Research on UAV Fixed Area Inspection based on Image Reconstruction

Kun Cao [a], Fei Wu [b]

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201600, China.

[a]ck1049053722@163.com,[b]yzk2015@yeah.net

Abstract

In order to reduce the amount of data in the acquisition and transmission of drone aerial video and further extend the UAV battery utilization rate, this article proposes a CsRGAN model (Color-super-Resolution Generative Adversarial Network) for aerial missions in fixed areas such as campus inspections. The low-resolution grayscale image is transmitted at the sky end by compression and the high-definition color video image is restored at the ground end by the model. Experimental results can prove that the method can greatly reduce the amount of data transmitted under the same video quality, and maximize the utilization of UAV battery.

Keywords

UVA, aerial photography, fixed Areas, GAN.

1. Introduction

Super-resolution reconstruction is an important technique to reduce the amount of data in image processing[1]. The solution is usually nonlinear, and its solution is often uncertain. For example, a low resolution image to be reconstructed may correspond to a number of high resolution images.

In order to get the optimal solution of high resolution image, the necessary prior information is usually introduced to constrain the process of image reconstruction. According to different solutions and technical means, image super resolution reconstruction methods mainly include.According to different solution ideas and techniques, image super-resolution reconstruction methods mainly include 1) interpolation-based methods. For example, Han et al. [1] proposed an edge-adaptive interpolation reconstruction method based on anisotropic Gaussian filters. This method is more advantageous in removing noise and the step chirp near the edge; Zhang and Wu[2] proposed a windowed normal Kriging interpolation reconstruction method, which can better protect the edges of the image.The interpolation based method does not take into account the content of the image, but simply calculates the pixels between pixels, which can easily lead to serious blur and lose a lot of details. Freeman et al. [3] proposed a learning based approach to model the spatial relationship between high and low resolution image blocks by Markov networks. On this basis, Gao et al. [4] simultaneously trains two projection matrices to map feature spaces of different resolution images into special subspace; Bevilacqua et al. [5] Based on neighborhood embedding, the robustness of the algorithm is enhanced by constraining the weights between neighborhood image blocks and using the factorization of semi non negative matrices to enhance the robustness of the algorithm. Kim et al. [6] is an image reconstruction method for structural analysis. The method selects the high frequency block for the reconstruction and improves the image based on the sharpness of the image block. The quality of the reconstruction of the edge.

Color reconstruction of grayscale image is the process of adding RGB value to grayscale image [7].At present, the main method of coloring grayscale images is mainly 1) the color diffusion method based on manual strokes[8] is based on the hypothesis that "if the brightness of adjacent pixels in space is similar, the color is generally similar". Levin et al.[8] put forward the optimal color diffusion method of energy equation, and its efficiency is relatively low. Yatziv et al.[9] put forward a similar quick image and video coloring method. Though the coloring speed has been improved, its input requirements are rather severe.2）Welsh et al.[10] propose a color transfer method based on the reference color image, that is, using the brightness and texture information, the pixels in the color image are locally matched with the pixels in the target gray image, and then the color is transferred to the most similar pixels. The method is time-consuming and can not achieve satisfactory results.

In this paper, the super resolution enhancement and gray color technology are fused, and the CsRGAN model is proposed to restore the low resolution gray image of the sky end into color HD video image.

2. Proposed Algorithm

2.1 CsRGAN Model

In this model, the gray image reduces the size of the image through 3 encoders and improves the number of channels. The purpose of this is to express the image with a smaller and high dimensional matrix, and to extract the features through 6 residual networks, and use the residual net to increase the depth of the whole network. It will not cause the convergence of gradient dispersion.Then, the decoder is deconvolution to restore it to 256 x 256 matrix, and to enlarge the matrix while maintaining its dimension.Finally, it is amplified by sub-pixel convolution. This method magnify the image on the basis of all the details of the image, so that the image will not be distorted or deformed.Then the grayscale image after the super-resolution is converted into three channels of the same gray value image, and then a series of encoders constantly reduce the size of the image and increase the number of channels. The purpose of this is to represent the picture with a smaller, high-dimensional matrix, and then compress it to a 512 dimensional pixel. Point to represent the whole image. Finally, the decoder reconstructs this high-dimensional pixel into a random color image using a deconvolution method. In the process, the entire network does not change too much about the overall structure of the image.

However, the deconvolution of a single network can easily lead to the overfitting state of the network, especially when the network has more than 6 layers of convolution, the image features contain a large number of high dimensional features and ignore too much low layer feature information.Therefore, this paper uses the method of Image translation, drawing on the idea of U-net, using jump layer connection, and sharing a lot of low level information between the input and output of each layer, so as to ensure that the low layer feature information will not be missing when the deconvolution is in the decoder.Specifically, the network adds the hop layer connection between each i layer and the N-1 layer, where n is the total number of layers. Each hop layer connection simply connects all channels to all channels in the i layer.

In this model, the discriminant network introduces the idea of self encoder. First, the image and the real picture are generated by convolution, and a matrix of 128, 256 * 256 dimensions is formed at the same time, and then the two are linked together.The purpose of this is because the size of the picture is too large to be sent directly into the decoding module, which makes the last layer too high, resulting in overfitting, so convolution reduces the size and dimension of the picture first, and then links it together.

After the comparison of the network, the production network is adjusted according to the loss function, and the above process is iterated continuously. The generated network can remember the corresponding structure of each image feature and complete the reconstruction of the high definition color image.

2.2 Loss Function

This article uses the generation of antagonism network to complete the reconstruction of the image, so the loss function generated by [11] is used to generate the antagonism of the network prototype, as shown in formula (2-1):

$$\mathrm{V}(D,G)=\min_{G}\max_{D}V(D,G)=E_{x\sim p_{data}(x)}[\log D(x)]+E_{z\sim p_{z}(z)}[\log(1-D(G(z)))] \tag{1}$$

The whole generation network is composed of two parts, super resolution reconstruction and color reconstruction, so we introduce two different losses G_{g1}、 G_{g2} .

$$G_{g1} = \arg\min \frac{1}{N}\sum_{n=1}^{N} l_{CR}(G, T_1) \tag{2}$$

$$l_{CR} = l_{SSIM}^{CR} + l_{VGG19}^{CR} \tag{3}$$

Magnifying part: the loss function of the magnifying part is shown in formula 2-2 and 2-3, in which the structure similarity of the image should be considered as the magnifying part is only for the gray level image, so this paper uses SSIM to replace the MSE. In addition, the trained VGG19 is still used as one of the network loss functions.

$G_{g2} = l_{L1}^{CR} + l_{HSV}^{CR}$ (4)

$$l_{L1}^{CR} = avg\|T - G\| \tag{5}$$

$$l_{HSV}^{CR} = avg\left[0.6\|T_H - G_H\|_{MSE} + 0.2\|T_S - G_S\| + 0.2\|T_V - G_V\|\right] \tag{6}$$

Coloring part: the overall loss function is shown in formula 2-4. In addition to the L1 norm, the value of HSV is added as the loss function of the second part. See formulas 2-5 and 2-6.

Finally, as shown in formula 2-7, the L1 norm is added, and the distribution alignment of the HSV values of the generated picture and the real picture is compared, and the similarity between the two pictures is considered from the whole and the two parts of the color.

$$G = G^* + \lambda\left[l_{L1}(G) + l_{HSV}(G)\right] \tag{7}$$

Where,

$$l_{L1}(G) = E_{T,Z\sim p_{data}}\left[\|T - G(Z)\|_1\right] \tag{8}$$

$$l_{HSV}(G) = E_{T,Z\sim p_{data}}\left[\|T_{HSV} - G_{HSV}\|_{MSE}\right] \tag{9}$$

3. Experiment and Analysis

Experimental environment in this paper：The Ubuntu16.04 platform is implemented by Python Programming with a Tensorflow1.20 framework. The processor is Intel Core i7-6300HQ,2.9 GHz,16 kernel CPU, memory 64G, graphics card as GTX1080, and 8G.

In experiment, we only consider the luminance channel of YCrCb color space, so c=1 is in the first / last layer. The two color channel is two-color, just for display, not for training and testing. Note that our method can be extended to direct training of color images by setting c=3. In this paper, we mainly use c=1 to compare with previous methods, because most of them only concern about luminance channels.

In order to avoid boundary effects during training, all convolution layers are not filled, and the network produces smaller output. In processing test images, convolutional neural networks can be applied to images of any size. During the test, all the convolutional layers are filled with enough zeroes, so the output image size is the same as the input size. In order to solve the boundary effect, in each convolutional layer, the output of each pixel (before ReLU) is normalized by the number of effective input pixels, which can be calculated in advance.

In order to solve the boundary effect, in each convolution layer, the output of each pixel (before ReLU) is normalized by the number of effective input pixels, which can be calculated in advance.

The filter weights of each layer are randomly selected from the random Gauss distribution, and the zero mean and standard deviation are 0.001. The learning rate is 10-4 of the first two levels, and the last level is 10-5. From experimental experience, it is found that a smaller learning rate in the last layer is very important for network convergence. The specific results are shown in fig.1 and fig.2.

Input Output Target

Fig. 1 sequence5

Input Output Target

Fig. 1 sequence18

The objective evaluation index of sequence 5 and sequence 18 is shown in Table 1.

Table 1 Performance Index

Index sequence	5	18
PSNR（dB）	28.617190	21.758541
NC	0.997791	0.997791
SSIM	0.981915	0.981915

4. Conclusion

This paper provides a novel deep learning method for image color and super resolution reconstruction. The experimental results show that CRGAN learns end to end mapping between color, low resolution and high resolution images, and there is little extra preprocessing / post-processing besides optimization. Through a lightweight structure, CRGAN has better performance. By exploring more hidden layers / filters in the network and different training strategies, we can get better and better performance. In addition, the structure has the advantage of simplicity and robustness, and it can be applied to other low-level visual problems, such as image blur or demising.

References

[1] Han J W, Suryanto, Kim J H, et al. New edge-adaptive image interpolation using anisotropic Gaussian filters[J]. Digital Signal Processing, 2013, 23(1):110-117.

[2] Zhang Q, Wu J. Image super-resolution using windowed ordinary Kriging interpolation[J]. Optics Communications, 2015, 336(336):140-145.

[3] Freeman W T, Jones T R, Pasztor E C. Example-Based Super-Resolution[J]. Computer Graphics &Applications IEEE, 2002, 22(2):56-65.

[4] Zhang K, Gao X, Tao D, et al. Single image super-resolution with non-local means and steering kernel regression.[J]. IEEE Transactions on Image Processing, 2012, 21(11):4544-56.

[5] Bevilacqua M, Roumy A, Guillemot C, et al. Neighbor embedding based single-image super-resolution using Semi-Nonnegative Matrix Factorization[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2012:1289-1292.
[6] Kim C, Choi K, Ra J B. Example-Based Super-Resolution via Structure Analysis of Patches[J]. IEEE Signal Processing Letters, 2013, 20(4):407-410.
[7] Chang Y, Saito S, Nakajima M. Example-based color transformation of image and video using basic color categories[J]. IEEE Trans on Image Processing,2007,16(2) : 329-336.
[8] Levin A, Lischinski D, Weiss Y. Colorization using optimization[J]. ACM Trans on Graphics,2004,23(3) : 689-694.
[9] Yatziv L, Sapiro G. Fast image and video colorization using chrominance blending[J]. IEEE Trans on Image Processing,2006,15(5) : 1120-1129.
[10] Welsh T, Ashikhmin M, Mueller K. Transferring color to greyscale images[J]. ACM Trans on Graphics,2002,21(3) : 277-280.
[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. arXiv preprint arXiv:1406.2661, 2014.Z.W. Zhang, J.N. Wang: Crane Design Manual (China Railway Press, China 1998), p.683-685. (In Chinese)
[12] C. Li, W.Q. Yin, X.B. Feng, et al. Brushless DC motor stepless speed regulation system based on fuzzy adaptive PI controller, Journal of Mechanical & Electrical Engineering, vol. 29 (2012), 49-52.
[13] China National Standardization Management Committee. Specifications of Crane Design (China Standardization Press, China 2008), p. 16-19.
[14] J. Liu, E.L. Chen and Z.T. He: Journal of Shi Jia Zhuang Railway Institute (Natural Science), Vol. 22 (2009) No. 4, p.40-42.
[15] Q. D. Zeng, Q. E. Li: Progress in Civil Engineering, Vol. 32 (2012) No. 9, p. 3077-3080.