Unsupervised Image Super Resolution Reconstruction Based on Generative Adversarial Network

Jiexin Zheng [1], Hezhi Zhang [2]

[1]School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China

[2]School of Science, Chang'an University, Xi'an 710064, China

Abstract

Generative adversarial network is one of the research hotspots in recent years. Compared with traditional convolution network, it can generate rich image details. SinGAN is a single image generation model, which can be applied to the super-resolution reconstruction of images. This paper proposes a new network structure based on SinGAN, which can learn effective information from a single image more fully. In order to ensure the deep supervision, effective gradient flow and feature reuse ability of convolutional neural network, the dense residual structure is introduced into the generator and discriminator of SinGAN to speed up the network convergence. Experimental results show that the proposed method has better performance and higher efficiency in image super-resolution reconstruction.

Keywords

Unsupervised, super resolution, SinGAN.

1. Introduction

With the development of deep learning, generate adversarial net (GAN) has become one of the research hotspots in recent years. Generative adversarial network [1] is a generative model proposed by Goodflow in 2014. It is based on the theory of statistics and probability, and draws lessons from the adversarial idea of game theory [2]. It is an efficient data generation method. At present, generative adversarial network has been successfully applied to a variety of computational vision tasks, such as image super-resolution reconstruction [3-6], image style migration [7] and image enhancement [8-9], and it also shows great development potential.

The initial GAN model has some limitations. In the process of training, it is difficult to ensure the stability and convergence, and it is easy to appear unstable training, gradient disappearance, mode collapse and other phenomena, which lead to the model can not be trained. Radford et al. [10] combined convolutional neural networks (CNN) with GAN and proposed deep convolution general adverse networks (DCGAN). CNN is used to replace the multi-layer perceptron structure in the traditional GAN, and the GAN is improved from the network topology. Finally, only a small number of categories and a large number of repeated images may be generated. In order to solve the problem of gradient disappearance and training instability, Arjovsky et al. [11] proposed WGAN (Wasserstein generative advanced networks), which uses Wasserstein distance to measure the distance between real samples and generated samples. Noise is inevitable in practical visual tasks. In reference [12], we train GAN network on noisy training set to generate noiseless images without pre-determined noise distribution. A more robust generation network is proposed. In reference [13], the reconstruction of invisible domain image in training set using trained GAN network is also tested on StyleGAN and ProGAN, and the realistic reconstruction of invisible image is completed. Zhang et al. [14] introduced the self attention mechanism into GAN network to deal with the long range and multi-level dependence of images, so as to make the generated images more realistic and detailed.

ISSN: 2664-9640

In order to make full use of the residual information and the relationship between channels in SinGAN [15], this paper adds a dense residual structure based on the network structure of SinGAN to enhance the learning ability and feature reuse of the network. Experimental results show that the proposed model has better performance in image super-resolution reconstruction.

2. Related work

2.1. SinGAN

SinGAN is an unconditional generative model which uses a single natural image to train. It can learn the internal data relationship of the input training data to generate high-quality and diverse samples consistent with the visual content of the training image. SinGAN adopts the form of cascaded generator discriminator pair, and forms a pyramid network structure through n GANs. Its structure is shown in Figure 1.

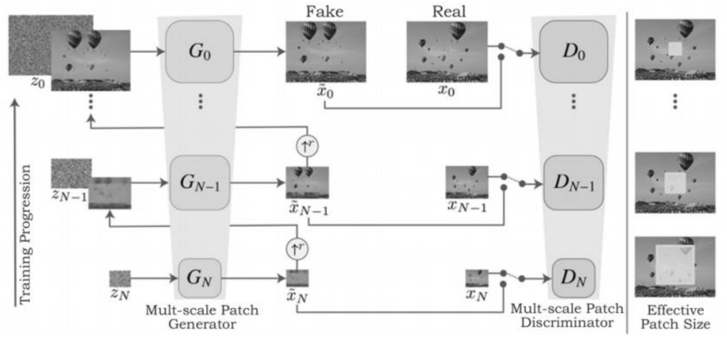

Figure 1 Network structure of SinGAN

From training to testing, it is based on the idea of coarse to fine. From bottom to top, the scale gradually changes from rough to fine, which allows the generation of new sample images with any size and aspect ratio. SinGAN generator (as shown in Figure 2) in the n scale. The output of G_{n+1} is added with the noise of the corresponding size after upper adoption as G_n. The D network is used to judge whether the generated image is true or false. Each level of the discriminator uses a Markov discriminator, which is completely composed of convolution layers. The final output is an $n \times n$ matrix, and the mean value of the output matrix is taken as the judgment result. By learning from a single image, SinGAN can generate images with high quality at any scale.

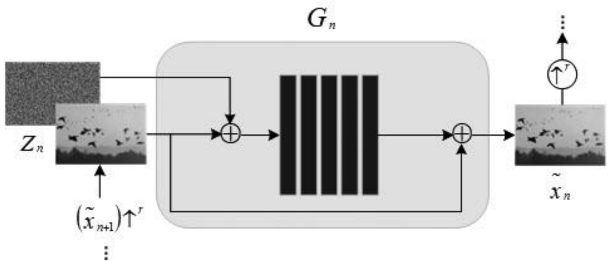

Figure 2 single scale formation process of SinGAN

3. Method

3.1. Overall network framework

At present, most image super-resolution reconstruction algorithms based on deep learning obtain network parameters by learning from external data sets, which brings a large training cost. The invention designs an unsupervised image super-resolution reconstruction algorithm based on the generation countermeasure network. Firstly, through image preprocessing, the

ISSN: 2664-9640

image pyramid is generated from the processed image; then a network framework with pyramid structure is designed, which starts from scale n to scale 1, and judges whether the generated image is true or false through the discriminator. Finally, the image to be processed is input into the generator of trained scale-1 network for super-resolution reconstruction.

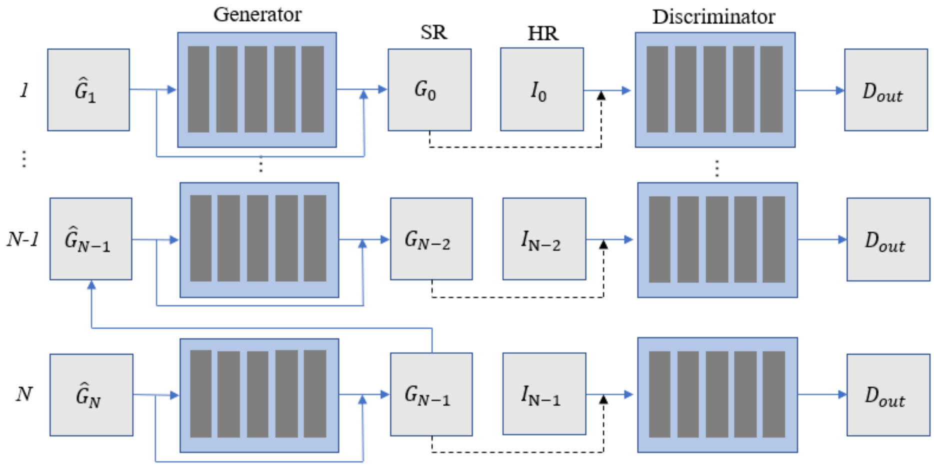

Figure 3 The network framework proposed in this paper

3.2. Loss function

A new loss function is used to optimize the network parameters. Let the output image of the generator be G and the real image be I. In the traditional image quality enhancement network, MSE loss is generally used. Although it is easy to get a high peak signal-to-noise ratio (PSNR), it is easy to cause image blur, resulting in the loss of high-frequency details. Therefore, the invention uses the pre-training VGG network to calculate the L1 distance between the generated image and the high-quality PNG image, which is recorded as l_{vgg}, as follows:

$$l_{vgg} = \frac{1}{WH}\sum_x\sum_y |VGG(G) - VGG(Y)|$$

In order to obtain high peak signal-to-noise ratio (PSNR), MSE loss is calculated as follows:

$$l_{MSE} = \frac{1}{WH}\sum_x\sum_y (G - I)^2$$

The counter loss is used to improve the discriminator's ability to distinguish G from I. When the discriminator can't distinguish G from I, the quality of the image generated by the generator is equivalent to that of I. The resistance loss is recorded as l_{adv}, as follows:

$$l_{adv} = \sum_n -\log\big(D(G)\big)$$

Where D is the discriminator.

To sum up, the loss function l proposed by the invention is as follows:

$$l = \alpha \cdot l_{vgg} + \beta \cdot l_{MSE} + \gamma \cdot l_{adv}$$

Where α, β and γ denote l_{vgg}, l_{MSE} and l_{adv}. The super parameter of is used to adjust l_{vgg}, l_{MSE} and l_{adv} in each group.

4. Experiment

4.1. Training

Generate pyramid network and parameter initialization. Suppose that the network has n scales, the composition of the network is as follows:

ISSN: 2664-9640

Table 1 Pyramid network

scale	Generator			Discriminator		
	input	output	conv	input	output	conv
1	$\hat{G}_1$	G_0	m+N	I_0	G_0	m+N
...	...	...	...	...	...	...
N-1	$\hat{G}_{N-1}$	G_{N-2}	m+1	I_{N-2}	G_{N-2}	m+1
N	$\hat{G}_N$	G_{N-1}	m	I_{N-1}	G_{N-1}	m

The network is trained from the nth scale. First, N generates $\hat{G}_1$ by up sampling I_N, $\hat{G}_N$ and G_{N-1}. The generator of the nth scale is used to generate G_{N-1}. G_{N-1} is input to the discriminator network as the input of the discriminator, and the network of the discriminator extracts the input image G_{N-1} and then the real image I_{N-1} is to input discriminator network, the parameters of the whole network are optimized by maximizing the loss of the discriminator.

Figure 4 Pyramid image in this example

4.2. Experimental analysis

SinGAN is a multitask network, which can be applied to image super-resolution reconstruction. The improved model based on SinGAN has better performance in super-resolution reconstruction. Compared with the existing deep learning model, the sample size and training times of the proposed network in the process of single image super-resolution reconstruction are only 1, and the pyramid model is used to achieve super-resolution reconstruction, which saves the training time and cost.

Set5 data set is a common data set in the field of super-resolution reconstruction. In this paper, two times and four times magnification experiments are carried out on set5, and PSNR and SSIM indexes are used to evaluate the image quality after reconstruction. The results are shown in the table below.

Table 2 Performance of SinGAN and our model on set5

method	scale	image	PSNR	SSIM
SinGAN	2	baby	27.421	0.843
		bird	30.153	0.936
		butterfly	24.792	0.921
		head	31.337	0.819
		woman	27.228	0.896
	4	baby	22.576	0.711

		bird	27.956	0.796
		butterfly	21.623	0.810
		head	27.825	0.687
		woman	24.142	0.813
Ours	2	baby	28.104	0.857
		bird	32.081	0.951
		butterfly	27.142	0.951
		head	31.917	0.837
		woman	29.824	0.923
	4	baby	25.161	0.746
		bird	29.976	0.817
		butterfly	22.876	0.811
		head	28.445	0.717
		woman	25.814	0.830

Experimental results show that the performance of the proposed model on set5 is better than that of SinGAN. PSNR and SSIM are commonly used image quality evaluation indexes. It can be seen from table 2 that the model proposed in this paper is better than SinGAN in these two indexes.

5. Conclusion

At present, the deep learning based super-resolution reconstruction network is trained based on external data sets, which requires much computing resources and cost. SinGAN is an unconditionally generated network based on a single image, which can be used for image super-resolution reconstruction. This paper presents a new network model based on SinGAN. The dense residual structure can make the network converge quickly and reduce the training time greatly.

References

[1] GOODFELLOW I J, POUGET-ABADIE J, M-IRZA M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
[2] RATLIFF L J, BURDEN S A, SASTRY S S. Characterization and computation of local nash equilibria in continuous games[C]//2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2013: 917-924
[3] ZHANG S, LIANG G, PAN S, et al. A fast medical image super resolution method based on deep learning network ［J］. IEEE Access, 2018, 7： 12319-12327
[4] Pengju Liu, Hongzhi Zhang, Yue Cao, et. Learning cascaded convolutional networks for blind single image super-resolution[J]. Neurocomputing,2020,417.
[5] Xu Jian,Gao Yan,Xing Jun,Fan Jiulun,Gao Qiannan,Tang Shaojie. Two-direction self-learning super-resolution propagation based on neighbor embedding[J]. Signal Processing,2021,183.
[6] Helal Uddin Mullah, Bhabesh Deka, A.V.V. Prasad. Fast multi-spectral image super-resolution via sparse representation[J]. IET Image Processing,2020,14(12).
[7] Biagiotti Giacomo,Purić Edvin,Urbančič Iztok,Krišelj Ana,Weiss Matjaž,Mravljak Janez,Gellini Cristina,Lay Luigi,Chiodo Fabrizio,Anderluh Marko,Cicchi Stefano,Richichi Barbara. Combining cross-coupling reaction and Knoevenagel condensation in the synthesis of glyco-BODIPY probes for DC-SIGN super-resolution bioimaging[J]. Bioorganic Chemistry,2021,109.

ISSN: 2664-9640

[8] Dankovich Tal M.,Rizzoli Silvio O.. Challenges facing quantitative large-scale optical super-resolution, and some simple solutions[J]. iScience,2021,24(3).
[9] Reitz Sydney J.,Sauerbeck Andrew D.,Kummer Terrance T.. SEQUIN: An imaging and analysis platform for quantification and characterization of synaptic structures in mouse[J]. STAR Protocols,2021,2(1).
[10] ADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. [2020-04-05]. https://arxiv.org/ abs/1511.06434. 10.1109/AIAR. 2018.8769811.
[11] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[EB/OL]. [2020-04].https:// arxiv.org/ abs/1701.07875.
[12] Noise Robust Generative Adversarial Networks, CVPR2020
[13] Image Processing Using Multi-Code GAN Prior
[14] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [EB/OL]. [2020-06-06]. https://arxiv.org/pdf/1805.08318.pdf.
[15] SHAHAM T R, DEKEL T, MICHAELI T. SinGAN: Learning a generative model from a single natural image[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 4570-4580.
[16] Qilong Wang, Banggu Wu, Pengfei Zhu, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. Computer Vision and Pattern Recognition 2020.
[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
[18] Tong T, Li G, Liu X, et al. Image Super-Resolution Using Dense Skip Connections[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017: 4799-4807.