Single Image Super-Resolution: Methods and Result

Kaggle link for the project

GitHub link for the project

Abstract

In the past few decades, the field of image super-resolution has gained immense significance across various domains, serving pivotal roles in applications spanning medicine, agriculture, remote sensing, defense security, and more. In this research, the focus is on Real-ESRGAN, an evolutionary leap from ESRGAN and a significant enhancement upon SRGAN—two influential frameworks in the realm of image super-resolution. Real-ESRGAN marks a notable advancement by addressing the limitations of its predecessors. Unlike ESRGAN and SRGAN, Real-ESRGAN introduces a groundbreaking capacity to upscale natural images while preserving their authenticity and detail. This capability to enhance natural images sets it apart in the realm of super-resolution techniques. The methodology of Real-ESRGAN showcases the potential to produce images of higher resolution, previously unattainable by its predecessors. By leveraging the advancements in Real-ESRGAN, this research delves into the unexplored territory of natural image enhancement, aiming to surpass the limitations faced by previous methods. The objective is to understand and highlight the groundbreaking potential of Real-ESRGAN in creating sharper, more detailed, and visually appealing high-resolution images, thereby paving the way for advancements in image super-resolution beyond the capabilities offered by ESRGAN and SRGAN. This exploration seeks to unveil the transformative impact of Real-ESRGAN in producing high-quality, natural-looking images that preserve their inherent details even after upscaling.

Introduction

Super resolution, a groundbreaking technology, holds the promise of elevating image and video quality beyond their original resolutions. Whether through Single-Image Super Resolution (SISR) techniques, which employ interpolation or advanced deep learning models, or through Multi-Image Super Resolution (MISR), which harnesses the power of multiple images, super resolution has emerged as a transformative force in the field of computer vision.

Single-Image Super Resolution (SISR) utilizes classical interpolation methods and the prowess of deep learning, including Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), to breathe new life into low-resolution or outdated visuals. The versatility of super resolution technology extends across various domains, including rejuvenating old photographs, enhancing surveillance footage for critical analysis, and notably, increasing the level of detail in medical imaging.

This paper investigates the ever-evolving field of super-resolution techniques, focusing on Real-ESRGAN. Building upon ESRGAN and further improving upon SRGAN, this method is capable of significantly enhancing natural images without compromising their inherent quality. The primary aim of this study is to comprehend the capacity of Real-ESRGAN in producing clearer images while exploring its potential applications in the future.

Related work

Image super-resolution

In the field of image processing, the quest for enhancing image and video quality has spurred numerous techniques, among which super resolution (SR) holds a pivotal role. This research paper aims to advance our understanding of single image super-resolution (SISR) techniques, while excluding discussions on approaches related to multi-image super resolution (MISR).

Historically, early researchers in the field pioneered prediction-based techniques to address SISR. However, these traditional filtering methods, such as linear, bicubic, or Lanczos filtering, often led to outputs with undesirably smooth textures.

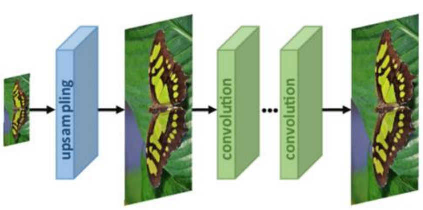

Fig:1 Predefined upsampling commonly uses the conventional interpolation, such as Bicubic, to upscale LR input images before entering the network.

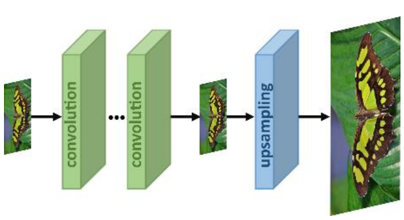

Fig:2 Post upsampling Upsampling layers are learnable and trained together with the preceding convolution layers in an end-to-end manner.

In contrast, methods that prioritize edge preservation have been proposed by experts in the field. One such innovative approach is the Deep Back-Projection Network (DBPN), which addresses mutual dependencies in low- and high-resolution images. DBPN employs iterative up- and downsampling layers with error feedback, representing various image degradations. Notably, its variant, Dense DBPN, which allows feature concatenation, has achieved significant improvements, setting new state-of-the-art results for large scaling factors, notably 8×, across multiple datasets.

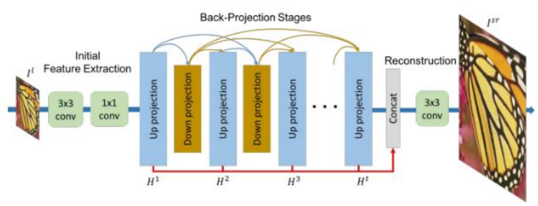

Fig:3 D-DBPN is an implementation for super-resolution.

Additionally, recent advancements in super-resolution have also been made with models like ESRGAN and SRGAN, which utilize deep learning for image resolution enhancement.

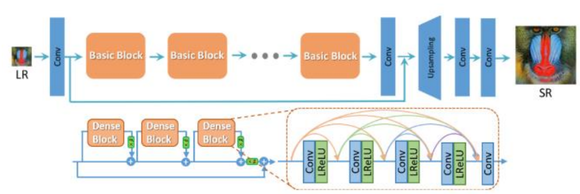

Fig:4 ESRGAN is the architecture which enhanced the SRGAN.

In the GAN architecture, there are two main models called Generative and Discriminative models. Discriminative models classify data, distinguishing differences. Generative models create new data, learning the underlying patterns and generating new examples based on learned distributions.

Martin Krasser and LEE Ho Ting have contributed significantly to the field of Single Image Super-Resolution by implementing cutting-edge techniques in TensorFlow 2.x. Their work includes the implementation of Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR). Additionally, they tackled the Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR). Moreover, their implementation of Photo-Realistic Single Image Super-Resolution using a Generative Adversarial Network (SRGAN) is a remarkable contribution. This offers a high-level training API and enables model fine-tuning, providing an essential resource for our research.

In recent years, convolutional neural network (CNN) based SR algorithms have shown remarkable performance. For instance, researchers incorporated a sparse representation prior into their feed-forward network architecture. Others utilized bicubic interpolation and deep fully convolutional networks, achieving state-of-the-art SR performance. Subsequently, it was demonstrated that allowing networks to learn upscaling filters directly can significantly increase both accuracy and speed. With their deeply recursive convolutional network (DRCN), researchers presented a highly performant architecture that balances long-range pixel dependencies with a compact model. Of relevance to this paper are the works by Johnson et al. and Bruna et al., emphasizing a perceptual similarity-based loss function for the recovery of visually compelling HR images.

Our Approach

SRGAN, ESRGAN and Real-ESRGAN are three types of artificial intelligence models that can enhance the resolution and quality of low-quality images. They use a technique called generative adversarial networks (GANs), which consist of two parts: a generator and a discriminator. The generator tries to create realistic high-resolution images from low-resolution inputs, while the discriminator tries to distinguish between the generated images and the real high-resolution images. The generator and the discriminator compete with each other, improving their performance over time.

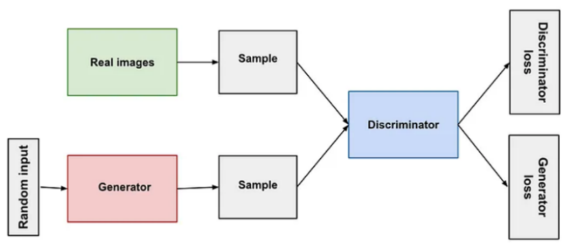

Fig:5 image source Both the generator and the discriminator are neural networks. The generator output is connected directly to the discriminator input. Through backpropagation, the discriminator's classification provides a signal that the generator uses to update its weights.

The main difference between SRGAN and Real-ESRGAN is that they are trained on different types of data and optimized for different types of degradation. SRGAN is trained on synthetic data, where the low-resolution images are obtained by downscaling the high-resolution images using a simple method called bicubic interpolation. This method produces smooth and blurry images that lack fine details. SRGAN is good at restoring these details and creating sharp and realistic textures. However, SRGAN is not very effective on real-world images, which may suffer from various types of degradation, such as noise, compression, blur, or unknown artifacts. These types of degradation are difficult to simulate and require more complex models to handle.

ESRGAN and Real-ESRGAN are two AI models that employ deep learning to augment resolution and correct degradation. They share similar names but serve subtly different purposes. Real-ESRGAN is an upgraded ESRGAN trained with pure synthetic data and is capable of enhancing details while removing annoying artifacts for common real-world images. Real-ESRGAN shows better results on faces compared to the original version.

The choice of model depends on the specific use case and the type of image degradation that needs to be corrected. ESRGAN is ideal for pristine bicubic downscaled images or weird noise from AI generations, while Real-ESRGAN is tailored to tackle real-world corruptions like the compression common in social media.

Real-ESRGAN is an extension of SRGAN and ESRGAN that is designed to tackle real-world degradation. It is trained on a large and diverse dataset of real low-quality images, which cover various scenarios and sources of degradation. Real-ESRGAN uses a more advanced network architecture and a more robust loss function to capture the complex characteristics of real degradation and restore the natural appearance of the images. Real-ESRGAN can handle both synthetic and real degradation and produce more visually pleasing results than SRGAN and ESRGAN.

Therefore, **Real-ESRGAN** was selected among various super resolution methods due to its training on natural images, offering a promising approach for converting real-world low-quality images to high-resolution ones without compromising their natural appearance.

In our approach to refining the Real-ESRGAN model for our specific task, we initially employed the existing pre-trained model to enhance mild and ×8 images. However, as we encountered more difficult and wild-type images, we realized the necessity of fine-tuning the model due to the high error rates associated with this method.

To address the challenges posed by these difficult image types, we resorted to finetuning the existing model. This finetuning process was conducted using a Kaggle notebook, specifically tailored to accommodate and train the difficult dataset, further refining the Real-ESRGAN model.

Fig:6 The comparisons on real-world samples(difficult image type). Real-ESRGAN protects the naturality of the image.(*Zoom in for best view*).

This systematic approach began with the assembly of a pertinent dataset containing low-quality real-world images. Leveraging the Real-ESRGAN model's pre-existing capabilities, derived from its initial training on natural images, we initialized the model using its pre-existing weights. The objective was to preserve the natural appearance of the images while enhancing their quality, leading us to conduct fine-tuning. This fine-tuning process involved adjusting the model's layers and parameters to better suit the unique characteristics of the low-quality images. Subsequently, the performance of the fine-tuned model was evaluated on a separate validation set to ensure its efficacy in enhancing images without compromising their innate natural qualities. This iterative process allowed us to tailor the model's performance specifically to our research task, catering to a wider range of image complexities.

SRGAN, ESRGAN and Real-ESRGAN are all powerful models for image super-resolution, but they have different strengths and weaknesses. SRGAN and ESRGAN are better for synthetic degradation, such as bicubic downscaling, while Real-ESRGAN is better for real-world degradation, such as compression or noise. Depending on the type and source of the low-quality images, we have chosen Real-ESRGAN as the most suitable model.

Image type	PSNR value	SSIM value
mild	28.60754393545078 dB	0.68769343
×8	30.492572358662233 dB	0.77257087
difficult	28.573498321667625 dB	0.69677128

Fig. 7 shows PSNR and SSIM comparisons between super-resolved and HR images based on input image category.

Real Image SRGAN ESRGAN Real-ESRGAN

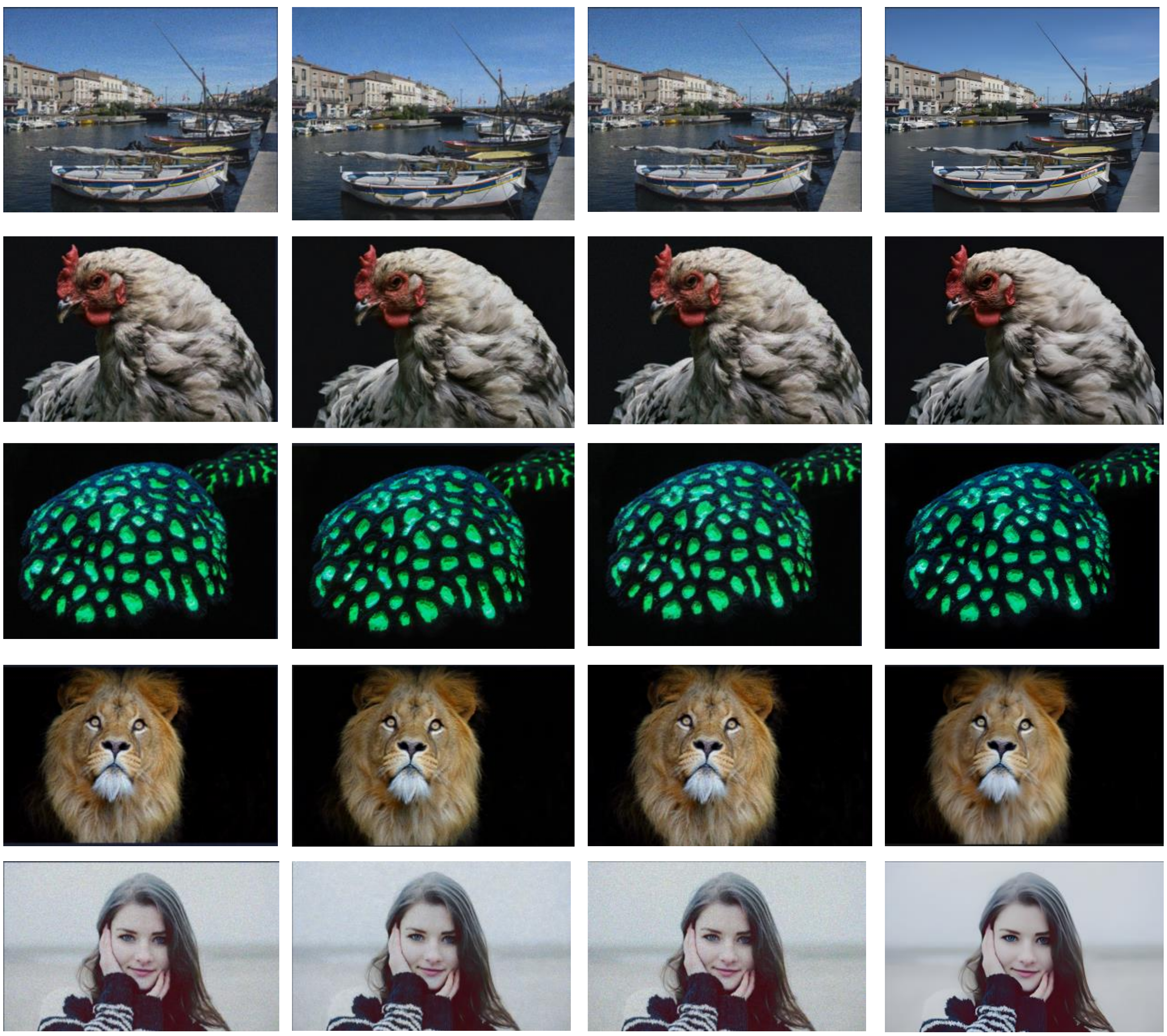

Fig:8. The comparisons of several representative real-world samples. Real-ESRGAN outperforms previous approaches in both removing artifacts and restoring texture details. Other methods may either fail to remove overshoot and fail to restore realistic and natural textures for various scenes. (*Zoom in for best view*).

Discussion

Given the constraints of time, our research primarily focused on exploring and gaining insights into various existing super resolution methodologies and their progressive advancements in enhancing image resolution. Consequently, our study led us to examine several prominent models such as SRGAN, ESRGAN, Real-ESRGAN, and diverse architectures like GAN and UNet. Throughout our exploration, an invaluable discovery was the identification of ChaiNNer, a node-based image processing Graphical User Interface (GUI). This tool significantly facilitated and streamlined our workflow, contributing to a more efficient research process.

Acknowledgements

We extend our sincere appreciation to Dr. Ranga Rodrigo whose exceptional guidance, unwavering support, and insightful feedback were instrumental in the successful completion of this project. Additionally, we would like to express our heartfelt gratitude to Tharindu Wickramasinghe for their valuable contributions.

Conclusion

The investigation into image super-resolution, particularly focusing on Real-ESRGAN, highlights its pivotal advancement over previous models, notably ESRGAN and SRGAN. Real-ESRGAN's unique ability to enhance natural images while maintaining their original quality marks a significant leap in image super-resolution. This research unveils the transformative potential of Real-ESRGAN, showcasing its capacity to generate sharper, more detailed, and natural-looking high-resolution images. Its surpassing of prior limitations suggests a promising future for image enhancement, with far-reaching implications across diverse fields like medicine, agriculture, remote sensing, and defense security.

References

[1] Jing L., & NTIRE. (2018). NTIRE 2018 Challenge on *Single Image Super-Resolution*: Methods and Results.

[2] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2017). *Generative Adversarial Networks*

[3] Haris, M., Shakhnarovich, G., & Ukita, N. *Deep Back-Projection Networks For Super-Resolution.*

[4] Krasser, M. (2018). *Super-resolution using Generative Adversarial Networks* (Version 1.0)

[5] Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). *Enhanced Deep Residual Networks* for Single Image Super-Resolution. arXiv:1707.02921 [cs.CV]

[6] Yu, J., Fan, Y., Yang, J., Xu, N., Wang, Z., Wang, X., & Huang, T. (2018). Wide Activation for Efficient and Accurate Image Super-Resolution. arXiv:1808.08718 [cs.CV].

[7] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017) *Photo-Realistic Single Image Super-Resolution* Using a Generative Adversarial Network. arXiv:1609.04802 [cs.CV]

[8] Jeewanjj02. (Year, Month Day). *ESRGAN & SRGAN: Image Resolution* Using Deep Learning. Medium.

[9] Polo Club. (Year, Month Day). GANLab - An Interactive Tool for Visualizing, Understanding and Comparing GANs.

[10] Boesch, G. (n.d.). *Generative Adversarial Networks* . VISO AI. Link

[11] Mishra, D., Dror, I., Hadar, O., & Choukroun, D. (2023). A Fusion-Based Framework for Unsupervised Single Image Super-Resolution. In Cyber Security, Cryptology, and Machine Learning (pp. 85-95). doi:10.1007/978-3-031-34671-2_7

[12] Corrêa de Araújo, G., Jordão, A., & Pedrini, H. (2023). Single Image Super-Resolution Based on Capsule Neural Networks. In Intelligent Systems (pp. 113-128). doi:10.1007/978-3-031-45392-2_8

[13] Al-Mekhlafi, H., & Liu, S. (2023). Single image super-resolution: A comprehensive review and recent insight. Frontiers of Computer Science, 18(1), 181702.