International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 2, June 2026, pp. 664~676
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i2.pp664-676

A new generation of artificial intelligence contributing to improving the image quality

Salwa A. Alagha, Hadeel N. Abdullah, Suad Khairi Mohammed
Department of Electronic Engineering, College of Electrical Engineering, University of Technology, Baghdad, Iraq

Article Info

Article history:

Received Jul 16, 2025
Revised Apr 23, 2026
Accepted May 22, 2026

Keywords:

Deep learning
Generative adversarial networks
Image improvement
SRGAN
Wavelet transformation

ABSTRACT

High-resolution (HR) images provide inclusive and critical information, which is substantial for many implementations. Production operation for high-quality images from low-quality images can be costly and time-consuming. The main advancement in this domain is produced by enhanced super-resolution generative adversarial network (ESRGAN); the ESRGAN and different deep learning (DL) models exhibit prominent advances in image super-quality. This research proposes introducing the discrete wavelet transform (DWT) as a multi-scale analysis stage that feeds into the network, whereby the frequencies are analyzed before being fed into the generative adversarial networks (GAN). The goal is to enhance the ability to recover edges and fine details, especially in low-resolution images. The performance of this proposed model is implemented, evaluated, and comparatively assessed. Key performance parameters, such as peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM), are calculated, which compare the proposed model with other image-improving models (Bicubic, SRResNet, and ESRGAN). The experimental results indicate that the proposed method ESRGAN new yields a good result in image improvement, with a PSNR of (26.22, 26.00, 25.51, and 23.89) and an SSIM of (0.6638, 0.6255, 0.5882, and 0.6286) for four datasets, respectively.

Corresponding Author:

Hadeel N. Abdullah
Department of Electronic Engineering, College of Electrical Engineering, University of Technology
Al Wahda-Neighborhood, Baghdad 19006, Iraq
Email: hadeel.n.abdullah@uotechnology.edu.iq

1. INTRODUCTION

In the era of artificial intelligence and the auric of big data, the demand for high-quality images and videos has been obvious in multiple domains, including medical imaging, satellite imagery, and military spy technology [1]. Multi-layer artificial neural networks designed to process two-dimensional input data for image classification are called convolutional neural networks (CNN). They perform well on various supervised learning tasks in terms of classification [1]. Conventional methods of enhancing videos are unsuccessful in handling complicated visual data. The typical techniques needed to improve the resolution of images and videos include enhancing sensors or developing optical systems. These systems are very expensive and represent a challenge, so the solution to this problem is going towards computational techniques for super-resolution (SR), deep learning (DL), and neural network methods, which have presented great progress [2].

In recent years, the field of image restoration has experienced remarkable progress driven by two primary streams: diffusion-based generative models and improved generative adversarial networks (GAN)-based architectures, complemented by advances in perceptual-loss design and learned image quality assessment (IQA) metrics. Diffusion models, originally proposed for high-quality image synthesis, have been

successfully adapted to low-level restoration tasks such as denoising, deblurring, inpainting, and SR. The key idea is to learn reverse of a gradual noising process, allowing iterative refinement of image. Early works such as SR3 demonstrated that denoising diffusion probabilistic models (DDPMs) could generate photo-realistic high-resolution (HR) images with superior texture realism compared to deterministic CNNs [3]. Subsequent research has introduced residual and predict-refine diffusion methods [4], latent diffusion to reduce sampling cost [5], and classifier-free conditional guidance for balancing fidelity and perceptual quality [6]. Recent surveys (2023–2024) highlight these models as state-of-the-art in perceptual image restoration while also stressing limitations such as high sampling cost, color inconsistency, and limited controllability [7], [8]. Moreover, foundational models such as AutoDIR (ECCV 2024) extend diffusion to "all-in-one" restoration pipelines, capable of handling multiple unknown degradations within a single framework [9].

Parallel to this, GAN-based image restoration methods remain highly relevant, especially in scenarios requiring real-time inference or deployment on resource-limited hardware. Enhanced GAN architectures, such as enhanced super-resolution generative adversarial network (ESRGAN), introduced residual-in-residual dense blocks (RRDBs), relativistic discriminators, and multi-scale attention modules, all of which improved perceptual realism while reducing artifacts [10]. More recent works integrate frequency-domain priors (e.g., wavelet transforms) into the GAN pipeline, providing better edge and texture preservation [11]. Surveys between 2020–2024 emphasize that GANs continue to achieve competitive results with much lower inference latency compared to diffusion models, though they suffer from training instability and occasional hallucination artifacts [12]. Emerging hybrid designs combine GANs with diffusion models—using GANs for coarse, fast prediction and diffusion for stochastic refinement—showing promising trade-offs between speed and perceptual quality [13].

Another significant trend is the optimization of perceptual loss functions. Traditional visual geometry group (VGG)-based perceptual loss [14] has been widely adopted, but recent works propose dual- or multi-term perceptual losses that explicitly balance structural fidelity and textural realism [15]. Furthermore, learned perceptual metrics such as learned perceptual image patch similarity (LPIPS) [16] and deep no-reference IQA models [17] are increasingly integrated directly into training objectives, aligning optimization with human perceptual judgments. ESRGAN has presented great progress in image SR. The basic concept of ESRGAN started from an earlier paradigm of super-resolution generative adversarial network (SRGAN). ESRGAN has introduced various innovations that have led to superior image quality and addressed popular issues, such as unrealistic textures, by enhancing network architecture, loss functions, and training strategies [10].

This paper presents SRGAN. Then, dive into the details of the ESRGN algorithm, exploring its architecture, training methodology, and the specific innovations that contribute to its superior performance. The main contributions of this work are as follows:

i) Combining ESRGAN with wavelet transformation. Most ESRGAN research focuses on improving the internal structure (RRDB, loss functions, and training strategy). This research proposes introducing the discrete wavelet transform (DWT) as a multi-scale analysis stage that feeds into the network, whereby the frequencies (coarse + fine details) are analyzed before being fed into the GAN. The goal is to enhance the ability to recover edges and fine details, especially in low-resolution (LR) images.
ii) Previous research (ESRGAN, ESRGAN+, and SRResNet) relied solely on a deep CNN. In this research, a new version, called ESRGANnew, is designed, which is an improved version of ESRGAN by adding multi-scale wavelet components while preserving the GAN structure.
iii) Simplifying the structure by using a multilayer perceptron (MLP) as a generator and discriminator to reduce computational complexity, while leveraging the advantages of Wavelets in feature extraction.

The rest of the paper is organized as follows. Section 2 details the ESRGAN architecture in recent literature. Section 3 describes the wavelet multi-scale transform, presents the ESRGAN network, and outlines the proposed model. Section 4 shows the results obtained from the experiments of this work and discusses and compares the results. Section 5 concludes the paper. Finally, section 6 presents the future scope of this work.

2. LITERATURE REVIEW

A very deep convolutional network modelled after VGG-net for ImageNet classification is used in the highly accurate single-image SR approach presented by Kim *et al.* [18]. Accuracy is greatly increased by increasing network depth, and the final model employs 20 weight layers. Residual learning and extraordinarily high learning rates are used to optimize a demanding network swiftly. Training stability is guaranteed by gradient clipping, and convergence speed is optimized. Frequently cascading small filters in a deep network design allows for efficient access to contextual information over large image regions. However, convergence speed becomes a crucial concern for intensive networks during training. Suggested a straightforward, efficient training process that leverages incredibly high learning rates and solely learns

residuals. On benchmarked photos, the approach performs significantly better than the current method. Other picture restoration issues, such as de-noising and compression artefact removal, can easily be solved with this method. This suggested approach outperforms current approaches in terms of accuracy, and the outcomes' visual enhancements are readily apparent. An enhanced deep super-resolution network (EDSR) was created by Lim *et al.* [19] and outperformed the currently available SR techniques at the time. This model's notable performance boost results from optimization by eliminating superfluous modules in traditional residual networks. Performance is additionally enhanced by increasing the model size while preserving the training technique. A novel multi-scale deep super-resolution system (MDSR) and training technique might be suggested to reconstruct HR images of various upscaling variables in a single model. On benchmark datasets, the suggested approaches outperform the cutting-edge techniques.

Song *et al.* [20] use adder neural networks (AdderNets) to investigate the single-picture SR problem. AdderNets circumvent the high energy consumption of traditional multiplications by using additions to compute the output features, as opposed to CNN, examining how an adder operation relates to identity mapping and adding shortcuts to improve SR model performance with adder networks. After that, a learnable power activation will be created to fine-tune specifics and modify the feature distribution. Experiments on a number of benchmark models and datasets show that SR model images utilizing AdderNets can perform and look as good as their CNN baselines while using roughly 2.5 times less energy.

SRGAN, a GAN for imagining SR, was developed by Ledig *et al.* [21]. The framework can infer photo-realistic natural images. This was accomplished by proposing a perceptual loss function that included both a content cost and an adversarial loss. They use a discriminator network trained to distinguish between the original photo-realistic images and the super-resolved images, and adversarial loss drives the solution to the natural image manifold. Furthermore, perceptual similarity—rather than pixel-space similarity—generates content loss. By using publicly available benchmarks, a deep residual network can restore photo-realistic textures from substantially down-sampled photos.

In place of the original ESRGAN, Rakotonirina and Rasoanaivo [22] created a network architecture using a new basic block. Additionally, noise inputs were added to the generator network to benefit from stochastic variation. The final photos showcase the ESRGAN, a perceptual-based method for SR of a single image that may create more realistic textures and photo-realistic images. The visual quality of these generated photos should be enhanced. This way, the model is expanded to improve the image quality further. A more objective way to profit from perceptual loss is suggested by Rad *et al.* [23]. They developed a targeted objective function for optimizing a deep network-based decoder that uses the related terms to punish images at various semantic levels. Using segmentation labels to construct object, background, and boundary (OBB) labels, the suggested method estimates an acceptable perceptual loss for boundaries while considering texture similarity for backdrops. Sharper edges and more realistic textures were the outcomes of the suggested method. According to the results of in-depth user research as well as qualitative results on common benchmarks, it fared better than other cutting-edge algorithms. Johnson *et al.* [14] coupled the advantages of optimization-based picture production techniques with feed-forward image transformation tasks by using perceptual loss functions to train feed-forward transformation networks. This technique was used for style transfer, where comparable performance and a much faster speed were achieved compared to previous approaches, in which training with a perceptual loss helps the model to rebuild small features and edges better. Try single-image SR, which produces aesthetically pleasant results by substituting a perceptual loss for a per-pixel loss.

Hssayeni and Ghoraani [24] looked at a promising yet underutilized framework called conditional generative adversarial networks (cGANs) for enhancing deep regression models for time-series data with an asymmetric and fragmented distribution. Initially, they looked into the feasibility of employing a vanilla cGAN as a data restoration tool to enhance the generality of the developed models compared to previously unreported data in such datasets. They then suggested a modified cGAN architecture that increased their regression models' extrapolation and generalizability. The developed cGAN architecture considerably enhanced extrapolation and generalizability for predicting regression scores.

To eliminate explicit motion compensation, Jo *et al.* [25] offer a novel end-to-end deep neural network that computes a residual image and dynamic up-sampling filters based on the local spatiotemporal neighborhood of each pixel. This method uses dynamic up-sampling filters to immediately reconstruct an HR image from the input image, and the computed residual is used to add the fine details. Networks may produce HR films that are significantly clearer and more consistent over time with the use of a novel data augmentation methodology. Through numerous studies, they have also offered network analysis to demonstrate how the network implicitly handles motions. An efficient SR network with excellent scalability was proposed by Zhang *et al.* [26] to handle numerous degradations using a single model.

To improve the consistency of the reconstructed films, Liu *et al.* [27] suggest an end-to-end temporal consistency learning network (TCNet) for video super resolution (VSR). Self-alignment was

learned from inter-frames using a spatiotemporal stability module. ESRGAN introduces RRDB, which enhances the learning ability and stabilization of the network. In ESRGAN, the original content loss is replaced by perceptual loss depending on features extracted by high-level layers of a pre-trained VGG network. ESRGN has a preference for quality over the usual pixel-wise precision. The relative average generative adversarial network (RaGAN) is used as the characteristic that makes the structure more factual by focusing on the proportional originality of generating images compared with real images, which can be indicated as ESRGN.

3. MATERIALS AND METHOD

The implemented model was tested only on natural images (DIV2K, Set5, Set14, BSD100, and Urban100). This section describes the DIV2K dataset: DIVerse 2K resolution high-quality images as used for the challenges @ NTIRE (CVPR 2017 and CVPR 2018) and @ PIRM (ECCV 2018). The DIV2K dataset has the following structure: 1000 2K resolution images divided into 800 images for training, 100 images for validation, and 100 images for testing. Hardware resources used include: DELL (64-bit), CPU: Intel (R) Core (TM) i3, RAM: 4 GB, Storage: 500 GB SSD. Software Frameworks include DL libraries. MATLAB (R2013b).

3.1. Multiscale wavelet

The two-dimensional DWT applies low-pass (L) and high-pass (H) filters along rows and columns, which break the image into frequency sub-bands, as shown in Figure 1. Instead of working in pixel space, the image geometry is represented in a multi-scale, orientation-sensitive space. This decomposition is applied recursively to obtain multiple scales, leading to a hierarchical representation of the video frame [28]. It also relies on the wavelet transform, which is a mathematical tool used to decompose a signal into different frequency components. Unlike the Fourier transform, which analyzes the signal in terms of sine and cosine functions, the wavelet transform uses wavelets that are localized in both time and frequency. This process involves transforming each frame into a set of sub-bands representing various levels of detail. The wavelet transform uses filter banks made up of biorthogonal high-pass and low-pass filters. This process is applied to an image at high resolution. The approximation and detail coefficients at each decomposition level are given in (1) to (4) [11].

$$C_j = \left(L_{ow} + C_j - 1L_{ow}\right) s\downarrow \quad (1)$$

$$D_{vj} = (H_{igh} + C_j - 1L_{ow})\, s\downarrow \quad (2)$$

$$D_{hj} = \left(L_{ow} + C_j - 1H_{igh}\right) s\downarrow \quad (3)$$

$$D_{dj} = \left(H_{igh} + C_j - 1H_{igh}\right) s\downarrow \quad (4)$$

Where j represents the level of detail and $s\downarrow$ indicates a reduction to 1/s of the original resolution. Cj represents approximation coefficients, which capture the coarse, large-scale structures. Dhj, Dvj, and Ddj represent details coefficients (horizontal (Dhj), vertical (Dvj), and diagonal (Ddj)), which capture the finer, high-frequency.

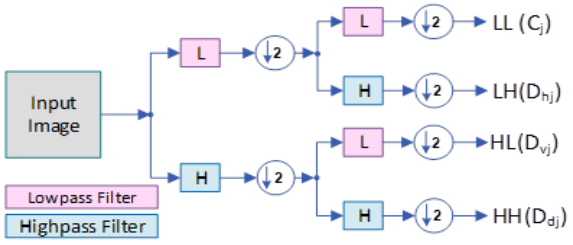


Figure 1. Structure of 2D-DWT

When wavelet transforms are introduced into ESRGAN, they bring an extra layer of geometry to how image information is represented and reconstructed. This makes edges, textures, and fine details more separable and easier for ESRGAN to reconstruct. Forces ESRGAN to match frequency bands directly, improving sharpness and texture fidelity. This is like aligning the frequency geometry of the image rather

than just pixels. After processing the individual scales, the enhanced image is reconstructed by combining the modified components using the inverse wavelet transform. This step ensures that the improvements made at each scale are integrated into the final video frame. The inverse wavelet transform recombines sub-bands back into the spatial image. After ESRGAN enhances details in wavelet space, the IDWT rebuilds the HR image. Ensures reconstructed geometry is consistent—fine textures and global structure.

3.2. ESRGAN architecture

The SRGAN architecture serves as basis for ESRGAN, adding significant enhancements. ESRGAN presents the RRDB as an alternative to the conventional residual block. Inspired by densely connected convolutional networks (DenseNet)—a feed-forward CNN architecture—every layer in this block is directly connected to every other layer. As shown in Figure 2, the architecture consists of these three main parts [2].

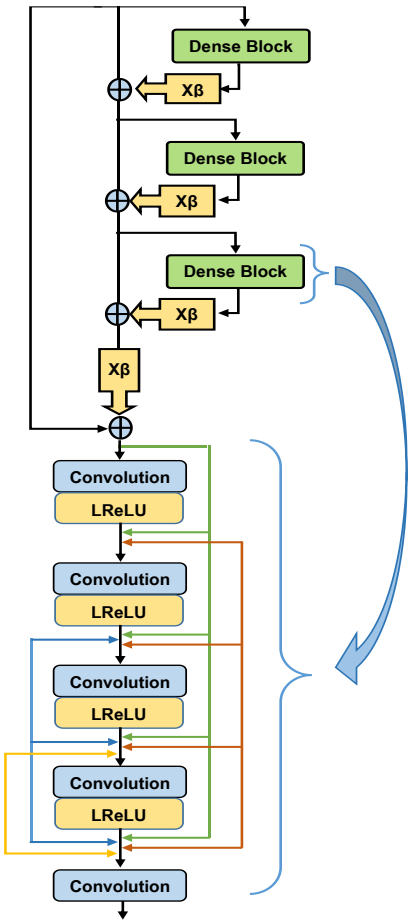


Figure 2. The basic architecture of SRGAN

3.2.1. Generator

The updated residual block is the main difference between the essentially same generators in ESRGAN and SRGAN. To do this, one just needs to scale the output of the present SRGAN model by swapping the residual block with the new RRDB block. Two major changes are made to the generator structure. First, remove all batch normalization (BN) layers. Second, replace the original basic block with the new RRDB. This block combines features from multi-level residual networks and dense connections. Removing the BN layers helps improve performance and reduces the computational load. The generator network in ESRGN uses RRDB to improve learning stability and capacity. These blocks enable the network to retain more contextual information, effectively capturing high-frequency details and complex textures. As shown in Figure 3, the key features include:

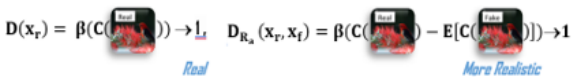
– Residual learning: it helps with training deeper networks by dealing with the problem of the gradient disappearing.
– Dense connections: enhance feature reuse and strengthen gradient flow [10], [12].

3.2.2. Discriminator

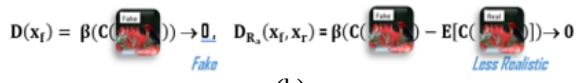
The relativistic discriminative agent assigns a truth value of whether an image is real or an ESRGAN. RaGAN assesses the relative authenticity of the generated image compared to real images, in contrast to traditional discriminators, which only concentrate on differentiating between real and fake images. This method lowers common artefacts and helps create more realistic textures. Where x is the real image, the standard discriminator is replaced by the relativistic average discriminator (RAD), denoted as D_{R_a}, SRGAN represents the standard discriminator as $\beta D(x) = (C(x))$, β is Beta function, and $C(x)$ is non-transformed discriminator output; RAD is constructed as (5).

$$D_{R_a}(x_r, x_f) = \beta(C(x_r - E_{x_f}[C(x_f)])) \quad (5)$$

Where x_r is a real image that is more realistic than x_f is a fake image, $Ex_f[.]$. The operation involves averaging all dummy data in the mini-batch, as illustrated in Figure 3. Figure 3(a) shows the standard GAN discriminator, while Figure 3(b) shows the relativistic GAN discriminator.



(a)



(b)

Figure 3. Discriminator architectures of (a) standard GAN and (b) relativistic GAN

3.2.3. Loss function

The proposed model uses hybrid loss combining structural, perceptual, and adversarial components:

i) Adversarial loss (GAN-based):
   – Ensures generated images are perceptually realistic.
   – Implemented with RaGAN discriminator, which compares realism between fake and real images.

ii) Perceptual loss (materials in context database or MINC/VGG-based):
   – Extracted from a VGG network fine-tuned for material recognition.
   – Prioritizes recovery of textures and fine details over pixel accuracy.

iii) Content loss (pixel or structural): ensures structural consistency with the ground truth.

The discriminator loss, as defined in (6) [29], [30].

$$L_D^{Ra} = -E_{x_r}\left[log\left(D_{R_a}(x_r, x_f)\right)\right] - E_{x_f}\left[log\left(1 - D_{R_a}(x_f, x_r)\right)\right] \quad (6)$$

The adversarial loss of the generator for a symmetrical, as defined in (7).

$$L_G^{Ra} = -E_{x_r}\left[log\left(1 - D_{R_a}(x_r, x_f)\right)\right] - E_{x_f}\left[log\left(D_{R_a}(x_f, x_r)\right)\right] \quad (7)$$

The discriminator and adversarial loss for ESRGAN, as defined in (8).

$$L_G = L_{percep} + \lambda L_G^{R_a} + \eta_{L_1} \quad (8)$$

3.3. Proposed model

Figure 4 displays the block diagram of the proposed ESRGANnew with wavelet-based multi-scale enhancement. The model begins with preprocessing through DWT, followed by feature extraction and RRDBs with leaky rectified linear unit (ReLU) activation. The generator reconstructs the HR output via inverse discrete wavelet transform (IDWT), while the RaGAN discriminator enforces perceptual realism. Loss optimization integrates adversarial, perceptual (VGG/MINC), and content losses. The process follows a sequence of steps to enhance images:

i) Preprocessing:
   a) Input: read colored images of any type into the program, crop a portion of it, and adjust it to a suitable (N × M) size.
   b) Normalizing: pixel intensity values are normalized to a suitable range before training.

c) Cropping: randomly selected patches from LR inputs were used to generate HR images.
d) Downsampling: LR images were produced using bicubic interpolation, the standard degradation model in single-image super-resolution (SISR).
e) Wavelet decomposition (DWT): each input was decomposed into approximation (low-frequency) and detail (high-frequency) sub-bands. This provided a multi-scale representation, enhancing the network's ability to reconstruct fine edges and textures.

ii) Network architecture: the proposed model builds upon the ESRGAN and integrates wavelet-based multi-scale analysis. The model consists of two main networks:
a) Generator network (G):
– Feature extraction layers: initial convolution layers extract spatial features.
– Each block integrates residual learning and dense (RRDBs) connections, allowing stable training of deeper networks without BN.
– Employed leaky ReLU activation after convolutional layers.
– Upsampling layers to recover the HR output.
– Apply IDWT to reconstruct the enhanced image from processed wavelet sub-bands.
b) Discriminator network (D):
– Based on a RaGAN framework, where the discriminator compares the relative realism of generated versus real images.
– Uses multiple convolutional layers with leaky ReLU activation to capture hierarchical texture features.

iii) Training parameters: the model was implemented in MATLAB with the following settings:
a) Optimizer: gradient descent with momentum parameter for training set to 0.9, weight is regularized to 0.0001, learning rate is set to 0.001.
b) Batch size: 64
c) Hidden units: 256
d) Epochs: ~20 (training converges after ~20 epochs).
e) Loss functions:
– Adversarial loss (RaGAN): enforces perceptual realism.
– Perceptual loss: derived from a VGG network fine-tuned for material recognition (MINC), emphasizing textures.
– Content loss: maintains structural similarity between input and ground truth.

iv) Evaluation metric: the proposed system evaluates the results of improving images with objective measurements. This approach is assessed based on comparing the image before and after the improvement operation, depending on the proposed method. The following measurements are employed to assess the outcomes:
a) Peak signal-to-noise ratio (PSNR): The quality of an improved image is evaluated with the PSNR value. For an M × N image, apply the PSNR for the original image ($O_{i,j}$), and improved or enhanced image ($E_{i,j}$), by using (9).

$$PSNR = 10\ log_{10}\ \frac{M \times N}{\sum_{i=1}^{M}\sum_{j=1}^{N}(O_{i,j} - E_{i,j})^2} \quad (9)$$

b) Structural similarity index metric (SSIM): SSIM is its core, a statistical measurement. It played a significant role in IQA in many application disciplines. It results from a product of three local dissimilarity factors (luminance, variance, and correlation), as defined in (10).

$$SSIM(x,y) = \big(l(x,y)\big)^{\alpha}.(c(x,y))^{\beta}.(r(x,y))^{\gamma} \quad (10)$$

Where $l(x,y)$ is related to luminance, $c(x,y)$ is contrast differences, and $r(x,y)$ is represented by the structure variations between x and y. α, β, and γ are parameters that define the significance of each component.

The network is considered converged when:
– Training stabilizes at ~20 epochs with the given setup.
– PSNR and SSIM values plateau on validation sets.
– Adversarial loss and perceptual loss balance (generator produce sharp, realistic details without artifacts).

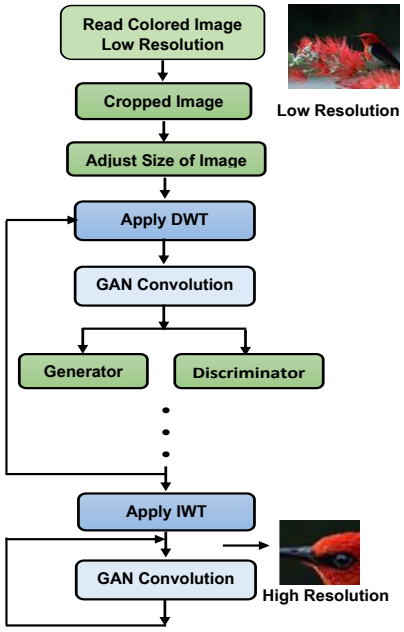


Figure 4. Block diagram for the proposed model

Figure 5 represents the proposed work design. The DWT Layer represents the transformation layer to the DWT following the use of the wavelet method to enhance the process of improving and developing the image. Wavelet-based multi-scale enhances the representation of an image's edge depending on the wavelet classification features. GAN layer 1 and GAN layer 2 represent the generator and discriminator processes operations; two MLP (feed-forward connection) networks are used in these layers, one for generation and the other for discrimination. The GAN processes at these levels are done for multiple bands of wavelet images. GAN layer 3 uses the GAN process to improve the inverse images of wavelets generated from layers 1 and 2. IDWT layer represents the inverse wavelet transformation operation for the final improved image.

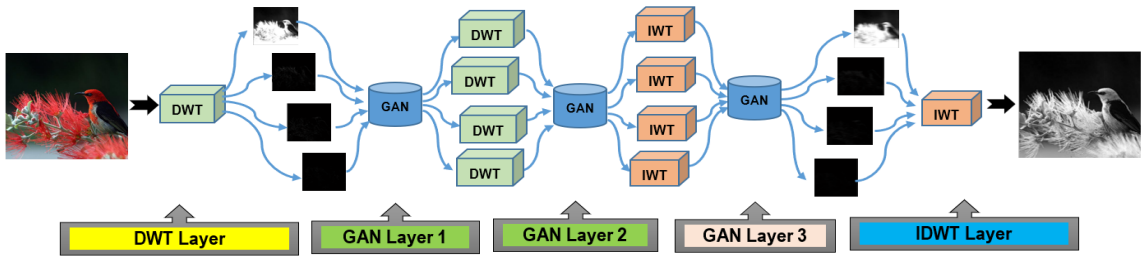


Figure 5. The proposed multi-frame wavelet with multi-scale GAN

4. RESULTS AND DISCUSSION

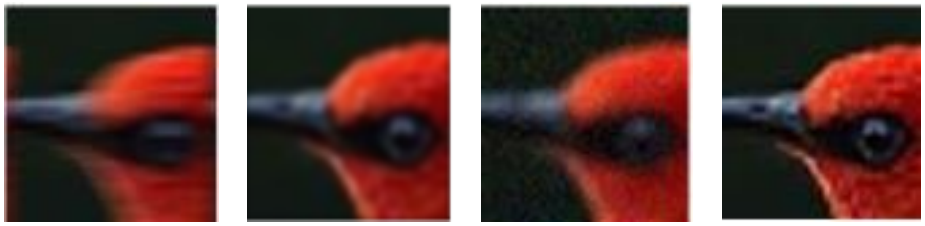
The proposed work was implemented using the MATLAB programming language. The DIV2K training data includes 40 images for training and 10 for testing; models were trained using the red, green, and blue channels. Benchmark Datasets include: (Set5, Set14, BSD100, and Urban100 datasets), as shown in Figure 6. In this proposed model, two images will be presented as an example of implementation, evaluation, and comparison of the results, as shown in the images later.

Figure 7 represents the first colored image from the DIV2K dataset, where Figure 7(a) shows the original image and Figure 7(b) shows the selected part of the image, which will be improved using the proposed method. Figure 8 represents the cropped portion of an image with different improved methods. These methods are: Figure 8(a) is Bicubic, Figure 8(b) is SRResNet, Figure 8(c) is ESRGAN, and Figure 8(d) is the proposed method ESRGANnew.

Figure 6. Examples of the DIV2K dataset

(a) (b)

Figure 7. First colored image from DIV2K: (a) original image and (b) original image with selected part



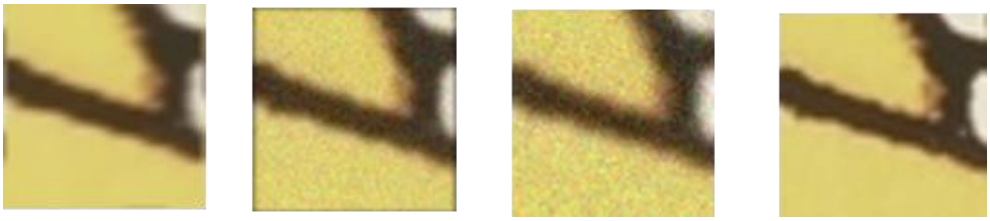
(a) (b) (c) (d)

Figure 8. First results for four improved image methods: (a) bicubic, (b) SRResNet, (c) ESRGAN, and (d) proposed ESRGANnew

Figure 9 represents the second colored image from the DIV2K dataset, where Figure 9(a) shows the original image and Figure 9(b) shows the selected part of the image, which will be improved using the proposed method. Figure 10 represents the cropped portion of the image with different improved methods. These methods are: Figure 10(a) is Bicubic, Figure 10(b) is SRResNet, Figure 10(c) is ESRGAN, and Figure 10(d) is the proposed method, ESRGANnew.

(a) (b)

Figure 9. Second colored image from DIV2K: (a) original image and (b) original image with selected part



(a) (b) (c) (d)

Figure 10. Second results for four improved image methods: (a) Bicubic, (b) SRResNet, (c) ESRGAN, and (d) proposed ESRGANnew

This study, which proposes an improved image SR model called ESRGANnew, is highly relevant to real-world deployment and user impact, primarily by providing a cost-effective, high-performance computational solution for producing high-quality images from low-quality sources. The impact of deploying the ESRGANnew model centers on superior visual quality and operational efficiency. The qualitative comparison of the results demonstrates that the proposed method of improving SRGANnew is better than other methods. Figures 8(a) to 8(c) are more blurred than Figure 8(d), which has more obvious image distinct boundaries. Figures 10(a) to 10(c) are also more blurred than Figure 10(d), which has more obviously distinct image boundaries. The quantitative comparison of the results demonstrates that the proposed improved method, SRGANnew, indicates the best performance. Table 1 shows the PNSR and SSIM values for four methods (Bicubic, dual regression networks (DRN), ESRGAN, and ESRGANnew). PSNR for ESRGANnew indicates the best performance (the highest) among others (29.52, 26.51, 26.45, and 29.21), the same as SSIM (0.8751, 0.7711, 0.7281, and 0.9523).

Previous studies and this SRGANnew proposed work have been compared in terms of the new SR generative network. Table 1 shows that the PSNR and SSIM values vary depending on the four datasets and four different SR methods (Bicubic, DRN, ESRGAN, and ESRGANnew). Table 2 compares ESRGANnew with baseline models, including bicubic interpolation, super-resolution convolutional neural network (SRCNN), ESRGAN, and transformer-based methods. The main limitations of this work are that it requires a large amount of data to train effectively, and the training convergence is comparatively slow due to the extensive database (time-consuming). Although noise and blur modeling were not directly implemented in this work, the framework can be extended to handle such degradations in future experiments.

Table 1. Comparison between previous ESRGAN methods and the proposed work

Dataset	Performance	SR methods			
		Bicubic [15]	DRN [30]	ESRGAN [21], [23]	Proposed ESRGANnew
Set5	PSNR	26.69	27.43	29.40	29.52
	SSIM	0.7736	0.792	0.8472	0.8751
Set14	PSNR	26.08	25.28	26.02	26.51
	SSIM	0.7466	0.653	0.7397	0.7711
BSDS100	PSNR	26.07	25.00	25.16	26.45
	SSIM	0.7177	0.606	0.6688	0.7281
Urban100	PSNR	24.73	22.99	28.413	29.21
	SSIM	0.7101	0.644	0.899	0.9523

Table 2. Comparison between ESRGANnew and baseline models

Method	Type	Strengths	Weaknesses	Relative to ESRGANnew
Bicubic	Interpolation	Fast, simple	Blurry, poor detail recovery	Lower PSNR/SSIM, blurry outputs
SRCNN	CNN-based	First DL SR model	Shallow, poor texture recovery	Weaker perceptual quality
ESRGAN	GAN-based	Realistic textures, strong perceptual loss	May hallucinate details, heavy computation	Slightly lower PSNR/SSIM, less sharp
Transformer-based	Attention-based	Captures global context, SOTA accuracy	Expensive, memory-heavy	Higher potential accuracy, but less efficient
Proposed (ESRGANnew)	GAN + Wavelet	Balanced fidelity and perceptual realism; improved edge/texture detail	Slower convergence, requires large data	Outperforms ESRGAN, more efficient than transformers

5. CONCLUSION

Artificial intelligence advancements have significantly improved video and image resolution methods, most notably with the development of the ESRGAN. ESRGAN provides noticeably higher performance and superior image clarity when compared to previous SRGAN models and older techniques like bicubic interpolation. This procedure is further enhanced by the wavelet multi-scale analysis, which efficiently manages different frequency components in pictures. The proposed method ESRGANnew, which combines the Wavelet technique with the ESRGAN method, introduces good results in terms of improving the image and approaching the improvement process at a very suitable time. The proposed method demonstrates that ESRGANnew outperforms the ESRGAN method by a large margin on benchmarked images. The values of PSNR and SSIM in this work make ESNRnew a distinguishable model. ESRGANnew integrates perceptual and adversarial losses, enabling sharper and more realistic textures than SRCNN, and improves structural boundaries and sharpness compared with ESRGAN.

6. FUTURE SCOPE

Future scope for this work: first, the model can be optimized by training another model of AI networks to reduce resource consumption and achieve a lightweight network structure; second, apply and evaluate another type of DWT like density dual tree discrete wavelet transforms (DDDT-DWT) method, or curvelet technique will be used to find out whether this developed version of Wavelet technique and curvelet will contributes to develop this model and improve the images or not; third, using another types of images like medical images, satellite images, and road observation images (vehicles). In future extensions, achieving real-time inference of this study would transform the model from a post-processing tool into a core component of live systems. A multimodal extension would leverage textual context to guide the enhancement process, leading to semantically accurate and visually consistent results. In future extensions of this study, the LPIPS metric will be further validated using larger and more diverse datasets. While the current evaluation with 40 test images provided useful indicative results, expanding the dataset will strengthen the statistical reliability of LPIPS and allow for more comprehensive comparisons with state-of-the-art methods. Additionally, future experiments may include domain-specific datasets (e.g., medical, satellite, or surveillance images) to evaluate the generalization of LPIPS-based perceptual assessment across different application scenarios.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to everyone who provided support and assistance in completing this research, particularly to the Department of Electronic Engineering, College of Electrical Engineering, University of Technology, Baghdad, Iraq, for their continuous encouragement and valuable contributions throughout the development of this work.

FUNDING INFORMATION

This research received no external funding.

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
Salwa A. Alagha	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Hadeel N. Abdullah	✓	✓				✓			✓	✓	✓			
Suad Khairi Mohammed		✓		✓	✓		✓	✓		✓				

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.

INFORMED CONSENT

This study does not include any personal information, identifiable patient data, or human subjects requiring informed consent. Therefore, informed consent was not necessary for this research.

ETHICAL APPROVAL

This study did not involve human participants, human data, or animal experiments. Therefore, ethical approval and informed consent were not required.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author, [HNA], on request.

REFERENCE

[1] R. W. Saleh and H. N. Abdullah, "Early diabetic retinopathy detection using convolution neural network," *Revue d'Intelligence Artificielle*, vol. 37, no. 1, pp. 101–107, 2023, doi: 10.18280/ria.370113.

[2] M. Mariani and Y. K. Dwivedi, "Generative artificial intelligence in innovation management: a preview of future research developments," *Journal of Business Research*, vol. 175, no. 3, pp. 1–21, 2024, doi: 10.1016/j.jbusres.2024.114542.

[3] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020, pp. 6685–6840, doi: 10.5555/3495724.3496298.

[4] S. Shang *et al.*, "ResDiff: combining CNN and diffusion model for image super-resolution," in *The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)*, 2024, vol. 38, no. 8, pp. 8975–8983, doi: 10.1609/aaai.v38i8.28746.

[5] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10674–10685, doi: 10.1109/CVPR52688.2022.01042.

[6] G. Baykal, H. F. Karagoz, T. Binhuraib, and G. Unal, "ProtoDiffusion: classifier-free diffusion guidance with prototype learning," in *Proceedings of Machine Learning Research*, 2023, vol. 222, pp. 106–120.

[7] X. Li *et al.*, "Diffusion models for image restoration and enhancement: a comprehensive survey," *International Journal of Computer Vision*, vol. 133, no. 11, pp. 8078–8108, 2025, doi: 10.1007/s11263-025-02570-9.

[8] Z. Luo, F. Gustafsson, Z. Zhao, J. Sjölund, and T. Schön, "Taming diffusion models for image restoration: a review," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 383, no. 2299, 2025, doi: 10.1098/rsta.2024.0358.

[9] Y. Jiang, Z. Zhang, T. Xue, and J. Gu, "AutoDIR: automatic all-in-one image restoration with latent diffusion," in *Computer Vision – ECCV 2024*, pp. 340–359, doi: 10.1007/978-3-031-73661-2_19.

[10] X. Wang *et al.*, "ESRGAN: enhanced super-resolution generative adversarial networks," in *Computer Vision – ECCV 2018 Workshops*, Cham, Switzerland: Springer, 2018, pp. 63–79, doi: 10.1007/978-3-030-11021-5_5.

[11] T. Wang, W. Sun, H. Qi, and P. Ren, "Aerial image super resolution via wavelet multiscale convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 769–773, May 2018, doi: 10.1109/LGRS.2018.2810893.

[12] K. Fu, J. Peng, H. Zhang, X. Wang, and F. Jiang, "Image super-resolution based on generative adversarial networks: a brief review," *Computers, Materials and Continua*, vol. 64, no. 3, pp. 1977–1997, 2020, doi: 10.32604/cmc.2020.09882.

[13] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2023, doi: 10.1109/TPAMI.2022.3204461.

[14] J. Johnson, A. Alahi, and L. F. -Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, Cham, Switzerland: Springer, pp. 694–711, doi: 10.1007/978-3-319-46475-6_43.

[15] J. Song, H. Yi, W. Xu, X. Li, B. Li, and Y. Liu, "Dual perceptual loss for single image super-resolution using ESRGAN," 2022, *arXiv:2201.06383*.

[16] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595, doi: 10.1109/CVPR.2018.00068.

[17] M. H. Eybposh, C. Cai, A. Moossavi, J. R. -Romaguera, and N. C. Pégard, "ConIQA: a deep learning method for perceptual image quality assessment with limited data," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-70469-5.

[18] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[19] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1132–1140, doi: 10.1109/CVPRW.2017.151.

[20] D. Song, Y. Wang, H. Chen, C. Xu, C. Xu, and D. Tao, "AddersR: towards energy efficient image super-resolution," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 15643–15652, doi: 10.1109/CVPR46437.2021.01539.

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

[22] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+ : further improving enhanced super-resolution generative adversarial network," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2020, pp. 3637–3641, doi: 10.1109/ICASSP40776.2020.9054071.

[23] M. S. Rad, B. Bozorgtabar, U. V. Marti, M. Basler, H. K. Ekenel, and J. P. Thiran, "SROBB: targeted perceptual loss for single image super-resolution," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2710–2719, doi: 10.1109/ICCV.2019.00280.
[24] M. D. Hssayeni and B. Ghoraani, "Deep regression modeling for imbalanced and incomplete time-series data," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 6, pp. 3767–3778, 2024, doi: 10.1109/TETCI.2024.3372435.
[25] Y. Jo, S. W. Oh, J. Kang, and S. J. Kim, "Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3224–3232, doi: 10.1109/CVPR.2018.00340.
[26] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3262–3271, doi: 10.1109/CVPR.2018.00344.
[27] M. Liu, S. Jin, C. Yao, C. Lin, and Y. Zhao, "Temporal consistency learning of inter-frames for video super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 4, pp. 1507–1520, 2023, doi: 10.1109/TCSVT.2022.3214538.
[28] S. A. A. Al-Hameed, H. N. Abdullah, N. H. Khalf, and J. M. Alghazo, "An enhanced steganography approach for concealing audio in images using double density-dual tree wavelet transform," *Revue d'Intelligence Artificielle*, vol. 37, no. 5, pp. 1237–1244, 2023, doi: 10.18280/ria.370516.
[29] M. Chu, Y. Xie, J. Mayer, L. L. -Taixé, and N. Thuerey, "Learning temporal coherence via self-supervision for GAN-based video generation," *ACM Transactions on Graphics*, vol. 39, no. 4, pp. 75:1--75:18, 2020, doi: 10.1145/3386569.3392457.
[30] Y. Guo *et al.*, "Closed-loop matters: dual regression networks for single image super-resolution," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5406–5415, doi: 10.1109/CVPR42600.2020.00545.

BIOGRAPHIES OF AUTHORS

Salwa A. Alagha is currently a lecturer and a researcher in the Department of Electrical Engineering, University of Technology. She holds a bachelor's degree from the Department of Computer Science at the University of Technology in 1994. She received her master's degree and Ph.D. degree in Computer Science at the University of Technology in 2023 and 2016, respectively. Her research interests include image processing and multimedia. She can be contacted at email: salwa.a.alagha@uotechnology.edu.iq.

Hadeel N. Abdullah is a professor in the Department of Electronic Engineering at the University of Technology, Iraq. She obtained her bachelor's degree in Control and Systems Engineering from the University of Technology in 1993, followed by a master's and Ph.D. in Electrical Engineering from the same institution in 2000 and 2005, respectively. Her research expertise spans signal and image processing, artificial intelligence, and object detection and tracking, where she has made significant contributions to these fields. She can be contacted at email: hadeel.n.abdullah@uotechnology.edu.iq.

Suad Khairi Mohammed is currently a lecturer and a researcher in the Department of Electrical Engineering at the University of Technology. She received her B.Sc. in 1997 from the University of Technology, Iraq. She received her M.Sc. degree in 2004 from the University of Technology, Iraq. She received her PhD degree in 2018 from the University of Malaysia Pahang in Electronic Engineering. Her research interests include optimization, neural networks, and digital design. She can be contacted at email: suad.k.mohammed@uotechnology.edu.iq.