Octave Residual Network for Image Motion Deblurring

An Qi, Jiang Lifen, Sun Huazhi, Liang Yan, Xu Tongtong, Xu Fan, Zhou Yongheng

College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China
xiaoanqi@uzoo.cn; wxxjlf@sina.com; sunhuazhi@mail.tjnu.edu.cn; mail_liangyan@163.com; tngt1206@163.com; 854955915@qq.com; 13132565021@163.com

Abstract. Image motion deblurring technology is an essential topic. Traditional methods with a high algorithm complexity when estimating the blur kernel require a lot of calculations. To solve this problem, we proposed Octave convolution residual block, and proposed Octave Residual Network (ORN) based on this block. ORN is a Generative Adversarial Networks, where the generator is a multi-scale recurrent network, and the discriminator is a deep neural network. ORN restore sharp images in an end-to-end manner where blur is caused by various sources. We present multi-scale loss function that mimics conventional coarse-to-fine approaches. To verify the validity of ORN, we trained our model on GOPRO dataset, the peak signal-to-noise ratio (PSNR) of the model reached 30.04, and the operation time of each step is reduced by 0.04 seconds. Experimental results show that our method achieves the state-of-the-art performance of image motion deblurring not only in performance but also in training speed.

1. Introduction

When acquiring images, relative motion between the shooting equipment and things will cause image quality degradation and produce uneven blurring, resulting in people being unable to obtain useful information from the images. Therefore, the image deblurring technology has attracted people's attention. There are many ways to reduce image motion blur. For instance, we can reduce motion blur by shortening the exposure time of the sensor, but it will cause high hardware cost, require stronger illumination, and cause higher noise. In recent years, with the improvement of computing equipment, Convolutional Neural Networks (CNN) have been used in many complex computer vision fields, at the same time, the Generative Adversarial Network performs well in the study of image degradation. Therefore, we build a deep learning model based on the Generative Adversarial Network to remove image motion blur.

Firstly, we propose an octave convolution residual block based on octave convolution, and then we propose a motion blur removal model ORN, which integrates the octave convolution residual block into the GAN network to enhance the effect of image deblurring. The octave convolutional residual block in the model pays more attention to high-frequency information and ignores low-frequency information. The introduction of this structure reduces the redundant parameters. We also introduce the context model as a basic block to expand the receptive field. Based on the above advantages, the ORN model we proposed to remove image motion blur has higher training efficiency, shorter test time and higher picture quality.
2. Related work

Traditional methods of removing image motion blur generally use complex blur functions. For instance, literature [1] proposes a method to estimate fuzzy kernels separately. Literature [2] derived fuzzy kernels based on the statistical law that natural image gradients obey the heavy-tail effect. Literature [3] provides reliable edge information for fuzzy kernel estimation based on the protruding edges and low rank priors of the blurred image, and low rank priors provides reliable priors for intermediate images. Levin et al. [4] proposed a method to remove image motion blur based on this method combined with Bayesian minimum mean square error sampling algorithm for deconvolution.

Although the above methods can remove noise to a certain extent and effectively improve the quality of motion blurred images, faced with motion blur images of unknown blur kernels, they require complex calculations to estimate the blur kernels, resulting in increased model computational overhead. Recently, CNN has made significant progress in deblurring single images. For instance, literature [5,6] used CNN to estimate the non-uniform fuzzy kernel. At the same time, there are many end-to-end CNN models [7-9] for image deblurring. For instance, Nah et al. [7] proposed a multi-scale convolutional neural network, which start with a blurred image with a coarser ratio at a lower resolution, and gradually restores the image at a higher resolution until it reaches full resolution. This method uses a multi-scale mechanism from rough to excellent. In recent research, it can be found that the generative adversarial network (GAN) performs well in dealing with image degradation, and it can produce clear and realistic effects [13,10]. However, the amounts of redundant parameters of the current defuzzification model is often very large, which results in a long operation time.

Aiming at the problems of high computational cost and large redundant parameter of deep learning model by traditional methods, we propose an octave convolutional residual block, and use it to construct a generator to generative adversarial network, and propose ORN for image deblurring. The experimental results show that the ORN takes less time and the blur removal effect is better.

3. Proposed model

Our proposed model ORN consists of two primary modules, as shown in Figure 1. The generator is used to generate image information. The blurred image B passes through the generator to generates the fake image G(B) to fool discriminator. The discriminator is used to prevent the generator from generating fake images. The clear sample image S and the generated image G(B) are both used as inputs to discriminator D to obtain a probability, which is expressed the performance of the generator G. The two modules are trained against each other until the generator generates a spurious image.

![Figure 1. GAN model framework](image)

3.1. Generator

The generator is the most important part in ORN. The blurred image is passed through the generator Network to generate a deblurred image. The structure of the generator in our model is shown in Figure 2, which uses a multi-scale network architecture.
To utilize the information both in coarse level, and in fine level, the network adopts a three-layer network structure of Gaussian Pyramids. The size of input and output at each scale level is different. Taking the coarsest level network as an example, the input image is first converted into a highly abstract feature map by the encoder. Then the features are passed through a TCN module, which will produce useful information to the next stage to help restore the clearer image. Finally, the estimated image is output by the decoder. The structure of the network with different scales is the same. The image generated by the network of coarser level is input as the input of the network of finer level. The finer the level of the network is, the clearer the images are generated.

3.1.1. Encoder

The core of neural network computing is to extract the characteristics of the data. The encoder structure is introduced in the generator to compress the original picture effectively. To get the most important semantic information, improve the efficiency of the model and reduce the operation cost, the Octave Convolution [11] is introduced into the residual block. The Octave Convolution Residual Block is proposed to be used as the basic block of the GAN generator's encoder network, as shown in Figure 3.

Figure 3. The structure of Encoder based on Octave Convolution Residual Block
The encoder includes three convolutional layers. Each convolutional layer is followed by an Octave Convolutional Residual Block. It is worth noting that the first and second convolutional layers are followed by three octave residual blocks, and the last convolutional layer is followed by two Octave Convolution Residual Blocks. To further expand the receptive field of the model, the context [12] module is introduced at the last layer of the encoder. The context module includes four layers of parallel dilated convolutions with different dilated rates. The four dilated rates are set to 1, 2, 3, and 4, respectively. The output of the previous layer is concatenated by four dilated convolutions so that the encoder can extract more diverse features, which makes the network to eliminate ambiguity.

3.1.2. Decoder

The processed encoded information must be gradually mapped to the original size by the decoder to obtain the image of original resolution size. The structure of decoder in our model is shown in Figure 4.

![Figure 4. the structure of Decoder Based on Octave Convolution](image)

The decoder and encoder are roughly opposite in structure, and every three stacked octave residual blocks will be connected to a deconvolution layer. These deconvolution layers double the dimensions of the input features and halve the number of channels to output the estimated clear image.

3.2. Discriminator

The list of authors should be indented 25 mm to match the abstract. The style for the names is initials then surname, with a comma after all but the last two names, which are separated by ‘and’. Initials should not have full stops—for example A J Smith and not A. J. Smith. First names in full may be used if desired. If an author has additional information to appear as a footnote, such as a permanent address or to indicate that they are the corresponding author, the footnote should be entered after the surname.

In GAN, the parameter update of the generator is not directly from the data samples, but from the backpropagation of the discriminator. GAN can often generate samples that are closer to real and clear images than other models. The discriminator of PatchGAN [37] is applied in our proposed model. Generally, GAN often outputs a "true" or "false" vector to represent the evaluation of the entire image. PatchGAN divides the input image into N×N blocks and outputs an N×N matrix. Each element in this matrix represents the "true" or "false" of each block of the input image. Finally, all the results in the matrix are averaged and output.

Since the size of a small block is smaller than the original image in the discrimination process, the number of parameters is smaller, and the running speed is faster. Furthermore, this kind of discriminator can run on pictures of any size. The discriminator network structure of our proposed model is shown in Figure 5.
3.3. Loss function
To optimize the network parameters, we adopt the combination of multi-scale content loss and antagonistic loss to train the model. The output of each layer can be the highest resolution image in this scale by adopting a coarse to fine method.

To prevent the occurrence of overfitting and obtain a better deblurring effect, this paper uses the L2 norm. The multi-scale content loss function is defined as shown in formula (1):

$$L_{cont} = \sum_{i=1}^{n} \frac{k_i}{N_i} \|I^i - J^i\|_2^2$$

where $I^i$, $J^i$ represent the model output and the corresponding clear image at each scale level, $k_i$ is the weight of each scale, and $N_i$ is the number of elements in $I^i$ to be standardized.

The definition of the anti-loss function is shown in formula (2):

$$L_{adv} = \sum_{n=1}^{N} -D_{\theta_{D}}(G_{\theta_{G}}(I^B))$$

where $G_{\theta_{G}}$ represents the generated sample, and $D_{\theta_{D}}$ represents the discrimination of the real sample. At training time, the generator tries to minimize the adversarial loss, while the discriminator tries to maximize the adversarial loss.

By combining multi-scale content loss and confrontation loss, the total loss function is expressed as formula (3):

$$L_{total} = L_{cont} + \lambda L_{adv}$$

where $\lambda$ represents the weight coefficient. After many experiments and comparisons, $\lambda = 1 \times 10^{-2}$ was finally obtained to obtain the best results.

4. Experiments
To verify the effectiveness of ORN model, we compare our method with previous state-of-the-art image deblurring approaches on both evaluation and real images. The methods of CNN-15 [14], l0 norm based deblurring method [13], and SRN [9] are selected.

4.1. Dataset
GOPRO dataset [8]. We use GOPRO dataset to train the network and selects 3,214 blurred/clear image pairs. We use 2103 pairs of images to train the network, and the evaluate model by remaining 1111 pairs.

4.2. Settings
We implement our framework on TensorFlow platform. The comparative experiments are conducted on a PC configured with Inter i7 CPU and NVIDIA RTX 2080Ti. All comparative experiments have the same training configuration and are performed on the same dataset.

The choice of parameters has a great influence on the accuracy of the experiment. In this paper, we adjust the parameters when training the model manually and explore different configurations of the...
parameters so that the model can achieve the best effect. The optimizer parameter settings are shown in Table 1.

| Optimizer | Adam solver |
|-----------|-------------|
| β1 (exponential decay rate of the first-order moment estimate) | 0.9 |
| β2 (second-order moment estimated exponential decay rate) | 0.999 |
| ε (prevent division by 0 during implementation) | 10^-8 |
| Learning rate | Initial value 0.0001 to 1e-6 exponentially decay |

4.3. Experimental Results and Analysis
We explore the performance of the proposed model for removing image motion blur from three aspects: visual effects, peak signal-to-noise ratio, and training time.

The ORN model converges after training 200,000 times on the above training set and tests the model effect in the test set. Figure 6 shows the effect of removing the motion blur of the image of the vehicle in the GOPRO dataset. The left image is the original blurred image in the dataset, and the right image is the clear image generated by the proposed model. It can be seen that our model can achieve a clear resolution in removing the visual effect of image motion blur. The runtime and test time of each step during training are reduced by 0.04 seconds.

Figure 6. Test results of removing image motion blur

We use the PSNR to measure the effect of image processing. The higher the PSNR value, the higher the similarity between the deblurred image and the original image. We compare PSNR of proposed model with that of CNN-15, the deblurring method based on the l0 norm, and our previous proposed SRN method. The experimental results are shown in Table 2.

| method | PSNR  |
|--------|-------|
| CNN-15 | 28.63 |
| Based on l0 norm | 28.32 |
| SRN | 29.57 |
| The algorithm of this article | 30.04 |

5. Conclusions
We proposed Octave Convolution Residual Block based on Octave Convolution, and introduced the Octave Convolutional Residual Block into the GAN network to obtain Octave Residual Network (ORN). The TCN and the context module in ORN model use multi-layer dilated convolution to increase the
receptive field and better capture multi-scale context information. Experimental results show that ORN not only achieves good results in visual effects, but also has a higher PSNA than other baselines, which reduces the redundancy of the network and speeds up the image processing speed of the network.

Acknowledgments
This work was supported by National Natural Science Foundation of China under Grant No.61703306, 61702370, Tianjin Natural Science Foundation under Grant No.18JCYBJC85900, 18JCQNJC70200, 17JCYBJC16400, Tianjin Science and Technology Project under Grant No.17ZLZXZF00530, 131 Three-Level Personnel of Tianjin Normal University under Grant No.043/135305QS20, the Doctoral Foundation of Tianjin Normal University under Grant No. 52XB1102, 52XB1302.

References
[1] LevinA, WeisY, DurandF, etal. Efficient marginal likelihood optimization in blind deconvolution [R]. Technical report, Weizmann-Inst-TR, 2011: 409-417.
[2] Fergus R, Singh B, Hertzmann A, et al. Removing camera shake from a single photograph[J]. ACM Transactions on Graphics, 2006, 25(3): 787-794.
[3] Shijie Sun, Huaici Zhao, Bo Li, Mingguo Hao, Jinfeng Lv. Blind Deconvolution for Noisy and Blurry Images Using Low-Rank Prior [J]. Journal of Electronics and Information Technology, 2017, 39 (08): 1919-1926.
[4] LevinA, WeisY, DurandF, etal. Efficient marginal likelihood optimization in blind deconvolution [R]. Technical report, Weizmann-Inst-TR, 2011: 409-417.
[5] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In CVPR, 2015.
[6] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian D Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In CVPR, 2017.
[7] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deburring. In CVPR, 2017.
[8] Jiawei Zhang, Jinshan Pan, Wei-Sheng Lai, Rynson WH Lau, and Ming-Hsuan Yang. Learning fully convolutional networks for iterative non-blind deconvolution. In CVPR, 2017.
[9] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In CVPR, 2018.
[10] Z. Liu, R. Yeh, X. Tang, Y. Liu and A. Agarwala. Video frame synthesis using deep voxel flow. In ICCV. IEEE, 2017.
[11] Chen LC, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.
[12] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.
[13] Song C, Lin X. [IEEE ICC 2015 - 2015 IEEE International Conference on Communications - London, United Kingdom (2015.6.8-2015.6.12)] 2015 IEEE International Conference on Communications (ICC) - Blurred License Plate Recognition based on a single snapshot from drive recorder[C]// IEEE International Conference on Communications. IEEE, 2015:7108-7113.
[14] Qi An, Lifen Jiang, Huazhi Sun, Yan Liang, Chunmei Ma. Model for removing motion blur of vehicle image based on the generative adversarial network [J]. Journal of Tianjin Normal University (Natural Science Edition), 2020, 40 (01): 76-80.