ISSN: 2664-9640

An image super-resolution reconstruction method based on adversarial generative network

Yanli Shi, Xing Gu, Tingting Gui

Changchun University of Science and Technology National Demonstration Center for Experimental Electrical, Changchun, Jilin, 130022, China

Abstract

As a major source of information and a medium for us to perceive the world, images are everywhere in our lives. Image processing technology is used in various fields, such as security monitoring, remote sensing imaging, medical imaging and so on. Due to insufficient device resolution, complex environmental factors, and loss in the image transmission process, the problem of image resolution degradation will all be caused. In order to reconstruct image details and obtain higher-quality clear images, super-resolution reconstruction technology emerged. This paper designs a depth generator structure with recursive attention network for the characteristics of low resolution, insufficient detail information, blurred boundaries, and poor perception quality. The recursive attention module is used to extract high-frequency information from the feature map, suppress useless information, and improve the expression ability of feature extraction, which is conducive to the reconstruction of infrared image texture details. In addition, we have carefully designed a discriminator, which consists of a deep convolutional neural network, which can better reconstruct the image. The experimental results on the infrared image data set show that the proposed method is higher than several typical methods and achieves higher image visual quality.

Keywords

Infrared image; Super-resolution reconstruction; Attention mechanism; Generative adversarial network.

1. Introduction

Due to the lack of a priori knowledge of the original scene and the imaging process, super-resolution reconstruction has become an intractable ill-posed inverse problem of imaging degradation. Image super-resolution reconstruction methods are divided into three types: interpolation-based, reconstruction-based and learning-based. The interpolation-based method is simple and intuitive, and can get the results quickly, but there are problems such as blurred edges after image reconstruction. The method based on reconstruction focuses on restoring the high-frequency information part of the image. This method is simple and low in calculation, but it ignores part of the high-frequency detail information of the image. The learning-based method is to establish the mapping relationship between low-resolution images and high-resolution images through a large amount of training data. Compared with other reconstruction methods, better image reconstruction quality can be obtained.

In order to improve the feature extraction performance and obtain satisfactory high-resolution infrared images, we are introducing the attention mechanism into SRGAN and propose an infrared image super-resolution reconstruction method based on the generative confrontation network. Because of the characteristics of low-resolution infrared images, such as low contrast and poor perceptual quality, we introduced a recursive attention network into the generator to extract more high-frequency features of the image and suppress useless information to enhance

the expressiveness of the features and help For the reconstruction of texture details. In addition, because the visual recognition of infrared images is not as good as natural images, they contain less key semantic information, and the loss function is optimized. One strategy is to use the pre-trained VGG-pre-activated 19 network features to calculate the perceptual loss, which can improve the accuracy of image reconstruction. The other is to use Wasserstein distance to guide GAN's confrontation training and ensure its convergence.

2. Related work

In 2014, Goodfellow and others first proposed the generative confrontation network model GAN, which is mainly composed of generator and discriminator. The generator module can generate sample data from the input data according to certain rules, and then input the generated sample data and the original real data into the discriminator module. The discriminator module is a two-classifier that can distinguish the input Is the generated data or the original real data, and output its probability value. The basic framework is shown in Figure 1. In Figure 1, the generator is represented by a function G, and the discriminator is represented by a function D.

First, a noisez, which was sampled from LR training images, is sent to G to generate a false-samplexthat is similar to HR images that were input into D. At the same time, HR training samples are also sent to D as real samples, and D tries to distinguish the difference between the false-samplex and the real-sample HR. The objective of the generator is to continuously produce high-resolution images that are as close to the real samples as possible until the generated high-resolution images can deceive the discriminator, namely, to make D(G (z)) as close to 1 as possible.

The objective of the discriminator is to make D(G) as close to 0 as possible, and eventually, it will realize a balance in the mutual game. By adding a discriminant loss to the traditional perceptual loss, SRGAN can generate texture details that are closer to those of the HR images, thereby making the images more realistic. However, due to the uniformity among the infrared image pixels, SRGAN still does not reconstruct some texture details sufficiently clearly, and it may produce fuzzy artifacts.

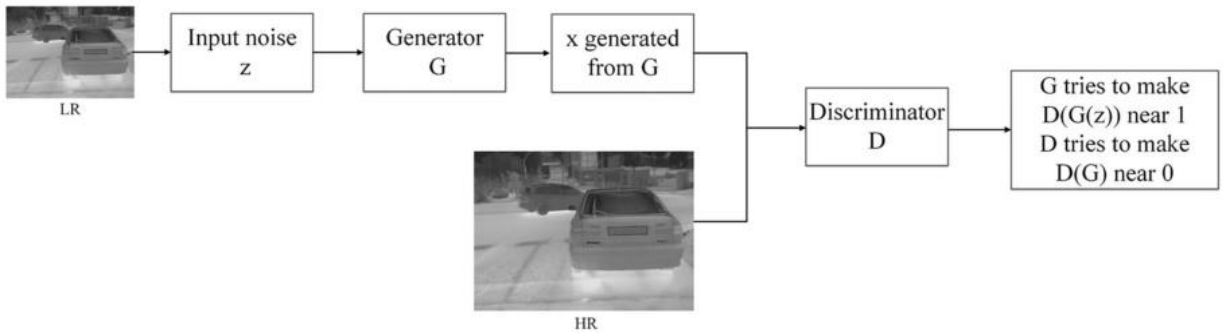

Fig. 1Basic framework of SRGAN

3. Proposed method

Based on the basic framework of SRGAN, we redesigned the generator to generate infrared images with higher visual quality and introduced a recursive attention module. In addition, the loss function is used as an optimization strategy to evaluate the reconstruction effect through objective evaluation indicators and subjective visual quality. The objective of infrared image super-resolution reconstruction is to estimate a high-resolution reconstructed image ISR according to a low-resolution input image ILR such that ISR is as close to the real high-resolution image IHR as possible, which is pursued by training a generator network G to

generate a high-resolution image that is as similar to the real high-resolution image IHR as possible. ISRis expressed as:

$$I_{SR} = G_\theta(I_{LR})$$

Where ILRis obtained from the corresponding high-resolution image IHR with scale down-sampling factor r and θ represents the parameters of the network, which can be obtained through continuous optimization of the loss function in the adversarial training. The parameters must satisfy the following expression:

$$\theta = \arg\min \sum L^{SR}(I_{SR}, I_{HR})$$

where $L^{SR}(I_{SR}, I_{HR})$represents the reconstruction error.

3.1. Design of the generator

Neural network feature mapping ability is very strong, it can map low-resolution feature maps to high-resolution space, which is often used for super-resolution reconstruction tasks. Infrared image pixel gradient range is small, and some weak details are not easy to extract. Therefore, we introduce the recursive attention module into the generator structure. The main function of the recursive attention model is to further extract detailed texture information for image reconstruction and suppress useless low-frequency information. These texture details are important for the generation of high-resolution images. The better the detailed information is extracted, the more accurate the reconstruction result will be. The generator structure is shown in Figure 2.

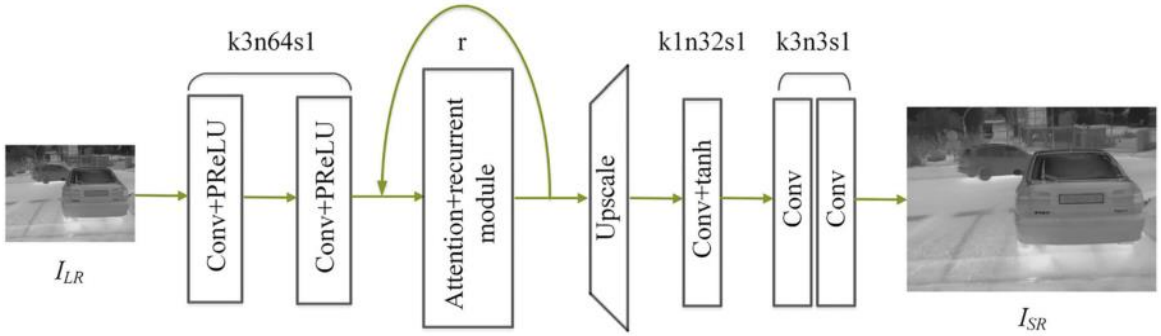

Fig. 2Generator structure: kis the size of the convolution kernel, nis the number of convolution kernels, sis the convolution step, andris the number of recursions

Then, the extracted shallow feature maps are sent into the recursion attention module as new inputs. The attention layer structure is illustrated in detail in Fig.3.

There are abundant low-frequency components and a few valuable high-frequency components in the low-resolution infrared image space. The low-frequency part is flatter, and the high-frequency part is typically full of edges, textures, and other details. In super-resolution tasks, high-frequency channel features are more important for reconstruction; thus, we introduced an attention mechanism to focus on such channel features. We can assign attention resources to each feature channel by recursively calling the attention layer. With the increase in the number of recursion, the trained attention maps can increasingly highlight the detailed texture and related structure.

3.2. Design of the discriminator

In addition to improving the generator, we have also made targeted improvements to the discriminator. The improved discriminator network structure and parameter settings are presented in Fig.4. In the figure, SR is a high-resolution image that is generated by the generator, and HR is the real high- resolution image. The discriminator adopts the deep convolution structure. A study [30] showed that in super-resolution reconstruction tasks, the batch norm

(BN) layer tends to destroy image spatial information and reduce the reconstruction performance; hence, we removed the BN layer and used Leaky_ReLU (λ= 0.2) as the activation function. After the 7-layer convolutions, the feature maps were input into two fully connected layers and classified by the sigmoid activation function to judge the output high-resolution image as true or false.

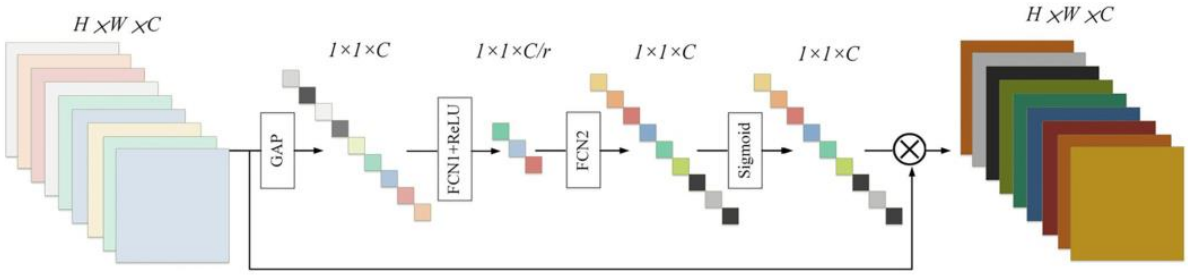

Fig. 3 Schematic diagram of the attention layer

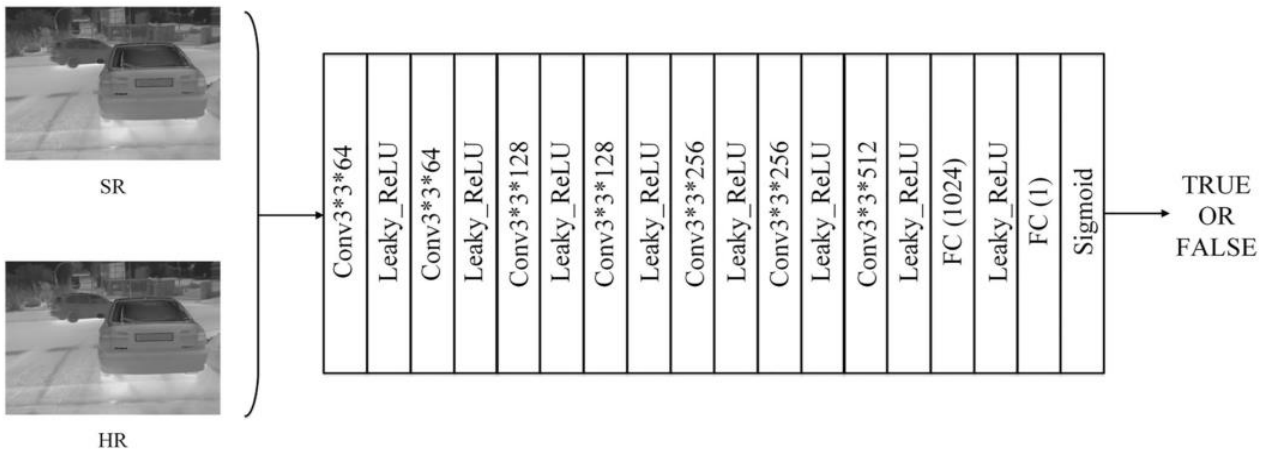

Fig. 4 Discriminator structure and parameter settings

4. Experiment

4.1. Experimental environment and parameter settings

The platform of the experiments is a Windows 10 operating system with an Intel 2.2 GHz i7-8750h CPU with 8 G memory that is configured with an NVIDIA GTX 1080 GPU, and we trained the model under the GPU-based TensorFlow deep learning framework. In the article, the MSRA method , which was proposed in reference , was used to initialize the weights, and the Adam optimizer with a momentum and weight decay of 0.9 and 0.0001, respectively, was used to optimize the network. The batch size was set to 16, and the initial learning rate was 10−4. Since the generator is completely convolutional, it can be applied to images of any size. The training consisted of two parts: First, the generator was pretrained with MSE to avoid the local optimization of the GAN network with direct training. The pre-training learning rate was 10−4for 200 iterations. Then, the generator and the discriminator were alternately trained, and the learning rate was 10−410,000 iterations. At the same time, the Wasserstein distance was used to optimize the adversarial training.

4.2. Datasets

Dataset: We train our network on two widely used super-resolution datasets: DIV2K and Flicker2K. DIV2K includes 800 LR-HR image pairs, and Flicker2K includes 2650 image pairs. The HR patch size is set to 256×256, while the minibatch size is 32. We only flip the image randomly (vertical or horizontal) without any other data enhancement. In order to compare

fairly with other methods, we use four widely used data sets: set5, set14, BSD100 and Urban100. All images are used to evaluate the performance of the algorithm, and the performance of the algorithm in each dataset is calculated separately.

4.3. Evaluation indices

Image quality evaluation is very important for judging performance. The evaluation methods of super-resolution reconstruction algorithms are divided into subjective evaluation and objective evaluation. Subjective evaluation means that the evaluation of image quality is based on people's subjective feelings. SRGAN proposed the Mean Opinion Score (MOS), where a specified number of raters are required to score the results for the evaluation of the image effect of the super-resolution reconstruction method. Because the subjective evaluation is not accurate enough, it is a more efficient image quality evaluation method . However, before the new evaluation method is proposed, there are two commonly used image quality evaluation methods, namely PSNR and SSIM, which are used to objectively evaluate the model. However, in super-resolution reconstruction scenes in real infrared images, only low-resolution images must be reconstructed, but there is no corresponding high-resolution to provide reference images; therefore, it is necessary to introduce methods for reducing quantitative free reference image quality evaluation.

The peak signal-to-noise ratio (PSNR) is an objective image evaluation index that is based on the error between corresponding pixels, which is defined as follows:

$$PSNR = 10 \times \lg \frac{(2^n - 1)^2}{MSE}$$

Where MSE represents the mean square error between images, which is expressed as:

$$MSE = \frac{1}{H \times W} \sum_{x=1}^{H} \sum_{y=1}^{W} [I_{HR}(x,y) - I_{SR}(x,y)]^2$$

The smaller the MSE value is, the larger the PSNR value is and, thus, the closer the image is to the comparison image.

The structural similarity (SSIM) measures the similarity of images from three aspects: brightness, contrast, and structure. Its value ranges from 0 to 1. The larger the value, the lower the distortion of the image.

$$SSIM(x,y) = L(x,y)C(x,y)S(x,y)$$

Where L(x,y), C(x,y) and S(x,y) represent the brightness, contrast, and structure, respectively.

Table 1 Average PSNR(dB)/SSIM values on Test1 and Test2 of the proposed method and five other SR methods

Dataset	Scale	Bicubic PSNR/SSIM	SRCNN PSNR/SSIM	DRCN PSNR/SSIM	SRGAN PSNR/SSIM	ESRGAN PSNR/SSIM	Ours PSNR/SSIM
	2	35.229/0.980	37.823/0.934	38.382/0.937	37.553/0.931	37.653/0.930	**38.689/0.948**
Test1	3	32.008/0.856	33.971/0.909	34.590/**0.919**	33.595/0.899	33.580/0.909	**34.703**/0.912
	4	30.242/0.795	31.503/0.840	32.142/0.867	31.039/0.856	31.033/0.861	**32.230/0.872**
	2	31.280/0.835	32.001/0.845	32.905/0.869	31.824/0.844	32.258/0.855	**33.176/0.874**
Test2	3	28.169/0.811	28.906/0.830	29.988/0.858	29.079/0.831	29.909/0.835	**30.566/0.862**
	4	22.683/0.758	24.067/0.794	25.005/0.821	24.055/0.791	24.066/0.807	**25.697/0.826**

4.4. Results and analysis

Fig. 5result comparison of various methods and the proposed model on Test1-car with an upscaling factor of 3. (a) Ground-truth HR; (b) Bicubic interpolation; (c)SRCNN ;(d) DRCN; (e)SRGAN ;(f)ESRGAN ;and(g)Ours

5. Summary

We present a novel model that uses recursive attention learning that is based on generative adversarial network to super-resolve infrared images. We design a generator network with recursive-attention modules, which can adaptively adjust the feature channel information and enhance the expressiveness of features. The recursive attention learning strategy can not only make the texture of infrared images more natural and realistic but also reduce the generation of pseudo textures. For the loss function, feature values of the VGG-19 network before activation are used to constrain the perceptual loss, which can be better monitored recursively. The Wasserstein distance is also used to optimize adversarial training and to increase the stability of the network training. The experimental results have demonstrated that the proposed model has a significant effect on the super-resolution reconstruction of vehicle infrared images. Compared with those of several advanced models, the objective indices of the reconstructed images have improved significantly, along with the visual texture details However, the proposed method did not perform well on other images, such as natural images and thermal infrared images. Increasing the general performance of our model will be the focus of our next study.

References

[1] Yu X, Ye X, Gao Q (2019) Pipeline image segmentation algorithm and heat loss calculation based on gene-regulated apoptosis mechanism. Int J Press Vessel Pip , 329-336.

[2] Zhang K, Zuo W, Zhang L (2018) Learning a single convolutional super-resolution network for multiple degradations. In Proc IEEE Conf Comput Vis Pattern Recognit (CVPR), 3262-3271.

[3] Mao Q, Sun H, Liu Y, Jia R (2019) Fast and efficient non-contact ball detector for picking robots. IEEE Access 7:175487–175498.

[4] He K, Cao X, Shi Y (2019) Pelvic organ segmentation using distinctive curve guided fully convolutional networks. IEEE Trans Med Imaging 38(2):585–595.

[5] Tian Y, Jia R, Xu S, Hua R, Deng M (2019) Super-resolution reconstruction of remote sensing images based on convolutional neural network. J Appl Remote Sens 13(4):046502.

[6] Liu Q, Jia R, Zhao C, Liu X, Sun H (2020) Face super-resolutio reconstruction based on self-attention residual network. IEE access 8:4110–4121.

[7] Qian R, Tan R, Yang W (2018) Attentive generative adversaria network for raindrop removal from a single image. In Proc IEE Conf Comput Vis Pattern Recognit (CVPR), 2482–2491.

[8] Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In Proc IEEE Conf Comput Vis Pattern Recognit (CVPR), 7132-7141.

[9] Wang X, Yu K, Wu S (2018) ESRGAN: enhanced super-resolution generative adversarial networks. In Proc Euro Conf Comp Vis (ECCV), 63-79.

[10] He K, Zhang X, Ren S (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In IEEE Int Conf Compu Vis (ICCV), 1026–1034.

[11] Kim J, Lee J, Lee K (2016) Deeply-recursive convolutional network for image super—resolution. In Proc IEEE Conf Comput Vis Pattern Recognit (CVPR), 1637-1645.

[12] Ledig C, Theis L, Huszár F (2017) Photo-realistic single image super-resolution using a generative adversarial network. In Proc IEEE Conf Comput Vis Pattern Recognit (CVPR), 105–114