Cartoon single-image super-resolution approach based on generative adversarial network

Guangxing Wang[1]**, Seong-Yoon Shin**[2]**, Jong-Chan Kim**[3]
[1]School of Computer and Big Data Science, Jiujiang University, Jiujiang, China
[2]School of Computer Science and Engineering, Kunsan National University, Gunsan, South Korea
[3]Department of Computer Engineering, Sunchon National University, Suncheon, South Korea

Article Info

Article history:

Received Dec 11, 2023
Revised Jan 8, 2023
Accepted Jan 11, 2024

Keywords:

Deep learning
Generative adversarial networks
Machine learning
Single image super-resolution
Transform learning

ABSTRACT

In recent years, the study of a single image super-resolution (SISR) is crucial to improving image resolution and using hardware technology to improve image resolution. SISR is widely used in satellite remote sensing, video surveillance, and medical image processing because it mainly relies on deep learning algorithms to realize the conversion from low-resolution (LR) images to high-resolution images. It has the advantages of low cost, simple operation, and high efficiency. This paper proposes an image super-resolution method based on a generative adversarial network named text localization generative adversarial nets (TLGAN) model. The method is improved based on super-resolution generative adversarial networks (SRGAN), and the batch normalization layer is removed, which significantly reduces the computational burden of the model. In TLGAN model, we used the transfer learning method to pre-trained the model on the large dataset ImageNet, and then apply the pre-trained model to the cartoon image data set animes to achieve image super-resolution. Experimental results report that the proposed method has the advantages of fast running speed and excellent visual perception of super-resolution images compared with bicubic interpolation and SRGAN method.

Corresponding Author:

Seong-Yoon Shin
School of Comuter Science and Engineering, Kunsan National University
Gunsan 54150, South Korea
Email: s3397220@kunsan.ac.kr

1. INTRODUCTION

Image super-resolution (ISR) technology is a computer software technology that recovers a high-resolution image from a low-resolution (LR) image or image sequence [1]. Generally, image super-resolution technology is divided into super-resolution restoration and super-resolution (SR) reconstruction. The single image super-resolution (SISR) is a computer software technology that recovers a high-resolution (HR) image from a LR image. The research of ISR is mainly divided into interpolation-based, reconstruction-based, and learning-based methods. Of course, ISR can also be achieved through hardware technology. The disadvantage is that the cost is too much, and the equipment accuracy requirements are rigorous.

In recent years, more and more researchers prefer software-based ISR research, especially reconstruction-based methods. The main reason is that software-based image super-resolution technology has the advantages of high efficiency and low cost. Yang *et al.* [2] proposed the use of sparse coding to achieve ISR. Yang *et al.* [3] used the sparse prior features of image statistics to achieve the restoration of LR images to HR images. Timofte *et al.* [4] proposed an image super-resolution method based on fast sample anchor point proximity restoration. Chavez-Roman and Ponamaryov [5] tried to restore high-resolution images by combining offline wavelet transfer and sparse representation. Although these methods are based on mathematical and statistical methods to achieve high-resolution image restoration, they have certain advantages under the current technical conditions. However, they are challenging to deal with the current image super-resolution projects of large data sets. These methods have disadvantages, such as long model operation time and low accuracy of the high-resolution images obtained.

In the past two decades, simulation research based on biological neurology has promoted machine learning development. By establishing a neural network model, the computer has a learning and thinking function similar to the human brain. Especially in past decade, the introduction of deep learning (DL) into computer vision has promoted a leap in graphics and image processing technology, and the ISR method based on DL has made remarkable achievements. Dong *et al.* [6] proposed an ISR algorithm named super-resolution convolutional neural network (SRCNN) based on a convolutional neural network (CNN). Although an image sensory evaluation method is proposed in SRCNN, the high-resolution image obtained has shortcomings such as artifacts and aliasing, and the visual effect is not ideal. Kim *et al.* [7] proposed a deep neural network (DNN) SR model named VDSR, which uses a deep recursive layer to achieve feature extraction. High-resolution recovery has achieved good results, but artifacts and aliasing still exist. Ledig *et al.* [8] proposed to achieve the SR of realistic scene based on a generative adversarial network (GAN) model named super-resolution generative adversarial networks (SRGAN). The SRGAN model uses a two-layer architecture, namely the generator and the discriminator. The generator generates the corresponding high-resolution image HR from the low-resolution image LR through continuous learning. The discriminator compares the HR with the original HR image GT, and if they are different, it is judged as false; otherwise, it is judged as real. The process of SR model training is the process of the game between the discriminator and the generator. As a result of model training, it is known that the HR image determined by the discriminator is actual.

In this paper, transform learning generative adversarial network (TLGAN) model is proposed that is an ISR method based on transfer learning (TL). Based on the SRGAN model, we first remove the batch normalization (BN) layer of the SRGAN model, which helps reduce the computational burden of the computer. Secondly, inspired by migration learning, the proposed model is pre-trained with extensive data set ImageNet, and the model is saved. Thirdly, the pre-trained model performs image super-resolution reconstruction on the cartoon dataset Animes to obtain high-resolution animes images. Finally, super-resolution experiments were performed on the proposed model in three cartoon data sets. Experimental results indicates that the proposed method is superior to bicubic interpolation and SRGAN method, and has the advantages of short running time and sound visual effects of high-resolution images. Moreover, the image artifacts are reduced to a certain extent, the aliasing is eliminated, the texture is clear, and the visual effect is realistic and natural. As shown in Figure 1, the super-resolution images generated by TLGAN models.

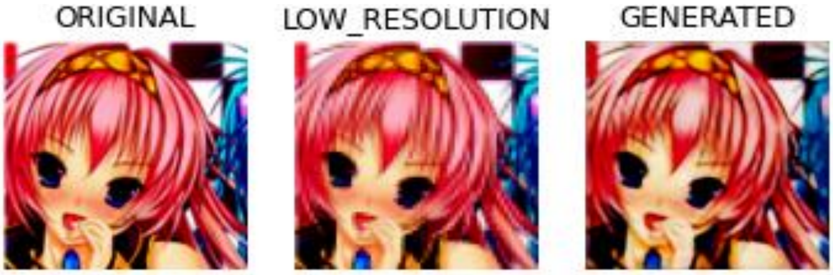


Figure 1. The image reconstructed by the TLGAN model is compared with the original and LR image

The rest of this paper is structured as follows. In section 2, related work on ISR is introduced. Section 3 presents the proposed method, including the model structure, parameter setting, and evaluation metric. The experimental process and results will be introduced in section 4. Section 5 summarizes the research work and gives future works.

2. RELATED WORK

ISR is a computer vision technology that improves the original image resolution through hardware or software. The process of obtaining a HR image from a series of LR images is called super-resolution reconstruction. The core idea of SR reconstruction is to exchange time-bandwidth (i.e., acquiring multiple frame image sequences of the same scene) for the spatial resolution to realize the conversion from time resolution to spatial resolution. Using hardware methods to improve image resolution, usually using higher-precision optical sensor originals, or increasing the shooting device's chip size, increasing image capacity, and improving image resolution [9], [10].

In recent years, the ISR method based on signal processing technology has been favored by researchers. This method is to obtain HR images from multiple observable LR images. In recent years, machine learning has used CNN algorithms to deal with image SR research and achieved astonishing results. Through the CNN, the LR image's feature information can be lifted and restored to generate the HR image. The HR image generated by the SRCNN model has artifacts, and aliasing. The model training time is extended. To solve the problems of SRCNN, researchers utilized GAN model in SR research [11]–[14]. This method means training the GAN model through paired LR-HR images, and using the trained model to establish the new HR images in a reconstructed way. The reconstructed HR image has realistic texture and natural visual effects. In addition, papers newly presented on image super resolution were dealt with in detail in [15]–[18].

3. TLGAN MODEL

3.1. Problem statement

The goal of image super-resolution is to recover the correct HR image from the LR image. Here, I_x and I_y represent the degraded LR and HR images, respectively. The mathematical formula for generating LR from HR images is shown in (1).

$$I_x = D\left(I_y \colon \delta\right) \tag{1}$$

D is marked as a degradation mapping function, and δ is marked as a degradation parameter, including scaling factor, and noise. Usually, only LR images are given during the image degradation process. Since the degradation process is unknown, the researchers can only recover a HR image $\hat{I}_y$, and it is similar to the real HR image from the LR by establishing a model. This process is expressed by (2).

$$\hat{I}_y = F(I_x \colon \theta) \tag{2}$$

here, F represents the SR model, and θ represents the parameter.

Since the image degradation process is affected by image compression artifacts, anisotropic degradation, sensor noise and speckle noise, to take these factors that affect the degradation process into consideration, most researchers use down-sampling to establish a degradation model. The mathematical formula of the downsampling process is shown in (3).

$$D\left(I_y; \delta\right) = \left(I_y\right) \downarrow_s, \{s\} \subset \delta \tag{3}$$

In where, $\downarrow_s$ means that the scaling factor is s down-sampling operation. Since most of the SR algorithms are based on image texture, and the usual down-sampling operations use bicubic interpolation algorithms, these can achieve image super-resolution but still cannot get rid of the effects of artifacts. Inspired by these researchers, the centralized sampling method is usually combined to perform image degradation operations. This processing process is represented by mathematical formula, as shown in (4).

$$D\left(I_y; \delta\right) = \left(I_y \cdot k\right) \downarrow_s + n_\varsigma, \{k, s, \varsigma\} \subset \delta \tag{4}$$

Among them, $I_y \cdot k$ is marked the convolution operation of the blur kernel and HR image. n_ς is marked the added white Gaussian noise which standard deviates from ζ. Compared with (3), (4) is closer to reality. The HR image produced by (4) has a more natural visual effect compared with (3). Therefore, a more practical mathematical expression for super-resolution is as in (5).

$$\hat{\theta} = \arg\min_{\theta} L\left(\hat{I}_y,\ I_y\right) + \lambda \Phi(\theta) \tag{5}$$

In where, $L\left(\hat{I}_y,\ I_y\right)$ is marked the loss function between the generated HR $\hat{I}_y$ and the real image I_y, $\Phi(\theta)$ represents regular term, and λ represents equalization parameter. Most loss functions are measured by pixel-by-pixel loss. Therefore, in order to enhance the robustness of the model, multiple loss functions are adopted. In the SRGAN model, the pixel-by-pixel loss is usually used to measure an image loss rate.

3.2. Network structure

The core of the SR algorithm that generates the adversarial network is reconstructing a HR image I_y from the input LR image I_x. In model training, such I_x and I_y appear in pairs; that is to say, I_y obtains I_x through convolution kernel and Gaussian noise degradation, and it is usually performed by down-sampling. Since the GAN network is mainly composed of generator *G* and discriminator *D*, as described by Ledig *et al*. [8], this paper defines generator *G* and reconstruct an SR image by parameter θ_G. The determiner *D* is defined, and the parameter θ_D determines the authenticity of the generated image. Therefore, the final GAN network is marked as the maximum and minimum problem, and the mathematical expression is shown in (6).

$$\min_{\theta_G}\max_{\theta_D} V(D,G) = E_{I_y \sim P_{train}(I_y)}\left[\log\left(D_{\theta_D}(I_y)\right)\right] + E_{I_x \sim P_G(I_x)}\left[\log\left(1 - D_{\theta_G}(I_x)\right)\right] \tag{6}$$

In the following section, the mathematical representation of the SR mechanism for a single image will be briefly introduced. Then we will illustrate the structure of our model based on transfer learning in detail, including loss function, activate function, and evaluation metric. In the GAN network, a particular sensory loss function is marked as L^{SR}, which is mainly adopted to calculate the content loss and counter-loss in the image reconstruction process. The generation process of SR loss L^{SR} is shown in (7).

$$L^{SR} = L^{SR}_{con} + \gamma L^{SR}_{adv} \tag{7}$$

In the GAN model, the content loss is mainly calculated based on the pixel-by-pixel to compensate for the lack of content caused by mean square error (MSE). The content loss is generally obtained by calculating the Euclidean distance between the features of the real image I_y and the reconstructed image $G_{\theta_G}(I_x)$. The mathematical expression of content loss is shown in (8).

$$L^{SR}_{con} = \frac{1}{W_{ij}H_{ij}}\sum_{x=1}^{W_{ij}}\sum_{y=1}^{H_{ij}}\left(\emptyset_{ij}(I_y)_{xy} - \emptyset_{ij}\left(G_{\theta_G}(I_x)\right)_{xy}\right)^2 \tag{8}$$

here, *W* and *H* are marked as the width and height of the image, respectively. (*i*, *j*) represents the feature map dimension, and φ_{ij} is marked as the feature map vector between the *i*-th pooling and the *j*-th convolutional layer.

Based on the probability of the discriminator on all training images to a high-dimensional space, and generates a data distribution that is difficult for the discriminator to distinguish. The loss caused by this process is called adversarial loss, also known as the GAN loss. The mathematical expression adversarial loss is shown in (9).

$$L^{SR}_{adv} = -\sum_{n=1}^{N}\log\left(D_{\theta_D}\left(G_{\theta_G}(I_x)\right)\right) \tag{9}$$

Generally, in order to calculate the GAN loss, the initial weight parameter Υ is set to le-3, so that when we minimize the adversarial loss, the generated adversarial loss is expressed as $\log\left(D_{\theta_D}\left(G_{\theta_G}(I_y)\right)\right)$. The advantage of this is that it can enhance gradient descent during initial training and accelerate model convergence.

3.3. Activate function

In the original SRGAN model, the generator is mainly composed of convolutional layer, batch normalization (BN) [19] layer, and activation function. The purpose of the BN operation is to regularize the size of the model input unit, avoid gradient disappearance or gradient explosion, and accelerate model training process. However, although the BN layer can normalize the image feature map and solve the problem of inconsistency in the input image feature map's size, compared to solving the classification problem, the BN operation appears to be inadequate in dealing with the image super-resolution problem. Therefore, in our model, we removed the BN operation and replaced it with a deep residual module.

ReLU [20] is adopted as the activation function in the GAN network. However, the PReLU function is adopted as activation function in proposed model. The advantage of this is that in the model training process, the activation function can be adjusted according to the dynamic parameters in the training process so that the image can retain the most extensive feature map, thereby reducing the feature loss during the training process. The mathematical expression of *PReLU* is shown in (10).

$$PReLU(x_i) = \begin{cases} x_i & if x_i > 0 \\ a_i x_i & if x_i \leq 0 \end{cases} \tag{10}$$

The above represents the loss function of the proposed TLGAN model structure. The loss function is introduced in detail from a mathematical perspective, including setting the activation function. The TLGAN model structure diagram is shown in Figure 2.

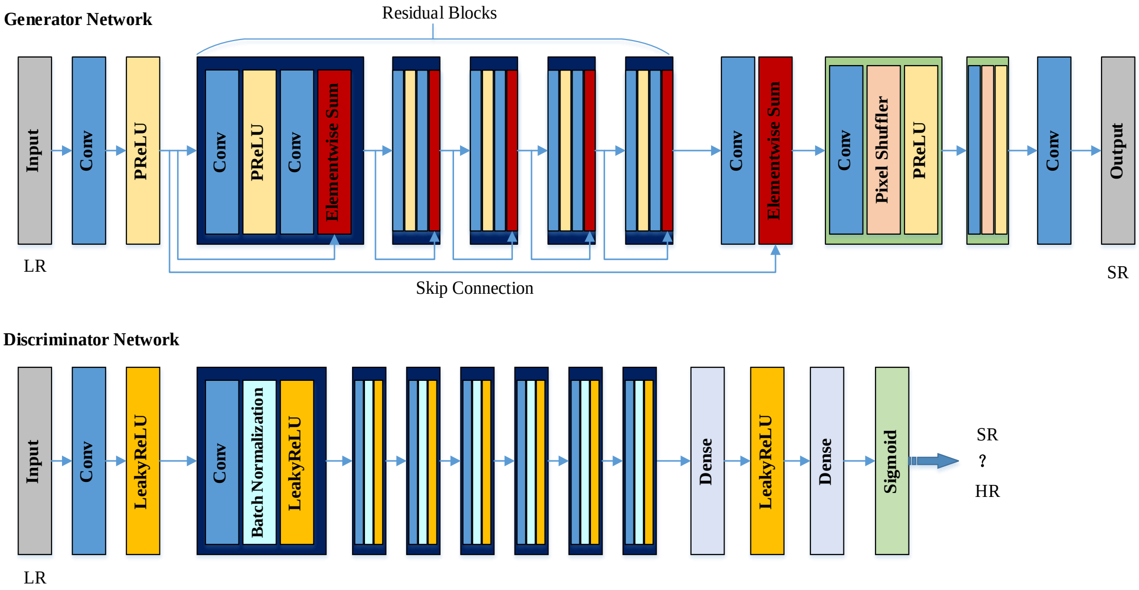


Figure 2. Proposed super-solution model architecture

3.4. Image qualify metric

For the quantitative evaluation of experimental results, we use two commonly used evaluation indicators which are peak signal-to-noise ratio (PSNR) and structural similarity metrics (SSIM) [21]. PSNR is often used as a metric to objectively evaluate images. The definition of PSNR is as follow.

Assume that I represents a clean image and K represents a noisy image of size m×n, the definitions of MSE and PSNR are shown in (11) and (12), respectively.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{11}$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{12}$$

In where, MAX_I^2 represents the image maximum possible pixel value. Assume that each pixel is represented by 8-bit binary, the maximum value is 255. Normally, if the pixel value of an image is marked as T-bit, then $MAX_I^2 = 2^T - 1$.

Of course, this method is to calculate the PSNR of the grayscale image. For color images, the PSNR of the three red, green, and blue (RGB) channels needs to be calculated, and then the average of their PSNRs is calculated. Finally, the PSNR value of the image can be objectively reflected.

Another evaluation metric is SSIM, which is a metrics of two images similarity. In the calculation of SSIM, one image is uncompressed and undistorted, and the other is distorted. SSIM is a common indicator used to measure the structural similarity of two images, and its value range is [0,1]. The larger the value of SSIM is, the more similar the two images are. When the value of SSIM is 1, it means that the two images are the same. The SSIM is calculated by (13).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{13}$$

Among them, μ_x is marked as the average of x, μ_y is marked as the average of y, σ_x^2 is maeked as the variance of x, σ_y^2 is marked as the variance of y, and σ_{xy} is marked as the covariance of x and y. $c_1 = (k_1L)^2, c_2 = (k_2L)^2$ are constants used to maintain stability. L is marked as the dynamic range of the pixel value, usually k_1=0.01, k_2=0.03.

4. EXPERIMENT

4.1. Experimental environment

In the actual experiment, the experimental platform we used was a workstation equipped with NVIDIA TITAN Windows 10 with Tensorflow and CUDA 11.0 [22], [23]. We use Python language to build models, and conduct model training and testing. In the experiment, the model visual geometry group 19 (VGG19) [24] pre-trained on the large dataset of ImageNet [25]–[27] was used to extract image features, and the pre-trained model was trained and tested on Anime cartoon data. We verify the model robustness through the Simsons-characters and Anime-Face [28], [29] datasets. Table 1 shows the details of the dataset.

Table 1. Detailed description of datasets

Dataset	Amount	Avg. Res.	Avg. Pix.	Format
ImageNet	143,1167	500×375	187,500	JPEG
Simsons-charts	9,878	1280×720	921,600	PNG
Anime-face	21,551	64×64	4,096	JPG

4.2. Experimental details

In the experiment, the size of the LR image input by the proposed model is 64×64, the batch size is 8, and the HR image size is 256×256, all of which are color images. The model is performed by up-sampling, and the scale scaling factor of the bicubic interpolation method is *X*4. The initial learning is *le*-4, optimizer is Adam, and the number of training rounds is 500. The generator uses 16 residual blocks, the convolution kernel value is 3, the step value is 1, and ReLU is adopted to activation function. The activation function used by the discriminator is LeakyReLU, the momentum is 0.8, and the sigmoid function is used for classification. Before the image is sent to the model for training, data expansion techniques such as cropping and rotation are used to ensure the sufficient amount of model training data and avoid the phenomenon of non-convergence or over-fitting due to too little data. The data normalization method is used to prevent the occurrence of gradient explosion and gradient disappearance problems. Following common practice, PSNR and SSIM metrics are adopted as evaluation indicators for the evaluation of model prediction results. Finally, the model training and prediction results are saved, and the evaluation scores of super-resolution images and the effective display of super-resolution images are compared. The training loss curve is shown in Figure 3. Figure 4 is the PSNR evaluation curve of the proposed model training. Figure 5 is the SSIM evaluation curve of the proposed model training. The comparison of super-resolution prediction results of the proposed model is shown in Figure 6. The super-resolution prediction results predicted by different methods on three different datasets are shown in Figure 7. Table 2 is a comparison of the super-resolution prediction scores of the TLGAN model.

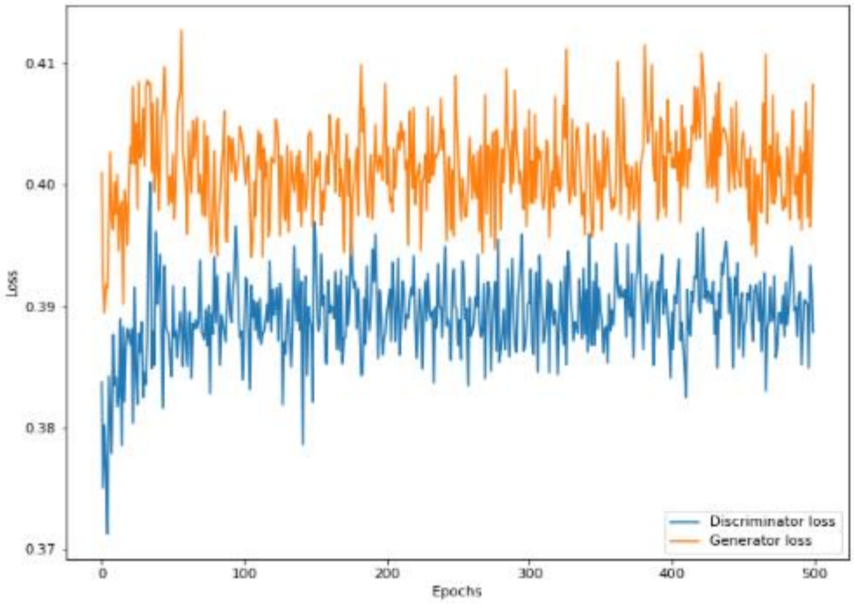


Figure 3. The trainling loss curve of the TLGAN model

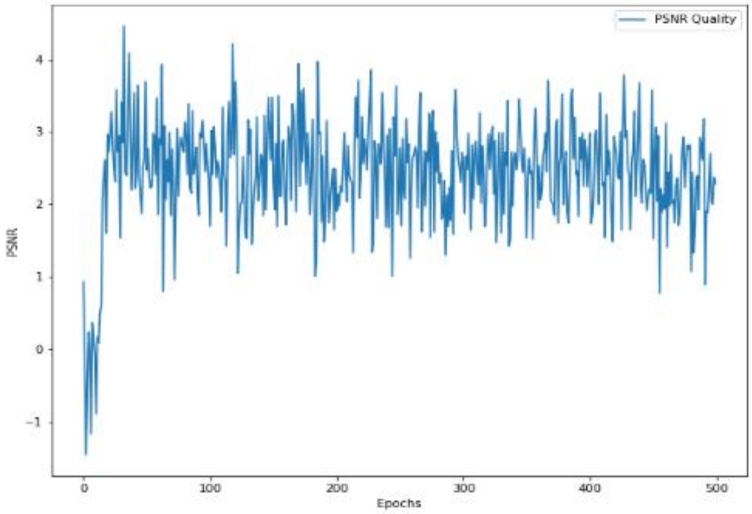


Figure 4. TLGAN model training PSNR curve

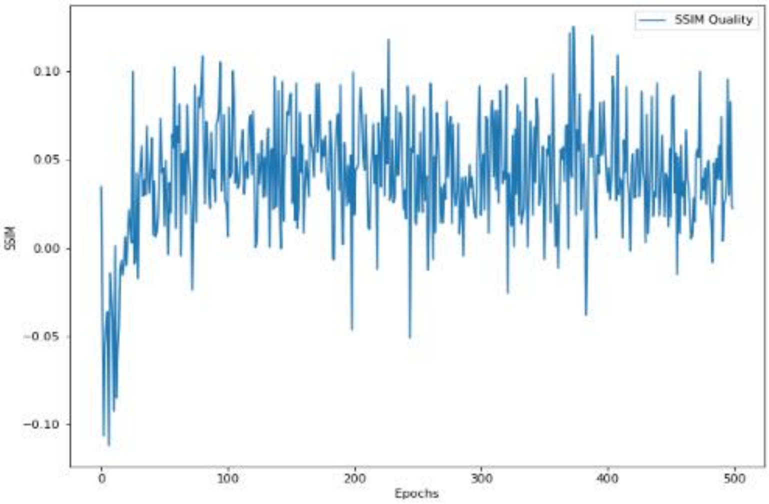


Figure 5. TLGAN model training SSIM curve

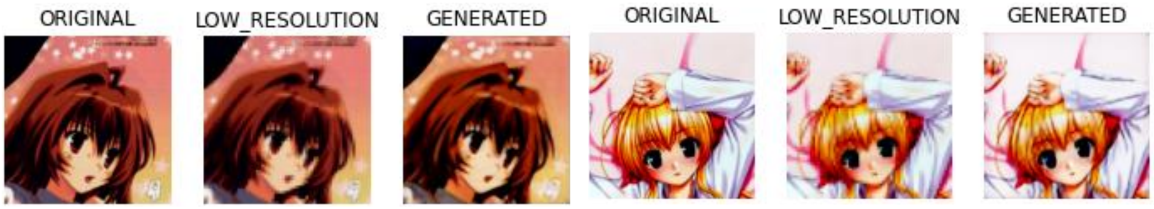
4.3. Experiment results analysis

Analysis of model training results. The training result of model training is directly related to the prediction result, so model training is essential. The details of the evaluation of model training are shown in Figures 3 to 5. Figure 3 reveals that at the beginning of training, the losses of both the generator and the discriminator are relatively low and have large fluctuations. As the training progresses, it stabilizes after about 50 iterations, the generator loss rate is about 0.40, and the determiner loss is about 0.39. Figures 4 and 5 are respectively the PSRN and SSIM of the SR and HR images generated by the model during the training phase. Figure 5 indicates that after 50 iterations of training, the SSIM numerical changes tend to be stable. These two evaluation indicators intuitively reflect the game process of the model training process generator and discriminator.

The visual effect evaluation analysis of the super-resolution model. The comparison of the original images, LR, and SR images generated by the model during training is shown in Figure 6. Among them, Figure 6(a) is the image generated during the model training process, and Figure 6(b) is the SR image generated by the generator. From the perspective of naked eye vision, the resolution image generated by the TLGAN model is very close to the real image, including texture details of the generated SR image. Experiments indicating that the model training results are excellent.

Robustness analysis of the TLGAN model. The prediction results of the TLGAN model on three different cartoon datasets are shown in Figure 7. At the same time, the prediction results of TLGAN, bicubic interpolation method, and SRGAN model are shown in Figure 7. Among them, LR is the model input image, and HR is the high-resolution image of the original image. First of all, our proposed model super-resolution results perform best in the naked-eye visual sensory evaluation indicators of SR images, which are specifically reflected in the high definition, natural texture, and light of SR images. The SR image generated by the TLGAN model is very close to the HR image of the original image.

However, experimental differences were noted in which the PSNR and SSIM indicators obtained by the bicubic interpolation method are generally higher in the three different data set tests, followed by the SRGAN method, and our model scores the lowest. Table 2 confirms this experimental result. The experimental results show that although our models' PSNR and SSIM scores are lower than the bicubic interpolation and SRGAN methods, the visual perception effect is the best. The super-resolution image obtained by our model can be close to the original HR image. This also representative that our model has better robustness in the prediction of cartoon characters' face super-resolution.



(a) (b)

Figure 6. Comparison of super-resolution samples reconstructed from the TLGAN model (a) is the image generated by the training process and (b) is the SR image during the test phase

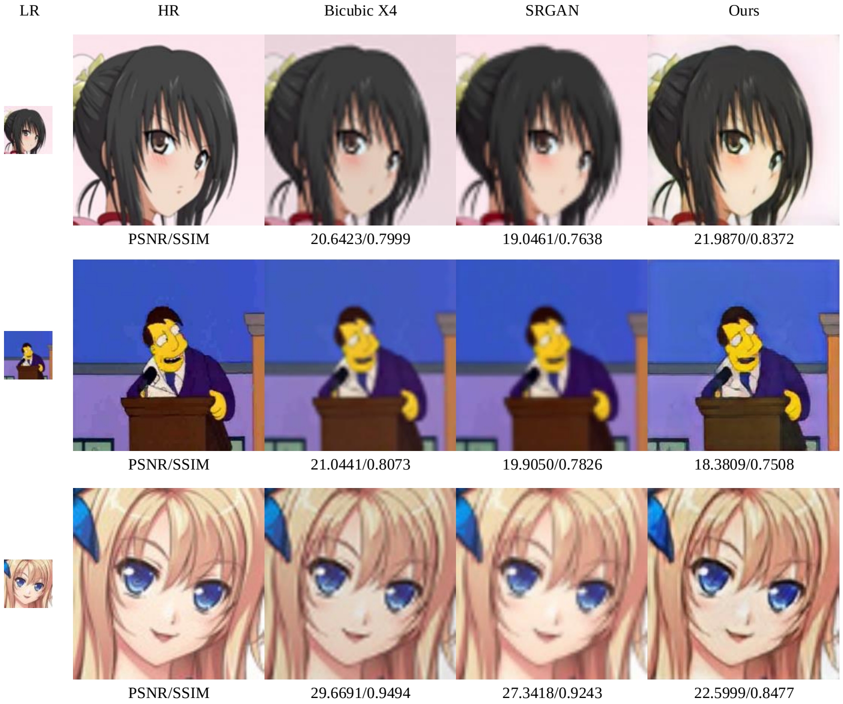


Figure 7. Sample evaluation results of the proposed model in different data sets

Table 2. Test scores of the proposed TLGAN model in different datasets

Dataset	Bicubic	SRGAN	TLGAN
	PSRN (dB)/SSIM	PSRN (dB)/SSIM	PSRN (dB)/SSIM
Anime-characters	21.4474/0.7800	19.8097/0.7289	19.9128/0.7493
Simpsons-characters	19.6024/0.6769	18.8215/0.6155	17.0138/0.5933
Anime face	24.6243/0.8471	23.2505/0.8121	19.9527/0.7720

5. CONCLUSION

This paper proposed an ISR model based on GAN network that uses transfer learning inspiration to improve and optimize the primary network. First, optimize the basic generative adversarial network structure. For example, cut off the BN layer in the generator network, and use upsampling to extract image features and use the LeckyReLU activation function determiner network and Sigmoid to achieve classification. Secondly, an extensive data set ImageNet pre-training model VGG19 is used to extract high-frequency features of images, accelerate model convergence and prevent the occurrence of gradient disappearance and gradient explosion problems. Finally, data expansion methods such as image flipping and clipping are used to fully increase the number of training samples and prevent model overfitting. We applied the proposed model to the super-resolution experiment of cartoon images, and carried out visual perception evaluation, PSNR and SSIM metric evaluation with the bicubic interpolation and SRGAN algorithm. Experimental results reveal that although the TLGAN model has low super-resolution PSNR and SSIM data in different data sets, it is best to evaluate visual perception. The super-resolution images produced by the proposed model have clear outlines and natural textures, which are almost close to the original high-definition images, which proves that the model has certain robustness. Since the PSNR and SSIM evaluation indicators of the models tested in different cartoon datasets are low, we think it may be caused by the uneven color distribution of cartoon images. In other words, it shows that the color of the input image is relatively single, and there is a partial monochrome phenomenon in the image. Therefore, we will use the proposed model for super-resolution research on real images to test the model prediction effect and portability in future work.

ACKNOWLEDGEMENTS

"This research was supported by the Ministry of Science, ICT and Future Planning (MISP), Korea, under the National Program for Excellence in SW supervised by the Institute for Information & communications Technology Promotion (IITP) (2023-0-00065)".

REFERENCES

[1] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014, doi: 10.1109/TMM.2014.2311320.

[2] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 1–8. doi: 10.1109/CVPR.2008.4587647.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010, doi: 10.1109/TIP.2010.2050625.

[4] R. Timofte, V. De Smet, and L. Van Gool, "A+: adjusted anchored neighborhood regression for fast super-resolution," in *ACCV 2014: Computer Vision -- ACCV 2014*, 2015, pp. 111–126. doi: 10.1007/978-3-319-16817-3_8.

[5] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1777–1781, Oct. 2014, doi: 10.1109/LGRS.2014.2308905.

[6] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1646–1654. doi: 10.1109/CVPR.2016.182.

[8] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 105–114. doi: 10.1109/CVPR.2017.19.

[9] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *ECCV 2014: Computer Vision – ECCV 2014*, 2014, pp. 63–78. doi: 10.1007/978-3-319-10584-0_5.

[10] O. Bowen and C.-S. Bouganis, "Real-time image super resolution using an FPGA," in *2008 International Conference on Field Programmable Logic and Applications*, 2008, pp. 89–94. doi: 10.1109/FPL.2008.4629913.

[11] I. Goodfellow, "NIPS 2016 tutorial: generative adversarial networks," *Preprint arXiv.1701.00160*, Dec. 2016.

[12] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," *Preprint arXiv.1805.08318*, May 2018.

[13] I. J. Goodfellow *et al.*, "Generative adversarial networks," *Prepr. arXiv.1406.2661*, Jun. 2014.

[14] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020, doi: 10.1109/TCSVT.2019.2920407.

[15] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Science Reviews*, vol. 232, Sep. 2022, doi: 10.1016/j.earscirev.2022.104110.

[16] Y. Chen, R. Xia, K. Yang, and K. Zou, "MFFN: image super-resolution via multi-level features fusion network," *The Visual Computer*, Feb. 2023, doi: 10.1007/s00371-023-02795-0.

[17] G. Wu, J. Jiang, and X. Liu, "A practical contrastive learning framework for single-image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2023, doi: 10.1109/TNNLS.2023.3290038.

[18] M. Zhou, K. Yan, J. Pan, W. Ren, Q. Xie, and X. Cao, "Memory-augmented deep unfolding network for guided image super-resolution," *International Journal of Computer Vision*, vol. 131, no. 1, pp. 215–242, Jan. 2023, doi: 10.1007/s11263-022-01699-1.

[19] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 448-45448–45.

[20] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *Preprint arXiv.1505.00853*, May 2015.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[22] G. Research, "An end-to-end open source machine learning platform," *TensorFlow*. https://tensorflow.google.cn/

[23] M. Abadi *et al.*, "TensorFlow: large-scale machine learning on heterogeneous systems," *Preprint arXiv.1603.04467*, Mar. 2016.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Preprint arXiv.1409.1556*, Sep. 2014.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[26] K. Yang, K. Qinami, L. Fei-Fei, J. Deng, and O. Russakovsky, "Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Jan. 2020, pp. 547–558. doi: 10.1145/3351095.3375709.

[27] Stanford Vision Lab, "ImageNet Dataset," *Image Net*. http://www.image-net.org (accessed Nov. 1, 2023).

[28] "Simpsons wiki," *Simpsons Wiki*. https://simpsons.fandom.com/wiki/Simpsons_Wiki (accessed Dec. 11, 2023).

[29] "Animeface-2009," 2009. https://github.com/nagadomi/animeface-2009 (accessed Dec. 12, 2023).

BIOGRAPHIES OF AUTHORS

Guangxing Wang received his M.S. degree in Computer Application Technology from Huazhong University of Science and Technology, Wuhan, China in 2009. From 2016 to the present, he has been an associate professor in the school of Computer and Big Data Science, Jiujiang University in China. received his Ph.D. degrees from the Dept. of Computer Information Engineering of Kunsan National University, Gunsan, Korea, in 2021. His research interests deep learning, image processing, and diagnosis. He can be contacted at email: wanggxrs@163.com.

Seong-Yoon Shin received his M.S. and Ph.D. degrees from the Department of Computer Information Engineering of Kunsan National University, Gunsan, Korea, in 1997 and 2003, respectively. From 2006 to the present, he has been a professor in the School of Software. His research interests include image processing, computer vision, and virtual reality. He can be contacted at email: s3397220@kunsan.ac.kr.

Jong-Chan Kim received his B.S. degree from Sunchon National University in 2000, his M.S degree from Department of Computer Science, Sunchon National University in 2002, his Ph.D. degree from Department of Computer Science, Sunchon National University in 2007. Full-time lecturer at Dong-Eui University, Department of Imaging Information Engineering (2007-2010). And a senior researcher professor of the Automation and System Research Institute at Seoul National University in 2013. Director, Korea Computer Graphics Society (2013-2020). Journal editor and director of the Korean Digital Contents Society (2018-2020). Journal editor of the Korea Multimedia Society (2012-2020). Vice Chairman of General Affairs in the Korea Convergence Software Society 2020. Chairman of Organizing Committee for Fall Conference in the Korea Multimedia Society 2020. Currently he is Assistant Professor (Tenure Track) of Department of Computer Engineering, Sunchon National University, Republic of Korea (Sep. 2021). He is the editor of the Processes special issue, and his current research interests are image processing, computer graphics, automatic control, machine learning, deep learning, computer vision, data analysis and data prediction. He can be contacted at email: seaghost@scnu.ac.kr.