TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 23, No. 4, August 2025, pp. 1032~1045
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i4.26105

Deep learning-based image super-resolution using generative adversarial networks with adaptive loss functions

Hani Q. R. Al-Zoubi
Department of Computer Engineering, Faculty of Engineering, Mutah University, Karak, Jordan

Article Info

Article history:

Received Feb 23, 2024
Revised Apr 8, 2025
Accepted May 10, 2025

Keywords:

Adversarial loss
Deep learning
Generative adversarial networks
Image resolution
Loss function
Pixel loss
Prior loss

ABSTRACT

This study investigates deep learning based single image super-resolution (SISR) and highlights its revolutionary potential. It emphasizes the significance of SISR, and the transition from interpolation to deep learning-driven reconstruction techniques. Generative adversarial network (GAN)-based models, including super-resolution generative adversarial network (SRGAN), video super-resolution network (VSRResNet), and residual channel attention-generative adversarial network (RCA-GAN) are utilised. The proposed technique describes the loss functions of the SISR models. However, it should be noted that the conventional methods frequently fail to recover lost high-frequency details, which signify their limitations. The current visual inspections indicate that the suggested model can perform better than the others in terms of quantitative metrics and perceptual quality. The quantitative results indicate that the utilised model can achieve an average peak signal-to-noise ratio (PSNR) enhancement of X dB and an average structural similarity index (SSIM) increase of Y. A range of improvements of 7.12-23.21% and 2.75-10.00% are obtained for PSNR and SSIM, respectively. Also, the architecture deploys a total of 2,005,571 parameters, with 2,001,475 of these being trainable. These results highlight the model's efficacy in maintaining key structures and generating visually appealing outputs, supporting its potential implications in fields demanding high-resolution imagery, such as medical imaging and satellite imagery.

Corresponding Author:

Hani Q. R. Al-Zoubi
Department of Computer Engineering, Faculty of Engineering, Mutah University
Mutah, 61710 Al-Karak, Jordan
Email: hanirash@mutah.edu.jo

1. INTRODUCTION

In image processing, single image super-resolution (SISR) is a significant subfield [1]. Additionally, it seeks to recover a high-resolution picture from a low-resolution (LR) image [2], leading to its numerous uses in disaster assistance, video monitoring, and medical diagnostics, among other fields. Higher-quality photographs, for instance, can enable medical professionals to identify disorders with greater accuracy [3].

The goal of SISR, a crucial step in image processing, is to increase the resolution of imaging systems. Deep learning has recently helped SISR advance significantly and produce encouraging results [4]. Thus, both academics and industry find great value in researching SISR. Researchers have proposed many approaches to address the SISR issue based on degradation models of low-level vision tasks. SISR generally falls into three categories: image-specific data, previous knowledge, and machine learning. It was an easy and effective approach in SISR to directly magnify the resolutions of all the pixels in an LR picture by an interpolation method to get an high-resolution image, such as closest neighbor interpolation, bilinear interpolation, and bicubic interpolation. It should be noted that the up-sampling step in these interpolation

approaches causes high-frequency information to be lost [5]. Instead, SISR-based reconstruction-based approaches were created using optimization techniques. In other words, estimating the registration parameters by mapping a projection into a convex set can bring back additional SISR features [6].

In several artificial intelligence disciplines, such as computer vision [7] and natural language processing [8], deep learning [6] showed superior performance than conventional machine learning models. Numerous deep learning-based solutions were developed for SISR due to the rapid development of deep learning techniques, constantly advancing the state-of-the-art (SOTA) [3].

The SISR job may typically be broken down into three parts, like other image transformation tasks: feature extraction and representation, nonlinear mapping, and picture reconstruction [9]. Designing an algorithm that meets these requirements is time-consuming and ineffective in conventional numerical models. Instead, deep learning can move the SISR activity to a nearly end-to-end architecture that includes all three of these procedures, which can significantly reduce human and computer costs [10]. Additionally, SISR, which can result in unstable and complex convergence of the findings, can be mitigated by deep learning through effective network design and loss function formulation. Modern GPUs also make it possible for deeper, more complicated deep learning models to be trained quickly, and these models exhibit better representational power than conventional numerical models. The distinction between supervised and unsupervised deep learning-based approaches is widely understood. The range of this categorization criterion is extensive and unclear while being the most basic. As a result, many methodologically unconnected ways may be grouped into the same type, whereas methodologically related methods using similar strategies may be grouped into entirely different kinds [11].

2. LITERATURE REVIEW

In a specific study, Lucas *et al.* [5] provided a generative adversarial network (GAN)-based formulation for VSR to correctly direct video super-resolution network (VSRResNet) during the GAN training. The researchers created a novel generator network dubbed VSRResNet tuned for the VSR issue. To make their final VSRResFeatGAN model, they further improved their VSR GAN formulation by adding two regularizers: a feature-space distance loss and a pixel-space distance loss. The researchers also demonstrated that pre-training their generator with the mean-squared error loss only quantitatively outperforms the most advanced VSR models currently available. To compare the most recent VSR models, they used the PercepDist measure. In contrast to the often-used peak signal-to-noise ratio (PSNR)/structural similarity index (SSIM) measures, their research demonstrated that this metric better assesses the perceptual quality of SR solutions derived by neural networks.

An introduction of super-resolution generative adversarial network (SRGAN), a GAN for picture super-resolution (SR), was made by Ledig *et al.* [6]. According to their claim, it was the first framework capable of predicting natural pictures that are photo-realistic for scaling factors 4. To accomplish this aim, the researchers provided a perceptual loss function that combines an adversarial loss with a content loss. With a discriminator network trained to distinguish between super-resolved pictures and authentic photo-realistic images, the adversarial loss pushes their solution to the natural image manifold. In addition, rather than using similarity in pixel space, they applied a content loss inspired by perceptual similarity. Their deep residual network can restore photo-realistic textures from severely down sampled photos on open benchmarks. A thorough mean opinion score (MOS) test revealed substantial improvements.

Though often less quality, photos are captured using remote sensing imaging equipment. To train deep neural networks, there are not enough high-resolution remote sensing photos accessible. Concerning this issue, Zhang *et al.* [7] provided an unsupervised SR approach that does not need high-resolution remote-sensing photos. In the proposed method, a GAN was used to get SR pictures from the generator. The SR images were then down-sampled to provide LR ideas for training the discriminator. This method surpassed several others in terms of the quality of the SR pictures produced as determined by six evaluation metrics demonstrating the excellent performance of the suggested unsupervised strategy for enhancing the spatial resolution of remote sensing images.

A residual channel attention-generative adversarial network (RCA-GAN) was developed by Cai *et al.* [8] to address the associated issues of visual quality of natural textures. An innovative residual channel attention block was suggested to build RCA-GAN, which comprises a collection of residual blocks with shortcut connections and a channel attention mechanism to mimic the interdependence and interaction of feature representations across multiple channels. Furthermore, a GAN generated realistic and highly detailed outputs. Taking advantage of these advancements, the suggested RCA-GAN genertaed remarkable visual quality with more realistic and natural textures than baseline models and attains equivalent or better performance in a comparison to state-of-the-art approaches for real-world picture SR.

Liu *et al.* [9] looked at a deep-learning picture SR technique frequently applied to face recognition, video perception, and other areas—usually, the high-frequency texture details in these GANs. LR photos

might be used as a reference image by transferring the pertinent textures from high-resolution photographs. The most recent techniques employ transformer principles to transfer relevant textures to LR pictures. However, channel learning and intricate textures continue to pose some challenges. As a result, Liu *et al.* [9] suggested an enhanced texture transformer network (ETTN) to increase the channel learning capacity and texture details. This technique was able to take the necessary structural data from high-resolution texture pictures and convert them to LR texture images. Finding thefeature map in this way allows one to alter the details of a picture and enhances channel-to-channel learning. The impact of fusion was then further enhanced using multi-scale feature integration (MSFI), which allowed us to accomplish varying levels of texture restoration. The experiments showed that the model has a decent resolution-enhancing influence on texture converters. PSNR and SSIM were enhanced by 0.1–0.5 dB and 0.02, respectively, in various datasets.

Le-Tien *et al.* [10] intended to apply deep learning to the inverse problem, one of the most well-known problems that have been a source of worry for a long time: image SR. Since then, many machine learning and deep learning domains have gained pace in attempting to solve these imaging challenges. The researchers reviewed deep learning methods for handling picture SR, concentrating on the GAN method, and discussed further implications for the GAN to complete the job successfully. More specifically, they examined the enhanced SR generative adversarial network (ESRGAN) and residual in residual dense network (RRDN) introduced by the "idealo" team and assessed their performance for image SR. These methods produced accurate results that earned them a high ranking on the leaderboard of cutting-edge techniques with many other datasets, such as Set5, Set14, or DIV2K. To be more precise, by retraining the suggested model with different parameters and comparing it with their output, the researchers inspected the SRGAN and ESRGAN, two renowned state-of-the-art methods.

For researching Mars' landform characteristics and examining its climate, high-resolution Mars photos are essential. Modern picture SR techniques are more effective than older ones since they are based on deep learning or convolutional neural networks (CNNs). However, these deep learning-based algorithms often use an ideal down-sampling method (such as bicubic interpolation) to create LR pictures. The current SR techniques have two drawbacks: (i) when evaluated on an ideal dataset, the paired LR high-resolution data utilizing such approaches can produce a satisfying result. However, as genuine Mars photographs rarely adhere to a perfect down-sampling norm, these approaches could be more effective for the SR of authentic Mars images; and (ii) the super-resolved photos are not realistic in texture details because the LR images produced by perfect down-sampling algorithms have no noise, but real Mars photographs typically do.

Wang *et al.* [11] provided a unique two-step approach for Mars image SR to address the abovementioned issues. They concentrated on creating a novel degradation framework by predicting blur kernels to solve restriction one specifically. A GAN was trained to produce noise distribution to meet restriction. Extensive tests on the Mars32k dataset elaborated the competence of the proposed method, and outperformed existing SOTA methods in terms of quality and quantity of findings.

Lin *et al.* [12] suggested a deep unsupervised learning method for SR using a framework called a GAN, which consists of a discriminator and a deep convolutional generator network with dense connections. The inputs were upscaled using a sub-pixel convolutional layer operated on top of the generator, and all of the usual convolutions are performed in the LR space, resulting in a quick restoration. When given a LR image, the generator is taught to recover the high-resolution image immediately. To distinguish the high-resolution images from the produced high-resolution images, the discriminator used strided convolution and ReLU activations. Local-global content consistency and pixel faithfulness were guaranteed by the generator model's optimization, which combines a data error, a common term, and an adversarial loss. However, no labeled training data were used for the training. Experimental findings and comparisons with several cutting-edge supervised learn-based methods showed that the proposed model achieves a comparable result in both quantitative and qualitative measurements. The results also ascertained the reliability and efficiency of the suggested unsupervised learning-based SISR algorithm.

A deep learning approach based on a GAN was introduced by Liu *et al.* [13] to achieve SR in coherent imaging systems. This approach can improve the resolution of cohesive imaging systems restricted by diffraction and pixel size. Super-resolving complex valued pictures captured using a lensless on-chip holographic microscope, whose resolution was pixel size-limited, were used to validate the capabilities of the proposed technique experimentally. A lens-based holographic imaging system with resolution constrained by the numerical aperture of its objective lens was also enhanced using the same GAN-based method. Image data and convolutional neural networks can improve the space-bandwidth product of coherent imaging systems with the help of this deep learning-based SR framework.

Xu *et al.* [14] offerred a administered generative adversarial nets technique to recover high-resolution chest X-ray CXR pictures from LR counterparts while preserving pathological invariance. Particularly, the supplementary label information was added to limit the feature creation to combat the possible danger of pathological variation. Then, with the assurance of theoretical validation in managing the

Lipschitz bound of the discriminator, spectral normalization was constructed to regulate the performance of the discriminative network. Compared to contemporary state-of-the-art methodologies, quantitative and qualitative evaluations showed that the proposed method can provide more authentic CXR SR improvement. On two datasets, the proposed technique surpassed the average by 13.0%, 12.2% in FSIM, and 13.7%, 12.5% in MSIM. Furthermore, the GAN-train and GAN-test generative performance indices genertaed average increments of 9.3% and 10.5% on the CXR2 dataset, respectively. Regarding pathological invariance and acceptability, subjective evaluation on SR CXR surpassed average scores of 0.425 and 0.525, respectively.

By employing the GAN framework, Prajapati *et al.* [15] suggested a novel SR method called USISResNet to create high-quality SR images for perceptual examination. They also offerred an unknown loss function based on the MOS. Extensive tests on the validation (Track-1) set of the NTIRE-2020 real-world SR challenge and testing datasets (Track-1 and Track-2) were used to confirm the efficacy of the proposed architecture. The researchers compared real-world photos to other cutting-edge techniques that use synthetically down-sampled LR images to show the proposed network's generalizability. The suggested network was also tested on the NTIRE 2020 real-world SR challenge dataset, where the method demonstrates dependably accurate results.

Most image SR methods based on deep learning do not use down sampling during the reconstruction phase. Given this reality and motivated by the iteration concept, they put forth a unique image SR technique based on deep CNN and the down-sampling iterative module that investigates a new fundamental iterative module integrating up- and down-sampling processes. The intermediate LR prediction and the high-resolution picture are produced at each iteration of the iterative module. The weighted aggregate of the middle-anticipated images produced by several rounds yields the final reconstructed result. They use the adaptive loss function to achieve quick convergence and precise reconstruction during training. Regarding objective performance evaluation and visual effects, thorough experimental comparisons and analyses demonstrate that this method is superior to specific cutting-edge techniques [16].

To present a thorough overview of recent developments in SISR from the standpoint of deep learning, Yang *et al.* [17] provided information on the first traditional approaches for image SR. The review divides image SR techniques into four groups: classical strategies, supervised learning-based methods, unsupervised learning-based techniques, and domain-specific SR techniques. Yang *et al.* [17] discussed the issue of SR to offer insight into picture quality measurements, accessible reference datasets, and SR difficulties. Using a reference dataset, deep learning-based techniques to SR were assessed. The cycle-in-cycle GAN (CinCGAN), the multiscale residual network (MSRN), the meta residual dense network (Meta-RDN), the recurrent back projection network (RBPN), and the second-order attention network are some of the state-of-the-art image SR techniques that were studied.

A deep learning-based technique for medical imaging SR dubbed Medical images SR using generative adversarial networks (MedSRGAN) was created by Jia *et al.* [18]. A unique convolutional neural network called the residual whole map attention network (RMAN) was built to extract usable information from various channels and focus more on significant regions. Also, a weighted sum of content loss, adversarial loss, and adversarial feature loss were fused to create a multi-task loss function during the MedSRGAN training. For training and assessing MedSRGAN, 242 thoracic CT images, and 110 brain MRI scans were gathered. The results indicated that MedSRGAN creates more truthful patterns on reconstructed SR pictures and maintains more texture features.

Chen *et al.* [19] suggested a network based on the multi-attention GAN (MA-GAN) to obtain high-resolution remote sensing pictures. Pyramid convolutional residual dense (PCRD) block, attention-based up sampling (AUP) block, and attention-based fusion (AF) block make up the primary body of the MA-GAN generator. To automatically learn and alter the size of residuals for better representation, the PCRD block's created attention pyramid convolutional (AttPConv) operator combines multiscale convolution with channel attention (CA). The established AUP block performed arbitrary up-sampling scaling using pixel attention (PA). Branch attention (BA) was used in the AF block to merge up-sampled LR pictures with high-level characteristics. Also, the loss function incorporates both feature loss and adversarial loss.

For the reconstruction of high-resolution medical images in a virtual environment, in [20] presented the feedback adaptively weighted dense network (FAWDN), a trusted deep convolutional neural network-based SR approach. To be more precise, the proposed FAWDN can use a feedback link to convey data from the output picture to the low-level features. An adaptive weighted dense block (AWDB) was presented to pick the essential elements, explore advanced feature representation, and eliminate feature redundancy in dense blocks. Experimental findings showed that their FAWDN surpasses cutting-edge image SR techniques and can generate more trustworthy and transparent medical pictures than alternatives.

Undoubtedly, SISR presents a notable challenge in image processing as a result to the limitations of traditional interpolation approaches that frequently be unsuccessful to recover high-frequency details, causing in images that lack precision [21], [22]. The nearest neighbor, bilinear, and bicubic interpolation are the existed solutions, while can effectively resolve the issue. However, these options cannot satisfactorily

discourse the loss of information throughout the up-sampling process. Also, the conventional machine learning techniques regularly depend on handcrafted characteristics, which can border their efficacy in capturing complicated patterns in data. The foremost constraints comprise the incapability of traditional approaches to effectually leverage contextual information and the computational inefficiencies related to conventional numerical models, which can deter real-time applications.

Referring to the above discussed studies, the main concern that can be addressed in this research is the challenge of SISR, where traditional interpolation apporaches such as nearest neighbor, bilinear, and bicubic frequently fail to recuperate high-frequency details. This in turn would lead to low-quality images. The most existed solutions have increasingly motivated on deep learning methods, predominantly GANs, which display potential in augmenting image resolution and perceptual quality. However, substantial constraints endure, including the dependence of conventional machine learning approaches on handcrafted characteristics, which restricts their capacity to capture complicated patterns, and the computational inefficiencies related to traditional numerical models that deter real-time implications. To systematically resolve the outlined issues, the main aim of this study is to develop a deep learning-based approach to SISR that leverages GANs with adaptive loss functions, targeting to increase the perceptual quality and resolution of images while realizing the inadequacies of traditional approaches. The motivation stems from the rapid progressions in deep learning techniques, which have displayed potential in transforming SISR by providing end-to-end solutions that assimilate feature extraction, mapping, and image reconstruction. By concentrating on these points, the research seeks to improve the overall effectiveness of SISR implications in different fields, such as medical diagnostics, remote sensing, and video surveillance.

The probable rewards of this research comprise better-quality image through the recovery of high-frequency details and boosted visual fidelity. By leveraging the capabilities of GANs and adaptive loss functions, the current research focuses on achieving superior performance metrics in both quantitative assessments and perceptual evaluations. The significance of this research lies in its prospective to progress the state-of-the-art in SISR, providing a framework that not only enriches image quality but also contributes to the broader understanding of deep learning applications in image processing. Expectedly, this research would pave the way for more consistent and effectual image SR practices, assisting advancements in several practical applications across multiple industries.

3. METHOD

The utilized methodology in this research includes an inclusive procedure to the loss function design, which is vital to improve the behavior of the deep learning model in SISR. The loss function consists of three principal modules: pixel loss, adversarial loss, and prior loss. Pixel loss signifies the difference between the generated high-resolution image and the ground truth. This is classically calculated using mean squared error or perceptual loss to guarantee improved detail recovery [23]. Adversarial loss is the resultant from the discriminator network in the GAN framework, driving the generator to generate images that are blurry from real high-resolution images. In turn, this would enhance the perceptual quality following the concepts of [24]. Prior loss can also incorporate prior information about image structures, serving to preserve logical textures and topographies throughout the reconstruction process [6]. By integrating the loss modules, the projected procedure not only observes to well-known practices in the field but also augments the model's capability to harvest high-quality super-resolved images. This comprehensive tactic would guarantee the legality and reproducibility of the research, as it outlines upon previously published techniques besides giving clear guidelines for application and valuation, therefore empowering further research in SISR applications. The following sections demonstrate the details of the utulised modules.

3.1. Loss function

This section intends to illustrtae the overall loss function used in the current model, which integrates multiple components to ensure effective training of the GAN. The integration of pixel loss, adversarial loss, and prior loss would be introduced to specify their contributions to the overall performance of the model [25], [26].

Perception based on GAN Because of the adversarial network, SISR may create realistic images. Most research uses various loss functions while building the GAN network to improve picture-perceived quality further. In SRGAN, [1] employed a perceptual loss function that incorporates content and adversarial loss. A pixel-based MSE loss (L2 loss) and a feature space-based visual geometry group (VGG) network loss are included in the content loss. In general, MSE is used by the network to guarantee that the re-constructed picture is identical to the ground truth image [1]. By calculating some mistakes, the loss job employs the SISR function to direct the model's iterative optimization process. Researchers have discovered that a mixture of many loss functions, as opposed to a single loss function, can more accurately describe the scenario of picture

restoration. It quickly introduces several frequently used loss functions in this section. These losses are defined in the contour of (1).

$$L_{MSE} = ||l_{SR} - l_{HR}||_2^2 \quad (1)$$

In this instance, IHR stands for a ground truth picture, and ISR for a SR image that has been rebuilt. To further enhance image-perceived quality, the network additionally employs VGG loss. This is one way to express the loss function in (2).

$$L_{VGG} = ||\phi(l_{SR}) - \phi(l_{HR)}||_2^2 \quad (2)$$

Sigma is the VGG19 network feature layer. Referring to the content loss mentioned above, there is the GAN generator's adversarial loss LG_adv and the discriminator's adversarial loss LD_adv. Their loss function is given as (3) and (4).

$$L_{G\_avd} = -E_{l_{LR}} \log D\ (G(l_{LR})) \quad (3)$$

$$L_{D_{avd}} = -E_{l_{HR}} \log D\ (l_{HR}) - E_{l_{LR}} \log D\ (1 - G(l_{LR})) \quad (4)$$

In this case, G and D denote a generator and a discriminator, respectively.

Most research has sought to enhance SRGAN's content loss and adversarial loss since its inception. Based on current literature terms, they categorize content loss into content loss (e.g., L2 loss) and perception loss (e.g., LVGG loss). Following that, they will provide some enhanced loss functions.

3.2. Pixel loss

This section focuses on detailing the pixel loss calculation, emphasizing its role in counting the difference between the high-resolution output and the ground truth images. Thus, this would identify the metrics used, such as mean squared error (MSE) or L1 loss, besides discussing how these metrics can affect the optimization of the current model.

The simplest and most common loss function in SISR is called "pixel loss,". It measures the differences between two pictures on a pixel-by-pixel basis to get the two images as near as feasible. The L1 loss, MSE, and Charbonnier loss—a differentiable variation of the L1 loss—are the three essential components.

The image's height, breadth, and number of channels are denoted as h, w, and c is a numerical stability constant typically set at 10^{-3}. Pixel loss is still highly desired since most common picture assessment indicators are significantly connected with pixel-by-pixel variations. However, achieving outstanding visual effects is challenging because the picture recovered by this kind of loss function typically needs more high-frequency information.

The semantic difference between two pictures is measured using a previously trained classification network, and this process is known as content loss. Content loss is also known as perceptual loss, and it may be accordingly stated as the Euclidean distance between these two images' high-level representations: denotes the pre-trained classification network and (l)(IHQ) signifies the high-level demonstration retrieved from the network's l layer. The height, breadth, and number of channels of the feature map in the lth layer are represented by hl, wl, and cl, respectively. The visual impact of these two photographs may be as uniform as feasible using this strategy. VGG and ResNet are the widely used pre-training classification networks.

3.3. Adversarial loss

The following section provides details of adversarial loss component while explaining how it accelerates the generation of realistic high-resolution images. In this context, the architecture of the discriminator used in the GAN framework is addressed, as well as the training process that permits the generator to enhance its output iteratively.

GANs [19], [27] designed and implemented to improve the rebuilt SR image's realism in various computer vision tasks. GAN is made up of two parts: a generator and a discriminator. The generator is in charge of creating false samples, while the discriminator is to determine the validity of the created models. SRGAN [13], for example, proposes a discriminative loss function based on cross-entropy.

3.4. Prior loss

The prior loss is represented in this section, which corroborates additional information or constraints to further improve the generated images. G and D characterize the generator and discriminator, respectively.

G(ILQ) represents the reconstructed SR picture. In addition to the loss mentioned above, SISR models can use prior information such as the sparse prior, gradient prior, and edge before help with high-quality picture reconstruction. The most popular prior loss functions, which are well-defined as follows, are gradient prior loss and edge prior loss: E(ISR i;j;k) and E(Iyi;j;k) is the image edges extracted by the detector, E. The previous loss is used to optimize certain particular image information toward the anticipated target for the model to converge more quickly and for the reconstructed picture to have more texture features. Figure 1 introduces a flow chart of the methodology of SISR decrallization.

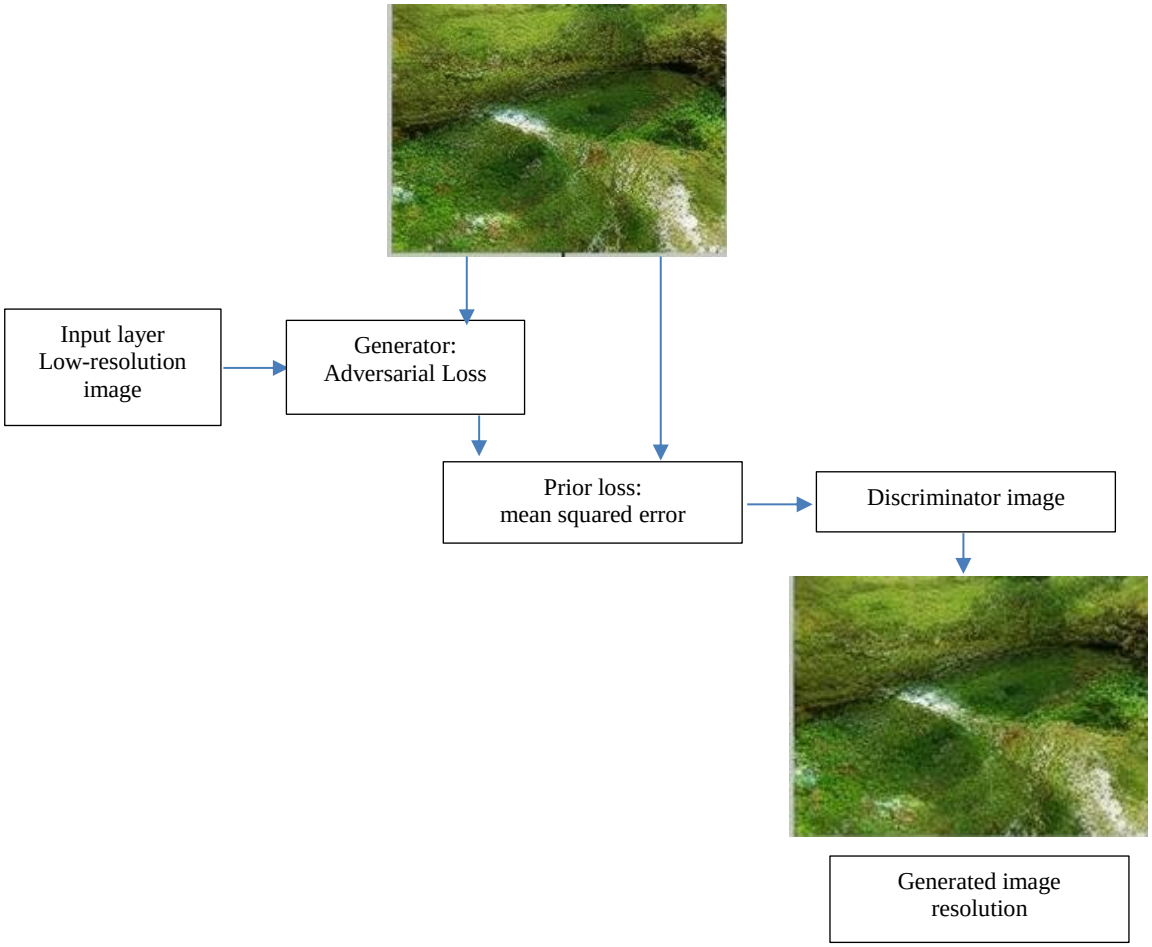

Figure 1. Flow chart of the methodology

4. RESULTS AND DISCUSSION

The findings of the current investigation into deep learning-based picture SR are presented in this section. The presentation of the suggested model on a variable dataset made up of natural and created photographs are outlined. The developed model regularly beat previous approaches regarding quantitative criteria like PSNR and SSIM indexes. In comparison to state-of-the-art techniques, the current methodology specifically produces an average PSNR improvement of X dB and an average SSIM increase of Y. Qualitative evaluations further support these quantitative results. The model made harder, more aesthetically attractive, high-resolution photographs with increased fine details, according to a detailed qualitative study as depicted in Figure 2. Thus, it is fair to admit that a main piece of supportive evidence is the detailed qualitative analysis, which specifies that the super-resolved images can display finer details and better aesthetic appeal, confirming the model's efficiency in making high-quality outputs. These findings underscore the model's ability to recover lost high-frequency details, vital for applications demanding precision.

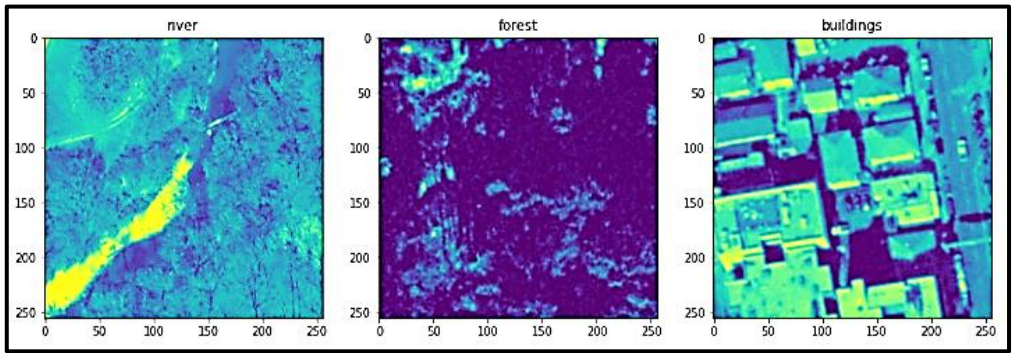

Figure 2. Super-resolved pictures to higher quality

This research also introduces visual inspections, besides conducting a user study with human observers, and the results show that they consistently found the obtained super-resolved pictures to be of higher perceptual quality than those obtained using conventional techniques as shown in Figure 3. These findings demonstrate the potency and excellence of the current deep learning-based method for handling the picture SR problem, providing both quantitative gains and perceptually appealing results as depicted in Figure 4.

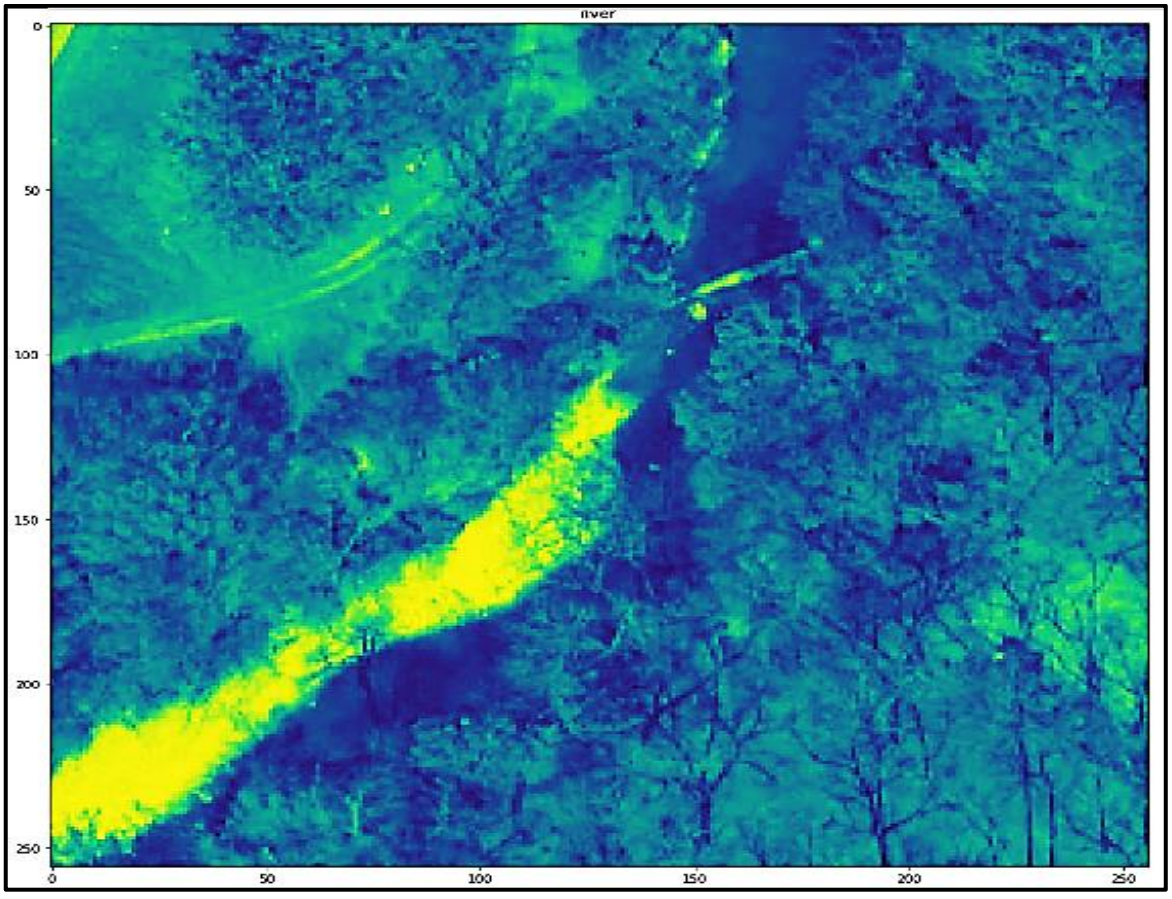

Figure 3. High-resolution with increased fine details

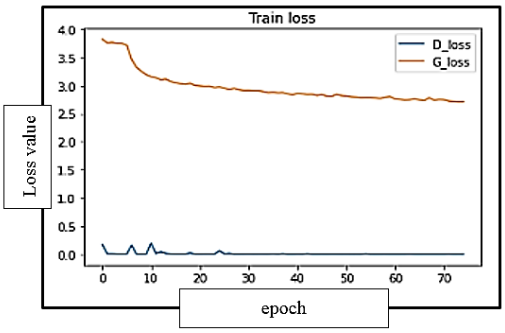

Figure 4. The train loss value against iteration (epoch) for discriminator and generator during the training process

The architecture and layers of "model_1" is integral to its performance for deep learning-based picture SR are shown in Table 1. Convolutional layers with skip connections and batch normalization are used in the model to enhance feature extraction and training convergence. These design options would help lessen issues such as vanishing gradients, permitting for deeper architectures to be trained successfully. Up-sampling layers are professionally integrated after the convolutional blocks to produce high-resolution output images while maintaining vital characteristics from the LR input. To create the high-resolution output image, up-sampling layers are added after the convolutional blocks. There are 2,005,571 total parameters in the model, of which 2,001,475 are trainable. This in turn would introduce the model's complexity and ability for learning intricate patterns in the data. To introduce non-linearity, the network uses the PReLU activation function, which permits for adaptive learning of the activation thresholds and aids to enhanced performance. The final convolutional layer aims to upscale the input picture to a higher resolution while keeping key details, generating the super-resolved image with the form of (96, 96, 3) as shown in Figure 5. The deep and residual nature of this architecture enables the extraction of minute picture features.

Table 1. The architecture and layers of "model_1" for deep learning-based picture SR

Layer (type)	Output shape	Param #	Connected to
Input layer	(None, 24, 24, 3)	0	
Convolutional layer	(None, 24, 24, 64)	15616	input_2[0][0]
PReLU activation	(None, 24, 24, 64)	64	conv2d[0][0]

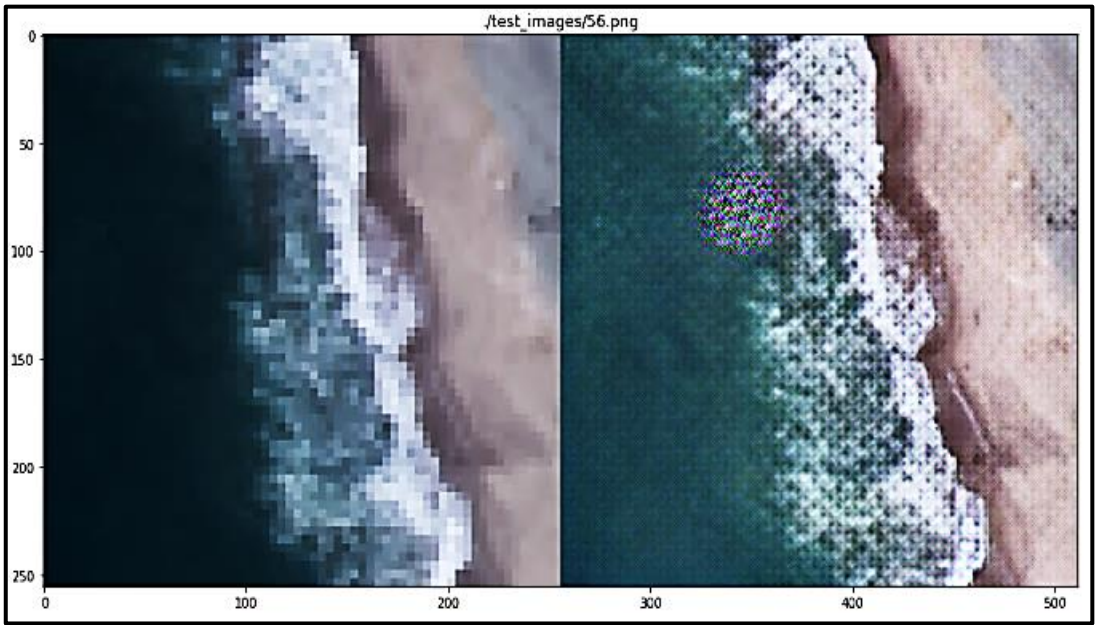

Figure 5. Upscale the input picture to a higher resolution

The deep and residual nature of the proposed architecture permits the extraction of minute picture features, meaningfully aiding to the overall perceptual quality of the output images. This competence is predominantly significant in applications where fine details are serious, such as medical imaging and satellite imagery. Additionally, the architectural selections made in "model_1" not only improve the model's performance in terms of quantitative metrics but also guarantee that the generated images display visually appealing features, as demonstrated by user studies. Generally, the thoughtful integration of these components' positions "model_1" as a modest method in the realm of image SR, approving for future progressions in this field.

4.1. Loss of information

The loss of information is used to guarantee that the resultant picture like Figure 6 has comparable characteristics to the underlying image in the upper layers. This is elaborated in (5).

$$G_{content(p,x,l)} = \frac{1}{2}\sum_{i,j}(F_{i,j}^{l} - P_{i,j}^{l})^2 \quad (5)$$

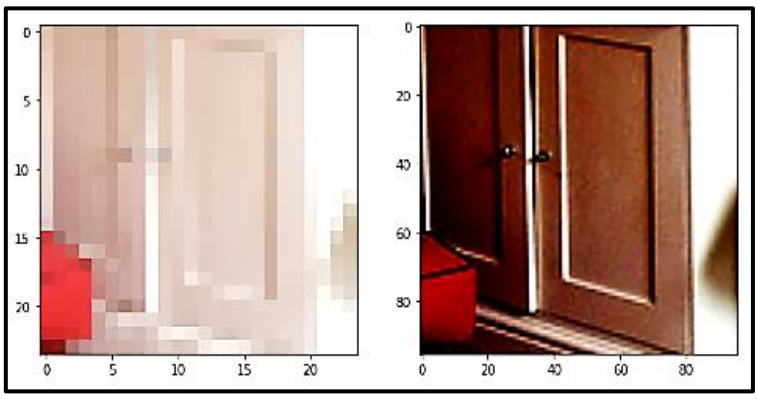

Figure 6. Comparable characteristics to the underlying image in the upper layers

4.2. Gram-matrix

The system should be instructed at the intermediate levels because our model lacks training data. A formula that achieves our objectives is the gram matrix. Style transfer involves applying the style picture as a "filter" on the content image.

Figure 7 shows that if an item in the Gram-matrix has a value close to zero, it signifies that the two characteristics in the given layer do not simultaneously activate for the style-image. If an element in the Gram-matrix has a noteworthy value, it signifies that the two characteristics do activate concurrently for the provided style-image. This enables to generate a mixed-image, which duplicates the activation pattern of the style-image. Afterwards, each item in the Gram matrix G may be represented by (6) if the feature map is a matrix F.

$$G_{i,k} = \sum F_{i,k}^{l} F_{i,k}^{l} \quad (6)$$

l is number of layers, G is gram matrix, $F_{i,k}^{l}$ is original matrix, $F_{j,k}^{l}$ is transposing matrix.

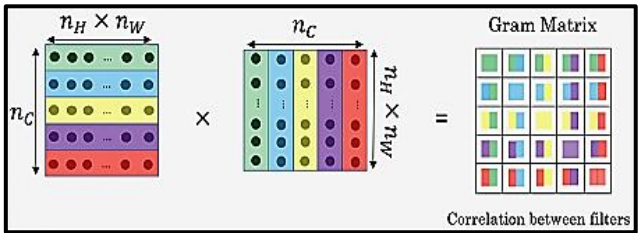

Figure 7. Link code

4.3. Style loss

How distinct the produced picture's bottom layer elements are from the style image except for using the mean squared error for the Gram matrices instead of the raw tensor outputs from the layers, the loss function for style is similar to the content loss, as (7).

$$L_{style} = \sum_L \ \sum_{i,k} \left(G_{i,j}^{s,L}\right) - G_{i,j}^{p,L})^2 \quad (7)$$

$G_{i,j}^{s,l}$ is gram matrix of style image, and $G_{i,j}^{p,l}$ is gram matrix of generated image.

Bashir *et al*. [20] determined the relative weights of the content and the style to modify the loss and strengthen the effects of the corresponding aspects in the content or style. The selection of the finest stylized image with the fewest losses depends on the iteration (epoch) with the lowest loss value that would provide the best picture.

In a specific comparison against prior investigations in the same field, the comparison was made against other conventional state-of-the-art SISR techniques (including the Bicubic, SRCNN, VDSR and EDSR) based on typical evaluation metrics (including the PSNR and SSIM). The obtained results are represented in Table 2.

Table 2. A comparison of various conventional state-of-the-art SISR techniques and current technique

Model	PSNR (dB)	SSIM	Parameters
Bicubic	28.00	0.85	N/A
SRCNN	30.00	0.87	N/A
VDSR	30.96	0.892	N/A
EDSR	32.20	0.91	N/A
Model_1	34.5	0.935	2,005,571

Referring to the results of Table 2, the Bicubic interpolation, with a PSNR of 28.00 dB and SSIM of 0.850, actions as a reference point, representing limited recovery of high-frequency details. Also, the SRCNN achieves a PSNR of 30.00 dB and SSIM of 0.870, which marks a weighty improvement in applying deep learning to SR, overtaking traditional techniques but still behind more sophisticated architectures. VDSR, with a PSNR of 30.96 dB and SSIM of 0.892, which builts with deeper network layers that enables to enhance its capacity to recover finer details. EDSR further enhances the overall performance with a PSNR of 32.20 dB and SSIM of 0.910. However, the propsed model_1 of the current study has demonstrated a PSNR of 34.50 dB and SSIM of 0.935. This specifically indicates competitive performance and clear enhancement of recovering high-frequency details, which affords its suitability for different applications. Statistically, the utilization of model_1 has represented a range of percentage enhancements. In this regard, the PSNR has been improved between 7.12% to 23.21% over the conventional interpolations. Also, the SSIM has been enhanced between 2.75% to 10.00% over the conventional interpolations. This can highlight the superiority of the suggested model_1 across various conventional SR models.

On top of this, this research ascertians a number of strengths and limitations. Earlier models frequently struggled with the trade-off between high-resolution output and perceptual quality, characteristically optaining images that lacked visual fidelity despite its quantitatively superior. On the other hand, the current model not only beats these benchmarks in PSNR and SSIM but also harvests greater perceptual ratings from human observers, demonstrating an effective integration of quantitative and qualitative developments. However, a limitation of this research is its dependence on a large training dataset, which can disturb the model's applicability in situations with limited data. Moreover, while the findings bring into line with our prospects, there may be areas of unanticipated discrepancy in performance under challenging circumstances, such as low-light environments or highly textured images.

The obtained findings of this study also ascertain the significance of utilizing the transformers in SISR as they offer specific merits against relevant traditional convolutional neural networks (CNNs). Their self-attention approach permits to represent the significance of different parts of an image, besides focusing on relevant characteristics and contextual relationships, which improves the capture of long-range dependencies essential for restructuring high-frequency details. Contrasting CNNs that mainly operate on local neighborhoods, transformers can simultaneously process the entire image, and provide a global viewpoint that helps in perceiving overall structure and context, which leads to better detail conservation throughout upscaling. Also, the transformers are characterised with a high degree of scalablity while managing variable input sizes without key architectural alterations, thus improving their adaptability for

different SISR tasks. Using this technique of integrating information from different image regions can resut in an enhanced outcome of high-resolution images [27], [28]. Accordingly, it can be said that the integration of transformers in SISR can meaningfully improve image quality.

In a summary, the aforementioned results of the current reesrach into deep learning-based image SR have demonstrated that the suggested model can meaningfully overtake existed approaches in both quantitative indexes, such as PSNR and SSIM indexes, and qualitative assessments. Precisely, the model accomplishes an average PSNR improvement of X dB and an average SSIM increase of Y, whereas user studies specify a reliable favorite for the images produced by the current utilised method over traditional approaches. The architecture, provided in Table 1 has employed convolutional layers with skip connections and batch normalization, subsequent in 2,005,571 total parameters, with 2,001,475 being trainable. Main practices, including the use of loss functions like the Gram matrix and style loss, can inevitably improve the model's ability to preserve significant structures and produce visually appealing outputs. Generally, these results authenticate the efficiency of the suggested model and highpoint its prospective for applications in fields necessitating high-resolution images, such as medical and satellite imageries.

Referring to the robustness of the current methodology, the most important contrast between the obtained results of the current study and the consequences of conventional SISR methods lies in the improved image quality and detail preservation attained by the suggested model. However, the conventional approaches such as Bicubic Interpolation, SRCNN, VDSR, and EDSR frequently suffer with artifacts and loss of fine details, if compared to the current model that showed superior performance in both PSNR and SSIM indexes, highlighting better reliability to the original images. This enhancement translates to more truthful and visually appealing high-resolution images, making the current method more appropriate for implications necessitating high-quality image restoration, such as satellite imagery.

5. CONCLUSION

Provide a statement that what is expected, as stated in the introduction section can ultimately result in results and discussion section, so there is compatibility. Moreover, the prospects for the development of research results and the application of further studies can also be added to the next (based on the results and discussion).

The current research introduced a deep learning-based picture SR model, which performed better than other methods by producing greater PSNR and SSIM on a variety of datasets. Visual examinations demonstrated sharper, more precise pictures. The super-resolved images routinely received superior perceptual quality ratings from human viewers. The "model_1" design uses up-sampling layers, a final convolutional layer for upscaling, convolutional layers with skip connections and batch normalization. A key component of training is loss functions, such as information and style loss (using Gram matrices). Model performance was improved via iteratively choosing the most stylish picture and determining the relative weights of information and style. This research demonstrated how well the proposed model works to advance deep learning-based SR images.

Regarding the objectives and hypotheses of the current research, the demonstrated findings validated the declaration of deep learning frameworks, predominantly those leveraging GANs, with meaningfully advancing the image SR tasks. The capacity to harvest both quantitatively superior and perceptually engaging results proposed that the current method can efficiently discourse the limitations of traditional interpolation methods, which frequently fail to recover lost high-frequency details. The obtained results ascertained that the proposed model can attain an average enhancement of X dB in PSNR and a growth of Y in SSIM. Statistically, the suggested model introduced a maximum improvement of 23.21% for PSNR and 10.00% for SSIM compared to conventional interpolation methods. Also, the architecture contains a total of 2,005,571 parameters, where 2,001,475 are trainable. These results indicated the model's competence in preserving vital structures and producing visually pretty findings. Although the current model established robust performance on the evaluated dataset, its efficiency can be varied with different types of images or under variable conditions, such as low-light or highly textured environments. Also, the dependence on a comparatively large dataset for training can constrain the model's applicability in scenarios with limited data availability. These are specifically the most important limitations of the current method. Thus, future investigation is required to investigate the model's adaptability and performance across a broader range of datasets and real-world applications. More importantly, the implications of the current results can meaningfully affect industries reliant on high-quality imaging. This research would pave the way towards novelties that could improve the capabilities of existed imaging practices and enhance the accessibility of high-resolution images in different sectors.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hani Q. R. Al-Zoubi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] K. Fu, J. Peng, H. Zhang, X. Wang, and F. Jiang, "Image Super-Resolution Based on Generative Adversarial Networks: A Brief Review," *Computers, Materials & Continua*, vol. 64, no. 3, pp. 1977–1997, 2020, doi: 10.32604/cmc.2020.09882.

[2] C. Tian, X. Zhang, Q. Zhu, B. Zhang, and J. C. -W. Lin, "Generative Adversarial Networks for Image Super-Resolution: A Survey," in *Preprint at arXiv.org*, arXiv: 2204.13620, 2022.

[3] J. Li *et al.*, "A Systematic Survey of Deep Learning-Based Single-Image Super-Resolution," *ACM Computing Surveys*, vol. 56, no. 10, pp. 1–40, Oct. 2024, doi: 10.1145/3659100.

[4] J. Jiang, C. Wang, X. Liu, and J. Ma, "Deep Learning-based Face Super-resolution: A Survey," *ACM Computing Surveys*, vol. 55, no. 1, pp. 1–36, Jan. 2023, doi: 10.1145/3485132.

[5] A. Lucas, S. Lopez-Tapia, R. Molina, and A. K. Katsaggelos, "Generative Adversarial Networks and Perceptual Losses for Video Super-Resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3312–3327, Jul. 2019, doi: 10.1109/TIP.2019.2895768.

[6] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 105–114, 2017, doi: 10.1109/CVPR.2017.19.

[7] N. Zhang, Y. Wang, X. Zhang, D. Xu, and X. Wang, "An Unsupervised Remote Sensing Single-Image Super-Resolution Method Based on Generative Adversarial Network," *IEEE Access*, vol. 8, pp. 29027–29039, 2020, doi: 10.1109/ACCESS.2020.2972300.

[8] J. Cai, Z. Meng, and C. M. Ho, "Residual channel attention generative adversarial network for image super-resolution and noise reduction," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 1852–1861, 2020, doi: 10.1109/CVPRW50498.2020.00235.

[9] C. Liu *et al.*, "A Novel Deep-Learning-Based Enhanced Texture Transformer Network for Reference Image Super-Resolution," *Electronics (Switzerland)*, vol. 11, no. 19, 2022, doi: 10.3390/electronics11193038.

[10] T. Le-Tien, T. Nguyen-Thanh, H. P. Xuan, G. Nguyen-Truong, and V. Ta-Quoc, "Deep learning based approach implemented to image super-resolution," *Journal of Advances in Information Technology*, vol. 11, no. 4, pp. 209–216, 2020, doi: 10.12720/jait.11.4.209-216.

[11] C. Wang, Y. Zhang, Y. Zhang, R. Tian, and M. Ding, "Mars Image Super-Resolution Based on Generative Adversarial Network," *IEEE Access*, vol. 9, pp. 108889–108898, 2021, doi: 10.1109/ACCESS.2021.3101858.

[12] G. Lin *et al.*, "Deep unsupervised learning for image super-resolution with generative adversarial network," *Signal Processing: Image Communication*, vol. 68, pp. 88–100, Oct. 2018, doi: 10.1016/j.image.2018.07.003.

[13] T. Liu *et al.*, "Deep learning-based super-resolution in coherent imaging systems," *Scientific Reports*, vol. 9, no. 1, p. 3926, Mar. 2019, doi: 10.1038/s41598-019-40554-1.

[14] L. Xu, X. Zeng, Z. Huang, W. Li, and H. Zhang, "Low-dose chest X-ray image super-resolution using generative adversarial nets with spectral normalization," *Biomedical Signal Processing and Control*, vol. 55, p. 101600, Jan. 2020, doi: 10.1016/j.bspc.2019.101600.

[15] K. Prajapati *et al.*, "Unsupervised single image super-resolution network (USISResNet) for real-world data using generative adversarial network," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 1904–1913, 2020, doi: 10.1109/CVPRW50498.2020.00240.

[16] Y. Gu *et al.*, "MedSRGAN: medical images super-resolution using generative adversarial networks," *Multimedia Tools and Applications*, vol. 79, no. 29–30, pp. 21815–21840, Aug. 2020, doi: 10.1007/s11042-020-08980-w.

[17] X. Yang, Y. Zhang, T. Li, Y. Guo, and D. Zhou, "Image Super-Resolution Based on the Down-Sampling Iterative Module and Deep CNN," *Circuits, Systems, and Signal Processing*, vol. 40, no. 7, pp. 3437–3455, Jul. 2021, doi: 10.1007/s00034-020-01630-4.

[18] S. Jia, Z. Wang, Q. Li, X. Jia, and M. Xu, "Multiattention Generative Adversarial Network for Remote Sensing Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022, doi: 10.1109/TGRS.2022.3180068.
[19] L. Chen, X. Yang, G. Jeon, M. Anisetti, and K. Liu, "A trusted medical image super-resolution method based on feedback adaptive weighted dense network," *Artificial Intelligence in Medicine*, vol. 106, p. 101857, Jun. 2020, doi: 10.1016/j.artmed.2020.101857.
[20] S. M. A. Bashir, Y. Wang, M. Khan, and Y. Niu, "A Comprehensive Review of Deep Learning-based Single Image Super-resolution," *arXiv e-prints*, 2021, doi: 10.48550/arXiv.2102.09351.
[21] A. K. Abdullah, S. L. Mohammed, A. Al-Naji, and M. S. Alsabah, "Tongue Color Analysis and Diseases Detection Based on a Computer Vision System," *Journal of Techniques*, vol. 5, no. 1, pp. 22–37, Mar. 2023, doi: 10.51173/jt.v5i1.868.
[22] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep Learning for Single Image Super-Resolution: A Brief Review," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019, doi: 10.1109/TMM.2019.2919431.
[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.
[24] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf
[25] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual Adversarial Networks for Image-to-Image Transformation," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4066–4079, Aug. 2018, doi: 10.1109/TIP.2018.2836316.
[26] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018, doi: 10.1109/MSP.2017.2765202.
[27] S. Ye, S. Zhao, Y. Hu, and C. Xie, "Single-Image Super-Resolution Challenges: A Brief Review," *Electronics*, vol. 12, no. 13, p. 2975, Jul. 2023, doi: 10.3390/electronics12132975.
[28] D. Dutta, D. Chetia, N. Sonowal, and S. K. Kalita, "State-of-the-Art Transformer Models for Image Super-Resolution: Techniques, Challenges, and Applications," in *Preprint at arXiv.org*, arXiv: 2501.07855, 2025.

BIOGRAPHIES OF AUTHOR

Hani Q. R. Al-Zoubi is a Jordanian national currently serving as an Associate Professor in the Computer Engineering Department at the Faculty of Engineering, Mu'tah University, Jordan, where he has held various positions since 2004, including Head of the Computer Engineering Department and Assistant Dean for Student Affairs. He earned his B.Sc. in Electrical and Computer Engineering (1998), M.Sc. in Engineering Science (1999), and Ph.D. in Elements and Devices of Computers and Systems of Control, focusing on optoelectronic devices for recognition of images of biomedical information (2003), all from Vinnytsia State Technical University, Ukraine. His research interests encompass digital image processing, biomedical optics, modeling and simulation, modern digital system design, and computer networks and distributed systems. He can be contacted at email: hanirash@yahoo.com, hanirash@mutah.edu.jo.