IoT-Enhanced GAN Framework for Multi-Modal Medical Image Enhancement with Residual Attention and Fusion

Janet Madhu*

*Associate Professor, School of Business, Woxsen University, India; janet.madhu@woxsen.edu.in

*Corresponding Author: janet.madhu@woxsen.edu.in

DOI: https://doi.org/10.30212/JITI.202604.005

Submitted: Nov. 24, 2025 Accepted: Jan. 29, 2026

ABSTRACT

Medical image enhancement is essential for improving diagnostic image quality and supporting clinical decision-making. With the increasing integration of the Industrial Internet of Things (IoT) in healthcare, real-time data from medical devices and sensors has become a key factor in enhancing multimodal medical images. In this paper, we propose a novel GAN-based model that integrates a residual attention mechanism, multi-scale convolutional networks, and adaptive upsampling modules. The model effectively enhances both local details and global structures in multimodal medical images, addressing the limitations of existing methods. We introduce a dual-discriminator design consisting of global and PatchGAN discriminators, to ensure the authenticity and quality of the generated images. Additionally, the integration of IoT data, such as real-time physiological monitoring, enables the model to dynamically adjust parameters for personalized and context-aware image enhancement. Experimental results on the BraTS and LIDC-IDRI datasets demonstrate that our model outperforms existing methods in terms of PSNR, SSIM, and visual quality. The results indicate that the model preserves fine details and maintains structural consistency, highlighting its potential for medical image enhancement. These findings suggest that the proposed approach could serve as a reliable solution for clinical imaging tasks, where high-quality image enhancement is critical for accurate diagnosis and treatment planning.

Keywords: Medical image enhancement, Multimodal imaging generative adversarial networks, Residual attention mechanism, Multi-scale convolutional networks, IoT

1. Introduction

Image enhancement methods play a significant role in modern healthcare system as they improve image quality and provide more accurate information for the diagnosis and treatment of diseases [1, 2]. In clinical medicine, medical imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET) —often suffer from low resolution, high noise levels, and poor contrast due to equipment constraints, imaging environments, and patient variability [3]. These limitations not only make lesion localization and analysis inaccurate

when performed by medical professionals, but also reduce diagnostic efficacy and increase the risk of misdiagnosis. Consequently, image enhancement has become an essential field of study in medical image processing to improve visual quality and detail, thereby aiding clinical diagnosis, treatment planning, and surgical navigation [4]. Historically, traditional image enhancement algorithms, including spatial filtering, histogram equalization, and frequency domain transformations have shown limited success in noise reduction and contrast enhancement [5]. Nevertheless, these approaches are normally designed for single-modality images and they lack the capability to address the complexities inherent in complex medical imaging. Furthermore, as clinical requirements for diagnostic accuracy grow, single-modality image enhancement, is increasingly insufficient [6, 7]. Applying a single modality in most complex cases may fail to provide a clear picture of lesions or their functionality. Therefore, multimodal medical image integration and enhancement, integrated with IoT technologies, has become an essential topic of research. Multimodal medical imaging technology integrates complementary data from multiple modalities, including MRI and CT, to facilitate a more holistic understanding of disease diagnosis [8]. For example, MRI is ideal for highlighting soft tissue contrast, which allows for the accurate definition of lesion morphology, whereas CT enables high-resolution visualization of dense structures, such as bones. The integration and improvement of multimodal images allow for the simultaneous display of anatomical and functional characteristics, providing clinicians with more comprehensive and detailed information. In addition, IoT devices are crucial to this advancement and provide real-time data on the physiological condition or mobility of the patient, which dynamically influences the parameters of the enhancement model. This allows the enhancement process to be adjusted according to the patient's specific medical needs [9, 10]. As Figure 1 demonstrates, healthcare systems have incorporated with IoT technologies to gather and analyze medical data, improving decision-making and patient monitoring.

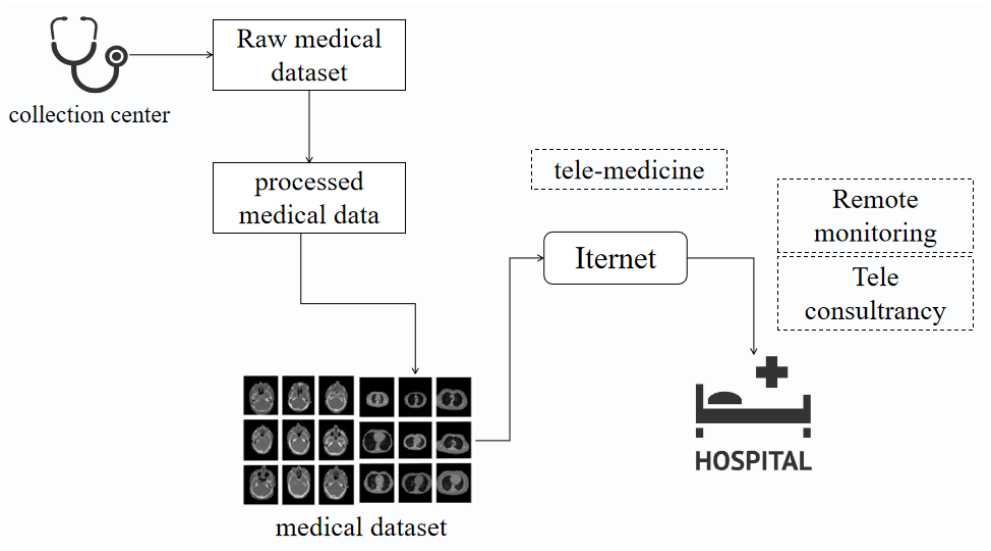

Figure 1. Illustration of iot in healthcare: collecting medical data from various sources to enable remote monitoring, telemedicine, and teleconsultation.

Generative Adversarial Networks (GANs) and deep learning technologies have emerged as transformative tools in medical image processing, offering effective solutions for medical image enhancement and multimodal image fusion [11]. Operating on an adversarial framework, GANs consist of a generator and a discriminator that function collaboratively to extract high-dimensional features and produce high-fidelity images. GANs have demonstrated impressive performance in super-resolution, denoising, and contrast enhancement in single-modality medical imaging tasks, showing remarkable success [12]. Inspired by this success, we integrate real-time sensor data from the Internet of Things (IoT) into the GAN model. This integration allows for the dynamic optimization of essential parameters, such as the image resolution and enhancement strength based on the patient physiological data (e.g., heart rate, blood oxygen level and mobility indicators). Such a personalized method can significantly increase the precision and efficacy of medical image evaluation. However, utilizing GANS to enhance multimodal medical images remains challenging. The major challenges include the efficient extraction and integration of complementary information across multiple modalities, maintaining global consistency, preserving local details, and avoiding the generation of artifacts. It is imperative to overcome these obstacles to realize the potential of GANs in multimodal medical imaging, which will make it possible to create more successful and effective diagnostic tools. To address these challenges, we propose a novel multimodal medical image enhancement method based on Generative Adversarial Networks (GAN). This approach utilizes a generator that incorporates a residual attention mechanism and multi-scale convolutional networks. This design enables the model to focus on critical image regions, such as tumors or vital organs, and improves its ability to retain texture details by dynamically collecting multi-scale features. Furthermore, we developed a hybrid discriminator that combines PatchGAN, to preserve local details, with a global discriminator, to maintain overall anatomical structural integrity. We also introduce a multi-level loss function consisting of perceptual loss, Structural Similarity Index Measure (SSIM) loss, and pixel-wise L1 loss — to ensure that the quality of generated images is maximized across multiple dimensions. Moreover, the use of the IoT data may affect the parameter corrections of the model, so that the enhancement of the image became individual and suited to the clinical environment of the individual. This integration of IoT data assists in optimizing the model performance with greater precision and clinical relevance of the improved medical images.

The primary contributions of this paper are as follows:

- This paper proposes a generator structure that combines a residual attention mechanism with multi-scale convolutional networks, dynamically emphasizing the critical regions in medical images, such as tumors and significant organs, to effectively enhance detail representation and texture quality.
- This paper designs a hybrid discriminator architecture that integrates the advantages of PatchGAN and global discriminators to simultaneously optimize local detail consistency and global anatomical structural integrity in generated images.
- This paper introduces a comprehensive multi-level loss function, leveraging perceptual loss,

SSIM loss, and pixel-wise L1 loss to ensure the visual quality and diagnostic relevance of the enhanced images across multiple dimensions.

2. Literature Review

2.1 Traditional Medical Image Enhancement Method

Traditional medical image enhancement methods have long served as a cornerstone of medical image processing, aiming to optimize image quality and provide clearer and more accurate information for clinical diagnosis and treatment planning. These methods address various aspects of image quality improvement — including detail preservation, noise suppression, and contrast enhancement— and have achieved significant results across these domains.

Among traditional medical image processing techniques, local enhancement methods play a pivotal role by focusing on the optimization of brightness and contrast within specific image regions. Adaptive contrast enhancement, for example, dynamically modifies contrast according to the local brightness distribution, effectively highlighting lesion boundaries and structures [12]. This capability makes it highly effective for processing low-contrast CT and X-ray images. Simultaneously, brightness adjustment algorithms, like dynamic range compression, aim to rectify overexposure and underexposure issues, thereby enhancing the overall visual quality of medical images [13, 14]. Texture enhancement and detail optimization are equally critical in traditional approaches. Median filtering and Gaussian filtering, two widely used noise suppression techniques, efficiently reduce noise in medical images. An advanced variant, bilateral filtering, considers both pixel spatial proximity between pixels and grayscale similarity, enabling effective background noise removal while preserving essential edge details[15]. While these methods demonstrate high robustness in handling ultrasound and MRI images, they occasionally encounter difficulties in retaining complex tissue structures and lesion features. Global optimization methods grounded in the grayscale distribution also see broad applications. Histogram equalization (HE) redistributes grayscale values to enhance overall contrast, while Adaptive Histogram Equalization (AHE) refines this process by focusing on local regions to improve detail visibility in low-contrast scenarios [16, 17]. However, both techniques carry the risk of introducing artifacts or over-enhancing certain areas, particularly in high-dynamic-range images. Light model-based enhancement, exemplified by the Retinex theory, has drawn increasing attention. By emulating the human eye's light adaptation mechanisms, Retinex methods mitigate environmental light interference to boost contrast while maintaining texture details [18]. Nevertheless, when applied to images containing intricate tissue structures, these methods often struggle to simultaneously maintain boundary sharpness and global consistency.

While traditional methods have achieved notable success in improving medical image quality, they often rely on manually defined parameters, which limits their adaptability. Furthermore, these methods exhibit constrained capabilities when dealing with multimodal data and complex imaging scenarios, making it difficult to meet the rigorous demands of modern medical imaging. Although the emergence of deep learning technologies has mitigated many of these limitations, traditional approaches continue to play a foundational role in initial image processing.

2.2 Applications of Generative Adversarial Networks in Medical Imaging

Since their inception, Generative Adversarial Networks (GANs) have demonstrated wide-ranging applications in the field of medical image processing. Through the adversarial learning mechanism between the generator and the discriminator, GANs identify complex feature relationships within datasets, achieving significant advancements in tasks such as super-resolution reconstruction, denoising, modality conversion, and multimodal fusion of medical images [19]. Additionally, numerous studies have proposed improved GAN-based architectures, providing more effective solutions for medical imaging tasks.

In the field of medical image super-resolution reconstruction, the classic SRGAN has demonstrated efficacy in generating high-resolution images by combining perceptual loss and adversarial loss. Researchers have modified SRGAN to address the specific characteristics of medical images, such as introducing residual networks and multi-scale feature extraction modules, enhancing the model's ability to capture fine-grained details [20]. In super-resolution tasks involving MRI and CT images, these methods substantially increase spatial resolution and boundary definition, rendering images more suitable for clinical diagnostic scenarios [21]. Moreover, GAN-based denoising techniques have emerged as a prominent research focus. Denoising GAN utilizes a generator to produce noise-free images, while the discriminator optimizes output fidelity [22]. To address the high-frequency noise in ultrasound images, several studies have combined convolutional autoencoders with adversarial learning frameworks to effectively denoise data while preserving clinically relevant medical information [23]. Further enhancements include incorporating attention mechanisms to dynamically emphasize regions affected by severe noise, improving the efficiency and quality of the denoising process [24]. In modality conversion, CycleGAN provides crucial support for cross-modality image generation. CycleGAN achieves unsupervised modality conversion through cyclic consistency constraints and has been successfully applied to tasks such as CT-to-MRI image conversion [25]. Researchers have refined the CycleGAN structure by integrating Conditional GANs (cGAN) and adaptive loss functions. This integration enables superior feature alignment and conversion accuracy across different modalities, assisting in tumor detection and organ segmentation tasks that rely on joint analysis of multimodal data. Multimodal image fusion is another domain where GAN technology is highly effective. By combining complementary information from various modalities into a unified image, fusion methods offer more comprehensive clinical diagnostic support. GAN-based multimodal fusion approaches utilize the generator to jointly model cross-modal features and the discriminator to ensure the consistency and detail quality of generated images [26, 27]. To further refine these methods, multi-scale convolutional modules and feature separation techniques have been introduced to effectively preserve key information from the original images. Consequently, the integration of feature extraction and detail enhancement has emerged as a focal point in multimodal image fusion research, facilitating the generation of fused images that more closely align with clinical requirements [28, 29].

Beyond the basic GAN framework, various enhanced variants have been tailored to the specific challenges of medical imaging. For instance, Attention GAN dynamically highlights clinically

significant regions, improving the precision of image generation and enhancement [7]. In contrast, Multi-Scale GAN captures image features across multiple spatial resolutions, optimizing image quality both globally and locally. Conditional GAN (cGAN) incorporates conditional constraints during the generation process, allowing for customized image creation [17, 30]. These advanced models, each tailored to address the unique characteristics of different medical imaging tasks, have significantly broadened the application range and practicality of GANs in the medical field.

GANs and their enhanced variants have played a vital role in tasks such as super-resolution reconstruction, denoising, modality conversion, and multimodal fusion of medical images. Through technological innovations, these methods have demonstrated significant capabilities in detail preservation, feature extraction, and modality information fusion. The continuous progress in research provides robust theoretical and technical support for multimodal medical image enhancement tasks and establishes a solid foundation for future studies.

3. Method

3.1 Applications of Generative Adversarial Networks in Medical Imaging

The paper proposes a multimodal medical-image enhancement approach, aimed at improving the previously described strategy through Generative Adversarial Networks (GANs). This framework focuses on elevating medical imaging quality, particularly for low-to-high resolution super-resolution tasks. The architecture integrates a residual attention mechanism, multi-scale convolutional networks and a hybrid multi-scale discriminator. This design is intended to balance local feature-extraction with global structural regularity in medical images, and it employs the Internet of Things (IoT) technologies when adapting to the specific conditions of the patient. This process does not only retain important structural information but also enhances the quality of the image. In the meantime, IoTs devices such as wearable sensors provide real-time patient information, allowing the attention mechanism to dynamically focus on the most meaningful parts of the anatomy, necessitated by anatomical complexity. Furthermore, contextual health information from IoT-enabled systems facilitates refined feature extraction aligned with the patient's physiological state. To convert low-resolution images into high-resolution outputs, a high-fidelity upsampling system is employed to improve textures and details while minimizing artifact generation. Real-time data, such as heart rate and temperature indications, are utilized to automatically optimize the upsampling process, ensuring customized and precise results. The discriminator employs a hybrid design integrating PatchGAN for local area assessment and a global discriminator to maintain anatomical consistency across the entire image. With the incorporation of IoT data, the system can better maintain real-time anatomical accuracy during image enhancement. We use loss function suite comprising of perceptual loss (learned VGG network), structural similarity index (SSIM) loss, and L1 loss. All these losses guarantee that the images produced are visually realistic and medically accurate with anatomical features maintained. The loss functions may further be refined based on the IoT data to meet the specific medical requirements of these individual patients. A combination of such methods and the IoT technologies results in an effective multimodal approach to medical image enhancement. This

solution enhances accuracy in diagnosis and clinical decision-making, and the added advantage is that it fits to real-time information of patients. Our model has a total architecture represented in Figure 2 and the different modules and how they relate and operate is explained in Figure 2.

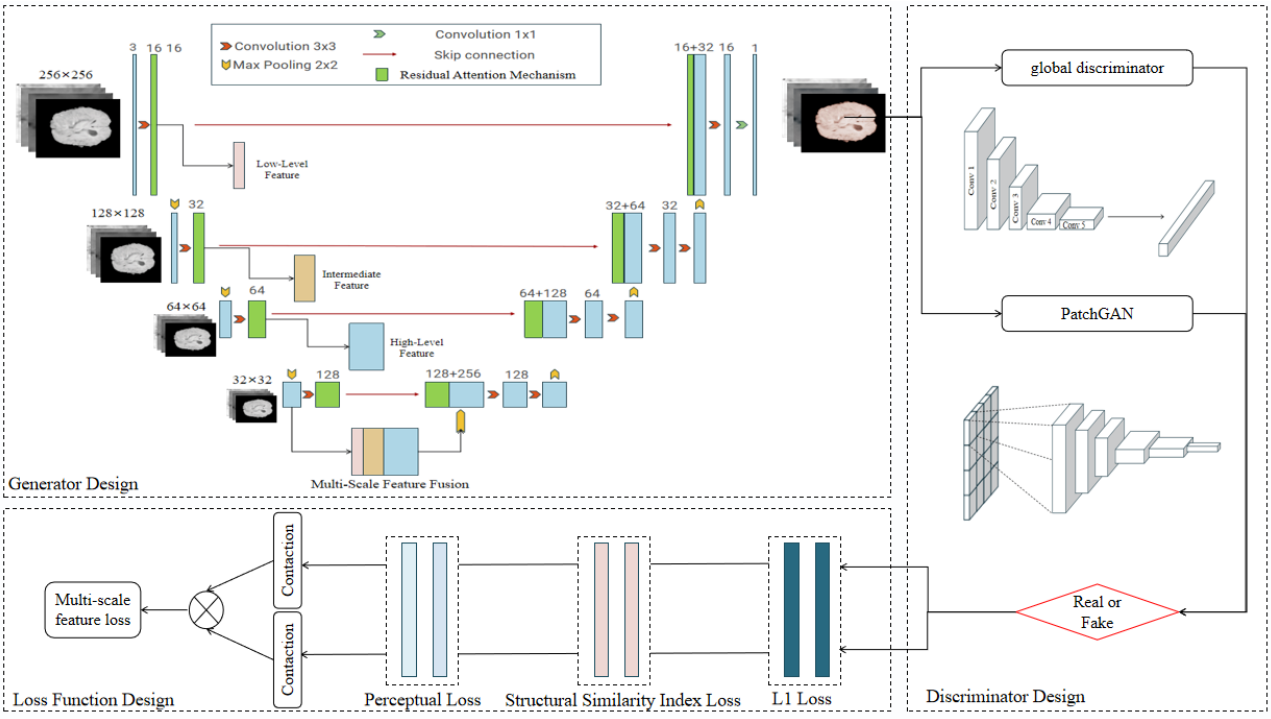

Figure 2. Overall, architecture of the proposed model, including the generator, discriminator, and loss function designs. The generator utilizes multi-scale feature fusion and a residual attention mechanism to enhance the resolution and detail of medical images. The discriminator consists of a global discriminator and a PatchGAN module, ensuring both local and global consistency in generated images. The loss function integrates multi-scale feature loss, perceptual loss, structural similarity index loss, and L1 loss to optimize the model for high-quality image generation.

3.2 Generator Design

In this study, the generator design combines Residual Attention Mechanism, Multi-Scale Convolutional Networks, and Adaptive Upsampling Modules. These modules not only optimize existing techniques but also introduce novel components to better handle the complexities of medical imaging. Each module contributes critically to detail preservation and image quality enhancement, with the ultimate goal of transforming low-resolution medical images into high-resolution outputs, while maintaining essential structural details.

3.2.1 Residual attention mechanism

The first key component of the generator is the Residual Attention Mechanism. This mechanism combines residual learning with attention-based weighting to preserve information across deep network layers. This mechanism introduces residual connections that guarantees efficient information flow in deep networks and mitigate gradient vanishing issues. Meanwhile, the attention mechanism adaptively assigns varying weights to different pixel regions based on clinical importance, such as

focusing on tumors or organ boundaries. The structure of the Residual Attention Mechanism is illustrated in Figure 3.

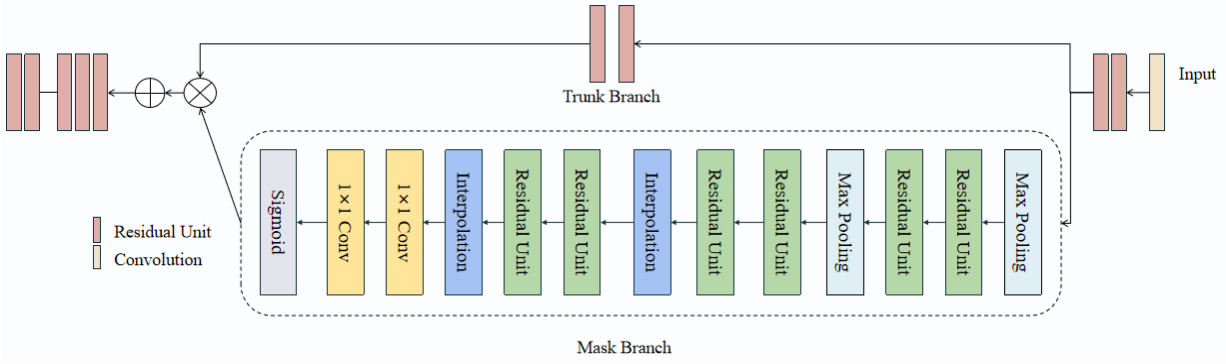

Figure 3. The architecture of the Residual Attention Mechanism.

The output after applying the Residual Attention Mechanism can be described as:

$$Y_i = X_i + A_i \cdot (W_i * X_i) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 1]}$$

Where X_i is the input feature map, A_i is the adaptive attention weight map learned by the network, W_i is the convolutional kernel, and * represents the convolution operation. The attention map A_i is computed as follows:

$$A_i = \sigma\left(\text{Conv}(X_i)\right) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 2]}$$

Where Conv (·) is the convolution operation, and σ is the sigmoid activation function, which ensures that the attention weight map is within the range of [0,1]. this allows the prioritization of important regions in the image. The second core module of the generator is the Multi-Scale Convolutional Network, which is designed to extract features across different scales. Medical images typically contain structural information at various scales, such as different details in bones and soft tissues, which are best captured at different resolutions. Multi-scale convolutional networks apply convolution operations with multiple kernel sizes, enabling the extraction of hierarchical feature representations from the image to ensure that both local and global structural details are preserved. For the k-th convolutional layer, the output feature map X_k is computed as:

$$X_k = \sum_{i=1}^{N_k}(W_i * X_{k-1}) + b_k \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 3]}$$

Where W_i is the i-th convolutional kernel at the k-th layer, N_k is the number of kernels used in the k-th layer, X_{k-1} is the input feature map from the previous layer, and b_k is the bias term. The multi-scale convolution operation enables the network to capture image features at varying receptive fields, improving the overall quality of the generated images.

To effectively fuse the features extracted at different scales, the network includes a scale fusion module, which aggregates the feature maps from different scales. The fused feature map X_{fused} can be expressed as:

$$X_{fused} = \sum_{k=1}^{K} \lambda_k \cdot X_k \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 4]}$$

Where λ_k is the fusion weight for the k-th scale, X_k is the feature map at the k-th scale, and

X_{fused} is the resulting fused feature map. This weighted sum enables the model to balance the contributions from each scale, ensuring that both local details and global structures are preserved in the generated image.

3.2.2 Upsampling and adaptive modules

To transform low-resolution images into high-resolution outputs, the generator incorporates an Upsampling Module, which increases the spatial resolution of the input image. The upsampling operation is implemented using transpose convolution (also known as deconvolution), which not only enlarges the image resolution while preserving image details. Let X represent the low-resolution input image, and $\hat{Y}$ denote the upsampled high-resolution output, the upsampling process is described as:

$$\hat{Y} = \text{TransConv}(X, W_{upsample}) + b_{upsample} \cdots\cdots\cdots \text{[Formular 5]}$$

Where $W_{upsample}$ is the convolutional kernel used in transpose convolution, $b_{upsample}$ is the bias term, and $\text{TransConv}(\cdot)$ represents the transpose convolution operation. This operation enlarges the low-resolution input, restoring high-frequency details in the process. In addition to upsampling, the generator includes an Adaptive Module, which modulates the strength of detail enhancement according to the content of the image. In medical images, different regions may require varying levels of detail enhancement, particularly in critical areas such as tumors or organs. The adaptive module dynamically regulates the enhancement based on the importance of each region. The enhanced output image $Y_{enhanced}$ is expressed as:

$$Y_{enhanced} = X + \alpha \cdot (F(X)) \cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 6]}$$

Where α is the adaptive coefficient, $F(X)$ represents the enhancement operation, and $Y_{enhanced}$ denotes the enhanced output image. The adaptive module ensures that important regions receive more detailed enhancement, while the rest of the image maintains a balance of quality.

3.3 Discriminator Design

The discriminator is designed to evaluate the authenticity of the generated images, thereby helping the generator toward improved performance. In this study, we propose a hybrid discriminator structure is proposed that integrates PatchGAN and global discriminators to ensure the optimization of local details and the maintenance of global structural consistency. Figure 4 presents the overall architecture of the discriminator, highlighting the interactions and functional roles of each component.

In traditional Generative Adversarial Networks (GANs), the discriminator typically outputs a global judgment regarding the entire image. However, in PatchGAN, the discriminator evaluates each local patch of the image. By adopting this strategy, PatchGAN can more thoroughly evaluate the local structure of the image, which improves the quality of the generated image' s details. In this design, each patch is treated as a binary classification task, and the output of the discriminator is the probability of the patch being "real." Given the input image $\hat{Y}$, the output for each patch i can be expressed as:

$$D_{patch}(\hat{Y}_i) = \sigma(W_D * \hat{Y}_i + b_D) \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 7]}$$

where $\hat{Y}_i$ represents the i-th patch, W_D is the discriminator's convolution kernel, $*$

represents the convolution operation, b_D is the bias term, and σ is the sigmoid activation function that outputs the probability that the patch is real. The core structure of the discriminator is a deep convolutional network that processes the image in layers, extracting multi-level features. In each convolution layer, the input feature map $\hat{Y}$ undergoes a series of convolution operations to derive higher-order image features. Let the input to the k-th layer be X_{k-1}, then the output feature map X_k of the k-th layer can be expressed as:

$$X_k = Conv_k(X_{k-1}, W_k) + b_k \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 8]}$$

where W_k is the convolution kernel at the k-th layer, b_k is the bias term, and $Conv_k(\cdot)$ represents the convolution operation. Each layer extracts different levels of image features, from low-level edge information to high-level semantic representations, progressively building a global understanding of the image. To ensure that the discriminator evaluates both fine-grained details and overall structural integrity, a global discriminator is introduced at the final stage. The global discriminator ensures that the generated image not only retains fine details but also maintains global consistency. The features of all local patches are aggregated and passed to the global discriminator to compute the overall authenticity score. The global judgment $D_{global}(\hat{Y})$ can be computed as:

$$D_{global}(\hat{Y}) = \sigma\left(FC\left(Flatten\left(\sum_{i=1}^{N} X_i\right)\right)\right) \cdots\cdots\cdots\cdots \text{[Formular 9]}$$

where $FC(\cdot)$ represents the fully connected layer operation, $Flatten(\cdot)$ refers to flattening all patch feature maps into a one-dimensional vector, and $\sum_{i=1}^{N} X_i$ denotes the sum of all patch feature maps. This design enables the discriminator to make a global judgment based on both local and global information. In the hybrid discriminator structure, the evaluation of local patches and global consistency are conducted in parallel paths. The local patches are evaluated using PatchGAN, while the global structure is further optimized by the global discriminator. This design allows the discriminator to focus on both the details and the global structure of the image, thus improving the quality and consistency of the generated image.

$$D_{final}(\hat{Y}) = \alpha D_{patch}(\hat{Y}) + \beta D_{global}(\hat{Y}) \cdots\cdots\cdots\cdots \text{[Formular 10]}$$

where α and β are weighting coefficients used to control the contribution of the local patch discriminator and the global discriminator, respectively. The discriminator can achieve a flexible balance between local detail and global consistency by adjusting these coefficients, thereby enhancing the overall quality of generated images. The architecture integrates local detail evaluation with global consistency assessment to determine the authenticity of generated images. Through the optimization of these two aspects, the discriminator ensures that the final output not only preserves fine details but also maintains global structural coherence, a crucial requirement for medical image generation. To further optimize the performance of the discriminator, the authenticity of the generated image is evaluated by averaging the results of all patches:

$$D_{authenticity}(\hat{Y}) = \frac{1}{N}\sum_{i=1}^{N} D_{patch}(\hat{Y}_i) \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 11]}$$

This represents the average evaluation of all patches, allowing for an overall authenticity score while still preserving detailed patch-level assessments.

Through the introduction of the hybrid discriminator architecture, a balanced optimization between local detail fidelity and global structural coherence is achieved, ensuring that the generated medical images are optimized in detail and maintain global consistency. This design is critical for medical image generation, as the accurate restoration of details (e.g., boundaries of tumors, shapes of organs) and global structure (e.g., organ positioning and proportions) are equally essential. The hybrid discriminator structure improves the quality of generated images and contributes to advancing the automation and precision of medical image analysis.

3.4 Loss Function Design

The loss function is formulated to ensure that the generator produces realistic images while preserving the details and structure, especially in tasks such as super-resolution and multi-modal fusion for medical imaging. To achieve this objective, several loss functions were designed, including Perceptual Loss, Structural Similarity Index (SSIM) Loss, and L1 Loss, which collectively guide the generator toward producing high-quality images.

Perceptual Loss is a feature-based loss function that measures the discrepancy between the generated and real images in a high-level feature space. Traditional pixel-based losses like L2 loss often overlook the semantic information in images, whereas Perceptual Loss facilitates the preservation of image details and structure by calculating the difference in high-level features extracted from a pre-trained network such as VGG-19. Let $\phi(x)$ be the features extracted by the pre-trained network, the Perceptual Loss can be expressed as:

$$\mathcal{L}_{\text{perceptual}} = \sum_l \left\| \phi_l(\hat{Y}) - \phi_l(Y) \right\|_2^2 \cdots\cdots\cdots\cdots\cdots \text{[Formular 12]}$$

where $\phi_l(\cdot)$ represents the features extracted from layer l, $\hat{Y}$ is the generated image, and Y is the real image. This loss function enables the generator better capture the global structure and details of the image by comparing the high-level feature maps. Structural Similarity Index (SSIM) Loss is a loss function specifically designed for image quality evaluation, measuring similarity in terms of luminance, contrast, and structural information. Compared to traditional L2 loss, SSIM better reflects the structural information of the image, which is essential in medical image generation tasks. SSIM Loss can be expressed as:

$$\mathcal{L}_{\text{SSIM}} = 1 - \frac{\left(2\mu_{\hat{Y}}\mu_Y + C_1\right)\left(2\sigma_{\hat{Y}Y} + C_2\right)}{\left(\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1\right)\left(\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2\right)} \cdots\cdots\cdots\cdots \text{[Formular 13]}$$

where $\mu_{\hat{Y}}$ and μ_Y are the mean values of the generated and real images, $\sigma_{\hat{Y}}^2$ and σ_Y^2 are their variances, $\sigma_{\hat{Y}Y}$ is the covariance between the generated and real images, and C_1 C_1 and C_2 are constants added for stability. The goal of SSIM Loss is to maximize the similarity between the generated and real images in terms of luminance, contrast, and structure. Third, L1 Loss is used to enforce pixel-level consistency between the generated and real images. L1 Loss avoids the blurring effect that may occur with L2 Loss and helps preserve more details in the generated image. L1 Loss computes the absolute difference between the generated and real images, and it can be expressed as:

$$\mathcal{L}_{L1} = \sum_i \left|\hat{Y}_i - Y_i\right| \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 14]}$$

where $\hat{Y}_i$ and Y_i are the pixel values of the generated and real images, respectively. L1 Loss encourages the preservation of fine-grained details by minimizing the pixel-level differences, making it particularly useful for tasks that require fine detail restoration, such as medical image super-resolution. Additionally, Adversarial Loss is utilized to optimize the generator and discriminator through adversarial training. In the GAN framework, Adversarial Loss forms the core competitive mechanisms between the generator and discriminator. The generator targets the generation of images that are indistinguishable from real ones, while the discriminator attempts to distinguish real images from generated ones. The generator's Adversarial Loss is expressed as:

$$\mathcal{L}_{GAN} = -E_{\hat{Y}\sim p_{data}}\left[\log D(\hat{Y})\right] - E_{\hat{Y}\sim p_{gen}}\left[\log\left(1 - D(\hat{Y})\right)\right] \cdots\cdots\cdots \text{[Formular 15]}$$

where $D(\hat{Y})$ represents the discriminator's estimated probability of the generated image being real, and p_{data} and p_{gen} represent the distributions of real and generated images, respectively. Through this adversarial training, the generator is iteratively improved to generate more realistic images, making them closer to real images.

In summary, the proposed loss function framework provides comprehensive guidance for optimal image quality, ensuring high consistency in both the details and structure of the image. Perceptual Loss and SSIM Loss assist the generator in capturing high-level features and structural information, L1 Loss ensures pixel-level consistency, and Adversarial Loss improves the realism of the generated images through adversarial training. These loss functions function in unison to enable the production of medical images with superior quality and detailed hierarchical feature representations.

4. Experiments

4.1 Datasets

In this study, we selected two publicly available medical image datasets: Brain Tumor Segmentation (BraTS) and Lung Image Database Consortium Image (LIDC-IDRI), which are suitable for brain tumor segmentation lung nodule detection and image super-resolution tasks. These datasets are widely used in medical image analysis and provide rich annotation data, offering an effective benchmark for performance evaluation across multiple tasks.

The BraTS dataset (Brain Tumor Segmentation Challenge) contains multimodal brain MRI images, including T1-weighted, T2-weighted, Fluid Attenuation Inversion Recovery (FLAIR), and post-contrast T1-weighted (T1ce) sequences. This dataset is particularly challenging for multimodal image fusion and tumor segmentation, providing research opportunities for automated brain tumor detection and analysis. For this experiment, we utilized the 2021 version, which contains multimodal MRI images from approximately 300 patients. It encompasses a diverse range of tumor lesions, rendering it suitable for super-resolution reconstruction, image generation, and segmentation tasks.

The Lung Image Database Consortium Image (LIDC-IDRI) dataset contains chest CT scans for

the detection and analysis of lung nodules. It includes more than 1,000 patient CT scans from various medical institutions, accompanied by expert annotations regarding nodule location, size, shape, and malignancy. The main task is automatic pulmonary nodule detection and segmentation. The dataset is well suited for super-resolution tasks, lung nodule detection, and image enhancement tasks, and its diverse nodule characteristics and relatively low-resolution CT images make it ideal for super-resolution reconstruction.

These datasets have a wide range of application values and are suitable for medical image super-resolution, image generation, multi-modal fusion and segmentation. Each dataset provides a large amount of labeled data that helps us comprehensively evaluate the performance of the proposed model through quantitative and qualitative analysis and verify its validity in real-world tasks.

4.2 Data Preprocessing

In our experiment, we conducted crucial preprocessing on the BraTS and LIDC-IDRI datasets to prepare the input data for the proposed super-resolution and image generation model.

For the BraTS dataset, which contains multi-modal MRI images (T1, T2, FLAIR, and post-contrast T1), we standardized the spatial resolution of all modalities first. This step ensured alignment and consistency among the different imaging sources. Then, using high-quality interpolation methods, we resized each modality to 256×256 pixels. This process maintained structural details while facilitating effective multi-modal image fusion and super-resolution tasks. To create data for the super-resolution task, the original high-resolution images were downsampled to generate corresponding low-resolution inputs, which served as inputs, while the original high-resolution images were set as the generator's target outputs. Additionally, to enrich training set diversity and improve the model's ability to learn hierarchical feature representations of complex anatomical structures, data augmentation techniques— including rotation, scaling, translation, and flipping— were applied.

Regarding the LIDC-IDRI dataset, we processed chest CT scans by slicing the 3D images and cropping each CT slice to a uniform 256×256 pixels. Pixel intensity values were normalized to the range of 0 to 1, standardizing the data distribution across the dataset. Similar to the BraTS dataset, low-resolution CT images were generated through downsampling, with the corresponding high-resolution CT slices acting as the target outputs.

To ensure accurate results in multi-modal image fusion and super-resolution, especially for the multi-modal MRI images of the BraTS dataset and the CT slices of LIDC-IDRI, image registration techniques were achieved to achieve precise spatial alignment. Moreover, all images from both datasets were normalized during preprocessing, ensuring consistency in scale and pixel intensity distribution. These comprehensive preprocessing steps were essential for dataset preparation in training and testing, enabling our model to generate high-quality images with improved resolution, structural consistency, and visual detail.

4.3 Experimental Setup

The experiments were conducted in a high-performance computing environment utilizing an

NVIDIA A100 GPU equipped with 16GB of memory to accelerate the training and processing of large-scale data sets. The Phyton programming language was employed, alongside with TensorFlow and PyTorch, for implementing and training the GAN models. PyTorch handled model construction and training, while TensorFlow assisted with image processing tasks. This setup facilitates efficient computation and parallel processing during training, enabling large-scale experiments to be completed in a reasonable amount of time.

For model optimization, the Adam optimizer was used with a learning rate of 0.0002, and an adaptive learning rate adjustment strategy was adopted. A batch size of 32 was selected to balance training efficiency and memory usage; training proceeded for 500 epochs to ensure adequate learning of hierarchical feature representations and structures. An early stopping mechanism was implemented to prevent overfitting; if the loss did not decrease significantly for 20 consecutive epochs, training was terminated. This strategy helps avoid overtraining and conserves computational resources while ensuring the model's ability to generalize. Loss and performance metrics were monitored after each epoch to facilitate early stopping. The data set was divided into a training set (70\%), a verification set (15\%) and a test set (15\%). The training set was used for parameter updates, the validation set for hyperparameter tuning and overfitting prevention, and the test set for final model evaluation. This division ensures a fair and scientific training and evaluation process. GPU acceleration was used throughout the training process, and fixed random seeds were applied in all experiments to maintain consistency in data segmentation and training.

This experimental setup maximized training efficiency and stability, leveraging hardware resources effectively while ensuring reliable model performance evaluation.

4.4 Evaluation Metrics

In this experiment, we used various metrics to assess the performance of the proposed GAN-based multimodal medical image enhancement method. For image quality, Peak Signal-to-Noise Ratio (PSNR)and Structural Similarity Index Measure (SSIM) were utilized. PSNR measures the pixel-level difference between the generated and ground-truth images, where a higher PSNR indicates superior quality. SSIM evaluates brightness, contrast, and structure, with values closer to 1 indicating higher structural and perceptual similarity. Furthermore, Normalized Cross-Correlation (NCC) and Mutual Information (MI) were employed to assess the correlation and information content between the generated and real image, respectively. These metrics facilitate the evaluation of structural information preservation. To measure the model's ability to recover image detail, Spatial Frequency (SF) and Mean Gradient (MG). SF reflects image detail density, while MG measures edge sharpness. The application of these metrics demonstrates whether the generation process restores fine detail and improves visual quality.

By integrating these evaluation metrics, the performance of the proposed model in image super-resolution, multimodal image fusion, and enhancement tasks was comprehensively analyzed from multiple perspectives and dimensions. These metrics not only facilitated a quantitative analysis of image quality but also provided a foundation for the further optimization and refinement of the architecture.

4. Results and Discussion

4.1 Comparison with State-of-the-Art Methods

Table 1. Performance comparison of ablation models on brats and lidc-idri datasets

Method	BraTS		LIDC-IDRI	
	PSNR	SSIM	PSNR	SSIM
FGAN[31]	27.8 ± 1.5	0.80 ± 0.05	29.5 ± 1.4	0.84 ± 0.03
FS-GAN[32]	25.9 ± 1.2	0.75 ± 0.06	27.2 ± 1.1	0.78 ± 0.05
SAG-GAN[33]	24.7 ± 1.1	0.73 ± 0.07	26.3 ± 1.3	0.76 ± 0.07
MedGAN[34]	26.2 ± 1.3	0.77 ± 0.05	28.1 ± 1.2	0.80 ± 0.04
U-Patch GAN[35]	25.8 ± 1.2	0.74 ± 0.06	27.5 ± 1.3	0.79 ± 0.05
CNN-FBB[36]	24.6 ± 1.2	0.72 ± 0.07	26.8 ± 1.4	0.77 ± 0.06
Ours	28.5 ± 1.0	0.84 ± 0.04	30.2 ± 1.1	0.88 ± 0.03

Table 1's results incontrovertibly showcase the efficacy of our proposed method in multimodal medical image enhancement. On the BraTS dataset, the approach attained a peak PSNR of 28.5 ± 1.0 and SSIM of 0.84 ± 0.04, surpassing the second-best performing model, FGAN, by 0.7 dB in PSNR and 0.04 in SSIM. Similarly, for the LIDC-IDRI dataset, our model achieved a PSNR of 30.2 ± 1.1 and SSIM of 0.88 ± 0.03, outperforming FGAN by the same margins in both metrics. Other methods, including FS-GAN and SAG-GAN, exhibited lower performance, with reduced PSNR and SSIM values across both datasets. The superiority of the proposed model was consistently evident across datasets, indicating a substantial improvement in overall image quality. It outperforms state-of-the-art models, excelling particularly in the retention of image details and the maintenance of structural integrity-critical aspects of medical image enhancement. These results indicate that the method not only yields higher PSNR and SSIM values but also ensures the superior preservation of fine details and structural consistency, making it suitable for clinical applications demanding high-fidelity image enhancement. In essence, these findings validate our method's effectiveness in multimodal medical image enhancement. With marked improvements in both quantitative metrics and visual quality relative to existing approaches, the method represents a reliable solution for safeguarding essential diagnostic image features.

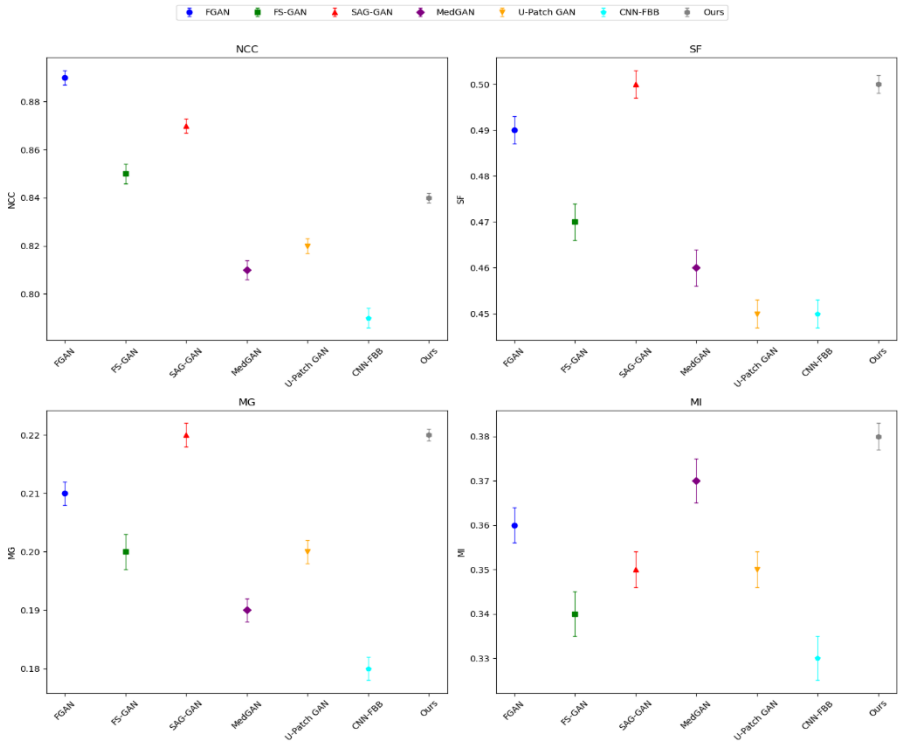

Figure 4. Quantitative comparison of the performance of the proposed ours method and other competing methods across various evaluation metrics: NCC, SF, MG, and MI.

The figure 4 illustrates the performance comparison of the proposed method with alternative approaches across multiple evaluation metrics. Based on the results for NCC, SF, MG, and MI, the proposed model consistently outperforms competing methods in all metrics. For the NCC metric, the proposed model achieved a value of 0.84, which represented the highest score among all methods. This result was accompanied by the shortest error bars, indicating greater stability and consistency in the medical image enhancement task. In contrast, the competing methods exhibited noticeably longer error bars, suggesting greater variability in their performance. In terms of SF, the proposed model demonstrated superior performance with a value of 0.50, accompanied by the shortest error bars. This finding highlights the architecture's ability to preserve fine image details and overall quality better than the competing methods. The alternative methods presented relatively longer error bars, further emphasizing the advantage in maintaining image details. Regarding the MG metric, the proposed approach demonstrated high performance with the smallest error bars, reflecting superior precision and consistency in the enhancement process. Although other methods attained reasonable results, the values remain inferior to the proposed model in both stability and accuracy. Regarding MI, the model maintained the top performance with a value of 0.38 and the shortest error bars, demonstrating a superior capability in preserving critical information within the image. In comparison. alternative methods showed longer error bars and lower scores, which underscores the ability of the proposed model to retain valuable image information more effectively. The proposed model excelled across all evaluated metrics, demonstrating a significant advantage over the comparison methods, particularly in terms of the minimizing variance and maintaining consistent image quality. This confirms the

robustness and effectiveness of the architecture for high-quality medical image enhancement.

Table 2. Comparison of training time and throughput across different methods on BraTS and LIDC-IDRI datasets.

Method	BraTS		LIDC-IDRI	
	Train Time (min)	Throughput (images/s)	Train Time (min)	Throughput (images/s)
FGAN	18.56	4	17.45	3.8
FS-GAN	19.34	3.8	18.52	3.7
SAG-GAN	20.67	3.9	19.21	3.6
MedGAN	21.56	4.1	20.43	3.5
U-Patch GAN	22.12	4.2	21.34	3.4
CNN-FBB	19.78	4	18.45	3.5
Ours	20.97	4	19.89	4

The training time and throughput results, summarized in Table 2, illustrate the computational efficiency of the evaluated methods on the BraTS and LIDC-IDRI datasets. The proposed model demonstrates a favorable balance between training time and throughput, underscoring its practicality for medical imaging applications. On the BraTS dataset, the training times of most models varied between 18.56 and 22.12 minutes, while the proposed model required 20.97 minutes. Although this training time is slightly higher than that of some approaches such as FGAN (18.56 minutes) and FS-GAN (19.34 minutes), the architecture maintained a throughput of 4 images per second, which is consistent with the highest-performing methods in the comparison. This stable throughput reflects the efficiency of the network architecture, particularly its ability to handle large-scale data while maintaining competitive computational performance. A similar trend was observed on the LIDC-IDRI dataset, where the proposed method recorded a training time of 19.89 minutes, which is comparable to the average training time of the other methods. For instance, FS-GAN and FGAN exhibited slightly lower training times of 18.52 and 17.45 minutes, respectively, while more complex models such as U-Patch GAN required up to 21.34 minutes. Despite these differences, the proposed model again maintained a throughput of 4 images per second, showcasing its reliability and computational scalability. The analysis of both datasets indicates that while some models achieve marginally shorter training times, they often compromise on throughput, leading to less efficient data processing. In contrast, the proposed model not only offers competitive training times and also achieves optimal throughput, ensuring a robust performance across various medical imaging tasks. These findings highlight the practical advantages of the method, particularly for real-world scenarios requiring both speed and accuracy in multimodal medical image enhancement.

4.2 Ablation Study

In Table 3, the performance of the proposed model was systematically evaluated after the

selective removal of different components, utilizing the BraTS and LIDC-IDRI datasets with PSNR and SSIM as the primary evaluation metrics.

Table 3. Performance comparison of ablation models on brats and lidc-idri datasets

Method	BraTS		LIDC-IDRI	
	PSNR	SSIM	PSNR	SSIM
Full Model (Ours)	28.5 ± 1.0	0.84 ± 0.04	30.2 ± 1.1	0.88 ± 0.03
No Residual Attention	27.0 ± 1.2	0.80 ± 0.05	28.7 ± 1.2	0.85 ± 0.04
No Multi-scale Convolution	26.5 ± 1.3	0.79 ± 0.06	28.2 ± 1.3	0.83 ± 0.05
No Upsampling Module	27.4 ± 1.0	0.82 ± 0.04	29.4 ± 1.1	0.87 ± 0.03
No PatchGAN	26.0 ± 1.3	0.77 ± 0.06	28.2 ± 1.4	0.83 ± 0.05
No Global Discriminator	26.6 ± 1.2	0.78 ± 0.05	28.5 ± 1.2	0.84 ± 0.04

The performance fluctuations of each variant illustrate the impact of component removal on image quality and structural consistency. The Full Model (Ours) outperformed all variants, underscoring the effectiveness of our comprehensive approach in preserving both image details and global consistency. When the Residual Attention Mechanism was removed (No Residual Attention), a significant performance decline occurred. On the BraTS dataset, PSNR dropped to 27.0 ± 1.2 dB and SSIM to 0.80 ± 0.05; on the LIDC-IDRI dataset, the values decreased to 28.7 ± 1.2 dB and 0.85 ± 0.04, respectively. These outcome highlights the mechanism's pivotal role in retaining fine image details. A more pronounced performance decline was observed when the Multi-scale Convolution module was removed. PSNR values plummeted to 26.5 ± 1.3 dB on BraTS and 28.2 ± 1.3 dB on LIDC-IDRI, while SSIM decreased to 0.79 ± 0.06 and 0.83 ± 0.05, respectively. These data indicate that multi-scale convolution is essential for enhancing image quality and ensuring detail retention.

The removal of the Upsampling Module (No Upsampling Module) resulted in a relatively minor decrease in both PSNR and SSIM. PSNR values declined slightly to 27.4 ± 1.0 dB on BraTS and 29.4 ± 1.1 dB on LIDC-IDRI, while SSIM was reduced to 0.82 ± 0.04 and 0.87 ± 0.03, respectively. This suggests that although the upsampling module contributes to image resolution, its impact is less pronounced compared to other components. For the variants without PatchGAN and the Global Discriminator, removing PatchGAN (No PatchGAN) triggered a more significant performance decline. PSNR values dropped to 26.0 ± 1.3 on BraTS and 28.2 ± 1.4 on LIDC-IDRI, with SSIM decreasing to 0.77 ± 0.06 and 0.83 ± 0.05, respectively. This finding demonstrates that PatchGAN is crucial for maintaining local consistency and preserving image details. Conversely, removing the Global Discriminator (No Global Discriminator) had a more limited impact on performance, with PSNR at 26.6 ± 1.2 dB on BraTS and 28.5 ± 1.2 dB on LIDC-IDRI, and SSIM at 0.78 ± 0.05 and 0.84 ± 0.04, respectively. These results indicate that while the global

discriminator primarily affects global structural consistency, its influence is less than that of PatchGAN. In summary, the removal of any module led to significant changes in PSNR and SSIM values, underscoring the importance of each component in ensuring image quality, detail preservation, and structural integrity. Among all components, the Residual Attention Mechanism and PatchGAN emerged as the most influential factors for enhancing image quality and preserving fine details, while the removal of the Multi-scale Convolution and Global Discriminator also substantially impacted performance, emphasizing their indispensable roles in handling complex medical images.

4.2 Visual Comparison of Image Enhancement Results

The visual results presented in Figure 5 clearly demonstrate the superiority of the proposed method over existing state-of-the-art models across both the BraTS and LIDC-IDRI datasets. On the BraTS dataset, the method successfully enhances the details of brain tumor regions, showing notable improvements in both structural integrity and hierarchical feature representations, especially in the regions highlighted by the red bounding boxes. This is in contrast to other methods, such as FGAN and FS-GAN, in which the tumor regions remain less distinct and blurry, indicating the challenges in preserving fine-grained structures during image enhancement. Similarly, on the LIDC-IDRI dataset, which comprises chest CT images, the method is effective at enhancing the details in critical regions like lung nodules.

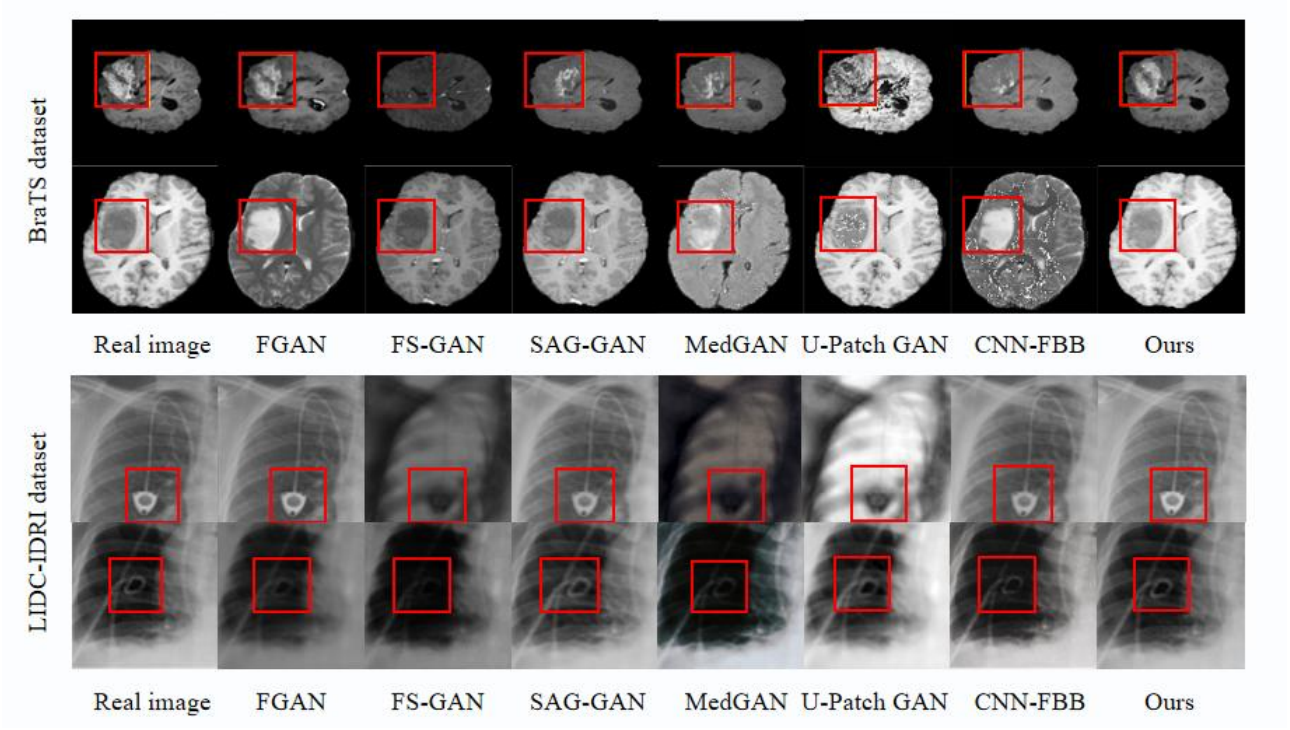

Figure 5. Visual comparison of the image enhancement results on the BraTS and LIDC-IDRI datasets. The red boxes highlight areas where significant improvements in image quality can be observed.

While other methods including SAG-GAN and U-Patch GAN showed limited improvement, the

image quality was not as consistently enhanced, and noise remained more prominent in certain areas. The method preserves the original image structures while effectively enhancing low-resolution regions, resulting in a more accurate and detailed reconstruction of the lung tissues and nodules. These visual results further confirm the effectiveness of the approach in addressing the key challenges in medical image enhancement, particularly the preservation of fine structural details and the reduction of noise. The consistent visual superiority observed across both datasets underscores the robustness and versatility of the proposed architecture, highlighting its potential for clinical applications that demand high-quality medical images with high spatial resolution and detailed preservation.

5. Conclusions

This study introduced an innovative IoT-based GAN framework tailored for multimodal medical image enhancement. By leveraging the real-time monitoring capabilities of IoT devices, the framework incorporates continuous, patient-specific data, facilitating the dynamic adaptation of the image enhancement process. This personalized approach ensures that the model responds to individual patient conditions, significantly improving the accuracy and clinical relevance of the resulting images.The model's generator integrates a Residual Attention Mechanism, Multi-Scale Convolutional Networks, and Adaptive Upsampling Modules. These components function synergistically to elevate image resolution while preserving fine details and maintaining global structural consistency. Specifically, the generator targets critical anatomical regions—such as tumors and vital organs—effectively transforming low-resolution inputs into high-fidelity outputs. Simultaneously, the hybrid discriminator, combining a global discriminator with PatchGAN, serves as a rigorous quality control mechanism. This dual-discriminator design distinguishes between real and generated images to guarantee the authenticity and superior quality of the enhancement results. Experimental evaluations on the BraTS and LIDC-IDRI datasets demonstrate the framework's superiority, as it outperformed existing state-of-the-art methods in both quantitative metrics and visual quality.

Despite these achievements, the model possesses certain limitations. It exhibits reduced effectiveness when processing severely noisy or heavily corrupted images, which may occasionally introduce minor artifacts. Additionally, the model's computational complexity results in relatively high processing costs, potentially restricting its application in immediate real-time clinical imaging scenarios. However, the integration of IoT data provides a viable solution by offering continuous feedback for real-time adjustments during the imaging process. Future research directions will focus on improving computational efficiency through lightweight network designs or model pruning techniques. Moreover, leveraging IoT-enabled real-time data to enhance the model's robustness against extreme noise or low-quality inputs could significantly improve generalization across diverse medical imaging contexts. Overall, the proposed method establishes a solid foundation for future advancements in medical image enhancement and holds significant potential for supporting accurate clinical diagnoses and effective treatment planning.

Acknowledgements

This article received no financial or funding support.

Conflicts of Interest

The author confirms that there are no conflicts of interest.

References

[1] Salem, N., Malik, H. and Shams, A. Medical image enhancement based on histogram algorithms. Procedia Computer Science, 2019, 163, 300–311.

[2] Yang, Y., Su, Z. and Sun, L. Medical image enhancement algorithm based on wavelet transform. Electronics Letters, 2010, 46(2), 120–121.

[3] Islam, S.M. and Mondal, H.S. Image enhancement based medical image analysis. In: Proceedings of 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India: IEEE, 2019, 1–5. DOI: 10.1109/ICCCNT45670.2019.8944910.

[4] Ma, Y., Liu, J., Liu, Y., Fu, H., Hu, Y., Cheng, J., Qi, H., Wu, Y., Zhang, J. and Zhao, Y. Structure and illumination constrained GAN for medical image enhancement. IEEE Transactions on Medical Imaging, 2021, 40(12), 3955–3967.

[5] Dinh, P.H. and Giang, N.L. A new medical image enhancement algorithm using adaptive parameters. International Journal of Imaging Systems and Technology, 2022, 32(6), 2198–2218.

[6] Chang, V., Hahm, N., Xu, Q.A., Vijayakumar, P. and Liu, L. Towards data and analytics driven B2B-banking for green finance: A cross-selling use case study. Technological Forecasting and Social Change, 2024, 206, 123542.

[7] Du, G., Liu, Z. and Lu, H. Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment. Journal of Computational and Applied Mathematics, 2021, 386, 113260.

[8] El-Shafai, W., Ghandour, C. and El-Rabaie, S. Improving traditional method used for medical image fusion by deep learning approach-based convolution neural network. Journal of Optics, 2023, 52(4), 2253–2263.

[9] Ullah, Z., Farooq, M.U., Lee, S.-H. and An, D. A hybrid image enhancement based brain MRI images classification technique. Medical Hypotheses, 2020, 143, 109922.

[10] Lundervold, A.S. and Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 2019, 29(2), 102–127.

[11] Ma, Y., Liu, Y., Cheng, J., Zheng, Y., Ghahremani, M., Chen, H., Liu, J. and Zhao, Y. Cycle structure and illumination constrained GAN for medical image enhancement. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2020, Cham, Switzerland: Springer, 2020, 667–677.

[12] Iqbal, S., Qureshi, A.N., Li, J. and Mahmood, T. On the analyses of medical images using traditional machine learning techniques and convolutional neural networks. Archives of Computational Methods in Engineering, 2023, 30(5), 3173–3233.

[13] Hagerty, J., Stanley, R.J. and Stoecker, W.V. Medical image processing in the age of deep learning-Is there still room for conventional medical image processing techniques? In: Proceedings of the 2017 International Conference

on Computer Vision Theory and Applications, 2017, 306–311.

[14] Ramesh, K., Kumar, G.K., Swapna, K., Datta, D. and Rajest, S.S. A review of medical image segmentation algorithms. EAI Endorsed Transactions on Pervasive Health & Technology, 2021, 7(27).

[15] Suganyadevi, S., Seethalakshmi, V. and Balasamy, K. A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval, 2022, 11(1), 19–38.

[16] Cao, Y., Shao, Y. and Zhang, H. Study on early warning of E-commerce enterprise financial risk based on deep learning algorithm. Electronic Commerce Research, 2022, 22(1), 21–36.

[17] Gong, M., Chen, S., Chen, Q., Zeng, Y. and Zhang, Y. Generative adversarial networks in medical image processing. Current Pharmaceutical Design, 2021, 27(15), 1856–1868.

[18] Bhutto, J.A., Tian, L., Du, Q., Sun, Z., Yu, L. and Tahir, M.F. CT and MRI medical image fusion using noise-removal and contrast enhancement scheme with convolutional neural network. Entropy, 2022, 24(3), 393.

[19] Yi, X., Walia, E. and Babyn, P. Generative adversarial network in medical imaging: A review. Medical Image Analysis, 2019, 58, 101552.

[20] Singh, N.K. and Raza, K. Medical image generation using generative adversarial networks: A review. In: Health Informatics: A Computational Perspective in Healthcare, 2021, 77–96.

[21] Chen, Y., Yang, X.-H., Wei, Z., Heidari, A.A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q. and Guan, Q. Generative adversarial networks in medical image augmentation: A review. Computers in Biology and Medicine, 2022, 144, 105382.

[22] Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., Chen, M., Wu, B., Zhang, H. and Chai, X. Generative adversarial networks in medical image segmentation: A review. Computers in Biology and Medicine, 2022, 140, 105063.

[23] Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q. and Shen, D. Medical image synthesis with context-aware generative adversarial networks. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017, Cham, Switzerland: Springer, 2017, 417–425.

[24] Ali, M., Ali, M., Hussain, M. and Koundal, D. Generative adversarial networks (GANs) for medical image processing: Recent advancements. Archives of Computational Methods in Engineering, 2025, 32(2), 1185–1198.

[25] Mahapatra, D., Bozorgtabar, B. and Garnavi, R. Image super-resolution using progressive generative adversarial networks for medical image analysis. Computerized Medical Imaging and Graphics, 2019, 71, 30–39.

[26] Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P. and Michalski, M. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Cham, Switzerland: Springer, 2018, 1–11.

[27] AlAmir, M. and AlGhamdi, M. The role of generative adversarial network in medical image analysis: An in-depth survey. ACM Computing Surveys, 2022, 55(5), 1–36.

[28] Zhao, J., Hou, X., Pan, M. and Zhang, H. Attention-based generative adversarial network in medical imaging: A narrative review. Computers in Biology and Medicine, 2022, 149, 105948.

[29] Islam, S., Aziz, M.T., Nabil, H.R., Jim, J.R., Mridha, M.F., Kabir, M.M., Asai, N. and Shin, J. Generative adversarial networks (GANs) in medical imaging: Advancements, applications, and challenges. IEEE Access, 2024, 12, 35728–35753.

[30] Jeong, J.J., Tariq, A., Adejumo, T., Trivedi, H., Gichoya, J.W. and Banerjee, I. Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation. Journal of Digital Imaging, 2022,

35(2), 137–152.

[31] Pan, Y., Liu, M., Xia, Y. and Shen, D. Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10), 6839–6853.

[32] Yu, Y.-F., Zhong, G., Zhou, Y. and Chen, L. FS-GAN: Fuzzy self-guided structure retention generative adversarial network for medical image enhancement. Information Sciences, 2023, 642, 119114.

[33] Qi, C., Chen, J., Xu, G., Xu, Z., Lukasiewicz, T. and Liu, Y. SAG-GAN: Semi-supervised attention-guided GANs for data augmentation on medical images. arXiv preprint arXiv:2011.07534, 2020.

[34] Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S. and Yang, B. MedGAN: Medical image translation using GANs. Computerized Medical Imaging and Graphics, 2020, 79, 101684.

[35] Fan, C., Lin, H. and Qiu, Y. U-Patch GAN: A medical image fusion method based on GAN. Journal of Digital Imaging, 2023, 36(1), 339–355.

[36] Qiu, T., Wen, C., Xie, K., Wen, F.Q., Sheng, G.Q. and Tang, X.G. Efficient medical image enhancement based on CNN-FBB model. IET Image Processing, 2019, 13(10), 1736–1744.