SYMMETRYBREAK: RESOLVING BILATERAL AMBIGUITY IN TRAINING-FREE BRAIN MRI ANOMALY LOCALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Localizing brain abnormalities from an unlabeled MRI without task-specific training, lesion annotations, or external healthy-brain references remains challenging. Although the brain's bilateral organization provides an intrinsic within-subject reference, conventional contralateral comparison can identify where the hemispheres differ without determining which side is abnormal, often producing mirrored responses. We introduce *SymmetryBreak*, a training-free framework that converts bilateral latent asymmetry into directional abnormality localization. Given a single MRI slice, a frozen vision encoder produces spatially organized latent representations, which are divided into corresponding left and right hemispheres through deterministic horizontal mirroring. *SymmetryBreak* first computes a bilateral cosine discrepancy map to identify regions of symmetry breaking. It then constructs an *image-internal reference* from low-discrepancy correspondences and measures side-specific latent deviations from this reference. A shared-scale directional comparison determines which hemisphere exhibits stronger abnormality, and a global hemisphere decision suppresses mirrored responses before constructing a localization map. *SymmetryBreak* requires no task-specific training or external healthy-brain atlases. Across **BraTS 2019**, **BraTS 2021**, and **BraTS-Africa**, *SymmetryBreak* achieves an average Dice Similarity Coefficient (DSC) of **67.4**% and an AUPRC of **78.4**%, outperforming the evaluated baselines. These results demonstrate that intrinsic anatomical symmetry, combined with image-internal directional inference, can provide a simple training-free paradigm for localized brain abnormality detection without an external normal reference.

1 INTRODUCTION

Brain tumors exhibit substantial heterogeneity in morphology, spatial distribution, and tissue composition Martucci et al. (2023). Magnetic resonance imaging (MRI) is central to assessment, diagnosis, and treatment monitoring Martucci et al. (2023); Villanueva-Meyer et al. (2017). Accurate localization typically relies on task-specific supervision, such as voxel-level annotations or reference datasets, which are restrictive when labeled data are scarce or non-transferable across protocols Deverdun et al. (2023); Fathi Kazerooni et al. (2023).

Large-scale foundation models offer transferable representations without task-specific retraining Zhang & Metaxas (2024); Ghamizi et al. (2026); Ma et al. (2024). However, a frozen representation alone does not define pathology. Without task-specific learning, can pathological tissue be localized using only the internal structure of an unlabeled MRI and a frozen visual representation?

The brain's bilateral symmetry provides a natural within-subject reference: a focal lesion manifests as a violation of expected bilateral similarity. Comparing corresponding latent representations from both hemispheres yields a discrepancy signal. Yet, bilateral discrepancy introduces a fundamental ambiguity: it establishes that two regions differ, but not *which side is abnormal*. This non-directional signal often causes mirrored responses, where pathology on one side induces false anomaly evidence on the opposite side. Normal anatomical variation further contributes to mismatch Toga & Thompson (2003); Martins et al. (2019). Thus, detecting symmetry breaking and attributing it to a specific hemisphere are distinct problems.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

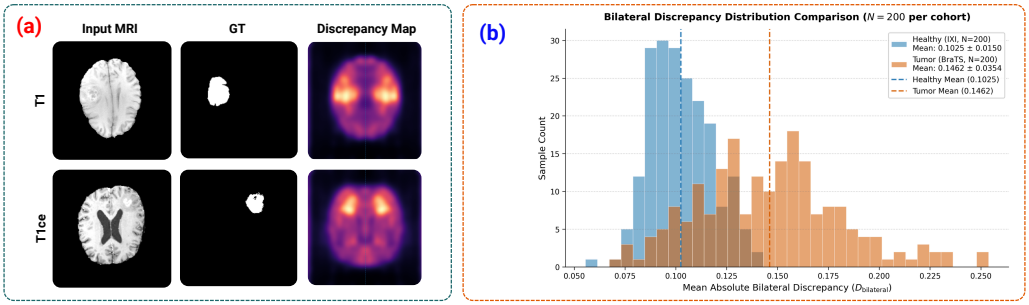


Figure 1: ***SymmetryBreak*** **addresses the directional ambiguity of bilateral latent discrepancy.** (a) A bilateral discrepancy map identifies regions where corresponding hemispheric representations differ but cannot determine which hemisphere is pathological, producing potentially mirrored responses. (b) Mean bilateral latent discrepancy is higher for tumor images than for healthy images, motivating bilateral discrepancy as a candidate signal for further directional attribution. Distributions are computed over 200 images per cohort from IXI BIA Group (2006) and BraTS-2021, respectively. Dashed lines denote cohort means.

We formulate this task as *directional anomaly attribution*, explicitly separating two questions: *where do corresponding representations disagree?* and *which hemisphere exhibits stronger abnormality?* We construct an *image-internal reference* from regions with low bilateral discrepancy. Because non-pathological tissue remains largely preserved, these regions provide a patient-specific baseline for typical latent behavior. Deviations from this reference are evaluated on a robust scale, and a global hemisphere decision aggregates this evidence to suppress mirrored responses, constraining localization to a single abnormal hemisphere.

We introduce ***SymmetryBreak*** **(SB)**, a training-free framework for brain tumor localization from frozen spatial representations. Given an unlabeled MRI slice, SB establishes left-right correspondence via horizontal mirroring, computes a bilateral cosine discrepancy map, and calculates side-specific deviations relative to the image-internal reference. These deviations yield directional evidence that guides global hemisphere selection and produces a one-sided anomaly map. SB requires no fine-tuning or external healthy atlases. Our contributions are:

- Formulate training-free tumor localization as a **directional anomaly attribution problem**, disentangling symmetry breaking from side attribution.
- Propose ***SymmetryBreak***, integrating bilateral discrepancy with an image-internal reference for localized attribution without external data or fine-tuning.
- Introduce a **global hemisphere selection mechanism** to constrain predictions to the pathological side, eliminating mirrored false positives.
- Evaluate SB across BraTS-2019, BraTS-2021, and BraTS-Africa Menze et al. (2014); Baid et al. (2021); Adewole et al. (2024), demonstrating strong performance and validating core components via ablations.

2 Related Work

Foundation Models for Training-Free Medical Image Analysis. Foundation models have increasingly enabled medical image analysis with transferable visual representations and promptable inference, reducing the dependence on task-specific model design and annotated datasets. Adaptations of the Segment Anything Model (SAM) have explored zero-shot and prompt-based segmentation of anatomical structures and pathological regions, including self-prompting, semantic-guided region selection, and domain-specific adaptation Kirillov et al. (2023); Wu & Xu (2024); Aleem et al. (2024); Jiang et al. (2024); Ravi et al. (2025); Oquab et al. (2023). Other foundation models provide general-purpose visual representations that can be transferred across tasks without conventional supervised training Radford et al. (2021); Caron et al. (2021). SaLIP combines SAM and CLIP to identify target-relevant regions through visual-semantic alignment and subsequently refines the selected regions through prompting Aleem et al. (2024). ZePT investigates zero-shot pan-tumor segmentation through query disentangling, self-prompting, and query–knowledge alignment Jiang

et al. (2024), while Malenia explores zero-shot tumor segmentation in volumetric medical imaging through tumor-level representations and semantic knowledge injection Jiang et al. (2025). These approaches demonstrate the potential of pretrained representations for reducing annotation requirements, but commonly rely on semantic guidance, learned queries, prompting, or decoder-based segmentation. In contrast, *SymmetryBreak* studies whether a frozen spatial representation alone can support patient-specific localization by exploiting intrinsic left-right correspondence.

Unsupervised and Generative Anomaly Detection in Brain MRI. A complementary line of research formulates pathology localization as deviation from a learned model of normal anatomical appearance. Reconstruction-based approaches identify abnormalities through discrepancies between an input image and its reconstruction, while more recent methods employ masked autoencoders, patch-based modeling, or diffusion processes to obtain localized anomaly responses Schlegl et al. (2019a); Wyatt et al. (2022); Bercea et al. (2023; 2024); Rashmi et al. (2024); Behrendt et al. (2024). Ano-swinMAE uses masked autoencoding to model brain MRI appearance and derives anomaly evidence from reconstruction behavior Rashmi et al. (2024). Patch-based diffusion methods similarly model local anatomical appearance and identify deviations using generative reconstruction or restoration Behrendt et al. (2024). Related formulations include AnoDDPM and AutoDDPM, which use diffusion-based generative models for unsupervised medical anomaly detection Wyatt et al. (2022); Bercea et al. (2023). Although these approaches reduce reliance on lesion annotations, their anomaly signal is generally determined by a learned normality model, making performance dependent on the training distribution, reconstruction quality, and learned generative representation. *SymmetryBreak* instead derives normality from the current subject itself: low-discrepancy contralateral locations provide an image-internal reference, enabling anomaly attribution in a frozen latent space without learning a task-specific normal distribution.

Anatomical Symmetry for Brain Abnormality Detection. The approximate bilateral organization of the brain has long motivated symmetry-based approaches to abnormality detection. Contralateral comparison can provide a patient-specific reference and has been explored using image intensities, structural correspondence, and learned representations Toga & Thompson (2003); Martins et al. (2019); Bianchi et al. (2013); Lian & Song (2014); Wu et al. (2021). Such approaches are attractive because they avoid requiring a separate healthy-subject reference for every imaging protocol or patient. However, brain symmetry is only approximate, and normal inter-hemispheric variation can produce differences even in the absence of pathology Toga & Thompson (2003); Martins et al. (2019). More fundamentally, a conventional bilateral discrepancy is inherently non-directional: it indicates that corresponding regions differ but does not establish which hemisphere accounts for the deviation. This creates an attribution problem when the discrepancy response appears on both sides or when normal anatomical differences compete with pathology-induced changes. *SymmetryBreak* builds on the patient-specific nature of contralateral comparison while explicitly separating *symmetry breaking* from *directional attribution*. Rather than treating bilateral discrepancy as the final anomaly score, our method uses low-discrepancy correspondences within the same image to construct an internal reference, estimates side-specific latent deviations relative to that reference, and compares the two sides on a shared scale. A global hemisphere decision then converts the resulting directional evidence into a localization map. This formulation differs from conventional symmetry-based detection in that bilateral correspondence is used not only to detect abnormality, but also to determine *which hemisphere is responsible for the observed deviation*, without task-specific training.

3 METHOD

3.1 PROBLEM FORMULATION AND OVERVIEW

We formulate training-free brain tumor localization as a *directional anomaly attribution* problem. Given an unlabeled MRI brain slice $x \in \mathbb{R}^{H \times W}$, our goal is to produce a spatial localization map identifying pathological tissue without training or relying on an external healthy-brain reference. Our approach is motivated by the approximate bilateral organization of the brain. Corresponding regions of the two cerebral hemispheres tend to exhibit related anatomical and visual characteristics, whereas a focal unilateral lesion can disrupt this correspondence. This provides an intrinsic reference that is available within the subject itself. However, bilateral comparison alone is insufficient. A discrepancy between corresponding left and right regions establishes *that* the two sides differ, but

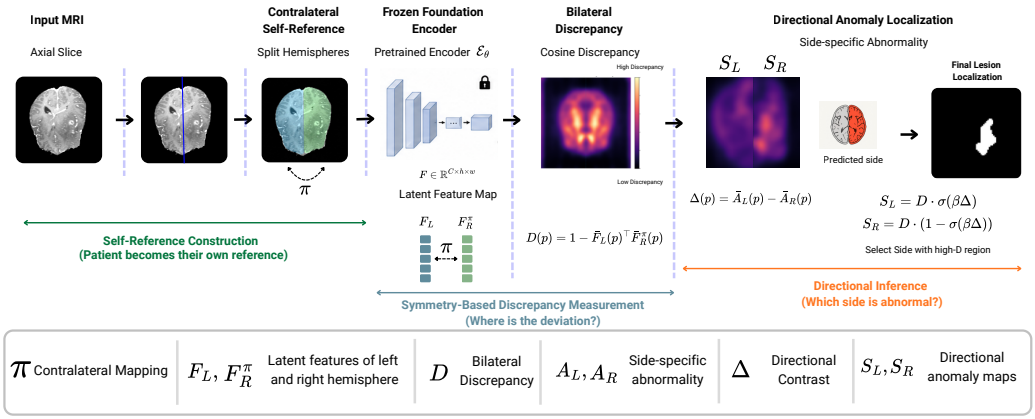


Figure 2: Overview of ***SymmetryBreak***. Given an unlabeled brain MRI slice, a frozen vision encoder produces a spatially organized latent representation. The representation is divided into corresponding left and right hemispheres, with the right hemisphere horizontally mirrored to establish deterministic contralateral correspondence. Bilateral cosine discrepancy identifies *where* the latent representation breaks symmetry. Low-discrepancy correspondences within the same image provide an image-internal reference for estimating side-specific latent deviations. The selected directional field is then returned to image space to obtain a localization map, without task-specific training and external healthy-brain atlases .

does not establish *which side* is abnormal. Thus, a conventional symmetry discrepancy is inherently non-directional and can produce mirrored responses.

SymmetryBreak explicitly separates these two questions. First, it detects spatial violations of bilateral symmetry using frozen latent representations. Second, it estimates which hemisphere exhibits stronger evidence of abnormality using a robust reference constructed from low-discrepancy regions of the same image. The resulting directional evidence is aggregated into a single hemisphere decision before constructing the final localization map.

3.2 Directional Anomaly Attribution

(1) **Frozen Spatial Latent Representation.** Let

$$F = \mathcal{E}_\theta(x), \qquad F \in \mathbb{R}^{C\times h\times w}. \tag{1}$$

F denote the spatial feature representation produced by a pretrained vision encoder $\mathcal{E}_\theta$. The encoder parameters remain frozen throughout inference.

Unlike a segmentation architecture, the encoder is not trained to predict tumor masks. The encoder serves only as a fixed transformation from image space to a spatially organized latent space.

We partition the latent feature map along its spatial width into two hemispheres $F_L, F_R \in \mathbb{R}^{C\times H\times W/2}$. assuming an even feature-map width w. The right hemisphere has the opposite left-right orientation relative to the left hemisphere. We therefore horizontally reflect the right feature map,

$$F_R^\pi = \mathit{flip}(F_R) \tag{2}$$

After this operation, corresponding spatial indices represent approximately homologous locations in the two hemispheres. Importantly, this correspondence is obtained deterministically from the spatial organization of the input.

For every corresponding location $p \in \Omega$, we therefore obtain $F_L(p), F_R(p) \in \mathbb{R}^C$. This simple construction provides the basis for all subsequent comparisons.

(2) **Bilateral Latent Symmetry Breaking.** We first identify *where* corresponding hemispheric representations disagree. For each spatial location p, we normalize the feature vectors in to $\bar{F}_L(p)$ and

$\bar{F}_R^{\pi}(p)$. The raw bilateral discrepancy is defined as:

$$D(p) = 1 - \bar{F}_L(p)^{\top} \bar{F}_R^{\pi}(p). \tag{3}$$

We refer to D as the **bilateral asymmetry map**. A large value indicates that the latent representations of corresponding locations are dissimilar. The use of bilateral discrepancy makes the bilateral comparison primarily sensitive to the direction of the latent representations rather than their absolute feature magnitude. More importantly, $D(p)$ is deliberately interpreted only as a *symmetry-breaking signal*. It does not determine the pathological side. For example, if a lesion occurs in the right hemisphere, the corresponding left-right comparison can produce a large value at the matched location even though the discrepancy is represented in a common coordinate system. We then normalize $D(p) \in [0, 1]$ with min-max values.

(3) **Patient-Specific Reference Region** A training-free localization method requires some notion of expected latent behavior. Instead of learning such a distribution from healthy subjects, *SymmetryBreak* estimates it directly from the current image. The key observation is that focal pathology typically occupies only a subset of the brain, leaving a substantial number of corresponding locations with relatively low bilateral discrepancy. We therefore define an image-internal reference set

$$\mathcal{N} = \{p \in \Omega_L : D(p) \leq Q_\alpha(D)\}, \tag{4}$$

where $Q_\alpha(D)$ denotes the α-th percentile of the bilateral discrepancy distribution. Importantly, $\mathcal{N}$ is not assumed to be a perfectly healthy set. It is only a set of *relatively low-discrepancy correspondences*. It may therefore contain normal anatomical variation, subtle pathology, or representation noise. We set α to 0.40.

For each hemisphere $s \in \{L, R\}$, the mean latent feature vector is estimated from the corresponding reference locations:

$$\mu_s(c) = \frac{1}{|\mathcal{N}|} \sum_{p \in \mathcal{N}} F_s(p, c),$$

where $c \in \{1, \ldots, C\}$ indexes the latent feature channels. The corresponding channel-wise standard deviation is

$$\sigma_s(c) = \mathrm{Std}_{p \in \mathcal{N}} \left(F_s(p, c)\right),$$

(4) **Directional Latent Abnormality.** The bilateral discrepancy D identifies asymmetric locations but does not assign responsibility to either hemisphere. We therefore estimate how strongly each side deviates from its corresponding image-internal reference.

For hemisphere $s \in L, R$, we first compute a feature-wise standardized deviation:

$$Z_s(p, c) = \frac{F_s(p, c) - \mu_s(c)}{\sigma_s(c)}. \tag{5}$$

where c indexes the latent feature dimension. We summarize the multidimensional deviation using the root-mean-square magnitude:

$$L_s(p) = \sqrt{\frac{1}{C} \sum_{c=1}^{C} Z_s(p, c)^2}. \tag{6}$$

Large $L_s(p)$ indicates that the latent representation on side s deviates strongly from the subject-specific reference. Because latent deviations can exhibit heavy-tailed responses, we apply logarithmic compression $A_s(p) = \log(1 + L_s(p))$. Here, $A_s(p)$ measures how unusual the local latent representation is relative to the patient-specific reference distribution for hemisphere s.

The two hemispheres must be compared on a common scale. Importantly, we do *not* independently normalize the left and right maps, since independent normalization could artificially remove the magnitude difference that carries the directional information. Instead, we construct a shared robust scale from the pooled distribution $\mathcal{A} = A_L(p), A_R(p) : p \in \Omega$. Let $m = median(\mathcal{A}), r = Std(\mathcal{A})$, we then define $\widetilde{A}_s(p) = \frac{A_s(p) - m}{r}$., Both hemispheres consequently share the same reference location and scale.

(5) **Directional Constrast.** The side-specific abnormality contrast is

$$\Delta(p) = \widetilde{A}_L(p) - \widetilde{A}_R(p). \tag{7}$$

A positive value indicates that the corresponding left-hemisphere location has a larger latent abnormality score, whereas a negative value indicates stronger abnormality on the right. Rather than making a hard decision independently at every pixel, we convert the signed contrast into soft directional evidence $W_L(p) = \sigma(\beta\Delta(p))$ and $W_R(p) = 1 - W_L(p)$, where $\sigma(z) = \frac{1}{1+\exp(-z)}$ is the sigmoid function and $\beta > 0$ controls the sharpness of the directional assignment. Thus, $W_L(p) + W_R(p) = 1$. The directional stage converts an unsigned symmetry violation into a soft directional assignment is responsible for the observed deviation.

The bilateral discrepancy map identifies where the two hemispheres differ, while the directional weights determine which side exhibits stronger evidence of abnormality. These two signals are combined as

$$S_L(p) = D(p)W_L(p), \qquad S_R(p) = D(p)W_R(p). \tag{8}$$

A high side-specific abnormality with weak bilateral discrepancy does not automatically generate a strong tumor response. Similarly, When directional evidence is ambiguous, the discrepancy is distributed approximately equally across the two sides; the subsequent patient-level laterality selection resolves this ambiguity. Consequently, a strong response requires simultaneous evidence for $D(p)$ as symmetry violation and $W_s(p)$ for side-specific evidence.

(6) **Global Abnormal Hemisphere Selection** Local directional evidence alone can remain ambiguous, particularly when normal anatomical variation generates several asymmetric locations. We therefore impose a global laterality constraint. We first identify the subset of strongly asymmetric locations $\mathcal{H} = \{p : D(p) \geq Q_\gamma(D)\}$, where Q_γ denotes a high percentile of the bilateral asymmetry distribution. In our experiments, $\gamma = 0.75$. We then summarize the directional evidence for each hemisphere using a high quantile $E_L = Q_\eta\left(S_L(p) : p \in \mathcal{H}\right), , E_R = Q_\eta\left(S_R(p) : p \in \mathcal{H}\right)$, where η is the hemisphere-evidence percentile. We use $\eta = 0.95$.

The inferred abnormal hemisphere is then $h^* = \arg\max_{s \in L,R} E_s.$, This is a deliberately global decision. Instead of allowing every location to independently label both hemispheres as abnormal, the method asks which hemisphere receives stronger directional evidence among the most asymmetric correspondences. The final selected directional field is

$$S(p) = \begin{cases} S_L(p), & h^* = L, \\ S_R(p), & h^* = R, \end{cases} \tag{9}$$

The selected map is then normalized to [0,1].

3.3 Training-Free tumor Localization

The anomaly field S is initially defined at the spatial resolution of the frozen encoder. We therefore upsample it to the original image resolution as S_{up}. Finally, tumor mask is obtained using a non-parametric threshold τ:

$$\hat{Y}(p) = \left[S_{\text{up}}(p) > \tau\right]. \tag{10}$$

4 Experiments and Results

4.1 Experiment Setup

Datasets. We evaluate *SymmetryBreak* on three dataset variants from the Brain Tumor Segmentation (BraTS) benchmark: BraTS-2019, BraTS-2021, and BraTS-Africa Menze et al. (2014); Baid et al. (2021); Adewole et al. (2024). These datasets provide multi-modal brain MRI scans along with expert-annotated tumor segmentations. To maintain a consistent single-modality evaluation setting, we evulate on all [FLAIR / T1ce / T1 / T2] modality as input. Ground-truth annotations are reserved exclusively for quantitative evaluation and are not utilized for model training, parameter optimization, or patient-specific adaptation. Our evaluation dataset consists of 649 scans in total,

Table 1: Quantitative evaluation on BraTS 2019 across MRI modalities. DSC (%) and AUPRC (%) are reported (↑: higher is better). ✓ denotes methods requiring training, while × denotes training-free methods. **Dark blue** indicates the best result, and **light blue** indicates the second-best result.

Method	Train	T1		T1ce		T2		FLAIR	
		DSC	AUPRC	DSC	AUPRC	DSC	AUPRC	DSC	AUPRC
Pixel Asymmetry	×	19.6	24.4	29.5	34.9	26.4	29.5	28.0	31.6
f-AnoGAN Schlegl et al. (2019b)	✓	22.6	23.5	36.0	42.1	39.4	45.1	49.4	53.8
DenoisingAE Kascenas et al. (2022)	✓	43.7	45.4	48.3	56.4	**64.2**	**75.6**	55.3	66.4
AnoDDPM Wyatt et al. (2022)	✓	**60.3**	**66.3**	63.4	69.9	58.3	65.4	64.5	73.0
AutoDDPM Bercea et al. (2023)	✓	54.3	53.2	**66.3**	**74.5**	63.4	69.3	**78.3**	**87.8**
SymmetryBreak	×	**67.9**	**78.9**	**68.1**	**81.5**	**73.6**	**87.7**	**77.8**	**87.0**

Table 2: Quantitative evaluation on BraTS 2021 across MRI modalities. DSC (%) and AUPRC (%) are reported (↑: higher is better). ✓ denotes methods requiring training, while × denotes training-free methods.**Dark blue** indicates the best result, and **light blue** indicates the second-best result.

Method	Train	T1		T1ce		T2		FLAIR	
		DSC	AUPRC	DSC	AUPRC	DSC	AUPRC	DSC	AUPRC
Pixel Asymmetry	×	7.6	16.5	10.0	20.8	24.8	26.5	28.6	31.1
f-AnoGAN Schlegl et al. (2019b)	✓	28.8	34.1	32.8	38.6	37.2	42.7	40.3	44.9
DenoisingAE Kascenas et al. (2022)	✓	39.4	47.5	46.3	59.3	61.0	69.3	71.3	78.6
AnoDDPM Wyatt et al. (2022)	✓	61.8	73.4	50.2	54.5	60.3	69.5	65.3	73.2
AutoDDPM Bercea et al. (2023)	✓	**67.4**	**79.3**	**64.3**	**72.4**	**68.2**	**83.4**	**74.3**	**85.3**
SymmetryBreak	×	**69.6**	**81.3**	**70.8**	**83.0**	**67.5**	**84.2**	**77.3**	**87.0**

randomly selected across three sources: BraTS-2019 (335 of 460), BraTS-2021 (219 of 1,251), and BraTS-Africa (95 scans).

Preprocessing. MRI volumes are processed slice-wise in the axial plane. Each slice is resized to 1024×1024 resolution to match the input requirements of the frozen feature extractor and normalized to zero mean and unit variance. High-density feature maps are extracted without task-specific fine-tuning.

Baselines. We compare *SymmetryBreak* with representative baseline methods spanning pixel-level statistics, reconstruction, adversarial, and diffusion-based anomaly detection approaches. Specifically, we benchmark against non-parametric structural baselines including Pixel Asymmetry, standard reconstruction models such as Denoising Autoencoders (DenoisingAE) Kascenas et al. (2022), and the adversarial pipeline f-AnoGAN Schlegl et al. (2019b). Additionally, we evaluate against diffusion-based frameworks, including AnoDDPM Wyatt et al. (2022), AutoDDPM Bercea et al. (2023).

Evaluation Metrics. We report the Dice similarity coefficient (DSC) and the area under the precision–recall curve (AUPRC) to evaluate tumor localization. DSC quantifies the spatial overlap between the predicted and ground-truth tumor masks, while AUPRC evaluates threshold-independent discrimination between tumor and non-tumor regions, which is particularly informative given the class imbalance between tumor and background. Higher values indicate better performance for both metrics.

Implementation Details. *SymmetryBreak* operates completely training-free, utilizing a frozen pre-trained MedSAM-ViT-B Ma et al. (2024) encoder as the backbone feature extractor, requiring no model training. All experiments are implemented in PyTorch and executed on an NVIDIA A100 GPU. The resulting anomaly map is min–max normalized to $[0, 1]$, smoothed with a Gaussian filter, and thresholded at $\tau = 0.5$ to yield the final binary tumor-mask. We select β to 4.

Table 3: Quantitative evaluation on BraTS-Africa across MRI modalities. DSC (%) and AUPRC (%) are reported (↑: higher is better). ✓ denotes methods requiring training, while × denotes training-free methods. **Dark blue** indicates the best result, and **light blue** indicates the second-best result.

Method	Train	T1		T1ce		T2		FLAIR	
		DSC	AUPRC	DSC	AUPRC	DSC	AUPRC	DSC	AUPRC
Pixel Asymmetry	×	11.2	16.7	21.2	23.5	18.7	20.1	14.6	27.5
f-AnoGAN Schlegl et al. (2019b)	✓	22.8	28.4	29.1	32.0	27.3	31.5	26.8	34.2
DenoisingAE Kascenas et al. (2022)	✓	35.6	42.1	38.4	43.8	39.0	44.2	41.5	48.7
AnoDDPM Wyatt et al. (2022)	✓	48.9	56.0	50.2	57.1	51.4	58.9	54.8	61.3
AutoDDPM Bercea et al. (2023)	✓	53.1	61.8	**56.8**	**66.4**	56.0	64.8	**63.5**	**70.1**
SymmetryBreak	×	**60.5**	**69.6**	55.3	65.2	**57.6**	**66.6**	62.4	69.2

4.2 Results

Quantitative Results. We evaluate our proposed training-free method, *SymmetryBreak*, against standard training-free baselines and supervised/generative models—including f-AnoGAN Schlegl et al. (2019b), DenoisingAE Kascenas et al. (2022), AnoDDPM Wyatt et al. (2022), and AutoDDPM Bercea et al. (2023)—across four MRI modalities (T1, T1ce, T2, FLAIR) on the BraTS 2019, BraTS 2021, and BraTS-Africa benchmarks (Tables 1–3). On BraTS 2019, *SymmetryBreak* establishes performance across T1 (67.9% DSC / 78.9% AUPRC), T1ce (68.1% DSC / 81.5% AUPRC), and T2 (73.6% DSC / 87.8% AUPRC), while achieving competitive second-best scores on FLAIR (77.8% DSC / 87.0% AUPRC) behind AutoDDPM (78.3% DSC / 87.8% AUPRC). This superior performance carries over to the BraTS 2021 benchmark, where our method secures top results in T1 (69.6% DSC), T1ce (70.8% DSC), and FLAIR (77.3% DSC), as well as the highest AUPRC on T2 (84.2%). Furthermore, on the challenging out-of-distribution BraTS-Africa dataset, *SymmetryBreak*demonstrates robust domain generalization by leading on T1 (60.5% DSC / 69.6% AUPRC) and T2 (57.6% DSC / 66.6% AUPRC), while maintaining close second-best performance on T1ce and FLAIR. Overall, while simple training-free methods like Pixel Asymmetry struggle ($< 30\%$ DSC) and deep generative pipelines suffer performance drops due to domain shifts or heavy training dependencies, *SymmetryBreak* consistently delivers top-tier accuracy across all datasets and modalities entirely without any training overhead.

Qualitative Results. Figure 3 presents a visual comparison of anomaly segmentation results generated by our proposed method, **Ours(SB)**, alongside baseline models against the ground truth (GT) across representative slices from the BraTS-Africa (FLAIR), BraTS-2019 (T2), and BraTS-2021 (T1ce) datasets. As observed, baseline methods frequently suffer from either under-segmentation or over-segmentation; for instance, f-AnoGAN and DenoisingAE consistently fail to capture the full extent of larger lesions (Row 1) or miss finer structural boundaries (Row 2), whereas diffusion-based methods like AnoDDPM and AutoDDPM occasionally produce imprecise contours near complex tissue margins.

4.3 Ablation Studies

Effect of each component. To evaluate the individual contributions of each architectural module in *SymmetryBreak*, we conduct systematic ablation experiments across three distinct MRI modalities and datasets (Table 4). Removing any single component yields a consistent performance drop, confirming their synergistic role in asymmetry-driven pathology localization. The *Bilateral Discrepancy* module proves most critical, as its removal induces a catastrophic failure mode, dropping DSC by up to 50.4% on BraTS'19 (73.6% $\rightarrow$ 23.2%) due to the loss of cross-hemispheric feature comparison. Stripping *Directional Attribution* similarly severely degrades performance, causing AUPRC to drop below 45.0% across all modalities by failing to disambiguate lesion presence from healthy brain structures. Omitting *Global Hemisphere Selection* causes substantial drops (51.7% DSC on BraTS'19), underscoring the necessity of establishing macro-level contextual references before fine-grained comparison. Finally, excluding the *Patient-Specific Reference* module yields consistent, notable drops—most prominently on the challenging BraTS-Africa cohort (60.5% $\rightarrow$ 54.8%

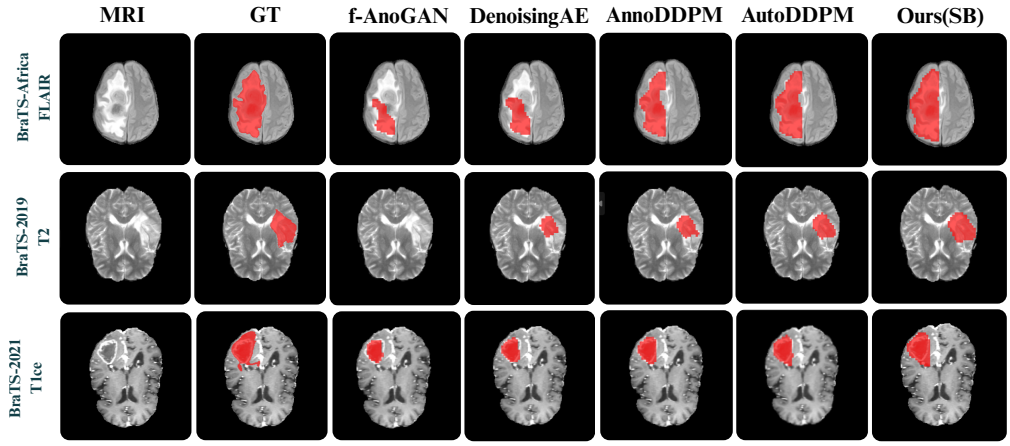


Figure 3: Qualitative evaluation and visual comparison of *SymmetryBreak* across three datasets: BraTS-Africa FLAIR (top row), BraTS-2019 T2 (middle row), and BraTS-2021 T1ce (bottom row). Columns display the input MRI, Ground Truth (GT) segmentation mask overlaid in red, and predictions from baseline methods.

Table 4: **Ablation study of *SymmetryBreak* architectural components.** Evaluation across MRI datasets demonstrates the contribution of each module.

Method / Variant	BraTS,19 (T2)		BraTS,21 (FLAIR)		BraTS-Africa (T1)	
	DSC	AUPRC	DSC	AUPRC	DSC	AUPRC
Full SymmetryBreak	**73.6**	**87.7**	**77.3**	**87.0**	**60.5**	**69.6**
w/o Directional Attribution	34.8	42.7	36.2	44.3	35.5	38.8
w/o Patient-Specific Ref.	69.8	85.2	75.4	84.2	54.8	66.0
w/o Global Hemisphere Sel.	51.7	46.0	53.4	46.1	45.6	41.6
w/o Bilateral Discrepancy	23.2	33.6	36.0	48.5	17.0	20.2

Table 5: Effect of different Frozen Backbone Representations on BraTS-2021 FLAIR.

Model	DSC (%)	AUPRC (%)
MedSAM-ViT-B	77.3	87.0
MedSAM-ViT-L	78.3	88.2
MedSAM-ViT-H	78.9	87.5
RAD-DINO-ViT	70.3	81.5
BiomedCLIP	69.4	76.7

DSC)—demonstrating its essential role in mitigating cross-site scanner heterogeneity. Overall, the integrated *SymmetryBreak* framework achieves peak performance across all evaluations (77.3% DSC on BraTS'21 FLAIR), proving every module is indispensable.

Effect of the Frozen Representation To evaluate the impact of the frozen backbone features on BraTS-2021 FLAIR segmentation (Table 5), we compared five model representations. The MedSAM series consistently outperforms broader vision-language models, with MedSAM-ViT-H achieving the highest Dice Similarity Coefficient (78.9% DSC) and MedSAM-ViT-L securing the top precision-recall score (88.2% AUPRC). Within the MedSAM family, scaling model capacity steadily improves results, with ViT-L gaining 1.0% DSC over ViT-B. Conversely, non-segmentation pretrained backbones like RAD-DINO-ViT (70.3% DSC) Pérez-García et al. (2025) and BiomedCLIP (69.4% DSC) Zhang et al. (2023) perform significantly worse, dropping up to 9.5% DSC compared to MedSAM-ViT-H. This demonstrates that domain-specific, segmentation-oriented features are crucial when using frozen representations.

5 Conclusion

We presented *SymmetryBreak*, a training-free framework for brain MRI tumor localization using bilateral latent symmetry and directional anomaly attribution. By combining a frozen foundation encoder, patient-specific robust references, and hemisphere-level aggregation, our method identifies both the location and responsible side of asymmetric pathology without task-specific training or external normal atlases. Although the approach is primarily suited to unilateral abnormalities that emerge at early stages, and depends on reliable contralateral correspondence. It provides a simple, interpretable foundation for training-free and patient-adaptive neuroimaging anomaly localization.

AI USE STATEMENT

In this work, we used generative AI tools for polishing text for readability, correcting grammatical errors, and refining style. We have not used generative AI tools for generating research ideas, initial drafts of core methodological claims, or creating experimental results, and code generation and data creation. Additionally, we used generative AI tools for rephrasing complex paragraphs to improve clarity and flow. We have reviewed all AI-assisted work. All LLM-suggested edits were manually verified by the authors to ensure mathematical precision, domain-specific terminological accuracy, and fidelity to our original findings. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. Our research does not involve human subjects, personal data collection, or sensitive private information, ensuring no direct privacy or security risks. While machine learning models present general potential for societal misuse, our methodological contributions pose no foreseeable harm, discrimination, or negative ethical impacts. All experiments rely on publicly available, standard benchmark datasets in full compliance with dataset licenses and legal standards.

REPRODUCIBILITY STATEMENT

All experiments use publicly available brain MRI datasets and a fixed, pretrained foundation model. We report the datasets, preprocessing steps, model configuration, hyperparameters, and evaluation metrics used in our experiments. The proposed method requires no task-specific training, fine-tuning, or learned parameters. Source code will be released upon acceptance.

REFERENCES

M Adewole, JD Rudie, A Gbadamosi, et al. Expanding the brain tumor segmentation (brats) data to include african populations (brats-africa)(version 1)[dataset]. *The Cancer Imaging Archive*, 2024.

Sidra Aleem, Fangyijie Wang, Mayug Maniparambil, Eric Arazo, Julia Dietlmeier, Kathleen Curran, Noel E. O'Connor, and Suzanne Little. Test-time adaptation with salip: A cascade of sam and clip for zero-shot medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 5184–5193, 2024.

Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021.

Finn Behrendt, Debayan Bhattacharya, Julia Krueger, Roland Opfer, and Alexander Schlaefer. Patched diffusion models for unsupervised anomaly detection in brain mri. In *Medical Imaging with Deep Learning*, volume 227 of *Proceedings of Machine Learning Research*, pp. 1019–1032, 2024.

Cosmin I Bercea, Michael Neumayr, Daniel Rueckert, and Julia A Schnabel. Mask, stitch, and re-sample: Enhancing robustness and generalizability in anomaly detection through automatic diffusion models. *arXiv preprint arXiv:2305.19643*, 2023.

Cosmin I. Bercea, Benedikt Wiestler, Daniel Rueckert, and Julia A. Schnabel. Generalizing unsupervised anomaly detection: Towards unbiased pathology screening. In *Medical Imaging with Deep Learning*, volume 227 of *Proceedings of Machine Learning Research*, pp. 39–52, 2024.

BIA Group. IXI Dataset. `https://brain-development.org/ixi-dataset/`, 2006. Accessed: 2026-09-23.

Anthony Bianchi, James V. Miller, Ek Tsoon Tan, and Albert Montillo. Brain tumor segmentation with symmetric texture and symmetric intensity-based decision forests. In *2013*

IEEE 10th International Symposium on Biomedical Imaging, pp. 748–751. IEEE, 2013. doi: 10.1109/ISBI.2013.6556583.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021.

Jenny Crinion, John Ashburner, Alex Leff, Matthew Brett, Cathy Price, and Karl Friston. Spatial normalization of lesioned brains: performance evaluation and impact on fmri analyses. *Neuroimage*, 37(3):866–875, 2007.

J. Deverdun et al. Challenges and opportunities for advanced neuroimaging of glioblastoma. *British Journal of Radiology*, 96(1141):20211232, 2023. doi: 10.1259/bjr.20211232.

Anahita Fathi Kazerooni, Sherjeel Arif, Rachel Madhogarhia, Nastaran Khalili, Debanjan Haldar, Sina Bagheri, Ariana M. Familiar, Hannah Anderson, Shuvanjan Haldar, Wenxin Tu, Meen Chul Kim, Karthik Viswanathan, Sabine Muller, Michael Prados, Cassie Kline, Lorenna Vidal, Mariam Aboian, Phillip B. Storm, Adam C. Resnick, Jeffrey B. Ware, Arastoo Vossough, Christos Davatzikos, and Ali Nabavizadeh. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. *Neuro-Oncology Advances*, 5(1):vdad027, 2023. doi: 10.1093/noajnl/vdad027.

Salah Ghamizi, Georgia Kanli, Yu Deng, Valérie Palissot, Magali Perquin, and Olivier Keunen. Foundation models for brain imaging: A systematic review. *NeuroImage*, 333:121877, 2026. doi: 10.1016/j.neuroimage.2026.121877.

Yankai Jiang, Zhongzhen Huang, Rongzhao Zhang, Xiaofan Zhang, and Shaoting Zhang. Zept: Zero-shot pan-tumor segmentation via query-disentangling and self-prompting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11386–11397, 2024.

Yankai Jiang, Wenhui Lei, Xiaofan Zhang, and Shaoting Zhang. Unleashing the potential of vision-language pre-training for 3d zero-shot lesion segmentation via mask-attribute alignment. In *International Conference on Learning Representations*, 2025.

Antanas Kascenas, Nicolas Pugeault, and Alison Q. O'Neil. Denoising autoencoders for unsupervised anomaly detection in brain mri. In *Proceedings of the 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research*, pp. 653–664, 2022.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *2023 IEEE/CVF international conference on computer vision (ICCV)*, pp. 3992–4003. IEEE, 2023.

Yanyun Lian and Zhijian Song. Automated brain tumor segmentation in magnetic resonance imaging based on sliding-window technique and symmetry analysis. *Chinese Medical Journal*, 127 (3):462–468, 2014. doi: 10.3760/cma.j.issn.0366-6999.20132554.

Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15:654, 2024.

Samuel Botter Martins, Barbara Caroline Benato, Bruna Ferreira Silva, Clarissa Lin Yasuda, and Alexandre Xavier Falcão. Modeling normal brain asymmetry in mr images applied to anomaly detection without segmentation and data annotation. In *Medical Imaging 2019: Computer-Aided Diagnosis*, volume 10950 of *Proceedings of SPIE*. SPIE, 2019. doi: 10.1117/12.2512873.

Matia Martucci, Rosellina Russo, Francesco Schimperna, Gabriella D'Apolito, Marco Panfili, Alessandro Grimaldi, Alessandro Perna, Andrea Maurizio Ferranti, Giuseppe Varcasia, Carolina Giordano, and Simona Gaudino. Magnetic resonance imaging of primary adult brain tumors: State of the art and future perspectives. *Biomedicines*, 11(2):364, 2023. doi: 10.3390/biomedicines11020364.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34 (10):1993–2024, 2014.

Mario Mustra, Kresimir Delac, and Mislav Grgic. Overview of the dicom standard. In *2008 50th international symposium ELMAR*, volume 1, pp. 39–44. IEEE, 2008.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Fernando Pérez-García, Harshita Sharma, Sam Bond-Taylor, Kenza Bouzid, Valentina Salvatelli, Maximilian Ilse, Shruthi Bannur, Daniel C Castro, Anton Schwaighofer, Matthew P Lungren, et al. Exploring scalable medical image encoders beyond text supervision. *Nature Machine Intelligence*, 7(1):119–130, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Kumari Rashmi, Ayantika Das, NagaGayathri Matcha, Keerthi Ram, and Mohanasankar Sivaprakasam. Ano-swinmae: Unsupervised anomaly detection in brain mri using swin transformer based masked auto encoder. In *Proceedings of the 7th International Conference on Medical Imaging with Deep Learning*, volume 250 of *Proceedings of Machine Learning Research*, pp. 1221–1236, 2024.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. In *International Conference on Learning Representations*, volume 2025, pp. 28085–28128, 2025.

Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019a.

Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019b. doi: 10.1016/j.media.2019.01.010.

Arthur W. Toga and Paul M. Thompson. Mapping brain asymmetry. *Nature Reviews Neuroscience*, 4(1):37–48, 2003. doi: 10.1038/nrn1009.

Javier E. Villanueva-Meyer, Marc C. Mabray, and Soonmee Cha. Current clinical brain tumor imaging. *Neurosurgery*, 81(3):397–415, 2017. doi: 10.1093/neuros/nyx103.

Junde Wu and Min Xu. One-prompt to segment all medical images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11302–11312, 2024.

Xinheng Wu, Lei Bi, Michael Fulham, David Dagan Feng, Luping Zhou, and Jinman Kim. Unsupervised brain tumor segmentation using a symmetric-driven adversarial network. *Neurocomputing*, 455:242–254, 2021. doi: 10.1016/j.neucom.2021.05.073.

Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 649–655. IEEE, 2022.

Shaoting Zhang and Dimitris Metaxas. On the challenges and perspectives of foundation models for medical image analysis. *Medical Image Analysis*, 91:102996, 2024. doi: 10.1016/j.media.2023.102996.

Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. *arXiv preprint arXiv:2303.00915*, 2023.

A EFFECT OF HYPERPARAMETER

To evaluate the robustness and sensitivity of the proposed framework on the BraTS2021 dataset ($T1ce$ modality), we systematically varied four key hyperparameters: α (Normal percentile), β (for sharpness of the directional assignment), γ (directional contrast control), and η (hemisphere-selection percentile). The evaluation metrics considered include (DSC), and (AUPRC) respectively. Baseline performance was evaluated around the anchor setting ($\alpha = 0.4, \beta = 4.0, \gamma = 0.75, \eta = 0.95$).

A.1 SENSITIVITY TO α

To evaluate the sensitivity of our proposed method to the hyperparameter α, which represents the percentile threshold scaled within the range of $[0, 1]$, we perform experiments across varying values of $\alpha \in [0.2, 0.7]$. As illustrated in Figure 4, both the (DSC) and (AUPRC) show consistent performance improvements as α increases from 0.2, reaching optimal performance in the range of $\alpha = 0.4$ to 0.5. Lower values of α ($\alpha < 0.3$) tend to incorporate excess background noise or uninformative features, which impairs overall localization precision. Conversely, setting α excessively high ($\alpha > 0.6$) overly restricts feature selection, leading to a mild drop in performance due to the potential loss of critical target boundary information. Based on these empirical observations, we select $\alpha = 0.4$ as our optimal default hyperparameter, as it delivers a robust balance between effectively filtering out irrelevant features and preserving essential localization details.

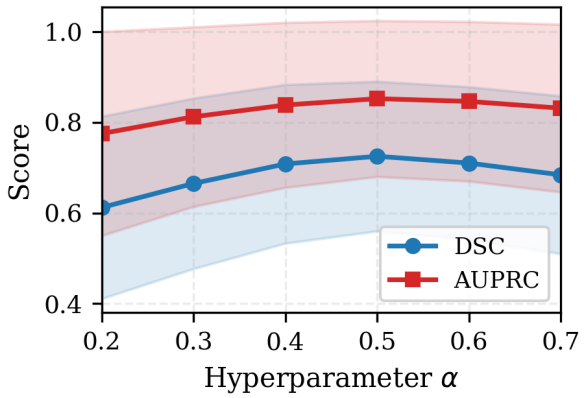


Figure 4: Sensitivity analysis with respect to hyperparameter α on the BraTS 2021 $T1ce$ dataset.

A.2 SENSITIVITY TO β

To analyze the sensitivity of our framework to the sharpness parameter β, which controls the sharpness of the directional assignment, we evaluate performance across values ranging from $\beta \in [2, 6]$ while keeping other hyperparameters fixed at their baseline settings. As depicted in Figure 5, both DSC and AUPRC show a steady performance increase as β increases from 2 to 4, peaking clearly at $\beta = 4.0$. When β is set too low ($\beta < 3$), the directional assignment remains overly smooth, failing to sufficiently differentiate sharp directional boundaries and resulting in suboptimal localization scores. On the other hand, increasing β beyond 4.0 ($\beta > 5$) leads to a gradual decline in performance, as an excessively sharp assignment function makes the feature representations overly sensitive to local noise and minor orientation fluctuations. Therefore, we select $\beta = 4.0$ as the optimal setting, ensuring sharp and reliable directional assignment without introducing instability.

A.3 SENSITIVITY TO γ

To examine the sensitivity of the proposed framework to the hyperparameter γ, which governs the directional contrast control, we evaluate performance across a range of values $\gamma \in [0.4, 0.9]$. As shown in Figure 6, both the DSC and AUPRC show a steady improvement as γ increases from 0.4, reaching optimal and stable performance in the range of $\gamma = 0.70$ to 0.75. Lower values ($\gamma <$

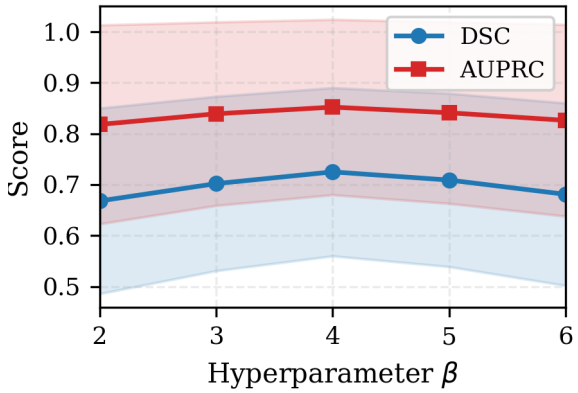


Figure 5: Sensitivity analysis with respect to hyperparameter β on the BraTS 2021 $T1ce$ dataset.

0.60) result in reduced segmentation accuracy due to insufficient directional contrast enhancement. Conversely, setting γ excessively high ($\gamma > 0.80$) leads to a slight performance degradation, as over-amplifying the contrast can exaggerate subtle feature variations and introduce noise. Based on these observations, we set $\gamma = 0.75$ as our default value to achieve an effective balance between directional contrast separation and noise suppression.

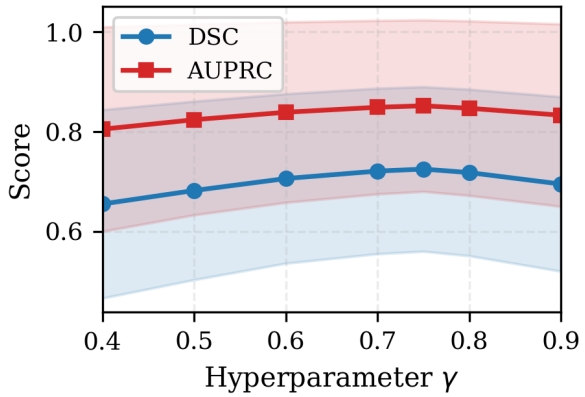


Figure 6: Sensitivity analysis with respect to hyperparameter γ on the BraTS 2021 $T1ce$ dataset.

A.4 Sensitivity to η

To assess the sensitivity of the framework to the hemisphere-selection percentile η, we evaluate model performance across values of $\eta \in [0.6, 1.0]$. As illustrated in Figure 7, both the DSC and AUPRC show a clear and steady upward trend as η increases from 0.60, reaching optimal performance around $\eta = 0.95$. Lower percentile thresholds ($\eta < 0.80$) lead to poorer localization results because a broader feature selection allows non-discriminative and noisy hemispheric features into the representation. Conversely, increasing η beyond 0.95 ($\eta = 1.0$) causes a minor drop in performance, as an overly restrictive threshold risks filtering out essential contextual cues near the selection boundaries. Consequently, we select $\eta = 0.95$ as the default hyperparameter, as it effectively isolates the most critical hemispheric regions while maintaining sufficient spatial context.

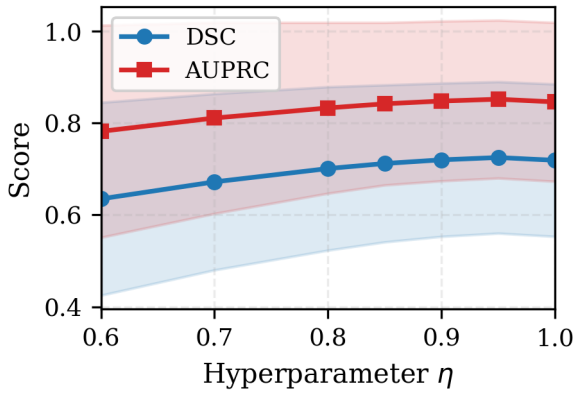


Figure 7: Sensitivity analysis with respect to hyperparameter η on the BraTS 2021 $T1ce$ dataset.

Table 6: Hemisphere-selection failure analysis evaluated across datasets for selected modalities. Evaluates DSC and AUPRC metrics when selecting the correct versus wrong hemisphere relative to the tumor side.

Dataset	**Modality**	**Correct Side**		**Wrong Side**	
		DSC ↑	**AUPRC** ↑	**DSC** ↑	**AUPRC** ↑
BraTS 2019	T2	74.0	87.6	2.9	5.2
BraTS 2021	FLAIR	79.8	88.6	3.6	6.0
BraTS-Africa	T1ce	58.9	68.5	7.2	7.9

B Failure Case Analysis

B.1 Impact of Hemisphere Selection

To evaluate the sensitivity of our method pipeline to initial localization, we conducted a failure analysis assessing performance under intentional hemisphere-selection errors across three datasets (Table 6).

When the correct tumor-bearing hemisphere is targeted, the model demonstrates robust segmentation performance, achieving up to 79.8% DSC and 88.6% AUPRC on BraTS 2021 (FLAIR). However, targeting the wrong hemisphere leads to a catastrophic collapse in performance: mean DSC drops from 70.9% to 4.6%, and mean AUPRC declines from 81.6% to 6.4%.

This severe degradation reveals two key insights into model behavior:

1. **High Precision Against False Positives:** The near-zero performance on the contralateral side indicates that the model heavily relies on localized feature cues rather than general intensity thresholds, successfully avoiding false-positive on healthy, symmetric brain tissue.
2. **Single Point of Failure:** Accurate upstream hemisphere routing is a strictly non-negotiable prerequisite; an uncorrected localization error in the initial stage completely disables effective downstream segmentation.

Notably, lower baseline scores on BraTS-Africa (58.9% DSC on the correct side) point to additional domain-specific challenges in contrast-enhanced T1 images, though the trend of total failure on the wrong side remains identical across all modalities and cohorts.

B.2 Qualitative Error Analysis.

A qualitative evaluation of challenging cases reveals distinct failure modes in our method (**SB**) when compared against the Ground Truth (**GT**) masks, as depicted in Figure 8. In cases involving asymmetrical or diffuse lesions (*Partial shared*), the **SB** exhibits significant over-localization, expanding beyond the anatomical tumor boundaries into non-tumoral brain parenchyma due to confounding

hyperintensities in the underlying structural MRI. In scenarios featuring mid-line or deeply seated pathology (*Centered Tumor*), the structural baseline misinterprets anatomical symmetry, resulting in whole-hemisphere false positives and severe lateral over-estimation. Furthermore, when dealing with focal or tiny pathological regions (*Small Tumor*), the proposed method fails to capture fine spatial details, producing massive half-brain false-positive regions instead of localized tumor. The corresponding heatmaps in the *Discrepancy Map* column highlight these spatial uncertainties and feature mismatches, demonstrating our method struggles on Bilateral tumor presence.

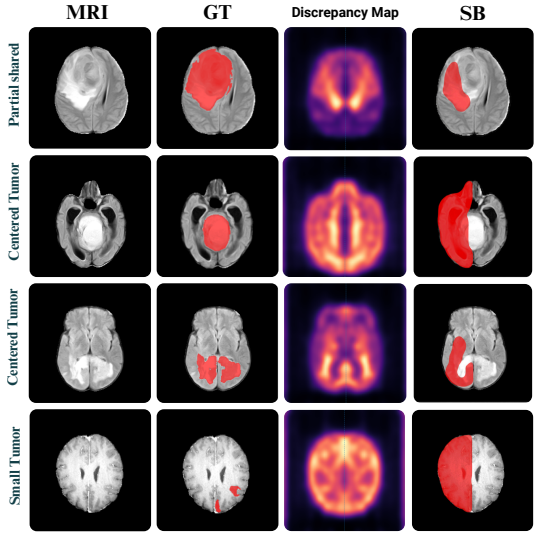


Figure 8: **Qualitative comparison and failure case examples across various tumor topologies.** Columns show the input brain **MRI**, Ground Truth (**GT**) annotations, the calculated **Discrepancy Map**, and Our proposed method (**SB**) model output. Rows illustrate difficult edge-case scenarios including partial spatial sharing (*Partial shared*), central ventricular brain tumors (*Centered Tumor*), and focal subtle tumor (*Small Tumor*).

C Algorithmic Summary

Algorithm 1 SymmetryBreak: Training-Free MRI Anomaly Localization

Require: Brain MRI image x, frozen encoder $\mathcal{E}_\theta$
Ensure: Anomaly mask $\hat{Y}$

1: Extract latent features: $F \leftarrow \mathcal{E}_\theta(x)$
2: Split F into left and right hemispheres: (F_L, F_R)
3: Flip the right hemisphere: $F_R^\pi \leftarrow \mathrm{Flip}(F_R)$ and normalized.
4: Compute bilateral discrepancy:

$$D(p) \leftarrow 1 - \cos(\hat{F}_L(p), \hat{F}_R(p))$$

5: Select low-discrepancy locations as the reference set
6: Compute latent abnormality maps $\hat{A}_L$ and $\hat{A}_R$
7: Compute directional contrast:

$$\Delta(p) \leftarrow \hat{(A)}_L(p) - \hat{(A)}_R(p)$$

8: Compute directional weights:

$$W_L(p) \leftarrow \sigma(\beta\Delta(p)), \qquad W_R(p) \leftarrow 1 - W_L(p)$$

9: Compute directional anomaly maps:

$$S_L \leftarrow D \odot W_L, \qquad S_R \leftarrow D \odot W_R$$

10: Estimate evidence for each hemisphere
11: Select the hemisphere with larger evidence
12: Keep only the selected anomaly map
13: Upsample and threshold the anomaly map
14: **return** $\hat{Y}$

D Limitations and Scope

SymmetryBreak is designed for brain MRI abnormalities that produce detectable left–right asymmetry and therefore may be less effective for tumors that are highly symmetric, diffuse, or located near the cerebral midline. Its performance also depends on sufficient anatomical correspondence between hemispheres; substantial motion, severe geometric distortion, incomplete coverage, or residual misalignment can introduce false asymmetries. In addition, the method operates slice-wise and uses a frozen foundation encoder, so its localization quality may depend on the representation quality of the selected backbone. The proposed framework should therefore be viewed as a training-free localization approach rather than a replacement for fully supervised clinical segmentation systems. Future work will investigate robustness across acquisition protocols, anatomical regions, stronger registration perturbations, and additional foundation-model backbones.

E Discussion on Assuming Alignment.

SymmetryBreak assumes approximate left–right correspondence between the two cerebral hemispheres, rather than perfect anatomical alignment or registration to an external atlas. Clinical MRI data are typically stored with patient-relative spatial orientation information, including image position and orientation, providing a consistent anatomical coordinate system Mustra et al. (2008); nevertheless, differences in head positioning, rotation, translation, motion, and acquisition geometry can introduce residual misalignment. Importantly, our method does not require nonlinear inter-subject atlas registration: it only requires sufficient within-subject correspondence for comparing approximately homologous locations across hemispheres. Such global pose differences can generally be addressed using lightweight rigid alignment, a standard preprocessing operation in brain MRI analysis Crinion et al. (2007). Thus, our assumption is not that clinical MRI scans are perfectly aligned at acquisition, but that approximate bilateral correspondence can be established without an external

healthy-brain atlas or task-specific learned registration. Imperfect alignment remains a limitation, since residual geometric mismatch may introduce spurious bilateral discrepancies and affect directional attribution, particularly in scans with substantial head rotation, motion artifacts, or incomplete brain coverage. Future work can further quantify sensitivity to controlled registration errors and incorporate automated mid-sagittal-plane estimation or robust registration when acquisition variability is substantial.