Progressive Upscaling Super Image Resolution

Ms. Swati Malik
Faculty of Computer Science of Technology
Maharaja Surajmal Institute of Technology
Delhi,India
Swatimalik@msit.in

Mr. Aman Kr. Karan
Computer Science of Technology
Maharaja Surajmal Institute of Technology
Delhi,India
Amankumarkarn87865@gmail.com

Mr.Ankit Khandelwal
Computer Science of Technology
Maharaja Surajmal Institute of Technology
Delhi,India
Ankitkhandelwalk0603@gmail.com

Mr. Gaurav Singh
Computer Science of Technology
Maharaja Surajmal Institute of Technology
Delhi,India
Gaurav10331233@gmail.com

***Abstract*— This paper presents a comparison of the performance of GAN, SRCNN, SRGAN, ESRGAN at various stages of super image resolution with the performance of PROGAN at every stage. The algorithms are evaluated based on their ability to enhance the resolution of images. The comparison is based on PSNR and Mean Square Value error. The results show that while SRCNN marked a foundational step, SRGAN and ESRGAN demonstrated commendable advancements, particularly in capturing finer details. However, ProGAN maintained a superior overall quality, setting a high standard for image super-resolution The paper concludes with a discussion of the strengths and weaknesses of each algorithm and recommendations for future research.**

***Keywords* — Super image resolution, algorithms, evaluation criteria.**

I. INTRODUCTION

This paper presents a comparison of various super image resolution algorithms. The algorithms are evaluated based on their performance in enhancing the resolution of images. The algorithms compared include the algorithm that contain SRCNN at 1st stage (2x stage), SRGAN at 2nd Stage and ESRGAN at final stage . The comparison is based on value of PSNR, value of mean square error and quality of image. The results show that while SRCNN marked a foundational step, SRGAN and ESRGAN demonstrated commendable advancements, particularly in capturing finer details. However, ProGAN maintained a superior overall quality, setting a high standard for image super-resolution.

The paper concludes with a discussion of the strengths and weaknesses of each algorithm and recommendations for future research.

In recent years, several algorithms have been proposed for super image resolution. These algorithms can be broadly classified into two categories: traditional methods and deep learning methods. Traditional methods include bicubic interpolation, Lanczos resampling, and nearest neighbor interpolation. Deep learning methods include generative adversarial networks (GANs), super-resolution generative adversarial networks (SRGANs), and enhanced super-resolution generative adversarial networks (ESRGANs)[1].

Image super-resolution is a process to recover an image of high-resolution (HR) from a low-resolution (LR) image [2]. In simple terms, it can also be referred to as image interpolation, scaling, upsampling, zooming, or enlarging [2]. The purpose of image super-resolution is to obtain a high pixel density and refined details from low-resolution (LR) image(s) that cannot be seen with the naked eye. Image super-resolution was very useful in applications that required recognition or detection purposes.

In the past, many image processing-based techniques have been used in image super-resolution before deep learning-based methods were started. Li et al. [5] and Nasrollahi et al. [6] classified image super-resolution methods into three different groups, namely, interpolation-based method, reconstruction-based method, and learning-based method. Several interpolation-based methods such as linear [7], bilinear [8], or bicubic [9,10] interpolations can be found in image super-resolution applications. These methods are simple, but the high-frequency details of the image are not restored and therefore more sophisticated insight may be required to recover the image [5,6].

The reconstruction-based method includes the sharpening of edge details [11], regularization [12,13], and deconvolution [14] techniques. Researchers also used these techniques for image reconstruction. The learning-based method has an advantage over the interpolation-based method by restoring the missing high-frequency details through a relationship established between the LR image and the HR image.

Conventional upscaling methods, in their haste to increase resolution, often stumble upon pitfalls that result in a noticeable degradation of image quality. These pitfalls manifest as blurriness, pixelation, and an overall loss of visual fidelity.

The core concept underpinning this project is the incremental enhancement of resolution. Rather than a single, abrupt transformation, progressive upscaling divides the process into smaller, more manageable increments. In each of these steps, the algorithm meticulously analyzes and enhances a portion of the image, with a relentless focus on preserving the fine details that might otherwise be lost.

Preserving the original characteristics of the image is a nuanced challenge. The algorithm must delicately balance the enhancement of resolution while safeguarding color fidelity, edge sharpness, and the overall aesthetic. Each type of content, whether it's a photograph, illustration brings its own set of characteristics and challenges that the algorithm must adapt to.

The paper is organized as follows. Section I is the introduction, Section II provides a brief overview of the existing literature on super image resolution algorithms.

Section III describes the methodology used in this study. Section IV presents the results of the comparison. Section V discusses the strengths and weaknesses of each algorithm. Finally, Section VI concludes the paper with recommendations for future research.

We hope that this study will provide valuable insights into the performance of various super image resolution algorithms and help researchers and practitioners in selecting the most appropriate algorithm for their specific needs.

II. Literature Review

The field of image super-resolution has witnessed significant advancements, driven by the continuous pursuit of achieving higher visual fidelity and enhanced details in reconstructed images. In this literature review, we delve into key datasets used for training and evaluation, followed by an exploration of influential methodologies, including the fully progressive approach and the perception-oriented network.

A Fully Progressive Approach to Single-Image Super-Resolution [3]

This paper proposes a novel single-image super-resolution (SISR) technique that employs a progressive approach to enhance image details at various scales. The proposed method, dubbed ProSR, consists of multiple sub-networks that progressively refine the low-resolution (LR) input image to produce a high-resolution (HR) output image. Each sub-network focuses on specific aspects of the image, such as restoring high-frequency details and preserving structural information. This progressive approach enables ProSR to achieve impressive resolution improvements, outperforming state-of-the-art SISR methods in terms of both quantitative metrics and visual quality.

Progressive Perception-Oriented Network for Single Image Super-Resolution[4]

This research introduces a novel SISR architecture, called PPON, that focuses on perceptual quality. PPON comprises multiple stages, each geared towards extracting and enhancing different levels of image features. The first stage focuses on minimizing pixel-wise error, while the second stage utilizes the features extracted by the previous stage to refine structural information. The final stage employs fine structure features distilled by the second phase to produce more realistic results. This progressive architecture, coupled with a perceptual loss function, enables PPON to achieve superior visual quality compared to existing SISR methods.

Enhanced Deep Residual Networks for Single Image Super-Resolution (Wang et al., 2018)[15]

This paper proposes an enhanced deep residual network (EDSR) architecture for SISR. EDSR introduces residual connections, densely connected convolutional layers, and a recursive residual block design, enabling it to achieve significant improvements in both PSNR and SSIM metrics.

Multi-Scale Residual Network for Single Image Super-Resolution (Zhang et al., 2018)[16]

This paper presents a multi-scale residual network (MSR) architecture for SISR. MSR employs a multi-scale approach, using multiple residual networks at different scales to capture image details at varying frequencies. This multi-scale design enables MSR to achieve superior results, particularly for high-resolution upscaling tasks.

Attention-Based Deep Fusion Network for Single Image Super-Resolution (Zhang et al., 2020)[17]

This paper proposes an attention-based deep fusion network (AFNet) for SISR. AFNet utilizes an attention mechanism to selectively fuse features from different scales, allowing it to focus on more relevant features and enhance the overall quality of the super-resolved image. AFNet demonstrates superior performance compared to previous SISR methods in terms of both PSNR and SSIM metrics.

III. Scope of work

This research project unfolds within a meticulously defined scope that intricately combines algorithmic development, dataset preparation, quantitative evaluation, comparative analysis, and documentation. The overarching goal is to advance our understanding of single-image super-resolution through the exploration of a progressive upscaling approach and its juxtaposition with a ProGAN-based iterative method.

Algorithmic Development:

- Formulate and implement a robust progressive upscaling algorithm, seamlessly integrating bicubic interpolation, SRGAN, and ESRGAN.
- Fine-tune algorithmic parameters to strike a balance between computational efficiency and optimal visual quality.
- Ensure algorithmic compatibility for a cohesive and effective multi-stage enhancement process.

Dataset Preparation:

- Employ the DIV2K dataset as the cornerstone for training, validation, and testing phases.
- Execute meticulous preprocessing to extract high-resolution images and corresponding low-resolution counterparts at 2x, 3x, and 4x downscaling factors.
- Establish a systematic division of the dataset, delineating specific subsets for training, validation, and final testing.

Evaluation Metrics:

- Implement stringent evaluation metrics, centering on the industry-standard Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE).
- Develop a standardized evaluation framework to meticulously gauge the performance of the

progressive upscaling approach at each developmental stage.

- Create a benchmark for comparative analysis, aligning with the ProGAN-based iterative method.

Comparison with ProGAN:

- Retrieve and preprocess data from a seminal research paper utilizing ProGAN for iterative super-resolution.

- Implement the ProGAN-based iterative method to facilitate a c

- Conduct an in-depth evaluation of PSNR and MSE values to delineate the relative performance of the progressive upscaling approach and ProGAN iterations.

Visual Analysis:

- Generate visually compelling comparisons between the progressive upscaling approach and ProGAN iterations, leveraging sample images from the designated test dataset

- Offer nuanced qualitative assessments, spotlighting visual enhancements or potential artifacts introduced by each method.

Scope Extension (If Applicable):

- Contemplate the extension of the progressive upscaling approach beyond the DIV2K dataset, potentially incorporating datasets like Flickr30k for a broader analysis.

- Investigate the generalizability of the proposed methodology across diverse image content, exploring its adaptability and efficacy.

Documentation and Reporting:

- Methodically document algorithmic intricacies, dataset preparation procedures, and evaluation methodologies.

- Synthesize results in a cogent and structured manner, presenting a comprehensive narrative of the research findings, and facilitating a deeper comprehension of the progressive upscaling approach's efficacy in contrast to ProGAN-based iterative methods.

Within this comprehensive scope, the research aspires to make meaningful contributions to the landscape of single-image super-resolution, unraveling the nuanced strengths and limitations inherent in progressive upscaling strategies and offering valuable comparative insights with ProGAN-based iterative approaches

IV. MATERIALS AND METHODS USED

Datasets:

DIV2K Dataset: The DIV2K dataset emerges as a pivotal resource in the realm of single-image super-resolution. Comprising 800 high-definition images for training, each accompanied by downscaled versions at 2x, 3x, and 4x resolutions, it provides a robust foundation for evaluating the effectiveness of upscaling techniques. Additionally, the inclusion of validation and test data, with low-res versions for feedback and final evaluation, respectively, ensures a comprehensive assessment of model performance.[18]

Flickr30k Dataset: Focusing on image description benchmarking, the Flickr30k dataset and its Entities subset contribute significantly to the exploration of textual entity localization. With 244k coreference chains and 276k annotated bounding boxes, it sets a new benchmark in this domain. The baseline, featuring image-text embeddings, object detectors, color classification, and a bias towards larger objects, establishes a strong foundation for evaluating complex models, particularly in the context of image-sentence retrieval.[19]

V. RESULTS AND DISCUSSIONS

The project aimed to enhance image resolution through a multi-stage approach employing deep learning techniques. Each stage introduced progressively advanced models to achieve superior results. The first stage focused on SRCNN, the Super-Resolution Convolutional Neural Network. The dataset selection was pivotal, requiring pairs of high and low-resolution images for effective training. During image preprocessing, pixel values were normalized to ensure consistent processing. Data augmentation techniques, such as rotation and flipping, were employed to diversify training samples.

In the training phase, the SRCNN model learned to map low-resolution images to their high-resolution counterparts. The architecture comprised three main layers: patch extraction, non-linear mapping, and reconstruction. This allowed SRCNN to effectively enhance details in low-resolution images. The output was an intermediate high-resolution image.

Fig 7.1 (a) SRCNN Input

Fig 7.1 (b) SRCNN Output

Stage two introduced SRGAN, the Super-Resolution Generative Adversarial Network. The objective was to not only improve resolution but also enhance visual quality through adversarial training. Images were preprocessed as needed, and SRGAN was trained with an adversarial loss in addition to the traditional super-resolution loss. The architecture included a generator for super-resolution and a discriminator for assessing realism. The output refined the intermediate result from SRCNN, producing a visually enhanced high-resolution image with heightened realism.

Fig 7.2 (a) SRGAN Input

Fig 7.2 (b) SRGAN Output

The third stage incorporated ESRGAN, the Enhanced Super-Resolution Generative Adversarial Network, aiming to further enhance resolution and fine details. Images were preprocessed specifically for ESRGAN training, and the model was trained with adversarial training, perceptual loss, and content loss. ESRGAN's architecture was deeper and more sophisticated compared to SRGAN, enabling it to capture finer details.

Fig 7.3 (a) ESRGAN Input

Fig 7.3 (b) ESRGAN Output

In the final stage, the outputs from SRCNN, SRGAN, and ESRGAN were compared against ProGAN (Progressive Growing of GANs), a benchmark known for its progressive growth architecture. ProGAN's stepwise refinement across different resolutions set a high standard. The evaluation considered quantitative metrics and visual inspection, revealing ProGAN's superior overall quality. While SRCNN served as a foundational step, SRGAN and ESRGAN exhibited notable improvements, with ESRGAN capturing finer details. ProGAN, however, showcased a more consistent and coherent refinement.

TABLE 1.2 Comparison Based on PSNR Value

PSNR VALUE	PROGAN	SRCNN	SRGAN
2X STAGE	35.80	34.60	
4X STAGE	30.39		42.34

This detailed project overview underscores the meticulous process of dataset selection, image preprocessing, and model training at each stage. It demonstrates the evolution from basic super-resolution techniques to advanced generative adversarial networks, with a final comparison against a state-of-the-art benchmark. The nuanced evaluation provides insights into the strengths and limitations of each approach, guiding future advancements in the field of image super-resolution through deep learning.

VI. CONCLUSION

In conclusion, this comprehensive image super-resolution project has traversed a transformative journey, showcasing the progressive refinement of models from SRCNN to SRGAN and ESRGAN. The meticulous dataset selection, image preprocessing, and innovative model architectures underscore the commitment to advancing the field of image enhancement through deep learning.

The initial stage with SRCNN laid a robust foundation by emphasizing the significance of a well-curated dataset and effective preprocessing techniques. As the project evolved, the introduction of SRGAN brought forth adversarial training, introducing a dynamic interplay between a generator and discriminator. This not only enhanced resolution but also significantly improved the visual realism of the generated images.

ESRGAN, in the third stage, further elevated the super-resolution process. Its deeper architecture, combined with adversarial training, perceptual loss, and content loss, showcased a commitment to capturing finer details and pushing the boundaries of image enhancement. The iterative refinement throughout each stage highlighted the project's dedication to continual improvement.

The final comparison against ProGAN, a benchmark in progressive growing architectures, provided nuanced insights into the relative strengths and weaknesses of each approach. While SRCNN marked a foundational step, SRGAN and ESRGAN demonstrated commendable advancements, particularly in capturing finer details. However, ProGAN maintained a superior overall quality, setting a high standard for image super-resolution.

This project's significance lies not only in its achievements in image super-resolution but also in its roadmap for future endeavors. It serves as an inspiring example of how the iterative refinement of deep learning models can lead to substantial strides in visual quality and fidelity. As the project concludes, it leaves a legacy of innovation, providing valuable insights for future research in the dynamic intersection of deep learning and image enhancement.

REFERENCES

[1] “Image Super Resolution | Deep Learning for Image Super Resolution,” Analytics Vidhya, May 26, 2021. https://www.analyticsvidhya.com/blog/2021/05/deep-learning-for-image-super-resolution/

[2] Anwar, S.; Khan, S.; Barnes, N. A deep journey into super-resolution: A survey. arXiv 2019, arXiv:1904.07523. [Google Scholar] [CrossRef]

[3] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, “A Fully Progressive Approach to Single-Image Super-Resolution,” Jun. 2018, doi: https://doi.org/10.1109/cvprw.2018.00131.

[4] Z. Hui, J. Li, X. Gao, and X. Wang, “Progressive perception-oriented network for single image super-resolution,” *Information Sciences*, vol. 546, pp. 769–786, Feb. 2021, doi: https://doi.org/10.1016/j.ins.2020.08.114.

[5] Li, X.; Wu, Y.; Zhang, W.; Wang, R.; Hou, F. Deep learning methods in real-time image super-resolution: A survey. J. Real-Time Image Process. 2020, 17, 1885–1909.

[6] Nasrollahi, K.; Moeslund, T.B. Super-resolution: A comprehensive survey. Mach. Vis. Appl. 2014, 25, 1423–1468.

[7] Tong, C.S.; Leung, K.T. Super-resolution reconstruction based on linear interpolation of wavelet coefficients. Multidimens. Syst. Signal Process. 2007, 18, 153–171. [Google Scholar] [CrossRef]

[8] Sun, N.; Li, H. Super Resolution Reconstruction of Images Based on Interpolation and Full Convolutional Neural Network and Application in Medical Fields. IEEE Access 2019, 7, 186470–186479. [Google Scholar] [CrossRef]

[9] Liu, J.; Gan, Z.; Zhu, X. Directional Bicubic Interpolation—A New Method of Image Super-Resolution. In 3rd International Conference on Multimedia Technology(ICMT-13); Atlantis Press: Paris, France, 2013; pp. 470–477. [Google Scholar]

[10] Kumar, G.; Singh, K. Image Super Resolution on the Basis of DWT and Bicubic Interpolation. Int. J. Comput. Appl. 2013, 65, 1–6. [Google Scholar]

[11] Dai, S.; Han, M.; Xu, W.; Wu, Y.; Gong, Y. Soft Edge Smoothness Prior for Alpha Channel Super Resolution. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8. [Google Scholar]

[12] Chang, K.; Ding, P.L.K.; Li, B. Single Image Super Resolution Using Joint Regularization. IEEE Signal Process. Lett. 2018, 25, 596–600. [Google Scholar] [CrossRef]

[13] Yu, L.; Cao, S.; He, J.; Sun, B.; Dai, F. Single-image super-resolution based on regularization with stationary gradient fidelity. In Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 14–16 October 2017; pp. 1–5. **[Google Scholar]**

[14] Shan, Q.; Li, Z.; Jia, J.; Tang, C.-K. Fast image/video upsampling. *ACM Trans. Graph.* **2008**, *27*, 1–7. **[Google Scholar]** **[CrossRef]**

[15] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced Deep Residual Networks for Single Image Super-Resolution,” 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jul. 2017, doi: https://doi.org/10.1109/cvprw.2017.151.

[16] Y. Liu, X. Zhang, S. Wang, S. Ma, and W. Gao, “Progressive Multi-Scale Residual Network for Single Image Super-Resolution.,” Jul. 2020.

[17] Zhang, Y., Li, K., Li, Y., Zhang, L., & Wei, Y. (2020). Attention-Based Deep Fusion Network for Single Image Super-Resolution. IEEE Transactions on Image Processing, 29(12), 3568-3581

[18] @InProceedings{Timofte_2017_CVPR_Workshops,

author = {Timofte, Radu and Agustsson, Eirikur and Van Gool, Luc and Yang, Ming-Hsuan and Zhang, Lei and Lim, Bee and others},

title = {NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results},

booktitle = {The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops},

month = {July},

year = {2017}

}

[19] P. Young, A. Lai, M. Hodosh, and J. Hockenmaier, "From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532-1543.

[20] Timofte, Radu and Gu, Shuhang and Wu, Jiqing and Van Gool, Luc and Zhang, Lei andYang, Ming-Hsuan and Haris, Muhammad and others,"NTIRE 2018 Challenge on SingleImageSuper-Resolution",TheIEEEConferenceonComputerVisionandPatternRecognition(CVPR)Workshops,June,2018

[21] Ignatov, Andrey and Timofte, Radu and others, "PIRM challenge on perceptual imageenhancementonsmartphones",EuropeanConferenceonComputerVision(ECCV)Workshops,January,2019

[22]] Ledig, C., Theis, L., Husz´ar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image superresolution using a generative adversarial network. In: CVPR. (2017)

[23] Jolicoeur-Martineau, A.: The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734 (2018)

[24] Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The pirm challenge on perceptual super resolution. https://www.pirm2018.org/PIRM-SR. html (2018)

[25] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV. (2014) [5.] Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: CVPR. (2016)