Super Resolution Reconstruction Model Based on Attention Mechanism and Generative Adversarial Network

Xia Miao
School of Electronic Information and Automation, Tianjin University of Science and Technology, 1038 Dagu South Road, Tianjin, China

Kunzhi Yang
School of Electronic Information and Automation, Tianjin University of Science and Technology, 1038 Dagu South Road, Tianjin, China

Xiaoyan Chen
School of Electronic Information and Automation, Tianjin University of Science and Technology, 1038 Dagu South Road, Tianjin, China

E-mail: miaoxia@mail.tust.edu.cn, cxywxr@tust.edu.cn
www.tust.edu.cn

Abstract

The purpose of image super resolution reconstruction is to recover high frequency details from low resolution images containing little information, so as to improve the visual effect of images. Based on the traditional SRGAN algorithm, this paper by importing the residual network structure and combining the channel attention mechanism constructs the basic network module, and this method is named CA-SRGAN. The proposed CA-SRGAN replace the base block of SRGAN with a residual dense block based on the channel attention mechanism. Test in Set5, Set14 and BSD100 data sets, the results of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of CA-SRGAN algorithm are better than those of the traditional SRGAN algorithm, and the details of the reconstructed image are clearer. The whole show better robustness and comprehensive performance.

Keywords: Super resolution reconstruction, Attention mechanism, Generative adversarial network

1. Introduction

With the rapid development of computer technology and the popularization of intelligent equipment, people's requirements for image clarity and resolution are constantly improving. However, in the real world scene, the image information obtained by the acquisition equipment often contains noises results in blur, phantom, sharpening. Sometimes this kind of image information cannot be directly applied to actual work, and cannot better complete the work of target recognition and detection, which increases the difficulty and cost of the work.

With the penetration of smart phones and mobile Internet in daily life, the cost of small high-resolution screens is decreasing year by year, and the computing power of miniaturized devices is also increasing rapidly. Benefiting from such external environment and the rapid development of deep learning, image super resolution reconstruction technology has been applied in many fields[1], such as security monitoring[2], satellite imaging, medical imaging, and mobile terminal image transmission. Image Super Resolution (SR) technology can recover detailed information. The improved reconstructed image by SR can be convenient for target recognition, target tracking and other work. Due to the fitting ability of superior neural network, image reconstruction method based on deep learning can extract the feature information of image, and generate images with richer texture details compared with traditional image reconstruction methods. The main purpose of this paper is to apply deep learning theory and method to, so as to obtain higher quality images.

2. Image super resolution reconstruction and generation of adversarial networks

2.1. *Image super resolution reconstruction*

Digital images stored in a computer usually contain attributes such as brightness, color space, resolution, etc. Image resolution reflects how many pixels an image has within a certain range, which determines how fine the detail of the image is. Generally, the higher the resolution, the more pixels are included in the image and the better the visual effect the image will present.

The optical signal in nature is continuous before entering the imaging system. In the process of being captured and entering the imaging equipment, it may go through various kinds of noise superposition, and finally get a digital image after discrete sampling. In this process, there may be some problems, such as imperfect transmission medium, noise interference, the object or the shooting equipment cannot achieve absolute stillness, and the degradation of the image is bound to be caused by the equipment performance, storage space limitation and other reasons, so that the output of the image is relatively low resolution. The process of image super resolution is exactly the opposite of the above imaging process, aiming to recover a higher resolution image with more detailed information from the degraded low resolution image.

2.2. *The adversarial generation network*

Inspired by the two-person zero-sum game in game theory, adversarial generative network was first proposed by Goodfellow et al. in 2014[3]. With the rapid development of adversarial generative network in theory and model, various improved variants keep emerging, which are not limited to the original research field of computer vision, but have extended and applied in natural language processing, human-computer interaction and other fields.

Adversary to generate network contains two models, which are generator and discriminator. The generator is used to capture a distribution network and generate a noisy image. A discriminator is a network that determines whether a picture is generated image or a real image. For the generator, the generated image tries to deceive the discriminator so that the discriminator cannot tell whether the source of the input image is a generator or real data[4]. For the discriminator, it can make use of prior knowledge in the training process to continuously acquire more powerful discriminant ability. The two models form competition and confrontation, promote each other, and finally reach a state of equilibrium maintained at a higher level.

Ledig et al.[5] first used the idea of adversarial generative network to solve the problem of image super resolution reconstruction and proposed the SRGAN model. Because of the particularity of the image hyper division problem, the input of the generator in SRGAN is the low-resolution image and the output is the super-resolution reconstructed image[6].

3. Super resolution reconstruction algorithm combined with channel attention mechanism

Based on SRGAN, we proposed an improved network named CA-SRGAN. In this paper, the residual network structure is introduced first, and the basic network module is constructed by combining the channel attention mechanism. Secondly, three key components of SRGAN, network architecture, loss resistance and loss perception, are studied and improved. This article then describes the experimental study environment, and the analyzes of the experimental results is introduced.

According to the different functions of the channels in the training process, the channel attention mechanism uses an adaptive way to assign different weights to each channel. The attention mechanism introduced into the generative network can make the network focus on facial features during the training process. In general, if an image has C channels, low resolution image (LR) can be represented as the real valued tensor of size $C \times W \times H$, high resolution image (HR) and super resolution image (SR) can be represented as the real valued tensor of $rH \times rW \times C$. Where H is the height of the image, W is the width of the image, is the number of channels in the image, C is the number of image channels and r is the down-sampling factor.

First, the backbone of the generated network is improved using residual channel attention blocks (RCAB) based on the channel attention mechanism to replace the base blocks[7] of SRGAN. RCAB is composed of residual channel attention blocks (RB) and channel attention mechanisms (CA). Each of these residual in residual (RIR) groups contains g residual blocks (RG) and long skip connections (LSC) each residual block contains b RCABs.

In the field of super resolution technology and image deblurring technology, it has been proved that the network performance and generalization ability will be enhanced when BN layer is removed, and the computational complexity of the network model will also be reduced. The generated network in this paper also enhances the generalization ability of the network model by removing the BN layer. By considering the dependency between channels, an attention mechanism for adaptive re-estimation of channel characteristics is introduced. The reconstructed face image has a more natural overall contour and local regional features.

The experimental results show that using this residual structure can train very deep CNN without the phenomenon of gradient disappearing, and the high-frequency information extracted by feature will not missing, which can better reconstruct the low-resolution image and improve the performance of the whole network model. The long jump connection can bypass the rich low frequency information and process the high frequency information, so as to improve the reconstruction effect. The overall network model structure is shown in Figure 1.

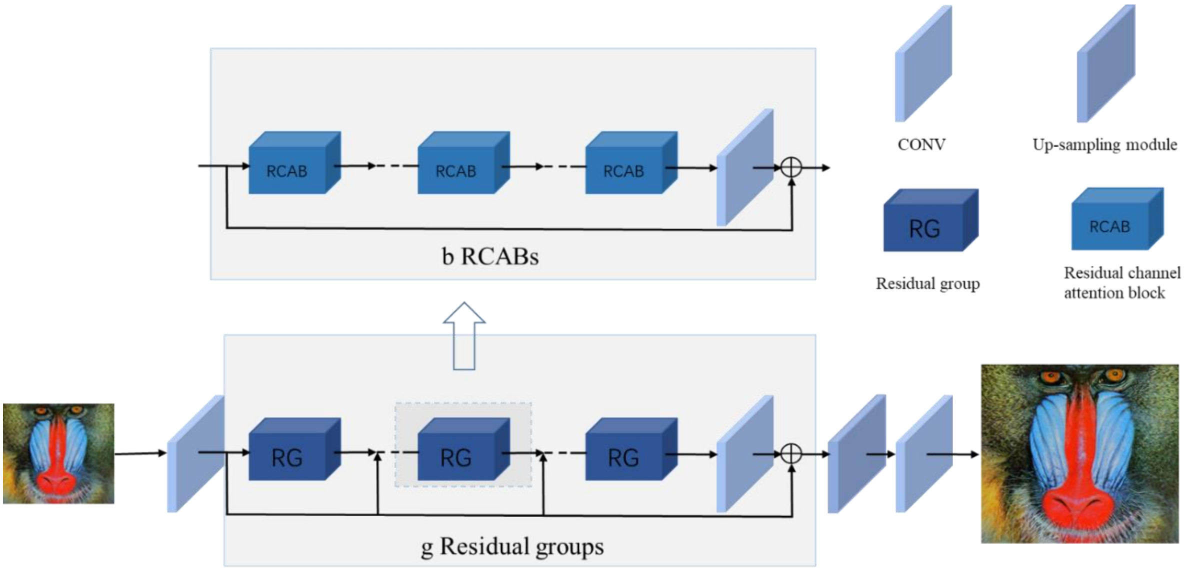

Fig. 1 Generation network improvements

The formula for Group i RGS is shown in Eq. (1)

$$F_i = H_i(F_{i-1}) = H_i(H_{i-1}(\ldots H_1(F_0)\ldots)) \tag{1}$$

The RCAB block for group i can be expressed as Eq. (2)

$$F_{i,b} = H_{i,b}(H_{i,b-1}(\ldots H_{i,1}(F_{i-1})\ldots)) \tag{2}$$

Where $F_{i,b}$ and $F_{i,b-1}$ are the input and output of the b RCAB in group i, respectively. The structure of RCAB is expressed as Figure 2,

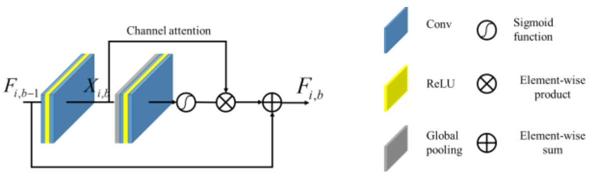

Fig. 2 The structure of RCAB

4. Experiment and Analysis

4.1. *Hardware Conditions and experiment set*

The training environment is Ubuntu 18.04 operating system, CPU is Intel Xeon E5-2360, graphics card is two NVIDIA GE-FORCE GTX1080Ti 11 GB, development language is Python, Pytorch framework. DIV2k data set was used for the training set of the network model, and SET5, SET14 and BSD100 were used for the test set.

4.2. *Experimental evaluation*

In order to further verify the detection performance of proposed CA-SRGAN, SRCNN, VDSR and SRGAN were used in this paper to compare the detection performance with CA-SRGAN proposed in this paper. The test data sets were Set5, Set14 and BSD100. The results are shown in the Table 1.

Table 1 Experimental results

	Set5		Set14		BSD100	
	PNSR	SSIM	PNSR	SSIM	PNSR	SSIM
SRCNN[8]	24.987	0.650	23.243	0.529	23.231	0.559
VDSR[9]	26.054	0.698	24.984	0.636	24.235	0.613
SRGAN[5]	27.971	0.740	25.318	0.682	24.237	0.631
CA-SRGAN	28.662	0.860	26.356	0.744	25.795	0.698

It can be seen from the table that compared with the original SRGAN model, the PSNR and SSIM of the improved CA-SRGAN model on SET5 data set are increased by 0.691 and 0.12 respectively. In SET14 data set, PSNR increased by 1.038, SSIM increased by 0.062; On BSD100 data set, PSNR increased by 1.558 and SSIM increased by 0.067. Compared with other models, the index also has a certain degree of improvement. To sum up, CA-SRGAN algorithm is superior to SRGAN algorithm.

5. Conclusion

In this chapter, an image super resolution reconstruction method based on attention mechanism and generative adversarial network is studied. The CA-SRGAN algorithm is compared with several classical super resolution reconstruction algorithms. The experimental results show that the CA-SRGAN algorithm proposed in this chapter has obvious improvement in the objective evaluation index and subjective visual effect, which shows the effectiveness and superiority of the proposed algorithm.

Acknowledgment

This work was supported by Shenzhen Softsz Co. Ltd., China.

References

1. Nasrollahi, K. , & Moeslund, T. B. . (2014). Super-resolution: a comprehensive survey. Machine Vision & Applications, 25(6), 1423-1468. https://sci-hub.st/10.1007/s00138-014-0623-4
2. Xiang, X. , Liu, W. , & Ling, L. . (2014). Low resolution face recognition in surveillance systems. Journal of Computer & Communications, 02(2), 70-77. https://www.scirp.org/journal/PaperInformation
3. Goodfellow, I. J., "Generative Adversarial Networks", 2014. https://arxiv.org/abs/1406.2661
4. Mao, X. , Li, Q. , Xie, H. , Lau, R. , & Smolley, S. P. . (2017). Least squares generative adversarial networks. IEEE. https://sci-hub.st/10.1109/ICCV.2017.304
5. Ledig, C. , Theis, L. , Huszar, F. , Caballero, J. , Cunningham, A. , & Acosta, A. , et al. (2016). Photo-realistic single image super-resolution using a generative adversarial network. IEEE Computer Society. https://arxiv.org/abs/1609.04802
6. Ma, C. , Jiang, Z. , Rao, Y. , Lu, J. , & Zhou, J. . (2020). Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation. IEEE. https://arxiv.org/abs/2003.13063v1
7. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y., "Image Super-Resolution Using Very Deep Residual Channel Attention Networks", 2018. https://arxiv.org/abs/1807.02758
8. C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016, doi: 10.1109/TPAMI.2015.2439281. https://arxiv.org/abs/1501.00092
9. J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Usinrence on Computer Vision and Pattern Recognition (CVg Very Deep Convolutional Networks," 2016 IEEE ConfePR), 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182. https://ieeexplore.ieee.org/document/7780551

Authors Introduction

Ms. Xia Miao

She is currently studying in Tianjin University of Science and Technology, majoring in electronic information, and main research field is computer vision.

Mr. Kunzhi Yang

He received his bachelor's degree in Engineering from Lanzhou University of Technology in 2019. He is currently studying control science and engineering at Tianjin University of Science and Technology.

Prof. Xiaoyan Chen

She, professor of Tianjin University of Science and Technology, graduated from Tianjin University with PH.D (2009), worked as a Post-doctor at Tianjin University (2009.5-2015.5). She had been in RPI, USA with Dr. Johnathon from Sep.2009 to Feb.2010 and in Kent, UK with Yong Yan from Sep-Dec.2012. She has researched electrical impedance tomography technology in monitoring lung ventilation for many years.