

Article

A Global-Local Blur Disentangling Network for Dynamic Scene Deblurring

Xiaoguang Li ^{1,2,*} , Feifan Yang ^{1,2}, Jianguo Huang ^{1,2} and Li Zhuo ^{1,2}

¹ Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China; yff@emails.bjut.edu.cn (F.Y.); huangjl1997@emails.bjut.edu.cn (J.H.); zhuoli@bjut.edu.cn (L.Z.)

² Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

* Correspondence: lxg@bjut.edu.cn; Tel.: +86-10-6739-2420

Abstract: Images captured in a real scene usually suffer from complex non-uniform degradation, which includes both global and local blurs. It is difficult to handle the complex blur variances by a unified processing model. We propose a global-local blur disentangling network, which can effectively extract global and local blur features via two branches. A phased training scheme is designed to disentangle the global and local blur features, that is the branches are trained with task-specific datasets, respectively. A branch attention mechanism is introduced to dynamically fuse global and local features. Complex blurry images are used to train the attention module and the reconstruction module. The visualized feature maps of different branches indicated that our dual-branch network can decouple the global and local blur features efficiently. Experimental results show that the proposed dual-branch blur disentangling network can improve both the subjective and objective deblurring effects for real captured images.

Keywords: dual-branch; disentangling network; branch attention



Citation: Li, X.; Yang, F.; Huang, J.; Zhuo, L. A Global-Local Blur Disentangling Network for Dynamic Scene Deblurring. *Appl. Sci.* **2021**, *11*, 2174. <https://doi.org/10.3390/app11052174>

Academic Editor: Francesco Bianconi

Received: 20 January 2021

Accepted: 24 February 2021

Published: 2 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Limited on the performance of capture devices and environmental conditions, images captured in a real dynamic scene are often blurry. Blur is one of the main degrading factors in the captured images. Blurry image not only affects subjective perception, but also affects the performance of the subsequent intelligent analysis. Insufficient camera resolution, long shooting distance, camera shake, and other factors may result in global image blur, while target motion, scene depth changes, and out-of-focus issues will lead to local blur. Different types of blur are randomly and complicatedly coupled. Therefore, the restoration of non-uniform blurry images in real dynamic scenes is a very challenging issue in low-level computer vision.

Due to the ill-posed essence of the image restoration problem, the conventional blind restoration methods [1–7] usually make assumptions about the blur kernel and then use the prior of natural images to restore the clear images. Most conventional methods [8–12] mainly focus on solving the motion blur caused by simple target movement, camera translation, rotation, and other factors. It is still difficult to handle the blurred images in real dynamic scenes. Early learning-based blind image restoration methods [13–16] used convolutional neural networks (CNNs) to estimate unknown blur kernels, such methods are difficult to deal with blur in a dynamic scene. Existing methods [17–21] adopt end-to-end deep networks to directly learn the mapping relationship between the blurry and clear images. This kind of method takes less consideration on the characteristics of the blurry images, which often leads to over smoothing.

According to whether the blur kernel is spatially varying, the blurry image is divided into the global blur and local blur [16], or called uniform blur and non-uniform blur. We have observed that real blurry images are often formed by the coupling of these two types

of blur. The background region usually appears global blur, or the change of the blur kernel is relatively gentle, while the foreground moving targets are local blur. The global blur appears on the whole image, while the local blur kernels may vary with respect to spatial position [22]. Accurately predicting the non-uniform blur area is of great significance to improve the quality of the blind restored images.

Only a few methods [16,23] classified blur regions according to the varying of blur. Liu et al. [23] proposed a Bayesian classifier to classify image patches into blurry and non-blurry patches. Yan et al. [16] proposed a Fourier transform-based blur kernel estimation prior. This method adopted the frequency domain characteristics and a deep neural network to solve the patch-based image deblurring. Although these two methods [16,23] can detect the local blur of real images, these methods only extract hand-craft features. Due to the diversity of dynamic scene blur, the performances of these methods are still limited. Reference [24] proposed a multi-branch network that can perceive human movement and achieved good results. However, because this method only considers the movement of people in the foreground, the applicability is still limited.

We designed a dual-branch network to resolve the problem of non-uniform blur image restoration. A phased training strategy is designed to disentangle the global and local blur features, and an attention mechanism is used to fuse the two kinds of features, thereby improving the quality of blind non-uniform blurry image restoration. Figure 1 shows the visualization of feature maps extracted via the global and local blur feature extraction branches, respectively. We can observe that the dual-branch network can capture the global and local blur features efficiently. Different from other multi-branch networks, in our model, two branches are designed to the same architecture. We show that the disentangling function can be achieved via a merely data-driven manner by a phased training strategy.



Figure 1. The feature map of disentangle branches for local and global blur feature maps. From left to right are the blurry image, local blur feature map, and global blur feature map.

The main contributions of this paper are as follows.

- A global-local blur feature disentangling network is proposed. The network adopts a parallel dual-branch architecture to decouple the global and local blur features. A branch attention mechanism is designed to dynamically fuse dual-branch features.
- A phased training strategy with task-specific datasets was introduced to train different branches and the branch attention model. We train the branch modules to extract the global and local blur features at first; then, we fix the branch parameters and train the attention module to adaptively restore the complex blurry images.
- A comprehensive comparison was conducted. With the guidance of the global and local blur feature disentangling branches and the branch attention mechanism, the network can reconstruct sharp edges and clear structures. The feature maps of the dual-branch network show that our method can successfully disentangle the global and local blur features. The restoration results show that the method proposed in this paper has achieved state-of-the-art subjective and objective performance compared with the existing dynamic scene deblurring methods.

2. Related Work

2.1. Image Deblurring for Dynamic Scene

Complex non-uniform blur results from camera shaking, object motion, varying in depth, and defocus, which make it a daunting task in computer vision. Conventional deblurring methods make efforts to estimate the blur kernel corresponding to each pixel, which is a serious ill-conditioned problem. They usually make some assumptions about

blur kernels. Methods [2,6,25] used a simple parameter prior model to quickly estimate local linear blur kernel. In Reference [7,26], different parameter prior models are employed to estimate blur kernel and restore images iteratively. However, most conventional methods [8–12] mainly focuses on solving the motion blur caused by simple target motion, camera translation, rotation, and other factors, while the blurry image of a real dynamic scene suffers from complex and non-uniform blurry degradation. Therefore, conventional methods are difficult to effectively solve the problem of non-uniform blur in real scenes. They often involve iteration, which results in time-consuming and limited performance.

Early learning-based methods [13–16] mainly used CNN to estimate unknown blur kernels to improve the accuracy in blind restoration, and then conventional deconvolution methods are employed to restore the blurry images. Sun et al. [13] and Yan et al. [16] parameterized and estimated the blur kernel through classification and regression analysis. These methods [13–16] improve the conventional deblurring framework with CNN kernel estimation. The quality of image restoration depends on the accuracy of estimated blur kernels.

Recently, some end-to-end deep learning-based deblurring methods [17–21] have emerged, inspired by research work, such as image transfer-based on Generative Adversarial Network (GAN) [27]. Kupyn et al. [20] regarded deblurring as a special case of image style transfer, that is, CNN is used to model the mappings from the blurry to clean image. GAN is used to generate images that are close to the real clear images. Nah et al. [18] designed a multi-scale network to extract the multi-scale information of the image in an iterative manner, and gradually restore the clear image. Tao et al. [28] proposed a scale recursive network with shared parameters. The experimental results show that these methods have achieved good results in both of the subjective objective quality compared with the conventional methods.

However, most of the above-mentioned methods pay less attention to the characteristics of the blurry images. We believe that an adaptive mechanism is necessary to handle the non-uniform spatial varying blur kernels. In order to adaptively restore different kinds of blur in the network, this paper disentangles the real complex blur image into near uniform global blur and non-uniform local blur. A dual-branch network with attention mechanism is introduced to disentangle the global-local blur features and reconstruct the real blurring image adaptively.

2.2. Multi-Branch Network in Image Restoration Task

The multi-branch network architecture has been widely used in many deep-learning-based algorithms, while there have been still a few attempts for image restoration. Different branches are usually designed as different architectures for specific tasks. Li et al. [29] proposed a deep guided network for image deblurring tasks, which includes an image deblurring branch and a scene depth feature extraction branch. The image deblurring branch is guided by the scene depth feature extraction branch to restore a clear image. The image restoration task usually contains two parts of information, namely image structure and details. Combining these features, Pan et al. [30] proposed a parallel convolutional neural network for image restoration tasks. The network includes two parallel branches to jointly estimate the image structure and detail information, and restore them in an end-to-end manner. Therefore, combining certain characteristics of the image itself can help to improve the quality of restoration.

From the view of signal processing, the global uniform blur is a linear shift invariant processing, while, in the local non-uniform blur, the blur kernel will vary with respect to spatial position. Therefore, different network mechanisms should be considered to deal with the two types of blur separately.

In this paper, a complex blur image is modeled from a novel perspective, and a disentangling network of the complex blur image is established from the perspective of the global uniform blur of the image background and the local blur of the foreground. Different network branches trained with task-specific datasets to disentangle the global and

local blur features and adaptively restore a clear image. Different from other multi-branch networks, in our model, two branches are designed to the same architecture. We show that the disentangling function can be achieved via a merely data-driven manner by a phased training strategy.

2.3. Attention Mechanism in the Image Restoration Task

The visual attention mechanism can detect the target in the image and capture the features of the region of interest quickly. Woo et al. [31] proposed a CBAM (Convolutional Block Attention Module) model to sequentially implement channel and spatial attention to extract features. It is widely applied to visual recognition and classification tasks. At present, it is still an active research topic in image restoration tasks. Reference [32,33] adopted the attention mechanism to rain removal and multi-degradation factor image restoration tasks, respectively. Qian et al. [32] employed the visual attention mechanism to the rain removal task to guide the network to focus on the raindrop. Sukanuma et al. [33] used channel attention to the restoration of various type of degraded images and improved the robustness of the algorithm by selecting different filters for different types of degrading factors, such as raindrops, blur, compression distortion, and noise. We use the attention mechanism to perceive non-uniform motion blur features, and have also achieved promising results [34]. Therefore, the attention mechanism has the ability to dynamically perceive blur features and improve the effect of deblurring tasks. Purohit et al. [35] proposed a self-attention module to handle varying spatially variant blur. Chen et al. [36] extended the CBAM model to adaptive learn the arrangement of the channel and spatial attention sub-modules in sequentially or in parallel.

Different from the channel and spatial attention model as CBAM, we introduced a branch attention to adaptively fused the global and local blur features and guide the network to generate a restored image with clear structures. In Reference [24], labeled datasets were adopted to constrain the network to focus on the movement of humans in the foreground. Different from Reference [24], we introduced a phased training strategy with task-specific datasets to train different branches and the branch attention model.

3. Proposed Method

We propose a dual-branch architecture-based global-local blurring feature disentangling network. The network contains two branches, a global blur feature extraction branch, and a local blur feature extraction branch. Firstly, the global blur feature branch and the local blur feature branch extract the global and local blur feature, respectively. Then, a branch attention module is used to dynamically fuse the global and local features to obtain the attention mask and apply on the local blur feature branch, which helps the local blur branch to capture the blur variance. Finally, the weighted feature of the local blur branch is combined with the global blur branch feature to jointly guide the generation of a restored image with sharp edges.

3.1. Network Architecture

The framework of the proposed global-local blur disentangling network is shown in Figure 2. The network is composed of 4 modules, a global blur feature extraction module, a local blur feature extraction module, a branch attention module, and a reconstruction module.

The network takes the blurry image as input. First, the two-branch network extracts the global and local blur features of the input, respectively. Then, the branch attention module dynamically fuses the dual branch features to obtain the attention mask and element-wise multiply the local branch features, so that the local branch module focuses on the local features. Finally, under the combined effect of the updated weighted local blur feature and the global blur feature, the reconstruction network is used to restore a clear image. The four modules are introduced in detail as follows.

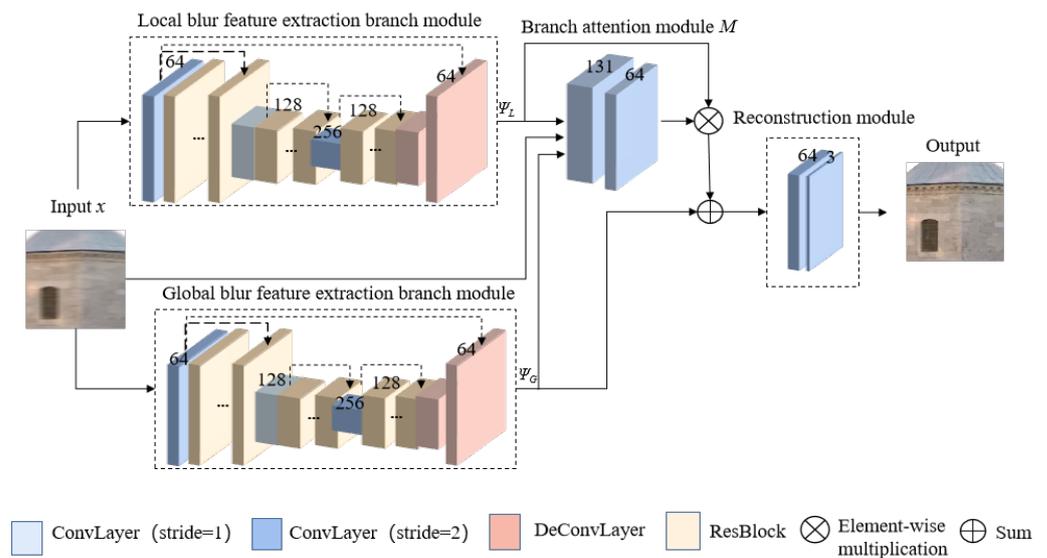


Figure 2. Dual-branch decoupling network architecture.

Local blur feature extraction branch module. The top branch of the framework is designed to extract the local blur features of the input image x . The architecture of this blur branch refers to Reference [37]. The module adopts an encoder-decoder architecture with multiple long and short skip connections. The encoder contains 3 scales, and the input and output of each scale are connected across layers through long-skipping connections. Each scale is composed of 6 residual modules, and each residual module is composed of two convolution layers with a kernel of 3×3 and a stride of 1. To enhance the fusion between low-level feature features and high-level features, short-skipping connections and long-skipping connections are adopted to fuse the feature maps of different layers. The decoder consists of two transposed convolutions and three convolutional layers. Each transposed convolution enlarges the spatial scale of the feature map with 2 times. Finally, we use 3 convolutional layers to reconstruct the restored image. We use the output Ψ_L of the first convolution of the decoder as the output of the local blur feature branch, and then as the input of the branch attention module.

Global blur feature extraction branch module. The architecture of the bottom branch is designed same as the top branch. To constrain the module to extract the global blur features of the input, we use global uniform blur data to train the branch and fix the weight of the branch after the network is converge to ensure that the branch pays more attention to global blur features. The specific training dataset and training strategy will be introduced in detail in Section 4.

Branch attention module. We do not simply adopt concatenation or multiplication on the output feature maps of the two branches. Instead, we designed a branch attention module to dynamically fuse the global and local blur features and guide the processing of restoration. As shown in Figure 2, the branch attention module is composed of two operations, namely element-wise multiplication and addition. Local blur is non-uniform in different spatial positions. Therefore, we use the concatenation of the three features as the input of the attention module, so that the attention module can obtain attention map via different feature maps, which includes the blurry image x , the output of the local blur branch Ψ_L , and the output of the global blur branch Ψ_G . There are two convolutional layers $M(\cdot)$ is adopted to extract the local weight mask feature map. It multiplies the corresponding elements with the output Ψ_L of the local blur branch to obtain the local blur features. The purpose of this step is to improve the ability to extract local blur features. Then, add the weighted local blur feature and the output Ψ_G of the global blur feature branch to obtain the output Ψ_B of the branch attention module. The process is shown in (1):

$$\Psi_B = \Psi_L \odot M(x, \Psi_L, \Psi_G) + \Psi_G, \tag{1}$$

where \odot represents element-wise multiplication. $M(\cdot)$ is the two convolutional layers for extracting the feature map of the local weight mask, the input is the concatenate feature of the three features of the blurry image x , the output of the local blur branch Ψ_L and the output of the global blur branch Ψ_G .

Different from the conventional CBAM, our branch attention is a simple element-wise attention model.

Reconstruction module. Our reconstruction module consists of 2 convolutions. We use Ψ_B , the output of the branch attention module, as the input of the reconstruction module. We use two convolutions with a kernel of 3×3 to reconstruct a restored image with the same size as the input image.

3.2. Phased Training and Loss Functions

As shown in Figure 2, the proposed framework contains two parallel branches, which estimate global and local blur feature from the input, respectively. To constrain the two branches to extract local and global blur feature, we use a branch loss function and the task-specific training data to constrain the branch network. There are 2 task-specific training sets involved in our phased training scheme, a global blur dataset and a local blur dataset, namely the GoPro dataset [18]. The details of these two datasets are described in Section 4.1. The training process is divided into 3 phases.

Phase 1: Train the global blur branch with the global blur dataset and a global loss function.

Phase 2: Train the local blur branch with the GoPro [18] training dataset and a local loss function.

Phase 3: Fix the network parameters of the global blur feature extraction branch, ensure that the global blur branch extracts the global blur feature, and then train the overall network. Therefore, the loss function of the network at this stage computed with the GoPro dataset.

We train the model by minimizing the dual-branch loss and an image content loss. The branch loss functions are shown in follows:

$$L_G(x) = \|\Psi_G(x) - G_{GT}\|_2^2, \quad (2)$$

$$L_L(x) = \|\Psi_L(x) - L_{GT}\|_2^2, \quad (3)$$

where x represents the input blurry image, and G_{GT} and L_{GT} represent the ground truth corresponding to the global and local blur branch outputs, respectively. $\Psi_G(\cdot)$ and $\Psi_L(\cdot)$ represent global and local blur branch networks, respectively. $L_G(\cdot)$ and $L_L(\cdot)$ represent global and local blur branch loss functions.

The content loss function is used to calculate the mean square error loss (MSE) between the output deblurred image and the corresponding clear image, that is, the L_2 loss:

$$L_\phi(x) = \|\phi(x) - X_{GT}\|_2^2, \quad (4)$$

where x and X_{GT} represent the input blurry image and its corresponding clear image, respectively. $\phi(\cdot)$ represents the overall network of the proposed method, which contains the feature information of the dynamic fusion of the two branches.

4. Experimental Results and Discussion

To evaluate the performance of our proposed method, we conducted intensive experiments in this section. We will introduce the dataset and experimental settings at first. To verify the effectiveness of the disentangling network and the branch attention, we conducted comprehensive ablation experiments. In addition, we compared the proposed method with the stat-of-the-art dynamic scene deblurring methods [18,20,21,34,38]. Experimental results and discussions are provided in this section.

4.1. Dataset and Experimental Settings

As demonstrated in Section 3.2, the training process includes three phases. Different training sets are employed in different training phases. First, to train the global blur feature extraction branch, we adopted a synthetic global blur dataset. Then, the GoPro dataset [18] is used to train the local blur feature extraction branch. Finally, the GoPro dataset is used to train the overall network. The following are the datasets and parameter settings used in training.

Global blur dataset. We have built a global blur dataset via blur convolution. The blur kernels in Reference [39] are adopted to generate the blur images, which includes 32 motion, 16 Gaussian, and 8 defocus blur kernels. The high-quality images are widely used Berkeley segmentation dataset (BSD68) dataset [40,41]. The global blur dataset contains 3808 blurry and clear image pairs, in which 2536 pairs are used for training and 1272 pairs are used for testing.

GoPro dataset. (GoPro: https://github.com/SeungjunNah/DeepDeblur_release accessed on 8 November 2017). The GoPro dataset [18] includes 3214 pairs of blur and clear images, covering a variety of scenes, and simulating non-uniform blur in dynamic scene. Instead of modeling a kernel to convolve on a sharp image, the blurry images are generated by recording the sharp information with a high-speed camera and integrating frames over time in the GoPro dataset. It is a realistic ground-truth blur dataset. We use the GoPro dataset to train the local blur extraction branch and the overall network. As the same as the settings in Reference [18], it is divided into 2103 training sets and the remaining 1111 images work as the testing set.

The data argumentation strategies in the training process include random flipping and rotation. The image blocks are as the input of the network, among which 120,000 image blocks are cropped out from the GoPro training set, and 20,000 image blocks are cropped out from the global blur training set.

Experimental settings. In the three network training phases, we adopted the ADAM optimizer [42] network training with default parameters. We adopted NVIDIA GeForce GTX 1080 Ti GPU for model training and testing, and the PyTorch is used to build our network framework.

4.2. Ablation Experiments

To verify the effectiveness of our proposed dual-branch module and branch attention module, we conducted 3 ablation experiments. First, the local blur feature extraction branch network is used as the baseline, which is referred to Local-Branch-Net (LB-Net). Then, a global blur feature branch network is added to the baseline, which is referred to Dual-Branched-Net (DB-Net). Finally, the branch attention network is added to the DB-Net, and the method we propose is called Dual-Branch Attention Fusion Net (DBAF-Net). We use the direct addition method to fuse the features extracted by the two parallel branches of DB-Net. Table 1 shows the average Peak Signal to Noise Ratio (PSNR) results of the 3 models on the GoPro dataset.

Effectiveness of dual-branch modules. Comparing the objective experimental results of the LB-Net and the DB-Net, we can see that the PSNR of the LB-Net is increased from 31.07 dB to 31.77 dB, a gain of 0.7 dB is achieved. It shows that the global and local blur feature extracted by the dual-branch network complemented each other to enhance restoration results. However, the simply addition of the two features in the baseline does not take full advantage of the dual-branch features.

Effectiveness of the branch attention module. Comparing the objective experimental results of the DB-Net and the DBAF-Net, we can see that the PSNR of the DB-Net has increased from 31.77 dB to 32.27 dB, namely 0.5 dB's gain. It shows that the branch attention module can effectively integrate dual-branch features.

Table 1. Average Peak Signal to Noise Ratio (PSNR) Results of ablation experiments.

Model	LB-Net	DB-Net	DBAF-Net
Local Branch	✓	✓	✓
Global Branch		✓	✓
Branch Attention			✓
PSNR (dB)	31.07	31.77	32.27

The subjective results of the ablation experiments are shown in Figure 3. From the zoom in regions, we can see that the restoration results of the DBAF-Net contain richer details. Aiming at the global blur text characteristics in the background, clear and recognizable characters are restored on the wall. For the blur motion in the foreground, the human arm can also get clear edge features.

Effectiveness of the disentangling network. The dual-branch architecture and the phased training scheme were designed to implement the global-local blur disentangling. Note that the LB-Net was end-to-end trained using the GoPro dataset. Therefore, it is a baseline framework without any disentangling design. While the DBAF-Net is the fully disentangling network. From Table 1, we can see that the DBAF-Net achieved 1.2 dB's gain. Generally, this gain is visually perceptible. We can see that our proposed global-local blur disentangling network, which was implemented via our dual-branch architecture and phased training scheme, is effective for dynamic scene image deblurring.



Figure 3. Subjective results of ablation experiments. From left to right are the blurry image, restoration results of the Local-Branch-Net (LB-Net), Dual-Branched-Net (DB-Net), Dual-Branch Attention Fusion Net (DBAF-Net), and the ground truth.

4.3. Comparisons with State-of-the-Art Deblurring Methods

To measure the effective of the proposed method, we compared our method with 9 other latest dynamic scene image deblurring methods in this subsection. They are method proposed by Nah et al. [18], DeblurGAN [20], DeblurGAN-v2 [21], BAG [34], and the method proposed by Gao et al. [38]. The method of Nah et al. [18] uses an end-to-end network to restore images and achieves a good deblurring effect. DeblurGAN [20] applies generative adversarial networks to image deblurring tasks, and this method can restore image details better. DeblurGAN-v2 [21] can recover rich edges and contours. The BAG [34] and Gao's method [38] are the latest dynamic scene deblurring algorithms, which achieve good subjective and objective results on the GoPro dataset. Jiang's method [43] generalizes better to handling real-world motion blur. Yuan's method [44] and Lei's method [36] can better restore blurry images in dynamic scenes. Shen's method [24] restored the blurry images with more semantic details. All objective results are reported in the papers.

Subjective evaluation. Figure 4 shows the subjective results of some of the compared methods. We can see that our method can restore the blurry image in high dynamic scene, and obtain the restored image with clear edges. It can be seen from the first image that the result of Gao's method [38] is blur in the zoom in regions. The characters on the wall cannot be recognized due to global blur. The result of DeblurGAN-v2 [21] is likely to cause averaging of restored images for local motion blur. Our method can better restore the dynamic blur of the sleeves and the uniform blur of the background text, which

indicated that our method can handle both global and local blur. We have released the representative results of the proposed method in Figure 4 through the following website: <https://doi.org/10.6084/m9.figshare.14068721.v1> accessed on 20 February 2021.



Figure 4. Subjective results of the GoPro test images. From top to bottom are the blurry images, the results of DeblurGAN (GAN = Generative Adversarial Network) [20], DeblurGAN-v2 [21], BAG [34], Gao's method [38], our method, and the ground truth.

Note that the ground truth images are taken by a high-speed global shutter camera as shown in the last row in Figure 4. According to the sampling theorem, most motion blur results in the shutter frequency is lower than 2 times the object motion. The proposed technique is used to relieve the frequency confusion effects.

Objective evaluation. The objective experimental results of different methods are listed in Table 2. From the results we can see that our proposed method achieved 32.27 dB, which is a significant improvement in the objective results. The SSIM is comparable to other methods.

Computation complexity. To evaluate the computational complexity of different methods, we have tested several compared methods with author-released source codes. The running time results in seconds for a 720×1280 image are listed in Table 2. We can see that the running time of our method is 0.72 s. It is the fastest method.

The basic operation unit of our proposed network is a 3×3 convolution on pixels. Therefore, the computation complexity is approximately linearly correlated to the size of input images. For example, for a 2 K input image (2048×1080), the running time is 2.1 s.

For 4 K or 8 K images, more memory is required for the GPU platform. Maybe it can be processed as different sub-images.

Table 2. Average PSNRs, SSIMs, and running time on the GoPro dataset.

Method	PSNR (dB)	SSIM	Time (s)
Nah et al. [18]	28.3	0.92	3.09
DeblurGAN [20]	27.2	0.95	0.97
DeblurGAN-v2 [21]	29.6	0.93	7.15
BAG [34]	29.4	0.89	1.13
Jiang et al. [43]	29.67	0.93	–
Yuan et al. [44]	29.81	0.94	–
Gao et al. [38]	30.92	0.94	1.60
Shen et al. [24]	30.26	0.94	–
Lei et al. [36]	31.42	0.94	–
Ours	32.27	0.92	0.72

4.4. Discussion

An extensive discussion of our proposed global-local blur disentangling network for dynamic scene deblurring is made in this section to provide further insights into the potential of further work.

- **Dynamic scene deblurring and its challenge**
In dynamic scene images, the complex blurs caused by various sources, such as object motion, camera shake, and scene depth variation. Different types of blurs are randomly coupled together with different parameters. For camera shake, it results from global blur. The blur kernel may be uniform everywhere. For object motion and scene depth variation, the blur kernel is local, which varies with respect to spatial position. Therefore, variation is the major challenge in dynamic scene image deblurring. We proposed to handle the complex variation of blur kernels with disentangling analysis. The different types of blur are roughly divided into a global and a local blur components. It is reasonable to handle the global and local blurs with different deblurring schemes. Intuitively, the idea of disentangling analysis provides adaptive mechanism to handle the variation of blur kernels.
- **Disentangling blur analysis and its implementation**
The motivation of our proposed method is to disentangle the blur to global and local components. The way to implement the disentangling operation depends on

two points, viz. the dual-branch architecture and the phased training scheme. The dual-branch network provides a framework to disentangle the two types of blur components. The phased training scheme forces the different branches to extract the global and local blur features, respectively. The advantages of our dual-branch network include (1) avoiding error accumulation and (2) partially interpretable, while the disadvantage is that it cannot be trained in an end-to-end manner.

- **Potential alternative implementation**
There are alternative approaches, such as the attention mechanism, to implement the idea of blur disentangling analysis. Using attention mechanism to disentangle blur features will be an interesting exploration. The attention-based blur disentangling network will depend on data correlation. Therefore, there is potential to discover more interesting disentangled blur factors. The drawback of the attention-based network is more parameters are involved, which leading to more training data required.
- **Limitations on performance**
The proposed dual-branch network benefits from the realistic ground-truth blur dataset, GoPro [18]. Compared to kernel convolution, integrating over time to the high frame rate videos provides more realistic dynamic blurry images. It enables efficient supervised deep learning and rigorous evaluation. However, there are two limitations that affect the performance of deblurring for real-world images. One is the inaccurate camera response function (CRF). As mentioned in Reference [18], there is no efficient CRF estimation algorithm available, the gamma correction with $\gamma = 2.2$ is used to estimate the CRF. Another one is domain mismatching. If the input blurry images are far away from that of the training set, the performance may decline. Domain adaptation methods may be helpful in such case.
- **Other potential applications and future work**
Idea of global-local blur disentangling analysis and its implementation, dual-branch network with phased training scheme, can be directly extended to video deblurring applications. For every single blurry frame in a video, the restoration task usually can be formatted as a deblurring issue for dynamic scene images. Note that the sharp images in the GoPro dataset are taken by a high-speed camera and selected from video frames. In addition, in video deblurring applications, there are more temporal priors that can be further explored to enhance the temple consistency of the restored frames. Towards real-world low-quality image restoration tasks, there are several other degradation factors that should be considered, such as noise, very high or low illumination, compressed artifacts, and so on. It is an interesting topic to solve the image restoration issues with the coexisting of different degradation factors via the idea of disentangling analysis. However, the degradation models for different factors are quilt different. For example, noise in real-world can be extra complex. To further explore this topic, we should pay more attention to prior knowledge of the degradation models and large-scale realistic training data.

5. Conclusions

We propose a parallel dual-branch disentangling network for decouple the global and local blur features. The network decomposes the feature extraction process into two branches. Through a phased training strategy, the network is trained to decouple and analyze global and local blur features. The attention fusion module is used to dynamically guide the reconstruction of restoration image. The feature map extracted by the branch network shows that the proposed dual-branch network can extract complementary features. The experimental results show that, compared with the existing dynamic scene deblurring methods, the proposed method significantly improved the subjective and objective performances, and the running speed has also been accelerated.

Our work insight that the task-specific branch training brings great promising to disentangle the degradation factors in real-world low-quality images. We provide a potential way to explore a partially interpretable framework for dynamic restore the real blur images.

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

Funding: This work in this paper is supported by the National Natural Science Foundation of China (No.61471013), the Beijing Education Committee Cooperation Beijing Natural Science Foundation (No. KZ 201910005007, No. KZ201810005002).

Data Availability Statement: The GoPro dataset is openly available in https://github.com/SeungjunNah/DeepDeblur_release accessed on 8 November 2017. The representative results of the propose method are openly available in <https://doi.org/10.6084/m9.figshare.14068721.v1> accessed on 20 February 2021.

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

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
GAN	Generative adversarial network
CBAM	Convolutional block attention module
LB-Net	Local-Branch-Net
DB-Net	Dual-Branches-Net
DBAF-Net	Dual-Branch attention fusion net

References

1. Bahat, Y.; Efrat, N.; Irani, M. Non-uniform blind deblurring by reblurring. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3286–3294.
2. Chan, T.F.; Wong, C.K. Total variation blind deconvolution. *IEEE Trans. Image Process.* **1998**, *7*, 370–375. [[CrossRef](#)] [[PubMed](#)]
3. Cho, S.; Lee, S. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 Papers*; ACM: New York, NY, USA, 2009; pp. 1–8.
4. Goldstein, A.; Fattal, R. Blur-kernel estimation from spectral irregularities. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 622–635.
5. Pan, J.; Hu, Z.; Su, Z.; Yang, M.H. Deblurring text images via L0-regularized intensity and gradient prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2901–2908.
6. Xu, L.; Jia, J. Two-phase kernel estimation for robust motion deblurring. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 157–170.
7. Xu, L.; Zheng, S.; Jia, J. Unnatural l0 sparse representation for natural image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1107–1114.
8. Whyte, O.; Sivic, J.; Zisserman, A.; Ponce, J. Non-uniform deblurring for shaken images. *Int. J. Comput. Vis.* **2012**, *98*, 168–186. [[CrossRef](#)]
9. Harmeling, S.; Michael, H.; Schölkopf, B. Space-variant single-image blind deconvolution for removing camera shake. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 829–837.
10. Levin, A. Blind motion deblurring using image statistics. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2007; pp. 841–848.
11. Chakrabarti, A.; Zickler, T.; Freeman, W.T. Analyzing spatially-varying blur. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2512–2519.
12. Tai, Y.W.; Kong, N.; Lin, S.; Shin, S.Y. Coded exposure imaging for projective motion deblurring. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2408–2415.
13. 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.
14. Gong, D.; Yang, J.; Liu, L.; Zhang, Y.; Reid, I.; Shen, C.; Van Den Hengel, A.; Shi, Q. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2319–2328.
15. Schuler, C.J.; Hirsch, M.; Harmeling, S.; Schölkopf, B. Learning to deblur. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 1439–1451. [[CrossRef](#)] [[PubMed](#)]
16. Yan, R.; Shao, L. Blind image blur estimation via deep learning. *IEEE Trans. Image Process.* **2016**, *25*, 1910–1921.
17. Noroozi, M.; Chandramouli, P.; Favaro, P. Motion deblurring in the wild. In *German Conference on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 65–77.

18. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3883–3891.
19. Ramakrishnan, S.; Pachori, S.; Gangopadhyay, A.; Raman, S. Deep generative filter for motion deblurring. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2993–3000.
20. 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, 18–22 June 2018; pp. 8183–8192.
21. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8878–8887.
22. Krishnan, D.; Fergus, R. Fast image deconvolution using hyper-Laplacian priors. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2009; pp. 1033–1041.
23. Liu, R.; Li, Z.; Jia, J. Image partial blur detection and classification. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
24. Shen, Z.; Wang, W.; Lu, X.; Shen, J.; Ling, H.; Xu, T.; Shao, L. Human-aware motion deblurring. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 5572–5581.
25. Babacan, S.D.; Molina, R.; Do, M.N.; Katsaggelos, A.K. Bayesian blind deconvolution with general sparse image priors. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 341–355.
26. Fergus, R.; Singh, B.; Hertzmann, A.; Roweis, S.T.; Freeman, W.T. Removing camera shake from a single photograph. In *ACM SIGGRAPH 2006 Papers*; ACM: New York, NY, USA, 2006; pp. 787–794.
27. Ledig, C.; Theis, L.; Huszár, 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 International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 4681–4690.
28. 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, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8174–8182.
29. Li, L.; Pan, J.; Lai, W.S.; Gao, C.; Sang, N.; Yang, M.H. Dynamic scene deblurring by depth guided model. *IEEE Trans. Image Process.* **2020**, *29*, 5273–5288. [[CrossRef](#)] [[PubMed](#)]
30. Pan, J.; Liu, S.; Sun, D.; Zhang, J.; Liu, Y.; Ren, J.; Li, Z.; Tang, J.; Lu, H.; Tai, Y.W.; et al. Learning dual convolutional neural networks for low-level vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3070–3079.
31. Woo, S.; Park, J.; Lee, J.Y.; So Kweon, I. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
32. Qian, R.; Tan, R.T.; Yang, W.; Su, J.; Liu, J. Attentive generative adversarial network for raindrop removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2482–2491.
33. Sukanuma, M.; Liu, X.; Okatani, T. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9039–9048.
34. Li, X.; Yang, F.; Lam, K.M.; Zhuo, L.; Li, J. Blur-Attention: A boosting mechanism for non-uniform blurred image restoration. *arXiv* **2020**, arXiv:2008.08526.
35. Purohit, K.; Rajagopalan, A. *Region-Adaptive Dense Network for Efficient Motion Deblurring*; AAAI: Palo Alto, CA, USA, 2020; pp. 11882–11889.
36. Chen, L.; Sun, Q.; Wang, F. Attention-adaptive and deformable convolutional modules for dynamic scene deblurring. *Inf. Sci.* **2021**, *546*, 368–377. [[CrossRef](#)]
37. Zhang, X.; Dong, H.; Hu, Z.; Lai, W.S.; Wang, F.; Yang, M.H. Gated fusion network for joint image deblurring and super-resolution. *arXiv* **2018**, arXiv:1807.10806.
38. Gao, H.; Tao, X.; Shen, X.; Jia, J. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3848–3856.
39. Zhang, K.; Zuo, W.; Zhang, L. Deep plug-and-play super-resolution for arbitrary blur kernels. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1671–1681.
40. Roth, S.; Black, M.J. Fields of Experts. *Int. J. Comput. Vision* **2009**, *82*, 205–229. doi:10.1007/s11263-008-0197-6. [[CrossRef](#)]
41. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)] [[PubMed](#)]
42. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
43. Jiang, Z.; Zhang, Y.; Zou, D.; Ren, J.; Lv, J.; Liu, Y. Learning Event-Based Motion Deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3320–3329.
44. Yuan, Y.; Su, W.; Ma, D. Efficient Dynamic Scene Deblurring Using Spatially Variant Deconvolution Network With Optical Flow Guided Training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3555–3564.