Advancements in Single Image Super-Resolution Techniques

Le Dai
Department of Mathematics, University of Washington, Seattle, U.S.A.

Keywords: SISR, SRCNN, SRGAN, CARN.

Abstract: The technology of image super-resolution has been widely used in the industry for data recovery, graphic rendering and enhancing image quality. This paper offers a detailed overview of the progress made in Single Image Super Resolution (SISR) techniques, tracking the shift from traditional interpolation methods to cutting edge deep learning approaches like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs). In the early stages of research, researchers proposed using interpolation calculations such as bilinear, cubic and b interpolations for image super-resolution, but these methods lacked to produce high-quality super-resolved images. With the advancement of machine learning, experts have introduced some super-resolution techniques like CNNs and GANs that have significantly improved the quality of super-resolved images. This paper delves into the impact of these advancements on various applications and explores future research avenues in SISR, emphasizing the potential for further enhancements in image quality and the development of new algorithms for diverse applications.

1 INTRODUCTION

Digital image processing has seen advancements over the years thanks to the progress in computing power. SISR is considered a significant invention in the realm of digital image processing and has garnered considerable interest among researchers. Image super-resolution entails enhancing the resolution of an image while preserving fine details, leading to an overall enhancement in image quality. As hardware computing power and algorithms continue to advance, SISR has the potential to enhance images significantly. This advancement not only enhances the visual appeal of images but also serves important functions across different applications by presenting clearer, more detailed images from lower-resolution sources. The motivation behind researching SISR is driven by its range of applications across various fields. For instance, in the realm of video games SISR enables graphics processing units to produce higher-quality images while using computational power (Mengistu 2019, Watson 2020). Nvidia Deep Learning Super Sampling (DLSS) which relies on SR techniques has made it possible for high end games to run smoothly on lower end Personal Computers. In imaging SISR can help enhance the clarity of diagnostic images leading to more accurate diagnoses and better patient outcomes (Plenge 2012). Additionally, in satellite imagery applications it plays a role in improving image quality, for environmental monitoring, urban planning and disaster management (Müller et al. 2020, Yu et al. 2021, Zhang et al. 2022, Liu et al. 2022).

In the early days of working on improving image resolution, researchers used techniques like bilinear and bicubic interpolations. While these methods were efficient in terms of computation, the resulting images often lacked quality, showing problems such as blurriness and artifacts. In recent times, thanks to advancements in deep learning (Siu & Hung 2012, Wang et al. 2020). Especially the introduction of CNNs and GANs. There has been significant progress in image super resolution. These modern data driven approaches have allowed the creation of models that can learn from large datasets to accurately reconstruct high resolution images. This has led to a remarkable enhancement in image quality and a reduction in common issues seen with traditional interpolation methods. Single Image Super Resolution (SISR) encounters difficulties in enhancing information, from low resolution images leading to some loss of data. Moreover, striking a balance between computation and the quality of enhanced images is crucial for applications that require real time processing. The emergence of Generative Adversarial Networks (GANs) such as SRGAN, aimed to tackle these challenges by enhancing image quality.

Dai, L.
Advancements in Single Image Super-Resolution Techniques.
DOI: 10.5220/0012832900004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 160-167
ISBN: 978-989-758-690-3

The paper is structured to reflect the significant developments in Single Image Super Resolution (SISR), focusing on balancing data-driven accuracy with computational efficiency. Chapter two introduces traditional SISR techniques, setting the stage for understanding foundational methods. Chapter three shifts to advanced models like CNNs, GANs, and specifically CARN, highlighting their impact on improving resolution while considering practical application constraints. The final chapter concludes with a summary of key findings and explores potential future research directions, emphasizing the ongoing quest for more efficient and higher-quality SISR methods. This paper provides a concise yet comprehensive overview of the field's evolution and current challenges.

2 PERFORMANCE EVALUATION METRICS

2.1 Peak Signal to Noise Ratio

Peak Signal to Noise Ratio (PSNR) is a metric commonly utilized to evaluate image quality. It quantifies the disparity between two images, typically an original image in low quality (I) and its reconstructed or super-resolved counterpart in high quality(K), through the following formula:

$$PNSR = 10log_{10}\frac{MAX_I^2}{MSE} \quad (1)$$

In this context, MAXI signifies the maximum possible pixel value in the image, and MSE(I, K) represents the Mean Squared Error between the original and the reconstructed images. A higher PSNR value is indicative of minor discrepancies between I and K, implying superior image quality. Nonetheless, despite the straightforward nature of this metric and its precise quantification of reconstruction errors, PSNR may not consistently align with human visual perception, hence possibly inaccurately representing the perceived quality of images.

2.2 Structural Similarity Index

The Structural Similarity Index (SSIM) is designed to overcome the limitations of traditional metrics like PSNR by considering more comprehensive aspects of image quality such as detail, brightness, and contrast. (9) SSIM evaluates the similarity between two images in a way that is more aligned with the eyes of humans. The formula for computing SSIM is given by(9):

$$\mathrm{SSIM}(I,K) = \frac{(2\mu_I\mu_K + c_1)(2\sigma_{IK} + c_2)}{(\mu_I^2 + \mu_K^2 + c_1)(\sigma_I^2 + \sigma_K^2 + c_2)} \quad (2)$$

The original and super-resolved images are denoted by I and K, respectively, and their average luminance values are μ_I and μ_K. Their variances are σ_I^2 and σ_K^2 , while the covariance between I and K is σ_{IK}. The constants c_1 and c_2 are added to stabilize the division with a weak denominator.

2.3 Learned Perceptual Image Patch Similarity

Learned Perceptual Image Patch Similarity (LPIPS) utilizes deep learning to assess image quality in a manner that aligns closely with human visual perception. It addresses the limitations of traditional metrics like PSNR and SSIM by incorporating variations in human perception. LPIPS calculates similarity by analysing image patches through deep neural networks, effectively capturing perceptual differences that may be overlooked by other metrics. This method offers a nuanced understanding of image quality, proving especially beneficial in applications requiring high visual fidelity, such as medical imaging, where preserving detail is paramount (Zhang et al. 2018).

The suitability and limitations of these metrics depend on the context. For example, while PSNR works well for quantifying signal reconstruction quality it may not be the reliable indicator when visual fidelity is crucial. On the hand SSIM and LPIPS provide a more nuanced evaluation of image quality, which is especially important in fields like medical imaging where preserving fine details is essential (Wang et al. 2020).

3 KEY DATASETS IN SISR RESEARCH

3.1 Set5 and Set14

The Set5 dataset, introduced in 2012, comprises five high-resolution images, including a variety of scenes and objects to test the robustness of super-resolution methods across different content types (Bevilacqua et al. 2012). As shown in FIGURE 1, images in Set5 are carefully selected to represent common photographic subjects, such as landscapes, animals, and urban scenes.

Figure 1. Image in SET5 (Bevilacqua et al. 2012).

On the other hand, as shown in FIGURE 2, the Set14 dataset (Zeyde et al. 2010), presented in 2013, extends the variety and challenges by including 14 high-resolution images. This dataset broadens the scope with a more diverse set of scenes and objects, ranging from text and graphics to natural landscapes and architectural elements. The images in Set14 are chosen to challenge super-resolution algorithms with a wider range of textures, details, and spatial complexities.

Both datasets provide images with resolutions varying from 200x200 to 500x500 pixels, catering to the need for evaluating algorithms at different scales and complexities. For training and testing purposes, these datasets are commonly used in their original high-resolution form to benchmark the quality of unsampled images against the ground truth.

Figure 2. Image in SET14 (Zeyde et al. 2010).

3.2 DIV2K

As shown in FIGURE 3, The DIV2K dataset is a high-quality resource introduced for advancing image super-resolution(SR) and other related tasks (Timofte 2017). Created by the research community, it features 2,000 diverse, high-resolution images sourced from a variety of scenes, including urban, rural, and natural environments. The images in DIV2K have a high resolution, ranging from 2K to 4K, making it particularly suitable for training and benchmarking SR algorithms.

This dataset is split into training sets with 800 images, validation sets with 100 images and test sets with 100 images, enabling a structured evaluation of model performance. For SR tasks, models are trained on lower-resolution images. The goal is to reconstruct the high-resolution image from its down-sampled version, with the original images serving as the ground truth for assessing the quality of the reconstruction.

DIV2K is notable for its large scale and high image quality, providing a challenging and comprehensive benchmark for super-resolution models. It has become a standard dataset in thefield, supporting not just SR research but also applications in image enhancement, compression, and computer vision at large.

Figure 3. Image in DIV2K (Timofte 2017).

3.3 Urban 100

As shown in FIGURE 4, The Urban100 dataset is widely acknowledged as a standard, for testing super resolution (SR) algorithms in settings. It consists of 100 high quality images showcasing landscapes architectural elements and detailed textures like buildings, bridges and street views. This dataset is renowned for its content that includes geometric shapes, straight lines and intricate details challenging for SR algorithms to faithfully recreate (Jb et al. 2015).

Urban100s image selection aims to assess the capabilities of SR methods in handling structures and textures effectively. Due to its resolution the dataset serves as a platform for evaluating how well SR models perform on real world scenes. Researchers utilize this dataset to gauge the model's effectiveness in enlarging low resolution images by factors of 2x, 3x or 4x while maintaining or enhancing the clarity and detail from the higher resolution images.

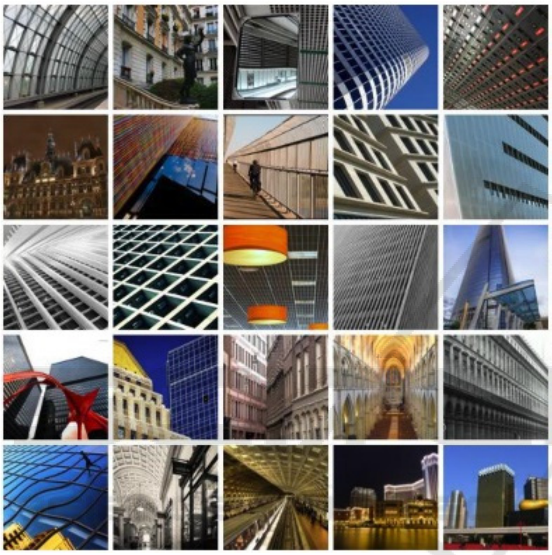

Figure 4. Image in Urban100 (Jb et al. 2015).

4 APPROACHES TO ENHANCING IMAGE RESOLUTION

4.1 Linear Approaches

The journey of SISR began with approaches that laid down the foundation for future advancements in this field. Traditional methods of interpolation, including bilinear, bicubic and nearest neighbor interpolation played a crucial role in shaping our understanding of upscaling images.

Linear interpolation, as shown is FIGURE 5, is one of the methods used estimates pixel values in high resolution images based on linear estimation techniques. It calculates pixels by considering the straight-line distance, between known pixel values. Linear interpolation although it is efficient in terms of computation can often lead to resolved images that have noticeable flaws, especially around edges where there are sudden changes in pixel values.

Bilinear interpolation, a more advanced technique than linear interpolation takes into con- sideration not only the linear distance but also the two-dimensional spatial relationship between pixels. It calculates an estimated value for a pixel by taking a weighted average of the four closest known pixels located diagonally. This method offers an improvement in the smoothness of upscaled images compared to linear interpolation. However, it still falls short when it comes to preserving high frequency details like edges and tends to result in slightly blurred outputs.

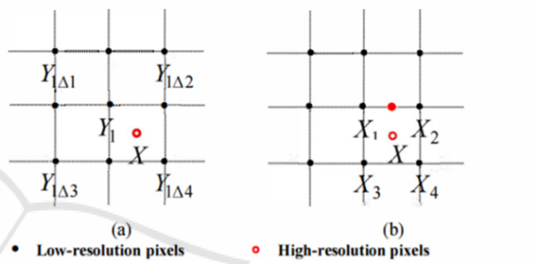

Figure 5. Graphical illustration of spatial positions of LR and HR pixels for linear interpolation (Siu & Hung 2012).

Nearest neighbor interpolation on the hand is the simplest method that assigns the valueof the nearest pixel to an unknown pixel. While this approach is extremely fast and doesn't require computation power it often produces blocky and pixelated images. It doesn't introduceany information during the upscaling process and therefore may not be suitable for applicationswhere image quality is crucial.

Although these initial approaches had limitations in their ability to generate quality and de- tailed images, they played a crucial role in paving the way, for more complex and sophisticatedtechniques. They emphasized the difficulties in SISR such as the need to improve detail preservation and reduce artifacts. These challenges have been the driving force behind the development of advanced super resolution techniques. Over time these traditional methods have served as benchmarks for more innovative approaches highlighting their ongoing importance in image processing.

4.2 Convolutional Neural Networks

The emergence of deep learning has had a significant impact on SISR. Among the deep learning methods CNNs have been particularly effective leading to a new era in SISR with models like the Super Resolution Convolutional Neural Network (SRCNN) being introduced. In this section we will explore how CNN based methods have transformed the field with a focus on SRCNN as a pioneering example. SRCNN, introduced by Dong et al. In 2014 was one of the earliest deep learning models that utilized CNNs for super resolution tasks. It represented a shift from interpolation-based methods by offering a fresh approach that learned an end-to-end mapping, between low-resolution images and high-resolution images. The structure of SRCNN comprises three layers; a layer that extracts and represents patches, a layer that performs nonlinear mapping and a reconstruction layer. This streamlined design allows SRCNN to directly learn the upscaling function from data, which's a major improvement compared to the manual feature engineering required in traditional methods (Dong et al. 2014).

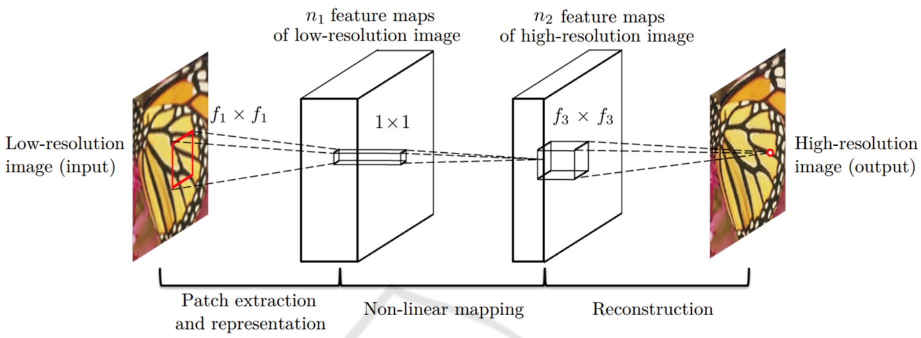

Figure 6. Graphical illustration of the process of SRCNN super resolution (Dong et al. 2014).

As Shown in FIGURE 6, In the layer of SRCNN overlapping patches are extracted from the low-resolution input and represented as high dimensional vectors. This process effectively captures the structure within these patches setting the foundation for subsequent layers. The second layer, which is the core of SRCNN performs a linear mapping of these vectors to another high dimensional space. In this space these vectors are expected to represent features of the high-resolution image. Finally in the layer the high-resolution image is reconstructed using these mapped vectors (Dong et al. 2014).

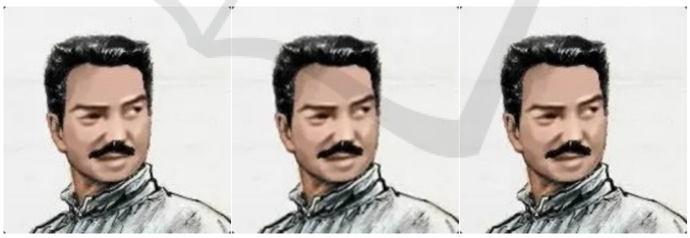

Figure 7. Graphical comparison of the process of Original Image Quality Bicubic interpolation and SRCNN super resolution (Picture credit: Original).

As shown in FIGURE 7, what distinguishes SRCNN from methods is its ability to learn intricate and hierarchical representations of image data. Unlike interpolation techniques that use fixed formulas for upscaling SRCNN adapts its approach based on the training data it receives. This adaptability leads to improved image quality, with enhanced details and sharper edges while reducing common artifacts observed in earlier methods. Furthermore, the remarkable achievements of SRCNN have paved the way for exploration and advancements in deep learning techniques for Single Image Super Resolution (SISR). It has served as a catalyst, inspiring models that build upon and refine the foundations laid by SRCNN. These models have continually elevated the efficiency, accuracy and quality of resolution methods showcasing the immense potential of Convolutional Neural Networks (CNNs) in image processing.

In essence the introduction of CNN based approaches like SRCNN has sparked a revolution in SISR. It has shifted the focus from feature engineering to data driven and adaptable learning methodologies. This paradigm shift not enhances the quality of super-resolved images but also drives continuous innovation in deep learning-based image super resolution techniques.

4.3 Generative Adversarial Networks

Generative Adversarial Networks (GANs) have emerged as an element in advancing Single Image Super Resolution (SISR) thanks to their ability to generate high resolution images that appear convincingly natural. The Super Resolution Generative Adversarial Network (SR-GAN) stands out as a model within this domain. It introduces a concept where the upscaling process is portrayed as a battle, between two networks: a generator and a discriminator (Ledig et al. 2016).

As shown in FIGURE 8, The primary objective of SRGANs generator is to produce high resolution images based on low resolution inputs. It has been trained to deceive the discriminator, which is createdto differentiate between resolved images and genuine high-resolution images. The role of the discriminator is crucial as it guides the generator in producing results that're increasingly difficult to distinguish from real high-resolution images. This adversarial training is what enablesSRGAN to bring back textural details that are often lost during the upscaling process.

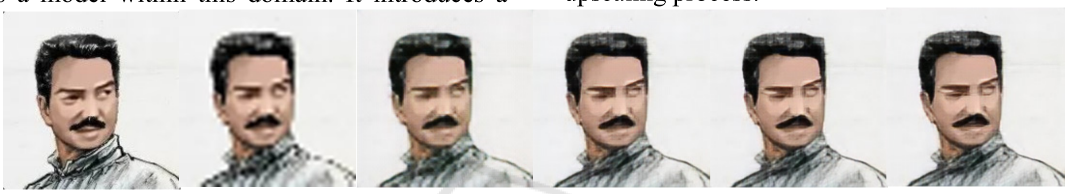

Figure 8. (Graphical comparison of the process of Original Image, Quality Bicubic interpolation, SRGAN in Epoch 0, 500, 1000, 1400(Picture credit: Original)

As shown in FIGURE 9, the architecture of SRGAN stands out for its convolutional neural networks that learn multiple layers of image features ranging from basic to complex. This allows for a restoration of details. Furthermore, SRGAN incorporates a loss function that goes beyond traditional pixel wise loss functions. It takes advantage of feature maps from trained networks capturing significant differences at a higher level and promoting visually pleasing solutions for human observers.

The success of SRGAN has been demonstrated in benchmarks where it has displayed superior performance in terms of both quantitative measurements and qualitative assessments. It particularly excels in situations where capturing details and textures are crucial such as medical imaging, interpreting satellite images and enhancing entertainment media.

However, despite these advancements GAN based methods, like SRGAN present their set of challenges. The process of training can sometimes be unstable. There is often a balance to be struck between achieving high resolution and preserving the natural statistics of images.Nevertheless, the development of SRGAN has inspired research and advancements in more sophisticated super resolution models based on GANs. These models continuously push the boundaries of what's possible in the realm of SISR. The influence of SRGAN and its successors extends into the future promising more realistic upscaling capabilities that have the potential to revolutionize various applications of digital imaging.

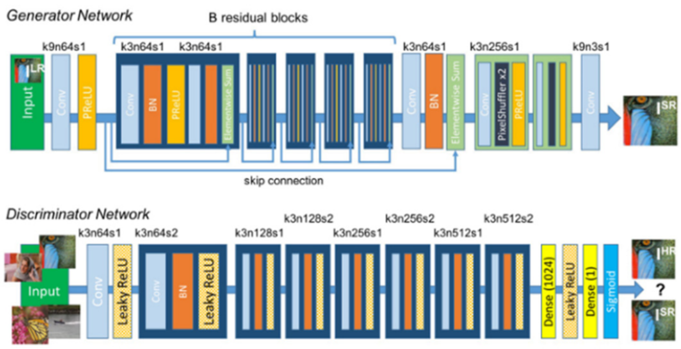

Figure 9. Graphical illustration of the process of SRGAN super resolution (Ledig et al. 2016).

4.4 Convolutional Anchored Regression Network

One noteworthy advancement in SISR is the Convolutional Anchored Regression Network (CARN). CARN prioritizes efficiency without compromising on the quality of images. It is designed to be a model that can easily operate on devices with limited computational resources. As shown in FIGURE 10, CARN utilizes a cascading mechanism where multiple levels of feature maps are extracted and combined. This allows the network to focus on details within an image at various scales resulting in high resolution reconstructions, with significant levels of detail (Ahn et al. 2018).

The foundation of CARNs architecture lies in its ability to establish connections across scales enabling the efficient utilization of hierarchical features. This plays a role in achieving its impressive performance. Moreover, CARN employs a method of residual learning to progressively refine its output. During training the network is guided by a loss function that promotes preserving the content of an image while enhancing its textural details. This delicate balance ensures that the super-resolved output maintains both fidelity and naturalness.

Numerous evaluations on benchmark datasets using commonly used SISR metrics such as PSNR and SSIM have demonstrated that CARN performs on par with, if not better than, more complex models. Its robust design marks a milestone towards practical real-world applications of SISR technology catering to the growing demand for high quality image upscaling on edge devices.

In contrast to approaches like bilinear and bicubic interpolation which are computationally efficient but often result in blurry images during upscaling due to their limited detail representation SRCNN introduced a revolutionary deep learning model for SISR. By learning an end-to-end mapping process SRCNN brought improvements in image quality while demanding considerable computational resources. On the hand SRGAN took things further by incorporating GANs into SISR with a focus on perceptual quality rather than pixel accuracy. This approach led to high resolution images, with realistic textures but came at the risk of potential training instability and artifacts. CARN, a recent approach provides a lightweight yet effective solution that achieves competitive performance. It strikes a balance between efficiency and image quality making it suitable for devices that have limited computational capabilities.

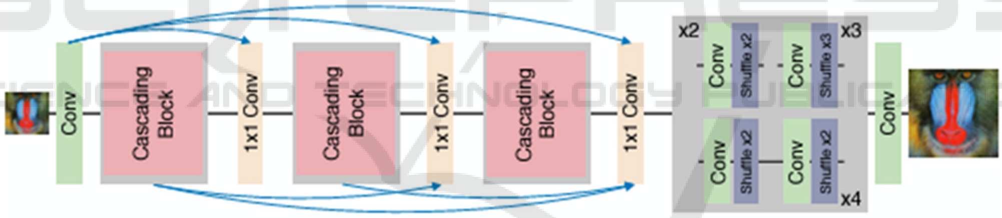

Figure 10. Graphical illustration of the process of SRGAN super resolution (Ahn et al. 2018).

5 CONCLUSION

This paper explores the progress, in enhancing image quality through technology examining the background of research methods for evaluating image quality datasets used and the principles behind algorithms. This paper talked about approaches like SRCNN, SRGAN and CARN when it comes to algorithms.

While deep learning-based techniques for enhancing image quality have made strides issues such as computational requirements and the necessity for efficient models hinder their use on edge devices. Future research should focus on making models lighter and faster to address these challenges reducing reliance, on hardware resources and enabling real world applications.

REFERENCES

A. Watson. Deep Learning Techniques for Super-Resolution in Video Games. preprint arXiv:2012.09810 (2020).

C. Dong, C. Loy, K. He, X. Tang, Image Super-Resolution Using Deep Convolutional Networks, 2014.

C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2016.

E. Plenge, Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time. 68, 1983–1993 (2012).

F. Liu, Q. Yu, L. Chen, Aerial image super-resolution based on deep recursive dense network for disaster area surveillance, Pers Ubiquit Comput, 26, 1205-1214, 2022.

H. Jb, A. Singh, N. Ahuja, "Single image super-resolution from transformed self-exemplars ", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (2015), pp. 5197-5206.

M. Bevilacqua, A. Roumy, C. Guillemot, A. More, Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding, 2012.

M. Müller, N. Ekhtiari, R. Almeida, Super-resolution of multispectral satellite images using convolutional neural networks. arXiv preprint arXiv:2002.00580 (2020).

Mengistu, Biruk, Deep-Learning Realtime Upsampling Techniques in Video Games, Classic Physiques, 10(2), 212–213 (2019).

N. Ahn, B. Kang, K. Sohn, Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network, 2018.

Q. Yu, F. Liu, L. Xiao, Z. Liu, X. Yang, Real-Time Environment Monitoring Using a Lightweight Image Super-Resolution Network, Int J Environ Res Public Health, 18(11), 2021.

R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results," in 2*017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Honolulu, HI, USA, 2017), pp. 1110-1121.

R. Zeyde, M. Elad, M. Protter, On Single Image Scale-Up Using Sparse-Representations, in Lecture Notes Comput, 6920, 711-730, 2010.

R. Zhang, P. Isola, A. Efros, E. Shechtman, O. Wang, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, Available: arXiv:1801.03924, 2018.

W. Siu, K. Hung, "Review of image interpolation and super-resolution", in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, (Hollywood, CA, USA, 2012), pp. 1-10.

Y. Zhang, R. Zong, L. Shang, D. Wang, On Coupling Classification and Super-Resolution in Remote Urban Sensing: An Integrated Deep Learning Approach, IEEE Transactions on Geoscience and Remote Sensing, 60, 1-17, 2022.

Z. Wang, J. Chen, S.C Hoi, Deep Learning for Image Super-resolution: A Survey, 2020.