Super-resolution reconstruction of unmanned aerial vehicle image based on deep learning

Lixuepiao Wan¹, Gongping Zhang²

¹College of Mechanical and Power, China Three Gorges University, No.8 Daxue Road, Yichang, CHN
²College of Computer and Information Engineering, Hubei University, 368 Friendship Avenue, Wuhan, CHN
2018105313@ctgu.edu.cn

Abstract. Due to the influence of weather, air flow and other environmental factors, unmanned aerial vehicle (UAV) aerial photography often leads to problems such as blur, defocus and so on. In order to solve the problem of low definition of UAV aerial images, this paper proposes a super-resolution reconstruction algorithm based on improved residual learning blocks (IRB) to reconstruct aerial images to achieve high definition. This method can effectively solve the problem of gradient explosion which is often encountered in deep learning network. At the same time, it can suppress the learning of useless information and make full use of important feature information. Through the establishment of UAV aerial data set for cross training of the network, the network can better adapt to the UAV aerial environment. The experimental results show that, compared with the existing neural network algorithm, the improved residual network algorithm can reconstruct UAV aerial image well, and the subjective visual effect of the reconstructed image is better while the edge information is greatly preserved.

1. Introduction

With the development of electronic technology, UAV aerial photography technology has been applied in many fields, such as military investigation, resource exploration and traffic scheduling, and has become a necessary means of solving industry problems. Compared with the traditional fixed point camera, UAV aerial photography technology has the advantages of mobility, small size and strong maneuverability. In the complex traffic scene, UAV aerial photography technology is used to obtain real-time images, which can provide important supplementary information for traffic scheduling. With the deepening of infrastructure construction, urban road traffic management will become an important field of UAV aerial photography technology development, and has broad application prospects.

In the actual situation, when the UAV vibrates or is interfered by birds, or it may encounter some unavoidable external adverse factors such as fog, strong light, rain and dim light, the quality of the collected image will be seriously degraded, resulting in low contrast, surface atomization, information loss and other problems. Due to the high uncertainty of UAV environment, the traditional image processing methods include image enhancement, restoration and super-resolution reconstruction algorithm, and the limitations in the scope of application, accuracy and speed are gradually highlighted. Therefore, it is very practical and necessary to seek a fast, accurate and applicable method for UAV image reconstruction.
In recent years, scholars have made great achievements in the field of UAV aerial image reconstruction. Chen G et al. [1] proposed an image restoration method based on physical model, established the model, calculated the depth ratio of each scenic spot, and carried out the restoration work according to the optical principle. Yan H et al. [2] proposed a new Lucy-Richardson (LR) image restoration algorithm, which makes up for the defects of the original algorithm by regularization processing and local neighborhood correlation method in non subsampled contourlet transform domain. Zhao L et al. [3] proposed a super-resolution blind restoration algorithm for Defocused blurred images based on differential image autocorrelation, which can reduce the factors to be considered in the calculation and improve the accuracy of the results. In order to improve the reconstruction speed, Fang H et al. [4] introduced the sub Gaussian random projection into the CS (compressed sensing) theory, which can accurately reconstruct under the condition of a certain number of measurements. Li Y et al. [5] proposed an improved parameter estimation and image restoration algorithm, which achieved good restoration effect and improved the anti noise performance of image restoration.

However, with the increase of data volume, the common problems of traditional image reconstruction methods are increasingly exposed: 1) the algorithm complexity is high, resulting in the actual image processing time is too long, and the effectiveness can not be met; 2) The image reconstruction algorithm is feasible, but the subjective visual effect of the actual reconstructed image is not ideal. 3) the intelligence of the algorithm is not high, and the requirement of hardware is high, which limits the application scope of the algorithm to a large extent. Therefore, image reconstruction based on deep learning has become the mainstream.

With the continuous in-depth study of scholars in the field of deep learning, the restoration and reconstruction of UAV aerial images has made an unprecedented breakthrough. Hu C et al. [6] proposed an image super-resolution reconstruction algorithm based on convolutional neural network, which avoids the over compression of important information of image and the shock of network gradient update through SRCNN (Super-resolution convolutional neural network) and NAG (Randomized rectified linear unit, RReLU) approaches. Xiao J et al. [7] combined with the existing image classification network model and visual recognition algorithm, adjusted the original convolution neural network, so that the subjective visual effect and objective evaluation index of the image were significantly improved. On the other hand, many scholars put forward new methods and ideas to solve the problems of accuracy and adaptability of traditional recognition network in the process of image recognition. Jiang Z et al. [8] proposed a learning network based on deep learning, aiming at the defects of flexibility and accuracy of traditional recognition methods, and analyzed it by using the established multi node three-layer structure, so as to improve the image recognition rate and realize the classification of UAV and non UAV. In order to solve the problem that convolution neural network (CNN) and local binary pattern (LBP) can only extract a single feature of facial expression image in the process of facial expression feature extraction, Li X et al. [9] proposed an expression recognition method based on feature fusion of deep learning. It can effectively improve the accuracy of expression recognition and is more robust to light changes.

In view of the above problems, this paper proposes a new deep network method, which is based on the important information features to suppress useless information, at the same time to achieve the purpose of reconstructing the image collected by UAV and improving the image resolution.

2. Method establishment

2.1. Super resolution reconstruction network based on residual learning structure

Different layers of convolutional neural network extract different image feature information. With the deepening of the network, the extracted information of different levels will increase, and the combination of information between different levels will also increase. We introduce residual learning structure into the super-resolution reconstruction network to solve the problem of gradient vanishing and gradient exploding. The specific network structure is shown in figure 1:
As shown in the figure above, this is the super-resolution reconstruction network model structure proposed in this paper. The model takes the unprocessed low resolution image (LR) as the input, and obtains higher resolution through network transmission. The structure of the model mainly includes:

1) The convolution layer of LR low-level information is composed of two conv;
2) A multi-scale residual learning network for learning high-level information consists of 23 improved residual blocks (IRB);
3) The feature fusion layer consists of one concat and one conv;
4) HR image reconstruction layer consists of one up sampling layer and two convolution layers.

2.2. Improved residual learning structure

In order to make the network have better learning effect, we build residual learning block to enhance the network's ability to extract information.

After applying convolution layer to the input LR image to learn low-level features, a group of RDB blocks are used to learn high-level features. Feature maps are concatenated in DenseNet instead of summation. Therefore, the i-th layer receives the characteristic maps of all the previous layers as input:

$$X_i = \max (0, w \ast [X_1, X_2, \ldots, X_{i-1}] + b)$$

In the above equation $[X_1, X_2, \ldots, X_{i-1}]$ represents the concatenation of the feature maps generated in the previous convolution layers 1, 2, ..., $i-1$. In the structure of IRB, short paths are created between one layer and other layers, which enhances the information flow through the deep network, thus alleviating the problem of vanishing gradient. In addition, the IRB structure proposed in this paper can greatly reduce the number of parameters through feature reuse, thus achieving higher performance through less memory and calculation. In the super-resolution reconstruction network, we use IRB structure as the building block of our network. The structure of each IRB learning block is shown in figure 2. In an IRB block, there are four convolution layers, three Relu layers and a concat fusion layer.

2.3. Training method

We cross train the network, divide 10000 sample pairs into 10 groups, each group contains 1000 sample pairs, and use cross training method for network training. In other words, each round of training will select 9 groups of sample pairs as the training set, and the remaining 1 group as the test set, which can adjust the network parameters. Each test set is different, and the cycle is 10 rounds. In this way, we can...
effectively screen out the data sets that are more conducive to network training, and prevent some poor training sets from affecting the overall effect of the network.

2.4. Selection of evaluation index

Peak signal to noise ratio (PSNR) and structural similarity (SSIM) are the two most widely used indicators to measure the quality of image reconstruction. The former quantitatively calculates the error between the processed result and the original image. The higher the PSNR is, the smaller the distortion is. The closer SSIM is to 1, it means that the structure of the processed image is very similar to that of the original image, that is, the resulting image is better. But in the field experiment, because there is no reference image, PSNR and SSIM can not be used. Therefore, this paper selects Blur Measure (BM) [10]-[11], Grayscale Mean Gradient (GMG) and Laplace sum (LS) [10] as evaluation indexes. The smaller the BM value is, the clearer the image is. The higher the GMG and LS values are, the higher the image quality is. The expression of BM is as follows:

\[ BM = \max(sD_{\text{vertical}}, sD_{\text{horizontal}}) \]  

\[ sD_{\text{vertical}} = \sum_{i,j=1}^{M-1,N-1} D_{\text{vertical}}(i, j) \]  

\[ sD_{\text{horizontal}} = \sum_{i,j=1}^{M-1,N-1} D_{\text{horizontal}}(i, j) \]

\[ i \in (0, M-1), j \in (0, N-1) \]

\[ D_{\text{vertical}} = |F(i, j) - F(i - 1, j)| \]

\[ D_{\text{horizontal}} = |F(i, j) - F(i, j - 1)| \]

In the equation, \( D_{\text{vertical}} \) and \( D_{\text{horizontal}} \) represent different images in the vertical and horizontal directions. \( F(i, j) \) is the pixel of \( (i, j) \) point on the image plane, \( (M, N) \) is the size of the image, BM can be normalized in the range of 0 to 1. The expressions of GMG and LS are as follows:

\[ GMG = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\alpha^2 + \beta^2} \]  

\[ \alpha = f(i, j+1) - f(i, j) \]

\[ \beta = f(i+1, j) - f(i, j) \]

\[ f(i, j) = f(i-1, j-1) - 8 \times f(i, j) - f(i+1, j-1) - \]

\[ f(i-1, j) - f(i+1, j) - \]

\[ f(i, j+1) - f(i-1, j+1) - \]

\[ f(i+1, j+1) - \]

\[ LS = \frac{1}{(M-2)(N-2)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left| f(i, j) - f(i, j+1) - f(i+1, j) - f(i, j+1) - f(i+1, j+1) \right| \]

2.5. Experimental configuration

The training platform of this experiment is: the operating system is Ubuntu 18.04 (Octa-core 2.0 GHz), and the graphics card is NIVID GTX 1080ti. The training and testing platform is CUDA8.0, cuDNN5.1, Pytorch 0.4 and python 3.6. In the setting of the main parameter learning rate, this paper adopts the method of learning rate attenuation, which can effectively prevent the training time from being too long.
The initial learning rate is set to be 0.001, and the learning rate is reduced to 1 / 10 of the original one with 100000 iterations. The Stochastic Gradient Descent (SGD) method is used to train the network.

2.6. Experimental result

In order to evaluate the effect of this algorithm on fuzzy image reconstruction more objectively and fairly, we first carry out the public data set experiment. In this experiment, this paper selects the images captured by UAV from VisDrone2019 public data set, performs down sampling operation, reconstructs the network through this paper for processing, and selects representative methods: ROBOST model [12], SRCNN model [13], FSRCNN [14] (Fast Super-Resolution Convolutional Neural Networks) model and EDSR [15] (Enhanced Deep residual networks for single image Super-Resolution) model are compared with the proposed method under the same experimental conditions.

For the test of super-resolution reconstruction network, we select two UAV images from VisDrone 2019 public data set to reconstruct UAV images with low contrast and low signal-to-noise ratio into clear images with high resolution and complete details. PSNR and SSIM are used to objectively evaluate the processing results of the five algorithms. PSNR and SSIM are calculated from the original image and the processed image. The specific test results are shown in figure 3.

| Algorithm   | PSNR/SSIM         | TIME |
|-------------|-------------------|------|
| ROBOST      | 23.56dB / 0.6432  | 3.2382s |
| SRCNN       | 25.14dB / 0.7024  | 3.1240s |
| FSRCNN      | 25.92dB / 0.7220  | 3.002s  |
| EDSR        | 26.96dB / 0.7432  | 2.8321s |
| Proposed    | 27.88dB / 0.7566  | 2.3240s |

Figure 3 Comparison of reconstruction effects of public data set.

From the perspective of image reconstruction effect, ROBOST model has not substantially improved the image blur problem, and the subjective visual effect is poor. SRCNN model eliminates some degree
of blur, but introduces a larger ringing effect, and the edge sharpness is not improved. FSRCNN model can improve the ringing effect of image, but the image has the problem of distortion. EDSR model improves the resolution of the image, but the contrast is low, and the image distortion is still not ideal. Compared with the existing methods, the proposed method can not only effectively deal with the problem of image blur, but also significantly improve the image distortion, which proves the effectiveness of the proposed reconstruction algorithm.

From the evaluation results, the PSNR value of ROBOST model is the smallest, which indicates that the image distortion is the most serious. The SSIM value of EDSR model is larger than that of FSRCNN model, but the increase is smaller. From the point of view of each index, this method achieves the best results in the quality of reconstructed image and processing time, and has obvious advantages.

In order to further test the image processing level of this method, the field experiment is carried out after the public data set experiment. The reconstruction effect of traditional ROBOST model in the public data set experiment is obviously lower than other reconstruction methods based on deep learning, so the ROBOST model is no longer selected in the field experiment. The experimental results are shown in figure 4, from which we can see that the proposed method can extract the characteristics of objects of different scales and reconstruct the effectiveness. From the perspective of subjective visual effect, it has ideal results. For example, in the comparison of the first row, the reconstructed image of this method has clear pedestrian contour, fine texture and good overall reconstruction effect. VDSR model [16] and EDSR model are also excellent in overall effect, but the restoration degree of local details is not high, and obvious ringing phenomenon appears. The overall effect of SRCNN model is not ideal, and the distortion can not be improved effectively.

Figure 4 Results of UAV image reconstruction.
Table 1  Comparison of test results of different methods.

| Test image | SRCNN    | VDSR    | EDSR    | Proposed |
|------------|----------|---------|---------|----------|
| Road sign  | 0.2588   | 0.2102  | 0.2033  | 0.1996   |
| (BM, GMG, LS) | 1746545  | 1998743 | 2050137 | 2133889  |
|            | 9236841  | 11564791| 1203219 | 1307831  |
| Road sign  | 0.2321   | 0.2006  | 0.2156  | 0.1933   |
| (BM, GMG, LS) | 2174652  | 2422315 | 1899376 | 2063721  |
| Canopy     | 0.2136   | 0.1903  | 0.1856  | 0.1987   |
| (BM, GMG, LS) | 2374665  | 2156434 | 2214151 | 2133209  |
| Building   | 0.2136   | 0.1903  | 0.1856  | 0.1987   |
| (BM, GMG, LS) | 1305479  | 15675391| 1632879 | 1736176  |

From the evaluation results, as shown in the test results of UAV image reconstruction in the field in table 1, it can be seen that the BM value of this method is smaller than the other three algorithms, and the GMG value and LS value of this method are obviously higher. The smaller the BM value is, the clearer the image is. The higher the GMG and LS values are, the higher the image quality is. It can be seen that our method can improve the image reconstruction quality.

3. Conclusion

Based on the idea of deep learning, we propose a super-resolution reconstruction algorithm based on the improved residual learning structure, which can extract and reconstruct the features from the blurred photos taken by UAV under unfavorable working conditions. We select the historical site in the public data set as an example. Compared with FSRCNN model, the PSNR and SSIM values of the reconstructed image are improved by 2.32dB and 0.0217, respectively. Compared with EDSR model, the PSNR value and SSIM value of the reconstructed image are improved by 0.79dB and 0.0076. In the field experiment, compared with other algorithms, the BM value of our algorithm is smaller. At the same time, GMG and LS values of our algorithm are evidently higher. It can be seen that our algorithm has a significant improvement on the extraction of local features of the image, the edge information has been greatly preserved, and the subjective visual effect of the overall image reconstruction is better.

Due to the limited selection of images in the data set, the applicability of our proposed super-resolution reconstruction algorithm in other fields needs to be further tested in order to improve the universality and better reconstruction effect of the algorithm.

References

[1] Chen G, Wang T and Zhou H 2008(05) Journal of Image Graphics (A new method of fog image restoration based on physical model) pp 888-93
[2] Yan H, Yan W and Li W 2010,36(15) Computer Engineering (Image restoration based on Lucy-Richardson algorithm) pp 204-5 + p 210
[3] Zhao L, Jin W, Chen Y and Su B 2008(09) Acta Optica Sinica (Blind restoration of defocused blurred image based on differential image autocorrelation) pp 1703-9
[4] Fang H, Zhang Q and Wei S 2008(08) Journal of Computer Research and Development (Image reconstruction method based on sub Gaussian random projection) pp 1402-7
[5] Li Y, Zhang L, Zhang Z and Zhao F 2011, 28(05) Computer Simulation (An improved motion blurred image restoration algorithm) pp 254-7
[6] Hu C, Zhan S and Wu C 2017, 43(05) Acta Automatica Sinica (Image super-resolution reconstruction based on deep feature learning) pp 814-21
[7] Xiao J, Liu E, Zhu L and Lei J 2017, 37(03) Acta Optica Sinica (Improved image super resolution algorithm based on convolutional neural network) pp 103-11
[8] Jiang Z, Cheng X, Peng Y, Wang J and Li Z 2017, 43(07) Electronic and Applied Technology (Research on UAV recognition algorithm based on deep learning) pp 84-7
[9] Li X and Niu H N 2020, 042(003) Computer Engineering and Science (Facial expression recognition based on feature fusion of vgg-net) pp 500-9
[10] Chen Y and Chen J 2014, 125(13) Optik - International Journal for Light and Electron Optics (Spatial adaptive regularized MAP reconstruction for LD-based night vision) pp 3162-5

[11] Chen Y, Yang W, Tan H, Yang Y, Hao N and Yang K 2016, 127(2) Optik - International Journal for Light and Electron Optics(Image enhancement for LD based imaging in turbid water) pp 517-21

[12] Chen Y, Niu K, Zeng Z and Pan Y 2020 A Wavelet Based Deep Learning Method for Underwater Image Super Resolution Reconstruction (IEEE Access vol 8) pp 117759-69

[13] Dong C, Loy C, He K, and Tang X 2014 European Conf. on Computer Vision, Springer, Cham (Learning a Deep Convolutional Network for Image Super-Resolution) pp 184-99

[14] Chao D, Chen C L and Tang X 2016 European Conf. on Computer Vision. Springer, Cham (Accelerating the Super-Resolution Convolutional Neural Network)

[15] Lim B, Son S, Kim H, Nah S and Lee K 2017 IEEE on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI (Enhanced Deep Residual Networks for Single Image Super-Resolution)

[16] Kim J, Lee J, and Lee K 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)(Accurate Image Super-Resolution Using Very Deep Convolutional Networks)