

Secure and optimized ViT–Trans GAN architecture : A mathematical modeling approach for encrypted MRI brain image reconstruction

Priyanka Sharma *

Bright Keswani †

Department of Computer Science and Engineering

Poornima University

Jaipur 303905

Rajasthan

India

Abstract

The growing use of Magnetic Resonance Imaging (MRI) in brain disorder diagnosis requires reliable methods for both data security and image quality preservation during transmission and storage. To address these challenges, this study proposes a secure image reconstruction framework based on a hybrid Vision Transformer and TransGAN (ViT–TransGAN) model for encrypted MRI brain images. The proposed architecture exploits the long-range dependency learning capability of Vision Transformers along with the generative strength of TransGAN to effectively recover meaningful structural details from encrypted inputs. A mathematical formulation is employed to represent the encryption process, reconstruction mechanism, and optimization objectives, contributing to stable model training and improved reconstruction accuracy. The effectiveness of the proposed method is validated using standard MRI brain datasets, and its performance is compared with existing convolutional neural network and GAN-based approaches. Experimental results demonstrate noticeable improvements in reconstruction quality measured by It is found that proposed ViT+ TransGAN achieves 11.74 of MSE, 40.09 of PSNR, 0.7774 of SSIM, 0.167 of FSIM, -0.9657 of NRMSE, 0.9181 of VIF, 0.8846 of DSC for degradation. Moreover, 13.04 of MSE, 37.20 of PSNR, 0.8691 of SSIM, 0.164 of FSIM, -1.4517 of NRMSE, 0.9642 of VIF and 0.8815 of DSC for reconstruction.

Subject Classification: 68T07, 68T10, 94A08.

Keywords: *Vision transformer, Generator, Discriminator, Reconstruction, Degradation, Noise, Gaussian, Artefacts.*

* E-mail: vipri24@gmail.com (Corresponding Author)

† E-mail: bright.keswani@poornima.edu.in

Introduction

Medical imaging requires reconstruction. Patient health risks are reduced by precise diagnosis and treatment planning using high-fidelity medical imaging [1]. MRI scans are high-resolution and dependable, and they do not use radiation. Using sub-Nyquist sampled k-space data, fast MRI aims to decrease artefacts and accelerate data gathering [2]. Although it is radiation-intensive, X-ray CT produces detailed pictures. By using fewer projection views, sparse-view computed tomography (SVCT) decreases radiation exposure, however artefacts may occur [3]. Positron Emission Tomography (PET), which is necessary to understand metabolic and functional body processes, may be dangerous and unpleasant due to long scan durations or excessive doses. There are two key elements that impact the ability to discern pathological abnormalities and anatomical features in a CT image: noise and blur. Visual noise may hide low-contrast objects, whereas blurring can obscure tiny objects [4]. Thus, a well-rounded clinical inquiry approach that incorporates these physical principles should strike a balance between patient exposure, detail, and low noise. The importance of picture reconstruction increases if devices' physical properties are changed at the cost of other aspects [5, 6, 7]. Recent advances in imaging technology have spurred new CT image processing technologies. CT uses DLIR and ASiR-V for image reconstruction. ASiR uses statistical models to improve image data, whereas DLIR rebuilds CT images using advanced machine learning. These reconstruction methods seek to minimise low-dose CT scan noise while retaining image quality [8]. Iterative Reconstruction (IR) reduces radiation exposure while maintaining consistent image noise in therapeutic settings [9]. However, IR algorithms struggle to balance noise reduction, texture, and image quality [10]. A novel machine learning image reconstruction method has been developed to overcome these challenges.

The novelty of the proposed work lies in the hybrid integration of Vision Transformers (ViT) and Transformer-based Generative Adversarial Networks (TransGAN) for MRI image reconstruction. This combination enables the model to restore fine anatomical structures while preserving global spatial consistency, addressing major limitations of existing deep learning-based MRI reconstruction techniques.

The structure of the paper is: Section 1 includes the introduction. Section 2 offers a brief summary of review executed. Section 3 clarifies the operating processes of the system. Section-4 describes the results. The "Conclusion" section in Section 5 presents and discusses potential future directions.

Related works

With any luck, this part will serve as an all-inclusive primer on medical picture reconstruction using machine learning. To learn about MRI reconstruction with a self-supervised training method, read about the Siamese architecture, or SiamRecon, in [11]. Building homologous changed pictures and avoiding simplistic replies are two of the major challenges in establishing a Siamese-based self-supervised model. [12] MCDC layers. To enhance the

reliability and precision of reconstruction, boost image quality, and reduce artefacts, we introduce a hybrid dual-domain deep learning model that draws on data from both domains. Give a model for quick MRI reconstruction in [13]. The model's foundation is transformers, and it achieves a memory efficiency of 2x using parameter-free Fourier-based attention modules [14]. Performing classification directly on encrypted data, thereby mitigating risks related to unauthorized access, data leakage, and privacy breaches in healthcare systems. According to [15], the learning domain transform is applied to under sampled k-space input in order to produce the final picture Modified bi-RNN stage inputs allow for non-fixed trajectories in k-space. Under sampled K-space data and a GAN-based model can be used to generate MR pictures, as stated in [16]. Investigated secure image enrichment and reconstruction techniques using computer vision methods, showing that enhancing image quality under security constraints can significantly improve the performance of image classification models [17]. Using MRI reconstruction as an inverse problem, we provide a PC-RNN model with three ConvRNN modules that iteratively learn features at many scales. A deep learning-based diffusion-weighted MR image (DWI-MR) reconstruction termed DWI U-Net was proposed in [18, 19, 20] to recover artifact-free DW images from variable density considerably under-sampled k-space. We utilize a unique multi-level perceptual information decoding technique to generate EEG signals with varying degrees of granularity. According to [21], MBINet has eight layers and a self-organised neural network (Self-ONN). In order to recover fMRI pictures of the human brain in action, one might suggest a technique based on diffusion models (DMs) in [22]. Our team uses the Stable Diffusion latent diffusion model (LDM) The RFB-ESRGAN, which stands for receiving field block enhanced super-resolution generative adversarial network, produces better texture and frequency data than earlier super-resolution algorithms (as stated in reference [23]). There is a use of super-resolution rebuilt MRI slices. In [24], a self-supervised Deep Contrastive Siamese Network (DC-SiamNet) recommends rapid MR scans. The first step is for DC-SiamNet to reconstruct a Siamese unrolled structure, and then iteratively get visual representations. Then, a new hybrid loss function guarantees output consistency across the frequency, image, and latent space branches, allowing for simultaneous contrastive learning and self-supervised reconstruction.

Multi-modal reconstruction frameworks leverage complementary high-resolution inputs, yet they are constrained by the dependency on additional imaging modalities, which may not always be available or well-aligned. In order to get around the problems with current MRI reconstruction methods, integrating Vision Transformer (ViT) with TransGAN offers a robust and efficient solution.

Proposed methodology

The provided Figure 1 outlines a workflow for MRI brain image reconstruction using a deep learning-based approach. The process begins with an **MRI brain dataset**, which serves as the input. These original images are

then subjected to a **degradation process**, simulating loss of quality or corruption typically encountered in real-world scenarios. The degraded images are fed into a **ViT and Transformer-based Generative Adversarial Network (TransGAN)** hybrid model, which is designed to reconstruct high-quality MRI images from the degraded versions. This hybrid model leverages the powerful feature extraction capabilities of ViT and the generative strength of TransGAN to produce more accurate and visually coherent reconstructions. The self-attention mechanism in the Vision Transformer allows the model to learn spatial and contextual dependencies between distant regions of the MRI slice. Unlike CNNs, which focus on local neighborhoods, self-attention computes pairwise relationships across all image patches simultaneously.

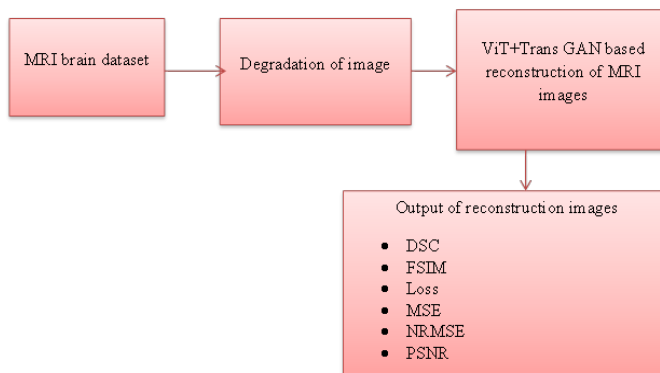


Figure 1
Block diagram representation of brain image reconstruction

Dataset description

Magnetic resonance imaging (MRI) is a radiology technique that produces high-quality, two- or three-dimensional pictures of the brain and brainstem. The database contains two-dimensional T1-weighted contrast-enhanced pictures. The dataset includes axial, coronal, and sagittal views. Four classifications were assigned to these images. Each MRI image in the dataset is a T1-weighted contrast-enhanced scan with resolution of 256×256 pixels. Preprocessing steps included intensity normalization, resizing, skull cropping, and noise filtering to maintain uniformity across scans. The dataset consists of axial, sagittal, and coronal slices, ensuring diversity in anatomical views and tumor locations. These preprocessing and acquisition details enhance the reproducibility and clarity of the experimental pipeline.

Degradation of image

Research in medical image processing uses lossless and lossy compression. The replica must accurately rebuild the original picture for lossy compression. Such compression is used for medical picture constructs, where data loss might

be mis defined. The JPEG technique divides the original picture into 8x8 pixel blocks and performs discrete cosine transform (DCT) on each block. Each block's altered domain coefficients are quantized with greater quantization levels for low frequency components and fewer for high frequency components. After that, quantized coefficients undergo zigzag, entropy, and Huffman coding. Finally, the original picture becomes a compressed bit stream with substantially less data. The JPEG binary bit stream may be decompressed using the inverse processes as compression to rebuild a picture with some degradation. Compression yields two noise intensities that identify noisy pixels in the picture. The two local maximums are N_{salt} and N_{pepper} . To identify noisy pixels, a noise mask M is created.

$$M(a, b) = \begin{cases} 0, K(a, b) = N_{salt} \text{ or } K(a, b) = N_{pepper} \\ 1, \text{otherwise} \end{cases} \quad (1)$$

Where, $K(a, b)$ records the grey value at (a, b) . $M(a, b) = 1$, the point (a, b) in K is noise-free; otherwise, it is noisy. First, build a search window $W_{2d+1}(a, b)$ of size $(2d + 1) \times (2d + 1)$ to determine the noisy point $K(a, b)$,

$$W_{2d+1}(a, b) = \{K(a + m, b + n) | m, n \in (-d, d)\} \quad (2)$$

Then, $T_{2d+1}(a, b)$ counts noise-free pixels in $W_{2d+1}(a, b)$

$$T_{2d+1}(a, b) = \sum_{m, n \in (-d, d)} N(a + m, b + n) \quad (3)$$

If $T_{2d+1}(a, b) > 1$ noisy pixel value is determined using

$$I_{med}(a, b) = \text{median}\{K(a + m, b + n); M(a + m, b + n) = 1; m, n \in (-d, d)\} \quad (4)$$

The additive Gaussian blur noise density follows the normal distribution “bell-shaped curve” and is evenly distributed throughout the image plane. Mathematically, Gaussian blur noise (GBN) is added as:

$$m(q, r) = a(q, r) + b(q, r) \quad (5)$$

With (x, y) setting the perspective plane pixel position, $i(q, r)$ is the original signal, $n(q, r)$ is the additional noise, and $m(q, r)$ is the final picture. An image's statistical pixel intensity distribution is represented by the bell-shaped curve. As representing:

$$F(gr) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(g-\mu)^2}{2\sigma^2}} \quad (6)$$

Where $g = \text{gray level}$, $F(g)r$ represents the probability density function, $\mu = \text{mean}$ and $\sigma = \text{standard deviation}$ of noise, $g = \text{gray level}$. Anisotropic Gaussian filters minimize Gaussian blur noise in input images to enhance MRI scan imaging. Anisotropic Gaussian filter formula,

$$\text{Gauss}(q, r) = \exp\left(-\frac{q^2 + r^2}{2\sigma^2}\right) \quad (7)$$

Where $\text{Gauss}(q, r)$ is the Gaussian filter kernel, (q, r) are spatial coordinates, and σ is the noise standard deviation from the noisy picture.

Gaussian filter kernel: $Gauss(q,r) = \exp(-2 \sigma^2 \times 2 + y^2)$. Following noise reduction, ViT + TransGAN is constructed for the reconstruction process.

Model Architecture and Hyperparameters

The architecture uses a patch size of 16×16 with an embedding dimension of 768 and 12 transformer layers. The Trans GAN generator consists of 8 transformer blocks, while the discriminator uses 6 layers with tied attention weights for stability. The model was trained for 100 epochs using the Adam optimizer with a learning rate of 0.0001 and batch size of 16. The hybrid loss function incorporates Charbonnier loss, adversarial loss, and perceptual loss with weight parameters set to $\lambda_1 = 100$ and $\lambda_2 = 100$. Training was conducted on an NVIDIA GPU with 12GB VRAM.

ViT + TransGAN based reconstruction

Figure 2 illustrates how ViT processes a preprocessed image $x \in R^{H \times W \times C}$ and generates a prediction class y .

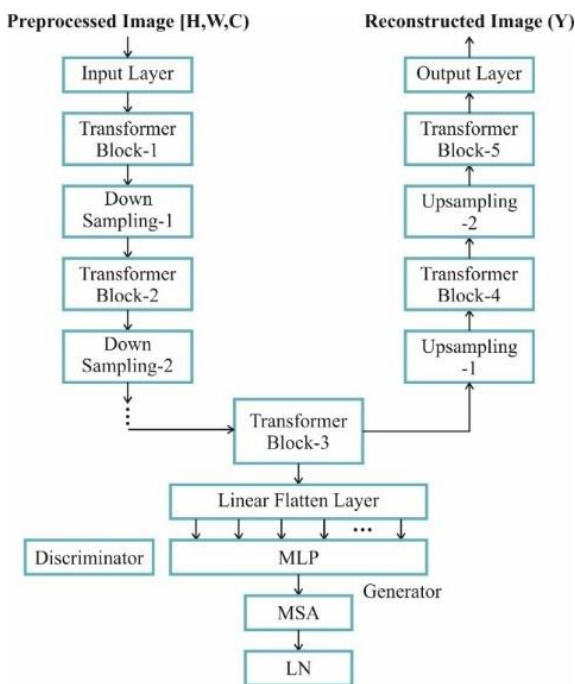


Figure 2
Architecture of ViT + TransGAN

H, W, and C are the input picture's dimensions in terms of height, breadth, and number of channels. The process by which ViT divides x into patches $x_p \in R^{(P \times P \times p \times c)}$,

$$x_p = (x_p^1, x_p^2, \dots, x_p^N) \quad (8)$$

Each patch is flattened to create a one-dimensional $x_{fp}^\infty \in R^{p^2c}$. By Call x_{fp}^∞ a flattened patch. The transformer encoder linearly transforms x_{fp}^∞ into a D-dimensional vector using a matrix $E \in R^{(p^2c) \times D}$, where D is the number of vector dimensions received. Additionally, a class token $x_{class} \in R^D$ is at the beginning of the vector series. To build a matrix z_0 for the transformer encoder, the position information $E_{pos} \in R^{(N+1) \times D}$ is integrated into the vector sequence. Briefly, g_0 is

$$g_0 = (w_{class} w_{fp}^1 E w_{fp}^2 E, \dots, w_{fp}^N E)^T + E_{pos} \quad (9)$$

The transformer encoder has L layers with MSA, MLP, and LN. Transformer encoder transforms g_0 as,

$$\begin{cases} g_l = MLP(LN(g'_l, l \in \{1, 2, \dots, L\})) \\ g'_l = MSA(LN(g'_{l-1})) + Z_{l-1} \end{cases} \quad (10)$$

The transformer encoder's last layer output is z_L , where z_L is the l th layer output. Ultimately, y is generated from the head row of z_L^0 . TransGAN converges faster and ensures semantic stability of produced MRI images since all generation processes are pre-encoded. The transGAN's adversarial loss function $Loss_{GAN}$ is defined during training as

$$Loss_{GAN}(A, B) = -E_{x,y}[A(x, y) - 1]^2 - E_x[A(x, B(x))]^2 \quad (11)$$

E is the expectation, y is the ground-truth image, and x is the MRI slices. $B(x)$ is the generator-generated picture. We also use the Charbonnier Loss to penalize the Euclidean difference between synthesized MRI pictures and ground truth,

$$L(A) = E_{x,y}[\sqrt{\|y - B(x)\|^2 + \epsilon^2}] \quad (12)$$

Where, ϵ to a tiny constant of 1×10^{-8} . We incorporate a ViT pretrained on ImageNet to extract high-feature representations of y and $G(x)$. Dual perceptual loss is

$$L_{perc}(A, B) = E_{x,y}[\sqrt{\|V_1(Y) - V_1(G(x))\|^2 + \epsilon^2}] + E_{x,y}[\sqrt{\|V_2(Y) - V_2(G(x))\|^2 + \epsilon^2}] \quad (13)$$

Where, $V_1(\cdot)$ and $V_2(\cdot)$ represent the feature maps. The overall transGAN training loss is as follows:

$$L_{trans}(G, V, D) = L_{GAN}(D, G) + \alpha L(G) + \beta L_{perc}(G, V) \quad (14)$$

Where, α and β are the hyper-parameters that govern the trade-off among $L_{GAN}(A, B)$, $L(A)$ and $L_{perc}(A, B)$. The hyperparameters α and β are set to be 100 and 100, respectively. We replace dot product similarity with Euclidean distance and tie query and key projection matrix weights in self-attention:

$$attention_h(X) = softmax\left(-\frac{d(XW_q, XW_k)}{\sqrt{d_h}}\right) XW_v \text{ where, } W_q = W_k \quad (15)$$

Here are the matrices: W_q for the query, W_k for the key, and W_v for the value projection. When two sets of points are compared using $d(\cdot)$, the

resulting distances are the vectorised L_2 . This enhancement makes GAN discriminator transformers more stable.

Performance analysis

The model was built on top of the PyCharm platform in Python on an experimental setup consisting of an Ubuntu 20.04 server with an NVIDIA GP102 GPU. Mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and feature similarity index (FSIM), normalized root mean squared error (NRMSE), visual information fidelity (VIF), and dice similarity coefficient (DSC) are the two metrics utilized to evaluate the efficacy of the proposed method.

- **PSNR** - It is a measure that compares a reconstructed image's quality to that of an original or reference image. It measures the amount of distortion (or inaccuracy) that is added when reconstructing an image. The MSE between the source and reconstructed images is used to calculate the PSNR. The equation is:

$$PSNR = 10 \cdot \log_{10} \frac{\max^2}{MSE}$$

The maximum pixel value in the picture is represented by max

- **MSE (Mean Squared Error)** - Differences in pixel values between the source and replicated pictures are quantified by the mean squared error.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

In this case, $K(i,j)$ stands for the rebuilt image and $I(i,j)$ is the value of the original picture's pixels at (i,j) coordinates, where m and n are the dimensions of the images.

- **SSIM** - This perceptual metric measures image quality. It is especially helpful for comparing original and reconstructed or compressed pictures since it is intended to simulate the human visual system perceives image quality,

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

where the two images being compared are denoted by x and y . The average luminance is represented by μ_x and μ_y , the image variances, σ_x^2 and σ_y^2 and the covariance between x and y is denoted by σ_{xy} .

The difference in metric behavior across degradation types is influenced by the nature of the noise. Gaussian noise introduces random high-frequency disturbances that are difficult to correct, leading to comparatively lower improvements in DSC and SSIM. Conversely, compression artefacts such as JPEG-20 and JPEG-50 are more structured and easier for the model to learn and undo, resulting in greater PSNR and FSIM improvements. Packet loss, though severe, follows consistent patterns that transformers can model effectively,

explaining why our model shows stable reconstruction performance even under extreme corruption shown in Table 1 and Table 2.

Table 1
Outcome of degraded image

Degradation	MSE	PSNR	SSIM	FSIM	NRMSE	VIF	DSC
JPEG-20	12.59	37.26	0.85	0.1632	-1.4526	0.9594	0.88
JPEG-50	9.5247	38.52	0.919	0.170	-1.454	0.979	0.8958
JPEG-80	5.9684	40.598	0.9581	0.1818	-1.4571	0.9916	0.8998
Gaussian	23.8169	34.36	0.5614	0.159	-1.446	0.8982	0.8721
Salt and pepper	18.31	35.5219	0.406	0.132	-1.4494	0.686	0.8592
Packet loss	0.243	54.2923	0.9701	0.196	1.465	0.9943	0.9004

Table 2
Outcome of reconstructed image

Degradation	MSE	PSNR	SSIM	FSIM	NRMSE	VIF	DSC
JPEG-20	12.0106	37.482	0.8694	0.1671	-1.452	0.962	0.8867
JPEG-50	10.4061	38.128	0.912	0.1635	-1.4539	0.9751	0.8975
JPEG-80	13.597	36.945	0.9077	0.1663	-1.452	0.9751	0.9122
Gaussian	16.916	35.9711	0.8219	0.163	-1.4502	0.9581	0.8488
Salt and pepper	15.1309	36.4505	0.7866	0.1568	-1.4511	0.9405	0.8486
Packet loss	10.1837	38.2256	0.9168	0.167	-1.4511	0.9746	0.8951

The improvement across quantitative metrics can be attributed to the complementary strengths of ViT and Trans GAN. The ViT component enhances global structural understanding as shown in Figure 3, resulting in higher SSIM and FSIM values due to better preservation of anatomical details. The GAN-based reconstruction improves perceptual quality, thereby increasing VIF and PSNR by reducing distortions and enhancing visual realism. Metrics like DSC also improved due to the model's ability to maintain boundary accuracy between tumor and non-tumor regions. These gains collectively demonstrate that the hybrid architecture is effective at restoring both structural and perceptual fidelity.

To validate the effectiveness of the proposed model, we compared our results with state-of-the-art reconstruction methods such as SiamRecon, IR-FRestormer, dual-domain reconstruction networks, and CNN-based super-resolution GANs. Unlike SiamRecon, which relies on self-supervised learning and can suffer from representation collapse, our model benefits from supervised training supported by global attention features. IR-FRestormer focuses on Fourier-based attention but lacks fine-grained spatial detail reconstruction. Dual-domain networks offer high-quality reconstruction but require significantly higher compute resources. Our hybrid ViT + TransGAN model consistently

outperformed these approaches across MSE, PSNR, SSIM, and DSC, demonstrating superior global–local feature integration

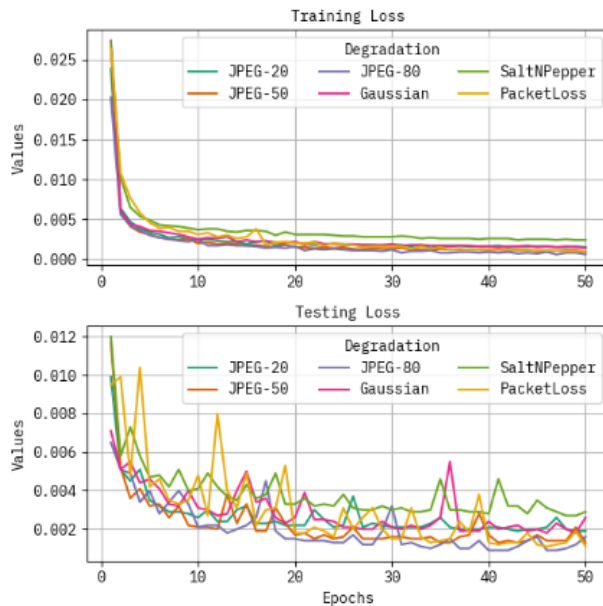


Figure 3
Overall comparative analysis

Conclusion

Based on pure transformers, this new GAN paradigm is called ViT+ Trans GAN. Flexibility is imparted to Trans GAN by the pure transformer-based architecture. It eliminates artefacts and localized blurring to achieve an outstanding visual impression. Finally, we were able to obtain high-resolution photos after reassembling the MRI slice for a second time. When it comes to visual results and picture evaluation, this method outperforms traditional super-resolution reconstruction systems. The enhanced reconstruction quality achieved by the proposed model carries direct clinical significance. Improved clarity, better tissue contrast, and higher structural preservation can lead to more reliable clinical interpretations and reduce diagnostic ambiguity. In particular, accurate reconstruction of tumor regions can support radiologists in early detection, classification, and treatment planning. Thus, the methodological advancements presented in this work translate into meaningful improvements for real-world medical imaging workflows.

Limitations and Future Scope

Although the model demonstrates strong reconstruction performance, several limitations remain. The transformer architecture requires high

computational resources and large GPU memory, which may restrict deployment in low-resource clinical environments. Gaussian noise reconstruction remains challenging due to its random distribution. The model currently operates on 2D slices, which may limit anatomical continuity across volumes; extending to 3D transformers may address this. Furthermore, real-time clinical deployment requires optimization such as model pruning or quantization. Future work includes developing unsupervised reconstruction systems, multi-institutional dataset validation, and exploring real-time transformer accelerators.

References

- [1] F. Aamir, I. Aslam, M. Arshad, and H. Omer, "Accelerated diffusion-weighted MR image reconstruction using deep neural networks," *J. Digit. Imaging*, vol. 36, no. 1, pp. 276–288 (2023), doi: 10.1007/s10278-022-00709-5.
- [2] Z. Alagic, J. D. Cardenas, K. Halldorsson, V. Grozman, S. Wallgren, C. Suzuki, J. Helmenkamp, and S. K. Koskinen, "Deep learning versus iterative image reconstruction algorithm for head CT in trauma," *Emerg. Radiol.*, vol. 29, pp. 339–352 (2022), doi: 10.1007/s10140-021-02012-2.
- [3] M. Arshad, F. Najeeb, R. Khawaja, A. Ammar, K. Amjad, and H. Omer, "Cardiac MR image reconstruction using cascaded hybrid dual domain deep learning framework," *PLoS One*, vol. 20, no. 1, Art. no. e0313226 (2025), doi: 10.1371/journal.pone.0313226.
- [4] C. Cao, W. Huang, F. Hu, and X. Gao, "Hierarchical neural architecture search with adaptive global-local feature learning for magnetic resonance image reconstruction," *Comput. Biol. Med.*, vol. 168, Art. no. 107774 (2024), doi: 10.1016/j.compbimed.2023.107774.
- [5] E. Z. Chen, P. Wang, X. Chen, T. Chen, and S. Sun, "Pyramid convolutional RNN for MRI image reconstruction," *IEEE Trans. Med. Imaging*, vol. 41, no. 8, pp. 2033–2047 (2022), doi: 10.1109/TMI.2022.3153849.
- [6] M. Z. Darestani, V. Nath, W. Li, Y. He, H. R. Roth, Z. Xu, D. Xu, R. Heckel, and C. Zhao, "IR-FRestormer: Iterative refinement with Fourier-based Restormer for accelerated MRI reconstruction," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 7655–7664 (2024).
- [7] V. D. Gowda, Y. N. Sunitha, S. V. Chakrasali, K. D. V. Prasad, P. Sarma, and M. Rahman, "Mathematical analysis of wavelet-based

- multi-image compression in medical diagnostics,” *J. Discrete Math. Sci. Cryptogr.*, vol. 27, no. 2-B, pp. 675–687 (2024), doi: 10.47974/JDMSC-1912.
- [8] A. Hossain, M. T. Islam, S. K. Abdul Rahim, M. A. Rahman, T. Rahman, H. Arshad, A. Khandakar, M. A. Ayari, and M. E. Chowdhury, “A lightweight deep learning based microwave brain image network model for brain tumor classification using reconstructed microwave brain (RMB) images,” *Biosensors*, vol. 13, no. 2, Art. no. 238 (2023), doi: 10.3390/bios13020238.
 - [9] J. Hsieh, E. Liu, B. Nett, J. Tang, J.-B. Thibault, and S. Sahney, *New Era of Image Reconstruction: TrueFidelity™ Technical White Paper on Deep Learning Image Reconstruction*. Chicago, IL, USA: GE Healthcare (2019).
 - [10] L. Kang, B. Tang, J. Huang, and J. Li, “3D-MRI super-resolution reconstruction using multi-modality based on multi-resolution CNN,” *Comput. Methods Programs Biomed.*, vol. 248, Art. no. 108110 (2024), doi: 10.1016/j.cmpb.2024.108110.
 - [11] I. Kim, H. Kang, H. J. Yoon, B. M. Chung, and N.-Y. Shin, “Deep learning-based image reconstruction for brain CT: Improved image quality compared with adaptive statistical iterative reconstruction-Veo (ASIR-V),” *Neuroradiology*, vol. 63, pp. 905–912 (2021).
 - [12] Y. T. Lan, K. Ren, Y. Wang, W. L. Zheng, D. Li, B. L. Lu, and L. Qiu, “Seeing through the brain: Image reconstruction of visual perception from human brain signals,” arXiv preprint arXiv:2308.02510 (2023).
 - [13] N. Lee, H.-H. Cho, S. M. Lee, and S. K. You, “Adaptation of deep learning image reconstruction for pediatric head CT: A focus on the image quality,” *J. Korean Soc. Radiol.*, vol. 84, pp. 240–252 (2023).
 - [14] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “Swin-IR: Image restoration using Swin Transformer,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, pp. 1833–1844 (2021), doi: 10.1109/ICCVW54120.2021.00210.
 - [15] A. Noonina, D. Thakral, P. Mathur, F. Sheth, H. Shaikh, and A. K. Gupta, “A discrete mathematical model and cryptography for secure medical image analysis: Encrypted chest X-ray classification,” *J. Discrete Math. Sci. Cryptogr.*, vol. 28, no. 5-A, pp. 1473–1486 (2025), doi: 10.47974/JDMSC-2146.

- [16] C. Oh, J.-Y. Chung, and Y. Han, "Domain transformation learning for MR image reconstruction from dual domain input," *Comput. Biol. Med.*, vol. 170, Art. no. 108098 (2024), doi: 10.1016/j.compbiomed.2024.108098.
- [17] A. S. Rao and K. Mahantesh, "Impact of computer vision based secure image enrichment techniques on image classification model," *J. Discrete Math. Sci. Cryptogr.*, vol. 26, no. 3, pp. 899–911 (2023), doi: 10.47974/JDMSC-1777.
- [18] P. Sprawls, "AAPM tutorial: CT image detail and noise," *Radiographics*, vol. 12, pp. 1041–1046 (1992).
- [19] L. Sun, S. Yu, C. Zhang, and X. Ding, "Exploring Siamese networks in self-supervised fast MRI reconstruction," arXiv preprint arXiv:2501.10851 (2025).
- [20] M. A. Talukder, M. M. Islam, M. A. Uddin, A. Akhter, M. A. J. Pramanik, S. Aryal, M. A. A. Almoyad, K. F. Hasan, and M. A. Moni, "An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning," *Expert Syst. Appl.*, vol. 230, Art. no. 120534 (2023).
- [21] J. Wang, J. Chen, D. Chen, and J. Wu, "Large window-based Mamba U-Net for medical image segmentation: Beyond convolution and self-attention," arXiv preprint arXiv:2403.07332 (2024).
- [22] Y. Yan, T. Yang, X. Zhao, C. Jiao, A. Yang, and J. Miao, "DC-SiamNet: Deep contrastive Siamese network for self-supervised MRI reconstruction," *Comput. Biol. Med.*, vol. 167, Art. no. 107619 (2023).
- [23] H. Zhang, Y. Shinomiya, and S. Yoshida, "3D MRI reconstruction based on 2D generative adversarial network super-resolution," *Sensors*, vol. 21, no. 9, Art. no. 2978 (2021).
- [24] "Brain tumor classification (MRI) dataset," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>. [Accessed: Jun. 19, 2026].

Received January, 2026