Artifact Aware Deep Learning with Diffuse Model for MRI Brain Tumor Image Segmentation

Mostafizur RAHMAN [1], *Wenmin WANG* [2], *Jiawei WANG* [3], *Yu WANG* [1,*]

[1] School of Computer and Artificial Intelligence, Beijing Technology and Business University, 100048, Beijing, PR China
[2] School of Computer Science and Engineering, Macau University of Science and Technology, 999078, Macao, PR China
[3] Dept. of Neurosurgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, PR China

2170021003@st.btbu.edu.cn, wmwang@must.edu.mo, jwwang_beijing@126.com, wangyu@btbu.edu.cn

Submitted January 11, 2025 / Accepted March 17, 2025 / Online first March 26, 2025

Abstract. *Brain tumor segmentation in MRI images is crucial for clinical diagnosis and treatment planning but those scans are usually affected by imaging artifacts which decrease the quality of data and hamper segmentation performance. To address these challenges, this study proposed a unique framework that seamlessly combines artifact correction with segmentation of tumors. The framework features a data preparation module which is able to prepare realistic artifact-contaminated and artifact-free MRI image pairs that have been used for training. It also includes a diffuse model which acts on MRI images and removes the artifacts thus giving high-quality inputs for segmentation. In addition, a modified 3D Convolutional Neural Network (CNN) architecture which integrates attention blocks and squeeze-and-excitation (SE) layers is used to segment the tumor subregions, including the enhancing tumor (ET), tumor core (TC), and whole tumor (WT). The framework was evaluated with artifact-corrupted data and clean data and achieved better results regarding the generation of artifact-free data and stable segmentation than the other baseline methods. This method emphasizes the magnitude of imaging artifacts on MRI-based segmentation and facilitates improvement in the clinical workflows. The code is available at https://github.com/Rahman3175/MMR*

Keywords

MRI artifacts, diffuse model, artifact-free images, attention blocks, brain tumor segmentation, 3D CNN

1. Introduction

In diagnosis, treatment planning, as well as monitoring neuro-oncology patients, brain tumor segmentation plays an important role in medical image analysis. Gliomas, which account for about 14.5% of total tumors in the central nervous system (CNS) and 48.6% of all malignant CNS tumors, comprise some of the most aggressive brain neoplasms. Apart from newer medical therapies, glioblastoma multiforme (GBM), which is the most aggressive glioma variant, barely survives for 15 months [1]. Considering that MRI is likely to produce excellent images of the brain and its abnormalities due to its high soft tissue contrast, this kind of imaging is invaluable during such procedures. However, manual evaluation of MRI is a time-consuming process, so automated computer-aided techniques are essential in the modern era to improve both the accuracy and efficiency of brain tumor segmentation [2]. More comprehensive data reporting is also key in minimizing the inequalities in the rates of tumor survival across populations of various ages [3].

Motion artifacts are the result of patient movement during MRI scans, as movement interferes with the flow of K-space data. Visual distortions such as blurring, ghosting, and general signal loss degrade the overall quality of the images produced, which blurs the view of the tumor structures [4], [5]. The artifacts impede the work of segmentation algorithms, which depend on accurate spatial and intensity metrics. Recent advances, such as joint field and motion control in T2-weighted imaging at 7T, have shown promise in mitigating motion-induced field fluctuations through simultaneous feedback loops [6].

Hybrid approaches that combine generalized auto calibrating partially parallel acquisitions (GRAPPA) with generative adversarial network (GANs) address the trade-offs of reduced signal-to-noise ratio in traditional PPI techniques, offering improved metrics like peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) for K-space reconstruction [7]. Deep learning models trained with diverse K-space subsampling techniques, accompanied by non-Cartesian trajectories, enable the use of extremely high acceleration rates without losing the quality of the reconstruction [8]. These examples prove the necessity of more comprehensive methods when dealing with motion artifact segmentation challenges.

Deep learning techniques, especially CNNs, have demonstrated outstanding performance in various tasks, including brain tumor segmentation, due to their ability to effectively localize multi-scale spatial features and efficiently handle 3D volume segmentation [9], [10]. Although architectures like 3D U-Net and its recent advancements are widely used, they often operate under highly constrained

DOI: 10.13164/re.2025.0166

conditions, assuming either artifact-free input or requiring preprocessing techniques such as bias correction [11]. Severe artifacts, such as outliers and illumination bias, remain significant challenges for segmentation in medical imaging [12].

To address these challenges, we present a unique framework that incorporates artifact removal and tumor segmentation using MRI scans. This framework consists of three well-defined modules including dataset preparation, training of a diffuse model, and training of a 3D CNN. In the data preparation subsection, artificial displacements, rotations and warping are added to the clean images to create paired artifact-affected and clean images, compared to the real-world counterparts, which were synthesized. This approach leverages synthetic datasets to address data scarcity and privacy concerns, ensuring robustness and generalizability [13].

To ensure that the generated artifacts accurately represent real-world MRI distortions, we selected displacement, rotation, warping, Gaussian noise, periodic ghosting, and K-space motion corruption due to their significant impact on brain tumor segmentation. The MRART dataset highlights the need for paired motion-corrupted and motion-free MRI scans to evaluate artifact correction techniques. However, the scarcity of such datasets necessitates synthetic motion artifacts to enhance segmentation model robustness and generalization [14]. Motion artifacts, including displacement, rotation, and warping, commonly occur in MRI scans as a result of involuntary patient movement during acquisition. These distortions introduce structural inconsistencies and spatial misalignment, leading to segmentation errors and reduced model robustness [15]. Gaussian noise is another frequent artifact caused by thermal noise and electronic interference in the imaging system, affecting contrast and tumor boundary clarity, thereby reducing segmentation performance [16]. Periodic ghosting arises from pulsatile motion, such as cardiac or respiratory activity, as well as gradient imperfections, resulting in the replication of structures that can confuse segmentation models [17]. K-space motion corruption occurs when patient movement distorts frequency-domain data, degrading the entire image reconstruction process, causing blurring and loss of anatomical details, and ultimately affecting segmentation accuracy [18]. These artifacts have been widely studied in MRI literature as major sources of degradation in automated analysis tasks. By simulating them in a controlled manner, our approach ensures that the segmentation model is robust to real-world imaging distortions, improving clinical applicability and reliability.

The diffusion model was trained using images from the paired dataset generated in the dataset preparation module in order to turn MRI images that contain artifacts into artifact-free images using advanced artifact correction techniques. Such models include motion artifact reduction through conditional diffusion probabilistic models (MAR-CDPM), which retain more structural details while reducing the motion artifacts and improving the image quality for segmentation tasks [19].

Tumor subregions, such as ET, TC, and WT, are accurately segmented using a modified 3D CNN-based segmentation module that incorporates enhanced attention mechanisms and SE layers to improve segmentation performance. U-Net’s attention models allow clinicians or operators to focus on specific regions while suppressing irrelevant data, thereby enhancing segmentation efficiency and reducing computational complexity [20]. The integration of SE layers with spatial-channel attention models further strengthens high-level feature extraction, improving segmentation accuracy and mitigating inconsistencies [21], [22], ensuring precise segmentation even in the presence of artifacts. Traditional artifact correction methods, such as bias field correction and filtering, have limited effectiveness when dealing with multiple artifacts in medical images, making them unsuitable for complex clinical applications [23]. In contrast, advanced algorithms that utilize a stack of head U-Nets with self-assisted priors can effectively correct rigid motion artifacts by leveraging information from neighboring slices, significantly improving image quality and structural similarity [24]. Additionally, GANs and diffusion models have demonstrated their ability to generate high-quality, artifact-free images. GANs correct motion distortions through adversarial learning, while diffusion models simulate artifact creation and iteratively restore critical anatomical structures [25], [26].

Recent models, such as the Denoising Autoencoder (DAE) [27], SwinMR [28], and RestormerGAN [29], have demonstrated notable effectiveness in artifact correction. DAE utilizes an encoder-decoder architecture to further reduce artifacts, resulting in clearer images. SwinMR leverages the Swin Transformer framework to enhance motion artifact correction while preserving structural integrity, whereas RestormerGAN combines GANs and transformer-based designs to generate artifact-free images.

Despite advancements in artifact segmentation and motion artifact correction, most existing methods in the literature have not been highly effective due to their lack of integration in addressing both tasks simultaneously. To overcome this limitation, our proposed framework merges motion artifact segmentation and motion artifact correction into a unified pipeline, training the segmentation task with high-quality outputs from a diffusion model. This approach significantly enhances model robustness and accuracy, particularly in complex cases. Furthermore, integrating attention mechanisms and SE layers into the 3D CNN architecture enhances tumor segmentation performance. Fine segmentation, combined with a hybrid loss function applied to 3D CNN-based models like SegResNet, results in higher Dice scores [30].

The remainder of this paper is organized as follows: In Sec. 2 the proposed method is described, including dataset preparation, diffuse model training, and 3D CNN training. In Sec. 3 the experimental setup, results, and a comparative analysis of the framework are given. Finally, conclusion with a summary of findings and future research directions is drawn in Sec. 4.

2. Method

2.1 Data Preprocessing

In this study, the preprocessing pipeline was designed to standardize the spatial resolution and normalize the intensity values of the MRI modalities while preserving key structural integrity. This involved T1ce, T1n, T2f, T2w, and segmentation masks, which were all reorganized to ensure a spatial resolution of 128 × 128 × 128 voxels. The resampling process is given by

$$I'(x',y',z') = \sum_{x,y,z} I(x,y,z)\,k\,(x-x',y-y',z-z') \tag{1}$$

where $I(x, y, z)$ is the source image, $I'(x', y', z')$ is the resampled image, and k is the interpolation kernel. To standardize the intensity values across datasets, voxel intensities were normalized to a range between [0, 1]. This normalization process is given by

$$I_{\text{norm}}(x,y,z) = \frac{I(x,y,z) - I_{\min}}{I_{\max} - I_{\min}} \tag{2}$$

where $I(x, y, z)$ represent the voxel intensity at position x, y, z, and $I_{\min}$ and $I_{\max}$ are the minimum and maximum intensity values in the image.

After pre-processing the data, all four MRI modalities for each patient were integrated into a single 4D array $C \times H \times W \times D$, with C denoting the four modalities that were included. We saved the segmentation masks as separate images, applying nearest-neighbor interpolation to maintain discrete labels. This approach ensures consistency across different MRI scans, preserving the integrity of tumor structures for accurate segmentation. In conclusion, all data that had been pre-processed was stored in .h5 format, which supports compression and is cost-effective in terms of both storage capacity and access time.

2.2 Architecture Overview

The proposed framework, as shown in Fig. 1, is structured into three interconnected modules, including dataset preparation, diffusion model training, and 3D CNN training. The dataset preparation module generates synthetic data by applying artifact-simulation techniques to clean MRI images and organizes them into paired datasets for training. The diffuse model is trained to transform artifact-affected images into artifact-free images, ensuring high-quality, artifact-free inputs for the segmentation model. Finally, the 3D CNN training module utilizes the artifact-free images to train a robust segmentation model capable of accurate brain tumor segmentation. Each module is described in detail in the following subsections.

2.3 Dataset Preparation Module

It can be seen in Fig. 1(a) that the dataset preparation module systematically generates a paired dataset of clean and artifact-affected images, as visualized in Fig. 2. This paired dataset serves as the foundation for training the diffusion model. As the first step, clean MRI data from the BraTS dataset is fetched, which is denoted as a tensor $I_c \in R^{C \times H \times W \times D}$ where C signifies the number of modalities, and H, W, D are the dimensions of the image in space. The clean MRI data is utilized to create-artifact affected images. In order to introduce realistic imaging artifacts, clean MRI images undergo a variety of transformations in a certain order. Displacement applies random linear shifts along each of the spatial axes and is given by

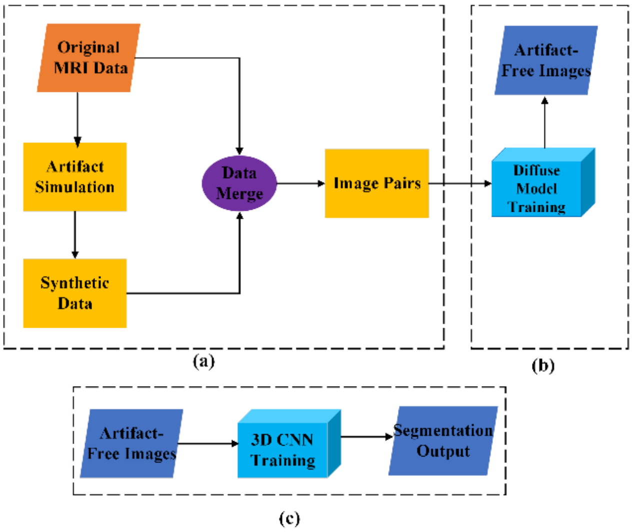


Fig. 1. Proposed architecture including (a) Dataset preparation, (b) Diffuse model training and (c) 3D CNN training.

$$I_{\text{d}}(x,y,z) = I_{\text{c}}(x-\Delta x, y-\Delta y, z-\Delta z) \tag{3}$$

where I_{c} is the clean MRI scan, I_{d} is the displaced image, and Δx, Δy ,Δz are random displacements sampled within a predefined range. In our pipeline, displacement shifts are randomly sampled between −5 and +5 voxels along each of the x, y, and z axes, ensuring realistically translational artifacts in multiple directions. Rotation is then applied to introduce angular displacements along the x, y, and z axes. The rotated image, denoted by I_{r}, is given by

$$I_{\text{r}} = R_{\text{z}}(\theta_{\text{z}})\left[R_{\text{y}}(\theta_{\text{y}})\left[R_{\text{x}}(\theta_{\text{x}})(I_{\text{d}})\right]\right] \tag{4}$$

where R_{x}, R_{y}, R_{z} are rotation operators along the x, y, and z axes, and θ_{x}, θ_{y}, θ_{z} are random angles sampled from maximum angle. In this study, the rotation angle is to $\theta = \pi/18$ radians or 10 degrees. Warping is then applied to mimic complex nonlinear deformations. The warped image denoted by I_{w} is given as

$$I_{\text{w}}(x,y,z) = \\ I_{\text{r}}\left(x + w_{\text{x}}(x,y,z), y + w_{\text{y}}(x,y,z), z + w_{\text{z}}(x,y,z)\right) \tag{5}$$

where w_{x}, w_{y}, w_{z} are random spatial displacements sampled from a normal distribution. In this implementation, warping is applied along all three axes, including x, y, and z, using sinusoidal functions to introduce structured distortions, given by

$$w_{\mathrm{x}}\left(y\right)=5\sin\frac{y}{10}, \tag{6}$$

$$w_{\mathrm{y}}\left(z\right)=5\sin\frac{z}{10}, \tag{7}$$

$$w_{\mathrm{z}}\left(x\right)=5\sin\frac{x}{10}. \tag{8}$$

These displacements introduce nonlinear distortions in the spatial coordinates, creating realistic artifact effects. Gaussian noise is added to the warped image to simulate stochastic imaging artifacts; the noisy image is given by

$$I_{\mathrm{n}}=I_{\mathrm{w}}+N\left(0,\sigma\right) \tag{9}$$

where I_{n} is the noisy image, and $N(0, \sigma)$ represents Gaussian noise with a mean of zero and a standard deviation of $\sigma = 0.005$. A periodic ghosting artifact is introduced by setting every 16th slice in the image to zero, creating a structured signal dropout, as follows

$$I_{\mathrm{g}}=I_{\mathrm{n}}, I_{\mathrm{g}}\left[\ ::16\right]=0. \tag{10}$$

This simulates periodic intensity dropouts, often observed in MRI due to sequence timing irregularities. Additionally, motion artifacts are simulated in the K-space domain. The Fourier transform of the noisy image is perturbed by a motion scaling factor, and the inverse transform is applied to produce the motion-affected image, given by

$$I_{\mathrm{m}}=F^{-1}\left(F\left(I_{\mathrm{n}}\right)\odot M\right) \tag{11}$$

where I_{m} is the motion-affected image, $\odot$ represents element-wise multiplication in the frequency domain, F and F^{-1} denote the Fourier and inverse Fourier transforms, respectively, and M is a motion-modulating factor sampled from a uniform distribution in the range [0.98, 1.02]. This step models artifacts resulting from patient motion during scanning, including blurring and ghosting effects**.**

The final artifact-affected image, I_{a}, is generated by combining these transformations. For every clean MRI image I_{c}, an artifact-affected counterpart I_{a} is created, forming a paired dataset and these pairs denoted as D_{p} are given by

$$D_{\mathrm{p}}=\left\{I_{\mathrm{c}}^{(i)},\ldots,I_{\mathrm{a}}^{(i)}\right\}_{i=1}^{N}. \tag{12}$$

This paired dataset is quite important in aiding the diffusion model by transforming artifact-affected inputs into high-quality, artifact-free outputs. Figure 2 illustrates the paired dataset, which encompasses all modalities such as T1ce, T1, T2f, and T2w, and displays the corrupted images alongside their cleaned versions.

2.4 Diffuse Model Training Module

As shown in Fig. 1, the diffusion model training module employs the paired dataset created during the dataset preparation process and aims to learn the mapping between artifact-affected MRI images and artifact-free images. This paired dataset, D_{p}, consists of clean MRI images I_{c} and their corresponding artifact-affected counterparts I_{a}. The diffusion model is designed to learn a mapping function f_θ, parameterized by θ, given by

$$f_\theta\left(I_{\mathrm{a}}\right)\approx I_{\mathrm{c}} \tag{13}$$

where f_θ is defined as the factor having a minimum inconsistency between the predicted artifact free output and the actual clean image.

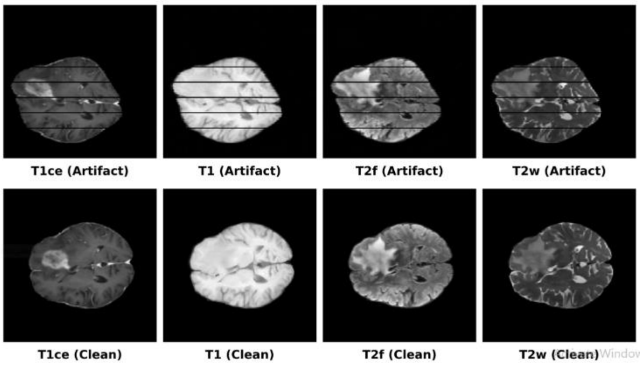


Fig. 2. Visualization of paired dataset, including artifact-affected and artifact-free images across all modalities.

In this study, we selected mean squared error (MSE) as the reconstruction loss function and structural similarity index (SSIM) as the perceptual loss to ensure both pixel-wise accuracy and perceptual quality. MSE is widely used in image restoration as it minimizes pixel-wise differences between predictions and ground truth images [31]. However, MSE alone often fails to capture structural distortions, which can be problematic in medical image reconstruction. As discussed by Wang and Bovik [32], MSE does not always align with perceptual image quality, necessitating SSIM for better structural fidelity in artifact removal tasks.

To address this, we incorporate SSIM, which evaluates structural similarity based on luminance, contrast, and texture, aligning more closely with human perception. Zhao et al. [33] analyzed various loss functions for image restoration, showing that perceptual losses like SSIM enhance both pixel-wise accuracy and structural fidelity. The combination of MSE and SSIM balances low-level fidelity with high-level perceptual coherence, making it well-suited for MRI artifact correction. The combined loss function, denoted as $L(\theta)$, is given by

$$L\left(\theta\right)= \\ \frac{1}{N}\sum_{i=1}^{N}\left(\left\|f_\theta\left(I_{\mathrm{a}}^{(i)}\right)-I_{\mathrm{c}}^{(i)}\right\|^2+\lambda\cdot\mathrm{SSIM}\left(f_\theta\left(I_{\mathrm{a}}^{(i)}\right),I_{\mathrm{c}}^{(i)}\right)\right) \tag{14}$$

where λ governs the trade-off between the two terms. Once the training is done, the diffuse model is used for artifact-free image generation, where it predicts clean MRI images, denoted by I_{p}, from artifact-affected inputs I_{a}, as given by

$$I_{\mathrm{p}}=f_\theta\left(I_{\mathrm{a}}\right). \tag{15}$$

This aspect guarantees the integrity and clarity of the produced images, and makes them additionally useful for the segmentation stage of the model. This module is critical in

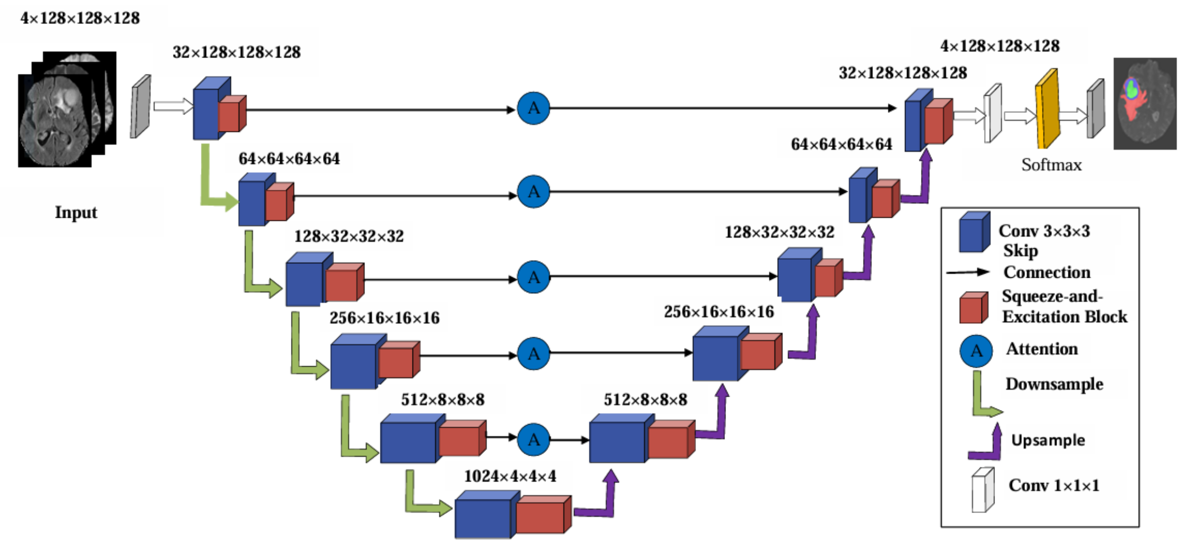


Fig. 3. Detailed workflow of the 3D CNN training module.

the proposed framework as it effectively bridges the gap between noisy, artifact-affected data and clean MRI images.

2.5 3D CNN Training Module

The 3D CNN training module, as shown in Fig. 1 and elaborated in Fig. 3, follows an improved encoder-decoder structure designed for precise brain tumor segmentation. It processes artifact-free MRI images generated by the diffusion model along with their corresponding ground truth segmentation masks. The architecture integrates an Attention Block (A), represented as a single unit in the figure, which consists of two types of attention mechanisms including SE with Spatial Attention and Axial Attention. SE-Spatial Attention enhances feature selection and localization by refining important regions, while Axial Attention captures long-range dependencies in volumetric MRI data, further strengthening feature representation.

As illustrated in Fig. 3, the encoder-decoder structure refines segmentation masks in the decoder after processing 3D MRI volumes in the encoder. During this process, spatial dimensions are progressively reduced while feature depth increases, optimizing feature extraction. The SE-Spatial Attention blocks reinforce both channel-wise feature importance and spatial relevance, dynamically recalibrating features while selectively enhancing crucial regions during upsampling. Axial Attention improves feature extraction by capturing complex tumor structures and dependencies across volumetric data. Additionally, skip connections transfer low-level spatial information from the encoder to the decoder, allowing for precise reconstruction of segmentation masks. This enhanced 3D CNN design, with its integrated SE, Spatial, and Axial Attention mechanisms, significantly improves segmentation performance, enabling accurate predictions of tumorous subregions such as ET, TC, and WT in volumetric MRI scans. The combination of these mechanisms not only enhances feature representation but also improves boundary precision, reducing segmentation errors. Moreover, the optimized architecture maintains a balance between accuracy and computational efficiency.

3. Experimental Results and Analysis

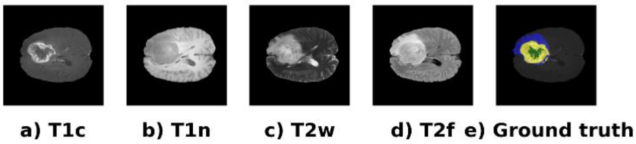
This section presents the experimental setup, covering the computational environment, datasets used, evaluation criteria, and results for both quantitative and qualitative assessment of the proposed framework.

3.1 Computing Environment

In this study, the experiments were conducted on an Intel Core i9-14900K high-end desktop processor with 24 cores running at a frequency of 6.0 GHz, Nvidia GeForce RTX 4090 GPU, 1 TB SSD and 48 GB RAM. The software range includes Python 3.9, CUDA 11.8, and PyTorch 2.0.0.

3.2 Dataset and Experimental Settings

In this study, the BraTS 2024 [34] dataset was used to train and test the model, containing 2251 training cases provided as MRI scans across multiple modalities. MRI modalities refer to different imaging sequences that capture various tissue contrasts by adjusting scanning parameters such as repetition time (TR) and echo time (TE). These modalities help in differentiating soft tissues, fluids, and abnormal growths, making them essential for accurate brain tumor segmentation. The dataset includes four key MRI modalities: T1-weighted (T1), which provides detailed anatomical structure; post-contrast T1-weighted (T1Gd), which enhances visualization of active tumor regions due to contrast agent uptake; T2-weighted (T2), which highlights edema and fluid-filled regions; and T2-FLAIR (Fluid-Attenuated Inversion Recovery), which suppresses cerebrospinal fluid (CSF) signals to better distinguish abnormal tissues, as shown in Fig. 4.



a) T1c b) T1n c) T2w d) T2f e) Ground truth

Fig. 4. Multi-modal MRI and expert-labeled tumor image from BraTS 2024.

Due to the unavailability of the BraTS 2024 official validation/test data, the experiments were validated using the five-fold cross-validation method, splitting the data in a 4:1 ratio of training to validation sets, ensuring robust evaluation of the proposed methodology. The MRI scans are provided in NIfTI format (. nii.gz) and sourced from multiple institutions using varied imaging protocols, ensuring diversity in scanner types and acquisition settings, which enhances the model's generalizability.

As seen in Fig. 4, all imaging datasets have been manually annotated by one to four raters, following the same annotation protocol, with approvals from experienced neuroradiologists. Annotations include the GD-enhancing tumor (ET, label 3), the peritumoral edematous/invaded tissue (ED, label 2), and the necrotic tumor core (NCR, label 1), as described in the latest BraTS paper. The ground truth data were created after pre-processing, including co-registration to a common anatomical template, interpolation to a uniform resolution (1 mm³), and skull stripping.

The diffusion model was trained using a patch-based Transformer architecture with an input size of $128 \times 128 \times 128$ voxels, where each patch was $8 \times 8 \times 8$ voxels. The Vision Transformer (ViT) model consisted of six transformer layers, each with eight attention heads, a hidden dimension of 512, and a multi-layer perceptron (MLP) dimension of 1024. The diffusion model employed 1000 timesteps with a linear beta scheduling strategy, where beta values were initialized from 1e–4 to 0.02. To ensure stable training and evaluation, we used a fixed 80% training and 20% validation split instead of cross-validation. This approach was chosen due to the high computational cost of training diffusion models, which makes repeated training over multiple folds impractical. The fixed split allowed for efficient training while maintaining a sufficient validation set to assess model performance.

The training process ran for 400 epochs with a batch size of 4. The Adam optimizer was used with a learning rate of 1e–4. The loss function incorporated a combination MSE for reconstruction accuracy and SSIM for perceptual quality, with an SSIM window size of 11. To ensure stability during training, noise was sampled using a normal distribution, and feature maps were projected using 3D convolution layers. The trained model generates artifact-free MRI images while preserving essential anatomical structures.

These artifact-free images generated by our proposed model were used to train a 3D CNN for brain tumor segmentation. The model was trained using a combination of cross-entropy and Dice loss with equal weighting to balance pixel-wise accuracy and segmentation consistency. The training process employed a five-fold cross-validation method, splitting the dataset in a 4:1 ratio for training and validation to ensure a robust evaluation. The model was trained for 400 epochs with an initial learning rate of 0.0001, optimized using Adam, and incorporated automatic mixed precision (AMP) for efficient computation. The batch size was set to 2, and early stopping was used to prevent overfitting. To further enhance training stability, gradient scaling with a dynamic loss scaler was applied. Model performance was assessed using the ET, TC, and WT Dice scores, which demonstrated its effectiveness in accurately segmenting the restored artifact-free dataset.

3.3 Evaluation Metrics

We evaluate the diffusion model's artifact correction and the 3D CNN's segmentation performance to ensure quantitative integrity in our framework analysis.

3.3.1 Artifact Correction Metrics

This subsection evaluates the diffuse model's ability to generate high-quality artifact-free images using metrics that assess image quality and structural preservation.

(1) Mean Squared Error (MSE)

Mean squared error (MSE) is used to measure the pixel-wise discrepancy between the generated artifact-free images and the ground truth clean images. It is defined as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left\| I_{\text{f}}^{(i)} - I_{\text{c}}^{(i)} \right\|^2 \tag{16}$$

where I_f represents the artifact-free images generated by the diffuse model, I_c denotes the ground truth clean images, and N is the total number of samples.

(2) Structural Similarity Index Measure (SSIM)

Regarding the previously issued observations, the structural similarity index measure (SSIM) quantifies the visual closeness of the artifact-free images to the original image, emphasizing structural features rather than just pixel values. It is defined as

$$\text{SSIM}(x,y) = \frac{\left(2M_\text{x}M_\text{y} + C_1\right)\left(2\,\text{Cov}_{xy} + C_2\right)}{\left(M_\text{x}^2 + M_\text{y}^2 + C_1\right)\left(\text{Var}_x + \text{Var}_y + C_2\right)} \tag{17}$$

where x and y are the two images being compared, M_x and M_y are the mean intensities of x and y, Var_x and Var_y are the variances of x and y, Cov_{xy} is the covariance between x and y, and C_1 and C_2 are small constants to stabilize the computation. SSIM values range from 0 to 1, where 1 indicates perfect structural similarity.

(3) Noise Level

The noise level is assessed using the variance of the intensity differences between I_f and I_c. It reflects the degree of residual noise in the corrected images. It is defined as

$$\text{NoiseLevel} = \text{Var}\left(I_\text{f} - I_\text{c}\right) \tag{18}$$

where $\text{Var}(I_\text{f} - I_\text{c})$ computes the variance of the intensity differences.

(4) Sharpness

Sharpness is a crucial aspect in the field of medical imaging. It focuses on the retention of edge structures in the generated images, which is critical for maintaining anatomical structures in medical imaging. It is measured using the variance of the Laplacian, defined as

$$\text{Sharpness} = \text{Var}\left(L\left(I_{\text{f}}\right)\right) \tag{19}$$

where $L(I_{\text{f}})$ represents the Laplacian of the image I_{f}, calculated as the sum of second-order differences across the spatial dimensions.

(5) Contrast-to-Noise Ratio (CNR)

Contrast-to-noise ratio (CNR) measures the contrast between regions of interest, such as tumor boundaries and the background, relative to the noise in the image. It is defined as

$$\text{CNR} = \frac{\left|M_{\text{ROI}} - M_{\text{b}}\right|}{\sigma_{\text{b}}} \tag{20}$$

where M_{ROI} is the mean intensity of the region of interest, M_{b} is the mean intensity of the background, and σ_{b} is the standard deviation of the background.

(6) Peak Signal-to-Noise Ratio (PSNR)

Peak signal-to-noise ratio **(**PSNR) evaluates the overall similarity between the generated artifact-free images and the ground truth by comparing the ratio of signal power to noise power. It is defined as

$$\text{PSNR} = 20\log_{10}\left(\frac{L}{\sqrt{\text{MSE}}}\right) \tag{21}$$

where L is the dynamic range of the pixel intensities.

3.3.2 Segmentation Metrics

To evaluate the segmentation performance of the proposed model, we consider the WT, TC, and ET regions, where WT encompasses TC, and TC encloses ET. The evaluation metrics include the Dice coefficient, Jaccard index, sensitivity, specificity, and the 95th percentile of the Hausdorff distance (HD95). Among these, the Dice coefficient is most significant as it measures overlap, while the Jaccard index quantifies the ratio of intersection to total area, making it particularly useful when only partial overlap exists. HD95 assesses boundary accuracy by determining the maximum surface distance between the predicted and ground truth regions. Sensitivity measures the proportion of correctly identified true positives, whereas specificity indicates the proportion of actual negatives correctly classified, reducing false positives. The mathematical definitions of these metrics are as follows:

$$\text{Dice} = \frac{2\left|P \cap T\right|}{\left|P\right| + \left|T\right|}, \tag{22}$$

$$\text{Jaccard} = \frac{\left|P \cap T\right|}{\left|P \cup T\right|}, \tag{23}$$

$$\text{Sensitivity} = \frac{\left|P \cap T\right|}{\left|T\right|}, \tag{24}$$

$$\text{Specificity} = \frac{\left|P_0 \cap T_0\right|}{\left|T_0\right|}, \tag{25}$$

$$\text{HD95}\left(A, B\right) = \max\left(\sup_{a\in A}\inf_{b\in B} d\left(a,b\right), \sup_{b\in B}\inf_{a\in A} d\left(a,b\right)\right) \tag{26}$$

where A and B represent the surfaces of T as the ground truth and P as the predicted segmentation, respectively, a and b are points on these surfaces, and $d\,(a, b)$ computes the Euclidean distance between them.

3.4 Comparative Results

3.4.1 Artifact Correction Results

Table 1 provides a comprehensive evaluation of sharpness, PSNR, MSE, SSIM, CNR, noise level, and other key metrics for assessing the performance of both artifact-affected and artifact-free images. Additionally, Table 2 presents a detailed comparison of our proposed method with state-of-the-art techniques, including DAE, SwinMR, and RestormerGAN, all tested on the artifact-free dataset generated by our pipeline.

As shown in Tab. 1, the proposed artifact correction framework significantly enhances imaging metrics, improving overall image quality. The sharpness of artifact-free images increases from an average of 0.003 to 0.006, leading to clearer and more detailed reconstructions. Similarly, the PSNR improves from 27.4 dB to 32.5 dB, indicating reduced reconstruction errors and higher image fidelity. The SSIM increases from an average of 0.77 to 0.93, demonstrating improved structural preservation and enhanced visual clarity. The CNR also improves from a near-negligible 0.0001 to 0.006, highlighting better differentiation in critical areas, such as tumor boundaries. Additionally, the framework ensures stable artifact suppression across various test cases, maintaining consistency in image quality. Notably, the noise level remains unchanged, confirming the framework's ability to remove artifacts without introducing additional noise.

The effectiveness of the proposed method in correcting motion artifacts is demonstrated by the PSNR and SSIM values as shown in Tab. 2. The degraded images exhibit poor quality, with a PSNR of 26.4 dB and an SSIM of 0.764, indicating significant structural distortions. Among existing methods, SwinMR achieves the highest SSIM of 0.913, while RestormerGAN attains a PSNR of 30.4 dB.

It can be seen from Tab. 2 that our transformer-based diffusion model surpasses existing methods, achieving a PSNR of 32.5 dB and an SSIM of 0.925. The superior performance is attributed to the model's ability to capture long-range dependencies and progressively refine structural details during artifact suppression. Unlike conventional models, the transformer architecture enhances motion artifact correction by leveraging patch-based attention and iterative

Metric	Artifact-Affected		Artifact-Free	
	Avg	Max	Avg	Max
Sharpness	0.003181	0.004450	0.006798	0.009891
PSNR (dB)	27.37939	31.16382	32.50965	35.61952
MSE	0.002173	0.008520	0.002106	0.008001
SSIM	0.773865	0.813834	0.925679	0.984184
CNR	0.000128	0.004096	0.006268	0.093660
Noise Level	0.011526	0.028101	0.011631	0.027616

Tab. 1. Comprehensive metrics for artifact-affected and artifact-free images.

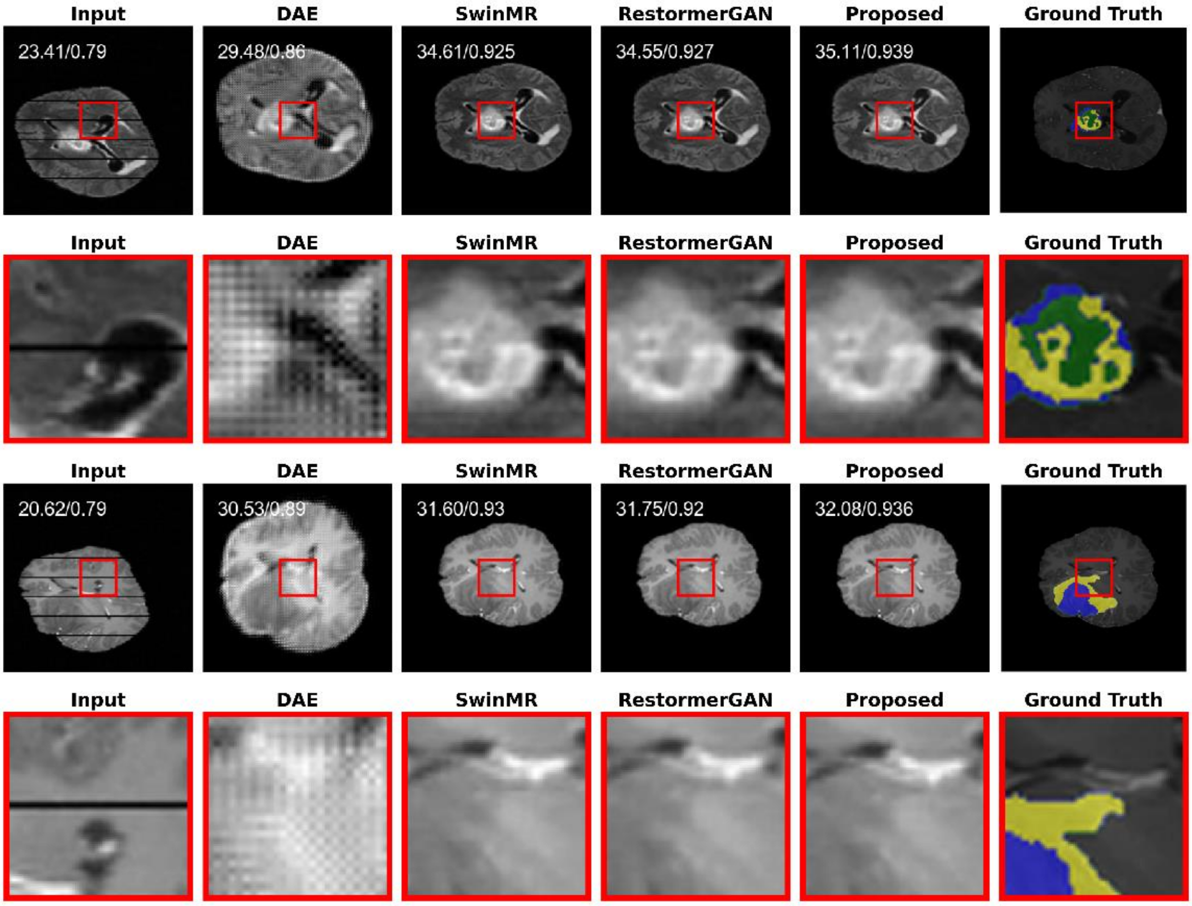


Fig. 5. Comparison of motion artifact correction results across models using pair dataset.

Method	PSNR (dB)	SSIM
Degraded Images	26.379	0.764
DAE [27]	28.325	0.875
SwinMR [28]	31.603	0.913
RestormerGAN [29]	30.423	0.892
Proposed	**32.509**	**0.925**

Tab. 2. Comparative results of artifact correction on synthetic and clean MRI image pairs.

denoising, ensuring high-fidelity image reconstruction. Figure 5 visually compares the motion artifact correction results across models using the synthetic and clean pair dataset, further demonstrating the effectiveness of our approach.

In Fig. 5, the outputs from different models that were artifact corrected and trained on the dataset containing synthesized and clean MRI image pairs are presented. The input images contain artifacts and have been windowed for clarity. Each image corner displays the PSNR and SSIM values of the respective image, providing a quantitative measure of reconstruction quality. Additionally, the difference images have been supersized up to three times for enhanced visibility. The figure showcases the results from Input, DAE, SwinMR, RestormerGAN, the proposed model, and the Ground Truth. Artifact-free images generated by the proposed model are shown in Fig. 6.

It can be seen from Fig. 6 that artifact-corrected images generated by the proposed model effectively demonstrate its capability in restoring high-quality MRI images. Each image's PSNR and SSIM values are annotated in the top-left corner, providing both visual and numerical assessments of the reconstruction fidelity. These results further highlight the model's ability to mitigate artifacts and enhance image clarity, reinforcing its effectiveness in MRI restoration.

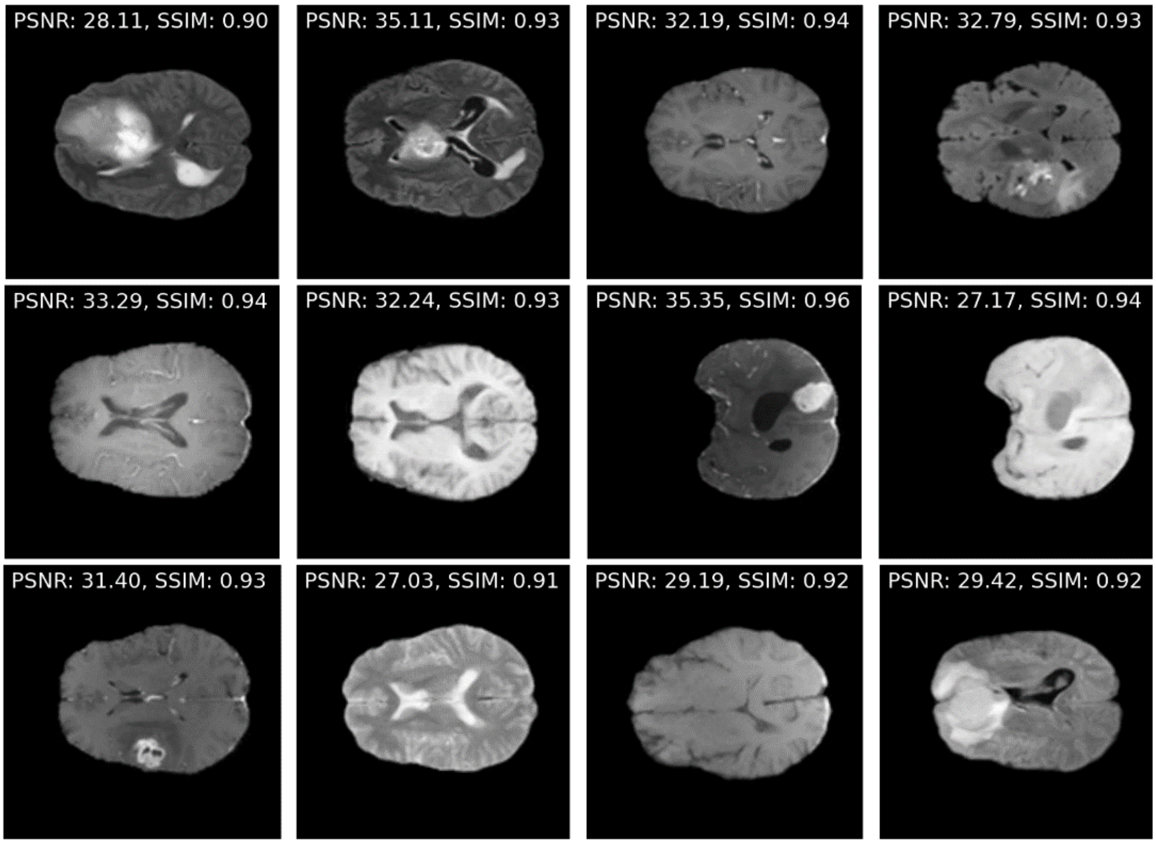


Fig. 6. Visualization of artifact-free images generated by the proposed model for representative cases.

Model Variant	Encoder-Decoder	SE Layers	Attention Blocks	Skip Connections	Mean Dice (%)			Mean Jaccard (%)		
					ET	TC	WT	ET	TC	WT
Baseline	✓	-	-	-	76.24	79.32	84.12	75.12	78.20	82.23
SE Layers	✓	✓	-	-	78.51	82.43	86.33	76.32	82.05	83.22
Attention Blocks	✓	-	✓	-	78.26	81.21	86.34	78.00	79.03	84.21
Skip Connections	✓	✓	✓	-	81.65	84.02	87.45	79.23	83.33	85.15
Proposed	✓	✓	✓	✓	84.23	89.20	91.85	83.45	88.34	89.05

Tab. 3. Ablation experiment results on artifact-free dataset from proposed pipeline.

Model Variant	GFLOPs (G)	Parameters (M)	Inference Time (s)
Baseline 3D CNN	1251.535	43.189	0.00576
SE+3D CNN	1251.892	43.410	0.01040
SE+3D CNN+Spatial Attention	1252.606	43.413	0.01302
SE + 3D CNN + Spatial Attention+ Axial attention	1337.057	51.29	0.13511

Tab. 4. Computational complexity analysis of the proposed pipeline.

3.4.2 Ablation Study

The ablation study, presented in Tab. 3 evaluates the impact of key components in our model, including SE layers, attention blocks, and skip connections, on segmentation accuracy. Conducted using five-fold cross-validation on the dataset generated through our pipeline, this analysis quantifies the contribution of each component to improving the Mean Dice and Mean Jaccard scores across ET, TC, and WT regions.

It can be seen from Tab. 3 that the baseline model achieves the lowest segmentation performance, highlighting the need for advanced architectural enhancements. The baseline struggles to delineate tumor regions, particularly in ET and TC, due to limited feature extraction capabilities. Adding SE layers improves feature recalibration, enhancing channel-wise attention and increasing Mean Dice scores, particularly for WT segmentation. The incorporation of attention blocks, consisting of spatial attention and axial attention, enhances feature representation by capturing both local and long-range dependencies within volumetric MRI data. This refinement improves tumor boundary delineation, reducing segmentation errors. Additionally, integrating skip connections further boosts accuracy by maintaining spatial consistency, improving gradient flow, and preventing feature loss. The proposed model, integrating SE layers, attention mechanisms, and skip connections, achieves the highest Mean Dice and Mean Jaccard scores across ET, TC, and

WT, demonstrating the effectiveness of this unified framework. In addition, the computational complexity of the proposed pipeline is shown in Tab. 4.

We can see from Tab. 4 that architectural enhancements impact GFLOPs, parameter count, and inference time, reflecting the computational complexity of the proposed 3D CNN pipeline. The baseline 3D CNN model has the lowest computational cost with 1251.535 GFLOPs and 43.189 million parameters but suffers from limited feature extraction, leading to suboptimal segmentation. Adding SE layers slightly increases complexity to 1251.892 GFLOPs and 43.410 million parameters while improving Mean Dice scores, particularly for WT segmentation. Spatial attention further refines local feature selection and tumor boundary delineation, enhancing segmentation while increasing GFLOPs to 1252.606. Despite minimal computational overhead, it helps preserve fine structural details. Integrating axial attention incurs the highest computational cost, increasing GFLOPs to 1337.057 and parameters to 51.29 million, due to its ability to capture both local and long-range dependencies, significantly improving feature representation and segmentation accuracy. The combination of SE layers, spatial attention, and axial attention strikes a balance between computational efficiency and segmentation accuracy, enhancing spatial awareness and ensuring accurate tumor region delineation without excessive overhead.

3.4.3 Segmentation Results

The proposed model's performance was evaluated using five-fold cross-validation, achieving superior results on the artifact-free dataset with higher Dice scores and lower HD95 values, as shown in Tab. 5. For baseline comparison, segmentation results on the BraTS 2024 dataset are shown in Tab. 6.

From Tab. 5, it can be observed that the proposed model delivers superior segmentation performance on the artifact-free dataset generated by our trained diffusion model. Specifically, it achieves a high Mean Dice score of 84.23% for ET and 91.85% for WT, surpassing Attention U-Net and V-Net, while performing comparably to TransBTS. Additionally, the proposed model attains the highest mean Sensitivity, with 86.75% for ET and 94.16% for WT, along with a mean Specificity of 99.86% for ET. In terms of boundary delineation, it achieves the lowest mean HD95 values, measured at 3.47 mm for TC and 3.51 mm for WT, demonstrating its effectiveness in accurate tumor boundary segmentation.

To ensure a fair comparison, all models, including the baseline and state-of-the-art methods, were retrained on the same artifact-free dataset generated by our trained diffusion model. This ensures that any performance differences arise from architectural variations rather than disparities in training data. Additionally, all methods underwent the same evaluation protocol, using five-fold cross-validation to ensure robust and unbiased performance assessment. These results highlight the robustness and effectiveness of the proposed approach compared to existing baselines. Unlike other frameworks that treat artifact correction and segmentation as separate tasks, often leading to suboptimal performance, our approach integrates both within a unified pipeline. Training the segmentation model on high-quality, artifact-free images from our diffusion model mitigates challenges of artifact-affected segmentation, improving structural integrity and predictive accuracy.

Furthermore, integrating spatial attention, axial attention, and SE layers within the 3D CNN enhances feature extraction, improving structural consistency and boundary precision in tumor segmentation. Combining artifact correction and segmentation within a single framework enables the model to handle complex imaging scenarios more effectively, ensuring accurate and reliable outcomes.

Table 6 presents the qualitative analysis of segmentation performance, demonstrating the superiority of the proposed model over state-of-the-art methods on the BraTS 2024 dataset. The model achieves a Dice score of 84.23% for ET and 91.85% for WT, surpassing Attention U-Net, 3D U-Net, V-Net, and TransBTS. It also maintains high sensitivity and specificity, ensuring reliable segmentation. Additionally, the model achieves the lowest HD95 values, including 10.61 mm for ET and 3.47 mm for TC, indicating superior boundary delineation and precise tumor region identification.

Model	Mean Dice (%)			Mean Sensitivity (%)			Mean Specificity (%)			Mean HD95 (mm)		
	ET	TC	WT	ET	TC	WT	ET	TC	WT	ET	TC	WT
AttU-Net [20]	80.12	82.42	88.68	82.26	86.76	90.35	98.65	98.95	99.11	17.71	7.46	5.56
3D U-Net [35]	82.82	87.67	90.06	83.13	87.54	**95.07**	99.57	98.49	99.74	**9.74**	6.68	8.26
TransBTS [36]	81.26	88.13	91.25	83.78	92.53	94.02	99.84	99.43	99.58	12.39	6.53	9.56
V-Net [37]	79.24	83.15	88.26	81.27	84.67	91.21	99.32	97.54	97.21	11.09	9.83	10.1
Proposed	**84.23**	**89.20**	**91.85**	**86.75**	**93.75**	94.16	**99.86**	**99.72**	**99.91**	10.61	**3.47**	**3.51**

Tab. 5. Quantitative comparison of segmentation results across models on the artifact-free dataset.

Model	Mean Dice (%)			Mean Sensitivity (%)			Mean Specificity (%)			Mean HD95 (mm)		
	ET	TC	WT	ET	TC	WT	ET	TC	WT	ET	TC	WT
AttU-Net	78.22	82.46	87.13	80.26	84.43	91.22	97.65	97.12	98.11	12.16	6.77	7.43
3D U-Net	81.52	85.13	88.95	82.86	88.35	93.17	99.89	98.78	98.25	**10.43**	7.28	9.01
TransBTS	81.07	86.55	89.23	83.24	91.62	**94.16**	99.90	99.70	**99.92**	13.67	8.23	9.92
V-Net	79.23	81.16	87.66	80.52	86.77	92.32	99.32	97.54	97.21	12.64	9.87	11.05
Proposed	**82.44**	**87.15**	**90.02**	**86.85**	**92.65**	93.76	**99.93**	**99.85**	99.90	10.85	**5.64**	**6.32**

Tab. 6. Quantitative comparison of segmentation results across models on the BraTS 2024 dataset.

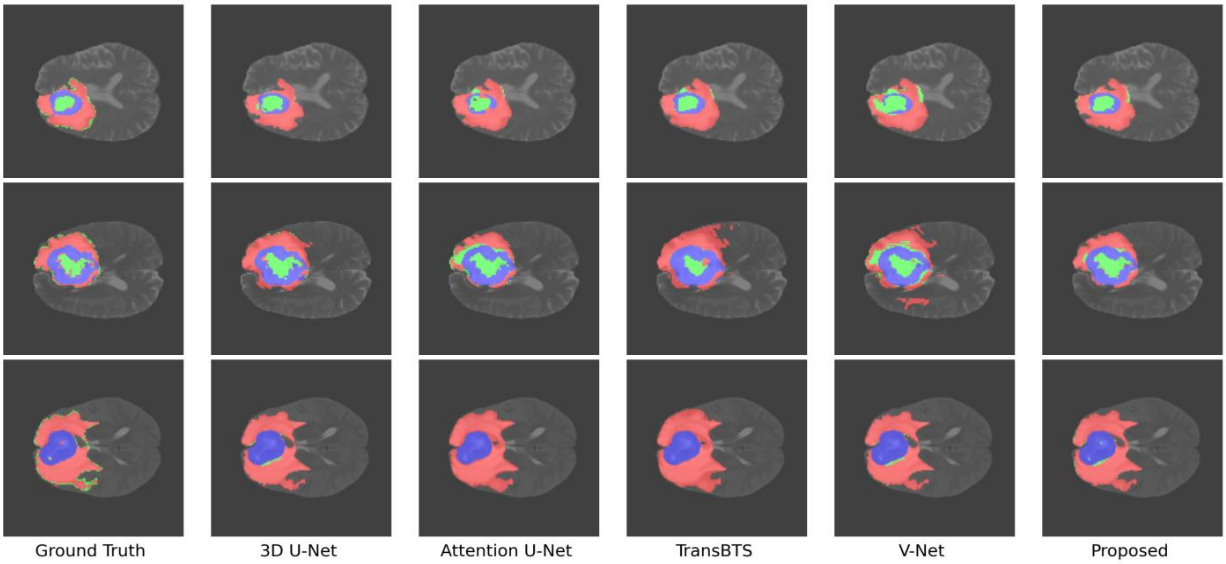


Fig. 7. 2D visual comparison of segmentation results across models on the artifact-free dataset generated by the proposed pipeline. Overlay colors: Red for WT, Green for TC, and Blue for ET (color figure available online).

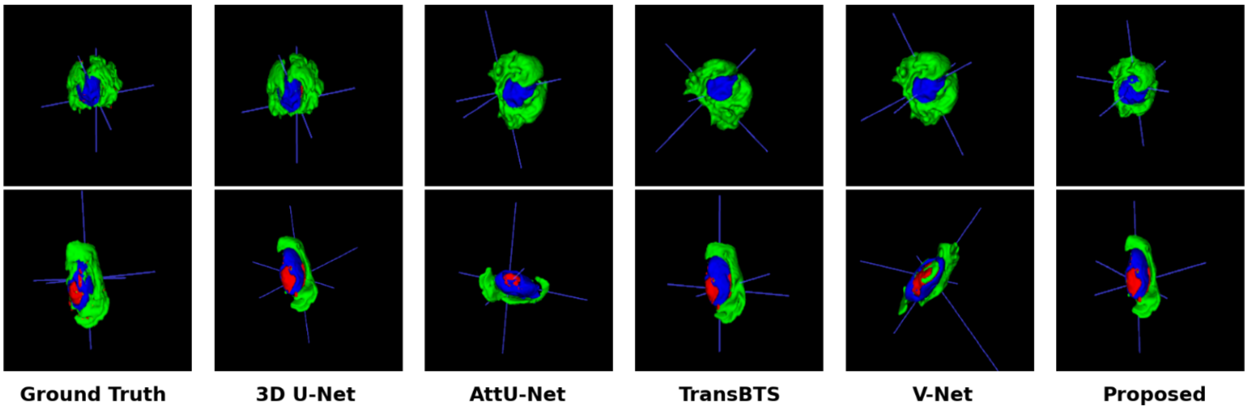


Fig. 8. 3D visual comparison of segmentation results across models on the artifact-free dataset generated by the proposed pipeline. Overlay colors: Red for WT, Green for TC, and Blue for ET (color figure available online).

The comparison between Tab. 5 and 6 shows that the proposed model achieves superior segmentation on the artifact-free dataset, validating the effectiveness of the unified framework. By integrating artifact correction and segmentation, the model enhances accuracy and robustness, overcoming the limitations of conventional models affected by imaging artifacts. This is further illustrated in Fig. 7 and Fig. 8, which provide 2D and 3D visual comparisons of segmentation results across models on the artifact-free dataset generated by the proposed pipeline.

Analysis of Fig. 7 and Fig. 8 shows that the second row, third-column image from Attention U-Net and the fifth column image from V-Net exhibit slight mispredictions, while other models demonstrate minor over or under segmentation, impacting tumor boundary precision. In Fig. 7, the 2D visual comparison highlights segmentation inconsistencies, where some models struggle with accurate tumor delineation. Figure 8 further illustrates these variations through 3D visualization, with Attention U-Net showing over-segmentation and TransBTS exhibiting slight overestimation. These differences stem from variations in model sensitivity to intensity, structural boundaries, and fine-grained tumor features. The proposed model, integrating artifact correction and attention-based feature extraction, achieves more consistent segmentation, effectively preserving structural integrity while enhancing accuracy and robustness.

4. Conclusion

This study introduces an artifact-aware deep learning framework for MRI brain tumor segmentation, integrating a diffusion-based artifact correction model with an advanced 3D CNN segmentation network. The proposed framework effectively enhances image quality and segmentation accuracy, as demonstrated by its superior performance compared to state-of-the-art methods such as SwinMR and Restormer-GAN. The model achieves a PSNR of 32.509 dB and an SSIM of 0.925, demonstrating its ability to restore high-quality, artifact-free MRI images. Additionally, the segmentation model consistently outperforms established baselines, achieving Mean Dice scores of 84.23% for ET, 89.20% for TC, and 91.85% for WT, while also obtaining lower HD95 values of 10.61 mm for ET, 3.47 mm for TC, and 3.51 mm for WT, indicating superior boundary accuracy.

Although the proposed framework significantly improves segmentation quality, it comes with computational complexity trade-offs. Incorporating SE layers, spatial attention, and axial attention increases GFLOPs and inference time, which may limit real-time clinical deployment. Future research will focus on optimizing computational efficiency through lightweight transformer architectures and model compression. Expanding the framework to other modalities, such as CT and ultrasound, will further validate its effectiveness. Additionally, integrating adaptive learning strategies will enhance model generalization, ensuring broader applicability in automated medical image analysis.

Acknowledgments

This work is supported by the Joint Project of Beijing Natural Science Foundation and Beijing Municipal Education Commission (No. KZ202110011015). The authors would like to thank CSC (China Scholarship Council) for funding to make this research possible lively and successful.

References

[1] GROCHANS, S., CYBULSKA, A. M., SIMINSKA, D., et al. Epidemiology of glioblastoma multiforme: Literature review. *Cancers,* 2022, vol. 14, no. 10, p. 1–32. DOI: 10.3390/cancers14102412

[2] ABD-ELLAH, M. K., AWAD, A. I., KHALAF, A. A., et al. A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned. *Magnetic Resonance Imaging*, 2019, vol. 61, p. 300–318. DOI: 10.1016/j.mri.2019.05.028

[3] PRICE, M., BALLARD, C., BENEDETTI, J., et al. CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2017–2021. *Neuro-oncology*, 2024, vol. 26, no. 6, p. vi1–vi85. DOI: 10.1093/neuonc/noae145

[4] HAACKE, E. M., BROWN, R. W., THOMPSON, M. R., et al. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. 1st ed. New York (USA): Wiley-Liss, 1999. ISBN 978-0471351283

[5] GODENSCHWEGER, F., KAGEBEIN, U., STUCHT, D., et al. Motion correction in MRI of the brain. *Physics in Medicine & Biology,* 2016, vol. 61, no. 5, p. R32–R56. DOI: 10.1088/0031-9155/61/5/R32

[6] VIONNET, L., ARANOVITCH, A., DUERST, Y., et al. Simultaneous feedback control for joint field and motion correction in brain MRI. *NeuroImage*, 2021, vol. 226, p. 1–18. DOI: 10.1016/j.neuroimage.2020.117286

[7] TAVAF, N., TORFI, A., UGURBIL, K., et al. GRAPPA-GANs for parallel MRI reconstruction. arXiv preprint, 2021, vol. 2101, p. 1 to 6. DOI: 10.48550/arXiv.2101.03135

[8] YIASEMIS, G., SÁNCHEZ, C. I., SONKE, J. J., et al. On retrospective k-space subsampling schemes for deep MRI reconstruction. *Magnetic Resonance Imaging*, 2024, vol. 107, p. 33–46. DOI: 10.1016/j.mri.2023.12.012

[9] BALAMURUGAN, T., GNANAMANOHARAN, E. Brain tumor segmentation and classification using hybrid deep CNN with LuNetClassifier. *Neural Computing and Applications*, 2023, vol. 35, no. 6, p. 4739–4753. DOI: 10.1007/s00521-022-07934-7

[10] ÇİÇEK, Ö., ABDULKADIR, A., LIENKAMP, S. S., et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Athens (Greece), October 2016, p. 424–432. DOI: 10.1007/978-3-319-46723-8_49

[11] ZOSSO, D., AN, J., STEVICK, J., et al. Image segmentation with dynamic artifacts detection and bias correction. *Inverse Problems and Imaging*, 2017, vol. 11, no. 3, p. 577–600. DOI: 10.3934/ipi.2017027

[12] SHAIK, F., PAVITHRA, P., SWARUPA RANI, K., et al. Image segmentation with complex artifacts and correction of bias. In *Proceedings of the 3rd International Conference on Communications and Cyber Physical Engineering (ICCCE 2020)*. Singapore, 2021, p. 519–526. DOI: 10.1007/978-981-15-7961-5_50

[13] KOETZIER, L. R., WU, J., MASTRODICASA, D., et al. Generating synthetic data for medical imaging. *Radiology*, 2024, vol. 312, no. 3, p. e232471. DOI: 10.1148/radiol.232471

[14] NARAI, A., HERMANN, P., AUER, T., et al. Movement-related artefacts (MR-ART) dataset of matched motion-corrupted and clean structural MRI brain scans. *Scientific Data*, 2022, vol. 9, no. 1, p. 1–6. DOI: 10.1038/s41597-022-01694-8

[15] ZAITSEV, M., MACLAREN, J., HERBST, M. Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging,* 2015, vol. 42, no. 4, p. 887–901. DOI: 10.1002/jmri.24850

[16] WANG, N. C., NOLL, D. C., SRINIVASAN, A., et al. Simulated MRI artifacts: Testing machine learning failure modes. *BME Frontiers*, 2022, vol. 2022, p. 1–16. DOI: 10.34133/2022/9807590

[17] BELJAARDS, L., PEZZOTTI, N., RAO, C., et al. AI-based motion artifact severity estimation in undersampled MRI allowing for selection of appropriate reconstruction models. *Medical Physics,* 2024, vol. 51, no. 5, p. 3555–3565. DOI: 10.1002/mp.16918

[18] OLSSON, H., MILLWARD, J. M., STARKE, L., et al. Simulating rigid head motion artifacts on brain magnitude MRI data–Outcome on image quality and segmentation of the cerebral cortex. *Plos One*, 2024, vol. 19, no. 4, p. e0301132. DOI: 10.1371/journal.pone.0301132

[19] SAFARI, M., YANG, X., FATEMI, A., et al. MRI motion artifact reduction using a conditional diffusion probabilistic model (MAR-CDPM). *Medical Physics*, 2024, vol. 51, no. 4, p. 2598–2610. DOI: 10.1002/mp.16844

[20] OKTAY, O., SCHLEMPER, J., FOLGOC, L. L., et al. Attention U-Net: Learning where to look for the pancreas. arXiv preprint, 2018, vol. 1804, p. 1–10. DOI: 10.48550/arXiv.1804.03999

[21] KARIMZADEH, R., FATEMIZADEH, E., ARABI, H., et al. Attention-based deep learning segmentation: Application to brain tumor delineation. In *Proceedings of the 2021 28th National and 6th International Iranian Conference on Biomedical Engineering (ICBME)*. Tehran (Iran), 2021, p. 248–252. DOI: 10.1109/ICBME54433.2021.9750374

[22] LIU, T., LUO, R., XU, L., et al. Spatial channel attention for deep convolutional neural networks. *Mathematics*, 2022, vol. 10, no. 10, p. 1–10. DOI: 10.3390/math10101750

[23] VAN LEEMPUT, K., MAES, F., VANDERMEULEN, D., et al. Automated model-based bias field correction of MR images of the

brain. *IEEE Transactions on Medical Imaging*, 1999, vol. 18, no. 10, p. 885–896. DOI: 10.1109/42.811268

[24] AL-MASNI, M. A., LEE, S., YI, J., et al. Stacked U-Nets with self-assisted priors towards robust correction of rigid motion artifact in brain MRI. *NeuroImage*, 2022, vol. 259, p. 1–17. DOI: 10.1016/j.neuroimage.2022.119411

[25] CHEN, G., XIE, H., RAO, X., et al. MRI motion correction through disentangled CycleGAN based on multi-mask K-space subsampling. *IEEE Transactions on Medical Imaging*, 2024, p. 1–16. DOI: 10.1109/TMI.2024.3523949

[26] CAI, T., LI, X., ZHONG, C., et al. DiffMAR: A generalized diffusion model for metal artifact reduction in CT images. *IEEE Journal of Biomedical and Health Informatics*, 2024, vol. 28, no. 11, p. 6712–6724. DOI: 10.1109/JBHI.2024.3439729

[27] VINCENT, P., LAROCHELLE, H., BENGIO, Y., et al. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning.* Helsinki (Finland), July 2008, p. 1096–1103. DOI: 10.1145/1390156.1390294

[28] HUANG, J., FANG, Y., WU, Y., et al. Swin transformer for fast MRI. *Neurocomputing,* 2022, vol. 493, p. 281–304. DOI: 10.1016/j.neucom.2022.04.051

[29] TSAI, T. H., LIN, Y. H., LIN, T. H., et al. Motion artifact correction in MRI using GAN-based channel attention transformer. In *Proceedings of the 2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. Toronto (Canada), 2023, p. 1–5. DOI: 10.1109/BioCAS58349.2023.10389083

[30] HSU, C., CHANG, C., CHEN, T. W., et al. Brain tumor segmentation (BraTS) challenge short paper: Improving three-dimensional brain tumor segmentation using SegResNet and hybrid boundary-dice loss. In *Proceedings of the International MICCAI Brainlesion Workshop*. Strasbourg (France), 2021, p. 334–344. DOI: 10.1007/978-3-031-09002-8_30

[31] ABBAS, R., GU, N. Improving deep learning-based image super-resolution with residual learning and perceptual loss using SRGAN model. *Soft Computing*, 2023, vol. 27, no. 21, p. 16041–16057. DOI: 10.1007/s00500-023-09126-4

[32] WANG, Z., BOVIK, A. C. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine,* 2009, vol. 26, no. 1, p. 98–117. DOI: 10.1109/MSP.2008.930649

[33] ZHAO, H., GALLO, O., FROSIO, I., et al. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging,* 2016, vol. 3, no. 1, p. 47–57. DOI: 10.1109/TCI.2016.2644865

[34] DE VERDIER, M. C., SALUJA, R., GAGNON, L., et al. *The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-Treatment MRI.* arXiv preprint, 2024, vol. 2405, p. 1–10. DOI: 10.48550/arXiv.2405.18368

[35] HENRY, T., CARRE, A., LEROUSSEAU, M., et al. Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: A BraTS 2020 challenge solution. In *Proceedings of Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: BrainLes 2020*. Lima (Peru), 2020, p. 327–339. DOI: 10.1007/978-3-030-72084-1_30

[36] WANG, W., CHEN, C., MENG, D., et al. TransBTS: Multimodal brain tumor segmentation using transformer. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention.* Strasbourg (France), 2021, p. 109 to 119. DOI: 10.48550/arXiv.2103.04430

[37] MILLETARI, F., NAVAB, N., AHMADI, S. A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 2016 4th International Conference on 3D Vision (3DV)*. Stanford (CA, USA), 2016, p. 565 to 571. DOI: 10.1109/3DV.2016.79

About the Authors ...

Mostafizur RAHMAN was born in 1986. He completed his B.Sc. in Electrical and Electronic Engineering (EEE) in 2009 and his M.Sc. in EEE in 2010, both at Islamic University, Kushtia, Bangladesh. He later received an additional M.Sc. in Computer Science from the same university in 2017. He is currently a Ph.D. candidate at Beijing Technology and Business University, Beijing, China. His research interests include pattern recognition, image processing, and computer vision.

Wenmin WANG received Ph.D. degree from Harbin Institute of Technology in 1989. Thereafter, he worked as an Associate Professor until 1991. He then gained overseas industrial experience for 18 years. Invited to return to China, he served from 2009 as a Professor at the School of Electronic and Computer Engineering, Peking University. Since 2019, he has been a Professor at the School of Computer Science and Engineering, Macau University of Science and Technology. His research interests include computer vision and multimedia processing.

Jiawei WANG, born in 1985, earned his M.D. and Ph.D. degree in 2013 from Nanjing University. He completed his postdoctoral fellowship from 2015 to 2017 at the Beijing Institute for Functional Neurosurgery at Xuanwu Hospital, Capital Medical University. Afterward, he joined the Department of Neurosurgery at the Cancer Hospital, Chinese Academy of Medical Sciences, where he currently serves as Associate Chief Physician. His research is centered on the study of brain tumors, brain functions, and the application of multimodal brain imaging techniques.

Yu WANG (corresponding author) was born in 1977. She received her Ph.D. degree from the University of Science and Technology Beijing in 2009. She was engaged in scientific research as a post-doctoral in the Beijing Key Laboratory of Multidimensional and Multiscale Computing Photography, Tsinghua University from 2009 to 2011. She is now a Professor and doctoral supervisor of Beijing Technology and Business University. Her research interests include pattern recognition, image processing and computer vision.