Super-Resolution Reconstruction of COVID-19 Images Based on Generative Adversarial Networks

Jiarui Zhou
Sino-European School of Technology, Shanghai University, Shanghai, China

Keywords: GAN, Image Enhancement.

Abstract: In low brightness or light conditions, the images generated by existing medical equipment for testing patients often have problems such as low clarity, feature loss, and excessive noise. The accuracy and timeliness of medical image processing directly affect doctors' diagnosis and treatment. A series of image enhancement technologies can viably make strides in the quality of low-brightness pictures, making image features more obvious, and thereby helping doctors identify lesions faster and more accurately. Traditional image enhancement techniques work for general cases, while super-resolution with Convolutional Neural Networks (CNNs) needs large labeled datasets and often misses high-frequency details during large-scale upscaling. In contrast, Generative Adversarial Networks (GANs), such as unsupervised learning neural networks, can effectively solve such problems. The objective of this work is to reproduce low-resolution COVID-19 images to super-resolution through the Super-Resolution Generative Adversarial Network (SRGAN) model. The results of the experiment show that the model can perform super-resolution reconstruction of COVID-19 images well.

1 INTRODUCTION

Image resolution reflects the pixel density of an image and is also a parameter for evaluating image detail information. High-resolution photos have more detailed information and denser pixels than low-resolution ones. However, in real life, image generation equipment is affected by many factors and cannot easily acquire perfect, high-resolution pictures. Therefore, the innovation of picture super-resolution reproduction from the perspective of algorithms and software is a popular area of study for computer vision and image processing.

Since December 2019, COVID-19 has spread rapidly around the world, with far-reaching impacts. It has dealt a severe blow to the global economy and changed the way people work and live. COVID-19 is the disease caused by the coronavirus, which essentially impacts the respiratory framework and lungs. Infection may cause severe respiratory symptoms such as pneumonia and has a high mortality rate. The main topic of this study is the application of super-resolution reconstruction in COVID-19 images. A typical technique for reconstructing images with high resolution is the interpolation method proposed by Gao et al. (2016). There are also reconstruction methods based on degradation models, the most notable of which is the iterative back-projection method proposed by Irani and Peleg (1991). Another significant method is the maximum a posteriori probability approach introduced by Schultz and Stevenson (1996). Additionally, Yang et al. (2010) proposed a method based on sparse representation. Among interpolation methods, the most commonly used is the bicubic interpolation method, which builds upon the bilinear interpolation technique. However, while this method enhances image reconstruction quality, it also increases computational complexity (AhmatAdil, 2021).

This paper mainly studies the application of Super-Resolution Convolutional Neural Network (CNN) technology based on generative adversarial networks (GANs) in COVID-19 images. The objective is to enhance the visual quality of COVID-19 pictures by performing super-resolution reconstruction using the Super-Resolution Generative Adversarial Network (SRGAN) approach. This will help to improve the precision and rapidity of subsequent diagnosis and treatment.

Zhou and J.
Super-Resolution Reconstruction of COVID-19 Images Based on Generative Adversarial Networks.
DOI: 10.5220/0013525000004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 406-411
ISBN: 978-989-758-754-2

2 LITERATURE REVIEW

Deep learning has started to be used in a variety of industries as artificial intelligence and computer technology have advanced. In the realm of image enhancement, CNNs have come to the fore. In 2015, Dong et al. first applied CNN to the field of image super-resolution reconstruction, namely Super-Resolution Convolutional Neural Network (SRCNN). Due to the slow running speed and high computational cost of SRCNN, Dong et al. improved SRCNN in 2016, increased the running speed, and named the model Fast SRCNN (FSRCNN). First, SRCNN enlarges the low-resolution image through interpolation and then restores it through the model. However, Shi et al. (2016) believe that research should start from the fundamentals and learn how to enlarge the data samples through the model. Based on this principle, they proposed an image super-resolution algorithm called Efficient Sub-Pixel Convolutional Network (ESPCN) with a sub-pixel convolutional layer. The ESPCN algorithm introduces a sub-pixel convolution layer, which indirectly achieves image magnification and greatly reduces the computational complexity of SRCNN. In 2020, Basak et al. introduced the channel attention mechanism method into SRCNN and achieved good results. Both SRCNN and ESPCN use Mean Square Error (MSE) as the loss function. This results in overly smooth images and insufficient details at high frequencies. To solve this problem, Christian et al. proposed a super-resolution reconstruction technique based on GAN, SRGAN, and innovatively defined a perceptual loss function (Ledig et al., 2017). The SRGAN method introduces a sub-pixel convolution layer to replace the traditional deconvolution layer and introduces the feature extraction module in the VGG19 model as the content loss for comparing super-resolution images with original high-definition images. The traditional content loss is obtained using the MSE method, which directly compares the pixel differences between two images. However, Christian et al. believe that this traditional method will only cause the model to over-learn pixel differences and ignore the deep intrinsic features of the reconstructed image (Ledig et al., 2017). Models such as the VGG19 model that specialize in extracting intrinsic features of images are just suitable for such tasks. At this point, the SRGAN model's general framework is now complete. When reconstructing super-resolution, the texture details of the images generated by SRGAN are much higher than those of SRCNN. Nevertheless, the images with super-resolution that were created by the original SRGAN model are still different from the original high-resolution images. In 2021, Wang et al. proposed an enhanced SRGAN based on the original SRGAN model, namely the Enhanced SRGAN (ESRGAN), which achieved good results.

3 EXPERIMENTAL DATASETS

This research uses experimental data that was obtained from the Kaggle website. This website is a professional machine learning platform website. The data set comes from the open-source high-definition COVID-19 data set shared by users and contains 5779 high-definition lung X-ray images. A training set and a test set are separated from the data set in a 9:1 ratio. The training set consists of 5216 photos, whereas the test set consists of 563 images. This paper first uses the resize function to unify the training set's picture size to 720×1280 and then uses the randint function to randomly crop the training set images, setting the crop size to 96×96 and randomly selecting the cropping position from the image. The cropped images are then converted into tensors through the torch. Tensor function and normalized to $[-1,1]$. Finally, the cropped images are quadrupled using bicubic interpolation to obtain images of size 24×24 and input into the generator.

4 METHODS

This paper reduces the number of residual modules of the generator in the SRGAN algorithm from 16 to 8 to lessen the complexity and quantity of computing.

At the same time, the dropout regularization technology is added to the discriminator to prevent model overfitting, strengthen the model's resilience and capacity for generalization. In the course of training, the high-resolution image is first changed using a method called bicubic downsampling to create a lower-resolution image, and then the image is input into the generator, and the output super-resolution image is obtained through the generator. Then high-resolution image and super-resolution image are input into the discriminator respectively, and the discriminator determines the authenticity and returns the outcome to the generator while optimizing the parameters of the discriminator itself. Finally, the high-resolution image and the super-resolution image's intrinsic feature differences are compared through the generator's loss function to optimize the generator parameters. The specific training flowchart is displayed in Figure 1, and the generator and discriminator's general structure can be seen in Figure 2 and Figure 3.

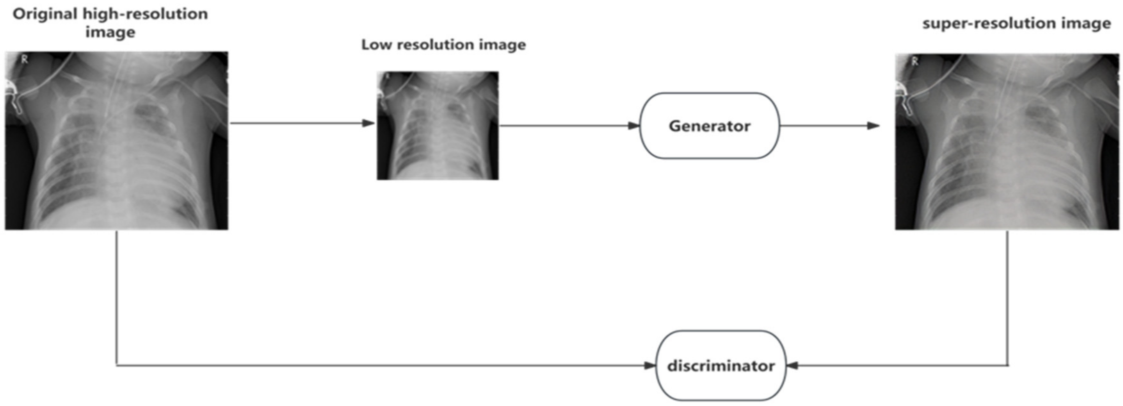


Figure 1: SRGAN Model Flowchart (Photo/Picture credit: Original).

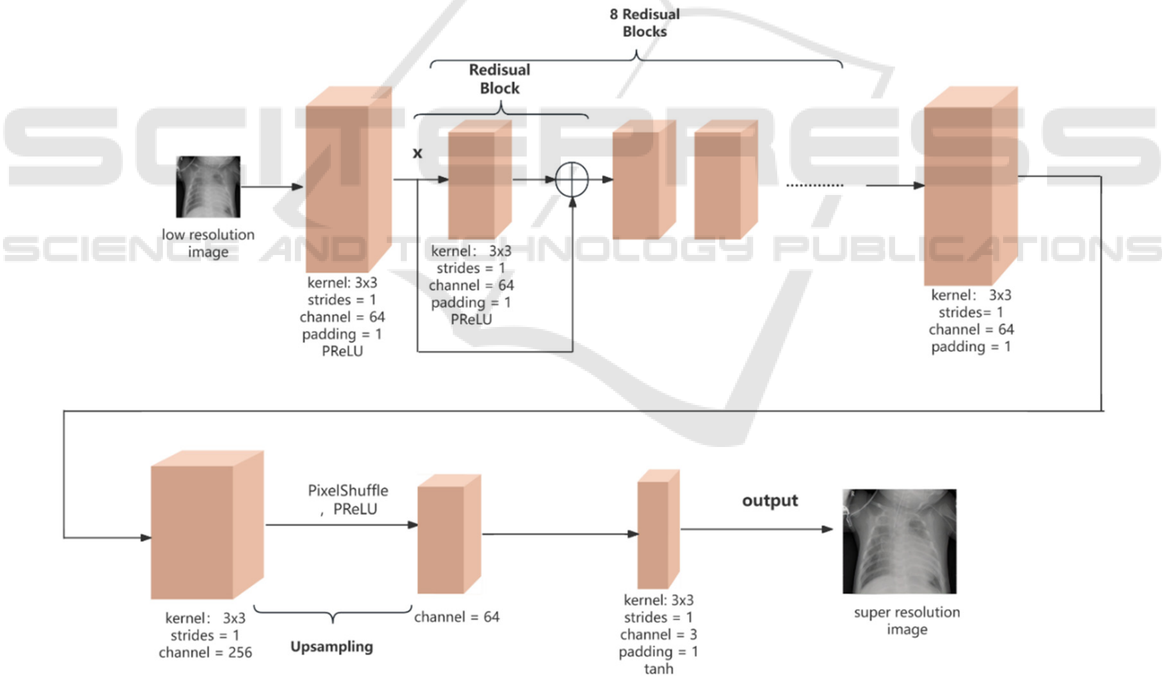


Figure 2: Generator flowchart (Photo/Picture credit: Original).

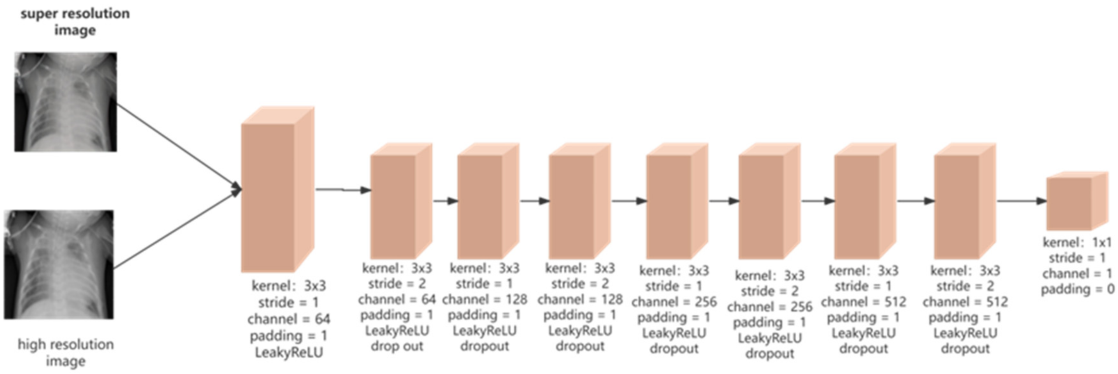


Figure 3: Discriminator flowchart (Photo/Picture credit: Original).

The generator's residual blocks are modified by this study, which is based on the original SRGAN model, from 16 to 8. First, the low-resolution image passes through a 3×3 convolution layer, and 64 output channels are configured to Obtain the shallow features of the image. Then, the image is passed through an 8-layer residual module to extract deeper features. The output result is then passed through a 3×3 convolution layer and normalized by the InstanceNorm2d function. Finally, an upsampling module and a reconstruction layer are used to reconstruct the super-resolution image.

In the discriminator network, the super-resolution image and the high-resolution image are first input into the 3×3 convolution layer to obtain the characteristics of the image that are shallow, 64 output channels are set. Then, after passing through 7 layers of 3×3 convolution layers, the number of channels rises to 512 from 64, and then passing through the judgment layer, the output channel is set to 1, and the output is the probability of the super-resolution image being true or false compared to the original high-resolution image.

5 EXPERIMENTAL RESULTS

5.1 Evaluation Methodology

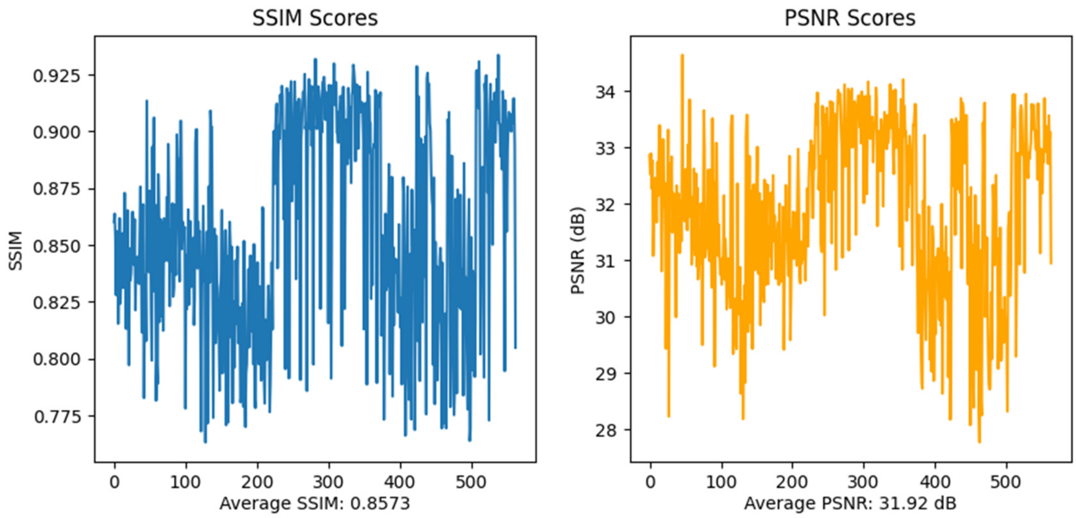
In order to assess the quality of super-resolution images, this paper employs structural similarity (SSIM) and peak signal-to-noise ratio (PSNR). PSNR is shown in Formula (1).

$$PSNR = 10 \times \log_{10}\left(\frac{MAX^2}{MSE}\right) \qquad (1)$$

Among these are the MSE of pixels between the original image and the reconstructed image and the maximum pixel value that can occur in an image. Better image quality is associated with a higher value. However, since it only compares the differences between image pixels and ignores the human eye's subjective experience of vision, the PSNR comparison method does not adequately reflect the differences in high-frequency details. The structural similarity SSIM is shown in Formula (2).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2\sigma_y^2 + C_2)} \qquad (2)$$

The calculation of SSIM is based on a sliding window. That is, each time a calculation is performed, a window of size $N \times N$ is captured from the image, and then the average SSIM of all the windows is determined after calculating the SSIM index for each window. In the formula, x represents the first image window data, and y represents the second image window data μ_x and μ_y are the means of x, y, σ_x and σ_y are the variances of x and y, σ_{xy} represents the covariance of x and y. C_1 , C_2 are two constants. Compared with PSNR, SSIM considers information on structure, contrast, and brightness, making SSIM closer to the evaluation of human visual perception. By analyzing the structural similarity of image patches, SSIM can capture the impact of distortion on image visual quality. In summary, PSNR provides error evaluation at the pixel level, while SSIM provides an evaluation of the structure and subjective visual perception of human eyes. Combining the two can provide a more comprehensive image quality evaluation.



(a) (b)

Figure 4: Experimental Results. (a) show the SSIM scores, (b) show the PSNR scores (Photo/Picture credit: Original).

5.2 Results Analysis

Figure 4 (a) displays the findings of the SSIM evaluation, while Figure 4 (b) presents the results of the PSNR analysis. The average SSIM value is 0.8573, indicating a high degree of similarity between the generated images and the high-resolution images in terms of structure and subjective perception by the human eye. The average PSNR value is 31.92 dB, suggesting that the generated super-resolution images have good reconstruction quality at the pixel level. Although the average values of SSIM and PSNR are high, this does not imply that all generated images meet this standard. Figures 4 (a) and 4 (b) clearly show that the values of SSIM and PSNR fluctuate. These variations are often due to the differing capabilities of the SRGAN model when processing various types of images. In terms of PSNR, the SRGAN model struggles to perfectly reconstruct pixel details in images with complex features or fine details. Conversely, for images with simple structures and smooth textures, the model is more adept at capturing these features, resulting in a higher SSIM value for such images. However, for images with complex structures, intricate details, and sharp edges, the model may introduce some distortion during reconstruction, leading to a decrease in the SSIM value.

6 CURRENT LIMITATIONS AND FUTURE PROSPECTS

Using the outcomes of experiments as a basis, this paper concludes that the SRGAN model has low PSNR and SSIM values. When processing complex COVID-19 images, and the overall results are volatile. This indicates that when processing images with complex textures and rich details, image details may be lost, and the generated pictures will produce artifacts in this case. Then, the SRGAN model is a unimodal model that provides a single set of information.

The SRGAN model introduces VGG19 to extract perceptual loss. In the future, the research can try to introduce other image feature extraction networks, such as Residual Neural Network (ResNet) or Inception module in Deep Neural Network (DNN), and compare the performance of these three networks to obtain an SRGAN model with better performance. In terms of training, the research can consider introducing data with richer details and more complex textures, improving the model structure, and fine-tuning the model separately when processing complex images to enhance the model's ability to handle complicated images. Finally, in the future, the research hopes to introduce image fusion technology based on GANs, for example, in medical images, a fusion of CT and MRI images to provide more comprehensive image information.

7 CONCLUSIONS

Because of advancements in science and technology, modern science has made the processing of medical images an indispensable part of this field. Through various image enhancement technologies, medical images with various problems can be effectively enhanced. This can not only improve the accuracy of doctors' diagnoses but also improve the speed of diagnosis. This paper's primary research goal is to improve the SRGAN model and apply it to COVID-19 images. Based on the original SRGAN model, in an effort to minimize the amount of computation while ensuring the quality of model reconstruction, the quantity of the generator's residual blocks in the SRGAN model was reduced from 16 to 8, and a regularized dropout method was added to the discriminator to enhance the model's capacity to be generalized. Through the above experimental results, the job of reconstructing super-resolution images can be successfully completed using the SRGAN model, and can effectively help doctors to make subsequent diagnosis and treatment after confirming the lesion. However, the model performs poorly when processing complex images. In the future, the research will consider introducing a perceptual loss extraction model with better performance, improving the model structure, and enhancing the model performance. At the same time, the research hopes to introduce image fusion technology to obtain medical images with more detailed information.

REFERENCES

AhmatAdil, A., 2021. Analysis of Medical Image Resizing Using Bicubic Interpolation Algorithm. JIK: Jurnal Ilmu Komputer.

Basak, H., Kundu, R., Agarwal, A., Giri, S., 2020. Single image super-resolution using residual channel attention network. In 2020 IEEE 15th international conference on Industrial and information systems (ICIIS) (pp. 219-224). IEEE.

Dong, C., Loy, C. C., He, K., Tang, X., 2015. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

Dong, C., Loy, C. C., Tang, X., 2016. Accelerating the super-resolution convolutional neural network. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14 (pp. 391-407). Springer International Publishing.

Gao, Y., Beijbom, O., Zhang, N., Darrell, T., 2016. Compact bilinear pooling. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 317-326).

Irani, M., Peleg, S., 1991. Improving resolution by image registration. CVGIP: Graphical models and image processing, 53(3), 231-239.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

Schultz, R. R., Stevenson, R. L., 1996. Extraction of high-resolution frames from video sequences. IEEE transactions on image processing, 5(6), 996-1011.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

Yang, J., Wright, J., Huang, T. S., Ma, Y., 2010. Image super-resolution via sparse representation. IEEE transactions on image processing, 19(11), 2861-2873.