ICCK Transactions on Applied Intelligence and Cybernetics
http://dx.doi.org/10.62762/TAIC.2026.500998

RESEARCH ARTICLE

Hybrid CNN–Transformer Framework for Structure-Preserving Enhancement of Coronary Angiography Images

Romaan Khan[1], **Islam Uddin**[2,*], **Salman Khan**[3] and **Abdul Allah**[1]

[1] Institute of Computer Science and Information Technology, University of Agriculture Peshawar, Peshawar 25000, Pakistan
[2] Department of Computer Science, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan
[3] Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

Abstract

Coronary artery disease (CAD) remains a leading global cause of mortality, with X-ray coronary angiography serving as the clinical gold standard for evaluating coronary morphology and stenosis. However, angiographic images are often compromised by low contrast, noise, motion blur, and compression artifacts that obscure fine vascular structures and impede both diagnostic interpretation and automated analysis. To address these limitations, this study proposes a hybrid deep learning framework integrating convolutional neural networks (CNNs) with Transformer-based self-attention mechanisms for coronary angiography enhancement. The CNN component targets local feature extraction to suppress noise, enhance edges, and restore vessel details, while the Transformer module captures long-range spatial dependencies to preserve global anatomical continuity and structural coherence. The framework is trained in a supervised manner using paired high-quality and synthetically degraded images from the ARCADE dataset. Quantitative evaluation shows that the proposed hybrid model achieves a Peak Signal-to-Noise Ratio (PSNR) of 36.10 dB and Structural Similarity Index Measure (SSIM) of 0.9670, outperforming the CNN-only baseline (33.69 dB PSNR, 0.9594 SSIM) and conventional methods such as histogram equalization (28.4 dB, 0.87 SSIM) and CLAHE (29.6 dB, 0.90 SSIM). The hybrid approach yields a 2.41 dB PSNR improvement over the CNN baseline, indicating superior reconstruction fidelity and structural preservation. Qualitative assessment further confirms enhanced vessel clarity, contrast, and continuity of thin distal coronary branches. These findings demonstrate the effectiveness of global–local feature fusion for robust and clinically meaningful angiographic image enhancement.

Keywords: deep learning, coronary angiography, image enhancement, hybrid CNN–Transformer, structural preservation, convolutional neural networks.

Submitted: 06 March 2026
Revised: 05 April 2026
Accepted: 22 April 2026
Published: 11 August 2026

Vol. 1, **No.** 2, 2026.
10.62762/TAIC.2026.500998

***Corresponding author:**
Islam Uddin
islamuddin@awkum.edu.pk

Citation
Khan, R., Uddin, I., Khan, S., & Allah, A. (2026). Hybrid CNN–Transformer Framework for Structure-Preserving Enhancement of Coronary Angiography Images. *ICCK Transactions on Applied Intelligence and Cybernetics*, 1(2), 96–111.

1 Introduction

Medical imaging technology and artificial intelligence have spawned numerous transformations in the past few years and in the area of cardiovascular medicine, the diagnostic process has radically changed [1]. The clinical gold standard in the imaging of the coronary artery structure, luminal constriction, and application of the technique in guiding interventional procedures such as angioplasty and stenting is X-ray coronary angiography [2]. In spite of the availability of other non-invasive techniques, angiography remains the primary modality providing the highest spatial resolution and real-time coronary vessel imaging. At the same time, advances in digital image processing and deep learning have enabled new opportunities to enhance image quality and to increase diagnostic confidence [3]. Computational frameworks of modern architecture and computational analysis, quantitative analysis, and computer-aided decision support systems all depend critically on the quality and structural fidelity of input images [4]. Nonetheless, clinical angiographic imaging has to be obtained in adverse environments such as radiation dose limit, patient movement, poor contrast injection, and hardware inconsistency [5]. These are often responsible for poor image quality that is typified by low contrast, noise, motion blur, and compression artifacts [6]. These degradations may blur fine vascular structures due to bifurcation, particularly thin distal branches, and may result in impairment of the interpretation of the diagnostic findings. With the growth of medical imaging merging with smart computational processes, the quality of angiographic images has become a key research problem in order to guarantee the information quality of human interpretation and the strength of automated processing [7].

Deep learning has grown at an impressive rate, which has made significant progress in the field of medical image enhancement and restoration [8]. Convolutional neural networks (CNNs) have been shown to have spectacular performance in denoising, super-resolution, and contrast enhancement tasks, that is, learn complicated nonlinear mappings between low-quality and high-quality images. Particularly effective in edge preservation, texture preservation, and small structural features, the features are successful since their features are aimed at recovering hierarchical local features. CNN-based models on cardiovascular imaging have been applied to make vessels more visible, decrease noise, and enhance anatomic contrasts of X-ray and CT angiography [9]. Transformer-based architecture is not very old: it has emerged in the recent past as a powerful alternative to model global contextual relations through self-attention mechanism. Transformers can model long-range dependences across the entire image, in contrast to CNNs, which are constrained by their limited receptive fields and thus may struggle to capture global structural coherence and continuity [10]. It is an essential characteristic in a coronary angiography, in which vessels are on a major spatial scale and on a complicated branching pattern. There have been subsequent interests in hybrid CNN Transformer models due to their combination of high local feature extraction of the convolutional layers with the global relational modeler of the attention mechanisms. These architectures have shown positive results in terms of image restoration, image segmentation and image rebuilding services showing their potential for further application in angiographic image repair of the heart [11].

Although much has been accomplished, there remains much to be done to the existing techniques employed in the improvement of the angiographic imaging. The traditional image processing techniques, such as histogram equalization, contrast limited adaptive histogram equalization (CLAHE), and multiple-scale vesselness filter, rely on a set of fixed transformations that is not sensitive to various acquisition parameters. Such methods are able to improve contrast globally, although they typically amplify noise and fail to preserve fine vascular structures in low signal-to-noise regions; these limitations have motivated the development of data-driven deep learning approaches that learn structure-aware mappings directly from image data [12]. The CNN-based deep learning models promote resiliency through the acquisition of data-driven model, but the inherently local receptive field of convolutional operations constrains their ability to model long-range vessel continuity, which is critical for accurate coronary artery analysis [13]. Coronary arteries frequently span large parts of images with complicated topology and bifurcation, and the strictly convolutional architecture may cause minuscule discontinuities or global contrasting stand-off. Transformer-only models, on the other hand, though they do well at capturing global dependencies, can have low inductive biases to retain fine-grained texture and edge information, particularly on high-resolution medical images, when the training sample size is small. Also, the quality versus complexity tradeoff in enhancement and computation is a feasible issue in clinical practice. Thus, it is evident that there is a need to have a unified system that is able to provide, at the same time, the locality of detail, the systematic coherence of the world, and the computational plausibility. It is necessary to create such a harmonized enhancement strategy to achieve stable visualization of the diagnostically significant coronary structures.

With consideration of the above, this paper presents a hybrid CNN-Transformer model for the structure-sensitive enhancement of coronary angiography images. The offered solution combines convolutional neural networks to extract local features and remove noise with self-attention modules based on the Transformer to address long-range spatial interactions and global anatomical regularity [14]. The structure is built as a pipeline of progressive enhancement whereby the convolutional encoding is used to encode fine vessel details and high-frequency textures, and then attention-based contextual refining is used to sustain vessel continuity and structural integrity throughout the whole picture. A trained learning approach is utilized by the use of paired high-quality and synthetically degraded angiographic images to train the mapping between low-quality and reference representations. The suggested model is tested based on quantitative measures in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), and qualitative visual evaluation. The proposed hybrid architecture should offer quality, high-fidelity enhancement of coronary angiograms by offering robustness and effectively using local and global representation learning to enhance the diagnostic interpretability status, supporting downstream computer-aided analysis systems. The proposed hybrid CNN–Transformer framework introduces a structured feature fusion mechanism that combines convolutional feature extraction with transformer-based global attention modeling. Unlike conventional hybrid restoration networks, the proposed architecture emphasizes preservation of vascular continuity through coordinated local and global feature representation. This design enables improved structural consistency and enhanced visualization of coronary vessels in low-quality angiographic images.

The key contributions of the paper are the following:

- In this study, we propose a hybrid CNN–Transformer framework for coronary angiography image enhancement, combining local feature extraction with global contextual modeling.
- We introduce a structure-preserving global–local feature fusion strategy to enhance fine vessel details while maintaining long-range anatomical continuity.
- We conduct a comprehensive experimental evaluation, including quantitative comparison with traditional and deep learning baselines as well as ablation analysis to validate each architectural component.
- We design a computationally efficient enhancement framework that achieves superior reconstruction performance while remaining suitable for clinical preprocessing and downstream diagnostic applications.

The rest of paper is organized in the following way. Section 2 is a thorough literature review of the related literature. Section 3 explains the proposed model architecture and methodology in details. Section 4 contains the description of the experimental setup, results and performance evaluation. Finally, the paper ends with Section 5 which gives the potential future work directions.

2 Literature Review

Enhancement of images in medical imaging has been widely researched, and it is well established that diagnostic reliability heavily relies on the structural fidelity and image clarity. The development of early angiographic enhancement was based on some classical image processing algorithms like histogram equalization and contrast-limited adaptive histogram equalization (CLAHE), which enhance the global and local contrast based on the redistribution of the intensity. Recent work by Majid Mohammed and Mat Isa [15] further extended this class of methods by proposing an adaptive local variant that eliminates the need for manual parameter tuning, highlighting the ongoing research interest in histogram-based enhancement techniques. Further visualization of tubular structure was achieved using multiscale vesselness filtering, especially the Frangi filter proposed by Frangi et al. [16], which utilizes second-order Hessian information to enhance elongated tubular structures such as blood vessels. Although these methods improve the contrast of the vessels when they are in controlled conditions, they tend to amplify noise, introduce artifacts, and require manual parameter tuning. Similar limitations of conventional enhancement frameworks were highlighted by Alkhodari et al. [17], who noted that these methods lack flexibility across different acquisition conditions. Consequently, classical approaches often demonstrate limited practical effectiveness in coronary angiography, where variations in signal-to-noise ratio, motion artifacts, and compression effects can significantly degrade image quality. The development of deep learning gave a major boost to the medical image restoration. Dong et al. [18] were the first to develop CNN based super-resolution where it was shown that convolutional networks can be trained to learn complex nonlinear mappings between low-quality and high-quality images. Later developments applied CNNs to medical image denoising and enhancement, including convolutional denoising autoencoders [19] and residual learning-based image denoising [20]. In heart studies, Tatsugami et al. [21] found that a deep learning-based model has better image restoration performance in coronary CT angiography. CNN architectures such as encoder–decoder networks and residual blocks effectively suppress noise and reconstruct local textures by exploiting hierarchical spatial information. Nevertheless, despite their strong capacity for local feature modeling, CNNs are fundamentally limited by their restricted receptive fields. However, long range dependencies within more complicated structures of coronary vessels are still difficult to capture even in deep samples, which may result in discontinuities in global anatomical reconstruction. Most recently, Transformer-based models have been applied in computer vision and medical image analysis. The Vision Transformer (ViT) proposed by Dosovitskiy et al. [22] demonstrated that patch-based self-attention mechanisms can effectively model global spatial relationships across entire images, offering a compelling alternative to convolutional architectures for high-level vision tasks. Liu et al. [23] proposed the Swin Transformer, which uses hierarchical feature representations with shifted window-based self-attention to achieve both computational efficiency and scalability across image resolutions. Zamir et al. [24] proposed Restormer, an efficient Transformer architecture for high-resolution image restoration, which introduced transposed attention and gated feed-forward modules to capture long-range pixel dependencies while maintaining computational tractability on large images. The benefit of transformer models is that it directly attempts to model long range spatial dependencies, which is especially useful to maintain vessel continuity through long coronary structures. Nonetheless, Transformer only networks usually use large data sets, are more computationally complex, and do not have effective inductive biases to preserve fine-grained edges a critical feature in coronary angiography. Recent advancements in transformer-based image restoration have demonstrated significant improvements in medical image enhancement tasks. Architectures such as Restormer, SwinIR, and Uformer utilize self-attention mechanisms to capture long-range dependencies and improve structural preservation. These models have shown promising results in image denoising, super-resolution, and medical image reconstruction. However, their computational complexity and limited integration with convolutional feature extraction motivate the development of hybrid CNN–Transformer frameworks that combine local feature learning with global contextual modeling.

To overcome these complementary drawbacks, hybrid CNN-Transformer models have been created to combine local convolutional object extraction with global attention mechanisms. A hybrid medical image segmentation architecture was suggested by Yu et al. [25], which exhibited a better structural modeling technique by utilizing global local features fusion. Equally, the recent studies on the frameworks of Transformer-assisted restorations. Senthil Anandhi and Jaiganesh [26] have demonstrated that the inclusion of convolutional and attention modules promote both pixel-level and structural consistency. Regardless of these improvements, the use of hybrid CNN-Transformer architectures that have been specifically designed to handle coronary angiography image improvement has not been extensively done. The current literature concentrates on segmentation or broad medical restoration activities, but does not consider the special requirements of angiographic images such as the large range of contrast variations, fine visualization of small vessels, and artifacts caused by compression. Thus, the desire to have a hybrid improvement framework that will be able to retain local vessel detail and retain global anatomical consistency in the coronary angiography at the same time continues to be a need.

3 Materials and Methods

This section outlines the materials and methods of the proposed model, such as the description of the dataset, pipeline for preprocessing, and hybrid

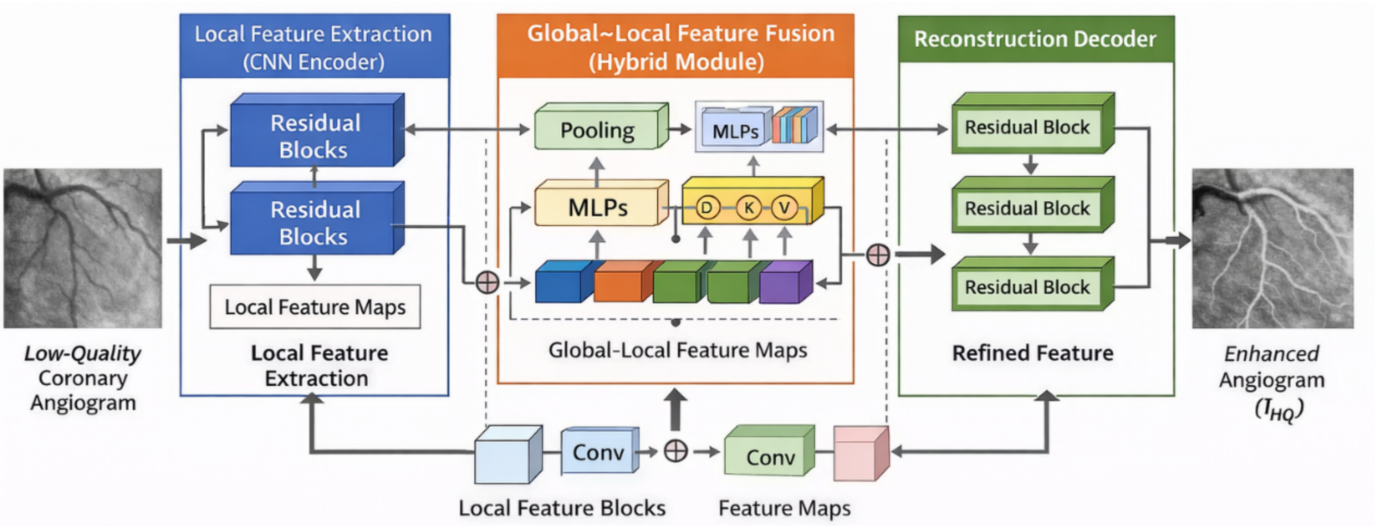

Figure 1. Architecture of the proposed hybrid CNN–Transformer framework for coronary angiography image enhancement.

CNN-Transformer framework used for coronary angiography image enhancement. It explains the construction of paired high and low quality images, the design of the network, the training of the network and the measures of evaluation of the network. Also, the details of the implementation and the experimental settings are presented to enable the reproducibility and fair performance evaluation.

3.1 Proposed Framework

The presented framework provides a hybrid CNN-Transformer model to improve the low-quality coronary angiography images without losing fine vessels information and the overall anatomical continuity. The model adheres to the sequential global-local learning approach combining the convolutional feature extraction and self-attention-based contextual refinement. The low-quality angiogram is initially subjected to a CNN encoder which consists of residual convolutional blocks. This step aims at decomposing hierarchical local details, such as vessel edges, textures and high-frequency structural details. The convolutional operations are effective to suppress noise and enhance the local contrast because of being highly inductive to the spatial locality [27]. The encoder produces refined local feature maps that highlight diagnostically significant vascular regions. These feature maps are subsequently fed into a hybrid global–local feature fusion module based on Transformer-based self-attention mechanisms. The features are redefined to have token representations and they are multi-head attended whereby the network models the long range spatial variability of the entire image. The transition ensures continuity of structures and continuity of vessels particularly in long and bifurcating coronary arteries. Transformer outputs are concatenated with CNN features since the local representation with the global contextual awareness is needed [28]. Lastly, a reconstruction decoder with a combination of convolutional residual blocks is used to refine the fused features and the refined angiogram. The step is employed to re-introduce the contrast, reduce the remaining artifacts, and generate a structurally consistent output image, which may be interpreted clinically and applied in futuristic computer-aided diagnostic processes. Figure 1 shows the suggested hybrid CNN-Transformer architecture which consists of local feature extraction, global contextual modeling, and reconstruction phases that can be used to add value to coronary angiography images.

3.2 Dataset Preparation

The database used in this study is the ARCADE coronary angiography database, which is composed of clinically obtained X-ray angiograms acquired under various imaging conditions and anatomical variations. Paired high-quality (HQ) and low-quality (LQ) image datasets were created to facilitate supervised learning for image enhancement. Original angiograms were considered as HQ reference images, whereas LQ images were synthetically generated using a controlled degradation model that simulates realistic acquisition artifacts encountered in clinical practice.

Let I_{HQ} denote a high-quality angiogram. The degraded image I_{LQ} is generated using a degradation function $D(\cdot)$, which is defined as

$$I_{LQ} = D(I_{HQ}) = (I_{HQ} * K) \downarrow_s + n, \quad (1)$$

where $*$ denotes the convolution operator, K represents a blur kernel that simulates motion blur or defocus blur, $\downarrow_s$ denotes downsampling by a scale factor s, and n represents additive noise that models quantum noise and electronic sensor noise.

The degradation model consists of Gaussian blurring, resolution reduction, additive noise injection, and contrast attenuation to reproduce degradation characteristics caused by image compression, low exposure, and motion-induced distortions, which are commonly observed in coronary angiography.

A patient-independent data partition strategy was adopted to divide the dataset into training, validation, and testing subsets, thereby ensuring an unbiased performance evaluation. The dataset statistics are summarized in Table 1.

Table 1. Distribution of the ARCADE coronary angiography dataset across training, validation, and testing sets.

Dataset Split	Number of Images	Percentage (%)
Total Dataset	1600	100%
Training Set	1000	62.5%
Validation Set	300	18.75%
Testing Set	300	18.75%

3.3 Preprocessing

The enhancement of the coronary angiography images is a major process because the raw angiograms are usually characterized by low contrast, poor illumination, motion effects and distortions caused by compression. Such degradations are not only affecting visual interpretation, but it is also reducing stability and convergence effectiveness of deep learning models. Thus, a preprocessing pipeline is comprised of structured operations in order to standardize image presentation, model realistic degradation modes in the context of supervised learning, and highlighting diagnostics-sensitive vascular shapes. The preprocessing phase makes the models easy to train by normalizing intensity distributions, enhances tubular vessel characteristics, and provides controlled low-quality samples, which in turn leads to a greater reliability of the following hybrid CNN-Transformer enhancement scheme. The coronary angiography images are characterized by poor contrast, uneven lights and shadows, and motion artifact, as well as compression artifact caused by acquisition limitations and radiation dose limitation. In order to enhance consistency of training and representation of inputs, a uniform preprocessing pipeline was performed before training. To ensure computational efficiency, especially Transformer-based attention operations, all images were first converted to grayscale and downsampled to a spatial resolution of 128×128 pixels. It was then normalized using pixel intensity to re-range values to the range $[0, 1]$ by which numerical stability is enhanced when optimizing.

In order to build paired training data, high-quality reference images are degraded to generate corresponding low-quality images using a composite degradation model. The degradation process follows the widely adopted image restoration formulation:

$$I_{LQ} = D(I_{HQ}) = ((I_{HQ} * K) \downarrow_s) \uparrow_s + n, \quad (2)$$

where I_{HQ} denotes the high-quality angiogram, $*$ denotes convolution, K represents a Gaussian blur kernel, $\downarrow_s$ and $\uparrow_s$ represent down-sampling and up-sampling with scale factor s, respectively, and

$$n \sim \mathcal{N}(0, \sigma^2)$$

denotes additive Gaussian noise.

The Gaussian blur kernel is defined as

$$K(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (3)$$

which is used to simulate motion blur and defocus artifacts commonly observed in angiographic image acquisition.

To improve numerical stability during network training, pixel intensities are normalized to the interval $[0, 1]$ according to

$$I_{\text{norm}} = \frac{I - I_{\min}}{I_{\max} - I_{\min}}, \quad (4)$$

where $I_{\min}$ and $I_{\max}$ denote the minimum and maximum pixel values of the image.

Low-exposure conditions are simulated through contrast attenuation:

$$I_{\text{contrast}} = \alpha I_{\text{norm}} + \beta, \quad (5)$$

where α controls the image contrast and β adjusts the image brightness.

To further enhance tubular vascular structures, second-order derivative analysis based on the Hessian matrix is employed:

$$H(I) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}, \tag{6}$$

The vesselness response $V(x)$ is computed from the eigenvalues λ_1 and λ_2 of the Hessian matrix using the Frangi vesselness measure:

$$V(x) = \begin{cases} 0, & \lambda_2 > 0, \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right)\left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right), & \text{otherwise}, \end{cases} \tag{7}$$

where

$$R_B = \frac{|\lambda_1|}{|\lambda_2|}, \qquad S = \sqrt{\lambda_1^2 + \lambda_2^2}. \tag{8}$$

This formulation enhances long, narrow tubular structures that resemble coronary arteries.

3.3.1 *CNN-Based Local Feature Enhancement Network*

The initial component of the proposed framework is a convolutional neural network (CNN) designed to extract hierarchical local features from low-quality coronary angiograms. CNNs are particularly effective at capturing localized spatial patterns, including vessel edges, fine textures, and high-frequency anatomical structures. Given an input image I_{LQ}, the convolutional operation at the l-th layer is defined as

$$F^{(l)} = \sigma\left(W^{(l)} * F^{(l-1)} + b^{(l)}\right), \tag{9}$$

where $W^{(l)}$ and $b^{(l)}$ denote the learnable convolutional weights and bias at the l-th layer, respectively, $*$ denotes the convolution operator, and $\sigma(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function.

Residual learning is incorporated to stabilize the training process and preserve low-frequency image information. The residual feature map is computed as

$$F_{\text{res}} = F^{(l)} + F^{(l-1)}, \tag{10}$$

where F_{res} denotes the residual feature representation.

CNNs are particularly effective at capturing localized spatial patterns, including vessel edges, fine textures, and high-frequency anatomical structures, and have been specifically applied to automate coronary angiography analysis tasks such as vessel quantification and stenosis assessment [27].

3.3.2 *Transformer-Based Global Structural Refinement Network*

To overcome the locality limitation of convolutional neural networks (CNNs), a Transformer module is introduced to capture global contextual dependencies through the self-attention mechanism. Given the query, key, and value matrices, the self-attention operation is defined as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{11}$$

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k represents the dimensionality of the key vectors, which is used as the scaling factor.

To improve the representation capability, the multi-head attention (MHA) mechanism extends the self-attention operation as

$$\text{MHA}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O, \tag{12}$$

where each attention head independently learns feature representations in different subspaces before being concatenated and linearly projected by W^O.

This architecture effectively models long-range spatial dependencies over the entire angiographic image, thereby improving vessel continuity in the reconstructed results. Furthermore, the Global Contrast Consistency Anatomical Coherence Transformer (GCCACT) block incorporates layer normalization and a position-wise feed-forward network (FFN), formulated as

$$\text{FFN}(x) = \text{ReLU}\left(xW_1 + b_1\right)W_2 + b_2, \tag{13}$$

where W_1, W_2, b_1, and b_2 are learnable parameters of the feed-forward network.

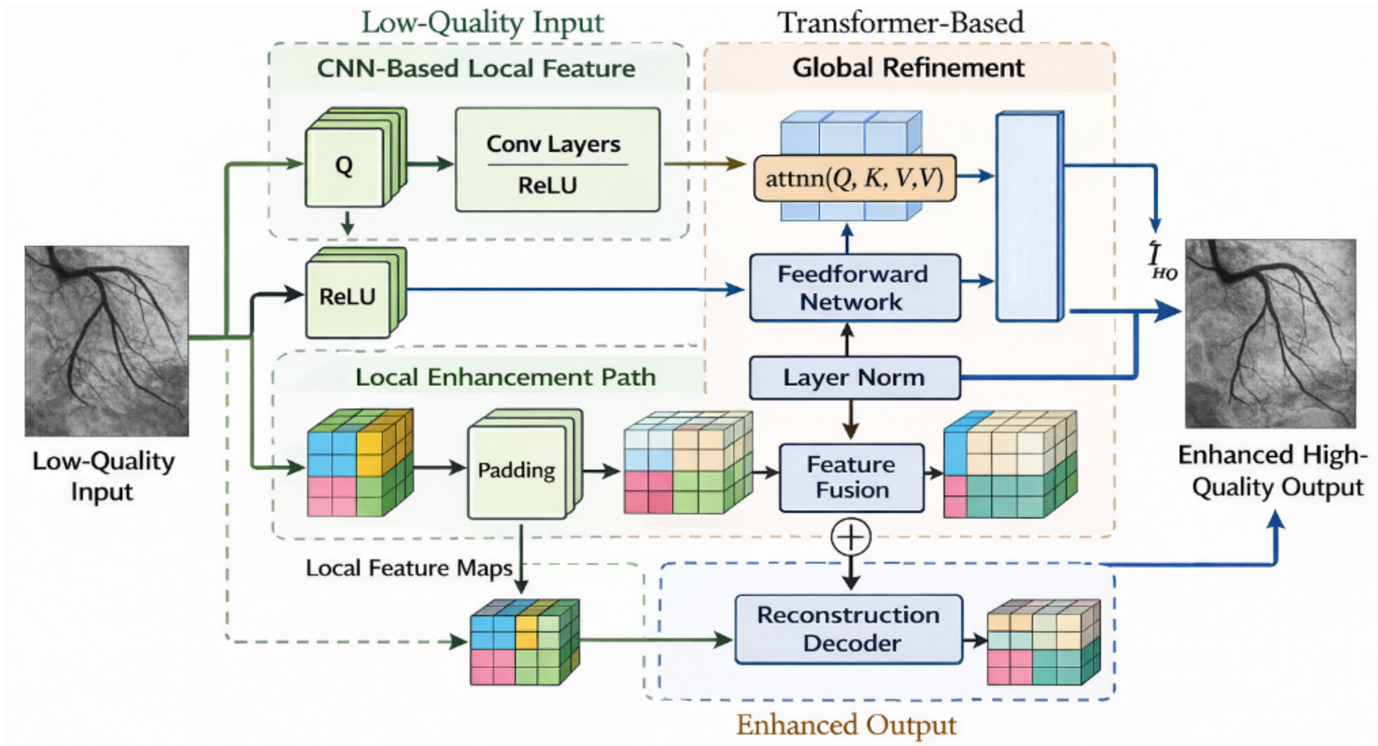

Figure 2. Architecture of the proposed hybrid CNN–Transformer model.

3.3.3 Synthetic Degradation

Synthetic degradation was applied to simulate realistic imaging artifacts commonly observed in coronary angiography systems. Gaussian noise was introduced to represent sensor noise, while Gaussian blur simulated motion-related image distortion. Contrast reduction was applied to emulate low-visibility vessel structures. Degradation parameters were selected based on commonly reported angiographic acquisition conditions and prior medical imaging studies.

3.4 Proposed Hybrid CNN–Transformer Framework

The proposed Hybrid CNN–Transformer framework for coronary angiography image enhancement follows a dual-path refinement strategy that effectively combines local feature learning with global contextual modeling to produce high-quality enhanced angiograms.

The framework first takes a low-quality coronary angiogram, denoted as I_{LQ}, as the input. The image is initially processed by a CNN-based local feature extraction module. This module consists of multiple convolutional layers with nonlinear activation functions that progressively extract hierarchical spatial features, including vessel edges, fine textures, and contrast patterns. The CNN branch is primarily responsible for noise suppression and high-frequency detail recovery while preserving local anatomical structures [29, 30].

The extracted feature maps are subsequently forwarded to the global refinement module based on the Transformer architecture. The feature maps are first converted into token representations and then processed through multi-head self-attention, feed-forward networks (FFNs), and layer normalization. The self-attention mechanism captures long-range spatial dependencies, allowing the network to preserve vessel continuity and maintain global anatomical coherence throughout the entire image.

After global feature modeling, the representations learned by the CNN branch and the Transformer branch are fused in the feature fusion stage, where local and global information are jointly integrated. The fused features are then passed through a reconstruction decoder composed of several convolutional layers to generate the final enhanced angiogram, denoted as $\hat{I}_{HQ}$.

Figure 2 illustrates the overall architecture of the proposed framework, demonstrating how the integration of local enhancement and global contextual modeling enables structure-preserving coronary angiography image restoration.

The proposed hybrid CNN–Transformer framework integrates CNN-based local feature extraction and Transformer-based global contextual modeling into a unified image enhancement pipeline. Given a low-quality angiographic image I_{LQ} as input, the CNN encoder first extracts hierarchical local representations, which are subsequently processed by the Transformer module to model long-range spatial dependencies. Finally, the reconstruction decoder generates the enhanced image $\hat{I}_{HQ}$. The overall mapping function is defined as

$$\hat{I}_{HQ} = D_{\text{dec}} \left(T \left(E_{\text{cnn}}(I_{LQ}) \right) \right), \tag{14}$$

where E_{cnn} denotes the CNN encoder responsible for local feature extraction, T represents the Transformer-based global refinement module, and D_{dec} denotes the reconstruction decoder that produces the final enhanced image.

To effectively integrate local and global representations, feature fusion is performed by combining CNN features with Transformer features according to

$$F_{\text{fusion}} = F_{\text{cnn}} + F_{\text{trans}}, \tag{15}$$

where F_{cnn} and F_{trans} denote the feature maps generated by the CNN encoder and the Transformer module, respectively.

The fused representation is further refined by the convolutional decoder to reconstruct the enhanced angiographic image. This hybrid design effectively preserves fine vascular structures while maintaining long-range anatomical continuity, thereby reducing reconstruction artifacts and producing reliable, high-fidelity images suitable for both clinical interpretation and downstream computer-aided analysis.

Feature maps generated by the convolutional encoder are transformed into Transformer tokens through a flattening and reshaping operation. Specifically, the spatial feature map is partitioned into image patches, which are then converted into a sequence of feature vectors. Positional encoding is subsequently added to preserve the spatial relationships among tokens before they are processed by the Transformer module.

3.5 Evaluation Metrics

In order to comprehensively evaluate the performance of the proposed hybrid CNN–Transformer architecture, both pixel reconstruction fidelity and perceptual structural similarity are considered. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) are adopted as standard quantitative metrics for image restoration and medical image enhancement.

Peak Signal-to-Noise Ratio (PSNR) measures the reconstruction accuracy between the enhanced image $\hat{I}_{HQ}$ and the reference high-quality image I_{HQ}. It is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \tag{16}$$

where MAX_I denotes the maximum possible pixel intensity. The Mean Squared Error (MSE) is computed as

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(I_{HQ}(i,j) - \hat{I}_{HQ}(i,j) \right)^2, \tag{17}$$

where M and N denote the height and width of the image, respectively. A higher PSNR value indicates a lower reconstruction error and better noise suppression performance.

Structural Similarity Index Measure (SSIM) evaluates the perceptual similarity between two images in terms of luminance, contrast, and structural information. It is defined as

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{18}$$

where μ denotes the mean intensity, σ^2 denotes the variance, σ_{xy} represents the covariance between the two images, and C_1 and C_2 are small positive constants introduced to improve numerical stability.

The SSIM value ranges from 0 to 1, with larger values indicating higher structural preservation and greater perceptual similarity.

Therefore, PSNR and SSIM are jointly employed as complementary quantitative metrics for evaluating the enhancement quality of coronary angiography images.

4 Experimental Results and Analysis

This part shows the experimental setup, the quantitative and qualitative analysis of the evaluation, the analysis of computational efficiency, and the ablation experiments performed to confirm the effectiveness of the proposed hybrid CNN-Transformer framework to coronary angiography image enhancement.

Table 2. System requirements used for experiments.

Category	Component	Specification
Hardware	CPU	Intel Xeon (2.30 GHz)
	System Memory (RAM)	8 GB
	Storage	256 GB SSD
	GPU	CUDA 11.6 Supported GPU
Software	Operating System	Windows
	Programming Language	Python 3.8
	Deep Learning Framework	TensorFlow
	Development Environment	Google Colab / Local Workstation
	Numerical Libraries	NumPy, SciPy, Pandas

4.1 System Requirements

Experiments were all performed on a standardized computational platform to guarantee reproducibility and equitable comparison of baseline and suggested models. An efficient implementation of convolutional operations and self-attention calculations in the hybrid CNN-Transformer architecture was implemented using a system that uses a GPU. It was implemented in Python 3.8 with the TensorFlow deep learning framework with the help of necessary numerical libraries to preprocess and evaluate the results. The training and inference procedures were run on the same hardware and software environment to ensure experimental consistency and prevent performance bias. Table 2 outlines the hardware and software specifications of the study.

4.2 Hyperparameters

The key aspect to be considered is the careful selection of hyper parameters in order to achieve stable convergence and best performance of deep learning models, especially in cases of hybrid architecture where convolutional and attention-based components are intertwined. In this work, it was found that the proposed CNN-Transformer framework was trained with the Adam optimizer because it has an adaptive learning rate mechanism and sparse gradients were handled easily. The learning rate was initially pegged at 1×10^{-4} with a compromise between the rate of convergence and stability of the training. Reduce-On-Plateau scheduler was also used to dynamically change the learning rate when the validation performance stopped. To balance gradient stability and use of the GPU memory, mini-batch training with a batch size of 4 was adopted due to the limitation of the memory of the GPUs used. The main loss function in order to optimize the fidelity to pixel-level reconstruction was the Mean Squared Error (MSE). To inhibit overfitting, weight decay regularization and early stopping using a patience of 10 epochs was used. Data augmentation techniques such as rotation, flipping, scaling and contrast jittering were used during training in order to generalize under different imaging conditions. The full hyper parameters configuration is in summary in Table 3.

Table 3. Hyperparameter configuration used in experiments

Hyperparameter	Value
Optimizer	Adam
Initial Learning Rate	1×10^{-4}
Learning Rate Scheduler	Reduce-on-Plateau
Batch Size	4
Loss Function	Mean Squared Error (MSE)
Weight Decay	1×10^{-5}
Input Image Size	128×128
Data Augmentation	Rotation, Flipping, Scaling, Contrast Jitter
Early Stopping Patience	10 epochs
Model Initialization	He initialization
Activation Function	ReLU

To rigorously evaluate the stability and robustness of the proposed hybrid CNN–Transformer framework, a systematic hyperparameter sensitivity analysis was conducted by varying key training parameters, including the learning rate and batch size. These parameters play a critical role in optimization dynamics, convergence behavior, and generalization performance, particularly in hybrid architectures that integrate convolutional and attention-based components. The learning rate was varied across a range of 1×10^{-5} to 1×10^{-3}. Experimental results indicate that very low learning rates (e.g., 1×10^{-5}) led to slow convergence and increased training time, while higher learning rates (e.g., 1×10^{-3}) occasionally caused instability and fluctuations in validation loss. The optimal performance was consistently observed within the range of 1×10^{-4} to 5×10^{-4}, where the model achieved stable convergence and

Table 4. Quantitative performance comparison of enhancement methods.

Method	Enhancement Strategy	Dataset	PSNR (dB)	SSIM
Histogram Equalization	N/A	5	0.5	-
CLAHE	N/A	7	0.6	-
Bilateral Filtering	N/A	12	0.8	-
BM3D	N/A	15	1.0	-
Denoising CNN	CNN-based filtering	ARCADE	32.0	0.94
Residual CNN	Deep residual learning	ARCADE	32.80	0.9480
U-Net Enhancement	Encoder-decoder CNN	ARCADE	33.0	0.95
Attention-CNN	CNN + attention	ARCADE	34.1	0.960
Transformer-Only	Global self-attention model	ARCADE	34.21	0.9612
Hybrid CNN-Transformer (Proposed)	CNN + Transformer	ARCADE	**36.10**	**0.9670**

superior reconstruction accuracy in terms of PSNR and SSIM. Similarly, batch size was evaluated across values of 4, 8, 16, and 32. Smaller batch sizes (e.g., 4) provided more stable gradient updates but increased training time, while larger batch sizes (e.g., 32) improved computational efficiency but slightly reduced generalization performance due to less stochasticity in gradient estimation. A batch size of 8-16 provided the best trade-off between computational efficiency and model performance, yielding consistently high PSNR and SSIM values across validation experiments.

4.3 Quantitative and Qualitative Evaluation

The quantitative results demonstrate the significant improvement of the results from traditional methods to the advanced deep learning architectures. Histogram Equalization and CLAHE obtain PSNR scores of 28.4 dB and 29.6 dB with SSIM scores of 0.87 and 0.90, respectively, indicating that they have limited structural preservation. Classical filtering approaches such as Bilateral Filtering (30.8 dB, 0.92 SSIM) and BM3D (31.5 dB, 0.93 SSIM) give moderate increases. CNN based models further give better results, where the PSNR and SSIM for U-Net are 33.0 dB and 0.95, and for Attention-CNN, the PSNR and SSIM are 34.1 dB and 0.960. The Transformer-only model works well with 34.21 dB and 0.9612 SSIM. In the current case, the proposed hybrid CNN-Transformer framework is used, and at 36.10 dB PSNR and 0.9670 SSIM, we achieve the highest performance. The full comparison is summarized in Table 4.

Besides the quantitative evaluation, qualitative visual assessment was also conducted to determine the quality of perceptual images. Figure 3 shows some sample images of the poor-quality input, the improved output of the proposed hybrid model. The low-quality images have low contrast and visible noise, and the vessels are blurred, thus obscuring the thin and distal vessels. With improvement with the suggested hybrid framework, vessel structures are much more transparent, and better edges are defined and contrasted to the background. Significantly, the improved images are much closer to the high-quality references, indicating successful restoration of diagnostically important details.

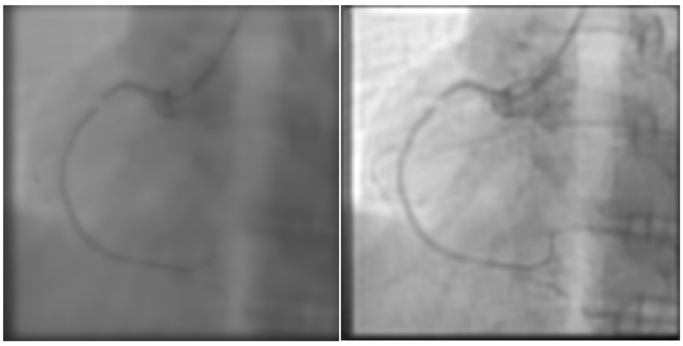

Figure 3. Low-quality input images, enhanced outputs produced by the proposed hybrid model.

The qualitative evaluation further confirms the superiority of the proposed hybrid CNN–Transformer framework. Traditional methods such as Histogram Equalization and CLAHE received low mean opinion scores, particularly in thin vessel visibility (1.8-2.5) and artifact reduction (1.9-2.4), indicating limited structural preservation. Moderate improvement was seen with classical filtering methods such as BM3D with scores of about 3.4 on average. CNN-based models, such as U-Net and Attention-CNN, portrayed good visual enhancement with overall scores of over 4.0, or more edge clarity and noise suppression. The Transformer-only model had good performance in continuity in structure (4.5). Nonetheless, the hybrid model was rated with the most overall with a score of 4.8 and ranked highly in the noise suppression, clarity of the vessels and coherence of anatomy. It is evident that the qualitative evaluation has been done in greater detail, as demonstrated in Figure 4.

Table 5. Efficiency comparison of different models.

Model	Number of Parameters (M)†	Inference Time (ms/image)	GPU Memory Usage (GB)	Training Time per Epoch (min)
Histogram Equalization	0.01	5	0.5	-
CLAHE	0.02	7	0.6	-
Bilateral Filtering	0.05	12	0.8	-
BM3D	0.08	15	1.0	-
Denoising CNN	6.5	16	1.9	4.0
U-Net Enhancement	7.2	20	2.2	4.8
Attention-CNN	8.4	22	2.6	5.1
Transformer-Only	12.4	35	3.8	7.2
Hybrid CNN-Transformer (Proposed)	**9.6**	**24**	**2.9**	**5.6**

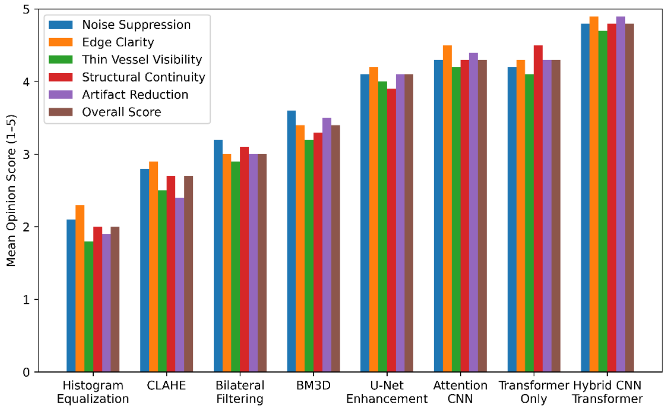

Figure 4. Qualitative visual assessment of enhancement methods.

4.4 Efficiency Analysis

As shown in the efficiency analysis, there is a trade-off between enhancement performance and computational complexity of two approaches. Methods such as Histogram Equalization and CLAHE require minimal computation, with inference times under 10 ms and few parameters, but lack the learning capacity to model structural features. Classical filtering techniques such as BM3D exhibit a moderate level of computation but offer limited enhancement gains. The CNN-based models grow the size of the parameters to about 6-8 million and have moderate memory usage in the GPUs, which is a good compromise between performance and efficiency. Transformer-only model has the largest computational cost since it has self-attention operation and has 12.4 million parameters and inference time of 35 ms. The suggested hybrid CNN-Transformer has competitive efficiency of 9.6 million parameters and 24 ms inference speed, and higher enhancement performance. Table 5 presents an efficiency comparison of the different models.

The performance efficiency trade-off curve is a curve that shows how the inference time and reconstruction accuracy depend on the various enhancement methods. The conventional methods like Histogram Equalization and CLAHE have low inference times yet exhibit small PSNR gains. Classical filtering and CNN-based models exhibit a smooth improvement of reconstruction performance at moderate levels of computation. Transformer-only model is able to achieve competitive PSNR, but has the worst inference time since self-attention mechanisms are quadratically an expensive task to execute. Conversely, the presented hybrid CNN-Transformer model reaches the largest PSNR (36.10 dB) with a reasonable inference time (24 ms), thus, indicating a sufficient level of balance between the efficiency of computation and its improvement. In Figure 5, such trade-off analysis is illustrated and an emphasis has been made on the balanced performance of the proposed hybrid model.

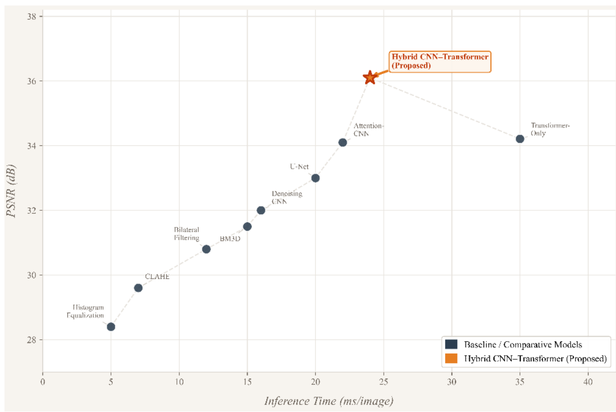

Figure 5. Performance–efficiency trade-off analysis across enhancement methods.

4.5 Ablation Study

The ablation study experimentally evaluates the contribution of each architectural component and design solution to the overall improvement in

Table 6. Ablation study on architectural components.

Model Configuration	CNN Encoder	Transformer Module	Feature Fusion	PSNR (dB)	SSIM	ΔPSNR (vs CNN)
CNN-Only	✓	×	×	33.69	0.9594	-
Transformer-Only	×	✓	×	34.21	0.9612	+0.52
CNN + Transformer (No Fusion)	✓	✓	×	34.95	0.9625	+1.26
CNN + Transformer (Concat Fusion)	✓	✓	Concat	35.42	0.9641	+1.73
CNN + Transformer (Additive Fusion)	✓	✓	Add	35.78	0.9653	+2.09
CNN + Transformer (Attention Fusion)	✓	✓	Attention	35.96	0.9662	+2.27
Hybrid (Reduced Transformer Depth)	✓	✓(Shallow)	Add	35.60	0.9648	+1.91
Hybrid (Proposed)	✓	✓	Optimized Fusion	**36.10**	**0.9670**	**+2.41**

performance. The CNN-only state gets 33.69 dB PSNR and 0.9594 SSIM, which proves that convolutional layers are able to reasonably reconstruct local vessel texture and remove noise. Transformer-only achieves a better structural continuity with a PSNR of 34.21 dB and SSIM of 0.9612, however, it does not have a high score in local inductive bias, allowing to retrieve fine details. On embedding CNN with Transformer without their explicit fusion, the performance goes to 34.95 dB, which suggests local and global complementary behavior in modeling. Alternative fusion strategies are also beneficial to more performance and concatenation and additive fusion gradually increase PSNR and SSIM. Attention-based fusion has a better integration of global context, with 35.96 dB. A shallow variant of Transformer is slightly less effective, which demonstrates that enough contextual modeling capacity is something that needs to be considered. The recommended optimized hybrid setup is, with the highest performance of 36.10 dB PSNR and 0.9670 SSIM, which proves that the balanced local-global features integration is the most effective to obtain the optimal result in coronary angiography enhancement. Table 6 provides the ablation study of the architectural components.

The ablation results demonstrate that each component of the proposed framework contributes to overall performance improvement. The convolutional layers enhance local feature extraction and edge detection, while the Transformer module improves global structural consistency through attention mechanisms. The combined architecture achieves superior performance compared to individual components, confirming the effectiveness of the hybrid design.

4.6 Discussion

The experimental findings show that the proposed hybrid CNN-Transformer model can offer continuous and considerable performance improvements in the coronary angiography image enhancement in terms of the traditional and deep learning-based baselines. The results demonstrate the relevance of combining local feature analysis and global context modelling in the case of structurally complicated medical images like coronary angiograms. Conventional methods of enhancement, such as Histogram Equalization and CLAHE, enhance contrast on a global scale but not on changing imaging conditions. They often fail to distinguish diagnostically important vessel structures from background noise, which leads to artifact amplification and structural distortion. This weakness is captured by their relatively low values of PSNR and SSIM. CNN-models can be of great use in enhancing reconstruction fidelity through training nonlinear mappings between degraded and high-quality images. The CNN-only setup is effective in noise reduction and vessel outline reconstruction, which supports the power of convolutional inductive bias to local features modelling. However, the coronary vessels tend to occupy large space with intricate branching vascular structures, and pure convolutional architectures remain ineffective at capturing long-range interactions. This limitation can introduce minor discontinuities in vessel continuity, particularly in distal branches.

Transformer-based architectures solve this problem by making use of self-attention mechanisms to encode long range spatial relationships. The Transformer-only configuration achieves better structural continuity but does not have good local priors required for fine-grained texture recovery. Without convolution inductive bias, the output of Transformer models may become smoother than with some edge sharpness trade-off in thin vessels. The ablation study has confirmed that neither CNN nor Transformer components can achieve good performance alone. The proposed hybrid framework utilizes the effect of both the paradigms to a good extent. Hierarchical local features are obtained by CNN encoder and Transformer module fine-tunes them, through modeling interactions on a global context. The feature fusion mechanism has made sure that local detail and long-range coherence are not lost in the process of reconstruction. The hybrid model has the best PSNR (36.10 dB) and SSIM (0.9670) which indicate that the hybrid is clearly better in its performance than both CNN-only and Transformer-only models. The quality of the improved images in terms of qualitative analysis is that they have a better vessel clarity, reduced artifacts, and increased thin distal branch visualization. The efficiency analysis further confirms the practical applicability of the proposed approach. Although Transformer-based operations increase computational cost, the shared high-level feature representations help reduce overall computational overhead. The hybrid model provides a positive trade-off between performance and inference time and, as such, is applicable to clinical preprocessing in the real world. Although these are encouraging outcomes, there are a number of limitations. The scale of the dataset is rather moderate, and the synthetic degradation pipeline may not fully capture the diversity of real-world acquisition artifacts. Furthermore, all images were downsampled to 128×128 pixels for computational efficiency; future work should evaluate the framework at higher resolutions to better assess its capacity for preserving thin vessel structures in clinical-resolution angiograms. Future directions would be to explore multi-center validation, use perceptual/ hybrid loss functions, and lightweight attention mechanisms to deploy attention in real-time. However, the suggested hybrid CNN-Transformer framework has a good potential to be a powerful and clinically significant improvement framework in coronary angiography imaging.

The improved performance of the proposed hybrid framework can be attributed to the complementary strengths of convolutional and transformer architectures. Convolutional neural networks effectively capture local spatial features such as edges and textures, while transformer modules model long-range dependencies through self-attention mechanisms. By integrating these components, the proposed model achieves improved structural preservation and global consistency in angiographic image enhancement. The proposed hybrid CNN–Transformer architecture is designed to be dataset-independent and can be applied to other angiography datasets without structural modification. Although the current study evaluates performance using the ARCADE dataset, the model relies on generic feature extraction and attention mechanisms that are applicable across different imaging environments. Future work will focus on validating the framework using multi-center datasets to further assess generalization capability.

5 Conclusion

The paper suggests the hybrid CNN-Transformer to improve the quality of coronary angiography images, which combines both local convolutional feature extraction and global self-attention-based contextual modeling. It was aimed at overcoming the drawbacks of the conventional enhancement approaches and single deep-learning frameworks in maintaining fine vascular structure details and long-range anatomical continuity. The experimental findings showed that traditional contrast enhancement and traditional filtering methods offer a low degree of structural preservation and can enhance noise. Models based on CNN contributed greatly to the reconstruction of local vessels but were limited by small receptive fields. Transformer-only models were found to be more globally coherent, but with a weak local inductive bias to fine texture restoration. These complementary capabilities were well integrated in the proposed hybrid framework to deliver higher quantitative results based on PSNR of 36.10 dB and SSIM of 0.9670. Qualitative analysis also established a greater vessel clarity, fewer artifacts and better visualization of thin distal branches. The analysis of efficiency showed that the hybrid model is in a good balance between the cost of computations and reconstruction performance, which makes it applicable in the context of practical clinical preprocessing. In summary, the combination of convolutional and attention-based mechanisms represents an effective solution for structure-sensitive enhancement of coronary angiography. Further development will be done on multi center validation on different clinical datasets to determine generalization in different conditions of acquisition. Model compression techniques and lightweight attention mechanisms will be investigated to enable real-time deployment. Also, the combination of the perceptual losses and assessment of the improvement structure in the lower-level tasks, including the vessel segmentation and stenosis detection, will further verify the clinical practicability.

Data Availability Statement

The datasets and source code supporting the findings of this study are publicly available at the following GitHub repository: https://github.com/islamuddinw/RK-IU.git

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study used only the publicly available ARCADE dataset. No new patient data were collected, and no additional ethical approval was required.

References

[1] Edpuganti, S., Shamim, A., Gangolli, V. H., Weerasekara, R. A. D. K. N., & Yellamilli, A. (2025). Artificial intelligence in cardiovascular imaging: Current landscape, clinical impact, and future directions. *Discoveries*, *13*(2), e211. [CrossRef]

[2] Gurav, A., Revaiah, P. C., Tsai, T. Y., Miyashita, K., Tobe, A., Oshima, A., ... & Serruys, P. W. (2024). Coronary angiography: a review of the state of the art and the evolution of angiography in cardio therapeutics. *Frontiers in cardiovascular medicine*, *11*, 1468888. [CrossRef]

[3] Asif, S., Wenhui, Y., ur-Rehman, S., ul-ain, Q., Amjad, K., Yueyang, Y., ... & Awais, M. (2025). Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. *Archives of Computational Methods in Engineering*, *32*(2), 853-883. [CrossRef]

[4] Liao, Q. M., Hussain, W., Liao, Z. X., Hussain, S., Jiang, Z. L., Zhu, Y. H., ... & Wu, D. D. (2025). Computer-aided application in medicine and biomedicine. *International Journal of Computational Intelligence Systems*, *18*(1), 221. [CrossRef]

[5] Benz, D. C., Benetos, G., Rampidis, G., Von Felten, E., Bakula, A., Sustar, A., ... & Buechel, R. R. (2020). Validation of deep-learning image reconstruction for coronary computed tomography angiography: impact on noise, image quality and diagnostic accuracy. *Journal of cardiovascular computed tomography*, *14*(5), 444-451. [CrossRef]

[6] Hu, J., & Law, K. E. (2025). A Comprehensive Review of Compressed Image Restoration and Compression Artifact Reduction Techniques. *IEEE Access*, *13*, 213824-213840. [CrossRef]

[7] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88. [CrossRef]

[8] Kumar, R. R., & Priyadarshi, R. (2025). Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches. *Multimedia Tools and Applications*, *84*(12), 10817-10875. [CrossRef]

[9] Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, *19*, 221-248. [CrossRef]

[10] Ali, A. M., Benjdira, B., Koubaa, A., El-Shafai, W., Khan, Z., & Boulila, W. (2023). Vision Transformers in Image Restoration: A Survey. *Sensors*, *23*(5), 2385. [CrossRef]

[11] Zhang, C., Wang, J., Shi, Y., Yin, B., & Ling, N. (2025). A CNN-transformer hybrid network with selective fusion and dual attention for image super-resolution. *Multimedia Systems*, *31*(2), 126. [CrossRef]

[12] Yang, H., Wang, Z., Liu, X., Li, C., Xin, J., & Wang, Z. (2023). Deep learning in medical image super resolution: a review. *Applied Intelligence*, *53*(18), 20891-20916. [CrossRef]

[13] Zhao, B., Peng, J., Chen, C., Fan, Y., Zhang, K., & Zhang, Y. (2025). Diagnosis of coronary heart disease through deep learning-based segmentation and localization in computed tomography angiography. *IEEE Access*, *13*, 10177-10193. [CrossRef]

[14] Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2023). Transformers in medical imaging: A survey. *Medical image analysis*, *88*, 102802. [CrossRef]

[15] Majid Mohammed, I., & Ashidi Mat Isa, N. (2025). Contrast Limited Adaptive Local Histogram Equalization Method for Poor Contrast Image Enhancement. *IEEE Access*, *13*, 62600-62632. [CrossRef]

[16] Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. (1998, October). Multiscale vessel enhancement filtering. In *International conference on medical image computing and computer-assisted intervention* (pp. 130-137). Berlin, Heidelberg: Springer Berlin Heidelberg. [CrossRef]

[17] Alkhodari, M., Alefishat, E., Jelinek, H. F., Kaabneh, A., & Liatsis, P. (2025). Enhancing CCTA image quality: a review of deep learning approaches for advanced artifact correction and denoising. *The Artificial Intelligence Review*, *58*(10), 331. [CrossRef]

[18] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and*

machine intelligence, 38(2), 295-307. [CrossRef]

[19] Gondara, L. (2016, December). Medical image denoising using convolutional denoising autoencoders. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)* (pp. 241-246). IEEE. [CrossRef]

[20] Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, *26*(7), 3142–3155. [CrossRef]

[21] Tatsugami, F., Higaki, T., Nakamura, Y., Yu, Z., Zhou, J., Lu, Y., ... & Awai, K. (2019). Deep learning–based image restoration algorithm for coronary CT angiography. *European radiology, 29*(10), 5322-5329. [CrossRef]

[22] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. [CrossRef]

[23] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021, October). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 9992-10002). IEEE. [CrossRef]

[24] Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., & Yang, M. H. (2022, June). Restormer: Efficient Transformer for High-Resolution Image Restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5718-5729). IEEE. [CrossRef]

[25] Yu, Z., Lee, F., & Chen, Q. (2023). HCT-net: hybrid CNN-transformer model based on a neural architecture search network for medical image segmentation. *Applied Intelligence*, *53*(17), 19990-20006. [CrossRef]

[26] Senthil Anandhi, A., & Jaiganesh, M. (2025). An enhanced image restoration using deep learning and transformer based contextual optimization algorithm. *Scientific Reports*, *15*(1), 10324. [CrossRef]

[27] Molenaar, M. A., Selder, J. L., Nicolas, J., Claessen, B. E., Mehran, R., Bescós, J. O., ... & Chamuleau, S. A. (2022). Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease. *Current cardiology reports*, *24*(4), 365-376. [CrossRef]

[28] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*. [CrossRef]

[29] He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE.

[30] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing. [CrossRef]

Romaan Khan Currently working as lecturer in institute of Computer Science & Information Technology, University of Agriculture Peshawar. Received BS Degree from Bacha Khan University Charsadda, Pakistan and received MS Degree from City University of Science & Information Technology Peshawar, Pakistan. (Email: romaan@aup.edu.pk)

Islam Uddin is currently serving as a Lecturer in the Department of Computer Science at Abdul Wali Khan University, Mardan, Pakistan. He is currently pursuing a Ph.D. degree in Computer Science at Abdul Wali Khan University, Mardan, Pakistan. He received his M.S. degree in Computer Science from Abdul Wali Khan University, Mardan, in 2021. His research focuses on the development of advanced computational and machine learning approaches for biological and medical data analysis. His research interests include bioinformatics, machine learning, deep learning, computational biology, and artificial intelligence for biomedical applications. He has contributed to several research publications in international journals, particularly in the area of deep learning–based prediction models for genomic and epigenetic sequence analysis. (Email: islamuddin@awkum.edu.pk)

Salman Khan is currently working as an Assistant Professor in the Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He received his Ph.D. degree in Computer Science in 2020 from Abdul Wali Khan University, Mardan, Pakistan. He obtained his M.S. degree in Mobile Computing from the University of Glamorgan, U.K., in 2009. Prior to joining King Saud University, he served as a Lecturer in the Department of Computer Science at Abdul Wali Khan University, Mardan, Pakistan. He has more than 15 years of combined experience in teaching and industry. His research interests include big data, mobile computing, bioinformatics, machine learning, and parallel programming. (Email: snawab@ksu.edu.sa)

Abdul Allah Currently working as lecturer in institute of Computer Science & Information Technology, University of Agriculture Peshawar. Received Bachelor and Master Degrees from University of Swat, Pakistan and received MPhil Degree from Islamia College University Peshawar, Pakistan. (Email: abdullahswat@aup.edu.pk)