Super-Resolution Image Generation for Diabetic Retinopathy Detection by SRGAN

Yi Zhao[a]

Data Science, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

Keywords: Super-Resolution Generative Adversarial Network, Diabetic Retinopathy Images, Medical Image Augmentation, Super-Resolution.

Abstract: As computer vision technology progresses, the Super-Resolution method is essential in medical image enhancement. In this article, Super-Resolution Generative Adversarial Network (SRGAN) is trained to produce high-resolution diabetic retinopathy images, aiming to assist numerous model training processes such as U-net and ResU-net. As a result of improving the original SRGAN framework, the resolution and quality of images reach a higher level, capturing more detailed information. Through this way, segmentation models can more accurately determine the location of lesions and tumor nodules, which enables early disease prediction and precise localization. Nowadays, many advanced segmentation studies work with high-resolution processing of medical images. The experiment results indicate that SRGAN has commendable efficacy in the APTOS-2019 dataset, achieving PSNR-43 SSIM-0.93, Precision-0.965, Recall-0.913, F1-Score-0.937, which demonstrates its superiority in detail restoration. SRGAN provides strong support for subsequent disease detection tasks, definitely facilitating more accurate diagnostic outcomes and ascending the reliability of medical image analysis.

1 INTRODUCTION

Super-resolution technology is capable of creating high-resolution pictures from low-resolution ones, tremendously improving the details of medical images. This technology provides more precise visual information for clinical analysis and diagnosis, especially diabetic retinopathy (DR) which requires clearer images to train the segment model. However, it is challenging for traditional methods to recover information from complex pictures, because of the limits of resolution increase and loss of detail.

The SRGAN was first presented in 2017. It combined traditional pixel-level loss and perceptual loss, utilizing a generative adversarial network framework. This model successfully enhances single-image super-resolution (Ledig, et al, 2017). To make the medical images look more realistic, SRGAN ascends the perceptual quality of generated images through adversarial training, meanwhile, maintaining high fidelity of detail and structure. Researchers reviewed the research on GANs and pointed out the pattern collapse problems, a kind of training instability that may occur during the training process of GANs (Gonog and Zhou, 2019). After that, many schemes were proposed to enhance the performance of SRGAN. For instance, scientists constructed the super-resolution algorithm that is based on the attention mechanism to better focus on key information areas of the image (Liu and Chen, 2021). This method achieved excellent results in MRI, CT, and retinal imaging. A more advanced SRGAN-based super-resolution method was created for CT images, optimizing network architecture to create certain medical photos (Jiang et al, 2020). Super-resolution residual network model can assist researchers in recovering high-level image semantic information (Abbas and Gu, 2023). In addition, two researchers discussed the challenges of adversarial training, suggesting some possible solutions, such as utilizing techniques for multimodal generation and enhancing the robustness of the model (Sajeeda and Hossain, 2022).

The core of this research is how to enhance the model quality based on SRGAN to provide more efficient and reliable data support for subsequent model training and clinical applications. In this

[a] https://orcid.org/0009-0000-6769-464X

Zhao, Y.
Super-Resolution Image Generation for Diabetic Retinopathy Detection by SRGAN.
DOI: 10.5220/0013698200004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 400-403
ISBN: 978-989-758-765-8

research, the powerful VGG-19 is used as the discriminator to better recognize the image realism. In the training process, the generator is trained for 50 epochs first, avoiding the mismatch problem between the generator and the discriminator. The code was uploaded to GitHub: https://github.com/ZhaoYi-10-13/Super-Resolution-Image-Generation-for-APTOS-2019-Dataset-Based-on-SRGAN

2 METHODOLOGIES

This section illustrates the SRGAN framework utilized to enhance diabetic retinopathy fundus images. This approach uses residual learning, a modified discriminator built upon a pre-trained VGG-19 network, and comprehensive data augmentation techniques to generate high-resolution images. These SR pictures will serve for clinical image analysis and diagnosis. In Figure 1, SRGAN dramatically increases the quality of fundus images.

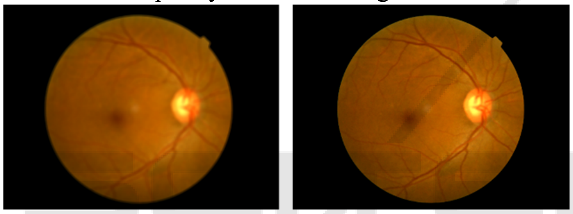

Figure 1: Comparison of fundus images before and after SRGAN processing. (Picture credit: Original)

2.1 Overview of SRGAN

The proposed SRGAN framework contains a generator and a discriminator. Generator G transforms a LR diabetic retinopathy image into SR images:

$$I^{SR} = G(I^{LR}) \tag{1}$$

In this model, discriminator D is replaced by pre-trained VGG-19, which helps to capture subtle textural and structural features in medical images, meanwhile, the adversarial training framework is formulated in a minimax method:

$$\min_G \max_D E_{I^{HR}\sim p_{data}(I^{HR})}[\log D(I^{HR})] + E_{I^{LR}\sim p_{data}(I^{LR})}\left[\log\left(1 - D\left(G(I^{LR})\right)\right)\right] \tag{2}$$

To assist SR images in maintaining important perceptual details, we incorporate a perceptual content loss computed using feature maps from the VGG-19 network:

$$L_{content} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G(I^{LR})) \right) \tag{3}$$

where $\Phi_{i,j}$ represents the feature map which is extracted from the j-th convolution layer before the i-th pooling layer in the VGG-19 model. $W_{i,j}$, $H_{i,j}$ are the feature map dimensions. Table 1 shows the structure of VGG-19.

Table 1: Structure of VGG-19

Layer	Output Shape	Parameters
Conv_2D	(224 224 64)	1792
Conv_2D	(224 224 64)	36928
MaxPooling_2D	(112 112 64)	0
Conv_2D	(112 112 128)	73856
Conv_2D	(112 112 128)	147584
MaxPooling_2D	(56 56 128)	0
Conv_2D	(56 56 256)	295168
Conv_2D	(56 56 256)	590080
Conv_2D	(56 56 256)	590080
……	……	……
Flatten	(25088)	0
Dense_1	(4096)	16781312
Dense_2	(1000)	4097000

The whole generator loss is a weighted sum of the content and adversarial losses:

$$L_G = L_{content} + \lambda L_{GAN} \tag{4}$$

And the adversarial loss defined as:

$$L_{GAN} = -\log\left(D\left(G(I^{LR})\right)\right) \tag{5}$$

where λ serves as a hyperparameter balancing the two terms.

2.2 Key Improvements

The improvements used in this SRGAN framework for enhancing diabetic retinopathy images are shown here:

1. Enhanced Residual Learning: The generator G contains numerous deep residual blocks so that it can effectively learn the mapping from LR to HR images. Each residual block is formulated as:

$$F(x) = x + H(x) \tag{6}$$

where $H(x)$is the residual function of the block.

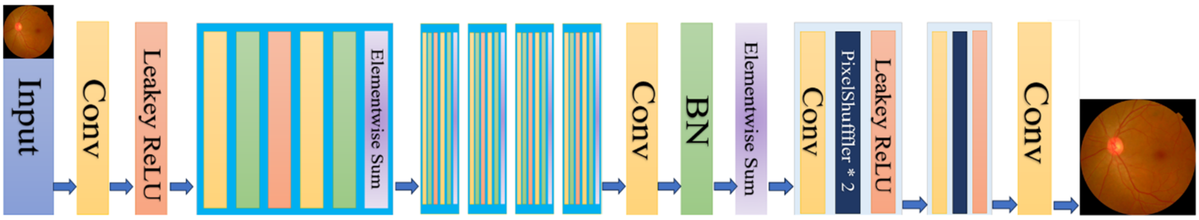

Figure 2: Structure of Generator (Picture credit: Original)

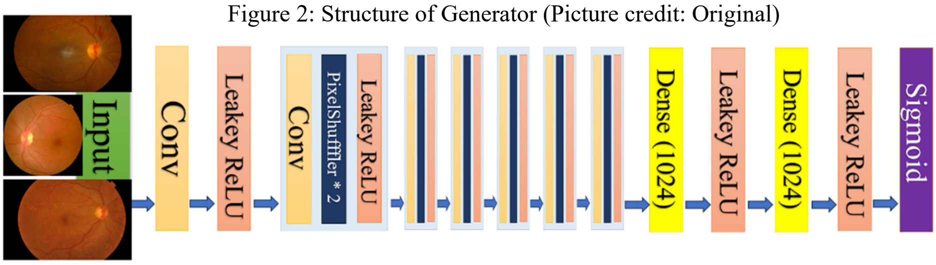

Figure 3: Structure of Discriminator (Picture credit: Original)

2. Modified Discriminator Architecture: This SRGAN model utilizes VGG-19 network with additional convolutional layers, and its architecture improves the network's capture of richer semantic features in diabetic retinopathy images.

3. Perceptual Loss Optimization: By computing the loss in the deep feature space of the VGG network, the images maintain more high-frequency details and realistic textures.

2.3 Model Structure

In Figure 2, the generator consists of multiple components, including convolutional layers, Leaky ReLU activations, element-wise summation, and batch normalization. Every layer inherits the parameter from previous parts and outputs the final SR images.

In Figure 3, the discriminator receives various images and processes them through a series of blocks, which were composed of two PixelShuffler layers, convolutional layers, and Leaky ReLU activations. After that, the extracted features are passed to a dense layer. Then, a sigmoid activation outputs the final probability to determine whether the image is real or generated.

3 EXPERIMENTS AND RESULTS

3.1 Data Preprocessing and Augmentation

Robust data preprocessing and augmentation are essential in model training, so each input image I is normalized as follows:

$$I_{norm} = \frac{I-\mu}{\sigma} \quad (7)$$

where μ and σ denote the mean and standard deviation of the training set, respectively.

Data augmentation is performed via a set of transformations T, to further enhance the diversity of the dataset and improve model robustness:

$$I_{aug}^{LR} = T(I^{LR}), I_{aug}^{HR} = T(I^{HR}) \quad (8)$$

with the augmentation set defined as:

$$\{\text{random crop}, \text{horizontal flip}, \text{vertical flip}, \text{rotation}\}$$

3.2 Training Process

3.2.1 Stage 1: Pre-training the Generator

To mitigate the mismatch between the generator and discriminator, the generator is pre-trained for 50 epochs using only the content loss.

This stage allows G to learn an initial mapping from LR to HR images perceptually and meaningfully.

3.2.2 Stage 2: Adversarial Training

Following pre-training, adversarial training is initiated. In each training iteration, the parameters of both the generator G and the modified VGG-19-based discriminator are updated using gradient descent:

$$\theta \leftarrow \theta - \eta \nabla_\theta L \quad (9)$$

where η is the learning rate, and the discriminator loss is defined as:

$$L_D = -E_{I^{HR} \sim p_{data}(I^{HR})}[\log D(I^{HR})] + E_{I^{LR} \sim p_{data}(I^{LR})}\left[\log\left(1 - D\left(G(I^{LR})\right)\right)\right] \quad (10)$$

3.3 Evaluation Metrics

The performance of our SRGAN framework is evaluated using both image quality metrics and downstream task metrics relevant to diabetic retinopathy analysis.

(1) Structural Similarity Index:

$$SSIM(I^{HR}, I^{SR}) = \frac{\left(2\mu_{I^{HR}}\mu_{I^{SR}} + C_1\right)\left(2\sigma_{I^{HR}I^{SR}} + C_2\right)}{\left(\mu^2_{I^{HR}} + \mu^2_{I^{SR}} + C_1\right)\left(\sigma^2_{I^{HR}} + \sigma^2_{I^{SR}} + C_2\right)} \quad (11)$$

Where μ and σ , $\sigma_{I^{HR}I^{SR}}$ denote the means, variances, and covariance of the images, and C_1, C_2 are constants for stability.

(2) Peak Signal to Noise Ratio

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

Where MAX is the maximum pixel value.

Table 2: The metrics of the model

Metric	Value
PSNR_dB	43.00
SSIM	0.93
Precision	0.965
Recall	0.913
F1-Score	0.937

Table 2 indicates that SRGAN performed exceptionally well in APTOS-2019, because of the outstanding PSNR_dB, SSIM and Precision Recall F1-score.

4 CONCLUSION

APTOS-2019 dataset includes 3662 diabetic retinopathy fundus images, which are used to downsample the LR images for model training. This method simulates the clinical image-generating challenges. The generator was first trained for 50 epochs to prevent the model from collapsing, after that, both G and D performed the adversarial training with a learning rate of 0.0001 and 500 epochs. Aiming to increase the model robustness, data augmentation is utilized at the beginning of training, such as random cropping, horizontal and vertical flipping, and rotation (every image is coped with the normalization process)

SRGAN is specially designed for generating SR retinopathy fundus images, because of the VGG-19-based discriminator. It greatly enhances the resolution of images and recreates the details. From the experimental results, SRGAN has a potential clinical image analysis application value, especially in segment.

Although SRGAN performs well in this case, there are still some challenges: certain areas in other pictures may appear blurred, and low-level detail reconstruction; Significant computational resources are used in training that obstacles to wider application; This dataset has limited variability of diabetic retinopathy images, which may limit the generalization ability of the model.

That is the reason why future research will focus on improving the model generalization ability and making it easier to be trained. In the end, integrating SRGAN will aid in early detection and intervention, aiming to develop medical technology.

REFERENCES

Abbas, R., Gu, N., 2023. Improving deep learning-based image super-resolution with residual learning and perceptual loss using SRGAN model. Soft Computing, 27(21), 16041-16057.

Deng, Z., Zhang, H., Liang, X., Yang, L., Xu, S., Zhu, J., & Xing, E. P., 2017. Structured generative adversarial networks. In Advances in neural information processing systems, 30.

Elanwar, R., Betke, M., 2024. Generative adversarial networks for handwriting image generation: a review. The Visual Computer, 1-24.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y., 2020. Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

Gonog, L., & Zhou, Y., 2019. A review: generative adversarial networks. In 2019 14th IEEE Conference on Industrial Electronics and Applications (ICTEA) (pp. 505-510). IEEE.

Jiang, X., Xu, Y., Wei, P., & Zhou, Z., 2020. CT image super resolution based on improved SRGAN. In 2020 5th International Conference on Computer and Communication Systems (ICCCS) (pp. 363-367). IEEE.

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

Liu, B., & Chen, J., 2021. A super resolution algorithm based on attention mechanism and SRGAN network. IEEE Access, 9, 139138-139145.

Sajeeda, A., & Hossain, B. M., 2022. Exploring generative adversarial networks and adversarial training. International Journal of Cognitive Computing in Engineering, 3, 78-89.

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F. Y., 2017. Generative adversarial networks: introduction and outlook. IEEE/CAA Journal of Automatica Sinica, 4(4), 588-598.