End-To-End Rotational Motion Deblurring Method Combining with Motion Information

Yudan Qiu¹ and Cheng Zhang¹,*

¹Key Laboratory of Dynamics and Control of Flight Vehicle, Ministry of Education, Beijing Institute of Technology, Beijing 100081, China

*E-mail: zhangcheng@bit.edu.cn

Abstract. We propose an end-to-end rotational motion deblurring method based on conditional generation adversarial networks. The proposed method calculates the blur path value of each pixel on the rotational motion blurred image to provide a priori information of its blur degree, and then connects it to the blurred image as the input of the network. In addition, a rotational motion blurred image dataset is produced, which contains different degrees of rotational motion blurred images, as an evaluation dataset for the method to the effect of rotational motion deblurring. Experiments show that the proposed method is superior to existing end-to-end deblurring methods in both qualitative and quantitative analysis when dealing with different degrees of rotational motion blur.

1. Introduction

When the camera is exposed, camera shake or motion relative to the target scene will cause the quality of the captured image to be degraded and visually blurred. Therefore, the images captured by a camera mounted on a rotational missile or a machine moving at high speed will appear blurred in the rotational motion. Deblurring methods are used to process such images to restore visually sharper images, which is not only conducive to human target recognition in the loop, but also can improve the effect of automatic target recognition methods.

Restoring a sharp image from a motion blurred image is essentially generating pixels of a sharp output image from the pixels of the input blurred image. This process can be defined as image translations. The conditional generation adversarial network has achieved good results in many image translations tasks [1]. Compared with ordinary CNNs, the generated images can better restore the texture and details.

At present, the mainstream methods for removing motion blur of an image are mainly aimed at images taken by a handheld camera. Such blurs are mainly caused by a slight shaking of the human hand when taking images. Rotational motion blurred images, compared with this type of blurred images, have a larger blur path, and the degree of blurring changes greatly with the change of the rotation angle of the camera during the exposure time. To solve this problem, the proposed method calculates the blur path value of each pixel of the image from the rotation angle of the image during the exposure time, and then connects it to the image as the input of the network to provide a priori information on the degree of blur. In addition, due to the lack of existing rotational motion blur image datasets, we also produced a rotation motion blur dataset containing 600 images with a rotation angle from 0.06 to 0.20.
2. Related work

Motion-induced image degradation can be modeled as

\[ B = kS + n \]  

(1)

Where \( B, S, k, n \) is vectorized blurred image, latent sharp image corresponding to the blurred image, the blur kernel matrix, and the random noise when collecting the image, respectively.

Early work usually restored the image from the blurred image by deconvolution when \( k \) is known. This process is called non-blind restoration. The commonly used methods are Lucy-Richardson method, Wiener filter, Tikhonov filter. The process of restoring a latent sharp image from a blurred image when \( k \) is unknown is called blind restoration. One type of blind restoration is to estimate the blur kernel \( k \) of the motion through priors such as image motion, and then apply the deconvolution method in non-blind restoration to restore. Finding a corresponding blur kernel for each pixel is a problem that is ill-posed and computationally intensive. Therefore, it is traditionally assumed that the blur kernel does not change with the change of spatial position. This is called space invariant blur or uniform blur.

Recently, end-to-end methods based on deep learning have begun to appear. This type of method no longer needs to explicitly estimate the blur kernel, but directly inputs the blurred image into the network, and then outputs the restored image. This type of method can be roughly divided into two categories, a multi-scale CNN method and an adversarial generation network (GAN) -based method. S. Nah [2] et al. imitated the traditional coarse-to-fine method and created a multi-scale network (DeepDeblur). Xin Tao [3] et al. used a recurrent neural network combined with the idea of multi-scale to propose a network called SRN. After that, Zhang et al. [4] combined the idea of space pyramid matching and multi-scale methods to propose a network named DMPHN, which further improved the deblurring effect, and its lightweight structure allowed real-time deblurring became possible. On the other hand, since Isola et al. applied conditional GANs to image translation tasks, they have made great achievements in the fields of image restoration, in-painting, and super-resolution. Kupyn et al. [5] first applied it to the field of image deblurring and created the DeblurGAN network. Later, Kupyn et al. used the FPN (feature pyramid network) as a generator to get an improved version of DeblurGAN-v2 [6] on the basis of DeblurGAN.

3. Rotational motion blur dataset

Although the deblurring of motion blur has gradually evolved from uniform blur to non-uniform blur. However, the research object of the existing deblurring methods is mainly the blurry images generated by the camera shake when people hold the camera. At present, the common datasets in the deblurring field (such as GoPro, Koehler [7], etc.) are also based on this assumption. Although this type of motion blur kernel changes with the change of spatial position, the degree of change is not great.

Rotational motion blur is a type of blur that changes drastically with the change of spatial position. Its blur type and general dataset are quite different. As can be seen from Figure 1, there is a big difference between the rotation motion blur images and the ordinary motion blur images.

![Figure 1. Rotational motion blur images and ordinary motion blur images](image-url)
As far as I know, there are currently no publicly available rotational motion blurred image datasets. For machine learning methods, the importance of data is self-evident, so we obtain the rotation motion blur dataset according to Hong's [8] method. Some pictures of the dataset show in figure 2. The rotation angle of the camera during the exposure time was changed from 0.06 to 0.20 (radians). A total of 600 images were randomly divided into training (540) and test (60) sets. This dataset was used to evaluate the effectiveness of the method in rotational motion deblurring. If you want to get the dataset, please contact us at 1047297617@qq.com.

Figure 2. Some images in the Rotational motion blur dataset

4. Proposed method
The disadvantage of the end-to-end method is that it cannot handle different degrees of rotation blur, which is caused by using the same inference process for different degrees of blur. The proposed method is an end-to-end deblurring method that combines motion information. It can better remove different degrees of rotational motion blur than the simple end-to-end method, and does not need to estimate the blur kernel explicitly like non-end-to-end methods.

4.1. Blur field generation
The degree of blur of a rotational motion blurred image increases with the increase of the camera's rotation angle during the exposure time. The blur path of each pixel in the rotational blurred image also increases as its distance from the center of rotation increases. As shown in figure 3, the existing end-to-end method becomes worse when processing blurred images with gradually increasing rotation angles. This is because in the absence of prior information about the degree of blurring of the image, it takes Caused by the same inference process. According to the rotation angle and the distance of each pixel from the center of rotation, the blur path in the exposure time is calculated to form a blur domain, which can be connected to the input network as a priori information and blur image, which can solve this problem well. When the exposure time of the camera is set, the rotational angular velocity of the camera carrier obtained in real time by the inertial guidance element, and the value of the blur path at each point on the image is calculated accordingly.

\[ p(x, y) = ((x - \bar{x})^2 + (y - \bar{y})^2)^{\frac{1}{2}} t \omega \]  

(2)

Where \((\bar{x}, \bar{y}), t, \omega\) is the coordinates of the rotation center, the exposure time of the camera, and the angular speed of the camera rotation, respectively. The value of the blur path is converted into a uint8 format to form a blur field that is consistent with the size of the input image, and is connected to the RGB channel of the image as an input to the network.

4.2. Network architecture
The network architecture in this paper is conditional generation adversarial networks (cGANs). The generator uses an FPN network capable of providing multi-scale information, and the discriminator uses triple-scales discriminators that operate on three scales. The overall loss of the generator consists of four items: MSE loss, content loss, adversarial loss, and gradient loss. And loss of discriminator is
the least square GAN (LSGAN) [9]. The generator uses (FPN-inception) in delburGANv2. This structure adds multi-scale information to the network and simplifies the network structure compared to ordinary multi-scale methods.

The discriminator uses the triple-scales discriminator that combines three scale information. The structure of the discriminator is the same as in the patch discriminator. The images are input into a discriminative network with a receptive field of $16 \times 16$, $70 \times 70$, $286 \times 286$, and the output is averaged to obtain the output of the discriminator. Note that the input image is resized to $256 \times 256$ during training, and the discriminator with a receptive field of $286 \times 286$ needs to be padding first. In addition, unlike the patch discriminator, the sigmoid layer is removed. Discriminators of three scales can provide information from local to global of the image, so that the restored image takes into account both local information and global content.

4.3. Loss function

We adopt LSGAN[9]. The discriminator loss is

$$L(D) = E_{x \sim P_{data}(x)}[(D(x) - 1)^2] + E_{z \sim P_{z}(z)}[(D(G(z))^2]$$

(3)

The generator's loss function consists of four parts, MSE loss, content loss, adversarial loss, and gradient loss.

The MSE loss is a commonly used loss function for image restoration tasks. The loss function is defined as follows:

$$L_{MSE} = \frac{1}{3wh} \|L-S\|^2$$

(4)

Where $L$, $S$, $w$, $h$ denotes the model output and ground truth image, width and height of images, respectively.

The content loss [10] is difference between the VGG-19 [11] conv3.3 feature maps of the ground truth and output images. The feature maps represent the abstract content of the images. The weights are pretrained on ImageNet. The loss function is defined as follows:

$$L_{content} = \frac{1}{w_{3,3}h_{3,3}} \left\| \phi_{3,3}(S) - \phi_{3,3}(L) \right\|^2$$

(5)

Where $\phi_{3,3}, w_{3,3}, h_{3,3}$ denotes feature maps obtained by the third convolution layer before the third maxpooling layer, $w_{3,3}, h_{3,3}$ are the dimensions of the feature maps.

The adversarial loss is defined as follows:

$$L_{adv} = E_{z \sim P_{z}(z)}[(D(G(z)) - 1)^2]$$

(6)

The reason for adding the gradient loss is the obvious difference in the gradient value between the sharp image and the blurred image. Specifically, the average gradient value of an image is calculated using a Sobel operator. Combining the four losses to get the total loss of the generator.

$$L(G) = 0.5* L_{MSE} + 0.006* L_{content} + 0.01* L_{adv} + 0.001* L_{gradient}$$

(7)

5. The experiments results

We use the deep learning framework Pytorch to implement the model, set the batch to 1 on the 1080Ti GTX GPU, and trained 300 epochs, which took about 3 days. The images were resized to $256 \times 256$ during training. The optimizer uses the Adam optimizer with an initial learning rate of $10^{-4}$. The learning rate is fixed at the first 150 epochs, and then decreases linearly to $10^{-7}$ in the next 150 epochs. Backbones chose inception-resnetv2, and its initial weights were trained on ImageNet.
The contrast models (DeblurGAN, DeblurGAN-v2, DMPHN) are all trained on the proposed rotational motion blur dataset according to the training method in their paper. Each model is inferred on the rotational motion blur test dataset, and the PSNR and SSIM of the restored image and the sharp image are calculated. The results are shown in Table 1. Combining Figure 3 can be found that when dealing with different degrees of rotational motion blur, the proposed method is better than the existing end-to-end methods in both quantitative and qualitative.

|         | DeblurGAN | DeblurGAN-v2 | DMPHN | Ours |
|---------|-----------|--------------|-------|------|
| PSNR    | 20.53     | 23.05        | 21.26 | 24.4 |
| SSIM    | 0.50      | 0.66         | 0.63  | 0.72 |

Table 1. PSNR and SSIM comparison

Figure 3. Rotational motion blur image when the rotation angle is 0.06, 0.08, 0.10, 0.12, 0.14 from top to bottom, as well as the end-to-end deblurring method and the deblurring effect of the proposed method. As can be seen from the figure, the proposed method has better visual effects when dealing with different degrees of rotational motion blur (especially when the rotation angle is larger).

6. Conclusion
This paper proposes an end-to-end rotation motion blur deblurring method based on cGANs. The blur field is calculated based on the motion information, and the gradient loss is added to the loss function. The rotational motion blur dataset is also established. The experimental results show that compared...
with other methods, the proposed method can better deal with the rotational motion blur in quantitative analysis and vision effects.

Reference

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros; Image-To-Image Translation With Conditional Adversarial Networks The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1125-1134

[2] Nah, Seungjun, T. H. Kim, and K. M. Lee. "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2017.

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

[4] Hong guang Zhang, Yuchao Dai, Hongdong Li, Piotr Koniusz; Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring.The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 5978-5986

[5] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Ji Matas.Deblurgan: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8183–8192, 2018.

[6] Orest Kupyn, Tetiana Martyniuk, Junru Wu, Zhangyang Wang; DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. The IEEE International Conference on Computer Vision (ICCV), 2019, pp. 8878-8887.

[7] Rolf Köhler, et al. "Recording and Playback of Camera Shake: Benchmarking Blind Deconvolution with a Real-World Database." European Conference on Computer Vision Springer, Berlin, Heidelberg, 2012.

[8] Hanyu Hong and Tianxu Zhang "Fast restoration approach for rotational motion blurred image based on deconvolution along the blurring paths," Optical Engineering 42(12), (1 December 2003).

[9] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang.Least squares generative adversarial networks, 2016. Cite arxiv:1611.04076. 4

[10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, 2016.

[11] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv e-prints, Sept. 2014. 4