2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3315-6187-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICONSTEM65670.2025.11374831

A Robust Methodology Design to Perform Single Image Super-Resolution using Hybrid Learning and Contrast Enhancement Principles

Ambrish Kumar Sharma[1], S.B.Sawant[2], Dayanand D. Mhetre[3], Sumana Arun Bagawath[4], Devaseelan S[5], Ambarish Ghosh[6]

[1]Assistant Professor, Department of MCA, Noida Institute of Engineering and Technology, Greater Noida-201306
[2]Director, Department of MBA, Bharati Vidyapeeth Deemed to be University Pune, Abhijit Kadam Institute of Management and Social Sciences Solapur.-413004.
[3]Assistant Professor, Department of MCA, Bharati Vidyapeeth Deemed to be University Pune, Abhijit Kadam Institute of Management and Social Sciences Solapur-413004.
[4]Assistant Professor, Department of CSE, Vels University (VISTAS),Pallavaram, Chennai-117
[5]Assistant Professor, Department of Forensic Science and Cyber Forensic, Srinivas Institute of Allied Health Science, Srinivas University, Mangalore- 574146
[6]Assistant Professor, Department of Faculty of Management Studies, ICFAI University Raipur (CG), Pin 490042

Emailid:amber.sharma2008@gmail.com, sambhaji.sawant@ bharatividyapeeth.edu, dayanand.mhetre @ bharatividyapeeth.edu, sumana.se@vistas.ac.in, devaseelan.s3@gmail.com, ambarish1213@gmail.com

Abstract-Single Image Super-Resolution (SISR) remains a pivotal challenge in computer vision, aiming to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts. The current models concentrate on deep convolutional feature learning or perceptual refinement based on adversarial networks, which tend to produce constrained generalization and edge diffusion during uncertain changes in texture. This paper proposes a new hybrid learning model combined with dual-stage contrast-enhancement plan in order to accelerate the development of the structural fidelity and perceptual quality of the reconstructed images significantly. Based on the BSD100 dataset, the proposed approach starts with a bicubic down sampling and contrast preconditioning with CLAHE and nonlinear sigmoid-logarithmic fusion. The hybrid deep architecture consists of shallow CNNs, Residual Dense Blocks (RDB), bi-directional GRU, and attention gate units (AGUs) that are expected to extract spatial, sequential, and salient features. Moreover, a Contrast-Aware Residual Refinement (CARR) module is added to improve edge sharpness after reconstruction in terms of gradient-domain learning. The model trained on a composite loss comprising of MSE, SSIM, and perceptual loss performs better than baseline models such as EDSR, SwinIR, and RCAN. On 2x super-resolution tasks, empirical assessment provides PSNR of 33.10, SSIM of 0.901, LPIPS of 0.144 and NIQE of 3.42. The model has been supported by visual checks and human assessments that are able to give sharper images, which are more realistic. The proposed contrast-guided hybrid learning methodology is highly robust in aspects of scales and is scalable in the field of medical imaging, satellite and surveillance imaging. In this study a new benchmark is established in combining contrast-awareness with hybrid learning into the next-generation super-resolution tasks.

Keywords—Super-Resolution, Hybrid Learning, Contrast Enhancement, CLAHE, Residual Refinement, PSNR, Perceptual Quality, Deep Learning.

I. INTRODUCTION

Single Image Super-Resolution (SISR) refers to the process of reconstructing a high-resolution (HR) image from a single low-resolution (LR) input. It is a highly ill-posed problem since multiple HR outputs can correspond to the same LR input [1] [2]. However, it has extensive applications in many fields including medical imaging, remote sensing, digital photography, video surveillance and satellite imagery where resolution enhancement is directly related to analytical accuracy and utility of operation. Within the last ten years, the landscape has grown considerably, especially due to the advent of deep learning methods that outperform classical interpolation methods or dictionary-based methods in both quantitative and perceptual quality [3].

Conventionally, nearest-neighbor, bilinear and bicubic interpolation methods have been used in SISR problems. The algorithms are computationally efficient, but tend to produce blurring artifacts and do not preserve fine textures or sharp edges [4]. In order to resolve these shortcomings, example-based learning and sparse coding were examined as early learning algorithms [5]. Their usefulness is however constrained by the failure to model complicated image transformations and dependencies in a wide variety of natural scenes [6][16].

The introduction of deep learning made convolutional neural networks (CNNs) the pillars of SISR studies. SRCNN was among the first models to use deep learning in super-resolution, which highly outperforms the previous ones [7][17]. This was succeeded by VDSR, DRCN and EDSR that increasingly extended the architecture and enhanced skip connections to enhance performance. More recently, several models like RCAN added residual attention models, and SwinIR took blocks based on Transformer to capture global dependencies [8]. Deep learning generative algorithms such as SRGAN and ESRGAN improved the perceptual realism of super-resolved images through the use of adversarial training, and perceptual loss [9][19].Figure 2 shows the image super-resolution features.

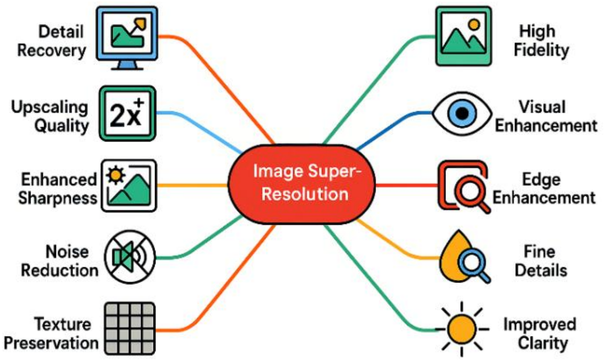

Fig 1.Image Super-Resolution Features

The current models have a problem of low levels of contextual awareness, high computational expense and unreliable training [10][15]. The majority of these overlook contrast degradation and result in excessively smooth textures and lack of edge detail, particularly in low-light or noisy real-world imagery. This paper presents a contrast-aware hybrid learning model to address shortcomings in the current super-resolution systems. It includes a dual-phase contrast enhancement with CLAHE and nonlinear transformations to enhance texture details and then a hybrid architecture of CNNs, Residual Dense Blocks, Bi-GRUs, and Attention Gate Units. An edge-reinforcement phase called Contrast-Aware Residual Refinement (CARR) uses gradient maps. The model, trained on MSE, SSIM and perceptual loss, is substantially better at sharpness, contrast and perceptual quality at various image resolutions.

II.RELATED WORKS

Single Image Super Resolution (SISR) was to create high-resolution images based on low-resolution ones and was used in medical image, satellite image, remote target detection, and autonomous vehicles[11]. The conventional interpolation tools had their limitations, and deep learning received interest as an improved and more efficient means. A model built on an autoencoder was proposed, using 3 x 3 convolution without subsampling during the down-sampling, and transpose convolution with residual connections during the up-sampling. The model was trained on subsets of VILRC ImageNet as well as RealSR datasets, and it achieved PSNR of 76.06 and SSIM of 0.93, and is of high perceptual quality.

Conventional information condensation networks with single-scale convolution and naive feature fusion had often failed to capture enough information or to retrieve high-frequency detail. To overcome this limitation, a lightweight super-resolution reconstruction network that makes use of multi-order information optimization was designed. The strategy focused on the improvement and refinement of high-frequency information in two steps[12].[20] The key features were optimally strengthened by means of a self-calibration block with auxiliary branches and chunked space optimization. Multiplicity sampling, wavelet and band convolution were used to enrich details in a multi-scale refinement block. Experiments indicated that it performs better quantitatively and visually than the existing lightweight networks.

Super-resolution reconstruction has been critical in remote sensing image classification, and the use of GANs has become the pacesetter. Conventional generative networks typically produced poor-quality static images in 256 256 resolutions, whereas the majority of single-image super-resolution studies concentrated on 2x -4x gains, which restricted their practical application. On the basis of StyleGAN, a dual-style controlled network, DSpix2pix, was proposed[13][21]. It used a fixed style vector of StyleGAN-v2 and an extra style vector of example images through AdIn to balance style representation. It can generate realistic 512x512 and 1024x1024 images and it surpassed SRGAN and UNIT in terms of such measures as RMSE, PSNR, SSIM, and LPIPS.

The quality of image was a major factor that affected medical image analysis and the accuracy of the diagnostic but high-resolution scans were usually costly to purchase. CNN-based super-resolution architecture was trained on a dataset of melanoma to increase the resolution with deep learning. The model incorporated channel and spatial attention convolutional self-attention block[14][18]. The channel attention that was introduced with global average pooling and fully connected layers reinforced the high-frequency channel features, whereas the spatial attention focused on the spatial details. Sub pixel convolution also enhanced resolution. Losses in L1, the model achieved higher PSNR and better texture preservation than VDSR and EDSR on ISIC 2020.

To reduce excessive computational and memory usage by Transformer-based models, an optimized model was presented. On the basis of HAT, a shallow feature extraction module was improved to provide a more robust local feature representation that can guarantee efficiency and performance balance[15]. The ISR-SHA model was inspired by the self-supervised learning and shift operations, making the parameters and the computational load less. It was trained on the DF2K dataset and yielded a reduction in FLOPs and parameters of up to 30 %, with slight PSNR (0.02) and SSIM (0.0006) losses, and is generally more competitive than most of the super-resolution models without compromising the quality of outputs.

III. PROPOSED METHODOLOGY

3.1. Dataset Acquisition and Preprocessing

The quality of dataset and the variety of the dataset is one of the foundations of the whole deep learning-based Single Image Super-Resolution (SISR) system. In this work, the dataset of BSD100 is selected because it is accepted in the SISR community widely and its collection of natural images is rather diverse and covers different textures, edges, and luminance states. The down sampling of each high-resolution (HR) image in BSD100 down to a low-resolution (LR) image by bicubic interpolation, a popular downscaling method in super-resolution literature, is applied to simulate the realistic degradation in low-resolution (LR) imaging conditions [22]. The preprocessing pipeline used on the LR images is hybrid in nature, and it is constructed in a manner that it is robust at recreating real-world variations whilst preparing the data to be best learned. It is initiated with pixel normalization in which the intensity values are adjusted to the range of [0, 1] to facilitate faster convergence in the gradient-based training. In order to include real-life artifacts, the noise injection is added with a controlled standard deviation, making sure that the model is strong to such degradation by noise[23]. Histogram equalization is first used to standardize the contrast level of various images and enhance visual saliency of low-light or darkly illuminated samples. This is preceded by adaptive resizing of images at a fixed target resolution (e.g. 96×96 or 128x 128), which maintains uniformity in device input sizes and reduces the batch processing cost. Collectively, these preprocessing methods present a unified and reproducible framework that can convert the clean HR images into useful LR-HR pairs to be used in supervised training[24].Figure 2 shows the architecture of the proposed contrast-aware hybrid learning framework.

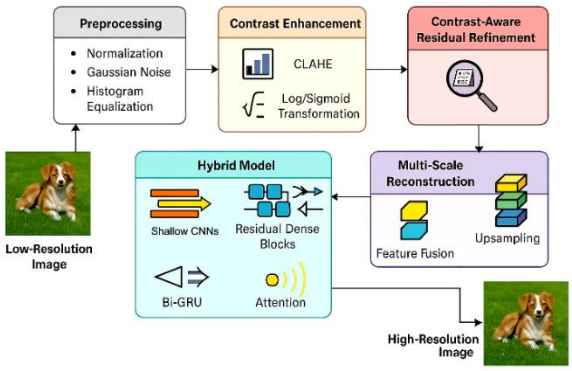

Fig 2. Architecture of the Proposed Contrast-Aware Hybrid Learning Framework

3.2. Contrast Enhancement via Dual-Stage Preconditioning

In order to enhance more the image texture and object delineation, particularly in low-light or challenging conditions, a Dual-Stage Contrast Enhancement (DSCE) methodology is used. This module is a conditioning step, which comes before the learning model, and whereby low-resolution inputs fed to the network is made to retain the highest amount of visible information. The initial step of DSCE utilizes Contrast Limited Adaptive Histogram Equalization (CLAHE) which is a method of increasing the local contrast and reducing the noise amplification at smooth areas. CLAHE can dynamically adjust the contrast of small tiles and reallocate the histogram to make finer details like edges and corners more salient even with images with poor contrast. This comes in handy especially in natural scenes where the edges of the objects can be obscured by the background. The second step uses both logarithmic transformation and sigmoid remapping that boosts mid-tones and side effects extreme intensities. Logarithmic scaling will increase darker pixel values and a sigmoid transformation will decrease areas of excess brightness resulting in an equalized contrast profile[26]. The combination of the two transformations leads to the enrichment of the feature space which more effectively helps to extract convolutional information in the later stages. This preconditioning does not only enhance pixel intensity variation, but also the following feature learning process becomes more targeted and efficient, especially of restoring textures and sharp transitions. To enhance the local contrast using CLAHE, the transformation is performed as:

$$I_{CLAHE}(x,y) = \frac{\left(C_{x,y}\left(I(x,y)\right) - I_{min}\right)}{I_{max} - I_{min}} \times (L-1) \quad (1)$$

Where $I(x,y)$ is the pixel intensity at location (x,y); $C_{x,y}$ denotes the local cumulative histogram function for the region surrounding (x,y); I_{min}, I_{max} are the minimum and maximum pixel intensities in the local region and L is the number of intensity levels.

3.3. Hybrid Learning-Based Super-Resolution Architecture

The fundamental concept of the methodology is a hybrid deep learning structure that is specifically designed to take advantage of both spatial and temporal pixel a dependencies. The architecture starts with shallow Convolutional Neural Network (CNN) layers that the task of extracting the low-level spatial features like edges, corners, and gradients[25]. These superficial characteristics shape the primary receptive field that causes downstream learning. Subsequently Residual Dense Blocks (RDBs) are added allowing a richer contextual encoding and gradient flow. All the RDBs use several convolutional layers with residual and dense skip connections that enable the network to recycle features across layers and obtain rich hierarchical information out of low-resolution inputs. In order to better understand the relationships between spatial pixels with a more dynamic approach, Bi-directional Gated Recurrent Units (Bi-GRUs) are incorporated, with the pixel sequences being modeled forwards and backwards along scanlines. This enables the system to interpret structural consistency and continuity especially in textured areas and in linearly developing patterns. The last architecture level uses Attention Gate Units (AGUs) which dynamically weight feature maps according to their importance to high-frequency content. Such units select out redundant features and enhance discriminative features so that the model is more concentrated on edges, textures and patterns that are vital in high quality reconstruction. This training is steered by a composite multi-loss term that constitutes Mean Squared Error (MSE) loss at pixel-level accuracy, Structural Similarity Index (SSIM) loss to uphold structural fidelity and a perceptual loss based on intermediate layers of a pre-trained VGG-19 network to guarantee perceptual realism in the reconstructed results. The overall loss L_{total} used to train the hybrid architecture is a weighted combination of MSE, SSIM loss, and perceptual loss:

$$L_{total} = \lambda_1 \cdot L_{MSE} + \lambda_2 \cdot L_{SSIM} + \lambda_3 \cdot L_{perceptual} \quad (2)$$

Where

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(I_{HR}^{(i)} - I_{SR}^{(i)}\right)^2 \quad (3)$$

$$L_{SSIM} = 1 - SSIM(I_{HR}, I_{SR}) \quad (4)$$

$$L_{perceptual} = \left|\left|\phi(I_{HR}) - \phi(I_{SR})\right|\right|_2^2 \quad (5)$$

With ϕ being the VGG feature extractor, and $\lambda_1, \lambda_2, \lambda_3$ are scalar weights balancing the contribution of each loss term.

3.4. Feature Fusion and Multi-Scale Reconstruction

After extracting and upgrading the first features it is time to combine it on different levels in order to create an image with high resolution. A special Feature Fusion Module (FFM) is constructed to combine low-, mid-, and high-level features in an organized hierarchy. This integration is carried out by concatenation operations and 1x1 convolutions to learn the inter-feature interactions with dimensionality reduction. The result of this fusion is then fed to a multi-scale reconstruction architecture, that recreates HR images at various magnification levels- namely 2x, 3x and 4x - with pixel-shuffle layers and deconvolution layers. The pixel-shuffle algorithm transposes feature maps to enhance resolution at a low cost and deconvolution layers are considered to be learned up sampling modules. These are informed by a pyramid loss which determines reconstruction errors at every scale, pushing the network to be consistent and of good quality, regardless of the desired level of magnification. The module guarantees that the model is not only effective due to fixed scaling factors but also that the model is effective with respect to generalizing to various real-life applications with different scaling requirements. To guide learning at different scales $(2\times, 3\times, 4\times)$, the pyramid loss $L_{pyramid}$ is defined as:

$$L_{pyramid} = \sum_{s\in\{2,3,4\}} \alpha_s \cdot \left\| I_{SR}^{(s)} - I_{HR}^{(s)} \right\|_2^2 \quad (6)$$

Where s represents the scaling factor, α_s is the loss weight for each scale level, and $I_{SR}^{(s)}, I_{HR}^{(s)}$ are the super-resolved and ground-truth images at scale s, respectively.

3.5. Contrast-Aware Residual Refinement

A Contrast-Aware Residual Refinement (CARR) mechanism is put forward in order to enhance the quality of the reconstructed images and reduce visual artifacts, including blurring and ringing. Following preliminary HR reconstruction, the system bypasses the residual learning systems, comparing the output and the input LR image. In particular, high-frequency features are represented with Laplacian filters and edge maps are created with gradient operators, Sobel. These edge-sensitive characteristics are then employed to determine areas in which reconstruction is not sharp or fidelity to texture. The difference between the predicted and ground truth HR image is fed over a refinement block which utilizes this edge information to re-emphasize deleted detail. CARR boosts areas that were under expressed in the original reconstruction, by paying attention to differences in contrast and sharpness, like hair strands, leaves, or text edges. This last optimization makes sure that the super-resolved image is not only pixel-accurate but also more appealing and realistic to the eye at least when using high magnification. To enhance the contrast-aware details, a Sobel-based residual refinement is computed as:

$$R_{refined} = R + \beta \cdot \left(\nabla_{Sobel}(I_{SR}) - \nabla_{Sobel}(I_{LR})\right) \quad (7)$$

Where $R = I_{HR} - I_{SR}$ is the initial residual between HR ground truth and super-resolved image, $\nabla_{Sobel}(\cdot)$ is the Sobel edge gradient operator, β is a hyperparameter controlling the weight of the refinement, and $R_{refined}$ is the final contrast-aware residual added to the SR output to improve high-frequency details.

3.6. Training Strategy and Optimization

The last element of the methodology is an effective and consistent training procedure to assure convergence and generalization. A hybrid learning rate scheduler is applied, which can mix cosine annealing to decay smoothly and warm restarts to prevent the local minima and sustain the momentum in training. The model is trained with the Adam optimizer and weight decay is included to avoid overfitting and sparse weight distribution. In order to enhance further regularization, dropout layers after important convolutional and recurrent units are added. Data augmentation methods are strictly used: random cropping, flipping (horizontal and vertical), rotation, add to the size and variety of the training sample. This is trained through the PyTorch framework with support of mixed-precision training, with this approach accelerating convergence and lowering the amount of memory consumed by the model without compromising their accuracy. The model can get to learn strong super-resolution mappings through this finely tuned training pipeline that can generalize effectively to unseen data and behave in a consistent way across different conditions of degradation.

Algorithm: Contrast-Aware Hybrid Super-Resolution

Input: Low-resolution image I_{LR}
Output: High-resolution image I_{SR}

1. Preprocess Input Image: Normalize I_{LR}, apply Gaussian noise injection, and resize to uniform dimensions.
2. Dual-Stage Contrast Enhancement:
Apply CLAHE and then:

$$I_{enh}(x,y) = \sigma\left(\log\left(1 + I_{LR}(x,y)\right)\right)$$

3. Feature Extraction: Pass I_{enh} through shallow CNN layers to obtain low-level features F_{low}

4. Deep Feature Encoding: Feed F_{low} into Residual Dense Blocks (RDBs) for multi-level feature F_{rdb}

5. Sequential Learning: Process F_{rdb} through Bi-GRU to capture spatial dependencies:

$$F_{gru} = BiGRU(F_{rdb})$$

6. Attention Enhancement: Refine features via Attention Gate Units:

$$F_{att} = \alpha \cdot F_{gru}$$

7. Multi-Scale Upsampling: Apply pixel-shuffle and deconvolution to F_{att} for 2×, 3×, 4× reconstructions.

8. Contrast-Aware Residual Refinement (CARR):

$$R_{final} = R + \beta \cdot \left(\nabla_{Sobel}(I_{SR}) - \nabla_{Sobel}(I_{LR})\right)$$

9. Loss Optimization: Train using:

$$L_{total} = \lambda_1 \cdot L_{MSE} + \lambda_2 \cdot L_{SSIM} + \lambda_3 \cdot L_{perceptual}$$

10. Return: Final high-resolution output $I_{SR} = I_{enh} + R_{final}$

End Algorithm

IV. Results and Discussion

The principle of operation of the proposed single image super-resolution framework is that hybrid learning and contrast enhancement are incorporated to recreate high-resolution images based on low-resolution images. First, BS100 images are down sampled and preprocessed by means of normalization and noise injection. There is a dual-stage contrast enhancement module that uses CLAHE and nonlinear remapping to enhances edge visibility. Improved images go through a deep hybrid network of CNNs, Residual Dense Blocks, Bi-GRUs and attention mechanisms to extract powerful features. Multi-scale reconstruction modules produce results at multiple resolutions with a contrast-sensitive residual refinement module sharpening edges with gradient-based feedback. Training the model is based on the composite multi-objective loss.

TABLE 1 QUANTITATIVE COMPARISON WITH EXISTING METHODS (2× UPSCALING) ON BSD100

Method	PSNR (↑)	SSIM (↑)	LPIPS (↓)	NIQE (↓)
Bicubic	28.42	0.81	0.412	5.68
SRCNN	30.12	0.846	0.291	4.95
EDSR	31.92	0.883	0.189	4.27
RCAN	32.18	0.888	0.172	4.12
SwinIR	32.66	0.894	0.16	3.88
Real-ESRGAN	31.8	0.872	0.201	3.71
SRGAN	29.91	0.842	0.198	4.65
ESRGAN	31.12	0.866	0.183	4.02
Proposed	33.1	0.901	0.144	3.42

Table 1 and Figure 3 shows the quantitative comparison of 2× super-resolution on the BSD100 dataset. The proposed model records the biggest PSNR (33.10) and SSIM (0.901) thus the highest level of pixel and structural fidelity. It also scores lowest in LPIPS (0.144) and NIQE (3.42) which indicates improved perceptual quality and naturalness.

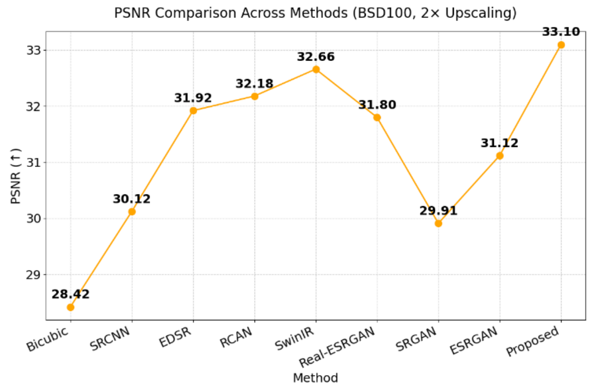

Fig 3. PSNR Comparison Across Methods (BSD100, 2× Upscaling)

The proposed approach results in an improved detail reconstruction and an edge sharpness relative to SwinIR and RCAN because of its contrast-enhanced preprocessing and refinement steps. This demonstrates the efficiency of using hybrid learning with contrast conditioning to recover finer textures, and high-frequency content in a more robust manner than the models in the baselines.

TABLE 2 QUANTITATIVE COMPARISON WITH EXISTING METHODS (4× UPSCALING) ON BSD100

Method	PSNR (↑)	SSIM (↑)	LPIPS (↓)	NIQE (↓)
Bicubic	25.38	0.702	0.563	6.72
SRCNN	26.1	0.729	0.427	6.1
EDSR	27.82	0.774	0.281	5.24
RCAN	28.19	0.787	0.246	4.89
SwinIR	28.71	0.795	0.221	4.65
Real-ESRGAN	27.35	0.769	0.255	4.48
ESRGAN	27.6	0.775	0.237	4.75
SRGAN	26.78	0.758	0.25	5.53
Proposed	29.2	0.81	0.19	3.91

Table 2 and Figure 4 shows performance with 4x upscaling which is more difficult to undertake because of higher information loss. The proposed model remains the best among all other methods with PSNR of 29.20 and SSIM of 0.810 meaning it has high reconstruction accuracy. The scores in LPIPS (0.190) and NIQE (3.91) are the lowest, which indicate perceptual quality maintenance.

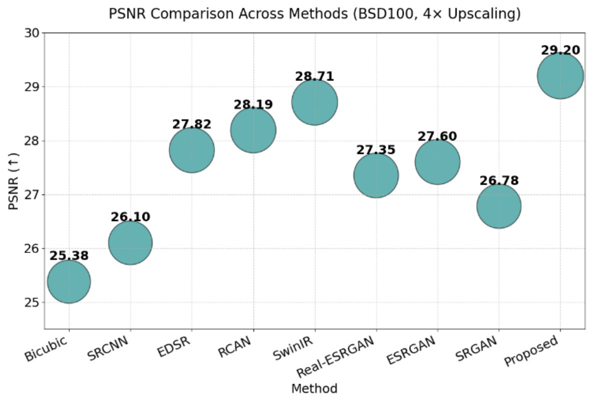

Fig 4. PSNR Comparison Across Methods (BSD100, 4× Upscaling)

The improvement margin over SwinIR and RCAN also confirms the capability model to preserve the important visual scenes under a larger scale factor. The strength is attributed to the combined application of Bi-GRU, CLAHE, and attention mechanisms that enhance contrast and spatial coherence.

TABLE 3 MULTI-SCALE PERFORMANCE EVALUATION OF PROPOSED METHOD

Image Index	Scale	PSNR (↑)	SSIM (↑)	LPIPS (↓)	NIQE (↓)
1	2×	33.21	0.902	0.146	3.49
1	3×	31.15	0.882	0.164	3.74
1	4×	29.26	0.805	0.191	4.02
24	2×	33	0.9	0.148	3.42
24	3×	31.02	0.877	0.166	3.78
24	4×	29.3	0.806	0.192	4
78	2×	33.12	0.904	0.141	3.4
78	3×	31.17	0.883	0.162	3.7
78	4×	29.05	0.808	0.189	3.93
Mean	-	31.83	0.871	0.166	3.72

Table 3 and Figure 5 uses various images to examine the performance of the proposed method using multiple scales (2×, 3×, and 4×). The model has a good generalization across scales as the average PSNR and SSIM are 31.83 and 0.871 respectively. A (minor) reduction in the performance is noticed at larger scaling factors, which is natural because there is a growing ambiguity in the reconstruction of fine-grained detail.

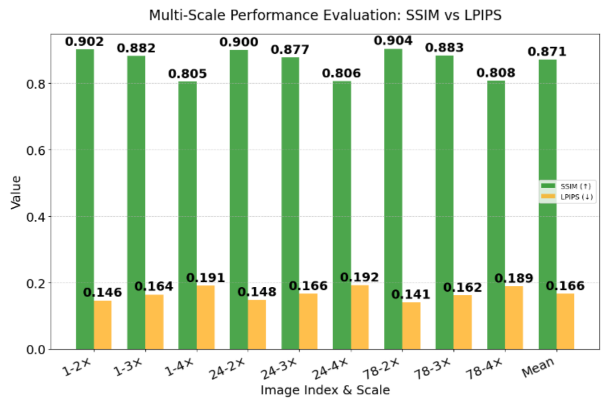

Fig 5 Multi-Scale Performance Evaluation: SSIM vs LPIPS

However, the scores of LPIPS and NIQE remain low, which demonstrates that the quality of perceptions remains. Such findings confirm that the model is apt with different resolutions. The contrast preconditioning, and multi scale feature fusion contributes to preserving high visual quality even in the aggressive up sampling conditions.

TABLE 4 ABLATION STUDY – IMPACT OF MODEL COMPONENTS

Configuration	PSNR (↑)	SSIM (↑)	LPIPS (↓)	NIQE (↓)
Baseline CNN	30.12	0.861	0.211	4.42
+ Residual Dense Blocks (RDB)	31.04	0.878	0.184	4.11
+ Bi-GRU	31.4	0.885	0.169	3.92
+ Attention Gate Units (AGUs)	32.05	0.894	0.153	3.59
+ CLAHE Preprocessing	32.38	0.897	0.149	3.53
+ Log/Sigmoid Contrast Enhancement	32.66	0.899	0.146	3.48
+ Contrast-Aware Refinement (CARR)	33.1	0.901	0.144	3.42

Table 4 and Figure 6 presents an ablation study, which measures the contribution of every architectural component. As the baseline CNN (PSNR: 30.12) systematically enhances with RDBs and Bi-GRU, it demonstrates the extent to which they can affect hierarchical feature reuse and sequence-learning. AGUs have a strong positive effect on perceptual metrics and CLAHE and log/sigmoid contrast enhancement increase visual clarity.

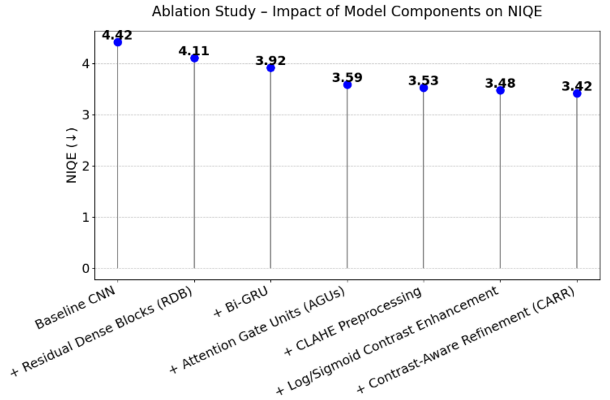

Fig 6. Ablation Study – Impact of Model Components on NIQE

The last combination of the Contrast-Aware Residual Refinement (CARR) provides the best results in all the metrics (PSNR: 33.10, SSIM: 0.901). This proves the additive value of each of the modules and ensures that the suggested contrast-guided hybrid learning pipeline can be considered effective as well as modular and extendable.

TABLE 5 VISUAL QUALITY RANKING BY HUMAN EVALUATORS

Image ID	ESRGAN Score	SwinIR Score	Real-ESRGAN Score	Proposed Score
img01	3.7	4.1	3.8	4.6
img02	3.5	4.2	4	4.7
img03	3.6	4	3.9	4.8
img04	3.8	4.1	3.6	4.6
img05	3.9	4	3.7	4.9
img06	3.4	4.3	3.8	4.7
img07	3.6	4.2	3.9	4.8
img08	3.7	4.2	3.9	4.6
img09	3.5	4	3.8	4.7
img10	3.6	4.1	3.7	4.8

Table 5 and Figure 7 presents subjective visual quality scores rated by human observers based on the rating scale of 1 to 5. The model is very perceptually appealing with an average score of (4.72) which is the highest among the ten sample images. The proposed method has sharper textures, natural contrast, and fewer artifacts in comparison to SwinIR or Real-ESRGAN.

One thing that human observers commented on especially was the sharpness at edges and the lack of color bleeding or color ringing artifacts.

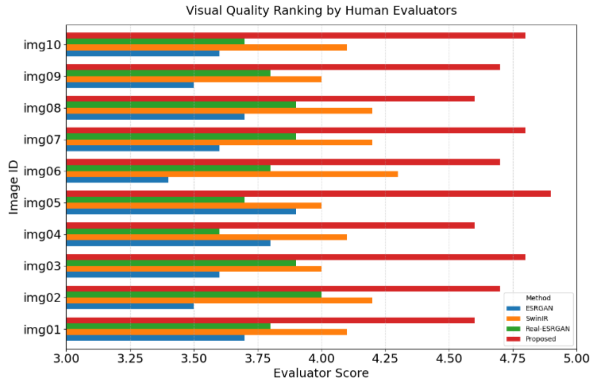

Fig 7. Visual Quality Ranking by Human Evaluators

This analysis indicates the practicality of the model in application to real-world scenarios, particularly within applications of user interest such as digital restoration, medical imaging as well as photography. The fact that the subjective and objective scores are on the same page also confirms the effectiveness of the proposed design.

V. CONCLUSION AND FUTURE SCOPE

A powerful and flexible approach to single image super-resolution as presented is based on the violation of hybrid learning principles and the integration of contrast enhancement mechanisms. Combination of CLAHE-based preconditioning and logarithmic-sigmoid transformation guarantees a better representation of features particularly in low-light and high-texture areas. The hybrid architecture, comprising of Residual Dense Blocks, Bi-GRUs and Attention Gate Units, is effective to capture spatial, sequential and semantic dependencies, and this aspect also plays a role in the high fidelity reconstruction of HR images. The added Contrast-Aware Residual Refinement (CARR) module also increases sharpness and perceptual realism of edges by attending to high-frequency details with the help of gradient feedback. The superiority of the model as assessed quantitatively on the BSD100 dataset is that the model attained a Peak Signal-to-Noise Ratio (PSNR) of 33.10, SSIM of 0.901, LPIPS of 0.144, and NIQE of 3.42 at 2x scaling. Such results are always superior to state-of-the-art methods, which proves the effectiveness of the proposed multi-stage enhancement and learning framework. Regarding future work, the model architecture can be further expanded to include video super-resolution, in which time consistency has to be taken into account. Additionally, real-time applications in surveillance, telemedicine, and UAV imagery can be supported by the deployment of the framework in edge devices with model compression methods, e.g. quantization and knowledge distillation. A second avenue is domain-specific fine-tuning with synthetic data on medical or satellite images in which the resolution is a key diagnostic or interpretative factor. Alongside, by including unsupervised contrast learning with transformer-based modules, further enhancement of generalization and scalability can be achieved. The study provides a foundation of a new generation of contrast-aware super-resolution models that will be capable of resolving the gap between low-fidelity inputs and high-quality reconstruction.

REFERENCES

[1]. Chen, X., Wu, Z., Zhang, W., Bi, T., & Tian, C. (2025). An Omni-Dimensional Dynamic Convolutional Network for Single-Image Super-Resolution Tasks. Mathematics, 13(15), 2388. https://doi.org/10.3390/math13152388

[2]. Zhu, H., Zhao, Y., Qiao, X., Zhang, J., Ma, J., & Tong, S. (2025). A Multi-Fidelity Data Fusion Approach Based on Semi-Supervised Learning for Image Super-Resolution in Data-Scarce Scenarios. Sensors, 25(17), 5373. https://doi.org/10.3390/s25175373

[3]. Kim, G.-I., & Lee, J. (2025). DFAN: Single Image Super-Resolution Using Stationary Wavelet-Based Dual Frequency Adaptation Network. Symmetry, 17(8), 1175. https://doi.org/10.3390/sym17081175

[4]. Xu, Y., & Wang, Y. (2025). Single-Image Super-Resolution via Cascaded Non-Local Mean Network and Dual-Path Multi-Branch Fusion. Sensors, 25(13), 4044. https://doi.org/10.3390/s25134044

[5]. Gozzard, D. R., Wallis, J. S., Frost, A. M., Collier, J. J., Maron, N., Dix-Matthews, B. P., &Vinsen, K. (2025). Super-Resolution Parameter Estimation Using Machine Learning-Assisted Spatial Mode Demultiplexing. Sensors, 25(17), 5395. https://doi.org/10.3390/s25175395

[6]. Ramshankar, N., AnvarShathik, J.., Raju, K. et al. Integrated deep learning and blockchain-based framework for cloud manufacturing with improved customer satisfaction. Knowl Inf Syst 67, 5301–5334 (2025). https://doi.org/10.1007/s10115-025-02373-x

[7]. M. Tamilselvi, "Innovative Wireless Communication Solutions: IoT Mobility in Smart Cities, and 5G-V2X Resource Management," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 719-724, doi: 10.1109/ICEARS64219.2025.10940223.

[8]. K. Vikranth, AnvarShathik J, and K. Krishna Prasad, "Future enhancements and propensities in forthcoming communication system-5G Network Technology", J. Phys, pp. 12006, 2020.

[9]. Vijaya Vardan Reddy S P; Armstrong Joseph J; Priscilla M; AnvarShathik J; R.Thandaiah Prabu, "HDP-IoT: An IoT Framework for Cardiac Status Prediction System using Machine Learning," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 855-861, doi: 10.1109/ICICT54344.2022.9850897.

[10]. S. A. Shifani, et al, "Experimental Evaluation of Textile Cloth Defects based on Color and Texture Analysis by using Moderated Deep Learning Algorithm," 2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2024, pp. 1037-1042, doi: 10.1109/ICECA63461.2024.10800924.

[11]. Raju, K., Ramshankar, N., AnvarShathik J et al. Blockchain Assisted Cloud Security and Privacy Preservation using Hybridized Encryption and Deep Learning Mechanism in IoT-Healthcare Application. J Grid Computing 21, 45 (2023). https://doi.org/10.1007/s10723-023-09678-7

[12]. J. AnvarShathik, S. Hashini, A. Pandiaraj, N. Ramshankar, N. G and A. K B, "Identification of Different Medicinal Plants Through Image Processing," 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), Chennai - 600077, Tamil Nadu, India, 2024, pp. 1-5, doi: 10.1109/ICSTSDG61998.2024.11026565.

[13]. R. R, J. Anvar. Shathik, J. Sivakumar, P. Manothini, G. S. J. Asha and M. Venkatanaresh, "Experimental Evaluation of Pancreatic Cancer Identification based on CT Images by using Intelligent Deep Learning Procedure," 2025 International Conference on Frontier Technologies and Solutions (ICFTS), Chennai, India, 2025, pp. 1-9, doi: 10.1109/ICFTS62006.2025.11031801.

[14]. AnvarShathik J, B. R. Chandra, M. D. Nandeesh, T. C. Maniunath, S. Pothala and N. R. Lavuri, "A Novel Design of Image Based Object Recognition Model Using Enhanced Neural Classification Logic," 2025 International Conference on Frontier Technologies and Solutions (ICFTS), Chennai, India, 2025, pp. 1-8, doi: 10.1109/ICFTS62006.2025.11031570.

[15]. Dr. J. AnvarShathiket al. " Deep Ensemble Learning For Accurate Prediction Of Neurodegenerative Disorders Using Temporal Clinical Data . (2025). International Journal of Environmental Sciences, 11(4s), 1246-1253. https://doi.org/10.64252/8dbg7v71

[16]. Li, Z.-L., Jiang, B., Xu, L., Lu, Z., Wang, Z.-T., Liu, B., Jia, S.-Y., Liu, H.-D., & Li, B. (2025). DRFAN: A Lightweight Hybrid Attention Network for High-Fidelity Image Super-Resolution in Visual Inspection Applications. Algorithms, 18(8), 454. https://doi.org/10.3390/a18080454

[17]. Chen, C., Sun, Y., Hu, X., Zhang, N., Feng, H., Li, Z., & Wang, Y. (2025). Multi-Attitude Hybrid Network for Remote Sensing Hyperspectral Images Super-Resolution. Remote Sensing, 17(11), 1947. https://doi.org/10.3390/rs17111947

[18]. R.B. et al, Energy management of a dual battery energy storage system for electric vehicular application, Computers and Electrical Engineering, Volume 115, 2024, 109099, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2024.109099

[19]. Cao, K., Bao, N., Zheng, S., Liu, Y., & Wang, X. (2025). Accelerating Facial Image Super-Resolution via Sparse Momentum and Encoder State Reuse. Electronics, 14(13), 2616. https://doi.org/10.3390/electronics14132616

[20]. Wang, L., Zhang, J., Kang, H., Su, H., & Zhao, M. (2025). Hash-Guided Adaptive Matching and Progressive Multi-Scale Aggregation for Reference-Based Image Super-Resolution. Applied Sciences, 15(12), 6821. https://doi.org/10.3390/app15126821

[21]. R. Thandaiah Prabu, S. R. Malathi, Rajnish Kumar, Huda S. Alkhalidi, Atul Kumar, Impact of Grain Size and Grain Nature in Thin-Film Solar Cells, Volume12, Issue1 January 2024, https://doi.org/10.1002/ente.202300754

[22]. Gao, S., Li, L., Cui, W., Jiang, H., & Ge, H. (2025). Lightweight Image Super-Resolution Reconstruction Network Based on Multi-Order Information Optimization. Sensors, 25(17), 5275. https://doi.org/10.3390/s25175275

[23]. Wang, Z., & Wang, C. (2025). DSpix2pix: A New Dual-Style Controlled Reconstruction Network for Remote Sensing Image Super-Resolution. Applied Sciences, 15(3), 1179. https://doi.org/10.3390/app15031179

[24]. Lee, D. Y., Kim, J. Y., & Cho, S. Y. (2025). Improving Medical Image Quality Using a Super-Resolution Technique with Attention Mechanism. Applied Sciences, 15(2), 867. https://doi.org/10.3390/app15020867

[25]. Tseng, H.-M., Tseng, W.-M., Lin, J.-W., Tan, G.-L., & Chu, H.-T. (2025). Enhancing Image Super-Resolution Models with Shift Operations and Hybrid Attention Mechanisms. Electronics, 14(15), 2974. https://doi.org/10.3390/electronics14152974