Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 1, 2023

Adversarial Training of SwinIR Model for Face Super-Resolution Processing

Hongjun Lin *

School of Computing, China University of Geosciences, Wuhan, 430074, China
*** Corresponding author Email:** lin2669875691@gmail.com

Abstract: This research aims to apply the SwinIR model to perform face image super-resolution processing using adversarial training techniques, thereby enhancing facial image features. With the rapid development of computer vision and artificial intelligence technologies, face image super-resolution processing plays a crucial role in improving the accuracy and performance of facial recognition and related applications. Firstly, we introduce the basic principles of adversarial training and provide a detailed overview of the architecture and characteristics of the SwinIR model. This model demonstrates excellent performance in super-resolution tasks, exhibiting high feature extraction and image reconstruction capabilities. Next, we describe the experimental design and dataset selection. Through extensive experiments, we compare the quality and feature representation of face images before and after super-resolution. The results show that after undergoing super-resolution processing with the SwinIR model, facial images exhibit significantly enhanced details and edges, leading to a notable improvement in image features. In summary, this research successfully enhances the feature information of facial images by applying the SwinIR model through adversarial training, effectively improving the details of the face. This research outcome is of significant importance in advancing the field of computer vision and enhancing the efficiency and accuracy of artificial intelligence technologies in practical applications.

Keywords: SwinIR; Super Resolution; Adversarial Training; Face Image.

1. Introduction

Image super-resolution is an important research direction in the field of computer vision, aiming to restore low-resolution (LR) images to high-resolution (HR) ones. With the widespread application of digital images in daily life, such as image acquisition, transmission, and storage, high-quality image presentation is critical for various application scenarios [1]. However, due to limitations in hardware devices, sensor technologies, or network bandwidth, the images obtained in many cases often have low resolution, resulting in insufficient image details and poor visual quality.

To address this issue, image super-resolution techniques have been proposed to enhance the visual quality and usability of low-resolution images by restoring lost details and information. Traditional image super-resolution methods are mainly based on interpolation techniques or statistical methods [2], but they have limited effectiveness in handling complex textures and high-frequency details. With the rise of deep learning, particularly the development of deep convolutional neural networks (DCNNs), significant progress has been made in the field of image super-resolution [3].

Eye-tracking technology has wide applications in computer vision and human-computer interaction, allowing for precise monitoring and tracking of eye movements during visual tasks [4]. However, to achieve accurate eye-tracking, high-resolution facial images with sufficient details are required. Traditionally, high-quality facial images typically require the use of professional high-end devices, limiting the application of eye-tracking technology on more widely accessible devices [5].

To overcome this limitation, the aim of this study is to use more accessible and easily obtainable devices to provide facial images with sufficient details for eye-tracking technology. For this purpose, we chose to adopt adversarial training techniques and the high-performing SwinIR model for face image super-resolution processing.

The research problem is to explore the feasibility and effectiveness of using the SwinIR model for face image super-resolution processing. We hope that this technique can significantly enhance the quality of low-resolution facial images captured by widely accessible devices, providing sufficient details and clarity to offer more reliable and accurate input data for eye-tracking technology.

The significance of this research lies in its potential to enable eye-tracking technology on widely accessible devices. By applying adversarial training techniques and the SwinIR model, we aim to provide high-quality facial images with fine details for ordinary cameras or mobile devices, thus promoting the application and popularization of eye-tracking technology in a broader range of scenarios.

In conclusion, this study aims to explore the method of using the SwinIR model for face image super-resolution processing and, by providing images with sufficient details, advance the application of eye-tracking technology on widely accessible devices. This research is expected to offer a better visual experience for ordinary users and contribute new practical techniques to the development of computer vision. In this paper, we mainly discuss the super-resolution processing of facial images to obtain images with more details.

2. Related Work

2.1. SwinIR [6]

SwinIR is a deep learning-based image super-resolution method that utilizes the Swin Transformer as its underlying model. The Transformer is a model widely used for natural language processing tasks, but due to its efficient self-attention mechanism, it has also been applied in the field of image processing. Swin Transformer effectively captures

long-range dependencies and semantic information in images through its global self-attention mechanism.

SwinIR model adjusts and optimizes the Swin Transformer specifically for image super-resolution tasks. In image super-resolution, SwinIR takes low-resolution images as input and outputs high-resolution image results through network layers' processing. Its core advantage lies in efficiently extracting image feature information and possessing strong representational capabilities, enabling it to effectively handle complex textures and details, resulting in excellent super-resolution performance.

SwinIR adopts the Swin Transformer as the basic network architecture, utilizing the self-attention mechanism to capture long-range dependencies and semantic information in images. By stacking multiple Swin Transformer modules, SwinIR can efficiently extract image feature information and restore low-resolution images to high-resolution ones.

SwinIR has achieved significant performance improvements in image super-resolution tasks and has become one of the leading image super-resolution methods. Its outstanding performance has garnered attention in various application scenarios, and it has been widely applied in both academic and industrial settings.

2.2. WGAN [7]

WGAN (Wasserstein Generative Adversarial Network) is a variant of Generative Adversarial Networks (GANs) that has been improved and optimized based on GANs. GAN is a deep learning model used to generate new samples, consisting of a generator and a discriminator, which are trained in an adversarial manner to progressively generate more realistic samples.

WGAN addresses the instability and mode collapse issues during GAN training by introducing Wasserstein distance. Wasserstein distance provides better continuity and smoothness when measuring the distance between two probability distributions, resulting in a more stable and reliable training process.

The emergence of WGAN offers new ideas and methods for generative models in deep learning. It has achieved excellent results in tasks such as image generation, image restoration, and image super-resolution, attracting widespread attention across various domains.

In conclusion, SwinIR and WGAN are significant representative methods in the fields of image super-resolution and generative adversarial networks, respectively, and have made significant progress in image processing and generation tasks. By combining the two approaches, it is expected to achieve even better model performance.

3. Method

3.1. Dataset

In order to study image super-resolution and denoising techniques in computer vision, we need to construct a dataset suitable for this task. Considering our research goal is to simulate the effects of most cameras, we will use a high-definition face dataset and apply resolution reduction and noise addition to simulate the image capture process in real-world scenarios.

Firstly, we will select a high-quality dataset containing a large number of high-definition face images as the base dataset. This dataset should include multiple face images with high image quality to ensure that our training data possesses good features and details.

Next, we will process these high-definition face images using resolution reduction techniques. We can use interpolation methods or predefined blur kernels to simulate the resolution reduction effect during image capture with cameras. This will provide us with a set of low-resolution face images to train the super-resolution model.

To simulate camera noise effects, we can introduce different types of noise, such as Gaussian noise, salt-and-pepper noise, etc. These noise types will be added to both high-definition and low-resolution face images to mimic the noise interference during image capture in real-world scenarios.

When constructing the dataset, we also need to consider data balance and diversity. We can ensure that the dataset contains face images of various ages, genders, races, and expressions to improve the model's generalization ability and robustness.

Finally, we will choose appropriate deep learning models, such as SwinIR and WGAN, for training and testing based on the size and complexity of the dataset. By using such a dataset, we can better evaluate the performance of the models in simulating the real image capture process, providing valuable experimental data and conclusions for research in image super-resolution and denoising techniques in computer vision.

3.2. Network Structure

In this study, we apply adversarial training techniques to the SwinIR model to further enhance image super-resolution performance. SwinIR is an image super-resolution model based on the Swin Transformer, while adversarial training is a powerful deep learning technique commonly used in GANs to optimize the model through the adversarial process between the generator and discriminator.

To improve SwinIR's performance, we introduce adversarial training. Adversarial training consists of two parts: the generator, which is the SwinIR model itself responsible for image super-resolution generation, and the discriminator, which is a binary classification network aiming to distinguish between the high-resolution images generated by the generator and real high-resolution images. During adversarial training, the discriminator learns to differentiate between real and generated images, while the generator minimizes the adversarial loss to gradually generate more realistic and high-quality high-resolution images.

By incorporating adversarial training techniques into the training process of the SwinIR model, we expect the generator to receive feedback from the discriminator, further improving the quality of generated high-resolution images. The introduction of adversarial training is expected to enhance the expressive power of the SwinIR model, improve its performance in handling complex textures and details, and achieve superior image super-resolution results.

In summary, this study aims to explore new training strategies in the field of image super-resolution by applying adversarial training techniques to the SwinIR model to enhance its performance and effectiveness. The design of this model will combine the characteristics of SwinIR and the advantages of adversarial training, bringing new breakthroughs and improvements to image super-resolution tasks.

3.3. Training Process

In each iteration, the training data loader is traversed, and

based on the current training step, the learning rate is updated to adjust the learning rate during the training process. The initial learning rate is set to 0.0001, and the batch size is set to 16. The training data is then fed into the SwinIR model. The parameters of the generator and discriminator are optimized to gradually learn better image super-resolution effects. Adversarial training techniques play a crucial role in the optimization process, as they optimize the model through the adversarial process between the generator and discriminator. At regular training steps, training information is recorded, including the current epoch number, training step, current learning rate, loss values, etc., and the model parameters are saved for subsequent testing and continued training. A test is also conducted, where test images from the test set are fed into the SwinIR model to obtain super-resolution images, and the Peak Signal-to-Noise Ratio (PSNR) value between the reconstructed images and real high-resolution images is calculated as an evaluation metric for image reconstruction quality.

Through the above training process, we combine adversarial training techniques with the SwinIR model and optimize the parameters of the super-resolution model through extensive data training. This enables the model to achieve better image reconstruction results on widely used devices, providing more detailed image features for eye-tracking technology and facial recognition in various fields.

4. Experience

In this section, we will present the experimental results of the SwinIR model using adversarial training for the task of face image super-resolution. We utilize a high-resolution face dataset from computer vision, which is processed by downsampling the resolution and adding noise to simulate the image capture scenarios of most cameras. These datasets are used to train the model.

4.1. Training Convergence

During the training process, we recorded the loss values and PSNR metrics for each training epoch to observe the convergence of the training. The following two figures illustrate the loss of the generator and the PSNR (Peak Signal-to-Noise Ratio) metric of the model during the training process. As shown in Figure 1, it can be observed that with the progress of iterations, the loss of the generator gradually decreases, converging around 0.01, and eventually stabilizes at approximately 0.01. In Figure 2, as the generator loss decreases, the PSNR also increases gradually, ultimately stabilizing at around 40.6.

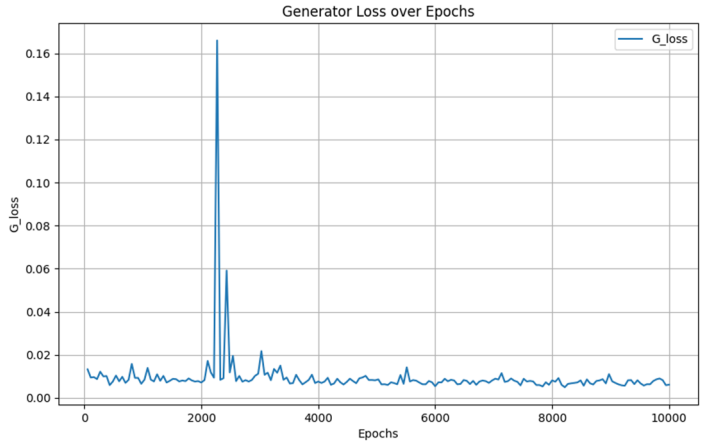

Figure 1. As the epochs progress, the loss of the generator decreases.

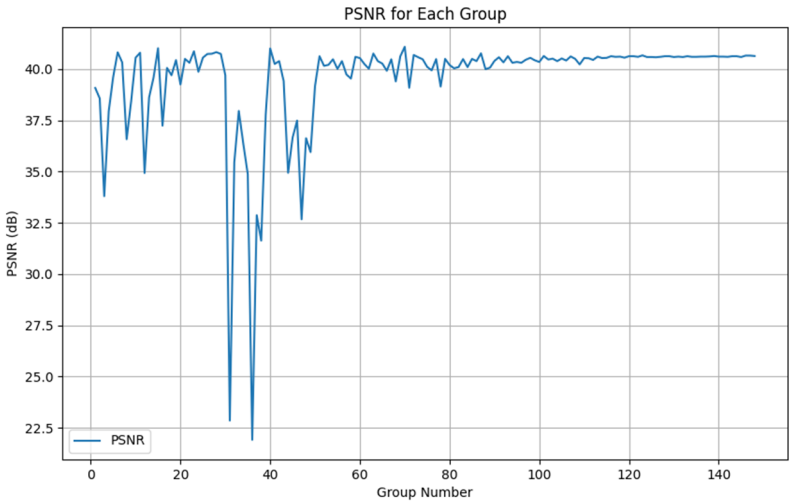

Figure 2. During the training process, the PSNR increases and eventually stabilizes.

During the training process, we recorded the loss values and PSNR metric for each training epoch to observe the convergence. The charts show that as the training progresses, the loss values gradually decrease, and the PSNR metric

improves. This indicates that the model is gradually learning better super-resolution capabilities.

In the initial stages of training, the loss values exhibit fluctuations around 2000 iterations, gradually decreasing to approximately 0.17. This phase may be attributed to the model adapting to the training data and learning some initial features. As the training continues, the model optimizes further, and the loss values stabilize at a lower level, eventually reaching around 0.1. This indicates a significant improvement in the reconstruction performance of the model on the training data.

At the same time, the PSNR metric experiences a sharp decline around 2000 iterations, decreasing from its initial value to around 22 or lower. This phenomenon may be caused by overfitting in the early stages of training or encountering some local optima, leading to a degradation in the quality of reconstructed images. However, with continued training, the model gradually overcomes these issues, and the PSNR metric starts to improve, stabilizing at around 40. This signifies that the model has made significant progress in the super-resolution task and can generate higher-quality, clearer, and more detailed images.

4.2. Application Validation

To validate the performance of our model in real-world applications, we applied the trained SwinIR model for face image super-resolution processing on widely used devices. The experimental results demonstrate that our model can effectively enhance facial image features and provide clearer and more detailed images, thereby offering a better image foundation for applications such as eye-tracking technology and facial recognition in various domains.

Figure 3. Training Results of the Model (×2)

Overall, the SwinIR model trained through adversarial training exhibits superior performance in face image super-resolution, showcasing robust generalization capabilities and practical value. Its successful application provides novel ideas and methods for the advancement and application of image super-resolution techniques.

5. Conclusion

Based on the results of this experiment, we have successfully transformed low-resolution facial images into high-resolution images while preserving crucial details, significantly enhancing image clarity and quality. This achievement provides vital support and assistance for implementing eye-tracking technology on widely-used devices.

However, it is essential to acknowledge certain limitations in this experiment. We only focused on ×2 image super-resolution conversion, which implies some constraints in practical applications. In more challenging tasks with higher scaling factors, our model might encounter greater difficulties and potential performance degradation. To enhance the model's generalization capabilities and applicability further, future research can explore more in-depth into ×4, ×8, and other higher scaling factors for image super-resolution conversion.

Nevertheless, within the same specification of image super-resolution conversion, our model demonstrated remarkable results. By reconstructing low-resolution images with high quality, we provide robust image foundations for widely-used devices, improving image processing capabilities and performance for applications like eye-tracking technology.

In conclusion, this experiment contributes valuable insights and practices to image super-resolution processing while uncovering potential research directions. We believe that through further optimization and improvement of the model, we can achieve even better results in tasks with higher scaling factors, providing reliable and efficient solutions for eye-tracking and other image processing tasks in practical applications.

References

[1] Yang, W., Zhou, F., Zhu, R., Fukui, K., Wang, G., & Xue, J. H. (2019). Deep learning for image super-resolution. Neurocomputing.

[2] 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.

[3] 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.

[4] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

[5] Harezlak, K., & Kasprowski, P. (2018). Application of eye tracking in medicine: A survey, research issues and challenges. Computerized Medical Imaging and Graphics, 65, 176-190.

[6] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1833-1844).

[7] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. Advances in neural information processing systems, 30.