Super-Resolution Analysis of Animal Images Based on ESRGAN Model

Shaoxu Li
College of Computer Science and Technology, Beijing Jiao Tong University, Weihai, China

Keywords: Super-Resolution, Animal, Images, Training Dataset.

Abstract: Image super-resolution (SR) plays a crucial role in enhancing the quality of images for society. This study introduces an Enhanced Super Resolution Generative Adversarial Network (ESRGAN) model designed specifically for improving the resolution of animal images. The objective of this paper is to explore the effect of different training datasets on the SR effect of specific animal target datasets by studying the models generated from different types of animal training datasets and similar animal training datasets. In addition, the effects of different types of animal image datasets on the performance of ESRGAN models are analyzed. Training datasets of multiple animal species are used to train different models which are trained under the same loss function. In addition, the target dataset is subjected to SR processing of species-specific animal images in this experiment to verify the effectiveness of this model in real-world applications. Finally, this study emphasizes the key role of dataset selection in the performance enhancement of ESRGAN models. This method provides an effective tool in the field of animal image processing that can be applied to a variety of real-world scenarios, thus contributing to the development of animal conservation, medical imaging, and scientific research.

1 INTRODUCTION

Image super-resolution (SR) has become more and more important with the development of image technology.SR technology can improve low-resolution (LR) images to high-resolution (HR) images, providing clearer and more detailed image information for various application scenarios (Sharma and Shrivastava 20222). With the wide application of SR technology in medicine, photo beautification, celestial body research, and so on (Frid-Adar et al 2018, Yang et al 2019 & Schawinski et al 2017). SR is becoming more important for its practical application. This study wants to explore the application of Enhanced Super Resolution Generative Adversarial Network (ESRGAN) technology in animal image SR. Animal images are of great significance in the field of ecology, animal behavior research and medicine. Due to the limitations of field environment and equipment, the acquired animal images are often limited by the resolution, resulting in information loss and analysis difficulties. By improving the resolution of animal images, the study can more accurately study and understand the ecological habits, behavior patterns and health status of animals, which provides a tool for the protection and research of animals.

In the field of image SR, researchers have proposed many methods, including traditional interpolation methods, deep learning-based methods and Generative Adversarial networks (GAN) (Keys 1981, Sharma and Shrivastava 2022 & Favorskaya and Pakhirka 2023). ESRGAN improves SRGAN by introducing composed of residual-in-residual Dense blocks (RRDB) architecture (Wang et al 2018). In recent years, ESRGAN technology, as a variant of GANs, has made significant breakthroughs and is widely used in image SR tasks. At the same time, many researches have studied the processing methods of animal images (Yang et al 2008). By integrating previous work and technological developments, the study can better understand the trends and limitations of the current field.

The main objective of this study is to introduce GAN technique to solve the animal image SR problem. The training effect of different animal image datasets is explored by introducing ESRGAN technique. Specifically, this experiment trains two models using species-specific animal image training set and mixed-species animal image training set

Li, S.
Super-Resolution Analysis of Animal Images Based on ESRGAN Model.
DOI: 10.5220/0012804200003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 433-437
ISBN: 978-989-758-705-4

respectively, and then compares their hyper-segmentation effects on species-specific animal images. Finally, this study explores the practical implications of this research through quantitative and qualitative analysis of the experimental results. The results show that the proposed model can effectively perform animal image SR. This study is expected to provide a tool for ecologists, animal behaviorists and medical researchers to help them study and analyze animal images more accurately.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The purpose of this study is to explore the hyper-segmentation effect of models trained on different animal datasets (Dataset). The training dataset for the first model in this study is a total of 100 images including various animals. The training dataset for the second model is 100 images of a single animal. To ensure generalization of the training models, the images are randomly cropped to specified sizes in this study before processing the original images. The images are additionally randomly horizontally inverted and rotated by 90 degrees with a 50% probability. In order to better compare the experimental results, the images provide with the dataset are processed in low resolution and high resolution in this experiment. In both cases, the image data is resized to the specified size to ensure that the generated images have the desired dimensions.

2.2 Proposed Approach

The main goal of this research is to explore how to maximize the SR effect of animal images using datasets of different species of animals. The research method of this study is to first define the model based on ESRGAN, then train the model separately with two different datasets. The first dataset contains various species of animals, and the second dataset contains only a single species of animal. In the training process, the study plot the change of loss value, and finally import the test set for testing and compare the results of various indicators. The research process is shown in the following Figure 1.

2.2.1 ESRGAN

The architecture used in this experiment is based on the ESRGAN network model. The model architecture of ESRGAN includes a generator and a discriminator (Jiang 2022). The generator accepts a LR input image and gradually improves the image resolution through multi-layer convolution and residual blocks. Its goal is to generate HR images with more details. The discriminator distinguishes between the real HR images and images generated by the generator. The effect of fake images generated by the generator is continuously improved according to the results of the discriminator during the training process, so as to achieve good sharpness ability.

2.2.2 Network Architecture

In this experiment, the architecture of the generator can be shown in the Figure 2.

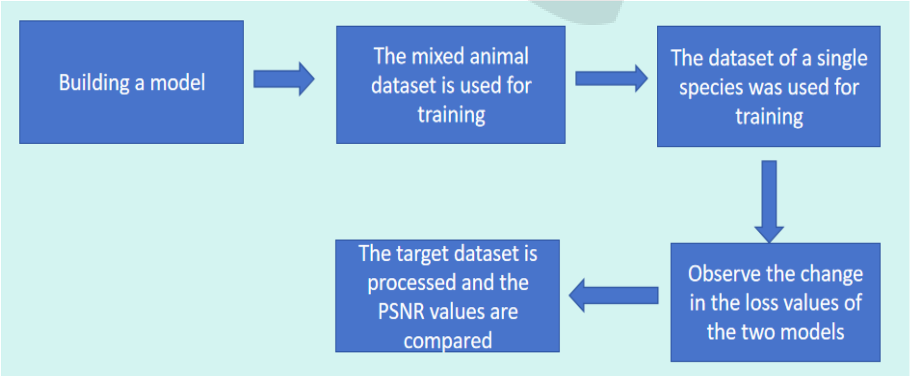
Building a model
The mixed animal dataset is used for training
The dataset of a single species was used for training
Observe the change in the loss values of the two models
The target dataset is processed and the PSNR values are compared

Figure 1: The research process (Original).

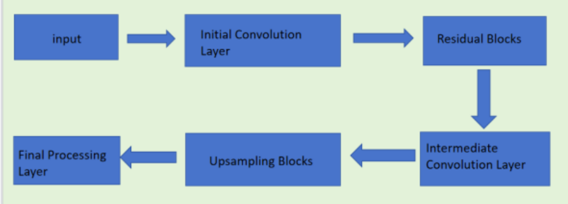

Figure 2: The architecture of generator (Original).

The input is convolutional and fed into the sequence of residual blocks. In an ESRGAN model the Residual sequence Block is a sequence containing multiple RRDB. RRDB is a kind of deep neural network block used for image processing. RRDB contains multiple Dense Blocks inside, usually five. Each dense block contains a series of convolutional layers for local feature extraction and enhancement. Each dense block contains the same structure, as shown in the Figure 3.

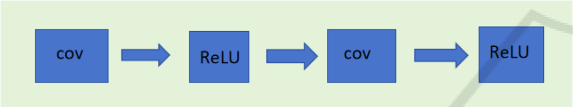

Figure 3: The architecture of RRDB (Original).

These residual blocks are stacked on top of each other. Each residual block performs deep feature extraction and augmentation. The intermediate convolutional layer is responsible for processing the output of the residual block. The upsampling block increases the resolution of the feature map. The final processing layer contains two convolutional layers. The architecture of the discriminator is shown in Figure 4.

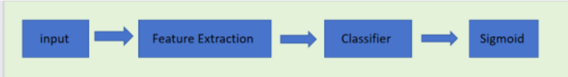

Figure 4: The architecture of Discriminator (Original).

The discriminator employs the visual geometry group (VGG) network to assess the likelihood that a real image is more authentic than a fake image. Here is the procedure for calculating the discriminator's output:

$$\mathrm{R}(\mathrm{x_r}) = \sigma(\mathrm{C(real)} - \mathrm{E[C(fake)]}) \to 1 \quad (1)$$

$$\mathrm{R}(\mathrm{x_f}) = \sigma(\mathrm{C(fake)} - \mathrm{E[C(real)]}) \to 0 \quad (2)$$

where $\mathrm{R}(\mathrm{x_r})$ and $\mathrm{R}(\mathrm{x_f})$ are the output of discriminator. When the discriminator determines that the real image is real, the output is $\mathrm{R}(\mathrm{x_f})$. Otherwise, the output is $\mathrm{R}(\mathrm{x_r})$. And C(real) and C(fake) represent the output score of the discriminator for the real and fake HR image, E[C(fake)] represents the expected value of the output of the discriminator for the fake HR image generated by the generator, E[C(real)] represents the expected value of the output for the real HR image , σ represents the standard deviation.

So the loss function of the discriminator is defined as follows:

$$L_d = -E_{x_r}[log(R(x_r))] - E_{x_f}[1 - log(R(x_f))] \quad (3)$$

On the contrary, the generator's adversarial loss function is formulated as follows:

$$L_g = -E_{x_f}[log(R(x_f))] - E_{x_r}[log(1 - R(x_r))] \quad (4)$$

where L_g means the loss function of the generator, L_d means the loss function of the discriminator, log means in logarithmic form, $\mathrm{R}(\mathrm{x_r})$ and $\mathrm{R}(\mathrm{x_f})$ is defined above, E_{x_r} is the expectation of the real image, E_{x_f} is the expectation of the fake image.

2.2.3 Loss Function

The perceived loss is calculated by utilizing pre-trained convolutional neural networks to assess the similarity in features between the generated image and the genuine image.In ESRGAN the perceptual loss compares the features with the real image using the activation function. After the above definition, for the generator G, its loss function is:

$$L_G = L_{percep} + \lambda L_g + \eta L_1 \quad (5)$$

where L_{percep} is perceptual loss, in the study it is L1 Loss, L_1 is pixel-wise Loss, that is:

$$L_1 = E_{x_i}||G(x_i - y)||_1 \quad (6)$$

In the study, $\lambda = 5 \times 10^{-3}$, $\eta = 0.01$. For the discriminator, the loss function is defined above:

$$L_D = L_d = -E_{x_r}\left[\log\left(R(x_r)\right)\right] - E_{x_f}[1 - log(R(x_f))] \quad (7)$$

2.3 Implementation Details

The python version used in this experiment is 3.11.5, the number of iterations is 2000, the learning rate is 1e-4, the total training cycle is 5, the batch size is 8, the gradient penalty weight is 10, the number of worker threads is 4, and the device used is GPU.

3 RESULTS AND DISCUSSION

This paper presents the experimental results of this experiment and discusses the significance of the experimental results in this section. In this experiment, the model is trained with a mixed data set of pictures of 100 animals. Then, then the resulting model is used to process the target dataset composed of different kinds of dogs, and the average Peak Signal-to-Noise Ratio(PSNR) value is obtained. Then, a second model is trained on a dataset consisting of images of different types of dogs (different from the target dataset), and the second model is used to process the same target dataset and obtain the average PSNR value.

3.1 Training Dataset of Mixed Animals

For the first model, the training dataset of mixed animals is used for training. Figure 5 and Figure 6 illustrate the fluctuations in the loss values of both the generator and discriminator.

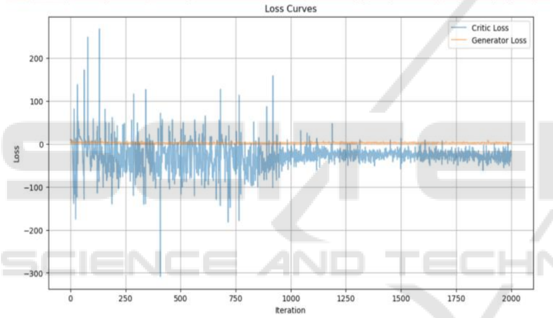

Figure 5: G-Loss (Generator Loss) and D-Loss (Discriminator Loss) (Original).

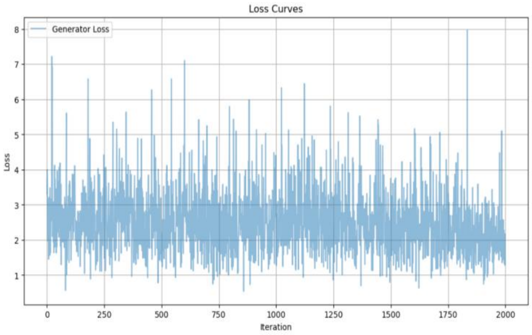

Figure 6: G-Loss (Picture credit: Original).

The results show that the discriminator loss fluctuates between -100 and 100 when some abnormal peaks are discarded, and after the number of training times reaches 900-1000 times, the fluctuation of the discriminator loss starts to become significantly smaller, and the value tends to 0 to -50. With some of the abnormal peaks discarded, the generator loss is between 1 and 6, and the fluctuation starts to become smaller when the number of training times reaches 1500 to 2000. Under this model, the average PSNR value of the target dataset image is 64.27290923861564.

3.2 Training Dataset of a Single Animals

For the second model, the training dataset consisted of images of dogs and is utilized for training. Fig. 7 and Fig. 8 depict the variations in the loss values for the generator and discriminator, respectively.

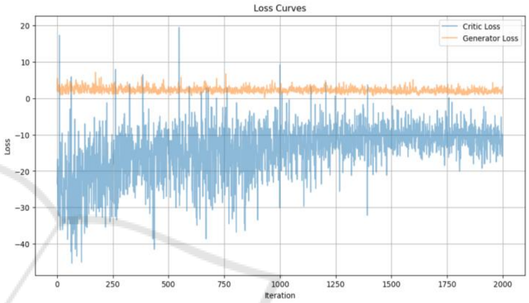

Figure 7: G-Loss and D-Loss (Original).

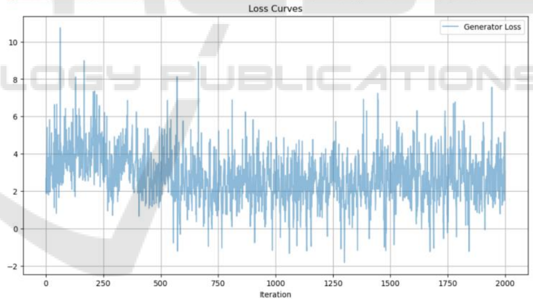

Figure 8: G-Loss (Original).

The results indicate that the discriminator loss falls within the range of -40 to 0, excluding certain abnormal peaks. After 1000-1250 training iterations, the fluctuation in the discriminator loss noticeably decreases, with values ranging from -20 to 0. In the case of discarding some abnormal peaks, the generator loss is between 0 and 6. Under this model, the average PSNR value is 65.84025702654814.

3.3 Comparison of Results

Throughout the training process, the absolute value of the discriminator loss for the first model is notably higher than that of the second model, and it exhibits greater fluctuations. Similarly, the generator loss for the first model shows more significant fluctuations and has a substantially higher absolute value compared to the second model. This means that the training dataset for the second model can be more conducive to training the model. Using PSNR index to measure the SR results of the two models, the first model is slightly lower than the second model. This demonstrates that the ratio between the SR result and the original image is greater, indicating an improved quality of the generated image. This means that training on a single animal dataset and using the generated model to SR process images of the corresponding species will be slightly better. In summary, the results suggest that it is better to use single-species datasets to train models and process images of that species.

4 CONCLUSION

The purpose of this study is to compare the SR effects of models trained on different training datasets on the same target dataset. Specifically, it is investigated whether the models trained on the training dataset of the same kind of animals have better SR effect on the target dataset of the same kind of animals. In this experiment, the SR model based on ESRGAN is established and a dataset consisting of pictures of various animals and a dataset consisting only of pictures of dogs are trained separately. By comparing the change of the loss value of the discriminator and the generator during the training process, this experiment can judge the training effect of the model and the training time required. Finally, SR processing is performed on the same target dataset composed of dog pictures and the SR effects are compared. The results show that the SR effect of the models trained with different training sets is not the same when the SR processing of a certain kind of animal image is needed. The model trained with the same type of animal images as the target dataset can have a better SR effect on the target animal images. At the same time, images consisting of a single species lead to faster training of models with less loss. In the future, this experiment will also consider removing the background of the dataset to obtain better experimental results, use more kinds of training dataset to compare the experimental effects and consider using more metrics to measure the SR effect of the images.

REFERENCES

A. Sharma, B. P. Shrivastava, "Different Techniques of Image SR Using Deep Learning: A Review," IEEE Sensors Journal, vol. 23, 2022, pp. 1724-1733.

M. Frid-Adar, I. Diamant, E. Klang, et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," Neurocomputing, vol. 321, 2018, pp. 321-331.

Q. Yang, Y. Ma, F. Chen, et al. "Recent advances in photo-activated sulfate radical-advanced oxidation process (SR-AOP) for refractory organic pollutants removal in water," Chemical Engineering Journal, vol. 378, 2019, pp. 122149.

K. Schawinski, C. Zhang, H. Zhang, et al. "Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit," Monthly Notices of the Royal Astronomical Society: Letters, vol. 467, 2017, pp. 110-114.

R. Keys, "Cubic convolution interpolation for digital image processing," IEEE transactions on acoustics, speech, and signal processing, vol. 29, 1981, pp. 1153-1160.

A. Sharma, B.P. Shrivastava, "Different Techniques of Image SR Using Deep Learning: A Review," IEEE Sensors Journal, vol. 23, 2022, pp. 1724-1733.

M.N Favorskaya, A.I. Pakhirka, "SF-SRGAN: Progessive GAN-based Face Hallucination," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 48, 2023, pp. 47-52.

X. Wang, K. Yu, S. Wu, et al. "Esrgan: Enhanced super-resolution generative adversarial networks," Proceedings of the European conference on computer vision (ECCV) workshops, 2018.

J. Yang, J. Wright, T. Huang, et al. "Image super-resolution as sparse representation of raw image patches," 2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008, pp. 1-8.

Dataset https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals

J. Jiang, L. Zhao, Y. Jiao, "Research on Image Super-resolution Reconstruction Based on Deep Learning," International Journal of Advanced Network, Monitoring and Controls, vol. 7, 2022, pp. 1-21.