Review of GAN-Based Image Super-Resolution Techniques

Huining Feng

Shenzhen MSU-BIT University, Shenzhen, China

1120210272@smbu.edu.cn

Abstract. The objective of Image Super-Resolution (ISR), a significant area of study in computer vision and image processing, is to produce high-resolution images from low-resolution images. The main objective of this paper is to explore the Image Super-Resolution methods based on Generative Adversarial Networks (GANs), especially SRGAN and ESRGAN. In the experimental part, the performance of SRGAN and ESRGAN will be evaluated by using PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) as evaluation metrics, and the results demonstrate the great potential of Generative Adversarial Networks in Image Super-Resolution tasks, especially in improving the quality of the generated images, and both SRGAN and ESRGAN perform well in recovering image details. Facing the current challenges, researchers can explore new methods and techniques to promote the development and application of image super-resolution technology. The prospect of wide application of GAN models in the field of image processing provides rich opportunities and possibilities for future research and innovation.

Keywords: Image Super-Resolution, Generative Adversarial Networks, Super-Resolution Generative Adversarial Network technique, Enhanced Super-Resolution Generative Adversarial Network technique.

1. Introduction

With the advancement of technology, the quality of images continues to improve, and higher image quality will bring people better visual effects and experience. However, various factors can lead to deterioration of image quality, such as image compression process. Therefore, the deterioration of image quality can be improved by employing super-resolution techniques.

The objective of Image Super-Resolution (ISR), a significant area of study in computer vision and image processing, whose aim is to generate high-resolution images from low-resolution ones. Specifically, it restores and reconstructs as much detail and clarity as possible in an image while enlarging the image size by utilizing computational methods and algorithms. The image super-resolution technique not only visually improves image quality, but also plays an important role in many specialized fields, it can be used in many different fields such as medical imaging, remote sensors [1], surveillance systems and video enhancement.

Although the image super-resolution technique has an important application value, it also faces many challenges. They are mainly reflected in the following aspects:

Estimation of degenerative model: The biggest challenge in image super-resolution restoration is the accurate estimation of the degenerative model. The degenerative model needs to extract motion information, blur information and noise information from low-resolution images. Because of the lack of

^{© 2024} The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

sufficient data, it is very difficult to estimate the degenerative model by relying only on low-resolution images. Most studies focus on designing image priors and ignore the influence of degenerative models on the results. Although many methods perform well under specific conditions, the results on natural images are often unsatisfactory, due to the inconsistency of the degenerative model of the natural image with the hypothetical model. This is one of the reasons why existing algorithms are difficult to be widely used.

Computational complexity and stability: The computational complexity and stability of image superresolution algorithms are the main factors restricting their wide application. The existing algorithms usually improve the reconstruction quality by increasing the amount of computation, especially when the magnification is large, the amount of computation increases exponentially, resulting in long processing time and limited practical applications. Although some methods can generate high-quality images, the detail reconstruction errors may occur when the input images do not conform to the algorithm's assumptions. In addition, the learning-based methods rely only on the generalization ability of the model to predict high-frequency details when the training dataset is incomplete, making it difficult to avoid the introduction of erroneous details. Therefore, improving the computational efficiency and stability of the algorithm is still an important research direction.

Compression degradation and quantization error: Many images super-resolution algorithms ignore the degradation of quality due to image compression. The transmission of images captured by consumergrade cameras and Internet images often compress images, which changes the degenerative model of images. The compressed images contain not only additive noise, but also content-dependent multiplicative noise. In addition, most reconstruction methods are based on continuous imaging models and do not consider the digital quantization errors, which can further affect the degenerative model and thus affect the reconstruction results.

Detail reconstruction and visual realism: A key challenge in the process of image upscaling is how to maintain or restore detail without introducing artifacts. The generated high-resolution image needs to be visually natural and realistic, avoiding oversmoothing or unnatural sharpening effects. This requires algorithms to be able to accurately reconstruct details and maintain overall visual consistency while upscaling the image.

Objective evaluation metrics: Currently used evaluation metrics of the super-resolution include Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [2]. These metrics rely on realistic high-resolution images as a reference to assess the quality of reconstructed images. However, these metrics do not fully reflect the actual results, and there are often cases where the subjective evaluation metrics are high and the objective evaluation metrics are low. In addition, the lack of realistic reference images for natural images limits the application of these metrics. For this reason, it is crucial to develop objective evaluation metrics that do not require reference images and are consistent with subjective evaluation.

The main objective of this paper is to explore image super-resolution methods based on Generative Adversarial Networks (GANs), especially Super-Resolution Generative Adversarial Network(SRGAN) and Enhanced Super-Resolution Generative Adversarial Network(ESRGAN). The results illustrate the great potential of Generative Adversarial Networks in image super-resolution tasks, especially in improving the quality of the generated images, both SRGAN and ESRGAN perform well in restoring image details.

2. Overview of super-resolution techniques

According to the different implementation methods, the super-resolution techniques can be separated into two methods: traditional super-resolution techniques and deep learning-based super-resolution techniques.

2.1. Traditional super-resolution techniques

The traditional image super-resolution techniques include interpolation methods, statistical-based methods and reconstruction-based methods. The interpolation methods [3], such as bilinear interpolation

and bicubic interpolation, which are calculated by the values of adjacent pixels, are simple and easy to use, but the generated images are limited in detail and prone to blurring. The statistical-based methods, such as sparse coding and dictionary-based methods, can reconstruct high resolution images by training sparse dictionaries or utilizing pre-trained dictionaries [4], which can generate more details but have the high computational complexity.

The reconstruction-based methods search for the optimal solutions in the high-resolution image space through optimization algorithms and constraints, such as regularization reconstruction and projective reconstruction [5]. This type of method is theoretically robust [6], but the computational complexity is also high. The traditional methods improve the image resolution to a certain extent, but they usually have limitations in detail reconstruction and computational efficiency, and cannot fully meet the needs of practical applications.

Before the emergence of GAN, the image super-resolution methods mainly include bicubic interpolation, sparse coding and dictionary-based methods. The bicubic interpolation [7] is a classical interpolation method that upscale an image by calculating the values of the adjacent 16 pixels through a cubic polynomial, and although the calculation is simple, the generated image usually lacks details and is prone to blurring. The sparse coding methods are based on the theory of sparse representation of images by training sparse dictionaries and using these sparse coefficients, in order to build high-resolution pictures from low-resolution ones, which can reconstruct certain details but have high computational complexity. The dictionary-based methods utilize the pre-trained dictionaries to represent image blocks, and reconstruct the high-resolution images by finding and matching image blocks in the dictionaries, which can effectively reconstruct details but require a large number of computational resources and storage space.

Although these early methods have improved image resolution to a certain extent, they all have their own limitations.

2.2. Deep learning-based super-resolution techniques

Due to the advancements in deep learning, particularly the growth of Convolutional Neural Networks (CNNs), image super-resolution techniques have made breakthroughs. The deep learning-based methods such as CNNs, GANs, Recurrent Neural Networks (RNNs), and Autoencoders have significantly improved the image quality. SRCNN is the first proposed CNN-based super-resolution model, which can learn the mapping from low resolution to high one through multilayer convolutional networks [8]. SRGAN utilizes generative adversarial networks to generate high visual quality images, while ESRGAN takes this a step further to enhance image detail and visual effects. RNNs and LSTMs improve the image reconstruction effect by capturing long-range dependencies, and the autoencoder generates the high-resolution images by learning high-level representations of images.

2.2.1. Overview of GAN. The introduction of GAN, the image super-resolution techniques have made significant progress after CNNs. Deep learning-based methods can generate more realistic and high-quality high-resolution images, greatly improving the restoration of image details and visual effects.

The GAN is a type of deep learning model and is proposed by [9]. Its basic principle is to generate realistic data models by two neural networks competing with each other: Generative Model and Discriminative Model. These two networks compete with each other during the training process, in order to continuously improve the generator's capacity to produce more accurate data, and also to continuously optimize the discriminator's ability to more correctly distinguish between real data and generated data, so they are therefore called "Adversarial" networks. Eventually, the GAN model reaches equilibrium when the data generated by the generators is realistic enough that the discriminators cannot easily distinguish it.

2.2.2. SRGAN Techniques. Despite the increasing depth and speed of CNNs and significant advances in single-image super-resolution as for accuracy and processing speed, how to restore fine texture details at large magnification remains a core challenge. The choice of objective function has a significant impact

Proceedings of CONF-MPCS 2024 Workshop: Quantum Machine Learning: Bridging Quantum Physics and Computational Simulations DOI: 10.54254/2753-8818/52/2024CH0134

on the performance of traditional optimization-based super-resolution methods. Recent research has focused on minimizing the Mean Squared Error (MSE) of reconstruction. Although this method achieves high PSNR, the generated images usually lack high-frequency details and are not as perceptually effective, as they fail to achieve the desired fidelity and detail richness at high resolution. Therefore, the authors of [10] proposed a GAN for image Super-Resolution (SR), namely SRGAN. Its generator employs a Residual Network (ResNet) structure [11] to mitigate the gradient vanishing problem through skip connections. The generator network inputs a low-resolution image, and then outputs a high-resolution image after multi-layer convolution and activation function processing. Its discriminator uses CNNs to gradually extract the features of the image through multi-layer convolutional layers, and outputs a value between 0 and 1 through fully connected layers and Sigmoid activation function, which indicates the probability of whether the input image is a generated image or a real image.

2.2.3. ESRGAN technique. SRGAN is the first model to apply GAN to image super-resolution, and it achieves significant improvement in visual effect. However, SRGAN still has shortcomings in detail restoration and visual realism. For this reason, the authors of [12] proposed ESRGAN to further improve the quality of super-resolution images. ESRGAN employs a GAN architecture to generate high-resolution images from low-resolution ones, along with several improvements. First, ESRGAN introduces a new perceptual loss function that compares not only the pixel-by-pixel differences, but also the high-level features of the generated images and target images. This method allows the model to generate clearer, which appeals more visually and has a natural feel. Besides, ESRGAN employs a Residual-in-Residual Dense Block (RRDB) architecture in the generator network, which enables the model to capture more complex nonlinear mappings between high-resolution and low-resolution images, thereby significantly improving the image quality. The relative discriminator loss is used to enable the discriminator to learn the relative differences between the generated images and real images, and improve the fidelity of the generated images.

With these improvements, ESRGAN further enhances the details and visual effects of the image while generating high-resolution images, overcomes some limitations of SRGAN and the generated images are visually more natural and realistic.

These methods based on deep learning have significant advantages over traditional methods and are capable of generating more realistic and detail-rich high-resolution images. However, these methods usually require large amounts of training data and computational resources. In practical applications, there is a trade-off between image quality and computational to meet the needs of different application scenarios.

3. Experimental results and analysis

3.1. Basic framework

Figure 1 shows the basic framework of SRGAN and ESRGAN.

The SRGAN framework can be divided into two parts: generator network and discriminator network. The generator network starts from a low-resolution image, passes through multiple Residual Blocks, goes through a convolutional layer, batch normalization and PReLU activation function processing, and finally generates a high-resolution image through an upsampling layer (e.g., Sub-pixel convolution). The generated high-resolution image is fed into a discriminator network that includes multiple convolutional layers, batch normalization and LeakyReLU activation function, and then passes through a fully connected layer and a Sigmoid activation function to output a probability between 0 and 1, which indicates the authenticity of the input image.

The ESRGAN framework has been improved on the basis of SRGAN. The generator network starts with a low-resolution image, passes through multiple RRDB, each of which RRDB contains dense blocks and residual connections, after these processes, generates a high-resolution image through an upsampling layer (e.g. Sub-pixel convolution). The generated high-resolution images are fed into a discriminator network that includes multiple convolutional layers, batch normalization and LeakyReLU

activation function, and then passes through a fully connected layer, and specifically employs Relativistic Discriminator Loss to output a probability between 0 and 1, which indicates the authenticity of the input image.

During the process of training, the generator and the discriminator are trained alternately. The generator tries to generate more detailed high-resolution images to fool the discriminator, while the discriminator not only distinguishes between real and generated high-resolution images, but also evaluates their relative differences.



- · Alternately train generator and discriminator
- · Generator: Generate finer high-resolution images to fool the discriminator
- Discriminator: Differentiate between real high-resolution images and generated images, and evaluate their relative differences

Figure 1. Basic framework of SRGAN and ESRGAN

3.2. Datasets

In order to evaluate the performance of SRGAN and ESRGAN, the datasets used mainly include Berkeley Segmentation Dataset 100(BSD100), Set5 and Set14. These three datasets are the standard datasets commonly used in the field of image processing, especially in image super-resolution tasks.

BSD 100 is a collection of 100 natural images extracted from BSD300, which is rich in image content, including a variety of different scenes and objects, such as natural landscapes, animals, buildings and so on. It provides accurate manual annotation to evaluate the performance of the algorithms in image segmentation and edge detection tasks. Set5 (Set5 Image Dataset) contains five high-resolution images, including close-ups of babies, birds, butterflies, hair, and women's faces. While Set14 (Set14 Image Dataset) contains 14 high-resolution images with a variety of images including animals, buildings, people, etc. Compared with Set5, Set14 provides more image samples, which are suitable for a wider range of evaluations.

In this task, it will be measured by comparing the difference between the original high-resolution image and the model-generated high-resolution image.

3.3. Evaluation Metrics

The paper utilize the PSNR and the SSIM as metrics to evaluate the performance of the two SRGANs and ESRGANs. PSNR is used to measure the similarity between the generated image and the original high-resolution image, and the higher the value of PSNR, the better the quality of the image. Whereas SSIM is employed to evaluate how similar images are in terms of brightness, contrast, and structure, the higher the SSIM value, the better the image quality.

3.4. Analysis of Results

Datasets	Value	SRGAN	ESRGAN
Set5	PSNR	29.4	29.8
	SSIM	0.9019	0.9102
Set14	PSNR	28.49	30.15
	SSIM	0.8184	0.845
BSD100	PSNR	27.68	27.96
	SSIM	0.7620	0.784

Table 1. Performances of SRGAN and ESRGAN in different datasets

As can be seen in Table 1, the PSNR and SSIM values of ESRGAN in all the datasets outperform those of SRGAN, indicating that the images generated by ESRGAN are of higher quality in terms of quantitative metrics. This is mainly due to the improvements of ESRGAN in model architecture and loss function. First, ESRGAN adopts a generative network structure based on residual blocks, which enables the model to better capture the detailed information of the image and reduce image blurring during the training process. Second, ESRGAN introduces that it combines with perceptual loss and adversarial loss, which effectively improves the perceptual quality of the image, so that the generated image not only performs well in objective metrics, but also is more natural and realistic in subjective visual effects. In addition, the ESRGAN generator can better adapt to the characteristics of different datasets during the training process, thereby improving its generalization ability in multiple scenarios.

4. Conclusion

SRGAN and ESRGAN perform well in the image super-resolution task. SRGAN achieves significant enhancement of super-resolution images by introducing generative adversarial networks, while ESRGAN further improves the quality and detail restoration of the generated images by improving the network architecture and the loss function on this basis. The experimental results demonstrate that these models exhibit strong potential in various evaluation metrics.

Future research can continue to go deeper in model optimization and application field expansion. By further optimizing the GAN models and improving the computational efficiency and image quality, they can be applied to more image processing tasks. In the face of the current challenges, researchers can explore new methods and techniques to promote the development and application of image super-resolution technique. The wide application prospect of GAN models in the field of image processing provides rich opportunities and possibilities for future research and innovation.

References

- Sumbul, G., Cinbis, R. G. & Aksoy, S. (2019). Multisource region attention network for finegrained object recognition in remote sensing imagery. IEEE Trans. Geosci. Remote Sens., 57(7), 4929-4937.
- [2] Wang, Z., Bovik, A. C., Sheikh H. R. & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process., 13(4), 600-612.
- [3] Tsai, R. & Huang, T. (1984). Multiple frame image restoration and registration. Advances in Computer Vision & Image Processing, 317-339.

- [4] Zhang, Y., Wu, W., Dai, Y., Yang, X., Yan B. & Lu, W. (2013). Remote Sensing Images Super-Resolution Based on Sparse Dictionaries and Residual Dictionaries. 2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing, 318-323.
- [5] Lorent, A. & Cierniak, R. (2014). Regularized Image Reconstruction from Projections Method. 2014 International Conference on Engineering and Telecommunication, 82-86.
- [6] Farsiu, S., Robinson, M. D., Elad, M. & et al. (2004). Fast and robust multiframe super resolution. IEEE Transactions on Image Processing, 13(10), 1327-1344.
- [7] Rujul, R. & Mehta, D. (2013). Survey on single image Super resolution techniques. IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), 5(5), 23-33.
- [8] Dong, C., Loy, C. C., He, K. & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell, 295-307.
- [9] Goodfellow, I., Pouget, A. J., Mirza, M. & et al. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 2672-2680.
- [10] Ledig, C., Theis, L., Huszr, F., Caballero, J., Cunningham, A., Acosta, A. & et al. (2017). Photorealistic single image super-resolution using a generative adversarial network. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 105-114.
- [11] He, K., Zhang, X. & et al. (2016). Deep residual learning for image recognition. Washington, DC: IEEE Computer Society, 770-778.
- [12] Wang, X. T., Yu, K., Wu, S. X., Gu, J. J., Liu, Y. H., Dong C., & et al. (2018). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. arXiv:1809.00219.