and resolution reduction and discuss the maximum likelihood (ML) solution of the degradation model. Then, based on the analysis, we discuss decoupling induction

multi-task learning and the CNN model construction method for multiple degradation image restoration. In addition, we obtain the first-order and second-order attention features of the decoupled structures for motion deblurring and SR, respectively, and obtain the third-order attention features by combining local series and parallel features. On this basis, for the sake of improving the feature redundancy and generalization ability of multi-order attention fusion, we utilize the drop-out gating integration method, which enhances the robustness and stability of the proposed multi-order attention mechanism.

An example result of the proposed method to deal with compound degeneration (motion blur as well as 4× down-sampling) is shown in Figure 1, where the comparisons to those results of RCAN [11] and SCGAN [6] are also demonstrated. The dominant contributions of the work are summed up as follows:

- We propose a novel joint motion deblurring and image SR model based on decoupling induction and multi-order attention drop-out gating. The proposed method can overcome the limitation of the single type degeneration assumption to achieve joint recovery with the aid of decoupling induction multi-task learning.
- We propose the use of decoupling dual-branch multi-order attention features for clear HR image reconstruction and select the drop-out gating learning method to enhance the robustness and the generalization of features' fusion.
- We validate and compare the presented model, not only with the publicly available and widely used natural image datasets, but also with synthetic images completely different from the training images. We show that, through decoupling induction and multi-order attention drop-out gating learning, our method can produce visual results of a quality that competes with the most advanced motion deblurring and image SR methods for LR and blurred images.



(**a**) (**b**)

(**c**) (**d**)

Figure 1. An example result of the presented decoupling induction and multi-order attention gating model for joint deblurring and 4× super-resolution. The details in the recovered image of our proposed method (**d**) are much clearer than those of RCAN [11] (**b**) and SCGAN [6] (**c**); the LR and blurred image is shown (**a**).

2. Related Work

2.1. Joint Image Deblur and SR

Compared with the traditional image SR methods, the first CNN-based image SR method, SRCNN [3,4], proposed by Dong et al., can generate more accurate HR details owing to powerful non-linear mapping. To extend three convolutional layers of SRCNN to a deeper level, Kim et al. [13] presented a true deep image SR model called VDSR via residual connection [14]. Recently, Liu et al. [15] also proposed a multi-scale deep encoder–decoder network called MSDEPC to super resolve LR images with the edge maps' prior information. In addition, Ledig et al. [16] proposed the application of a generative adversarial network (GAN) [17] for image SR, called SRGAN. SRGAN takes the perceptual loss and the adversarial loss to supervise the reconstruction of super-resolved images and can obtain more realistic SR results.

CNNs also play an effective role in motion deblurring. Xu et al. [18] and Sun et al. [1] developed some CNN-based methods to recover blurred images based on blur kernel estimation. Besides these non-blind deep methods, some deep blind deblurring methods [2,5] are also proposed. Nah et al. [2] applied a multiple scales CNN to recover clear images directly. Motivated by the work, Tao et al. [5] designed a simple structure motion deblurring network characterized by scale recursion. Moreover, inspired by the work of image translation [19], Ramakrishnan et al. [20] first applied GAN for motion deblurring. Then, Kupyn et al. [21] proposed DeblurGAN for blind motion deblurring, which utilizes the WGAN [22] with a gradient penalty to avoid the mode collapse issue in the classical GAN. Subsequently, Kupyn et al. [23] presented a new and very efficient GAN-based model for single image motion deblurring, named DeblurGAN-v2, which is based on a relativistic conditional GAN with a double-scale discriminator. Furthermore, for meteorological prediction application, Manzo et al. [24] adopted a pretrained deep network-based architecture for clouds' image description and classification. Recently, in order to address the problem that blurred images suffer from other degradation such as down-scaling and compression, Xu et al. [25] proposed the enhanced deep pyramid network (EDPN) model for blurry image restoration, by fully exploiting the self-scale and cross-scale similarities.

Few works can use CNNs for simultaneous motion deblurring and SISR. Xu et al. [6] solve the problem of super-resolving blurred facial images by SCGAN. However, their method is restricted to facial images, and it is not easy to obtain a good performance in real scenarios. Zhang et al. [7] proposed using a deep encoder–decoder model to perform joint motion deblurring and image SR. Zhang et al. [8] once again proposed a gated fusion method for concurrent motion deblurring and image SR. Recently, Liang et al. [26] utilized the dual supervised network to address this issue. However, they did not achieve satisfactory results. In addition, for plug-and-play image SR, Zhang et al. [27] proposed a new blind SR framework to achieve the processing of arbitrary blur kernels. In addition, Zhang et al. [28] proposed a dual supervised learning strategy to fully exploit the representation capacity of their deep model, which imposes constraints between LR and HR images.

2.2. Attention

In addition to feature transfer by residual connection, the attention mechanism is another widely used method for feature preservation and enhancement used in many image SR models [11,12,29,30]. Zhang et al. presented the RCAN [11] (residual channel attention network) model that utilizes channel attention with residual blocks to adjust the task adaptability of channel features and to strengthen their expression ability. Since RCAN only uses the first-order statistical information of channel features, Dai et al. [12] presented the so-called SAN (second-order attention network) model, which replaces the global average pooling with the global covariance pooling (second-order statistics) to obtain a better effect of channel features' selection and enhancement. Very recently, Niu et al. [31] designed a novel pixel-guided dual-branch attention network (PDAN) to jointly restore image details and the spatial scale.

In addition, Wang et al. [29] proposed the extraction and fusion of temporal and spatial attention features for video restoration. Furthermore, Fu et al. [30] introduced a dual attention network—containing one spatial branch and one channel branch for scene segmentation, which can adaptively extract and integrate the local and non-local features of spatial and channel attention.

3. Methodology

3.1. Multiple Degradation Decoupling Induction

For motion deblurring and image SR, we used the following equations to describe the corresponding degradation models, which are used to generate the LR and blur images for training.

$$\mathbf{y} = \left(\sum_{i=1}^{N} \boldsymbol{x_i}\right)/N + \boldsymbol{n} \tag{3}$$

$$\boldsymbol{y} = (\boldsymbol{x} \downarrow_{\boldsymbol{s}}) + \boldsymbol{n}, \tag{4}$$

where Equation (3) represents one typical motion degradation of a certain image sequence-averaging blur and Equation (4) denotes the process of image resolution reduction. Here, $\boldsymbol{x_i}$ and $\boldsymbol{y}$ in Equation (3) represent a sharp image of one clear HR image sequence (the image number of the sequence is N) and the corresponding blur image, respectively; and $\boldsymbol{x}$ and $\boldsymbol{y}$ in Equation (4) are the HR image and the corresponding LR image, respectively. $\boldsymbol{N}$ in the equation denotes the additional noise (normally it is Gaussian white noise). $\downarrow_{\boldsymbol{s}}$ is $\boldsymbol{s}\times$ the down-sampling operator (can be bicubic sub-sampling).

Based on Equations (3) and (4), the motion blur and the resolution reduction compound degeneration may be formulated as

$$\boldsymbol{y} = \left(\left(\sum_{i=1}^{N} x_i\right)/\boldsymbol{N}\right) \downarrow \boldsymbol{s} + \boldsymbol{n} \tag{5}$$

Obviously, averaging $\boldsymbol{N}$ frame images lead to blurring degradation and the subsequent down-sampling operation also reduces the resolution of the generated blur image.

Theoretically, if the frame averaging blur kernel and the spatial down-sampling kernel are denoted as $\boldsymbol{h}$ and $\boldsymbol{k}$, respectively, Equation (5) can be generalized as the following:

$$\boldsymbol{y} = (\boldsymbol{x} * \boldsymbol{h}) * \boldsymbol{k} + \boldsymbol{n}, \boldsymbol{y} = (\boldsymbol{x} * \boldsymbol{S}) + \boldsymbol{n} \tag{6}$$

Here the kernel convolution $\boldsymbol{h} * \boldsymbol{k}$ is defined as a new kernel S. This equation means the comprehensive function of multi-degradation basically equals one blur operation. Moreover, according to Equation (6), the residual $\boldsymbol{r}$ between the sharp HR image $\boldsymbol{x}$ and the degraded observation $\boldsymbol{y}$ is easily calculated. Assuming the image data obey the Gaussian distribution, a solution of maximum likelihood estimation (MLE) for Equation (6) can be obtained by $\widetilde{\boldsymbol{x}} = \boldsymbol{y} + \boldsymbol{r}$. Naturally, if the residual $\boldsymbol{r}$ is looked upon as the high-frequency details of $\boldsymbol{x}$, the observation $\boldsymbol{y}$ becomes its approximation component. Here, if assuming the details $\boldsymbol{r}$ can be decoupled into the deblurring details $\boldsymbol{r_{db}}$ and the SR details $\boldsymbol{r_{sr}}$ that is $\boldsymbol{r} = \boldsymbol{r_{db}} + \boldsymbol{r_{sr}}$, the MLE solution is further expressed as

$$\widetilde{\boldsymbol{x}} = \boldsymbol{y} + \boldsymbol{r_{db}} + \boldsymbol{r_{sr}} \tag{7}$$

According to Equation (7), if we can obtain the deblurring and the SR details individually through deep decoupling induction learning, then the original clear and HR image can be recovered. Moreover, although changing the sequence between motion blur and resolution reduction will lead to the multiple degraded models being different from Equation (5), the MLE solution with decoupling details (Equation (7)) remains the same. This indicates that our proposed decoupling induction method is robust to different sequences of multiple degenerated images.

3.2. Multi-Order Attention Gating

The decoupled deblurring features and SR features were then exploited to calculate the first-order channel attention (FOCA) and the second-order channel attention (SOCA), respectively. Meanwhile, their SOCAs were concatenated to calculate the FOCA again, which acquires the so-called third-order channel attention (TOCA). Then, all the acquired multiple order attentions were fused with multi-routes gating. Closing a route means that the corresponding feature attention is blocked and cannot be used for subsequent reconstruction. In fact, we used the drop-out mechanism—a probability of 0.5 was used to turn off some feature attentions randomly. The above processes are called multi-order attention drop-out gating. We used a similar method to calculate the FOCA and the SOCA, as explored in RCAN [11] and SAN [12]. Based on the principles of the SOCA and FOCA, we give the mathematical description of the third-order channel attention (TOCA) and multi-order attention drop-out gating learning in the following.

Given the deblurring feature maps x_{db} and the SR feature maps x_{sr}, assume they are with C feature channels and size $H \times W$. Note that the channel size of x_{db} and x_{sr} does not need to be equal and in the following we just take x_{db} as an example. We reshape the feature map x_{db} to a matrix X with the size $H \times W$; each element of which is C dimension. Here, if we treat the feature elements as samples, then the covariance matrix may be calculated and decomposed by the eigenvalue decomposition (EIG) as:

$$\sum X\bar{I}X^T = U\Lambda U^T \tag{8}$$

where $\bar{I} = \frac{1}{s}\left(I - \frac{1}{s}1\right)$, $s = H \times W$, and I and $\mathbf{1}$ are the identity matrix and the all-ones matrix, respectively. In addition, U is an orthogonal matrix and $\Lambda = diag(\lambda_1, \ldots, \lambda_C)$ is a diagonal matrix with eigenvalues in decreasing order. Then, the normalized covariance matrix can be acquired as $\hat{Y} = \sum^{\alpha} = U\Lambda^{\alpha}U^T$; α is a positive real number. Obviously, the normalized covariance contains the correlations of channel-wise features. Let $\hat{Y} = y_1, \ldots, y_C$; the c-th channel-wise statistics z_c can be obtained by global pooling $\hat{Y}$ as:

$$z_c = \frac{1}{C}\sum_{i=1}^{C} y_c(i) \tag{9}$$

Based on the equation, the feature weighting coefficient can be obtained through a simple sigmoid gating function [32] as:

$$\omega_c = f(W_U\delta(W_D z_c)) \tag{10}$$

where W_D and W_U are usually the convolution layers to adjust the number of feature channels to C/r and C, respectively. $f(\cdot)$ and $\delta(\cdot)$ are individually the sigmoid functions and RELU function. Thus, for deblurring feature x_{db} the second-order channel attention (SOCA) weighting is represented as:

$$\bar{x}_{db} = \omega_c \cdot x_{db,c} \tag{11}$$

Based on the equation, the SOCA weighting for image SR features x_{sr}, can be similarly described as $\bar{x}_{sr} = \omega_c \cdot x_{sr,c}$. Then, $\bar{x}_{db,c}$ and $\bar{x}_{sr,c}$ are concatenated and passed through the FOCA to obtain the final TOCA. Let $x_{cat} = concat(\bar{x}_{db}, \bar{x}_{sr}) = [x_{cat,1}, \ldots, x_{cat,2C}]$; we calculate the global average pooling along each channel dimension and then transform the statistics with channel scaling convolution layers and proper activation functions to obtain the FOCA weighting, which can be described as:

$$z_{toa,c} = \frac{1}{H \times W} = \sum_{i=1}^{H}\sum_{j=1}^{W} concat(\bar{x}_{db}, \bar{x}_{sr})_c(i,j), \tag{12}$$

$$S_{toa,c} = f(W_S\delta(W_I z_{toa,c})) \tag{13}$$

where $x_{cat,c}(i,j)$ is the value at the position (i,j) of the c-th concatenated SOCA features x_{cat}, and W_S and W_I are the channel up-scaling and down-scaling convolution layers, similar to W_U and W_D in Equation (10). Finally, the third-order channel attention (TOCA) weighting can be denoted as:

$$\hat{x}_{toa,c} = S_{toa,c} \cdot x_{ccat} \tag{14}$$

If the FOCA of the deblurring features and SR features are denoted as $\dot{x}_{db}$ and $\dot{x}_{sr}$, respectively, then all the multi-order attention features, $\dot{x}_{db}$, $\bar{x}_{db}$, $\dot{x}_{sr}$, $\bar{x}_{sr}$, and $\hat{x}_{toa}$, are sent to one five-routes gate for fusion. The gate works with the drop-out mechanism. Let the j-th route switch be a random variable r_j and obey the Bernoulli distribution with the parameter p (which is set to 0.5 in our practice)—that is $r_j \sim Bernoulli(p)$—and then, all the attention that can pass through will be fused by concatenation as:

$$\tilde{x} = concat\left(r_1\dot{x}_{db}, r_2\bar{x}_{db}, r_3\dot{x}_{sr}, r_4\bar{x}_{sr}, r_5\hat{x}_{toa}\right) \tag{15}$$

3.3. Network Architecture

Based on Equation (7), we can design two CNN branches to learn the deblurring details r_{db} and the SR details r_{sr} separately. This step is called decoupling induction learning. Moreover, we can individually calculate their multiple orders attention features, and utilize the drop-out gating method to fuse them. Here the step is named multi-order attention drop-out gating. The fused attention features concatenated with the LR and blur input images are then sent to the subsequent reconstruction module to obtain the final SR result.

The overall architecture of the proposed model is illustrated in Figure 2. Our model contains four dominant modules: the first one is the deblurring features extraction module, which can be used to predict the sharp LR image; the second one is the SR features extraction module, which can be utilized to obtain the super-resolved blur images; the third is the proposed multi-order attention drop-out gating module, which calculates different order attentions and fuses them with the drop-out gating mechanism; and the fourth one is the reconstruction module to recover the final clear and SR result. In the figure, the four modules mentioned are indicated by dashed boxes of different colors.

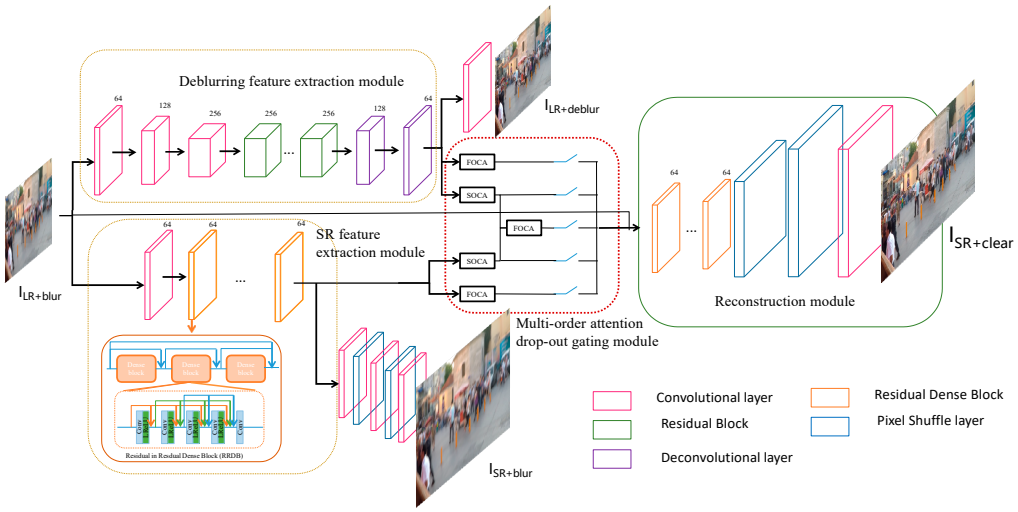


Figure 2. The overall architecture of the proposed model. Our model mainly contains four modules—deblurring feature extraction, SR feature extraction, multi-order attention drop-out gating, and reconstruction. An LR and blur input image is first passed through the separate SR and deblurring branches to obtain the decoupled features; then, they go through a multi-order attention drop-out gating fusion, before being reconstructed to output a super-resolved and clear image.

3.3.1. Deblurring Feature Extraction

This module aims to acquire the decoupled deblurring features, and henceforth, sharp LR images from blurry LR images $\boldsymbol{I_{LR+blur}}$. Inspired by [21], we adopted a residual encoder–decoder structure in this module. The encoder part is composed of several convolution layers which reduce the size of feature maps to a quarter of the input image size. We then added nine residual blocks between the encoder and decoder to refine the deblurring features. Then, the decoder exploits two deconvolutional upscaling layers to raise the resolution of the deblurring feature maps. Additionally, based on the deblurring features, we can use another two convolution operations to obtain a deblur LR image $\boldsymbol{I_{LR+deblur}}$ (see Figure 2).

Here, we denote the output deblurring features of the decoder as $\boldsymbol{x_{db}}$, which were later sent to the multi-order attention gating module. All the used activation layers are the leaky rectified linear units (LeakyReLU), and we used IN (instance normalization) operations in the residual blocks instead of the BN (batch normalization) ones, because the BN layer may reduce the flexibility of the network and undermine the scale information by normalizing the features and increasing computation. The mapping relationship learned from this module between the input $\boldsymbol{I_{LR+blur}}$ and the output $\boldsymbol{x_{db}}$ can be described as:

$$\boldsymbol{x_{db}} = \boldsymbol{H_{\uparrow 2}}\left(\boldsymbol{H_{\uparrow 1}}\left(\boldsymbol{RB}\left(\boldsymbol{H_{\downarrow 2}}\left(\boldsymbol{H_{\downarrow 1}}\left(\boldsymbol{H_c}\left(\boldsymbol{I_{LR+blur}}\right)\right)\right)\right)\right)\right) \tag{16}$$

where $\boldsymbol{H_{\downarrow 2}}$ and $\boldsymbol{H_{\downarrow 1}}$ are the down-scaling convolution layers of an encoder, $\boldsymbol{H_{\uparrow 1}}$ and $\boldsymbol{H_{\downarrow 1}}$ are the deconvolution layers of a decoder, RB represents the nine residual blocks, and H_c is the first convolution layer acting on the input $\boldsymbol{I_{LR+blur}}$. The activation and normalization operations are included in the layers by default.

3.3.2. SR Feature Extraction

The purpose of this module is to obtain decoupled SR image details. We utilized eight residual dense blocks [10] (each block contains five convolution operations with four LeakyReLU layers; see Figure 2 for reference) and one convolution layer to construct the deep structure to extract the high-frequency spatial detail features. From this, the super-resolved blur image $\boldsymbol{I_{LR+blur}}$ can also be acquired through two consecutive pixel shuffle layers and several convolution layers. To maintain the spatial information, neither the pooling layer nor stride operation is used in the module. At the same time, no normalization operations are applied. If denoting the extracted SR features as $\boldsymbol{x_{sr}}$, then the mapping relationship learned from the module between the input $\boldsymbol{I_{LR+blur}}$ and the output $\boldsymbol{x_{sr}}$ can be described as:

$$\boldsymbol{x_{sr}} = \boldsymbol{RDB_8}(\boldsymbol{H_c}(\boldsymbol{I_{LR+blur}})) \tag{17}$$

where $\boldsymbol{RDB_8}$ represents the eight consecutive residual dense blocks.

3.3.3. Multi-Order Attention Drop-Out Gating

This module summarizes the multiple orders attention of the learned deblurring features $\boldsymbol{x_{db}}$ and the SR features $\boldsymbol{x_{sr}}$ to obtain high-frequency image recovering details. $\boldsymbol{x_{db}}$ and $\boldsymbol{x_{sr}}$ are the inputs of the module and their first-order, second-order, and common third-order feature attention maps are calculated, respectively. Then, all these attention features are concatenated and sent to the drop-out layer to obtain the final feature maps $\widetilde{\boldsymbol{x}}$. The mapping relationship of this module and its processing details can be referred to in the previous Section 3.3.2 and Figure 2.

3.3.4. Reconstruction Module

In this module, the gated attention features $\widetilde{\boldsymbol{x}}$ and the blur LR image are sent into 16 residual dense blocks [10] and the result is further fed to two-pixel shuffle layers to improve the spatial resolution to 4×. After that, two convolution layers are used to acquire the final SR and clear image $\boldsymbol{I_{SR+clear}}$. Since most operations of our model are performed in

the LR low dimension functional space, the computation cost both in training and in the testing stages is quite low. The mapping relationship of the module is described as:

$$I_{SR+clear} = H_{2c}(P_2(P_1(RDB_{16}(concat(\widetilde{x}, I_{LR+blur})))))) \quad (18)$$

where RDB_{16} is the 16 consecutive residual dense blocks, P_1 and P_2 are the two-pixel shuffle layers, and H_{2c} represents two convolution operations.

3.4. Loss Functions

Our proposed model has three outputs: the LR deblurring image I_{LR+db}, the SR blur image $I_{SR+blur}$, and the clear SR image $I_{SR+clear}$. Then, the total loss of our model contains three parts: the LR but clear image loss, the HR but blur image loss, and the final HR and clear image loss. In our case, we usually calculate the difference between a certain output and its expectation with the ℓ_1 norm and treat it as the loss. The three losses of our model can be described as:

$$\ell_1 = \sum_{i=1}^{N} \|y_{HR+clear,i} - I_{SR+clear,i}\|_1 \quad (19)$$

$$\ell_2 = \sum_{i=1}^{N} \|y_{LR+clear,i} - I_{LR+db,i}\|_1 \quad (20)$$

$$\ell_3 = \sum_{i=1}^{N} \|y_{HR+blur,i} - I_{SR+blur,i}\|_1 \quad (21)$$

where $y_{HR+clear,i}$, $y_{LR+clear,i}$, and $y_{HR+blur,i}$ are the expectations of the three outputs, respectively. N is the number of training samples. Thus, the total loss is the sum of the above three losses:

$$L = \ell_1 + \alpha\ell_2 + (1 - \alpha)\,\ell_3 \quad (22)$$

where α is the loss balance factor, which is set to be 0.5 in our experiments.

In addition, sometimes in order to generate a more realistic image, we also consider introducing an SSIM [33] measure into the loss ℓ_1. At this time, the loss ℓ_1 can be modified as:

$$\ell_1 = \sum_{i=1}^{N} (\beta SSIM\left(y_{HR+clear,i}, I_{SR+clear,i}\right) + (1 - \beta)\|y_{HR+clear,i} - I_{SR+clear,i}\|_2 \quad (23)$$

where the β is used to balance these two terms, which is set to 0.84.

4. Experimental Results

4.1. Datasets and Training Details

Many experiments and performance comparisons are performed on the well-known public blur datasets: the GOPRO dataset [2] and the dataset developed by Lai et al. [34]. Originated from some natural video sequences, the GOPRO [2] dataset contains 2103 high-resolution training pairs (the sharp image and the blurry image) and 1111 test images. The size of every image in the dataset is 1280 × 720. The motion-blurred image is obtained by averaging several neighboring frame images and the LR image can be acquired by bicubic down-sampling on the corresponding HR image. In contrast to GOPRO, the dataset of Lai et al. [34] is composed of many man-made generated blur images, in which each degenerated image is the convolution result of the sharp image with a blur kernel. Here, the size of the degraded kernel may range from 21 × 21 to 75 × 75. Note that Lai et al.'s dataset [34] contains both uniform and non-uniform blurred images. The main characteristics of the two datasets are summarized in Table 1.

The training of the proposed model can be divided into two steps. In the first step, the model is trained by supervision with the LR blurry patches $I_{LR+blur}$, the sharp LR patches $I_{LR+clear}$, and the clear HR patches $I_{HR+clear}$. During training, the loss of our model (Equation (22)) is minimized. In the second step, the trained model is finetuned by using Equation (23) to replace the original ℓ_1 in Equation (22). The training procedure is implemented by the SGD solver from Pytorch [35] and the learning rate decreases from 0.01 to 0.00001 and the decay is set to be 0.5. In addition, the moment of the used solver

is 0.9 and the batch size of the training samples is 12. It takes about two days to train the proposed model if using an Nvidia Titan GTX1080ti GPU.

Table 1. Basic dataset characteristics of GOPRO [2] and Lai et al. [34].

Dataset	Lai et al. [34]	GOPRO [2]
Synthetic/Real	Synthetic and Real	Real
Blur type	Uniform and Non-uniform	Uniform
Ground-truth images	125	3214
Blurred images	300	3214
Depth variation	Yes	No

4.2. Experiments and Comparisons

Based on numerous LR and blurry input images on different test datasets, we performed lots of joint image deblurring and SR experiments and made comparisons with some recent SOTA (state-of-the-art) image SR models [10–12], the deblurring method [5], and the multiple degradations recovery approaches [6–8,36]. We also compared the combination method of the SR algorithm [10] and the deblurring method [5]. For fair play, all the comparisons were made by using the public codes provided by these methods. For those ones which cannot be publicly acquired (such as ED-DSRN [7]), we used our dataset to retrain the original networks. The comparisons with these related methods using the datasets of GOPRO [2] and Lai et al. [34], in terms of the PSNR, the SSIM, the model parameters, and the test time, are demonstrated in Tables 2 and 3. The visual results of these methods are also compared in Figures 3–5.

(**a**) HR (PSNR/SSIM) (**b**) RCAN (26.01/0.87) (**c**) SCGAN (24.81/0.80) (**d**) GFN (28.86/0.90) (**e**) Ours (28.79/0.912)

(**f**) HR (PSNR/SSIM) (**g**) RCAN (19.695/0.58) (**h**) SCGAN (19.97/0.54) (**i**) GFN (19.2/0.54) (**j**) Ours (20.21/0.59)

Figure 3. The details in the deblurred and super-resolved (4×) images generated by the presented decoupled induction and multi-attention drop-out gating model on GOPRO [2] and Lai et al. [34]; using our method, the image details are clearer than the ones acquired from RCAN [11], SCGAN [6], and GFN [8].

Table 2. The comparisons with SOTA methods of the quantitative performance on GOPRO dataset [2]. Best results are marked in bold.

Measures	RDN [10]	SRN [5]	SCG-AN [6]	RCAN [11]	RDN [10] + SRN [5]	ED-DSRN [7]	Zhang et al. [33]	GFN [8]	Our Proposed
PSNR	24.370	25.829	22.791	25.328	26.211	26.331	25.80	27.81	**27.82**
SSIM	0.739	0.782	0.783	0.804	0.792	0.810	0.768	0.83	**0.848**
Parameters	178 M	28.8 M	15 M	**1.5 M**	305 M	25 M	7 M	11 M	27 M
Training/Inference time	1.0 day/2.8 s	3 days/0.4 s	1.5 days/0.68 s	1.5 day/0.55 s	3.8 days/4 s	1.5 days/0.22 s	2 days/1.3 s	**2 days/0.07 s**	2 day/0.33 s

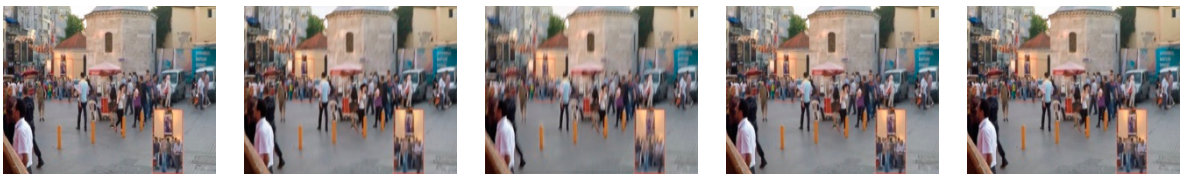
Table 3. The comparisons with SOTA methods of the quantitative performance on Lai et al. dataset [34]. Best results are marked in bold.

Measures	RDN [10]	SRN [5]	SCG-AN [6]	RCAN [11]	RDN [10] + SRN [5]	ED-DSRN [7]	Zhang et al. [33]	GFN [8]	Our Proposed
PSNR	17.780	17.444	18.572	17.729	18.861	18.791	19.003	19.12	**19.17**
SSIM	0.416	0.408	0.460	0.471	0.423	0.473	0.466	0.574	**0.59**
Inference time	2.3 s	0.3 s	0.50 s	0.9 s	2.2 s	**0.20 s**	1.1 s	0.42 s	0.5 s

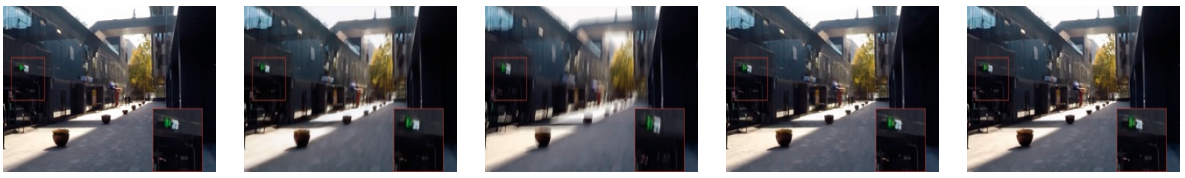
(**a**) HR (PSNR/SSIM) (**b**) RCAN (27.874/0.821) (**c**) SCGAN (24.61/0.727) (**d**) GFN (30.271/0.869) (**e**) Ours (30.316/0.872)

(**f**) HR (PSNR/SSIM) (**g**) RCAN (26.948/0.881) (**h**) SCGAN (23.10/0.786) (**i**) GFN (31.904/0.931) (**j**) Ours (31.886/0.935)

Figure 4. More visual comparison of our model with other methods on GOPRO [2].



(**a**) HR (PSNR/SSIM) (**b**) RCAN (24.555/0.72) (**c**) SCGAN (24.39/0.67) (**d**) GFN (25.297/0.713) (**e**) Ours (25.279/0.731)



(**f**) HR (PSNR/SSIM) (**g**) RCAN (21.446/0.61) (**h**) SCGAN (21.295/0.56) (**i**) GFN (21.358/0.578) (**j**) Ours (21.78/0.61)

Figure 5. More visual comparison of our model with other methods on Lai et al. [34].

According to Tables 2 and 3, it is clear that in most cases our model achieves the best multi-degradation recovery effects, and only in certain special scenarios, it is slightly inferior to GFN [8] (see Figure 3d,e), which seems to be the best joint image SR and the deblurring algorithm available at present. In Figure 3d,e and Figure 4, although the PSNR is slightly lower, the image we recovered looks better than the image generated by GFN [8]. Such contradictions may stem from the fact that the calculation of PSNR or SSIM only requires the neighborhood operations of certain image pixels and cannot reflect the true perception of human vision. In light of the quantitative metrics in Tables 2 and 3, it is easy to see that, compared to the other methods, even under multiple different blurs and LR datasets, the proposed method can achieve the best or the second best PSNR and SSIM performance.

According to Figure 5, we can easily see that on the Lai et al. [34] dataset, our approach shows a significant improvement. Although adjustments have been made to RCAN by fine-tuning the dataset, it still cannot compete with our trained network (see Figure 5b,g). It is clear that Figure 5b,g contains less texture detail than Figure 5e,j. This performance gap is mainly due to the lack of an encoder–decoder structure, which is a key architecture when designing a blind deblurring network. Although the performance of the retrained SCGAN is better than its pre-trained model, because of its small model capacity, this method cannot handle complex non-uniform blurs well.

In general, compared with other methods, especially GFN, the superiority of our method lies in (1) our two branches (super-resolution and motion deblurring), which are fully disentangled, whereas GFN's are not; and (2) we use multi-order attention to obtain the attention features of the two branches at different orders separately, and perform gated fusion through the drop-out mechanism, whereas GFN computes the correlation of different branches for fusion. Due to the simpler structure, the GFN method has fewer parameters and a faster computation speed than our approach. However, in practical applications, assuming no particular requirements for machine memory or computing speed, our method can be used in preference if the scene is rich in significant textures and the objects have multi-scale variations. Benefiting from joint attention learning, our method produces clearer and higher resolution images with good perceptual quality.

5. Ablation Study

For the sake of dissecting the role of the key components of the proposed decoupling induction and multi-order attention gating model, several variants were developed and tested: (1) deblurring alone, (2) SR alone, (3) without TOCA, and (4) no drop-out gating. These variants were trained with almost the same hyper-parameters as our original model. For the variants of deblurring alone and SR alone, the FOCA and SOCA features were concatenated and pushed to the reconstruction module. For the variant without TOCA, there were only four attention routes ($\dot{x}_{db}$, $\overline{x}_{db}$, $\dot{x}_{sr}$, $\overline{x}_{sr}$) for drop-out gating. The final variant used direct concatenating to replace drop-out gating. The results are shown in Table 4.

Table 4. Ablation study on GOPRO [2] dataset. The best results are indicated in bold.

Methods	GOPRO [2]	
	PSNR	SSIM
Deblurring alone	26.97	0.815
SR alone	25.84	0.791
Without TOCA	27.51	0.833
No drop-out gating	27.20	0.827
Ours	**27.82**	**0.848**

From Table 4, it is clear that without drop-out gating, the performance of the proposed approach is much suppressed. At the same time, the high-order attention TOCA really can

help to improve the reconstruction effects. In addition, it seems that the deblurring branch contributes more than the SR forking in multiple degradation decoupling reconstruction. Thus, we can conclude that the proposed mechanism of multi-order attention and drop-out gating is very effective for joint deblurring and super-resolution.

6. Conclusions

In this work, we proposed an effective end-to-end deep model which can deal with multiple degeneration problems for concurrent motion deblurring and image SR. Inspired by the idea of decoupled learning and multi-order attention features selection, our model firstly manages to construct the discrete network structures of motion deblurring and image SR respectively, and then realizes selective features' enhancement and fusion through multi-order attention drop-out gating. Many experimental results and comparisons to other SOTA methods were carried out to demonstrate the superior performance of our method in compound degradation recovery and generalization power.

Future work will focus on two aspects. The first one is to investigate why the deblurring branch matters more than SR forking in the proposed multiple degradation reconstruction approaches. Secondly, based on blur and resolution reduction, if more degeneration action (such as noise interference) is also introduced, a way to obtain a good image recovery effect will be investigated.

Author Contributions: Conceptualization, Y.C. and H.L.; methodology, Y.C. and H.L.; software, Y.C. and X.Z.; writing—original draft preparation, Y.C.; writing—review and editing, H.L.; visualization, X.Z.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant No. 61971004, the Natural Science Foundation of Anhui Province, grant No. 2008085MF190, the Key Project of Natural Science of Anhui Provincial Department of Education, grant No. KJ2019A0083 and KJ2021A1289, and the Open Project Fund of the Key Laboratory of Computational Intelligence and Signal Processing of the Ministry of Education (Anhui University), grant No. 2020A002.

Data Availability Statement: The links to the public datasets used in the paper are as follows: GOPRO dataset: https://github.com/SeungjunNah/DeepDeblur_release (accessed on 1 December 2021), Lai's dataset [34]: http://vllab.ucmerced.edu/wlai24/cvpr16_deblur_study/ (accessed on 1 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, J.; Cao, W.; Xu, Z.; Ponce, J. Learning a convolutional neural network for non-uniform motion blur removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 769–777. [CrossRef]
2. Nah, S.; Kim, T.H.; Lee, K.M. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 21–26 July 2017; pp. 3883–3891. [CrossRef]
3. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 184–199. [CrossRef]
4. Xi, S.; Wei, J.; Zhang, W. Pixel-Guided Dual-Branch Attention Network for Joint Image Deblurring and Super-Resolution. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 532–540. [CrossRef]
5. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8174–8182. [CrossRef]
6. Xu, X.; Sun, D.; Pan, J.; Zhang, Y.; Pfister, H.; Yang, M.-H. Learning to super-resolve blurry face and text images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 21–26 July 2017; pp. 251–260. [CrossRef]
7. Zhang, X.; Wang, F.; Dong, H.; Guo, Y. A deep encoder-decoder networks for joint deblurring and super-resolution. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Alberta, AB, Canada, 15–20 April 2018; pp. 1448–1452. [CrossRef]
8. Zhang, X.; Dong, H.; Hu, Z.; Hu, Z.; Lai, W.-S.; Wang, F.; Yang, M.-H. Gated fusion network for joint image deblurring and super-resolution. In *British Machine Vision Conference (BMVC)*; Springer: London, UK, 2018. [CrossRef]

9. Ying, T.; Jian, Y.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 21–26 July 2017; pp. 3147–3155. [CrossRef]
10. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481. [CrossRef]
11. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 August 2018; pp. 286–301. [CrossRef]
12. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.-T.; Zhang, L. Second-order attention network for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, NY, USA, 15–20 June 2019; pp. 11065–11074. [CrossRef]
13. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654. [CrossRef]
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
15. Liu, H.; Fu, Z.; Han, J.; Shao, L.; Hou, S.; Chu, Y. Single image super resolution using multi-scale deep encoder-decoder with phase congruency edge map guidance. *Inf. Sci.* **2019**, *473*, 44–58. [CrossRef]
16. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690. [CrossRef]
17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Farley, W.D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
18. Xu, L.; Ren, J.S.; Liu, C.; Jia, J. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*; Cornel University: Ithaca, NY, USA, 2014; pp. 1790–1798, Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.709.7888&rep=rep1&type=pdf (accessed on 1 November 2021).
19. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 21–26 July 2017; pp. 1125–1134. [CrossRef]
20. Ramakrishnan, S.; Pachori, S.; Gangopadhyay, A.; Raman, S. Deep generative filter for motion deblurring. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 2993–3000. [CrossRef]
21. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8183–8192. [CrossRef]
22. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Gan. *arXiv* **2017**. Available online: https://arxiv.org/abs/1701.07875 (accessed on 1 November 2021).
23. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 8878–8887. [CrossRef]
24. Manzo, M.; Pellino, S. Voting in transfer learning system for ground-based cloud classification. In *Machine Learning and Knowledge Extraction*; Cornel University: Ithaca, NY, USA, 2021; Volume 3, pp. 542–553. [CrossRef]
25. Xu, R.; Xiao, Z.; Huang, J.; Zhang, Y.; Xiong, Z. EDPN: Enhanced deep pyramid network for blurry image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 414–423. [CrossRef]
26. Liang, Z.; Zhang, D.; Shao, J. Jointly solving deblurring and super-resolution problems with dual supervised network. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 790–795. [CrossRef]
27. Zhang, K.; Zuo, W.; Zhang, L. Deep plug-and-play super-resolution for arbitrary blur kernels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1671–1681. [CrossRef]
28. Zhang, D.; Liang, Z.; Shao, J. Joint image deblurring and super-resolution with attention dual supervised network. *Neurocomputing* **2020**, *412*, 187–196. [CrossRef]
29. Wang, X.; Chan, K.C.; Yu, K.; Dong, C.; Loy, C.C. Edvr: Video restoration with enhanced deformable convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1954–1963.
30. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 18–20 June 2019; pp. 3146–3154. [CrossRef]
31. Niu, W.; Zhang, K.; Luo, W.; Zhong, Y. Blind motion deblurring super-resolution: When dynamic spatio-temporal learning meets static image understanding. *IEEE Trans. Image Process.* **2021**, *30*, 7101–7111. [CrossRef] [PubMed]
32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

33. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
34. Lai, W.-S.; Huang, J.-B.; Hu, Z.; Ahuja, N.; Yang, M.-H. A comparative study for single image blind deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1701–1709. [CrossRef]
35. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8024–8035.
36. Zhang, K.; Zuo, W.; Zhang, L. Learning a single convolutional super resolution network for multiple degradations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3262–3271. [CrossRef]