Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

FR-GAN: A Fusion Refine with Dual Attention Architecture for Denoising of Dental Panoramic X-ray Images

Sameena Begum[1], Nagaraj Yamanakkanavar[2]

Department of Electronics and Communication Engineering, Central University of Karnataka, Kalaburgi, India.

E-mail: [1]sameenabegum678@gmail.com, [2]nagrajpy@cuk.ac.in

Abstract

Denoising dental panoramic X-ray images has become a major concern in medical imaging and computer vision, particularly when the Gaussian noise is high. To denoise the noisy images while preserving image features, we introduce a fusion-refine generative adversarial network (FR-GAN). The FR-GAN comprises a generator, a dual-attention U-Net, and a ResNet enhanced discriminator. The generator takes noisy images as input and produces coarse representations for initial reconstruction. Then, these coarse images are used as a preliminary denoised image, which is again refined to restore structural features for image clarity. In addition, we incorporate the U-Net with a dual-attention mechanism linking the generator and discriminator, which enables the refinement of coarse features into smooth features to create meaningful representations. Meanwhile, the U-Net refiner incorporates progressive feature fusion attention and low-rank local window attention to complement multi-scale feature extraction and local texture refinement. Lastly, the ResNet block-enhanced discriminator implements perceptual realism by differentiating genuine and reconstructed images and directing the network to yield high-quality images. In our experiment, 100 dental images collected from Al-Badar Dental College and Hospital, Kalaburagi, were subjected to Gaussian noise at 20 dB and 40 dB using ImageJ software. To evaluate the effectiveness of the proposed approach, experiments were conducted with Gaussian noise at 20 dB and 40 dB. The experimental results show that the proposed method is superior to other state-of-the-art approaches for Gaussian noise of 20 dB and 40 dB. It achieves a PSNR of 31.705, SSIM of 0.826, MSE of 0.0007, and MAE of 0.0195 at 40 dB, and a PSNR of 32.897, SSIM of 0.861, MSE of 0.0007, and MAE of 0.0184 at 20 dB. The framework provides effective high-quality denoising with clean edges, clear textures, and few artifacts, and is applicable to precision-critical tasks, such as medical and scientific imaging.

Keywords: Denoising, Dual-Attention Mechanism, GAN, ResNet.

1. Introduction

Image denoising is a critical problem in computer vision and medical imaging, aiming to recover a pure, original image from a noisy or distorted image generated by various sources. To generate realistic images and enable the model to understand complex data distributions, generative adversarial networks (GANs) are used for adversarial training [1]. Furthermore,

DOI: https://doi.org/10.36548/jiip.2026.1.020
Received: 19.01.2026, received in revised form: 17.02.2026, accepted: 03.03.2026, published: 11.03.2026

extensive reviews provide an overview of GAN variants that include significance and challenges in medical image fusion [2]. However, concerns about stability across various GAN architectures are prevalent when enumerating application areas that will help future research, especially in medical image fusion and analysis [3]. To overcome unstable training and stabilize the training techniques, a simple GAN model has been proposed that facilitates convergence by using a simplified distributional model [4]. The U-Net architecture provides an encoder-decoder structure for medical image analysis and segmentation [5]. Additionally, the residual learning technique helps to develop deeper networks to address the vanishing gradient problem [6]. However, the conditional adversarial network has been developed, which can be extended to image-to-image translation by significantly expanding the range of supervised learning applications of GAN [7]. Furthermore, due to the scarcity of original reference images in clinical environments, self-supervised denoising methods like Wavelet have been developed that require a pair of noisy images during training, on which wavelet transformation is applied to obtain the denoised images [8]. Han et al. [9] proposed the dropped projection strategies (DPS) for denoising low-dose CBCT images to solve the over-smoothing problem encountered in the analysis of dental radiographs. Amani Almalki et al. [10] introduced the self-supervised masked image model (SimMIM) to leverage self-supervised learning and address the problem associated with the limited availability of panoramic radiographs. Based on these concepts, a clinically driven denoising framework, such as an image domain denoising network (IDDNet), has beendeveloped for low-dose cone-beam CT images [11]. Chen et al. [12] proposed the dual diffusion brownian bridge model (DBBCS) to capture the distributions of clean images and structural noise simultaneously using two interconnected diffusion processes. Additionally, to improve efficiency, a dual-diffusion model for efficient structural noise suppression and progressive feature fusion with reverse attention has been developed to understand the complex features [13]. While attention mechanisms are also used to improve image restoration across different areas by combining features from two branches of attention, this approach enables better pan-sharpening for remote sensing images [14]. Low rank and locality-constrained self-attention make sequence modelling tasks more efficient [15]. At the same time, CNN-based denoising methods, such as FFDNet, focus on computational efficiency and adaptability across different noise levels [16]. Transformer-based architectures such as restormer offer scalable solutions for high-resolution image restoration [17]. Kawar et al. [18] proposed the denoising diffusion restoration model to efficiently implement unsupervised denoising and eliminate the nonlinear inverse problem. However, unsupervised and self-supervised denoising methods are improving with the help of the Noise2Void method, which allows the model to learn from a single noisy image [19]. Additionally, diffusion posterior sampling techniques are developed that ensure asymptotically accurate results and eliminate inverse problems as stochastic filtering tasks [20]. P. Nandal et al. [21] developed a quantum deep convolutional generative adversarial network, which is merged with a quantum computing layer to enhance its ability through quantum-generated inputs. Iqbal et al. [22] introduced the data-driven column dropping masking strategy to ensure stability while preserving a smooth dropping rate. Gautam et al. [23] proposed a Pureformer that uses a feature enhancer block to expand the latent space with the help of a spatial filter bank, improving the fusion and restoration of features. Furthermore, a real-time biomedical image denoising network is a domain-specific denoising task developed to enhance gradients and high-frequency information while suppressing noise [24]. However, diffusion models have been demonstrated to be effective for restoring high-resolution microscopy images [25]. Kim et al. [26] conducted a systematic review of low-dose CT denoising and highlighted its growing clinical importance from a perceptual quality perspective. Besides image restoration, deep learning is being used in fully automated diagnostic pipelines, such as dental radiograph quality assessment [27]. A. Lee et al. [28]

developed a hybrid CNN-transformer network for denoising real-world images that systematically balances performance and efficiency under practical deployment constraints. Eventually, with the most recent studies pointing to perceptual, adversarial, and task-specific objectives, the design of loss functions remains a major factor in optimizing denoising performance [29]. Xiao et al. [30] developed window-based attention mechanisms that use various window shapes and sizes in the form of image information that aim to achieve a balance between representational capacity and computational efficiency. Together, these changes indicate that generative, diffusion, and attention-based models are evolving rapidly, and their role in medical image denoising and restoration is growing steadily. In contrast, image denoising has always been a critical challenge due to the scarcity of clean training data and the complexity of noise patterns. Various deep learning techniques have been proposed, but most of them work well either for prioritizing diagnostic detail or for computational efficiency. Each of GANs, CNNs, transformers, and diffusion models has its merits, but none can fully address the problems of stability, scalability, and generalization. Therefore, a unified efficient denoising architecture that is capable of enhancing structural fidelity and perceptual realism is required. We propose a fusion-refine generative adversarial network (FR-GAN). This generative adversarial network integrates a generator, U-Net with dual attention mechanisms, and a residual-block-enhanced discriminator for robust image denoising. The contributions of this paper are summarized as follows:

- We propose a novel FR-GAN framework consisting of a GAN generator, U-Net refiner with dual attention mechanisms, and a residual block-enhanced discriminator structure for powerful image denoising.
- The generator produces preliminary denoised images for initial reconstruction before proceeding to further refinement to recover structural details.
- The U-Net refiner utilizes a dual attention mechanism, such as progressive feature fusion attention (PFFA) and low-rank local window attention (LRLWA), to aggregate multi-scale features and apply the learned features from the coarse to smooth image and move the features in a meaningful way.
- The discriminator has enhanced residual blocks, which ensure perceptual realism by distinguishing real from fake images.
- The proposed method was evaluated on dental panoramic X-ray datasets under Gaussian noise levels of 20 dB and 40 dB, surpassing state-of-the-art methods in terms of performance evaluation metrics.

This paper is organized as follows: Section 2 presents an overview of the proposed method. The experimental analysis is discussed in Section 3. Finally, Section 4 presents the conclusion and a summary of the final observations.

2. Proposed Method

We proposed the FR-GAN architecture for denoising panoramic dental X-ray images. Figure 1 shows the proposed framework developed for denoising the dental X-ray images. The proposed FR-GAN framework can be discussed in three stages: (i) a generator, (ii) a U-Net refiner with a dual attention mechanism, which comprises PFFA and LRLWA, and (iii) a ResNet-based discriminator. The proposed framework uses a GAN generator to generate coarse

images and a U-Net encoder-decoder architecture to gradually capture contextual information from input images while progressively reducing spatial resolution. The LRLWA block is located at the bottleneck, enabling the U-Net architecture to locally refine features more efficiently while decreasing processing overhead. The PFFA module, integrated at skip connections, selectively highlights important features while discarding irrelevant noise. Fusion-refine signifies that the refinement process is inherently embedded in the fusion process itself. The arrangement of the PFFA block in the proposed architecture is shown in Figure 1.

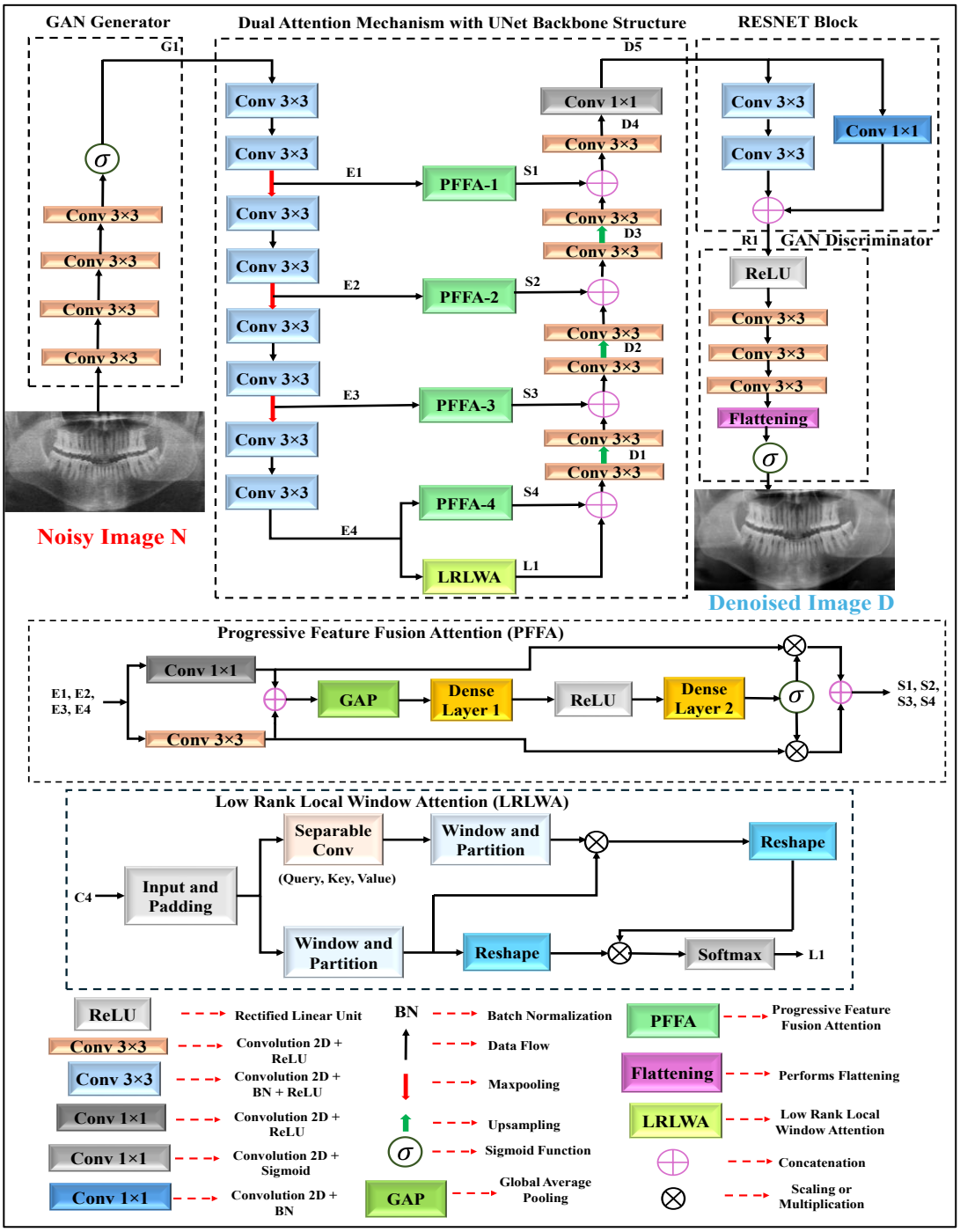

Figure 1. Architecture of the Proposed FR-GAN Framework

In addition, the combination of PFFA and LWLRA in a U-Net architecture serves as a collective refiner. It reflects the intended workflow in which fusion and refinement are

performed simultaneously rather than treated as separate steps in the proposed method. The decoder aims to ensure that the contextual information learned during the compression phase is combined with spatial accuracy. Furthermore, it reconstructs the high-resolution outputs by gradually upsampling the feature maps and combining them with the standard encoder skip connections. Finally, the ResNet-based discriminator comprises a ResNet block connected to the GAN discriminator. The ResNet block enables the proposed model to focus more on learning refinement for identity mapping rather than relearning the entire transformation. Additionally, it stabilizes the optimization and speeds up convergence. The discriminator in the proposed method serves as the supervisory stage of FR-GAN, guiding the generator and refiner toward producing clinically realistic denoised outputs. A detailed explanation of each block is as follows.

2.1 GAN Generator

In the FR-GAN model, the GAN generator acts as the first component in the denoising process. The GAN generator is designed to receive the noisy panoramic X-ray images and produce a rough yet structurally consistent output. A standard convolutional neural network (CNN) is often developed using a regression model, which is pixel-based. However, the adversarial learning component is used to ensure the generator produces images that are perceptually similar to the clean radiographs. The GAN generator component consists of a stacked convolutional layer and a ReLU activation component. A residual connection is also added to the model to allow the acquisition of finer features and proceed with the training component. The convolutional component is used to transform the feature maps and produce new feature maps that exhibit both structure and texture. A residual connection is also used to avoid the vanishing gradient problem and allow the model to utilize the features learned in the previous component. This approach makes the learning stable. The generator component may be viewed as a mapping function, which may be expressed as:

$$G_1 = \sigma(\delta(conv_{3\times3}(\delta(conv_{3\times3}(\delta(conv_{3\times3}(\delta(conv_{3\times3}(N)))))))) \quad (1)$$

The output of the GAN generator is G_1 the sigmoid function, and the ReLU function, ReLU the 2D convolution function conv 3×3 with kernel size 3×3 and N denotes the noisy image applied as input image to GAN generator. When it comes to the task of noise removal of dental panoramic radiographs, the generator can efficiently filter out large-scale noise artifacts such as graininess and streaks from low doses. However, the generator can effectively remove global noise. Nonetheless, leftover errors appear as fine Gaussian noise that is not completely eradicated. Fine anatomical details, such as enamel margins, root canals, and small lesions can be smudged. The texture may be distorted, making it look over-smoothed or artificially created. The U-Net refiner is employed to solve this issue, which mainly serves to retrieve fine structures. Consequently, the generator is designed as a first stage that creates a very rudimentary denoised image to lessen the load on the other modules while ensuring structural consistency for the refiner. The proposed framework's GAN generator model is effective in distilling input features through stacked convolutions and extracting rich local patterns. Ultimately, the sigmoid layer ensures that outputs are normalized. Therefore, the model can also be used for segmentation or mask generation tasks. To train the generator, we have utilized an adversarial loss, which encourages the generator to produce sharper and more realistic images of the denoised input by learning the data distribution from the discriminator's feedback. This will improve the fine features and the realistic look of the images beyond the structural similarity index.

2.2 U-Net Refiner with Dual Attention mechanism

The U-Net refiner with the Dual Attention Mechanism is the second stage of the proposed model, named FR-GAN. The purpose of this stage is to improve the structural accuracy and recover the details of the anatomy of the image, which might have been lost during the coarse denoising of the generator. The U-Net Refiner with the Dual Attention Mechanism consists of the following: (a) U-NET Encoder, (b) LRLWA, (c) PFFA, and (d) U-NET Decoder.

2.2.1 U-Net Encoder

The U-Net encoder progressively captures contextual information from the input image at different levels while simultaneously reducing spatial resolution. The encoder is composed of a chain of convolutional blocks. Each block typically consists of two 3×3 convolutions, followed by a ReLU activation and batch normalization, with the same padding to preserve spatial dimensions. Then, a 2×2 max pooling operation is performed after each stage, reducing the spatial resolution by half while doubling the number of feature channels. This pyramid system allows the network to obtain more abstract representations. Nevertheless, the first layers are more concerned with semantic patterns, and the deeper layers concentrate on fewer edges and textures. In addition, the encoder produces intermediate feature maps at each resolution via skip connections which are used by the decoder. These skip connections contain fine grained localization information that may be lost during downsampling. The U-Net encoder was an important element of our proposed method, as it provides the network with a sequence of multi-scale contextual features extracted from the generator output and gradually decreasing spatial resolution. The repetition of the convolutional and pooling layers is successful in learning both the minute details of the local structures and the high-level semantic patterns. Moreover, the feature representation that was produced was colorful and informative. The skip connections added at every encoder step locally preserve essential cues that are subsequently added at the decoder to produce high-quality, refined outputs. Such a design makes the encoder a potent feature extractor and a generator of spatial richness, which raises the accuracy, robustness and effectiveness of the model directly. The U-Net encoder's first stage is expressed as,

$$E_1 = maxpool_{2\times2}(BN(\delta(conv_{3\times3}(BN(\delta(conv_{3\times3}(G_1))))))) \quad (2)$$

where E_1 is the output of the first stage of the U-Net encoder, $maxpool$ represents the maxpooling operation, and BN denotes batch normalization. The E_1 is fed to the second stage of the U-Net encoder, and the output of the second encoder block is defined as,

$$E_2 = maxpool_{2\times2}(BN(\delta(conv_{3\times3}(BN(\delta(conv_{3\times3}(E_1))))))) \quad (3)$$

where E_2 denotes the output of the second encoder block, which is further applied to the third stage of the U-Net encoder, the production of the third stage of the encoder is described as,

$$E_3 = maxpool_{2\times2}(BN(\delta(conv_{3\times3}(BN(\delta(conv_{3\times3}(E_2))))))) \quad (4)$$

where E_3 is an output of the third stage of the U-Net encoder. The E_3 is fed to the fourth stage of the encoder, and the output of the fourth stage of the encoder is expressed as,

$$E_4 = maxpool_{2\times2}(BN(\delta(conv_{3\times3}(BN(\delta(conv_{3\times3}(E_3))))))) \quad (5)$$

where E_4 represents the output of the fourth stage of the encoder. The U-Net encoder output is given to LRLWA.

2.2.2 Low Rank Local Window Attention Block (LRLWA)

The feature output from the last stage of the U-Net encoder served as input to the LRLWA block. Padding is applied to the input so that the image's spatial dimensions align with a window size of about 7×7, enabling uniform and efficient window-based processing. Furthermore, three separable convolutions are applied to generate the query (Q_{sep}), key ($K^T{}_{sep}$), and value (V_{sep}). Feature maps with the query and key reduce the channel dimension to enforce a low-rank representation. These feature maps are partitioned into non-overlapping local windows, and each window is reshaped into a sequence of tokens. Within each window, attention weights are computed by taking the scaled dot product of queries and keys, followed by a Softmax normalization, which is described as follows:

$$A = softmax\left(\frac{Q_{sep}K^T{}_{sep}}{\sqrt{d}}\right)V_{sep} \tag{6}$$

where A denotes the attention mechanism, softmax is the softmax function, T variable represents the transpose operation applied to the key vector, and d is the dimension of the key vector. These weights are then used to aggregate the values, producing context-aware features for each window. Finally, the outputs from all windows are reshaped and merged back into the original spatial layout, yielding an attention-enhanced feature map that preserves locality while reducing computational cost. The output of the LRLWA block is expressed as,

$$L_1 = R_w(A(D_{low}(P_w(E_4)))) \tag{7}$$

where L_1 is an output of the LRLWA block, R_w denotes low rank approximation, D_{low} represents low rank projection, and P_w is a partitioning operation. The LRLWA module plays a significant role in the proposed model, as it helps refine highly efficient local features at a lower computational cost. However, it uses separable convolutions to produce low-rank queries and keys, achieving lightweight attention without sacrificing representational capacity. Dividing the input into windows enables the model to capture spatial dependencies at a fine-grained level, which is essential for vision tasks. Finally, this module enhances both the model's precision and its scalability, serving as a pivotal creative feature that harmonizes efficiency and performance.

2.2.3 Progressive Feature Fusion Attention (PFFA)

The PFFA block is integrated with the skip connections of the U-Net network to better enhance salient features and reduce extraneous noise in the U-Net's intermediate outputs. Figure 2 illustrates the progressive feature fusion attention (PFFA) block for feature extraction from noisy images. In the PFFA block, all encoder skip connections are sequentially fed into a feature refinement framework that begins with 1×1 and 3×3 separable convolutions. A 1×1 separable convolution followed by a ReLU activation is used to concisely compress and mix channel information. The 3×3 separable convolution followed by ReLU is then used to extract local spatial patterns.

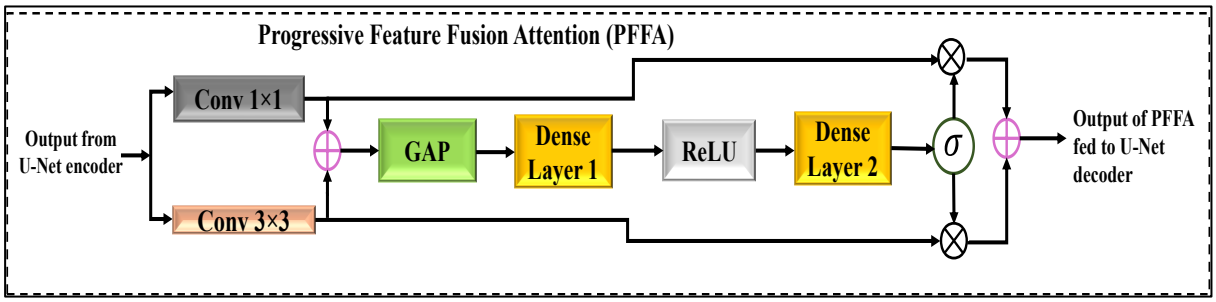

Figure 2. Progressive Feature Fusion Attention (PFFA) Block

These combined operations make up the progressive feature extractor. Smoothed features are then pooled using global average pooling (GAP), which serves as a channel descriptor by averaging across the spatial dimensions of the input. The descriptor output is then passed to two dense layers: the first, followed by the ReLU function, reduces dimensionality, while the second restores the original channel dimensionality. A sigmoid activation yields normalized weights in the range of (0, 1), indicating the level of each channel. These weights are then multiplied by each channel's corresponding weight to learn the importance of specific features. Finally, the module includes a residual connection by adding the scaled features. This ensures stability, retention of original information, and ease of gradient flow during training. The input feature map initiates the operation necessary for progressive convolution refinement and channel-wise attention expressed as,

$$S_1 = [conv_{1\times1}(E_1) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_1) \oplus conv_{3\times3}(E_1))))] \oplus [conv_{3\times3}(E_1) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_1) \oplus conv_{3\times3}(E_1))))] \quad (8)$$

where S_1 is the combined output of the first PFFA, $conv$ 1×1 denotes a 2D convolution with kernel size 1×1, $\otimes$ represents element-wise multiplication, $W2$ is the second dense layer, $W1$ is the first dense layer, GAP is global average pooling, $\oplus$ and denotes element-wise addition. The output S_1 from the first PFFA is provided to the last stage of the U-Net decoder via a skip connection. The same workflow applies to other PFFAs and their respective stages of the U-Net decoder. The output of the second PFFA is expressed as,

$$S_2 = [conv_{1\times1}(E_2) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_2) \oplus conv_{3\times3}(E_2))))] \oplus [conv_{3\times3}(E_2) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_2) \oplus conv_{3\times3}(E_2))))] \quad (9)$$

where S_2 denotes the output of the second PFFA, which is connected to the third stage of the decoder, the output of the third PFFA is mathematically described as,

$$S_3 = [conv_{1\times1}(E_3) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_3) \oplus conv_{3\times3}(E_3))))] \oplus [conv_{3\times3}(E_3) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_3) \oplus conv_{3\times3}(E_3))))] \quad (10)$$

where S_3 represents the output of the third PFFA block connected to the second stage of the decoder. The output of the fourth PFFA block is expressed as,

$$S_4 = [conv_{1\times1}(E_4) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_4) \oplus conv_{3\times3}(E_4))))] \oplus [conv_{3\times3}(E_4) \otimes \sigma(W_2\delta(W_1 GAP(conv_{1\times1}(E_4) \oplus conv_{3\times3}(E_4))))] \quad (11)$$

where S_4 denotes the output of the fourth PFFA block, which is connected to the first stage of the decoder. This concise description underlines the modules in two ways: progressive convolutional refinement through separable convolution and channel-wise attention via the GAP, MLP, and sigmoid pathway. By merging these two methods, PFFA increases the

representational power to the extent that the network can highlight informative channels through stable residual learning. The PFFA block plays a crucial role, facilitating feature enrichment by combining progressive convolutional refinement and adaptive channel attention. Therefore, it enables the network to focus on the most informative features and discard less useful ones, thereby enhancing its discriminative power. The residual connection ensures stability and preserves the original data, making the model more robust during training. In addition, PFFA boosts feature fusion and thus allows for performance gains in the segmentation task. The combined use of LRLWA and PFFA ensures balanced denoising, preserving both global fidelity and local diagnostic clarity.

2.2.4 U-Net Decoder

The decoder generates the segmentation map by progressively increasing the encoded features to the original resolution. Initially, the feature map is upsampled using a simple interpolation method at each stage. Then, this upsampled version is combined with the corresponding encoder feature map via skip connections, thereby preserving the fine details of the localization. The merged features are then passed through several $conv_{3\times3}$layers with ReLU activations, which refine the representation and remove artifacts. The same procedure is performed at all the decoder stages and the input resolution is restored stepwise. Finally, a $conv_{1\times1}$ layer converts the refined features into the required number of output channels, yielding pixel-wise predictions. Using ReLU-activated convolutions allows the decoder to bypass checkerboard artifacts and deliver smoother, more stable reconstructions while still leveraging the encoder's contextual information. The workflow of the U-Net decoder is mathematically expressed as,

$$D_1 = \delta\big(conv_{3\times3}(S_4 \oplus L_1)\big) \tag{12}$$

where D_1 is the output of the first stage of the decoder, which is fed to the second stage of the U-Net decoder. The output of the second stage of the decoder is given as,

$$D_2 = \delta(conv_{3\times3}(S_3 \oplus (\delta(conv_{3\times3}(D_1))))) \tag{13}$$

where D_2 denotes output from the second stage of the decoder. The $D_2$is applied to the third stage of the U-Net decoder. The output of the third stage of the decoder is mathematically represented as,

$$D_3 = \delta(conv_{3\times3}(S_2 \oplus (\delta(conv_{3\times3}(D_2))))) \tag{14}$$

where D_3 is an output from the third stage of the decoder, which is further applied to the fourth stage of the decoder. The output from the fourth stage of the decoder is expressed as,

$$D_4 = \delta(conv_{3\times3}(S_1 \oplus (\delta(conv_{3\times3}(D_3))))) \tag{15}$$

where D_4 represents the output of the fourth stage of the decoder, D_4 is fed to the classification layer to provide completely refined dental images. The output from the classification layer of the U-Net decoder is expressed as,

$$D_5 = \sigma(conv_{1\times1}(D_4)) \tag{16}$$

The significance of using the U-Net decoder in our proposed method is that it helps the model produce high-resolution outputs while preserving localized details. When the decoder receives the feature maps from the PFFA, integrated with a skip connection at a lower

resolution, it first upsamples them in steps and then fuses the upsampled features with the corresponding skip connection from the encoder. Therefore, the decoder not only receives the context learned during compression but also the spatial precision of the original signal via skip connections. Using ReLU-activated convolutions instead of transposed convolutions makes the proposed model's training more stable and reduces artifacts, resulting in smoother reconstructions. Such a design directly improves the precision and reliability of our technique. Further, it can be used for tasks where boundary precision and pixel-level predictions are essential.

2.3 ResNet-based Discriminator

The ResNet-based discriminator is augmented with two blocks: a ResNet block and a GAN discriminator block.

2.3.1 ResNet Block

The ResNet block serves as a primary component of residual networks. Its main aim is to reduce the vanishing gradient issue by enabling the training of very deep network architectures. The input feature map is processed inside the ResNet block through two different parallel paths, the main path, and the shortcut path. The shortcut path uses $conv_{1\times1}$ convolution proceeds by batch normalization to ensure the dimension of an input must match the main path output. The main path comprises of two consecutive $conv_{3\times3}$ convolutions. Each convolution layer followed by a batch normalization. The ReLU activation added after the convolution to add nonlinearity. The output of the shortcut and the main paths are combined elementwise and constitute the residual connection after which is follows a final ReLU activation of the compounded feature map and the output of a block is obtained. By facilitating the direct passage of gradients through the shortcut path, the ResNet block not only accelerates network convergence but also stabilizes training and enhances the network representational power. This is highly useful in intense networks, where one must preserve both contextual information and fine details. The ResNet block is expressed as,

$$R_1 = (BN(\delta(conv_{3\times3}(BN(\delta(conv_{3\times3}(D_5)))))) \oplus (BN(conv_{1\times1}(D_5))) \quad (17)$$

where, R_1 represents the output of the ResNet block. In our proposed method, the ResNet block is important because it improves both the model's learning efficiency and representational capacity. Adding a residual connection block enables the model to focus on learning as a variation of the identity mapping, rather than relearning the entire transformation to achieve optimization and faster convergence. The shortcut route maintains dimension alignment and provides a direct gradient flow, whereas the main path refines features through successive convolutions and ReLU activations. This block significantly reduces the vanishing gradient problem and allows deeper layers to have a greater effect. Further, this leads to higher accuracy, better robustness, and generalization across different tasks.

2.3.2 GAN Discriminator

In FR-GAN, the discriminator offers adversarial supervision to the generator and refiner. Therefore, the denoised images produced are indistinguishable from genuine original radiographs. The ResNet-inspired discriminator stabilizes adversarial training and prevents vanishing gradients. Such a combination helps the discriminator learn deep hierarchical representations without any performance degradation as the network depth increases. The final

classification layer applies a sigmoid activation to produce a probability score, which indicates whether the input image is real or generated. The final denoised image is described as,

$$D = \sigma(flattening(\delta(conv_{3\times3}(\delta(conv_{3\times3}(\delta(conv_{3\times3}(\delta(R_1)))))))))) \quad (18)$$

The generator and refiner are trained to minimize the adversarial loss, thereby increasing the realism of the output. The discriminator is crucial for ensuring perceptual realism in dental panoramic radiographs. While reconstruction losses, such as mean squared error, help promote pixel-wise similarity. However, such losses tend to generate very smooth images that lack texture. The discriminator rebalances this by disallowing outputs that are far from the distribution of real radiographs, hence indirectly encouraging the generator and refiner to make images with real-like textures and structural details. Residual learning in the discriminator also helps stabilize and enhance the adversarial training procedure. Therefore, the discriminator is the supervisory stage of FR-GAN that keeps the generator and refiner on the right track to produce clinically realistic denoised output.

3. Experimental Analysis

3.1 Materials and Methods

The original dataset consists of 100 images collected from Al-Badar Dental College and Hospital, Kalaburagi.We limited the dataset to 100 panoramic dental X-ray images due to availability constraints and the need for consistent quality and annotation across samples. For this study, we have used the collected 100 images that was corrupted to two different Gaussian noise level of 20 dB and 40 dB which create two different datasets. For better quantitative comparison of various deep learning models and our proposed method. Specifically, Gaussian noise levels of about 20 dB and 40 dB are used in this study because these values are widely adopted benchmarks in image denoising research.

3.2 Experimental Setup

All models were implemented and executed in a Jupyter Notebook environment. The experiments were performed using TensorFlow and ran on a GPU desktop computer with an NVIDIA GeForce RTX 3070 and CoLab. The CUDA deep neural network library (cuDNN) was used for the GPU implementation. The same hyperparameters were used for all the models, including our proposed model. During training, we chose a batch size of 2, a patch size of 16, an Adam optimizer, and an epoch size of 100. We tested two batch sizes of 1 and 2. Our experimental results show that a batch size of 2 performs better. However, GAN convergence is not strictly a matter of this parameter. Smaller batches result in noisier but faster updates, while larger batches yield smoother gradients. A batch size of 2 is usually chosen for our experiment because it offers better performance and because GPU memory is limited. Further, full image denoising with a patch size of 16 is used to balance computational efficiency and local detail preservation.

3.3 Results and Discussion

Table 1 shows the performance metrics, such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), mean squared error (MSE), and mean absolute error (MAE), that are used to compare the proposed method with other existing approaches.

Table 1. Represents the Description of Performance Metrics Used to Evaluate the Models

Metrics	Definition	Mathematical Formula
Mean Squared Error (MSE)	Measures the average of the squared difference between the original image (Ii, j) and the noisy image $K(i,j)$. MN is the image dimension.	$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j) - K(i,j)]^2$
Peak Signal-To-Noise Ratio (PSNR)	It is a logarithmic measure of the ratio between the maximum signal and noise power that is derived from MSE	$PSNR = 10\log_{10}\frac{(Peakval)^2}{MSE}$
Structured Similarity Index (SSIM)	Measures perceptual similarity by considering luminance (μ), contrast (σ), and structure (C). L is a dynamic range of pixel values.	$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x{}^2+\mu_y{}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$
Mean Absolute Error (MAE)	Measures the average of the difference between the original image (Ii, j) and the noisy image $K(i,j)$.	$MAE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\lvert I(i,j) - K(i,j)\rvert$

In this paper, computational time was not reported because this study prioritized validating denoising quality and architectural effectiveness over runtime analysis. The proposed FR-GAN structure consists of dual attention fusion and a ResNet is employed as a discriminator to support the training process and assist the model in enhancing the feature refinement. The model also progressively incorporates multi-scale characteristics and noise reduction while preserving fine anatomical details in dental panoramic X-rays, outperforming conventional CNN and GAN approaches. The lower the values of MSE and MAE, the less the difference in pixels between the denoised and original images can be seen. The increase in PSNR values proves that the presented approach is efficient in reducing noise without affecting the sharpness and clarity of diagnostic structures. Nevertheless, the improved values of SSIM denote the similarity of images after denoising and confirm the maintenance of the boundary of the tooth. Moreover, the performance metrics focus on the strength and accuracy of FR-GAN regarding clinical denoising problems. The proposed method provides a high level of clarity in diagnosis compared to the current models and offers a valid interpretation of complicated dental structures.

Conversely, FR-GAN is more appropriate for denoising medical images because it optimally balances accuracy and perceptual quality in cases where the source images were corrupted with Gaussian noise whose variance is about 40 dB. However, true clinical realism requires expert validation by radiologists or clinicians, which goes beyond automated metric evaluation. Thus, while noisy–clean pairs support quantitative assessment, clinical realism must be judged through human expertise. Figure 3 illustrates the qualitative results for denoising dental images, comparing the proposed method with existing methods. The illustration is separated into two sections: (a) represents the results of Sample 1, and (b) shows the results of Sample 2. Three different images are shown in each case: the original image, the noisy image corrupted by Gaussian noise, and the denoised images obtained by the proposed method and existing methods. For Samples 1 and 2, the quality of the noisy images is drastically degraded, and the structural details of the original images are lost due to the overlay of random fluctuations. The traditional denoising methods can partially reconstruct the original image. However, they leave behind some noise, blurring edges, and lowering sharpness. To compare and visualize, small regions of the denoised images obtained using the proposed method and existing methods are zoomed in as depicted in Figure 3. The proposed approach, in contrast with existing methods will give reconstructed images that resemble the original images much more closely while maintaining a balance between noise elimination and preservation of detail.

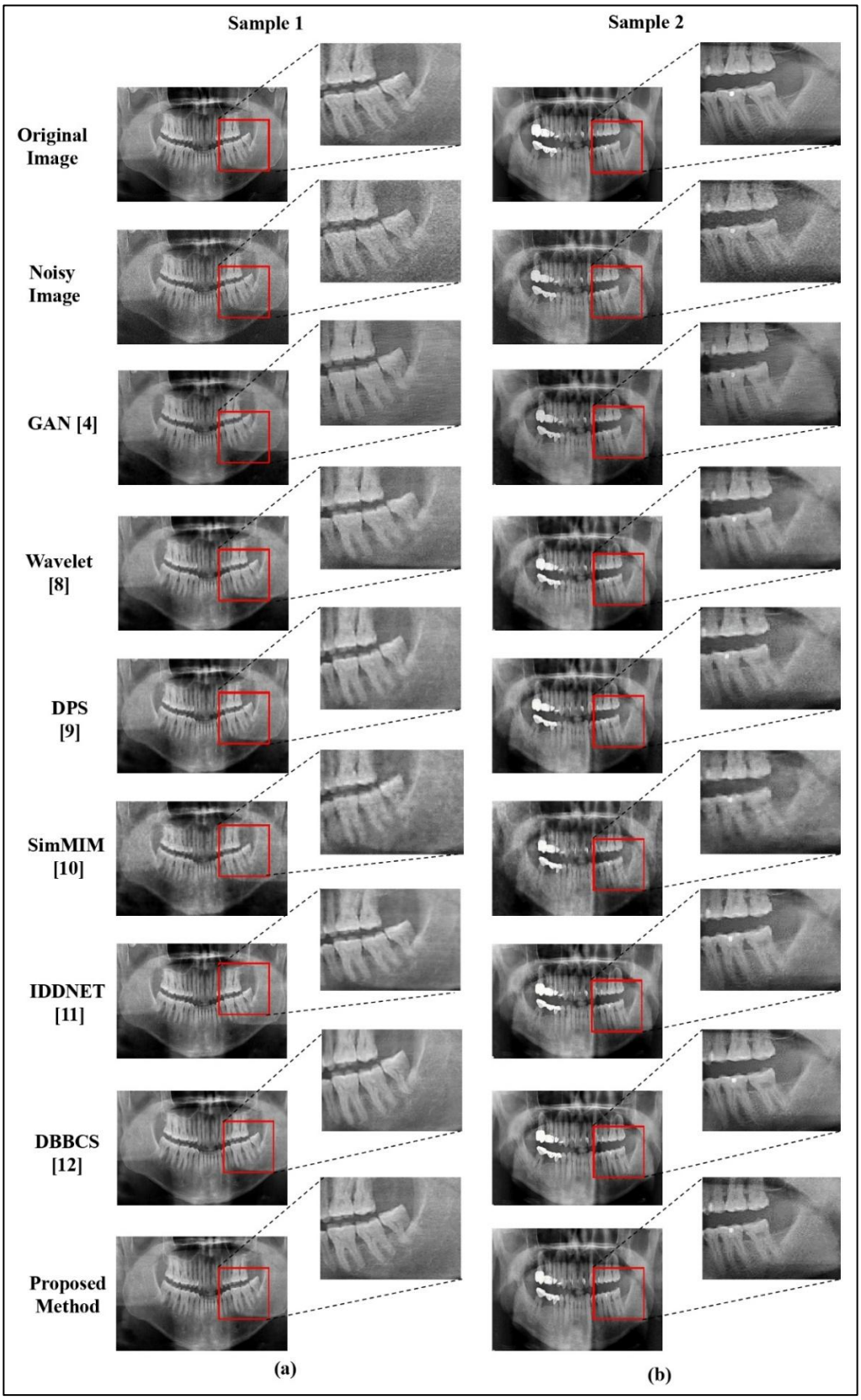

Figure 3. Denoising Comparison at 40 dB Gaussian Noise

Table 2 provides the quantitative demonstration of the proposed method against some of the existing methods in the presence of 40 dB Gaussian noise. Combined, these metrics are a measure of quantitative accuracy and perceptual quality of the reconstructed images. Evaluation metrics confirm that the proposed method is better than all other models in noise reduction and signal retention, and it has the highest PSNR of 31.705. Similarly, the proposed method has more structural consistency and perceptual similarity with the original image than other methods, which is supported by the SSIM value of 0.826. The proposed technique has the least MSE of 0.0007 and the least MAE of 0.0195, meaning that it minimizes pixel-level error and preserve the local fine detail of the image. The proposed method has 7.8 million parameters, whereas large models, like DBBCS with 31.4M parameters and Wavelet with 17.5M parameters, illustrate a trade-off between performance and the number of parameters that highlights the efficiency of the proposed model.

Table 2. Represents the Evaluation Metrics for Proposed Methods with Respect to Various Existing Methods for Gaussian Noise of About 40 dB

Sl.No.	Model	PSNR	SSIM	MSE	MAE	Parameters
1.	GAN [4]	27.780	0.737	0.0013	0.0278	1,604,034
2.	Wavelet [8]	28.410	0.748	0.0014	0.0288	17,502,596
3.	DPS [9]	29.315	0.758	0.0012	0.0255	1,946,305
4.	SimMIM [10]	27.580	0.788	0.0017	0.0317	2,436,993
5.	IDDNET [11]	29.850	0.792	0.0011	0.0240	15,807,457
6.	DBBCS [12]	30.300	0.797	0.0009	0.0230	31,401,946
7.	Proposed Method	31.705	0.826	0.0007	0.0195	7,879,419

Figure 4 illustrates the denoised images from the proposed method and other existing methods at a Gaussian noise level of approximately 20 dB. Figure 4 also consists of two samples: (a) represents denoised images for Sample 1, and (b) illustrates denoised images for Sample 2. Each sample shows the original image, the noisy image with Gaussian noise, and the denoised images produced by the proposed method and other standard methods. In both samples, the noisy images are severely degraded, with high-intensity noise obscuring fine structural details. Most denoising methods recover some parts of the image, but they also introduce blurring artifacts. As a result, the tooth edges become less sharp, and the texture loses its fidelity. In contrast, the proposed method clearly outperforms existing methods in reconstruction quality, with noise almost completely removed. Moreover, the details are extremely fine, and the accentuated edges are more defined. Additionally, the proposed approach provides a better trade-off between the minimization of noise and the preservation of detail. Table 3 presents a comparative analysis of the proposed method and various state-of-the-art methods for images corrupted by Gaussian noise of approximately 20 dB. The evaluation metrics assess the capability of the proposed model to eliminate noise in images and generate denoised images that are most similar to the original images. Our proposed method achieves the highest PSNR value of 32.897, surpassing other competing methods such as DBBCS with 32.640 and DPS with 32.105. This implies that the offered approach yields superior quality images, as more original details are restored despite the presence of significant noise. Moreover, the SSIM of our proposed method is equal to 0.861, which is quite competitive with current models and allows us to conclude that both structural and perceptual features have been retained quite well. The value of the residual also demonstrates the great performance of our proposed method, which achieves a very low MSE of 0.0007 and an MAE of 0.0184, indicating that it can decrease the distance between original and denoised images at the pixel level while preserving texture information. An important aspect of this assessment is the quantity of parameters, which is an indicator of effectiveness in the model. As shown in Table

3, the evaluation metric values for DBBSC and the proposed method are nearly identical, but the proposed method uses far fewer parameters than the DBBSC model. Furthermore, it is able to reconstruct very high-quality images while utilizing only a moderate number of parameters.

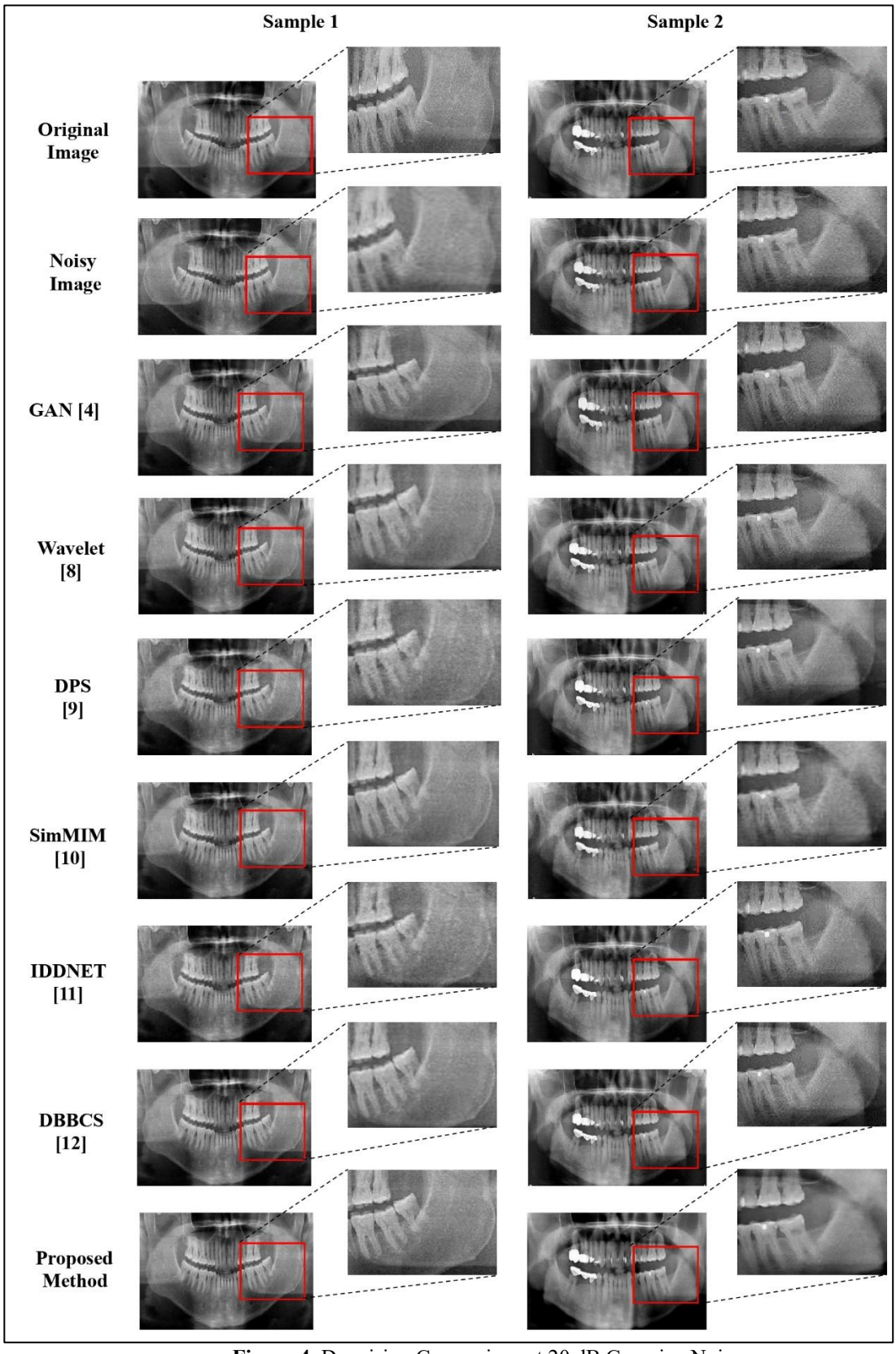

Figure 4. Denoising Comparison at 20 dB Gaussian Noise

Table 3. Demonstrates the Evaluation Metrics for the Proposed Methods Relative to Various Existing Methods for Gaussian Noise at About 20 dB

Sl.No.	Model	PSNR	SSIM	MSE	MAE	Parameters
1.	GAN [4]	28.490	0.822	0.0014	0.0292	1,604,034
2.	Wavelet [8]	31.350	0.856	0.0007	0.0209	17,502,596
3.	DPS [9]	32.105	0.860	0.0006	0.0187	1,946,305
4.	SimMIM [10]	30.950	0.893	0.0008	0.0217	2,436,993
5.	IDDNET [11]	31.872	0.848	0.0007	0.0196	15,807,457
6.	DBBCS [12]	32.640	0.857	0.0005	0.0182	31,401,946
7.	Proposed Method	32.897	0.861	0.0007	0.0184	7,879,419

Consequently, the proposed method is demonstrated to be efficient and effective, making it well-suited for situations where image restoration is needed even when the noise is as strong as Gaussian noise.

3.4 Ablation Study

The ablation study compared the effects of architectural variations, especially with or without attention mechanisms, in the UNet refiner. Table 4 demonstrates the case study of the proposed FR-GAN framework tested at a 20 dB Gaussian noise level. Performance is measured using standard metrics such as PSNR, SSIM, MSE, MAE, and the number of parameters. The baseline setup of the GAN generator, UNet refiner, and GAN discriminator achieves a PSNR of 32.1597 and an SSIM of 0.851, indicating reasonable denoising. When the PFFA is removed, performance degrades slightly, with the PSNR falling to 32.0249 and the SSIM to 0.848. This shows that PFFA helps to achieve better integration of multi-scale features and structural preservation. On the other hand, eliminating the LRLWA causes the performance to drop significantly, whereby the PSNR goes down to 30.8754 even though the SSIM of 0.8611 goes up slightly. Further, LRLWA is a significant factor in acquiring highly detailed local texture and in removing fine-scale noise. However, the proposed method, which integrates both PFFA and LRLWA into the U-Net refiner, produces the highest-quality image across all criteria, with a PSNR of 32.897, an SSIM of 0.861, an MSE of 0.0007, an MAE of 0.0184 and parameters of about 7,879,419.

Table 4. Represents the Case Study of the Proposed Method on Noisy Images with 20 dB Gaussian Noise

Sl. No	Model	PSNR	SSIM	MSE	MAE	Parameters
1.	GAN generator+ UNet refiner+GAN discriminator	32.159	0.851	0.0007	0.0215	7,775,579
2.	GAN generator+ UNet refiner without PFFA+GAN discriminator	32.024	0.848	0.0007	0.0209	7,551,619
3.	GAN generator+ UNet refiner without LRLWA+GAN discriminator	30.875	0.8611	0.0008	0.0217	7,156,099
4.	Proposed Method	32.897	0.861	0.0007	0.0184	7,879,419

4. Conclusion

In this paper, we have presented FR-GAN, which is a novel architecture for denoising noisy dental panoramic X-ray images corrupted by Gaussian noise. Moreover, progressive feature fusion attention (PFFA) is used for improving multi-scale feature aggregation via skip connections, while low-rank local window attention (LRLWA) is applied to refine local textures in the bottleneck. We have evaluated the performance of FR-GAN on 100 noisy dental panoramic radiographs corrupted by Gaussian noise with 20 dB and 40 dB intensities. Under

40 dB Gaussian noise, the performance of FR-GAN is evaluated as PSNR = 31.705 dB, SSIM = 0.826, with low reconstruction errors of 0.0007 for MSE and 0.0195 for MAE. Compared with the best baseline method, namely DBBCS, which has PSNR = 30.300 dB and SSIM = 0.797, FR-GAN improves PSNR by 1.405 dB and SSIM by 0.029 while maintaining low pixel-level reconstruction errors. In addition, under 20 dB Gaussian noise, FR-GAN has achieved PSNR = 32.897 dB, SSIM = 0.861, with low reconstruction errors of 0.0007 for MSE and 0.0184 for MAE. The number of parameters in the proposed model is approximately 7.88 million, which is much fewer than the parameters in the diffusion-based method, such as the 31.4 million parameters in the DBBCS method. This proves the efficiency of the model. FR-GAN achieves a proper trade-off between the accuracy of the model, the structure, and the efficiency, which makes the model applicable to medical image restoration tasks.

Acknowledgement

This study received funding from the Anusandhan National Research Foundation, Government of India, under project number ANRF/IRG/2024/000403/ENS.

References

[1] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative Adversarial Networks." Communications of the ACM 63, no. 11 (2020): 139-144.

[2] Zhou, Tao, Qi Li, Huiling Lu, Qianru Cheng, and Xiangxiang Zhang. "GAN Review: Models and Medical Image Fusion Applications." Information Fusion 91 (2023): 134-148.

[3] Nayak, Ankitha A., P. S. Venugopala, and B. Ashwini. "A Systematic Review on Generative Adversarial Network (GAN): Challenges and Future Directions." Archives of Computational Methods in Engineering 31, no. 8 (2024): 4739-4772.

[4] Zhang, Shufei, Zhuang Qian, Kaizhu Huang, Rui Zhang, and Amir Hussain. "Simplegan Stabilizing Generative Adversarial Networks with Simple Distributions." In 2019 International Conference on Data Mining Workshops (ICDMW), IEEE, 2019, 905-910.

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In International Conference on Medical image computing and computer-assisted intervention, Cham: Springer international publishing, 2015, 234-241.

[6] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770-778.

[7] Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-Image Translation with Conditional Adversarial Networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 1125-1134.

[8] Kang, Seung-Kwan, Si-Young Yie, and Jae-Sung Lee. "Noise2Noise Improved by Trainable Wavelet Coefficients for PET Denoising." Electronics 10, no. 13 (2021): 1529.

[9] Han, Young-Joo, and Ha-Jin Yu. "Self-Supervised Noise Reduction in Low-Dose Cone Beam Computed Tomography (CBCT) Using the Randomly Dropped Projection Strategy." Applied Sciences 12, no. 3 (2022): 1714.

[10] Almalki, Amani, and Longin Jan Latecki. "Self-Supervised Learning with Masked Image Modeling for Teeth Numbering, Detection of Dental Restorations, and Instance Segmentation in Dental Panoramic Radiographs." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2023, 5594-5603.

[11] Zhao, Xuzhi, Xinyi Wang, Yi Du, and Yahui Peng. "CBCT-IDDNet: A Three-Dimensional Res-Unet Based Image Domain Denoising Network for Clinical Dose Cone-Beam Computed Tomography—Winner of the International Conference on Acoustics, Speech, And Signal Processing-2024 Challenge." Quantitative Imaging in Medicine and Surgery 15, no. 10 (2025): 9844.

[12] Chen, Long, Changan Yuan, Huafu Xu, Ye He, and Jianhui Jiang. "Robust Denoising of Structure Noise Through Dual-Diffusion Brownian Bridge Modeling and Coupled Sampling." Electronics 14, no. 21 (2025): 4243.

[13] Qu, Zhong, Haoming Qu, Haonan Yin, Xuejuan Han, and Shufang Xia. "A Method of Object Detection Network with Progressive Feature Fusion and Reverse Attention for Traffic Scenes: Z. Qu et al." Signal, Image and Video Processing 19, no. 11 (2025): 942.

[14] Zhong, Xiwu, Yurong Qian, Hui Liu, Long Chen, Yaling Wan, Liang Gao, Jing Qian, and Jun Liu. "Attention_FPNet: Two-Branch Remote Sensing Image Pansharpening Network Based on Attention Feature Fusion." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021): 11879-11891.

[15] Guo, Qipeng, Xipeng Qiu, Xiangyang Xue, and Zheng Zhang. "Low Rank and Locality Constrained Self-Attention for Sequence Modeling." IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol, 27(12), 2019, 2213-2222.

[16] Zhang, Kai, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a Fast and Flexible Solution for CNN-based Image Denoising." IEEE Transactions on Image Processing 27, no. 9 (2018): 4608-4622.

[17] Zamir, Syed Waqas, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. "Restormer: Efficient Transformer for High-Resolution Image Restoration." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, 5728-5739.

[18] Kawar, Bahjat, Michael Elad, Stefano Ermon, and Jiaming Song. "Denoising Diffusion Restoration Models." Advances in neural information processing systems 35 (2022): 23593-23606.

[19] Krull, Alexander, Tim-Oliver Buchholz, and Florian Jug. "Noise2void-Learning Denoising from Single Noisy Images." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, 2129-2137.

[20] Dou, Zehao, and Yang Song. "Diffusion Posterior Sampling for Linear Inverse Problem Solving: A Filtering Perspective." In The Twelfth International Conference on Learning Representations. 2024.

[21] Dou, Zehao, and Yang Song. "Diffusion Posterior Sampling for Linear Inverse Problem Solving: A Filtering Perspective." In The Twelfth International Conference on Learning Representations. 2024.

[22] Iqbal, Md Tauhid Bin, Shamsul Alam Imon, Nurun Nahar, and Sung-Ho Bae. "DDD2Self: Self-Supervised Image Denoising with Dynamic Data-Driven Dropout." IEEE Access (2025).

[23] Gautam, Arnim, Aditi Pawar, Aishwarya Joshi, Satya Narayan Tazi, Sachin Chaudhary, Praful Hambarde, Akshay Dudhane, Santosh Vipparthi, and Subrahamanyam Murala. "Pureformer: Transformer-based Image Denoising." In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 1441-1449. 2025.

[24] Shi, Lemin, Xin Feng, Ping Gong, Dianxin Song, Hao Zhang, Langxi Liu, Yuqiang Zhang, and Mingye Li. "BDNet: A Real-Time Biomedical Image Denoising Network with Gradient Information Enhancement Loss." Biosensors 16, no. 1 (2026): 26.

[25] Osuna-Vargas, Pamela, Maren H. Wehrheim, Lucas Zinz, Johanna Rahm, Ashwin Balakrishnan, Alexandra Kaminer, Mike Heilemann, and Matthias Kaschube. "Denoising Diffusion Models for High-Resolution Microscopy Image Restoration." In 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, 2025, 4320-4330.

[26] Kim, Wonjin, Sun-Young Jeon, Gyuri Byun, Hongki Yoo, and Jang-Hwan Choi. "A Systematic Review of Deep Learning-Based Denoising for Low-Dose Computed Tomography from a Perceptual Quality Perspective." Biomedical Engineering Letters 14, no. 6 (2024): 1153-1173.

[27] Ameli, Nazila, Masoud Miri Moghaddam, Hollis Lai, and Camila Pacheco-Pereira. "Automated Quality Evaluation of Dental Panoramic Radiographs Using Deep Learning." Imaging science in dentistry 55, no. 2 (2025): 175.

[28] Lee, Ahhyun, Eunhyeok Hwang, and Dongsun Kim. "A Practical CNN–Transformer Hybrid Network for Real-World Image Denoising." Mathematics 14, no. 1 (2026): 203.

[29] Tian, Yingjie, Duo Su, Stanislao Lauria, and Xiaohui Liu. "Recent Advances on Loss Functions in Deep Learning for Computer Vision." Neurocomputing 497 (2022): 129-158.

[30] Xiao, Xianwei, and Baojiang Zhong. "Window Attention with Multiple Patterns for Single Image Super-Resolution." In 2023 IEEE 35th international conference on tools with artificial intelligence (ICTAI), IEEE, 2023, 731-738.