SPHERE: From MRI Sampling Mechanisms to Spatial Priors for Generalizable Brain Tumor Segmentation

Jiacheng Lu[1*], Shiyu Zhang[1*], Hui Ding[1✉], Junhui Xin[1], and Guoping Huo[2✉]

[1] College of Information Engineering, Capital Normal University, Beijing, China
[2] School of Artificial Intelligence, China University of Mining and Technology-Beijing, Beijing, China
jchengl@foxmail.com, sh1yuzh@163.com, dhui@cnu.edu.cn✉, xinjunhui721@163.com, kuoping@cumtb.edu.cn✉

Abstract. Brain tumor MRI segmentation is essential for clinical diagnosis and treatment planning. Although deep learning methods have achieved significant progress, most models operate solely in the image domain and overlook the influence of k-space sampling on spatial structure formation. This physical information gap often limits model generalization when encountering varying scanning conditions across different clinical centers. To address this issue, we propose **SPHERE**, **S**ampling-**P**rior **H**armonized r**E**construction-aware **R**epresentation l**E**arning, a sampling-aware framework for brain tumor MRI segmentation. SPHERE-Recover estimates k-space sampling–related geometric parameters directly from reconstructed MRI and transforms discrete scanning conditions into a continuous deformation vector field (DVF) that represents sampling-induced structural perturbations. Building on this prior, we develop SPHERE-Seg and introduce a Deformation Prior Module (DPM) to inject the DVF into feature modeling, enabling sampling-driven structural consistency learning. Experiments on BraTS and MSD show that SPHERE improves segmentation accuracy, boundary stability, and cross-dataset generalization. These results demonstrate that explicitly modeling MRI sampling mechanisms yields robust, generalizable, and efficient brain tumor segmentation.

Keywords: MRI Segmentation · K-space Sampling · Deformation Vector Field · Sampling-aware Learning · Cross-dataset Generalization

1 Introduction

Magnetic Resonance Imaging (MRI) is widely used for brain disease diagnosis and treatment planning, where automatic segmentation plays an important role in tumor delineation and lesion analysis [20, 27]. Recent deep learning methods have achieved strong performance on public datasets, particularly

* Equal contribution, ✉Co-corresponding authors

encoder–decoder architectures with long-range dependency modeling [23]. Subsequent studies further improved structural modeling using Transformer-based networks [29] and state space models such as Mamba [10]. Notably, the sliding-window attention of Swin Transformer [11] provides an efficient alternative to flattening or four-directional scanning strategies [34].

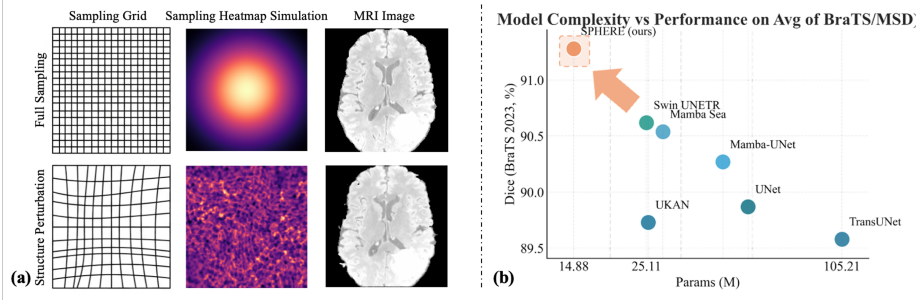

Fig. 1: Sampling-induced structural perturbations in MRI and their impact on segmentation. (a) Perturbed sampling causes spatially non-uniform patterns and geometric distortions. (b) Accuracy–complexity comparison on BraTS and MSD, showing that SPHERE achieves higher accuracy with fewer parameters.

MRI images are reconstructed by applying an inverse Fourier transform to sampled k-space signals. Acquisition parameters such as sampling trajectory, density, and phase-encoding order determine how spatial frequencies are preserved, often appearing as directional artifacts, edge blurring, and structural distortions in the image domain [15]. As a result, spatial structures are inherently shaped by the sampling mechanism before entering the segmentation network, and variations in sampling conditions may lead to different structural representations (Fig. 1(a)). However, most segmentation models implicitly assume statistically consistent inputs, which often break under varying scanning conditions and lead to reduced generalization [16].

A practical challenge is that most medical imaging datasets provide only reconstructed images, while complete k-space sampling information is available only in raw acquisition data. Without explicit sampling parameters, recovering imaging mechanisms and converting them into structural priors for segmentation remains largely unexplored. Recent studies suggest that MRI images implicitly encode acquisition information, making it possible to recover sampling mechanisms through learning [9, 19].

Motivated by these observations, we propose **SPHERE** (**S**ampling-**P**rior **H**armonized r**E**construction-aware **R**epresentation l**E**arning), a sampling-aware framework for brain tumor MRI segmentation. SPHERE recovers sampling mechanisms from MRI images and models sampling-induced structural perturbations as deformation priors, harmonizing sampling-aware priors with feature representations to improve cross-dataset generalization (Fig. 1(b)).

The main contributions of SPHERE are summarized as follows.

1. SPHERE-Recover. We propose a sampling parameter recovery method supervised by raw MRI acquisition data. The model predicts k-space sampling geometry directly from reconstructed MRI images. The recovered parameters are further used to simulate sampling-induced artifact distributions and derive a continuous deformation vector field (DVF), transforming discrete scanning conditions into a computable structural perturbation prior.

2. SPHERE-Seg. We develop the SPHERE-Seg segmentation framework with a Swin-RetNet backbone, which preserves local structural modeling while enhancing long-range dependency modeling, enabling effective representation of complex anatomical structures in high-resolution MRI.

3. Deformation Prior Module (DPM). We introduce DPM to inject the sampling-induced DVF into the segmentation network. Through feature alignment and retention modulation, the model learns structurally consistent representations and improves cross-dataset generalization under varying sampling conditions.

2 Related Work

2.1 MRI Sampling and K-space Artifacts

MRI images are reconstructed from signals sampled in the frequency domain (k-space), which corresponds to the Fourier representation of the image. Let $I(x, y)$ denote the spatial-domain image and $K(k_x, k_y)$ its k-space representation:

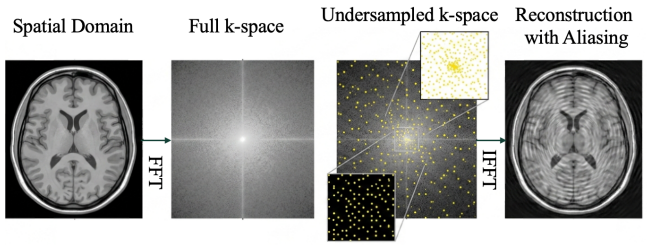

Fig. 2: Illustration of MRI image formation from k-space sampling. The spatial-domain image is transformed into k-space via Fourier transform. Undersampling in k-space leads to aliasing artifacts in the reconstructed image.

$$K(k_x, k_y) = \mathcal{F}(I(x, y)), \tag{1}$$

where $\mathcal{F}(\cdot)$ denotes the Fourier transform. In practical MRI acquisition, the sampling process can be written as

$$y = U_{\boldsymbol{\theta}}\, \mathcal{F}(x), \tag{2}$$

where x denotes the artifact-free image and $U_{\boldsymbol{\theta}}$ is a sampling operator determined by acquisition parameters $\boldsymbol{\theta}$ (e.g., sampling density and phase-encoding direction). When sampling is incomplete, the reconstructed image becomes

$$\hat{x} = \mathcal{F}^{-1}\big(U_{\boldsymbol{\theta}}\mathcal{F}(x)\big). \tag{3}$$

Undersampling in k-space often produces structured artifacts, such as aliasing, blurring, and ghosting, which propagate along the phase-encoding direction and degrade spatial structure representation [6, 17], as shown in Fig. 2. These artifacts arise from discontinuous sampling patterns and insufficient sampling density [30]. Variations in sampling parameters $\boldsymbol{\theta}$ therefore introduce systematic structural perturbations in reconstructed images, which can be interpreted as implicit deformation fields widely used in medical image registration [13].

Most existing studies focus on reconstructing missing k-space measurements or suppressing reconstruction artifacts using deep learning. In contrast, recovering sampling mechanisms from reconstructed images and exploiting them as structural priors for downstream vision tasks remains largely unexplored, especially in medical datasets that only provide reconstructed images (e.g., NIfTI or PNG).

2.2 RetNet and Long-Range Structural Modeling

Transformers achieve global modeling via self-attention, but their quadratic complexity with respect to sequence length limits their use on high-resolution medical images. RetNet [26] introduces a retention-based sequence modeling framework that maintains global dependency modeling with significantly lower computational cost [8].

The retention mechanism is formulated as

$$H_t = \sum_{i=1}^{t} D_{t-i} \odot (Q_t K_i^T) V_i, \tag{4}$$

where Q, K, and V are linearly projected features, D_{t-i} models relative distances through exponential decay, and $\odot$ denotes element-wise multiplication.

However, applying sliding-window strategies to RetNet remains underexplored. Moreover, existing methods have not considered integrating physical imaging priors, such as sampling-induced deformation fields, to improve robustness under varying scanning conditions.

3 Method

The spatial structure of MRI images is influenced not only by anatomy but also by the k-space sampling mechanism during acquisition. To model this physical prior, we propose the SPHERE segmentation framework (Fig. 3). SPHERE-Recover(Fig. 3(a)) first predicts sampling geometry parameters from reconstructed MRI images using a lightweight model. The predicted parameters are then used

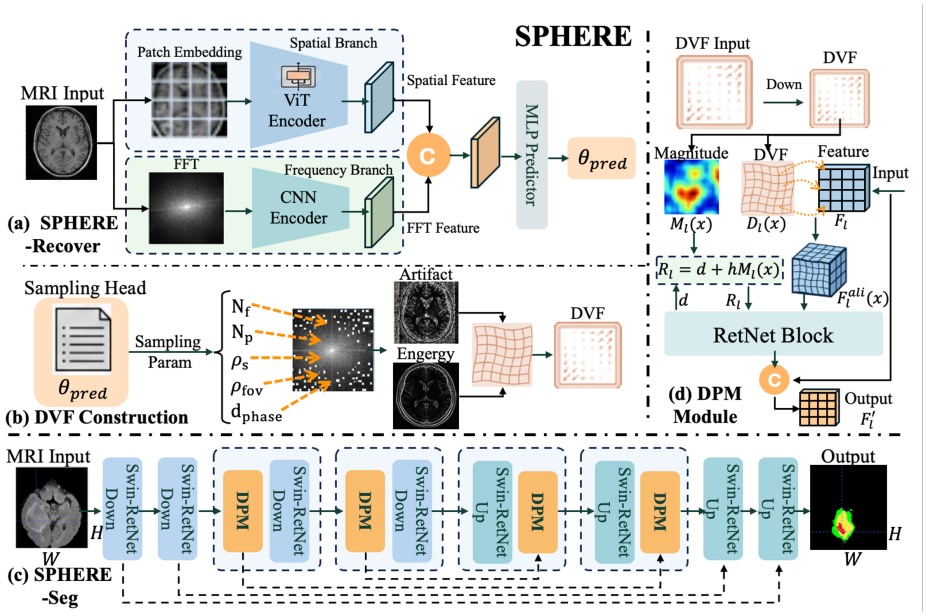

Fig. 3: Overview of **SPHERE**. (a) **SPHERE-Recover** estimates MRI sampling parameters from reconstructed MRI images by combining spatial and frequency-domain features. (b) The recovered parameters are used to simulate sampling-induced artifacts and construct the deformation vector field (DVF) as a structural prior. (c) **SPHERE-Seg** adopts a Swin-RetNet encoder–decoder architecture, where DPM modules inject the DVF prior into multi-scale feature modeling. (d) The **DPM** module integrates DVF information through feature alignment and retention modulation, enabling sampling-aware structural representation learning.

to construct a sampling-induced deformation vector field (DVF) as a structural prior(Fig. 3(b)). During segmentation, the DVF is injected into multi-scale feature modeling through the SPHERE-Seg(Fig. 3(c)) network and its core Deformation Prior Module (DPM)(Fig. 3(d)), enabling deformation-consistent structural representation.

3.1 SPHERE-Recover

The spatial structure of MRI images is influenced not only by anatomy but also by the sampling mechanism during acquisition. However, common segmentation datasets (e.g., NIfTI, PNG, or NumPy) do not provide sampling information. To recover this missing prior, we use DICOM MRI data with complete acquisition metadata to learn the sampling mechanism.

For each MRI image, we extract sampling-related parameters from the corresponding DICOM header:

$$\boldsymbol{\theta} = [N_f, N_p, \rho_s, \rho_{fov}, d_{phase}], \tag{5}$$

Table 1: MRI sampling parameters and their physical implications.

Param.	DICOM Tag	Physical Meaning	Image Impact
N_f	AcquisitionMatrix (Col.)	Frequency encoding dim.	Resolution / blur
N_p	AcquisitionMatrix (PhaseSteps)	Phase encoding dim.	Directional details
ρ_s	PercentSampling	Sampling ratio	Artifact strength
ρ_{fov}	PercentPhaseFOV	Phase FOV ratio	Structural compression
d_{phase}	PhaseEncodingDirection	Phase encoding dir.	Artifact orientation

where the physical meaning and sources of each parameter are summarized in Table 1.

To recover sampling parameters directly from images, we design a lightweight dual-branch model, **SPHERE-Recover**(Fig. 3(a)). The spatial branch uses a ViT encoder to capture global structural representations, while the frequency branch encodes the real and imaginary components of the 2D FFT using convolutional layers. The two features are concatenated and passed through a multilayer perceptron to predict the sampling parameters $\hat{\boldsymbol{\theta}}$.

During training, we adopt a joint regression and classification objective. For continuous sampling parameters, we use a mean squared error loss:

$$\mathcal{L}_{reg} = \sum_{i \in \mathcal{C}} \|\hat{\theta}_i - \theta_i\|_2^2, \tag{6}$$

where $\mathcal{C}$ denotes the set of continuous parameters. For the phase-encoding direction d_{phase}, we use a cross-entropy loss:

$$\mathcal{L}_{cls} = -\sum_{k} y_k \log \hat{y}_k. \tag{7}$$

The final objective is defined as

$$\mathcal{L}_{traj} = \mathcal{L}_{reg} + \lambda \mathcal{L}_{cls}, \tag{8}$$

where λ balances the regression and classification terms. Through this training process, SPHERE-Recover can infer MRI sampling mechanisms directly from reconstructed images, enabling estimation of scanning parameters even when DICOM metadata is unavailable for downstream segmentation tasks.

3.2 Sampling-Induced Artifact and Deformation Vector Field

The sampling parameters recovered from MRI images describe acquisition conditions but cannot be directly used for segmentation modeling. Therefore, we map the sampling mechanism into a continuous spatial perturbation field to explicitly characterize its impact on image structures, as shown in (Fig. 3(b)) and Fig. 4. Let the input image be

$$I \in \mathbb{R}^{H \times W}. \tag{9}$$

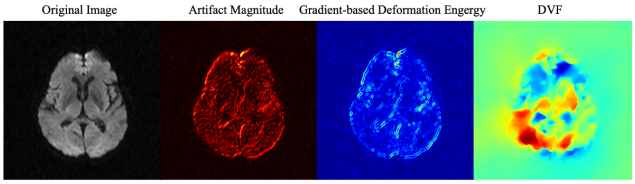

Fig. 4: Sampling-induced spatial structures. From left to right: original MRI image, artifact magnitude, gradient-based deformation energy, and the resulting deformation vector field (DVF), illustrating the geometric influence of the sampling process.

Given sampling parameters $\boldsymbol{\theta}$, the k-space sampling process can be written as

$$\tilde{I} = \mathcal{F}^{-1}(\mathcal{M}_{\boldsymbol{\theta}} \odot \mathcal{F}(I)), \tag{10}$$

where $\mathcal{M}_{\boldsymbol{\theta}}$ denotes the sampling operator determined by the sampling dimensions, sampling ratio, and phase-encoding direction. The reconstructed image $\tilde{I}$ thus represents a structurally perturbed version of I induced by the sampling mechanism. Based on the discrepancy between the original and sampled images, we define the sampling-induced artifact distribution as

$$A(\mathbf{x}) = |I(\mathbf{x}) - \tilde{I}(\mathbf{x})|. \tag{11}$$

To further characterize local structural variations, we compute the artifact gradient energy

$$E(\mathbf{x}) = \|\nabla A(\mathbf{x})\|_2. \tag{12}$$

Furthermore, the sampling-induced perturbation can be modeled as a continuous deformation between the original and sampled images. Let

$$\phi(\mathbf{x}) = \mathbf{x} + \mathbf{D}(\mathbf{x}) \tag{13}$$

denote the deformation mapping, where $\mathbf{D}(\mathbf{x})$ represents the displacement vector at location $\mathbf{x}$. The displacement field is obtained by minimizing brightness consistency with a smoothness constraint:

$$\mathbf{D} = \arg\min_{\mathbf{D}} \int_{\Omega} |I(\mathbf{x}) - \tilde{I}(\phi(\mathbf{x}))|^2 d\mathbf{x} + \lambda \int_{\Omega} \|\nabla \mathbf{D}(\mathbf{x})\|_2^2 d\mathbf{x}. \tag{14}$$

The resulting displacement field is referred to as the **sampling-induced deformation vector field (DVF)**. This structural prior will be injected into the subsequent segmentation network to guide feature modeling. The deformation field estimated in this work is not intended to represent an exact physical deformation caused by MRI acquisition.

In practice, k-space sampling artifacts may appear as aliasing, blurring, or ghosting. However, segmentation is mainly affected by spatial consistency,

boundary stability, and local structural changes. The DVF recovered by SPHERE is designed to capture these segmentation-relevant variations.

3.3 SPHERE-Seg and Deformation Prior Module

In the segmentation stage, we construct a U-shaped encoder–decoder network, termed **SPHERE-Seg**(Fig. 3(c). The network adopts a Swin-RetNet backbone with a hierarchical structure similar to Swin Transformer. The input image is first partitioned into local patches, and multi-scale representations are progressively built through hierarchical feature extraction. Within each stage, local window self-attention models fine-grained spatial structures, while the retention mechanism captures long-range dependencies, enabling both local detail preservation and global structural consistency in high-resolution MRI images.

Given an input MRI image I, the encoder extracts multi-scale features

$$F_l \in \mathbb{R}^{H_l \times W_l \times C_l}, \quad l = 1, \ldots, L. \tag{15}$$

To incorporate acquisition-related priors, we introduce the sampling-induced deformation vector field (DVF) $\mathbf{D}(\mathbf{x})$ derived in Sec. 3.2 as a geometric prior. The DVF is first downsampled to match the resolution of each feature scale:

$$\mathbf{D}_l(\mathbf{x}) \in \mathbb{R}^{H_l \times W_l \times 2}. \tag{16}$$

Accordingly, the deformation mapping is defined as

$$\phi_l(\mathbf{x}) = \mathbf{x} + \mathbf{D}_l(\mathbf{x}), \tag{17}$$

where $\mathbf{x}$ denotes spatial coordinates in the feature map. This mapping describes local structural displacements induced by the sampling mechanism.

To explicitly incorporate this geometric prior into feature modeling, we introduce a **Deformation Prior Module (DPM)** (Fig. 3(d))at the middle and high stages of the encoder–decoder (Stage 2–4). First, DPM performs DVF-guided spatial alignment:

$$F_l^{\text{ali}}(\mathbf{x}) = \mathcal{W}(F_l, \phi_l)(\mathbf{x}), \tag{18}$$

where $\mathcal{W}(\cdot)$ denotes a differentiable bilinear warping operator. This operation resamples features according to the DVF, improving geometric consistency of corresponding anatomical structures. The aligned features F_l^{ali} are also used as enhanced skip connections for the decoder. We further inject deformation information into cross-region feature aggregation. Let

$$M(\mathbf{x}) = \|\mathbf{D}(\mathbf{x})\|_2 \tag{19}$$

denote the magnitude of the DVF, and $M_l(\mathbf{x})$ its scale-specific version. The retention update is formulated as

$$F_l^{\text{ret}} = ((QK^T) \odot R_l)V, \tag{20}$$

where Q, K, and V denote query, key, and value projections of F_l^{ali}, and $\odot$ denotes the Hadamard product. R_l is the retention weight matrix defined as

$$R_l(\mathbf{x}, \mathbf{p}) = \gamma^{d(\mathbf{x},\mathbf{p})} + \eta\, h(M_l(\mathbf{x})), \tag{21}$$

where $d(\mathbf{x}, \mathbf{p})$ denotes the spatial distance function. The first term corresponds to the distance-based decay in RetNet, while the second term modulates the retention strength using the DVF magnitude. This design enhances structural consistency modeling in regions with significant deformation.

Finally, DPM fuses the updated features through a residual connection:

$$F_l' = F_l + \beta F_l^{\text{ret}}, \tag{22}$$

where β is a learnable parameter. The resulting features F_l' are then propagated through subsequent encoder and decoder stages.

Notably, SPHERE does not rely on DICOM metadata during inference. Given reconstructed MRI images (e.g., PNG, NPY, or NIfTI), the SPHERE-Recover module predicts sampling parameters and generates the corresponding deformation prior, which guides SPHERE-Seg for robust cross-dataset segmentation.

4 Experiments

4.1 Dataset

To learn MRI sampling mechanisms, we use the DICOM-based *MRI of the Brain to Recognize Pathologies* dataset [28], which provides complete acquisition metadata. For each image, the sampling parameter vector

$$\boldsymbol{\theta} = [N_f, N_p, \rho_s, \rho_{fov}, d_{phase}] \tag{23}$$

is extracted from the DICOM header and used to supervise the training of the SPHERE-Recover module.

Table 2: Datasets used in this study.

Dataset	Type	Slices	Modalities
MRI BRP	.dicom	351	T1/T2
BraTS-2019	.nii	51925	T1/T1ce/T2/FLAIR
BraTS-2023	.nii	193905	T1/T1ce/T2/FLAIR
MSD-Task01	.nii	75020	T1/T1Gd/T2/FLAIR

Segmentation experiments are conducted on the BraTS and MSD datasets. BraTS-2019 and BraTS-2023 [3,4,22] provide multi-modal MRI scans (T1, T1c, T2, and FLAIR) with annotations for enhancing tumor (ET), tumor core (TC), and whole tumor (WT). MSD-Task01 (Brain Tumor) [2, 25] follows a similar setting but labels ET as class 3.

All datasets are split into training and testing sets with a ratio of 8:2 by cases. Data preprocessing follows the nnUNet 2D pipeline [1,14]. Detailed dataset statistics are summarized in Table 2.

4.2 Implementation Details

Our framework is implemented in PyTorch and trained on an NVIDIA RTX 2080Ti GPU. For SPHERE-Recover, we use an ImageNet-1K pretrained ViT-B/16-224 with 768-dimensional features and fine-tune it on the DICOM dataset for θ regression. All experiments are repeated three times and the average results are reported. Segmentation performance is evaluated using the Dice coefficient and Hausdorff95 distance. The loss function is a weighted combination of Dice loss and binary cross-entropy (BCE) loss with a ratio of 2:8. All models are trained for 300 epochs using the Adam optimizer with an initial learning rate of 1×10^{-3} and early stopping with a patience of 50 epochs.

All models are trained on BraTS 2023 and evaluated on BraTS 2023, BraTS 2019, and MSD Task01 to assess both segmentation accuracy and cross-dataset generalization.

4.3 Evaluation Metrics

We evaluate SPHERE from three aspects: segmentation accuracy, sampling-parameter recovery, and deformation-prior consistency.

For segmentation, we report the Dice coefficient and the 95% Hausdorff distance (HD95). Dice measures region overlap:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}, \tag{24}$$

where TP, FP, and FN denote true positives, false positives, and false negatives. HD95 measures the 95th percentile of the bidirectional surface distance between predicted and ground-truth boundaries:

$$\text{HD95}(A, B) = \max\left(\max_{S_a \in S(A)} d(S_a, S(B)), \max_{S_b \in S(B)} d(S_b, S(A))\right), \tag{25}$$

where $S(\cdot)$ denotes boundary points and $d(\cdot)$ is the minimum point-to-set distance.

For sampling-parameter recovery, we report MAE, RMSE, and MedAE in the normalized parameter space. For ratio parameters, MAPE is additionally used, while parameters with physical units are denormalized and evaluated by Phys-MAE.

For deformation-prior consistency, we compare the predicted DVF with the DVF generated from ground-truth sampling parameters. Given the ground-truth and predicted displacement fields (u, v) and $(\hat{u}, \hat{v})$, we compute vector-field L1/L2 errors, magnitude errors (Mag-L1/Mag-L2), and SSIM on DVF magnitude maps.

Table 3: Dimension-wise sampling parameter prediction on the test set.

Param	MAE↓	RMSE↓	MedAE↓	MAPE↓	Phys-MAE↓
N_f	0.08531	0.10516	0.07212	–	21.50
N_p	0.08264	0.10443	0.06822	–	17.52
ρ_s	0.01614	0.02282	0.01049	2.02%	–
ρ_{fov}	0.01806	0.02008	0.01844	2.28%	–
d_p(row)	0.06871	0.10044	0.04254	–	–
d_p(col)	0.06994	0.10144	0.04666	–	–
Avg	**0.05680**	**0.08492**	**0.04308**	**2.15%**	**19.51**

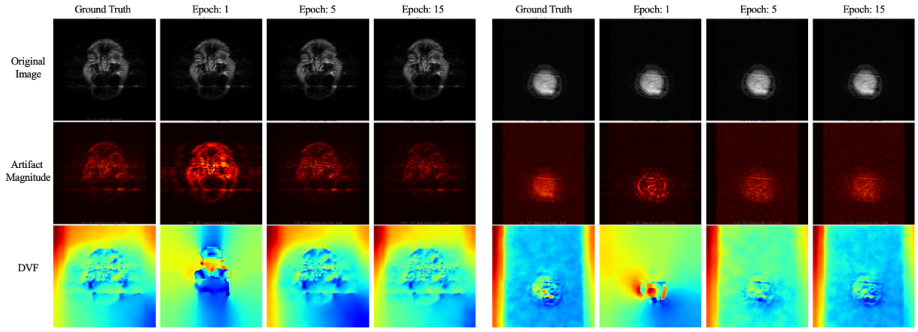

Fig. 5: Artifact magnitude and DVF evolution during training, showing convergence to sampling-consistent spatial patterns.

4.4 Feasibility Analysis of Sampling Mechanism Recovery

This section evaluates the recoverability, stability, and physical consistency of **SPHERE-Recover**.

Quantitative results on the independent test set are summarized in Table 3. The overall MAE, RMSE, and MedAE are 0.05680, 0.08492, and 0.04308. The physical-space MAE for frequency and phase dimensions are 21.50 and 17.52, while the MAPE for sampling ratio and phase FOV ratio are 2.02% and 2.28%. For the discrete phase-encoding direction, SPHERE-Recover achieves 100% classification accuracy.

Table 4 further reports physical interpretability and deformation-field consistency metrics. The recovered parameters generate stable deformation fields with consistent spatial patterns and high structural similarity (SSIM = 0.7844).

Fig. 5 visualizes artifact magnitude and DVF during training. Both gradually converge to spatially coherent patterns consistent with the sampling mechanism, indicating that the recovered parameters provide physically consistent geometric priors for subsequent DPM-based segmentation.

In addition, to validate the effectiveness of the joint spatial-frequency design in SPHERE-Recover, Table 5 compares it with the ViT-only and frequency-only variants. Its consistent superiority demonstrates that SPHERE-Recover

Table 4: Physical interpretability and DVF consistency evaluation.

Physical		**DVF**	
Metric	Value	Metric	Value
Phase Acc↑	100%	Vector L1↓	0.1334
Sampling MAPE↓	2.02%	Vector L2↓	0.1174
FOV MAPE↓	2.28%	Mag L1↓	0.0553
Freq MAE↓	21.50	Mag L2↓	0.0679
Phase MAE↓	17.52	Mag SSIM↑	0.7844

Table 5: Ablation of SPHERE-Recover designs using MAE ↓.

Recover Design	$\mathbf{N_f}$	$\mathbf{N_p}$	$\boldsymbol{\rho_s}$	$\boldsymbol{\rho_{fov}}$	$\mathbf{d_p(row)}$	$\mathbf{d_p(col)}$
Frequency-only	0.10373	0.16151	0.03820	0.03572	0.09575	0.12816
ViT-only	0.09428	0.09135	0.01920	0.02272	0.08196	0.08534
SPHERE-Recover	**0.08531**	**0.08264**	**0.01614**	**0.01806**	**0.06871**	**0.06994**

can effectively exploit both spatial cues and frequency-domain cues from k-space sampling for θ estimation.

4.5 Ablation Studies

To evaluate sampling-aware priors and DPM fusion, all models are trained on BraTS2023 and tested on BraTS2023 / BraTS2019, as shown in Table 6.

Without any sampling prior, the Base model obtains average Dice scores of 92.25 / 89.79. Using the recovered sampling parameter as a theta prior brings only marginal gains, and Aux-ViT also fails to improve cross-dataset performance. This indicates that the improvement does not simply come from extra inputs, an auxiliary ViT branch, or increased model capacity. In contrast, DVF-based modeling provides more consistent gains.

Under the same multiply fusion setting, Artifact Magnitude (A) provides limited improvement, Gradient-based Deformation Energy (G) achieves moderate gains, and the DVF prior performs best with 92.52 / 91.03. This suggests that DVF better captures segmentation-relevant spatial and structural variations induced by k-space sampling.

We further verify the necessity of structured DVF. E2E-DVF, which removes the explicit $\theta \rightarrow$ DVF construction, performs worse than full SPHERE. Shuffle-DVF further drops to 91.65 / 89.36, confirming that the spatial organization of DVF carries meaningful structural information rather than acting as simple regularization.

Finally, with DVF as the prior, Multiplication, SE, and convolution fusion achieve 92.52 / 91.03, 92.32 / 90.72, and 92.66 / 90.94, respectively. RetNet fusion performs best with 92.87 / 91.27. Overall, SPHERE benefits from explicit sampling recovery, structured DVF prior, and RetNet-based fusion, rather than simple module stacking.

Table 6: Extended ablation study of sampling-aware priors and DPM fusion strategies. All models are trained on BraTS2023 and evaluated on BraTS2023 / BraTS2019.

Method	Prior				Fusion				Dice Score ↑			
	θ	A	G	D	Mul	SE	Conv	RetNet	WT	TC	ET	Avg
Control Baselines												
Base	–	–	–	–	–	–	–	–	90.83 / 89.44	93.05 / 90.29	92.88 / 89.64	92.25 / 89.79
θ Prior	✓	–	–	–	–	–	–	–	90.72 / 89.74	92.87 / 90.48	93.26 / 91.11	92.28 / 90.44
Aux-ViT	–	–	–	–	–	–	–	–	90.68 / 89.20	92.85 / 90.06	93.42 / 89.63	92.32 / 89.63
E2E-DVF	–	–	–	✓	–	–	–	✓	91.01 / 89.96	93.44 / 90.72	93.30 / 91.88	92.58 / 90.85
Shuffle-DVF	–	–	–	✓	–	–	–	✓	90.40 / 88.82	92.61 / 89.70	91.95 / 89.57	91.65 / 89.36
Prior Comparison, fusion = Multiply												
Artifact Prior	–	✓	–	–	✓	–	–	–	90.55 / 89.67	92.83 / 90.51	92.83 / 89.60	92.07 / 89.93
Gradient Prior	–	–	✓	–	✓	–	–	–	91.04 / 89.63	93.20 / 90.87	93.11 / 90.85	92.45 / 90.45
DVF Prior	–	–	–	✓	✓	–	–	–	91.04 / 90.09	93.25 / 90.96	93.28 / 91.64	92.52 / 91.03
Fusion Comparison, prior = DVF												
DVF + SE	–	–	–	✓	–	✓	–	–	90.87 / 89.93	93.04 / 90.89	93.04 / 91.35	92.32 / 90.72
DVF + Conv	–	–	–	✓	–	–	✓	–	91.15 / 90.26	93.36 / 91.17	93.47 / 91.62	92.66 / 90.94
SPHERE	–	–	–	✓	–	–	–	✓	**91.25** / **90.25**	**93.67** / **91.33**	**93.69** / **92.24**	**92.87** / **91.27**

θ: Direct sampling-parameter prior. **A**: Artifact Magnitude. **G**: Gradient-based Deformation Energy. **D**: Deformation Vector Field. **Aux-ViT**: Auxiliary ViT branch. **E2E-DVF**: End-to-end DVF prediction. **Shuffle-DVF**: Shuffled DVF structure.

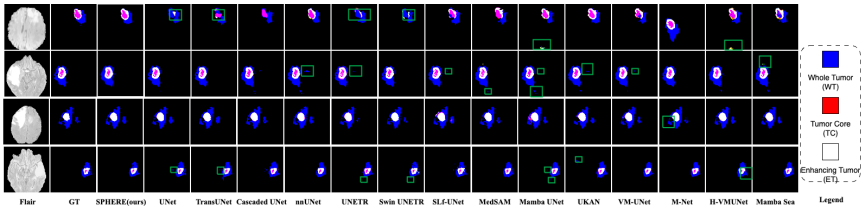

Fig. 6: Examples of segmentation results from multiple methods. From left to right: Flair modality input image, Ground Truth (GT), and segmentation results from various comparison algorithms.

4.6 Comparison with State-of-the-Art Methods

To evaluate the effectiveness of the proposed method, we compare SPHERE-Seg with representative approaches, including UNet [23], Cascaded UNet [18], TransUNet [5], nnUNet [14], UNETR [12], Swin UNETR [11], MedSAM [32], Mamba-UNet [31], M-Net [21], VM-UNet [24], H-VMUNet [33], and Mamba Sea [7]. All models are trained on BraTS 2023 and evaluated on BraTS 2023, BraTS 2019, and MSD Task01 to assess both in-domain performance and cross-dataset generalization.

Dice results are reported in Table 7. On BraTS 2023, SPHERE-Seg achieves the best performance on TC and ET (93.67% and 93.69%), while maintaining competitive WT performance. More importantly, under cross-dataset evaluation, SPHERE-Seg consistently achieves the best or second-best results on BraTS 2019 and MSD Task01. For example, it reaches 90.25%, 91.33%, and 92.24% on BraTS 2019, and 88.82%, 89.75%, and 90.51% on MSD Task01, outperforming existing

Table 7: Comparison with state-of-the-art methods when trained on BraTS 2023 and evaluated on BraTS 2023, BraTS 2019, and MSD Task01 datasets (Dice score %).

Model	Year	Dice Score (%)		
		WT↑	TC↑	ET↑
UNet	2015	90.71 / 88.16 / 87.03	93.05 / 90.48 / 87.85	93.36 / 89.97 / 88.25
Cascaded UNet	2019	90.32 / 87.77 / 87.49	92.85 / 89.67 / 87.32	92.47 / 89.50 / 88.19
TransUNet	2021	90.71 / 87.16 / 87.20	92.52 / 90.30 / 87.21	92.92 / 90.19 / 88.02
nnUNet	2021	90.34 /**89.44**/ 88.22	92.74 /**91.01**/ 88.36	92.37 /**91.33**/ 89.83
UNETR	2022	88.35 / 87.80 / 87.35	89.16 / 87.87 / 88.28	91.43 / 90.59 / 88.73
Swin UNETR	2022	91.11 / 89.61 / 87.90	93.20 / 90.81 / 88.51	93.42 / 91.28 /**89.77**
SLf-UNet	2024	90.81 / 87.26 / 87.44	93.18 / 89.24 / 87.31	93.30 / 90.37 / 88.64
MedSAM	2024	88.55 / 88.04 / 87.30	91.55 / 90.35 / 87.15	90.30 / 89.92 / 87.52
Mamba-UNet	2024	91.03 / 87.58 / 87.66	93.32 / 90.95 / 88.32	93.31 / 90.88 / 89.38
UKAN	2024	90.64 / 88.66 / 87.24	93.04 / 90.27 / 88.10	93.14 / 87.90 / 88.62
VM-UNet	2024	90.52 / 88.07 / 87.33	93.40 / 90.88 / 88.39	**93.50**/ 90.41 / 88.77
M-Net	2025	**91.33**/ 89.23 / 88.04	**93.55**/ 91.00 /**88.42**	93.42 / 90.43 / 89.45
H-VMUNet	2025	90.77 / 88.92 / 88.15	93.05 / 90.50 / 88.27	92.88 / 90.77 / 89.12
Mamba Sea	2025	91.12 / 88.87 /**88.41**	93.41 / 90.64 / 88.26	93.47 / 91.08 / 89.61
SPHERE-Seg (ours)	–	**91.25**/**90.25**/**88.82**	**93.67**/**91.33**/**89.75**	**93.69**/**92.24**/**90.51**

methods overall. Hausdorff95 results (Table 8) further confirm this advantage. SPHERE-Seg achieves the lowest HD95 across datasets and tumor regions, indicating more accurate and stable boundary localization. Qualitative comparisons in Fig. 6 show that SPHERE-Seg produces more continuous and accurate tumor boundaries, better capturing complex tumor structures while reducing false predictions.

We further analyze the computational complexity of the proposed framework. CNN-based models such as UNet require 39.40M parameters and 321.19G FLOPs, while Transformer-based TransUNet contains 105.21M parameters. Recent architectures reduce computation to some extent, e.g., Swin UNETR (25.11M parameters, 106.80G FLOPs), Mamba-UNet (72.44G FLOPs), and VM-UNet (61.42G FLOPs). In contrast, SPHERE-Seg uses only 14.88M parameters and 44.28G FLOPs, achieving the fastest segmentation inference time of 8:09. The ViT-based SPHERE-Recover module is used only for preprocessing to estimate θ, and SPHERE-Seg performs inference using only the DVF derived from θ. Thus, SPHERE-Recover adds no parameter or inference overhead to SPHERE-Seg.

These results demonstrate that the proposed sampling-aware structural prior improves cross-dataset generalization, boundary stability, and computational efficiency.

Table 8: Comparison with state-of-the-art methods when trained on BraTS 2023 and evaluated on BraTS 2023, BraTS 2019, and MSD Task01 datasets (Hausdorff95).

Model	Year	Hausdorff95		
		WT↓	TC↓	ET↓
UNet	2015	1.1863 / 1.4921 / 1.5117	0.7329 / 1.0152 / 1.5838	0.6730 / 0.8843 / 0.9129
Cascaded UNet	2019	1.2091 / 1.4698 / 1.5232	0.7488 / 1.0039 / 1.6492	0.7591 / 0.8341 / 0.8747
TransUNet	2021	1.1810 / 1.5381 / 1.5328	0.7276 / 1.1284 / 1.5383	0.6869 / 0.8693 / 0.9026
nnUNet	2021	1.2100 /**1.4352**/**1.4492**	0.7358 / 0.9554 / 1.2341	0.6722 / 0.8120 /**0.8402**
UNETR	2022	1.2427 / 1.4981 / 1.4870	0.8926 / 0.9824 / 1.4995	0.7211 / 0.9102 / 0.8672
Swin UNETR	2022	1.1629 / 1.4571 / 1.4681	0.7088 / 0.9312 / 1.1289	0.6631 / 0.8145 / 0.8443
SLf-UNet	2024	1.1748 / 1.4733 / 1.5111	0.7100 / 0.9808 / 1.5703	0.6709 / 0.8174 / 0.9175
MedSAM	2024	1.3155 / 1.5319 / 1.5203	0.8003 / 1.0975 / 1.6643	0.8153 /**0.7956**/ 0.9387
Mamba-UNet	2024	1.1734 / 1.4719 / 1.4733	0.7008 / 0.9647 / 1.1580	0.6764 / 0.8532 / 0.8751
UKAN	2024	1.1862 / 1.4724 / 1.5220	0.7234 / 0.9845 / 1.1593	0.6824 / 0.8674 / 0.8930
VM-UNet	2024	1.1806 / 1.4554 / 1.4839	0.7079 / 0.9632 / 1.1266	0.6781 / 0.8422 / 0.8502
M-Net	2025	**1.1534**/ 1.4383 / 1.4599	0.7069 / 0.9490 /**1.1155**	**0.6600**/ 0.8307 / 0.8443
H-VMUNet	2025	1.1833 / 1.4414 / 1.4914	0.7229 / 0.9543 / 1.1209	0.6931 / 0.8535 / 0.8736
Mamba Sea	2025	1.1837 / 1.4892 / 1.4882	**0.6977**/**0.9258**/ 1.1337	0.6728 / 0.8422 / 0.8675
SPHERE-Seg (ours)	–	**1.1533**/**1.3852**/**1.4237**	**0.6948**/**0.8876**/**0.9789**	**0.6471**/**0.7659**/**0.8202**

5 Conclusion

We propose SPHERE, a sampling-aware framework for MRI segmentation. Unlike conventional methods that rely solely on image-domain features, SPHERE explicitly models the MRI acquisition mechanism. SPHERE-Recover estimates k-space sampling parameters from reconstructed images and converts discrete scanning conditions into a continuous deformation vector field (DVF). Based on this prior, the SPHERE-Seg network integrates DVF through the Deformation Prior Module (DPM) to enable sampling-driven structural consistency modeling.

Experiments on BraTS and MSD demonstrate that SPHERE improves segmentation accuracy, boundary stability, and cross-dataset generalization. These results suggest that explicitly incorporating MRI sampling mechanisms into visual models can enhance robustness to varying scanning conditions and provide a promising direction for integrating imaging physics with medical vision models.

Acknowledgments

This work was supported by the Interdisciplinary Project of Capital Normal University (Grant No. 2026JCYY04), the Beijing Natural Science Foundation (Grant No. QY26478), and the National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University (Grant No. HMHAI-202407).

References

1. Al Shalabi, L., Shaaban, Z., Kasasbeh, B.: Data mining: A preprocessing engine. Journal of Computer Science **2**(9), 735–739 (2006)
2. Antonelli, M., Reinke, A., Bakas, S., et al.: The medical segmentation decathlon. Nature Communications (2022). https://doi.org/10.1038/s41467-022-30695-9
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
4. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Jambawalikar, S.R.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint (2018)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
6. Chen, L., Bao, L., Li, J., et al.: An aliasing artifacts reducing approach with random undersampling for spatiotemporally encoded single-shot mri. Journal of Magnetic Resonance **237**, 115–124 (2013)
7. Cheng, Z., Guo, J., Zhang, J., et al.: Mamba-sea: A mamba-based framework with global-to-local sequence augmentation for generalizable medical image segmentation. IEEE Transactions on Medical Imaging (2025)
8. Fan, Q., Huang, H., Chen, M., et al.: Rmt: Retentive networks meet vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5641–5651 (2024)
9. Gösche, E., Eghbali, R., Knoll, F., et al.: Domain influence in mri medical image segmentation: Spatial versus k-space inputs. In: International Workshop on Machine Learning in Medical Imaging. pp. 310–319. Springer (2024)
10. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
11. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Cham: Springer International Publishing (2021)
12. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
13. He, Z., He, Y., Cao, W.: Deformable image registration with attention-guided fusion of multi-scale deformation fields. Applied Intelligence **53**(3), 2936–2950 (2023)
14. Isensee, F., Jaeger, P.F., Kohl, S.A.A., et al.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
15. Jhamb, T.K., Rejathalal, V., Govindan, V.K.: A review on image reconstruction through mri k-space data. International Journal of Image, Graphics and Signal Processing **7**(7), 42–49 (2015)
16. Krupa, K., Bekiesińska-Figatowska, M.: Artifacts in magnetic resonance imaging. Polish Journal of Radiology **80**, 93–106 (2015)
17. Lee, D., Yoo, J., Ye, J.C.: Deep artifact learning for compressed sensing and parallel mri. arXiv preprint arXiv:1703.01120 (2017)

18. Li, X., Luo, G., Wang, K.: Multi-step cascaded networks for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 163–173. Springer International Publishing, Cham (2019)
19. Li, Y., Qi, Y., Hu, Z., et al.: A novel automatic segmentation method directly based on magnetic resonance imaging k-space data for auxiliary diagnosis of glioma. Quantitative Imaging in Medicine and Surgery **14**(2), 2008–2020 (2024)
20. Liang, Z.P., Lauterbur, P.C.: Principles of magnetic resonance imaging. SPIE Optical Engineering Press, Bellingham (2000)
21. Lu, J., Ding, H., Zhang, S., et al.: M-net: Mri brain tumor sequential segmentation network via mesh-cast. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 20116–20125 (2025)
22. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Van Leemput, K.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer International Publishing (2015)
24. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. ACM Transactions on Multimedia Computing, Communications and Applications (2024)
25. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms (2019)
26. Sun, Y., Dong, L., Huang, S., et al.: Retentive network: A successor to transformer for large language models. arXiv preprint arXiv:2307.08621 (2023)
27. Tan, A.C., Ashley, D.M., López, G.Y., Malinzak, M., Friedman, H.S., Khasraw, M.: Management of glioblastoma: State of the art and future directions. CA: a cancer journal for clinicians **70**(4), 299–312 (2020)
28. UniData: Brain mri dataset (dicom images). `https://unidata.pro/datasets/brain-mri-image-dicom/` (2025), accessed: 2026-03-01
29. Vaswani, A.: Attention is all you need. In: Advances in Neural Information Processing Systems (2017)
30. Wang, M., Li, P.: A review of deformation models in medical image registration. Journal of Medical and Biological Engineering **39**(1), 1–17 (2019)
31. Wang, Z., Zheng, J.Q., Zhang, Y., Cui, G., Li, L.: Mamba-unet: Unet-like pure visual mamba for medical image segmentation. arXiv preprint (2024)
32. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
33. Wu, R., Liu, Y., Liang, P., et al.: H-vmunet: High-order vision mamba unet for medical image segmentation. Neurocomputing **624**, 129447 (2025)
34. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint (2024)