Single Image Super-Resolution

Sangana Sai Venkata Surya Teja (EE23MTECH11032)
Gedela Madhu Babu (EE23MTECH11027)
Ritesh Kushwaha (EE23MTECH11031)
Kiran Yadav (EE23MTECH11026)
Manisha Margrate Malto (EE23MTECH11006)

Abstract

Recent years have witnessed significant advancements in the field of Single Image Super-Resolution (SISR), a key area within image processing focused on reconstructing high-resolution (HR) images from their low-resolution (LR) counterparts. With the advent of deep learning, super-resolution techniques have undergone a transformative evolution, leading to substantial enhancements in reconstruction quality, especially on synthetic datasets. This report offers a comprehensive overview of the latest developments in super-resolution, emphasizing the role of deep learning in pushing the boundaries of this domain. It chronologically charts the journey from traditional, classical methods to cutting-edge deep learning approaches, highlighting key milestones and breakthroughs. Some of the most notable advancements discussed are the latest diffusion models, which represent a paradigm shift in super-resolution technology. By providing an in-depth analysis of these innovative techniques alongside a historical perspective on super-resolution methods, this report aims to shed light on the evolution, current state, and future directions of super-resolution research.

1. Introduction

Image super-resolution (ISR) represents a pivotal challenge within the realm of computer vision, aiming to synthesize high-resolution (HR) images from their low-resolution (LR) counterparts. As the demand for high-definition visual content escalates—fueled by the proliferation of advanced display technologies and superior image capture devices—the need for efficient and effective high-resolution imagery is more pronounced than ever. Despite the allure of directly capturing HR images, such endeavors often entail prohibitive costs, extensive time investments, and the necessity for sophisticated equipment. Conversely, super-resolution (SR) techniques, rooted in signal processing, offer a cost-effective and versatile solution to augment the resolution of LR images. These methods enable the reconstruction of HR outputs from mere LR observations, pushing the boundaries of what's achievable with existing imaging systems.

To navigate the complexities of ISR, researchers have devised numerous methodologies capable of producing high-quality HR images from LR inputs. Broadly, these techniques are categorized into single-image super-resolution (SISR) and multi-image super-resolution (MISR). SISR approaches rely on a singular LR image to fabricate an HR image, while MISR techniques harness multiple LR images to achieve the same goal. The advent of deep learning has markedly shifted the landscape of ISR, positioning it as the preeminent methodology. Leveraging convolutional neural networks (CNNs), generative adversarial networks (GANs), and attention mechanisms, deep learning-based SR methods have set new benchmarks, delivering HR images with unparalleled detail and authenticity. These models adeptly learn the intricate mapping from LR to HR images, showcasing their capacity to produce results that were previously unattainable.

This study aims to furnish a concise overview of traditional SR methodologies before delving into an exhaustive examination of the latest deep learning-based SR research. Emphasis will be placed on dissecting the innovations and contributions of recent studies, analyzing the strengths and limitations of current approaches, and shedding light on the myriad challenges and prospects that lie ahead in the evolving domain of ISR. Through this exploration, we seek to offer valuable insights into the ongoing advancements and potential future directions in SR research, a field that continues to captivate and inspire due to its significant implications for image processing and beyond.

1.1. OBJECTIVE

The primary objective of image super-resolution (SR) is to reconstruct a high-resolution (HR) image from its low-resolution (LR) counterpart. In this context, let I_L denote the LR image and I_H the HR image it corresponds to. The transformation from I_H to I_L involves a degradation process, which can be mathematically modeled as:

$$I_L = D(I_H; \delta)$$

Here, D symbolizes the degradation process, and θ encapsulates the parameters characterizing this process, such as reduction in size, introduction of noise, and blurring factors. Through this degradation, I_L invariably loses significant detail and high-frequency information, necessitating a process to recover this lost information.

Nonetheless, a notable challenge arises when the specifics of the degradation process D remain elusive or entirely unknown, making it infeasible to directly ascertain how I_H deteriorates into I_L. The crux of super-resolution, therefore, lies in devising a method to effectively invert the degradation process D using the available I_L, such that the reconstructed image $\hat{I}_H$ approximates the original HR image I_H in definition and detail. This inverse operation can be succinctly described by the following equation:

$$\hat{I}_H = F(I_L; \theta)$$

In this formula, F signifies the super-resolution process, essentially serving as the inverse of the degradation function D, and θ represents various parameters integral to the SR process. These parameters may include the upscaling factor and specific weights within a deep learning architecture, among others.

Understanding and optimizing the SR process involves not only reversing the effects of degradation but also intelligently interpolating and inferring missing high-frequency content. Deep learning models, particularly those employing convolutional neural networks (CNNs), have proven exceptionally adept at this task, leveraging vast datasets to learn the complex mapping from I_L to I_H and achieving remarkable success in restoring images to their high-definition glory.

1.2. EXISTING WORKS

In [4],Khaledyan et al. (2020) address the challenge of real-time image super-resolution through the implementation of bilinear and bicubic interpolation techniques. Their work focuses on achieving cost-effective super-resolution, emphasizing the importance of efficiency and accessibility in processing low-resolution images to higher resolutions. The study demonstrates a pragmatic approach to enhancing image quality suitable for real-time applications, highlighting the balance between computational expense and super-resolution performance.

In [2],Dong et al. (2016) pioneer the use of deep convolutional neural networks (CNNs) for image super-resolution (SR). They introduce a model that learns an end-to-end mapping between low and high-resolution images, marking a significant advancement over traditional methods. Their findings showcase the superior capability of deep CNNs to recover high-frequency details from low-resolution images, setting a new benchmark in the SR field.

In [5],Anwar et al. (2020) provide an extensive survey of super-resolution techniques, with a particular focus on deep learning-based methods. This comprehensive review traverses the evolution of SR, from early interpolation methods to sophisticated deep neural networks, offering insightful analysis into the progress and trends within the domain. The paper serves as a valuable resource for researchers seeking to understand the depth and breadth of super-resolution technologies.

In [1], Bashir et al. (2021) compile an exhaustive review of single image super-resolution (SISR) techniques driven by deep learning. Their work evaluates various deep learning architectures and their efficacy in SISR, providing a thorough overview of the state-of-the-art methods. This review is instrumental for researchers and practitioners aiming to grasp the current capabilities and limitations of deep learning in enhancing image resolution.

In [3], Ledig et al. (2017) introduce a groundbreaking approach to single image super-resolution using generative adversarial networks (GANs). Their method, known for generating photo-realistic textures, transcends traditional SR techniques by producing images that are not only high in resolution but also impressive in perceptual quality. This work stands as a milestone in leveraging adversarial learning for the refinement of super-resolution processes.

In [6],Wang et al. (2018) present ESRGAN, an enhanced version of GAN for super-resolution that improves upon prior models by introducing novel network architectures and training techniques. ESRGAN is celebrated for its ability to generate high-resolution images with exceptional detail and realism, pushing the boundaries of what is achievable in SR. This paper is pivotal for those looking to understand advancements in GAN-based super-resolution techniques.

2. Methodology

Extensive research in the literature has delved into the Single Image Super-Resolution (SISR) problem employing diverse deep learning approaches. Existing methodologies are broadly classified into nine groups based on their distinctive model designs. The overall taxonomy employed in this body of literature is depicted . Among these categories, we initiate our discussion with the initial and straightforward Interpolation techniques.

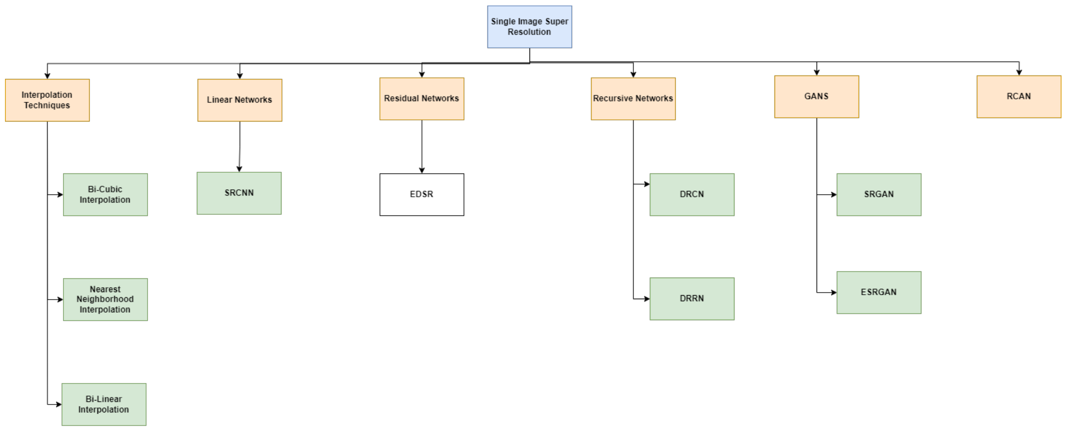

Figure 1. Methodology.

2.1. Interpolation Techniques

2.1.1 Bicubic interpolation:

This technique, widely utilized for image super-resolution, entails the fitting of a bicubic polynomial to the pixel values within a neighborhood surrounding the pixel undergoing interpolation. Bicubic interpolation typically yields smoother outcomes compared to alternative interpolation methods, although there is a risk of occasionally generating excessively smoothed images

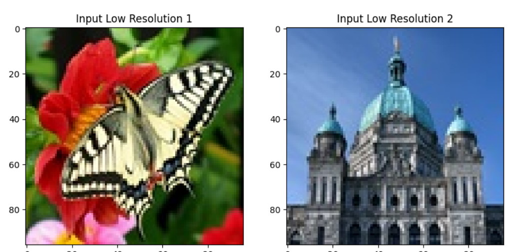

Figure 2. Input low Resolution.

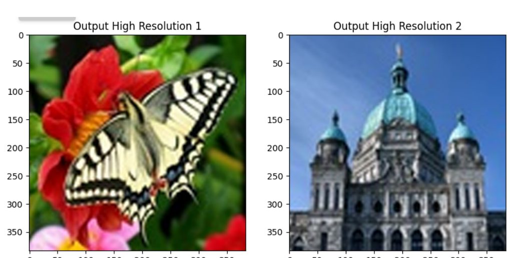

Figure 3. Bi-Cubic Interpolation.

2.1.2 Nearest neighbor interpolation:

This method entails straightforwardly copying the value of the nearest pixel to the pixel undergoing interpolation. Although it is a simple and quick technique, it may result in the creation of images with blocky and jagged edges.

2.1.3 Bilinear interpolation:

This method involves calculating the weighted average of the four nearest pixels to the pixel undergoing interpolation. It typically yields smoother results compared to nearest neighbor interpolation, although there remains a risk of producing blurry images.

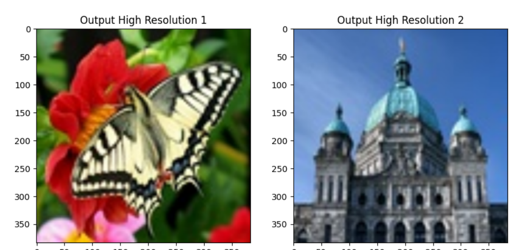

Figure 4. Bi-Linear Interpolation.

2.2. Linear Networks

Linear networks represent a category of deep learning architectures characterized by a straightforward structure featuring a singular pathway for signal propagation, devoid of skip connections or multiple branches. Their distinction lies in the method employed for the up-sampling operation, which can occur either through early or late upsampling mechanisms.

2.2.1 SRCNN:

The SRCNN (Super-Resolution Convolutional Neural Network), stands as a groundbreaking contribution to deep learning-driven super-resolution, serving as a source of inspiration for subsequent endeavors. This network exclusively incorporates convolutional layers and ReLU nonlinearity. Comprising three convolutional layers and two ReLU layers, arranged linearly, it undergoes training on a synthesized dataset utilizing the Mean Squared Error (MSE) loss function. The objective is to minimize the distinction between the output reconstructed high-resolution images and the ground truth high-resolution images.

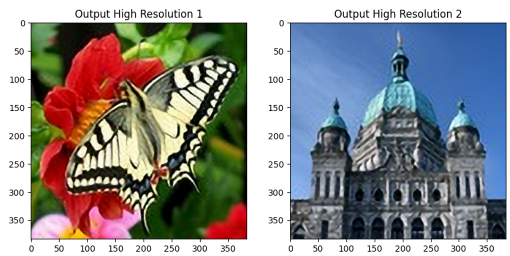

Figure 5. SRCNN(Super-Resolution Convolutional Neural Network).

2.3. Residual Networks

Residual learning, a technique employing skip connections in network architecture, serves to circumvent gradient vanishing and facilitates the design of deep networks. Initially applied in image classification, it has demonstrated enhanced performance in super-resolution tasks as well. In this approach, algorithms learn the discrepancy or residue between the input and the ground truth. These networks are classified into single-stage and multi-stage based on the number of stages utilized in their design.

2.3.1 EDSR:

The Enhanced Deep Super-Resolution (EDSR) technique adapts the ResNet architecture for super-resolution tasks, achieving enhancements by eliminating Batch Normalization layers and ReLU activation. In addition, they proposed the Multi-scale Deep SR (MDSR) architecture, which reduces parameter count through shared parameters and scale-specific layers. These methods employ a single loss for training and utilize data augmentation to create a self-ensemble, providing a comparable improvement to traditional ensemble-based models. EDSR and MDSR demonstrate superior performance compared to earlier architectures such as SRCNN, VDSR, and SRGAN, particularly in terms of quantitative metrics like PSNR.

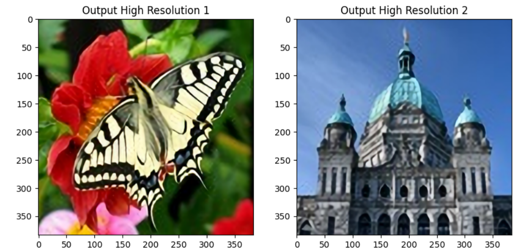

Figure 6. Enhanced Deep Super-Resolution (EDSR) .

2.4. Recursive Networks

Recursive networks utilize recursively connected convolutional layers or units to decompose the challenging super-resolution problem into a series of simpler sub-problems.

2.4.1 DRCN:

The Deep Recursive Convolutional Network (DRCN) applies identical convolution layers iteratively and comprises three distinct networks: the embedding net, the inference net, and the reconstruction net. The inference net executes super-resolution recursively through a single layer comprising convolution and ReLU operations, while the output is transformed to grayscale or color by the reconstruction net.

2.4.2 DRRN:

The Deep Recursive Residual Network (DRRN) presents a deeper architecture featuring up to 52 convolutional layers. It reduces network complexity by integrating residual image learning with local identity connections between compact blocks of layers within the network. DRRN employs recursive learning and duplicates a fundamental skip-connection block multiple times to establish a multi-path networks.

2.5. GANS:

GAN, short for Generative Adversarial Networks, represents a category of deep learning models utilized for generating new data resembling a provided dataset. GANs comprise two neural networks trained adversarially. The first network, known as the generator, generates new data samples. The second network, termed the discriminator, discerns between real data samples and fake data samples generated by the generator.

2.5.1 SRGAN

SRGAN, introduced by Ledig C., Theis L., and Huszar F., marks the pioneering application of Generative Adversarial Networks (GAN) in image super-resolution. Departing from previous approaches that overly prioritize PSNR, SRGAN addresses the issue of losing high-frequency details in reconstructed super-resolution images, aligning better with human perceptual expectations. The SRGAN architecture comprises two principal components: a generator network and a discriminator network. The generator network takes a low-resolution image as input and outputs a high-resolution image, leveraging deep convolutional neural networks (CNN) with skip connections to preserve image details and enhance resolution. Conversely, the discriminator network, also a CNN, distinguishes between the generated high-resolution image and the original high-resolution image.

SRGAN employs adversarial training, engaging in a two-player game between the generator and discriminator networks. The generator aims to produce high-quality images that deceive the discriminator, while the discriminator endeavors to discern between the generated and original high-resolution images. Through this adversarial process, the generator learns to generate high-quality, high-resolution images nearly indistinguishable from the originals.

SRGAN has exhibited impressive performance in generating high-quality images with fine details, finding applications in various domains such as medical imaging, satellite imaging, and digital photography. Unlike previous methods, SRGAN incorporates a perceptual loss function consisting of an adversarial loss and a content loss. The former guides the generator to produce high-resolution images based on extracted features, while the latter assesses the realism of the generated images compared to real high-resolution images. Utilizing VGG-based loss yields superior texture details compared to MSE-based loss, resulting in enhanced visual effects. Countermeasure loss and VGG loss have become prevalent techniques for improving super-resolution performance.

2.5.2 ESRGAN:

The Super-Resolution Generative Adversarial Network (SR-GAN) represents a landmark advancement capable of producing realistic textures in single image super-resolution. Nevertheless, the generated details often come with undesirable artifacts. In pursuit of enhancing visual quality, we meticulously scrutinize three crucial elements of SRGAN – network architecture, adversarial loss, and perceptual loss – and refine each to develop an Enhanced SRGAN (ESRGAN).

Specifically, we introduce the Residual-in-Residual Dense Block (RRDB) devoid of batch normalization as the fundamental building unit of the network. Additionally, we integrate the concept from relativistic GANs, enabling the discriminator to predict relative realness rather than absolute values. Furthermore, we enhance the perceptual loss by leveraging features before activation, providing robust supervision for brightness consistency and texture recovery.

With these enhancements, ESRGAN consistently delivers superior visual quality with more realistic and natural textures compared to SRGAN. ESRGAN clinched the top position in the PIRM2018-SR Challenge. It improves upon SRGAN by introducing RRDB without batch normalization, employing a relativistic discriminator for relative judgment, and enhancing the loss function with features before activation. Moreover, ESRGAN achieves a balance between PSNR evaluation and perceptual quality by blending a PSNR-oriented network with a GAN-based network.

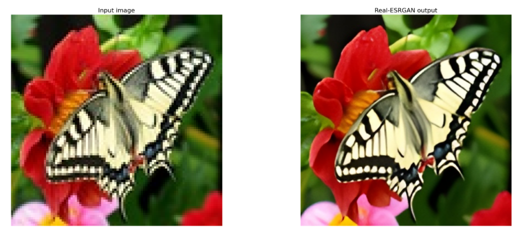

Figure 7. Enhanced$_{Super} - Resolution_{G}enerative_{A}dversarial_{N}etwork_{1}$.

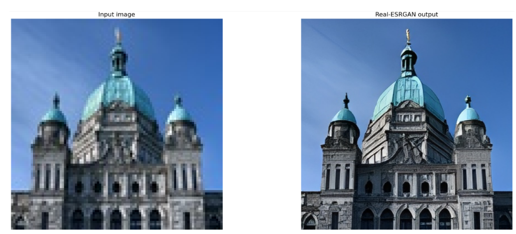

Figure 8. Enhanced$_{Super} - Resolution_{G}enerative_{A}dversarial_{N}etwork_{2}$.

2.6. RCAN:

RCAN, a deep learning-based image super-resolution technique, has demonstrated remarkable effectiveness in producing high-quality super-resolved images. The RCAN architecture incorporates residual connections and channel attention mechanisms to enhance the accuracy and visual quality of output images. In RCAN, a low-resolution image serves as input, undergoing a sequence of convolutional layers, residual blocks, and channel attention mechanisms to generate a high-resolution output image.

RCAN introduces a residual in residual (RIR) structure to construct a deeply layered network. This structure comprises several residual groups with long skip connections. Each residual group contains multiple residual blocks with short skip connections. The RIR design facilitates the bypassing of abundant low-frequency information through multiple skip connections, enabling the main network to focus on learning high-frequency details. Additionally, RCAN integrates a channel attention mechanism to adaptively rescale channel-wise features by considering interdependencies among channels.

Extensive experiments demonstrate that RCAN achieves superior accuracy and visual improvements compared to state-of-the-art methods.

3. RESULTS

The results obtained from the application of different methods are presented and analyzed in this section. The performance of each method was assessed using standard image quality metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The PSNR values provide a measure of the fidelity of the reconstructed images compared to the ground truth, while SSIM quantifies the similarity in structural information between the reconstructed and reference images.

Method	PSNR(Image1)	SSIM(Image1)	PSNR(Image2)	SSIM(Image2)
Bi-cubic Interpolation	20.09	0.71	21.23	0.7
Bi-linear Interpolation	19.47	0.69	20.93	0.67
SRCNN	21.48	0.73	21.5	0.72
EDSR	22.35	0.77	21.67	0.74
ESRGAN	21.36	0.75	18.92	0.66

Figure 9. PSNR and SSIM values.

4. CONCLUSION

It delineates the significance of SISR in meeting the demand for high-resolution images while navigating the challenges of direct capture. The survey covers classical interpolation methods and transitions to deep learning architectures.

Key methodologies, including linear networks, residual networks, recursive networks, and Generative Adversarial Networks (GANs), are explored. Seminal works like SRCNN, EDSR, ESRGAN, and RCAN are discussed, highlighting their architectural innovations and contributions.

Insights emerge regarding the efficacy of perceptual loss functions, adversarial training, and attention mechanisms in enhancing image quality. Notably, GAN-based models like SRGAN and ESRGAN leverage adversarial training for high-fidelity image generation.

The report emphasizes continual innovation and the transformative impact of deep learning on SISR, with implications spanning various domains. Overall, it underscores the evolution from classical techniques to cutting-edge methodologies in achieving high-resolution image reconstruction.

References

[1] Wang Y. Khan M. Arsalan Bashir, S. M. and Y Niu. “a comprehensive review of deep learning-based single image super resolution”,, 2021. arXiv e-prints. 2

[2] K. He C. Dong, C. C. Loy and X. Tang. ”image super-resolution using deep convolutional networks,”. 38(2):295–307, 2016. 2

[3] Ferenc Huszar Jose Caballero Andrew Cunningham Alejandro Acosta Andrew P. Aitken Alykhan Tejani Johannes Totz Zehan Wang Wenzhe Shi Christian Ledig, Lucas Theis. ”photo-realistic single image super-resolution using a generative adversarial network”, 2017. CVPR 2017: 105-114. 2

[4] K. Jafari M. H. Moaiyeri A. Z. Khuzani D. Khaledyan, A. Amirany and N. Mashhadi. ”low-cost implementation of bilinear and bicubic image interpolation for real-time image superresolution”, 2020. 2

[5] S. Khan S. Anwar and N. Barnes. ”a deep journey into super-resolution,”, 2020. ACM Comput. Surv. 53, 3 (2020), 1–34. 2

[6] Shixiang Wu Jinjin Gu Yihao Liu Chao Dong Yu Qiao Chen Change Loy Xintao Wang, Ke Yu. ”esrgan: Enhanced super resolution generative adversarial networks”, 2018. ECCV Workshops (5) 2018: 63-79. 2