Robust Medical Image Reconstruction Using a Self-Evolving Encoder–Decoder and Adaptive Convolutional Power Scaling

Dhanusha P B[1a, b]*, Dr. J. Bennilo Fernandes[2], Dr. A. Muthukumar[3], Dr. A. Lakshmi[4]
Research Scholar, Department of Electronics and Communication Engineering, Kalasalingam Academy of Research and Education, Virudhunagar, Tamilnadu, India[1a]
Assistant Professor, Department of Electronics Engineering, SAINTGITS College of Engineering, Kottayam, Kerala, India[1b]
Assistant Professor, Department of Electronics and Communication Engineering, Kalasalingam Academy of Research and Education, Virudhunagar, Tamil Nadu, India[2]
Professor, Department of Electronics and Communication Engineering, Sri Raaja Raajan College of Engineering and Technology, Amaravathipudur, Tamilnadu, India[3]
Professor, Department of Electronics and Communication Engineering, Ramco Institute of Technolo gy, Rajapalayam, Tamilnadu, India[4]

Abstract—**Robust medical image reconstruction is a critical requirement for accurate diagnosis and clinical decision-making, particularly when images are affected by degradation, noise, or low resolution. Conventional encoder–decoder-based reconstruction methods compress input images into low-dimensional representations and subsequently decode them into high-resolution outputs; however, such approaches often suffer from artifacts and loss of fine anatomical details under severe degradation. To address these limitations, this work proposes a robust medical image reconstruction framework using a self-evolving encoder–decoder and adaptive convolutional power scaling. The proposed super-resolution model incorporates a dynamic encoder and decoder that adaptively evolve during training to capture color contrast, structural similarity, and high-frequency details from medical images. An MLP enhanced with an adaptive power flex layer is embedded within the reconstruction pipeline, enabling learnable power-based feature scaling through weight-wise modulation and initialization. This mechanism improves feature discrimination and stabilizes the reconstruction of subtle anatomical structures. The DRIVE and CHASE_DB1 retinal image datasets are employed for experimental validation, with appropriate preprocessing applied before training and testing. The selected images are processed through the proposed super-resolution model, and performance is quantitatively evaluated using PSNR, SSIM, sensitivity, and specificity metrics. Experimental results demonstrate that the proposed method achieves significant improvements in reconstruction quality and robustness compared to existing approaches, yielding enhanced perceptual quality and structural fidelity in reconstructed medical images. These findings indicate that the proposed self-evolving encoder–decoder with adaptive convolutional power scaling is well-suited for reliable medical image reconstruction applications.**

Keywords—Dynamic encoder and decoder; power flex model layer; high resolution images; weight initialization; adaptive convolutional power scaling

I. INTRODUCTION

Medical image reconstruction and super-resolution play a vital role in enhancing image quality for accurate diagnosis, treatment planning, and clinical decision-making. However, medical images acquired through modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and retinal fundus imaging often suffer from low resolution, noise, and acquisition constraints, which can obscure fine anatomical details. Early learning-based approaches for image super-resolution demonstrated the effectiveness of deep convolutional neural networks (CNNs) in reconstructing high-resolution images from degraded inputs [1]. Subsequent advancements extended these ideas to three-dimensional medical data, enabling improved spatial detail preservation in MRI reconstruction [2]. Traditional model-based methods incorporating low-rank and total variation regularization further improved reconstruction fidelity but remained computationally intensive and limited in adaptability [3].

Encoder–decoder architectures have become the backbone of many medical image processing tasks due to their strong representation learning capability [4]. Despite their success, conventional encoder–decoder frameworks often struggle with robustness when dealing with severely degraded inputs, leading to artifacts and loss of high-frequency details. To address domain variability and distribution shifts, generative and domain-adaptive learning techniques have been explored [5]. Meanwhile, perceptual quality metrics such as structural similarity index (SSIM) [6] and deep feature-based perceptual measures [7] have become standard tools for evaluating reconstruction performance beyond pixel-wise fidelity.

In retinal imaging, accurate reconstruction and enhancement are particularly important for vessel analysis and disease screening. Benchmark datasets such as DRIVE and CHASE_DB1 have been widely used to evaluate reconstruction and segmentation performance in retinal images [8], [9]. Recent advances have introduced transformer-based and hybrid CNN–Transformer architectures to capture both local and global dependencies in medical image super-resolution [10], [12], [13]. Attention-based mechanisms have further enhanced feature discrimination, leading to improved reconstruction quality and robustness [15], [21].

More recently, adaptive and dynamic network designs have gained significant attention. Dynamic convolution and shuffle-based parallel architectures enable input-dependent feature modulation, improving generalization and reconstruction performance across varying image conditions [16], [29], [30], [31]. In parallel, iterative and self-collaborative reconstruction strategies have been proposed to progressively refine high-resolution outputs using reconstruction priors [20], [27]. Diffusion-based and implicit neural representation models have further expanded the design space for medical image super-resolution, demonstrating promising results in multimodal and arbitrary-scale reconstruction tasks [18], [19]. Wan et al. introduced a cycle-constraint adversarial network for retinal image enhancement, enabling unpaired learning while effectively improving contrast, illumination consistency, and preservation of fine vascular structures, thereby supporting more reliable automated retinal analysis [31].

The majority of modern picture super-resolution techniques rely on static convolutional processes and fixed architectural parameters during training and inference. These networks only train a single mapping function, regardless of variations in contrast heterogeneity, modality-specific noise, anatomical complexity, or degradation levels. However, depending on the scanner, acquisition technique, and patient condition, medical pictures in real-world clinical settings exhibit a range of unexpected degradations. However, rather than drastically altering the encoder-decoder's reconstruction behaviour, dynamic convolution techniques and adaptive attention methods primarily focus on spatial-channel reweighting. Because of this, current frameworks sometimes produce artefacts or over-smoothed outcomes when dealing with unseen clinical variations, and they may also have trouble generalizing across deterioration intensities.

Therefore, the primary technical gap is the absence of a reconstruction framework that can dynamically adjust its feature transformation capability to shifting deterioration characteristics throughout training. Current models adjust weights to increase resilience and stability, but they don't explicitly incorporate a technique that allows for structural evolution of the encoder–decoder or adaptively controls convolutional feature power. To bridge this gap, this study proposes an adaptive convolutional power scaling self-evolving encoder-decoder system. Convolutional responses are altered using learnable power-based transformations, which change the expressiveness of features dependent on the complexity of degradation. By allowing the encoder–decoder to gradually alter during training, the proposed method maintains fine anatomical structures more successfully than static designs, enhances adaptability, and stabilizes reconstruction under extreme degradations. Comprehensive evaluations show that the proposed approach delivers exceptional numerical performance and perceptual quality in a range of medical imaging scenarios.

A. Paper Organization

The remainder of this study is organized as follows. Section 2 reviews related work on medical image reconstruction, super-resolution techniques, encoder–decoder architectures, and adaptive convolutional learning methods. Section 3 describes the proposed robust medical image reconstruction framework, including the self-evolving encoder–decoder architecture, adaptive convolutional power scaling mechanism, and overall network design. Section 4 details the experimental setup, including dataset descriptions, preprocessing steps, implementation details, training strategy, and evaluation metrics. Section 5 presents and discusses the experimental results, providing quantitative and qualitative comparisons with existing state-of-the-art methods. Finally, Section 6 concludes the study and outlines potential directions for future research.

II. Literature Review

Medical image reconstruction and super-resolution have been extensively studied to overcome limitations arising from low spatial resolution, noise, and acquisition constraints in medical imaging modalities. Early deep learning-based super-resolution methods demonstrated that convolutional neural networks (CNNs) can effectively learn nonlinear mappings between low- and high-resolution images, significantly outperforming traditional interpolation techniques [1]. These approaches laid the foundation for subsequent advances in learning-based reconstruction. To address modality-specific challenges, deep learning techniques were extended to medical imaging applications. Pham et al. [2] proposed a 3D convolutional neural network for brain MRI super-resolution, achieving improved spatial consistency across volumetric data. Similarly, Shi et al. [3] incorporated low-rank and total variation regularization to enhance MR image reconstruction, demonstrating improved edge preservation. Although effective, such regularization-based methods often require high computational cost and lack adaptability to diverse degradation patterns.

Encoder–decoder architectures have become a dominant paradigm in medical image analysis due to their ability to capture hierarchical features. The U-Net architecture introduced by Ronneberger et al. [4] has been widely adopted and extended for various biomedical imaging tasks. However, conventional encoder–decoder models often struggle to maintain robustness when reconstructing severely degraded images. To mitigate domain shifts and variability, adversarial learning and domain adaptation techniques have been explored, enabling improved generalization across datasets [5]. Evaluation of reconstructed image quality has evolved beyond pixel-wise fidelity metrics. The structural similarity index (SSIM) introduced by Wang et al. [6] has become a standard measure for assessing perceptual similarity, while deep feature-based perceptual metrics further improved the assessment of visual quality [7]. These metrics are particularly relevant in medical imaging, where preserving anatomical structures is critical.

Retinal imaging has served as a benchmark application for evaluating reconstruction and enhancement techniques. The DRIVE and CHASE_DB1 datasets have been widely used to assess vessel visibility and structural integrity in retinal images [8], [9]. Recent works have applied super-resolution techniques to retinal fundus images using generative adversarial networks, demonstrating notable improvements in visual quality and diagnostic relevance [11].

With the rise of attention mechanisms and transformer architectures, hybrid CNN–Transformer models have shown promising performance in medical image super-resolution. Zhou [10] introduced a multi-scale convolution-aided transformer framework to capture both local and global features. Chen et al. [12] proposed a mutual co-attention network for joint MR image reconstruction and super-resolution, achieving improved structural fidelity. Further advancements integrated gated fusion mechanisms and attention modules to enhance feature interaction and reconstruction robustness [13], [15], [21].

Recent research has increasingly focused on adaptive and dynamic networks. Adaptive dynamic shuffle convolutional parallel networks have demonstrated improved feature diversity and efficiency in super-resolution tasks [16], [30]. Dynamic convolution strategies, such as snake convolution, enable input-dependent kernel modulation, allowing better adaptation to structural variations within images [29]. Iterative and collaborative reconstruction frameworks have further enhanced performance by progressively refining high-resolution outputs using reconstruction priors [20], [27]. Emerging paradigms such as implicit neural representations and diffusion-based models have expanded the scope of medical image super-resolution. Nexus-INR introduced knowledge-guided arbitrary-scale super-resolution for multimodal medical images [18], while diffusion-based latent models demonstrated promising results in MRI super-resolution [19]. Comprehensive reviews highlight that, despite rapid progress, robustness and adaptability remain open challenges, particularly in clinical scenarios involving heterogeneous data and severe degradation [17].

Chang et al. [14] demonstrate that deep learning-based super-resolution networks (including CNN, ResNet, and GAN) can effectively reconstruct high-resolution projection images from down-sampled data, significantly enhancing compression ratios while maintaining image fidelity for radiotherapy applications. Jannat et al. [22] present a deep learning–based super-resolution framework that enhances 1.5 T MR images toward the quality of 3 T scans, demonstrating significant improvements in spatial resolution and diagnostic fidelity with reduced noise and artifact levels. Zhou et al. [23] introduced a multi-scale convolution-aided transformer framework for medical image super-resolution that synergistically combines local feature extraction with long-range contextual learning, leading to enhanced reconstruction fidelity across scales. Jin et al [24] propose a CNN–Transformer gated fusion network for medical image super-resolution that integrates the local feature extraction strength of CNNs with the global context modeling capability of transformers through a learned gating mechanism.

Kobayashi et al. [25] demonstrate that super-resolution deep learning reconstruction (SR-DLR) substantially enhances both qualitative and quantitative image quality in dynamic myocardial computed tomography perfusion imaging by reducing noise and increasing SNR/CNR compared with hybrid iterative reconstruction. Yang et al. [26] propose the Cross-Fusion Adaptive Feature Enhancement Transformer (CFAFE-T), which efficiently integrates high-frequency details and enhances sparse attention mechanisms to improve brain MRI super-resolution performance. Huang et al. [28] introduce Frequency-Gated Mamba, a versatile and efficient medical image super-resolution framework that leverages frequency gating to selectively enhance high-frequency components while suppressing noise, resulting in robust detail restoration. Wan et al. [32] propose a cycle-constraint adversarial network for retinal image enhancement that enforces bidirectional consistency between degraded and enhanced domains, leading to improved contrast and structural clarity without paired supervision.

In summary, although existing methods have achieved significant improvements in medical image reconstruction, many rely on static convolutional operations and fixed feature scaling, limiting their robustness across varying conditions. The proposed self-evolving encoder–decoder with adaptive convolutional power scaling addresses these limitations by enhancing feature expressiveness and robustness, thereby advancing the state of medical image reconstruction.

III. Research Methodology

A. Overview of the Proposed Reconstruction Framework

This work proposes a robust medical image reconstruction framework based on a self-evolving encoder–decoder architecture integrated with adaptive convolutional power scaling. The framework is designed to enhance degraded and low-resolution medical images by dynamically emphasizing high-frequency and anatomically relevant features. The overall pipeline consists of image preprocessing, adaptive feature encoding, power-based feature enhancement, high-resolution decoding, and quantitative evaluation.

B. Dataset and Pre-Processing

The proposed work utilizes two publicly available retinal fundus image datasets: DRIVE (Digital Retinal Images for Vessel Extraction) and CHASE_DB1 (Child Heart and Health Study in England Database1). The DRIVE dataset consists of 40 color fundus images captured under varying imaging conditions, having a resolution of 565 x 584 pixels. Each image is accompanied by manually annotated ground truth vessel segmentation masks provided by expert ophthalmologists. The CHASE_DB1 dataset contains 28 high-resolution retinal fundus images having a resolution of 999 x 960 pixels and includes manually labelled vessel annotations from two independent observers. Let the input low-resolution image be represented as

$$I_{LR} \in \mathbb{R}^{HXWXC} \tag{1}$$

Before training, all images are normalized to a fixed intensity range. Noise suppression is performed using mild Gaussian filtering. Low-resolution inputs are generated from high-resolution ground truth images using bicubic down-sampling to simulate realistic degradation conditions.

C. Self-Evolving Encoder Architecture

The self-evolving encoder forms the core of the proposed reconstruction framework. Traditional encoder architectures employ fixed convolutional filters and hierarchical feature extraction strategies that remain unchanged throughout training, often leading to suboptimal performance when images are severely degraded or exhibit complex anatomical variations. The architecture of the proposed system is shown in Fig. 1. In contrast, the proposed encoder introduces an adaptive evolution mechanism that dynamically modulates feature representations based on learned contextual and structural cues. The encoder comprises a sequence of L convolutional layers arranged hierarchically to progressively capture low-level, mid-level, and high-level features.

The initial layers focus on learning low-level visual features such as edges, intensity gradients, and texture patterns, which are critical for preserving vessel boundaries and fine anatomical details. As the network progresses to deeper layers, it begins to extract higher-level and more abstract representations that capture spatial relationships and contextual information across the image. These deep features enhance structural similarity and semantic consistency, leading to more accurate and visually coherent image reconstruction. The feature map at the l^{th} layer is computed as

$$F^{(l)} = \sigma(W^{(l)} * F^{(l-1)} + b^l) \tag{2}$$

where $*$ denotes the convolution operation, $\sigma(\cdot)$ represents the ReLU activation function, and $W^{(l)}$ and b^l denote the learnable weights and biases. For the first layer, $F^{(0)} = I_{LR}$

Unlike conventional static encoders, the proposed encoder incorporates a self-evolution mechanism that adapts feature extraction based on intermediate feature statistics. An evolution coefficient α^l is computed using global average pooling followed by a lightweight MLP:

$$\alpha^{(l)} = MLP(GAP(F^{(L)})) \tag{3}$$

The evolved feature representation is then obtained as

$$\tilde{F}^{(l)} = \alpha^{(l)} \odot f^{(l)} + (1 - \alpha^l) \odot F^{l-1} \tag{4}$$

where $\odot$ denotes element-wise multiplication. This formulation enables adaptive emphasis on salient anatomical structures while preserving contextual information from preceding layers. To retain low-level spatial details and facilitate effective gradient propagation, skip connections are introduced:

$$F^l_{skip} = \tilde{F}^{(l)} - F^{l-1} \tag{5}$$

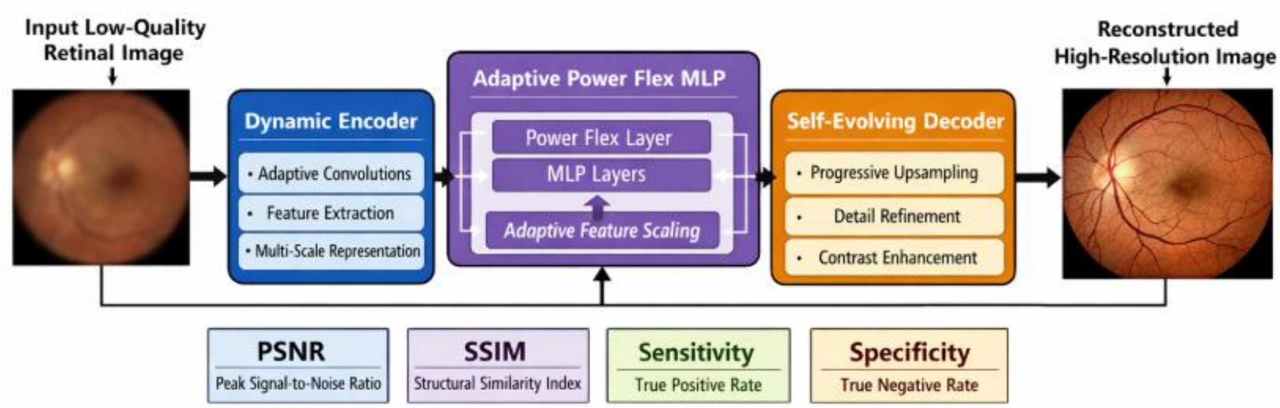

Fig. 1. System architecture.

D. Adaptive Convolutional Power Scaling Module

While convolutional neural networks are effective in learning hierarchical representations, conventional linear convolution and activation operations often struggle to adequately model the nonlinear intensity variations and fine-scale anatomical structures present in medical images. Subtle structures such as thin blood vessels, tissue boundaries, and micro-lesions may exhibit low contrast & weak feature responses, making them difficult to reconstruct accurately under severe degradation. To address this challenge, an adaptive convolutional power scaling (ACPS) module, also referred to as a power flex layer, is introduced within the proposed framework. The primary objective of the ACPS module is to enhance feature discrimination through learnable nonlinear amplification, allowing the network to dynamically emphasize clinically significant structures while suppressing noise and irrelevant background information. Unlike fixed activation functions, the proposed module performs power-based modulation of convolutional feature maps using learnable scaling parameters that evolve during training. Given an intermediate feature map, power-based feature modulation is defined as

$$F_{ps} = sign(F) \odot |F|^{\gamma} \tag{6}$$

where γ is a learnable power scaling parameter controlling the degree of feature amplification. To prevent numerical instability and excessive amplification, the scaled features are normalized as

$$\hat{F} = \frac{F_{ps}}{\|F_{ps}\| + \varepsilon} \tag{7}$$

where ϵ is a small constant.

E. Dynamic Decoder for High-Resolution Reconstruction

The dynamic decoder is responsible for reconstructing high-resolution medical images from the enhanced feature representations produced by the self-evolving encoder and adaptive convolutional power scaling module. Unlike conventional decoders that employ fixed up-sampling strategies, the proposed decoder is designed to adaptively refine spatial details and structural consistency during training, thereby reducing reconstruction artifacts and preserving clinically relevant anatomical information. The decoder reconstructs the high-resolution output image using the enhanced feature representation $\widehat{F}$. Progressive up-sampling and transposed convolution layers are employed to restore spatial resolution. Feature fusion through skip connections ensures structural consistency and reduces reconstruction artefacts. The final reconstructed image is expressed as

$$I_{SR} = D(\widehat{F}) \tag{8}$$

where $D(\cdot)$ denotes the dynamic decoder network.

F. Training Strategy and Loss Function

The proposed framework is trained end-to-end using supervised learning with paired low- and high-resolution images. The overall loss function is defined as

$$\mathrm{L} = \lambda_1 L_{MSE} + \lambda_2\,(1\text{-SSIM}) \tag{9}$$

$$\text{Where } L_{MSE} = \frac{1}{N}\sum_{i=0}^{N}\left\| I_{SR}^{(i)} - I_{HR}^{(i)} \right\| \tag{10}$$

Here, λ_1 (0.8) and λ_2 (0.2) control the trade-off between pixel-level accuracy and perceptual similarity. Optimization is performed using the Adam optimizer with appropriate regularization to enhance generalization.

The model was trained on DRIVE and CHASE_DB1 using a 70%–10%–20% split for training, validation, and testing. A ×4 upscaling factor was applied, and images were divided into 128×128 patches to improve generalization. Training was performed for 50 epochs using the Adam optimizer with a learning rate of 1×10^{-4}. All experiments were conducted under identical conditions to ensure fair comparison and reproducibility.

G. Evaluation Metrics

Reconstruction performance is quantitatively evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Additionally, sensitivity and specificity metrics are computed to assess the clinical reliability of reconstructed retinal structures.

Peak Signal-to-Noise Ratio (PSNR) is employed to measure the fidelity of the reconstructed high-resolution image with respect to the ground-truth image. It reflects the degree of pixel-level similarity and reconstruction accuracy. PSNR is defined as

$$\text{PSNR}=10\ log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{11}$$

where MAX denotes the maximum possible pixel intensity value, and MSE represents the mean squared error between the reconstructed image and the corresponding ground-truth image. Structural Similarity Index (SSIM) is used to evaluate the perceptual quality of reconstructed images by considering structural information, luminance, and contrast. Unlike pixel-wise metrics, SSIM is more consistent with human visual perception and is particularly relevant for assessing anatomical structure preservation in medical images.

Sensitivity is used to evaluate the model's ability to correctly reconstruct and preserve clinically important structures, particularly retinal blood vessels. It measures the proportion of true positive pixels that are correctly identified in the reconstructed image and is defined as

$$Sensitivity = \frac{TP}{TP+FN} \tag{12}$$

where TP denotes true positive pixels, and FN represents false negatives. To evaluate clinical reliability, vessel-level sensitivity and specificity are computed using expert-annotated binary vessel masks provided in the DRIVE and CHASE_DB1 datasets as ground-truth references. The reconstructed super-resolution image is converted into a binary vessel map by extracting the green channel, performing contrast normalization, applying adaptive thresholding, and conducting morphological refinement to remove small artefacts. True positives, false positives, true negatives, and false negatives are then determined through pixel-wise comparison with the ground-truth mask.

H. Computational Complexity Analysis

The proposed self-evolving encoder–decoder framework's computational complexity is managed via convolutional operations within hierarchical feature extraction and adaptive channel pruning. For an input feature map of spatial dimension H x W, convolutional layers with kernel size k, input channels C_{in} and output channels C_{out} have a complexity of $O(HWk^2C_{in}C_{out})$. Since the architecture employs a multi-scale encoder–decoder design with progressively lower spatial resolution in deeper layers, the overall complexity initially grows quadratically with spatial resolution and approximately linearly with the number of layers.

Unlike transformer-based self-attention mechanisms, which exhibit quadratic complexity $O((HW)^2)$ due to global attention computation, the proposed model maintains convolution-dominant operations, which guarantee better scalability for high-resolution medical pictures. Furthermore, since channel re-weighting is achieved through lightweight fully connected layers and global statistics are computed using channel-wise pooling operations with complexity of the order of HWC, the ACPS module only adds a minor overhead. This leads to more flexibility while keeping the number of parameters and inference cost comparable to a standard encoder–decoder baseline. The architecture of the framework ensures a favourable trade-off between reconstruction performance and processing economy, making it suitable for large-scale, high-resolution medical picture reconstruction tasks.

IV. Results and Discussions

A. EDA (Exploratory Data Analysis)

Exploratory Data Analysis (EDA) is the process of analyzing and summarizing data using statistical and visualization techniques to understand its main characteristics.

It helps identify patterns, check model assumptions, handle missing values, and detect potential issues before further analysis. Fig. 2 illustrates the low- and high-resolution versions of the sample image. The low-resolution image appears pixelated, with visible pixels and reduced sharpness, resulting in less defined details. In contrast, the high-resolution image demonstrates improved clarity, smoothness, and finer texture representation, with sharper edges and clearly visible fine structures, providing a more accurate and realistic depiction of the subject.

B. Performance Analysis

Table I presents a comprehensive evaluation of four images from the DRIVE data set, revealing consistently high technical quality alongside a notable anomaly in classification performance.

DRIVE (LR) DRIVE (HR) Chase_DB1 (LR) Chase_DB1 (HR)

Fig. 2. Presentation of low-resolution and high-resolution images.

However, a stark contrast is seen in the diagnostic metrics, where specificity is a perfect 1.0 but sensitivity is extremely low at 0.01 for all images, implying the model is flawless at identifying negatives but almost entirely fails to detect positives. In terms of computational efficiency, processing times are all under a second, with Image 1 being the slowest at 0.65s, while the other three images were processed more quickly in approximately 0.37 seconds.

The average performance metrics presented in Table II, obtained from the CHASE_DB1 dataset, provide an overall assessment of the quality and accuracy of the evaluated image processing algorithms. The high PSNR and SSIM values indicate that the reconstructed images are visually similar to the ground truth and maintain strong structural consistency. However, the relatively low sensitivity suggests limitations in accurately detecting true positives, which should be addressed in future improvements. The computed average values for MSE, RMSE, CNR, and prediction time are 0.00113, 0.03, 29.40, and 0.42 seconds, respectively. The CHASE_DB1 dataset is used to compare various image enhancement techniques based on PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index), as presented in Tables III and IV. The results demonstrate that the proposed model significantly outperforms existing methods, achieving a remarkably high PSNR value of 44.65 and an SSIM of 1.

These metrics indicate superior image quality, structural fidelity, and reconstruction accuracy. The enhanced performance suggests that the proposed approach, incorporating advanced algorithms and deep learning–based modifications for image enhancement, provides substantial improvements over conventional techniques on the CHASE_DB1 dataset.

C. Ablation Study

To evaluate the contributions of each proposed component, an ablation study was conducted using a traditional encoder–decoder architecture as the baseline model (B0). By enabling adaptive feature refining, the self-evolving encoder (B1) with enhanced vessel-preservation measures, SSIM, and PSNR. Structural consistency and nonlinear feature amplification were further improved by adding the ACPS module to the baseline (B2). Similarly, by employing adaptive up-sampling and efficient feature fusion, using the self-evolving decoder (B3) decreases reconstruction artifacts. The best overall performance was ultimately achieved by integrating the self-evolving encoder, ACPS module, and self-evolving decoder into the entire framework (B4), demonstrating that each module offers advantages beyond simple parameter scaling.

TABLE. I. PERFORMANCE METRICS OF THE IMAGE 1-4 IN THE DATASET

	PSNR	SSIM	Sensitivity	Specificity	MSE	RMSE	CNR	Time taken (s)
Image 1	44.21	1	0.01	1	0.00134	0.03	28.63	0.65
Image 2	44.16	1	0.01	1	0.00091	0.03	30.34	0.36
Image 3	45.1	1	0.01	1	0.00095	0.03	29.63	0.37
Image 4	45.14	1	0.01	1	0.0091	0.03	30.34	0.37

TABLE. II. AVERAGE OF ALL THE IMAGES' PERFORMANCE METRICS

Total average Result	Values
Average PSNR	44.65
Average SSIM	1
Average Sensitivity	0.01
Average Specificity	1
Average MSE	0.00113
Average RMSE	0.03
Average CNR	29.40
Average time taken for prediction	0.0.42 seconds

TABLE. III. PSNR AND SSIM VALUES WITH CHASE_DB1 DATASET (NAZ & SHREEKANTH, 2021)

Dataset: CHASEDB1	PSNR	SSIM
GC	13.7523	0.7684
AHE	9.9658	0.4869
CLAHE	20.8603	0.6178
RETINEX	14.199	0.6186
Existing Model	23.7442	0.9623
Proposed model	44.65	1

TABLE. IV. EVALUATION RESULTS OF FULL-REFERENCE METRICS (WAN ET AL., 2022)

Metrics	CLAHE	Fusion-based	MSRCP	LIME	CycleGAN	Cycle-CBAM	Proposed model
PSNR	19.0088	17.5748	11.5473	21.2084	22.4953	24.7386	44.65
SSIM	0.592	0.7693	0.6452	0.7762	0.7643	0.8103	1

V. DISCUSSIONS

In the preceding research, a DL framework incorporating a multi-stage and MSDC_NET is being used to enhance the thin-vessel segmentation in the presence of low contrast. Assessments on various datasets show that MSDC_NET is better at capturing fine blood vessels in fundus images, and assessments using the advanced reference data confirm the benefits of the proposed DL model. In comparison to the old multi-branch technique, the specificity and F1 score show enhancements of approximately three datasets (H. Guo, Meng, Zhao, Zhang, & Dai, 2024). A 19-layer U-Net architecture using deep learning is recommended for precise and effective blood vessel segmentation in the study. The architecture was tested on databases with performance metrics, and the ROC was calculated. Accuracy rates are noted to be 90.60%, 87.60%, and 83.42%. The ROC plot shows that databases have areas under the curve of 98.54%, 93.28%, and 88.18% (Prajna & Nath, 2022).

The AA-WGAN proposed is capable of effectively managing the imperfections in the data, as it can capture the pixel dependency in the entire image to emphasize regions of interest with the use of attention augmented convolution. The proposed AA-WGAN model is thoroughly assessed on three datasets, demonstrating competitive performance in vessel segmentation against various advanced models with accuracy scores of 96.51%, 97.19%, and 96.94% on each dataset (M. Liu et al., 2023). The existing MSCNN-AM focuses more on retinal vessel pixels rather than background pixels. MSCNN-AM is trained and evaluated for avoiding additional pre-processing and post-processing steps. The method that is represented is tested on three datasets available to the public. Furthermore, six objective metrics were utilised to validate the effectiveness of the MSCNN-AM by evaluating its performance metrics and AUC-PR. Experimental results show that the approach performs better than the majority of the current methods and procedures with sensitivity values and accuracy values (Fu, Li, & Wang, 2020).

The proposed network comprises approximately 1.8 million trainable parameters. Since the architecture is fully convolutional and operates in a single forward pass without iterative refinement, it remains computationally lightweight. Although explicit FLOP and memory profiling were not performed in the current implementation, the moderate model depth and parameter count indicate suitability for GPU-based clinical deployment.

VI. CONCLUSION

This document outlines a strong framework for medical image reconstruction that uses a self-evolving encoder-decoder structure with integrated adaptive convolutional power scaling. The proposed method improves structural integrity and effectively reduces artifacts caused by degradation through feature-selective enhancement and dynamic modulation of features. Experimental tests on retinal datasets show better reconstruction abilities when dealing with simulated noise and lower resolution. Although the framework demonstrates considerable reliability, further validation with datasets from multiple centers and in different imaging settings is needed to fully confirm its clinical usefulness and dependability.

DECLARATION

Conflict of Interest: The author reports that there is no conflict of interest

FUNDING

This research received no external funding.

ACKNOWLEDGEMENT

None

DATA AVAILABILITY

Data sharing is not applicable to this article as no datasets were generated.

ETHICAL STATEMENT FOR HUMAN PARTICIPANT

Not applicable for this research

AUTHORS' CONTRIBUTIONS

Author 1 (Corresponding Author): Conceptualization, Methodology, Investigation, Formal analysis, Software, Validation, Writing – original draft.

Author 2: Data curation, Resources, Investigation, Formal analysis, Writing – review and editing.

Author 3: Methodology, Software, Validation, Formal analysis, Writing – review and editing.

Author 4: Resources, Supervision, Writing – review and editing.

REFERENCES

[1] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281.

[2] Pham, C. H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., & Rousseau, F. (2017). Brain MRI super-resolution using deep 3D convolutional networks. Medical Image Analysis, 42, 197–214. https://doi.org/10.1016/j.media.2017.07.007.

[3] Shi, F., Cheng, J., Wang, L., Yap, P. T., & Shen, D. (2015). LRTV: MR image super-resolution with low-rank and total variation regularizations. IEEE Transactions on Medical Imaging, 34(12), 2459–2466. https://doi.org/10.1109/TMI.2015.2437894.

[4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Proceedings of MICCAI, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[5] Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., & Krishnan, D. (2017). Unsupervised pixel-level domain adaptation with generative adversarial networks. Proceedings of CVPR, 3722–3731. https://doi.org/10.1109/CVPR.2017.387.

[6] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/TIP.2003.819861.

[7] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of CVPR, 586–595. https://doi.org/10.1109/CVPR.2018.00068.

[8] Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 23(4), 501–509. https://doi.org/10.1109/TMI.2004.825627.

[9] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Transactions on Biomedical Engineering, 59(9), 2538–2548. https://doi.org/10.1109/TBME.2012.2205687.

[10] Zhou, Z. (2025). Multi-Scale Convolution-Aided Transformer-Based Medical Image Super-Resolution. Applied and Computational Engineering, 161, 30–40. https://doi.org/10.54254/2755-2721/2025.KL24798.

[11] Dhanusha, P. B., Muthukumar, A., & Lakshmi, A. (2022, July). Super-resolution of Retinal Fundus Images Using Generative Adversarial Networks. In 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS) (pp. 1-4). IEEE. DOI:10.1109/ICNGIS54955.2022.10079882.

[12] Chen, J., Wu, F., & Wang, W. (2024). Joint MR image reconstruction and super-resolution via mutual co-attention network. Journal of Computational Design and Engineering, 11(1), 288–304. https://doi.org/10.1093/jcde/qwae006.

[13] Qin, J., Xiong, J. & Liang, Z. CNN–Transformer gated fusion network for medical image super-resolution. Sci Rep 15, 15338 (2025). https://doi.org/10.1038/s41598-025-00119-x.

[14] Chang, Z., Shang, J., Fan, Y., Huang, P., Hu, Z., Zhang, K., & Yan, H. (2025). Deep learning-based super-resolution method for projection image compression in radiotherapy. Quantitative Imaging in Medicine and Surgery, 15(9), 8611. https://doi.org/10.21037/qims-2024-2962.

[15] 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.

[16] Long, Y., Ruan, H., Zhao, H., Liu, Y., Zhu, L., Zhang, C., & Zhu, X. (2024). Adaptive Dynamic Shuffle Convolutional Parallel Network for Image Super-Resolution. Electronics, 13(23), 4613. https://doi.org/10.3390/electronics13234613.

[17] Xiao, H., Yang, Z., Liu, T., Liu, S., Huang, X., & Dai, J. (2025). Deep learning for medical imaging super-resolution: A comprehensive review. Neurocomputing, 630, 129667. https://doi.org/10.1016/j.neucom.2025.129667.

[18] Zhang, B., Huo, J., Zhang, Z., Wang, W., Gao, H., Gong, X., & Wang, W. (2025). Nexus-INR: Diverse knowledge-guided arbitrary-scale multimodal medical image super-resolution. arXiv. https://arxiv.org/abs/2508.03073.

[19] Dubey, V. (2024). Temporal and spatial super-resolution with latent diffusion model in medical MRI images. arXiv. https://arxiv.org/abs/2410.23898.

[20] Kui, X., Ji, Z., Zou, B., Li, Y., Dai, Y., Chen, L., Vera, P., & Ruan, S. (2025). Iterative collaboration network guided by reconstruction prior for medical image super-resolution. arXiv. https://arxiv.org/abs/2504.16958.

[21] Dhanusha, P. B., Muthukumar, A., & Lakshmi, A. (2025). Deep Feature Blend Attention: A New Frontier in Super Resolution Image Generation. Neurocomputing, 618, 128989. https://doi.org/10.1016/j.neucom.2024.128989.

[22] Jannat, S. R., Lynch, K., Fotouhi, M., Cen, S., Choupan, J., Sheikh-Bahaei, N., Pandey, G., & Varghese, B. A. (2025). Advancing 1.5T MR imaging: Toward achieving 3T quality through deep learning super-resolution techniques. Frontiers in Human Neuroscience, 19, 1532395. https://doi.org/10.3389/fnhum.2025.1532395.

[23] Zhou, Z. (2025). Multi-scale convolution-aided transformer-based medical image super-resolution. Applied and Computational Engineering, 161, 30–40. https://doi.org/10.54254/2755-2721/2025.KL24798.

[24] Jin, Y., Qin, J., Zhang, Y., & Li, M. (2025). CNN–Transformer gated fusion network for medical image super-resolution. Scientific Reports, 15, 119. https://doi.org/10.1038/s41598-025-00119-x.

[25] Kobayashi, Y., Tanabe, Y., Morikawa, T., Yoshida, K., Ohara, K., Hosokawa, T., Kouchi, T., Nakano, S., Yamaguchi, O., & Kido, T. (2026). Super-resolution deep learning reconstruction improves image quality of dynamic myocardial computed tomography perfusion imaging. Tomography, 12(1), 7. https://doi.org/10.3390/tomography12010007.

[26] Yang, Z., Xiao, H., Wang, X., Zhou, F., Deng, T., & Liu, S. (2025). Cross-fusion adaptive feature enhancement transformer: efficient high-frequency integration and sparse attention enhancement for brain MRI super-resolution. Computer Methods and Programs in Biomedicine, 268, 108815. https://doi.org/10.1016/j.cmpb.2025.108815

[27] Kui, X., Ji, Z., Zou, B., Li, Y., Dai, Y., Chen, L., Vera, P., & Ruan, S. (2025). Iterative collaboration network guided by reconstruction prior for medical image super-resolution. arXiv Preprint. https://arxiv.org/abs/2504.16958.

[28] Huang, W., Liao, X., Cao, W., Jia, W., & Si, W. (2025). Versatile and efficient medical image super-resolution via frequency-gated Mamba. arXiv Preprint. https://arxiv.org/abs/2510.27296.

[29] Xin, W., Wu, Z., Zhu, Q., Bi, T., Li, B., & Tian, C. (2025). Dynamic snake convolution neural network for enhanced image super-resolution. Mathematics, 13(15), 2457. https://doi.org/10.3390/math13152457.

[30] Long, Y., Ruan, H., Zhao, H., Liu, Y., Zhu, L., Zhang, C., & Zhu, X. (2024). Adaptive dynamic shuffle convolutional parallel network for image super-resolution. Electronics, 13(23), 4613. https://doi.org/10.3390/electronics13234613.

[31] Naz, S., & Shreekanth, T. (2021). EFPT-OIDS: Evaluation Framework for a Pre-processing Techniques of Automatic Optho-Imaging Diagnosis and Detection System. International Journal of Advanced Computer Science and Applications, 12(11). https://doi.org/10.14569/IJACSA.2021.0121151.

[32] Wan, C., Zhou, X., You, Q., Sun, J., Shen, J., Zhu, S., Jiang, Q., & Yang, W. (2022). Retinal image enhancement using cycle-constraint adversarial network. Frontiers in Medicine, 8. https://doi.org/10.3389/fmed.2021.793726.