digital

MDPI

Article

CMYD-SurfaceNet: Scale-Aware Cascaded Multimodal MRI Segmentation via Representation-Level Structural Decoupling and Boundary-Constrained Learning

Chaymae El Mechal [1,*], Mostefa Mesbah [2,*], Loubna Mazgouti [1], Fatima Zahra Ammor [3] and Najiba El Amrani El Idrissi [1]

[1] Signal, Systems and Components Laboratory, Faculty of Sciences & Technologies, Sidi Mohamed Ben Abdellah University, Fez 30000, Morocco; loubna.mazgouti@usmba.ac.ma (L.M.); najiba.elamrani@usmba.ac.ma (N.E.A.E.I.)
[2] Department of Electrical & Computer Engineering, Sultan Qaboos University, Muscat 123, Oman
[3] Faculty of Medicine, University IBN ZOHR, Agadir 80000, Morocco; dr.a.ammor@gmail.com
* Correspondence: chaymae.elmechal@usmba.ac.ma (C.E.M.); m.mesbah@squ.edu.om (M.M.)

Abstract

Reliable delineation of brain tumor boundaries in multimodal magnetic resonance imaging (MRI) remains challenging despite substantial advances in deep learning–based segmentation. Although modern encoder–decoder architectures achieve strong volumetric overlap, precise geometric alignment of tumor contours remains inconsistent, particularly for small lesions and heterogeneous clinical cases. In neuro-oncology, even minor boundary deviations may influence surgical planning, radiotherapy targeting, and longitudinal treatment assessment. These limitations suggest that segmentation performance is not determined solely by network depth or loss design, but also by how multimodal information is structured prior to learning. We introduce CMYD-SurfaceNet, a scale-aware cascaded framework that restructures multimodal MRI inputs at the representation level to enhance boundary-sensitive segmentation. Rather than treating modalities as independently concatenated channels, selected sequences are first organized into a task-guided pseudo-RGB projection. This intermediate representation is subsequently transformed into the CMYK color space to disentangle shared luminance structure from modality-specific contrast dominance. To further encode geometric priors, a gradient-derived boundary density channel is incorporated to explicitly emphasize spatial discontinuities corresponding to tumor margins. The resulting CMYD representation is integrated within a two-stage nnU-Net cascade, where global tumor localization is followed by high-resolution region-of-interest refinement with auxiliary contour supervision. This scale-aware design improves sensitivity to small tumor components while stabilizing contour delineation. Extensive evaluation on the BraTS benchmark demonstrates consistent improvements in boundary-sensitive metrics. Compared with baseline nnU-Net, the proposed framework reduces HD95 from 3.6 mm to 2.4 mm and increases Surface Dice at 1 mm tolerance from 0.82 to 0.89, while maintaining competitive Dice performance. These findings suggest that representation-level structural decoupling, when combined with scale-aware refinement, may provide clinically relevant boundary-aware multimodal MRI segmentation support without increasing architectural complexity.

Keywords: brain tumor segmentation; multimodal MRI; nnU-Net; boundary-aware segmentation; cascaded segmentation; representation-level learning; structural disentanglement

Check for updates

Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 28 April 2026
Revised: 10 June 2026
Accepted: 11 June 2026
Published: 16 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Brain tumor segmentation from multimodal magnetic resonance imaging MRI is a central task in neuro-oncologic image analysis. Accurate delineation of tumor subregions such as the whole tumor (WT), tumor core (TC), and enhancing tumor (ET) is essential for surgical planning, radiotherapy contouring, and longitudinal treatment monitoring [1]. Although deep learning methods have significantly improved automated segmentation performance, achieving clinically reliable boundary precision remains challenging. Encoder–decoder architectures derived from U-Net have become the dominant paradigm in medical image segmentation because of their ability to integrate multiscale contextual information with spatial localization [2]. Among these frameworks, nnU-Net has demonstrated strong robustness across diverse biomedical datasets by automatically adapting preprocessing, architectural configuration, and training strategies to dataset characteristics [3]. Despite these advances, improvements in volumetric metrics such as Dice do not consistently translate into equivalent improvements in geometric boundary accuracy. This discrepancy is partly related to the widespread reliance on region-based optimization objectives. Metrics such as the Dice similarity coefficient emphasize global volumetric overlap but remain relatively insensitive to localized contour deviations [4]. Consequently, a segmentation model may achieve high Dice scores while still exhibiting boundary irregularities or subtle geometric misalignment. In clinical contexts, even small surface deviations can influence radiation dose planning and surgical margin assessment [1]. To improve boundary sensitivity, previous studies have introduced surface-based loss formulations and Hausdorff-inspired optimization strategies that explicitly penalize contour mismatch [5,6]. While these methods enhance geometric awareness during training, they primarily operate at the level of loss design or architectural modification. Comparatively limited attention has been directed toward how multimodal MRI inputs are structured prior to feature extraction. In most multimodal segmentation pipelines, MRI modalities including T1, contrast-enhanced T1 (T1c), T2, and FLAIR are concatenated as independent channels and directly used as network input [7]. Although this approach is straightforward, it assumes that early convolutional layers can internally disentangle shared anatomical intensity structure, modality-specific contrast variations, and structural discontinuities. When these heterogeneous signals are entangled within a single tensor, the network must resolve multiple representational subproblems during early training stages. Tumor morphology further increases segmentation difficulty. Brain tumors vary substantially in size, internal heterogeneity, and contour complexity [8]. Small tumor components and thin peripheral enhancing regions may occupy only a limited number of voxels, making them particularly vulnerable to omission during repeated downsampling operations in encoder–decoder architectures. In contrast, larger tumors often exhibit irregular geometries and heterogeneous intensity distributions that require refined spatial modeling. A representation that does not explicitly organize contrast relationships and structural cues may therefore struggle to maintain stability across such scale variability.

Motivated by these observations, we propose CMYD-SurfaceNet, a representation-driven and scale-aware framework for multimodal brain tumor segmentation. Instead of relying solely on architectural expansion, the proposed approach restructures multimodal inputs before feature learning begins. First, selected tumor-relevant modalities are organized into a task-oriented pseudo-RGB configuration to spatially align complementary contrast information. This intermediate representation is then transformed into the CMYK color space in order to separate shared luminance structure from modality-dominant chromatic variations. Although the luminance component is used during decomposition, it is not retained in the final input. Instead, it is replaced by a gradient-derived boundary density channel that explicitly encodes spatial intensity transitions associated with tumor

margins. The resulting four-channel CMYD representation, composed of chromatic components C, M, and Y together with a structural boundary descriptor D, defines a structured feature manifold in which contrast dominance and geometric discontinuities are disentangled into complementary dimensions. This organization reduces inter-channel redundancy and introduces deterministic structural priors prior to optimization. To further address tumor scale variability, the CMYD representation is integrated within a two-stage cascaded nnU-Net architecture. The first stage performs coarse global tumor localization across the full brain volume. The second stage refines segmentation within a region of interest extracted around the predicted lesion, enabling higher-resolution boundary modeling and improved preservation of small tumor components. This cascaded strategy redistributes computational focus from global detection to localized contour refinement, thereby enhancing sensitivity across different lesion scales.

Unlike most existing multimodal segmentation approaches, which mainly focus on network architecture or loss design, this work takes a different perspective by rethinking how the input itself is structured. Instead of simply concatenating MRI modalities, we reorganize them into a representation that separates contrast-related information from structural boundary cues before learning begins. This perspective suggests that segmentation performance is not only a function of model capacity, but also of how multimodal information is presented to the network. By introducing structured input decomposition together with a scale-aware cascaded refinement strategy, the proposed framework improves boundary precision without increasing architectural complexity.

The main contributions of this work can be summarized as follows:

- Unlike conventional multimodal segmentation approaches that rely primarily on architectural design, this work establishes a representation-driven perspective, highlighting the importance of input structuring as an effective alternative for improving boundary-aware segmentation.
- It introduces the CMYD representation, which explicitly separates contrast-related information from structural boundary cues, enabling more effective learning of tumor appearance and geometry.
- It proposes a scale-aware cascaded segmentation framework that improves sensitivity to small tumor components while preserving global context.
- It demonstrates, through extensive experiments, that the proposed approach consistently improves boundary-sensitive metrics, particularly HD95 and Surface Dice, without increasing model complexity.
- The results further indicate improved boundary stability and robustness across heterogeneous imaging conditions, demonstrating potential for real-world clinical applicability.

2. Related Work

Multimodal MRI has long been recognized as essential for reliable glioma segmentation because each imaging sequence captures complementary tissue characteristics [9]. T1-weighted imaging primarily provides anatomical context, contrast-enhanced T1 highlights enhancing tumor components, T2 reflects fluid abnormalities, and FLAIR improves visualization of edema and infiltrative regions. The integration of these modalities has therefore become a foundational element in brain tumor segmentation research [10].

Large-scale benchmark initiatives such as the Brain Tumor Segmentation (BraTS) challenge have further standardized multimodal evaluation protocols and stimulated methodological development [10]. These benchmarks have revealed that segmentation performance must remain robust under substantial inter-patient and inter-institution variability, including heterogeneous tumor morphology and imaging conditions. Prior to deep learning, segmentation approaches relied on handcrafted features combined with clustering,

probabilistic models, and graph-based optimization [11]. Although these methods represented important early contributions, they struggled to generalize across diverse tumor appearances. The introduction of convolutional neural networks substantially improved robustness and accuracy.

Encoder–decoder architectures, particularly U-Net and its 3D extensions, became central to volumetric medical image segmentation due to their multiscale representation learning capabilities [12,13]. Subsequent developments introduced residual variants, V-Net, and transformer-based models to further enhance contextual modeling [14–16]. Among these, nnU-Net demonstrated that a carefully self-configured framework can achieve strong baseline performance across diverse biomedical tasks without extensive manual tuning [17]. Despite architectural advancements, most multimodal pipelines continue to rely on simple channel concatenation of MRI sequences [18]. While effective in practice, this strategy assumes that the network can internally disentangle shared intensity baselines and modality-specific contrast information. To improve multimodal interaction, attention-based fusion mechanisms and feature-level integration strategies have been proposed [19,20]. These approaches enhance adaptive weighting of modalities but generally operate within the network rather than restructuring the input representation itself. Another major direction concerns boundary-aware learning. Because overlap-based metrics such as Dice are relatively insensitive to local contour deviations, researchers have proposed surface-based losses and Hausdorff distance–inspired optimization strategies to improve geometric precision [21,22]. Auxiliary boundary prediction branches have also been incorporated to reinforce contour sensitivity during training [23]. These methods confirm that structural accuracy is critical when segmentation results are used for clinical planning. Closely related to boundary precision is tumor scale variability. Brain tumors vary significantly in spatial extent, ranging from small enhancing nodules to large heterogeneous masses. Small lesions are particularly susceptible to omission during repeated downsampling operations in encoder–decoder networks. To mitigate this limitation, cascaded segmentation strategies have been proposed, where an initial network performs coarse localization and a second stage refines predictions within a restricted region of interest [24,25]. Cascaded architectures have demonstrated improved sensitivity to small lesions and enhanced boundary refinement by redistributing computational focus toward high-resolution modeling. In parallel, representation-oriented transformations have been explored as alternative means of organizing multimodal information. Pseudo-color strategies that map grayscale MRI modalities into RGB-like configurations have been investigated to improve contrast separation [26]. While these approaches suggest that structured channel organization can influence feature learning, they were largely designed for visualization or heuristic enhancement rather than principled representation-level disentanglement within end-to-end segmentation systems.

Recent studies have followed different directions to improve medical image analysis. Some works have focused on new network architectures, such as NeuroMamba, which uses state-space modeling to support efficient and explainable brain tumor segmentation [27]. Other methods, including ESE-UNet, aim to design lighter encoder–decoder models that can maintain good segmentation performance while reducing computational complexity [28]. In addition, multi-task learning has been explored by combining glioma segmentation with survival prediction, showing the potential benefit of learning related clinical tasks together [29]. Edge-aware convolutional networks have also been proposed to better capture tumor and organ boundaries, which is important for accurate segmentation [30]. Outside the segmentation task, systems such as RADAI highlight the broader use of deep learning for radiological image classification and reflect the increasing demand for robust and interpretable medical imaging models [31].

Although these studies are related to the present work, they mainly focus on modifying the model architecture or extending the learning objective. In contrast, the proposed approach does not primarily introduce a new segmentation backbone. Instead, it concentrates on reorganizing multimodal MRI inputs before feature extraction, with the goal of providing the network with more informative and consistent representations.

Overall, existing research has primarily focused on improving network architecture, loss formulation, and feature-level multimodal fusion. Comparatively limited work has examined whether multimodal MRI inputs themselves can be reorganized into a structurally disentangled and scale-aware representation prior to learning. This gap motivates the representation-driven and cascaded framework proposed in this study.

3. Methodology

3.1. Framework Overview

Brain tumor segmentation from multimodal magnetic resonance imaging (MRI) remains challenging due to substantial variability in lesion morphology, spatial extent, and internal heterogeneity. Although modern encoder–decoder architectures have achieved strong volumetric performance, most existing pipelines rely on direct modality concatenation, where T1, T1c, T2, and FLAIR sequences are stacked as independent input channels [32]. While effective, this strategy assumes that early convolutional layers can internally disentangle modality correlations, normalize intensity distributions, and extract boundary-relevant features during the initial stages of optimization. In practice, shared anatomical intensity structure, modality-specific contrast dominance, and geometric discontinuities are embedded within a single tensor. Such representational coupling increases the burden on early feature extraction layers and may limit robustness, particularly for small enhancing components and irregular tumor margins [33]. These limitations motivate restructuring multimodal inputs before segmentation learning begins. In this work, we address this limitation by restructuring multimodal MRI inputs prior to network learning. The proposed framework achieves this through representation-level transformation combined with scale-aware refinement. Unlike prior works that use these transformations independently, the proposed framework integrates pseudo-RGB structuring, CMY-based contrast decoupling, and explicit gradient-driven boundary encoding into a unified representation space tailored for segmentation learning.

Rather than increasing network depth or introducing complex fusion modules, the method reorganizes the input feature space to explicitly separate contrast information from structural cues.

First, tumor-sensitive MRI modalities are organized into a task-driven pseudo-RGB configuration. Structured cross-channel representations have been shown to influence feature separability and multimodal interaction [34]. By spatially aligning complementary contrast sources within a compact three-channel representation, this step reduces inter-channel redundancy prior to convolutional processing. Second, the pseudo-RGB representation is transformed into the CMYK color space to disentangle luminance structure from chromatic contrast variations. Color-space decomposition has been explored as a mechanism for reducing channel correlation and enhancing discriminative learning in medical imaging contexts [35]. In our formulation, the luminance component is used only for intermediate separation and is not retained in the final input. Instead, the chromatic channels C, M, and Y encode modality-dominant contrast deviations. To reinforce geometric sensitivity, a boundary density descriptor derived from gradient magnitude analysis is incorporated as an additional channel. Gradient-based structural cues have been shown to improve contour stability and boundary-aware feature learning [36]. The resulting CMYD representation—composed of C, M, Y, and D—forms a structured feature space

in which contrast imbalance and boundary transitions are encoded along complementary dimensions. To address tumor scale variability, the CMYD representation is integrated within a two-stage cascaded nnU-Net architecture. Cascaded segmentation strategies have demonstrated improved detection of small lesions and enhanced boundary refinement through coarse-to-fine modeling [24,25,37]. In the first stage, a coarse network performs whole-brain tumor localization. In the second stage, segmentation is refined within a region of interest centered on the predicted lesion, enabling higher effective resolution and improved preservation of small tumor components [38,39]. By combining representation-level structural decoupling with scale-aware cascaded refinement, the proposed framework enhances boundary-sensitive segmentation without increasing architectural complexity. The overall workflow is illustrated in Figure 1.

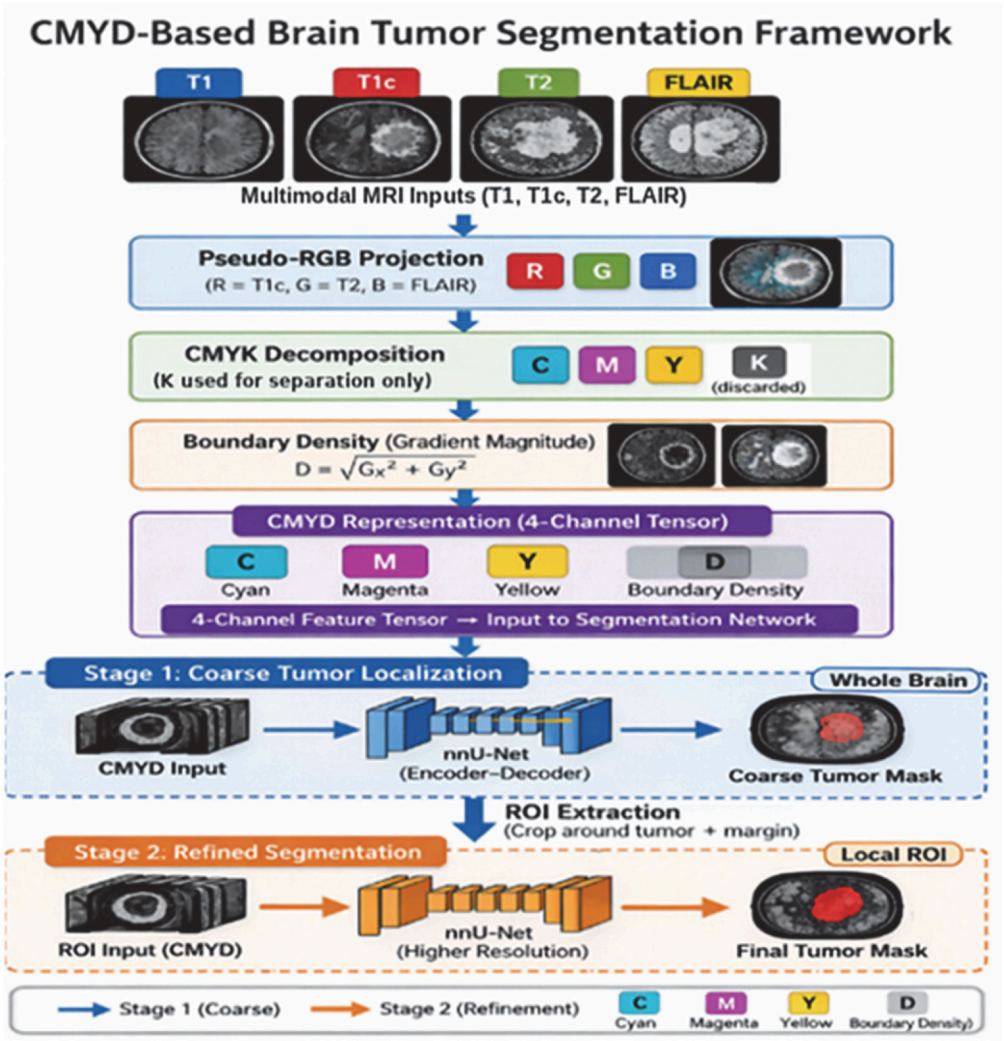


Figure 1. Overview of the proposed CMYD-SurfaceNet framework for scale-aware cascaded multimodal MRI brain tumor segmentation. Colors represent different modality and feature transformations: red/green/blue correspond to T1c/T2/FLAIR mapping, while cyan/magenta/yellow denote CMY decomposition channels. The D channel represents boundary density derived from gradient magnitude. Blue and orange blocks indicate Stage 1 (coarse localization) and Stage 2 (refined segmentation), respectively.

3.2. Multimodal MRI Inputs

Multimodal magnetic resonance imaging (MRI) remains the standard imaging protocol for glioma evaluation because different sequences reveal complementary aspects of tumor morphology and tissue alteration. Large collaborative efforts such as the BraTS challenges

have repeatedly shown that combining multiple MRI modalities improves the delineation of tumor subregions compared to single-sequence analysis [40,41]. This multimodal integration is therefore considered essential for robust and clinically meaningful segmentation. In this study, four routinely acquired MRI sequences are considered: T1-weighted (T1), contrast-enhanced T1-weighted (T1c), T2-weighted (T2), and fluid-attenuated inversion recovery (FLAIR). Each modality emphasizes a different biological or structural characteristic. T1-weighted imaging provides anatomical reference and overall structural context. After contrast injection, T1c highlights regions of blood–brain barrier disruption, typically corresponding to actively enhancing tumor tissue. T2-weighted images are sensitive to fluid content and therefore capture edema and tumor-associated swelling. FLAIR suppresses cerebrospinal fluid signals, improving the visibility of infiltrative tumor margins and peritumoral abnormalities. The complementary diagnostic roles of these modalities have been consistently validated in curated glioma MRI datasets and radiomic studies [42–44]. Although all four sequences are clinically relevant, the representation restructuring stage of the proposed framework focuses on tumor-discriminative contrast behavior. Specifically, the pseudo-RGB projection is constructed using T1c, T2, and FLAIR images. These modalities collectively encode enhancement patterns, edema distribution, and infiltrative characteristics, forming a contrast-oriented basis for the subsequent structural decoupling step. The T1 modality remains part of the preprocessing pipeline to preserve anatomical consistency but is not directly included in the pseudo-color mapping process. Prior to representation-level transformation, all volumes undergo standardized preprocessing procedures, including spatial co-registration, resampling to isotropic voxel spacing, and intensity normalization. These steps ensure cross-modality alignment and numerical stability, allowing the subsequent representation transformation to operate on harmonized data. The characteristics and functional roles of the MRI modalities considered in this study are summarized in Table 1.

Table 1. MRI modalities and their role in brain tumor analysis.

MRI Modality	Description	Tumor-Related Information
T1	Anatomical MRI sequence	Provides structural reference of brain anatomy
T1c	Contrast-enhanced T1-weighted MRI	Highlights enhancing tumor regions associated with contrast uptake
T2	Fluid-sensitive MRI sequence	Emphasizes edema and fluid-related abnormalities
FLAIR	Fluid-attenuated inversion recovery	Enhances visualization of infiltrative tumor regions and peritumoral edema

3.3. Pseudo-RGB Representation

In conventional multimodal MRI segmentation frameworks, imaging modalities are commonly concatenated as independent input channels and directly processed by convolutional neural networks. Although this strategy is straightforward and computationally efficient, it does not impose any structural alignment between modality-specific contrast mechanisms before feature extraction. Since early convolutional layers perform weighted cross-channel aggregation, the initial organization of modalities influences how enhancement patterns, edema distributions, and infiltrative regions interact during the first stages of representation learning. In glioma MRI, T1c typically exhibits stronger discriminative

contrast compared to T2 and FLAIR, motivating an ordered channel structuring rather than arbitrary stacking.

The purpose of this transformation is not to create a visual color representation, but to reorganize multimodal MRI signals into a structured form that facilitates learning. By mapping tumor-sensitive modalities into a pseudo-RGB configuration, complementary contrast information is spatially aligned within a compact three-channel representation, reducing arbitrary inter-channel interactions and improving feature consistency during early convolutional processing. To introduce explicit structural organization at the representation level, we define a pseudo-RGB embedding that reorganizes selected tumor-sensitive modalities into a compact three-channel tensor. Let the multimodal input be defined as

$$X = \{X_{T1c}, X_{T2}, X_{FLAIR}\}, \quad X_m \in \mathbb{R}^{H \times W}, \tag{1}$$

where each modality X_m represents a spatially aligned and intensity-normalized MRI slice defined over a two-dimensional spatial domain of height H and width W. Since these modalities are spatially co-registered and share identical resolution, they can be jointly represented as a three-channel tensor. Based on this unified representation, we introduce a deterministic embedding operator:

$$\Phi: \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{H \times W \times 3} \tag{2}$$

such that

$$X_{RGB} = \Phi(\mathrm{X}) \tag{3}$$

The operator is defined voxel-wise as

$$X_{RGB}(i, j, 1) = X_{T1c}(i, j) \tag{4}$$

$$X_{RGB}(i, j, 2) = X_{T2}(i, j) \tag{5}$$

$$X_{RGB}(i, j, 3) = X_{FLAIR}(i, j) \tag{6}$$

for all spatial coordinates (i, j) within the image domain.

The mapping Φ is deterministic and linear, performing a fixed channel re-indexing without modifying spatial geometry or voxel intensities. Mathematically, it can be interpreted as an identity-preserving permutation operator in tensor space. No additional information is introduced; rather, the existing modality-specific signals are reorganized along predefined channel axes. Although structurally simple, this embedding has representational implications for convolutional processing.

A convolutional layer computes feature responses according to

$$\mathrm{F} = \sigma(W \times X_{RGB} + b) \tag{7}$$

where W denotes the learnable convolutional kernels, b represents the bias term associated with each output feature map, and $\sigma(\cdot)$ is a nonlinear activation function. The bias term b shifts the linear response before activation, allowing the network to learn features that are not constrained to be zero-centered. Because convolution performs linear combinations across channels prior to nonlinearity, the statistical arrangement of modalities directly affects cross-channel interaction patterns during early feature extraction.

By embedding enhancement-dominant (T1c), edema-dominant (T2), and infiltration-sensitive (FLAIR) signals into orthogonal channel dimensions, the operator Φ imposes a contrast-aligned basis before optimization. This structured embedding reduces arbitrary modality entanglement and facilitates more stable cross-channel aggregation in the

first convolutional layers. Importantly, the pseudo-RGB transformation is not intended for perceptual colorization or visual enhancement. Instead, it serves as a representation-conditioning stage within the segmentation pipeline. By reorganizing multimodal information into a structured embedding space, it prepares the input tensor for the subsequent CMY-based structural decoupling stage, where shared intensity structure and modality-dominant contrast components are further separated. In summary, the pseudo-RGB embedding constitutes a mathematically well-defined, identity-preserving transformation that reorganizes tumor-relevant modalities into a contrast-aligned representation. While computationally lightweight, this re-indexing step influences early convolutional dynamics and provides a structured foundation for boundary-sensitive learning in the proposed CMYD-SurfaceNet framework.

3.4. CMYK Transformation

Although the pseudo-RGB embedding introduces structured channel alignment, residual inter-channel redundancy may persist due to shared anatomical intensity baselines across MRI modalities [45]. In multimodal brain imaging, non-pathological tissues frequently exhibit comparable brightness distributions in different sequences, which introduces correlated luminance components across channels. Such redundancy may bias early convolutional filters toward common structural intensity rather than modality-specific contrast deviations that are more indicative of tumor subregions [46–48].

The CMYK transformation is introduced to reduce redundancy across channels by separating shared luminance from modality-specific contrast components. To mitigate this effect, the pseudo-RGB representation is transformed into the CMYK color space. Color-space decomposition techniques have long been used in image processing to separate luminance and chromatic components in order to reduce channel correlation and enhance contrast discrimination [49]. In the present work, this transformation is employed not for perceptual color modeling, but for structural intensity decoupling prior to segmentation learning.

Let the normalized pseudo-RGB tensor be defined as

$$X_{RGB} = \{R, G, B\}, \quad R, G, B \in \left[0, 1\right]^{H \times W} \tag{8}$$

The shared luminance component K is computed voxel-wise as

$$K(i,j) = 1 - max(R(i,j), G(i,j), B(i,j)) \tag{9}$$

for all spatial coordinates (i, j).

This formulation corresponds to the standard RGB-to-CMYK conversion model and estimates the dominant brightness contribution at each voxel [49].

For voxels where $K(i,j) \neq 1$ the chromatic components are defined as

$$C(i,j) = \frac{1 - R(i,j) - K(i,j)}{1 - K(i,j)} \tag{10}$$

$$M(i,j) = \frac{1 - G(i,j) - K(i,j)}{1 - K(i,j)} \tag{11}$$

$$Y(i,j) = \frac{1 - B(i,j) - K(i,j)}{1 - K(i,j)} \tag{12}$$

When $K(i,j) = 1$, chromatic values are set to zero to ensure numerical stability.

Through this normalization, the chromatic channels encode relative contrast deviations with respect to the shared luminance baseline. From a statistical perspective, removing the

dominant luminance component reduces inter-channel covariance and improves feature conditioning in early convolutional layers [48]. This is particularly relevant in tumor segmentation, where pathological regions are often characterized by modality-specific contrast dominance rather than uniform intensity elevation. Although the CMYK transformation yields four channels, the luminance component K is not retained in the final representation. Its role is restricted to intermediate structural decoupling. The final representation therefore consists of the chromatic components $\{C, M, Y\}$, which provide contrast-sensitive inputs for subsequent segmentation learning. By separating common anatomical brightness from modality-dependent contrast imbalance, the CMYK-based transformation reshapes the input feature space prior to optimization. This structured intensity decoupling supports more stable multimodal feature interaction and facilitates boundary-aware refinement within the proposed CMYD-SurfaceNet framework.

3.5. Boundary Density Estimation

Although the CMY decomposition reduces modality entanglement by separating chromatic contrast components, accurate tumor segmentation additionally requires explicit sensitivity to spatial discontinuities. In brain MRI, tumor margins often correspond to local intensity transitions between pathological and healthy tissues. Several recent studies have emphasized that improving geometric consistency requires incorporating boundary-aware modeling into segmentation pipelines [50–52]. The boundary density channel is introduced to explicitly encode spatial discontinuities corresponding to tumor margins.

To explicitly encode structural transitions, we introduce a boundary density descriptor derived from gradient magnitude analysis. Let the structural intensity map be defined as

$$I(i,j) = \frac{1}{3}(R(i,j) + G(i,j) + B(i,j)) \tag{13}$$

where R, G, and B denote the normalized pseudo-RGB channels prior to CMY decomposition.

The spatial gradient of I is computed as

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right) \tag{14}$$

and the boundary density map is defined as the Euclidean norm of the gradient:

$$D(i,j) = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \tag{15}$$

In practice, the partial derivatives are approximated using discrete convolutional operators such as Sobel filters, which provide stable edge estimation under moderate noise conditions. Gradient-based boundary modeling has been widely adopted in recent segmentation research to enhance contour localization and reduce Hausdorff distance–related errors [50,51].

The resulting channel D captures regions of strong spatial intensity variation and encodes geometric discontinuities independently of chromatic contrast dominance. Unlike learned boundary branches that require additional trainable parameters, this formulation introduces a deterministic structural prior at the representation level. Such structural guidance complements boundary-aware optimization strategies previously proposed in the literature [52], while maintaining architectural simplicity. Importantly, the boundary density channel replaces the luminance component K obtained during CMYK decomposition. As a result, the final representation is defined as:

$$X_{CMYD} = \{C, M, Y, D\} \tag{16}$$

This four-channel tensor integrates contrast-dominant chromatic information with explicit geometric sensitivity. By combining modality disentanglement and structural gradient encoding, the CMYD representation provides a scale-consistent and boundary-aware input space for the subsequent cascaded nnU-Net segmentation stages. Recent evidence suggests that improving boundary modeling can significantly reduce HD95 and surface-related discrepancies even when Dice scores remain comparable [50–52]. The proposed boundary density formulation therefore aligns with contemporary segmentation research trends while preserving computational efficiency and interpretability.

3.6. Cascaded Multi-Stage Segmentation Framework

Brain tumor segmentation is inherently challenging due to substantial inter-patient variability, irregular lesion morphology, and severe class imbalance between tumor subregions and healthy tissue. In full-resolution segmentation settings, a single-stage model may capture global anatomical context but often lacks sufficient sensitivity to precisely delineate fine tumor boundaries, particularly in small enhancing or infiltrative regions. To address this limitation, we adopt a cascaded coarse-to-fine segmentation framework that explicitly separates global localization from local structural refinement.

Let the CMYD representation be defined as

$$X_{CMYD} \in \mathbb{R}^{H\times W\times 4} \tag{17}$$

where H and W denote the spatial dimensions of the input slice. A first segmentation network $f_1(\cdot)$, parameterized by θ_1, produces a voxel-wise probability map.

$$P^{(1)} = f_1(X_{CMYD};\theta_1) \tag{18}$$

For multi-class prediction, voxel probabilities are computed using the Softmax function.

$$P^{(1)}_{(i,j,c)} = \frac{\exp(z_{i,j,c})}{\sum_{k=1}^{C}\exp(z_{i,j,c})} \tag{19}$$

where $z_{i,j,c}$ denotes the logit corresponding to class c, and C is the number of segmentation classes. To ensure both voxel-wise classification accuracy and overlap consistency, a hybrid loss combining Cross-Entropy and Soft Dice is employed. The Cross-Entropy term enforces probabilistic consistency.

$$\mathcal{L}^{(1)}_{CE} = -\sum_{i,j}\sum_{c=1}^{C} Y_{i,j,c}\log\left(P^{(1)}_{(i,j,c)}\right) \tag{20}$$

while the Soft Dice loss mitigates class imbalance and directly optimizes region overlap.

$$\mathcal{L}^{(1)}_{Dice} = 1 - \frac{2\sum_{i,j} P^{(1)}_{(i,j)}\, Y_{i,j}}{\sum_{i,j} P^{(1)}_{(i,j)} + \sum_{i,j} Y_{i,j} + \varepsilon} \tag{21}$$

The overall loss is given by

$$\mathcal{L}^{(1)} = \mathcal{L}^{(1)}_{CE} + \lambda\mathcal{L}^{(1)}_{Dice} \tag{22}$$

where the parameter λ controls the contribution of the overlap constraint.

The coarse prediction $P^{(1)}$ provides reliable global tumor localization across the full image domain. Based on this prediction, a region of interest Ω is extracted as the minimal

bounding region enclosing the detected tumor area. Restricting the spatial domain to this region reduces irrelevant background influence and allows higher-resolution modeling of structural details. The cropped input is therefore defined as

$$X_{ROI} = X_{CMYD}|_{\Omega} \tag{23}$$

A second segmentation network $f_2(\cdot)$, parameterized by θ_2, refines the segmentation within the ROI:

$$P^{(2)} = f_2(X_{ROI}; \theta_2) \tag{24}$$

The same hybrid loss function is applied at this stage.

$$\mathcal{L}^{(2)} = \mathcal{L}_{CE}^{(2)} + \lambda \mathcal{L}_{Dice}^{(2)} \tag{25}$$

Both stages are jointly optimized through

$$\mathcal{L}_{total} = \alpha\, \mathcal{L}^{(1)} + \beta\, \mathcal{L}^{(2)} \tag{26}$$

where α and β balance global localization and local refinement.

Finally, the refined prediction within the region of interest (Ω) is mapped back to the original image space, where voxels inside Ω are assigned $P^{(2)}$, while those outside Ω retain $P^{(1)}$:

$$P_{final}(i;j) = \begin{cases} P^{(2)}(i;j) & (i;j) \in \Omega \\ P^{(1)}(i;j) & otherwise \end{cases} \tag{27}$$

Through this cascaded formulation, the first stage ensures structural awareness at the global level, while the second stage enhances boundary precision and sensitivity to small tumor subregions. The separation of localization and refinement improves robustness without introducing excessive architectural complexity.

3.7. Representation-Space Statistical Analysis

Although the proposed CMYD representation is inspired by color-space decomposition, its role in this work is not related to visual perception or printing-based color modeling. In other words, CMYD is not used to produce a visually meaningful color image. Rather, it is adopted as a structured preprocessing step that reorganizes multimodal MRI information before it is processed by the segmentation network.

The main purpose of this representation is to separate two types of information that are usually mixed in conventional multimodal inputs. The first is the shared intensity structure, which often appears across different MRI modalities and may introduce redundant information between channels. The second is the modality-specific contrast information, which is more useful for distinguishing tumor-related regions. By reducing this redundancy, the network receives a more organized input representation, where each channel contributes more distinctive and complementary information.

To evaluate this effect quantitatively, we measure the statistical dependence between the representation channels using two indicators: the mean pairwise cross-channel covariance and the normalized mutual information. Lower covariance values suggest that the channels are less linearly correlated, while lower normalized mutual information indicates that they share less redundant information. Therefore, a decrease in both measures suggests that the representation provides a more decorrelated and informative input space for segmentation learning.

This analysis is carried out on several input representations, including raw channel concatenation, pseudo-RGB, HSV, LAB, CMY, and the proposed CMYD representation.

HSV and LAB are included because they also separate intensity-related information from chromatic components, making them relevant alternatives for comparison. The goal of this comparison is to determine whether CMYD offers a more effective way to reduce inter-channel redundancy while also improving boundary-sensitive information. Unlike the other representations, CMYD incorporates a boundary density channel, which helps emphasize spatial discontinuities corresponding to tumor margins and may therefore support more accurate contour delineation.

4. Data Description and Preprocessing

To evaluate the proposed CMYD-SurfaceNet under both standardized benchmark conditions and real-world clinical settings, experiments were conducted on the public BraTS 2023 benchmark dataset [40] and on a retrospective clinical cohort collected at the University Hospital Center (CHU) of Fez, Morocco. This dual evaluation protocol enables standardized comparison while also assessing robustness in routine clinical practice.

4.1. Data Acquisition

The BraTS 2023 dataset consists of multi-institutional MRI scans collected from glioma patients across several clinical centers, reflecting variability in imaging protocols and acquisition settings. Each case includes four co-registered MRI modalities (T1, T1c, T2, and FLAIR). Ground-truth annotations are provided for three tumor subregions—Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET)—and were generated by expert neuroradiologists following standardized BraTS labeling guidelines [40].

4.2. Data Preprocessing

All MRI volumes were processed using the official BraTS preprocessing pipeline, including skull stripping, bias-field correction, spatial co-registration across modalities, and resampling to an isotropic voxel resolution of 1 mm^3. These steps ensure spatial consistency and intensity standardization across all cases. No additional preprocessing procedures were introduced beyond this standardized pipeline.

4.3. Training Protocol

All 1251 available cases from the BraTS 2023 training set were used in this study. A five-fold cross-validation scheme was adopted, ensuring strict patient-level separation across folds to prevent data leakage and provide a reliable evaluation of the model performance.

4.4. CHU Fez Clinical Dataset

To evaluate how the BraTS-trained model performs under real local clinical conditions, we included a retrospective clinical cohort from the University Hospital Center (CHU) of Fez, Morocco. This cohort included multimodal MRI scans from 86 glioma patients acquired during routine clinical practice. Unlike the standardized BraTS setting, these data reflect heterogeneous acquisition protocols and scanner configurations, making them suitable for assessing model behavior under local clinical variability.

Tumor masks were manually delineated by two experienced neuroradiologists following established clinical guidelines. The annotations were then reviewed to ensure consistency and reliability. Ethical approval for the use of this cohort was obtained from the institutional review board of CHU Fez, and all patient data were anonymized before analysis.

The CMYD-SurfaceNet model was trained exclusively on the BraTS 2023 training data and directly evaluated on all 86 CHU Fez cases without additional training, fine-tuning, or cohort-specific adaptation. Therefore, the CHU Fez experiment represents an external zero-shot evaluation under local clinical acquisition conditions. This protocol was used to assess how the BraTS-trained model behaves on retrospective clinical data acquired during

routine practice. In several cases discussed in a preoperative surgical context, the resulting boundary-aware segmentation outputs were shared with the radiological and surgical teams as complementary decision-support material for expert review. These outputs were used to support visualization and assessment of tumor boundaries before surgery.

4.5. Ethical Statement

This study used two datasets. The BraTS 2023 dataset is publicly available and consists of fully anonymized data; therefore, no additional ethical approval was required. The second dataset (CHU Fez) is a retrospective clinical cohort collected under routine practice. It consists of previously collected and anonymized data, with no direct interaction or involvement of human participants in our work.

4.6. Further Preprocessing

To ensure consistency between datasets, the same preprocessing pipeline applied to the BraTS dataset was also used for the CHU Fez data. This includes spatial co-registration, resampling to isotropic voxel resolution (1 mm^3), and per-modality z-score normalization, defined as:

$$X_{norm} = \frac{X - \mu}{\sigma} \tag{28}$$

where μ and σ represent the mean and standard deviation computed independently for each modality. This harmonized preprocessing pipeline reduces inter-dataset variability and ensures fair cross-domain evaluation, as summarized in Table 2.

Table 2. Summary of datasets and evaluation protocols.

Dataset	Patients	Modalities	Annotated Regions	Evaluation Protocol
BraTS 2023	1251	T1, T1c, T2, FLAIR	WT, TC, ET	5-fold cross-validation
CHU_Fez	86	T1, T1c, T2, FLAIR	WT, TC, ET	External zero-shot testing of the BraTS-trained model

4.7. Statistical Analysis

Quantitative findings are presented as mean values with their corresponding standard deviations. For the BraTS 2023 dataset, model performance was assessed through a five-fold cross-validation strategy, while ensuring that patients included in the training folds were completely separated from those used for validation. For the CHU Fez cohort, the final evaluation was carried out on all 86 cases as an external zero-shot test using the BraTS-trained model, without additional training or fine-tuning on CHU Fez data.

Confidence intervals were calculated using bootstrap resampling with 1000 iterations, providing a robust estimate of result variability. Since the revised manuscript focuses on descriptive performance comparison rather than formal statistical significance testing, no *p*-value claim is reported.

5. Environment and Hyperparameter Configuration

All experiments were implemented in PyTorch 2.1.0 using a 3D nnU-Net backbone. The experimental environment and software configuration are summarized in Table 3. The network architecture was kept unchanged across all experiments. Instead, performance improvements arise from restructuring the input representation as described in Section 3. The

number of trainable parameters (approximately 31.2 million) remains identical to the baseline nnU-Net, confirming that gains are attributable to representation-level modifications rather than architectural expansion.

Table 3. Environment Configuration.

Component	Specification
GPU	NVIDIA A100 40 GB
CPU	AMD EPYC 7742 (64-core)
RAM	256 GB DDR4
OS	Ubuntu 22.04 LTS
Python	3.10
PyTorch	2.1.0
CUDA/cuDNN	12.1/8.9
XAI Library	Captum 0.6
Validation	5-fold cross-validation (BraTS) + external zero-shot testing (CHU Fez)
CI Computation	Stratified bootstrap (n = 1000)

Training was performed using the AdamW optimizer with an initial learning rate of 3×10^{-4} combined with cosine annealing scheduling. The training hyperparameters and data augmentation settings are summarized in Table 4. Due to GPU memory constraints, the batch size was set to 2, and models were trained for 100 epochs. Early stopping with a patience of 5 epochs was applied based on the validation loss to prevent overfitting and improve generalization. Data augmentation included random rotations ($\pm 15^{\circ}$), scaling (0.9–1.1), elastic deformation, Gaussian noise injection, and intensity perturbations to improve generalization. An auxiliary boundary-aware supervision term with weighting coefficient $\lambda = 0.1$ was incorporated to encourage improved contour consistency during optimization.

Table 4. Hyperparameter Configuration.

Parameter	Value
—Training—	
Optimizer	AdamW
Initial LR	3×10^{-4}
LR schedule	Cosine annealing
Batch size	2
Max epochs	100
Early stopping	Patience 5
Weight decay	0.01
Loss function	Cross-Entropy + Dice
—Data Augmentation—	
Rotation	$\pm 15^{\circ}$
Flip	H + V
Brightness/Contrast	$\pm 20\%$
Gaussian noise	sigma = 0.02
—Architecture/Model—	
Input resolution	384×384
Scales	4 (stride 4, 8, 16, 32)
Boundary loss weight	$\lambda = 0.1$

6. Results

6.1. Main Results on BraTS 2023

The quantitative evaluation of CMYD-SurfaceNet was conducted on the BraTS 2023 [40] validation set and compared with the baseline 3D nnU-Net. Evaluation metrics included Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD95), and Surface Dice at 1 mm tolerance (SD@1), enabling joint assessment of overlap accuracy and contour fidelity.

As shown in Table 5, CMYD-SurfaceNet achieves moderate Dice improvements across all tumor regions, while producing clearer gains in boundary-sensitive metrics. For WT, HD95 decreases from 3.62 mm to 2.41 mm and Surface Dice at 1 mm increases from 0.821 to 0.892. Similar reductions in HD95 are observed for TC and ET, indicating improved contour conformity. These results suggest that the proposed framework primarily improves geometric refinement rather than substantially increasing volumetric overlap.

Table 5. Quantitative segmentation performance on BraTS 2023 (baseline vs. proposed method).

Method	**nnU-Net**			**CMYD-SurfaceNet**		
Region	WT	TC	ET	WT	TC	ET
DSC	0.913	0.869	0.851	0.921	0.873	0.858
HD95 (mm)	3.620	4.15	3.98	2.41	2.98	2.74
SD@1	0.821	0.765	0.743	0.892	0.828	0.816

6.2. Representation-Space Comparison

To justify the proposed representation, we compared CMYD with alternative representations, including raw channel concatenation, pseudo-RGB, HSV, LAB, and CMY, as presented in Table 6. The goal of this analysis is to evaluate whether CMYD reduces inter-channel redundancy and improves boundary-sensitive segmentation. All representation comparisons were conducted using the same preprocessing pipeline, training protocol, and validation split to ensure that performance differences are attributable to the input representation rather than experimental variation.

Table 6. Representation-space statistical analysis and segmentation performance on the WT region.

Representation	**Mean Covariance**	**NMI**	**WT DSC**	**WT HD95**	**WT SD@1**
Raw channel concatenation	0.486	0.612	0.913	3.62	0.821
Pseudo-RGB	0.421	0.553	0.915	3.28	0.845
HSV	0.337	0.441	0.917	3.05	0.862
LAB	0.318	0.415	0.918	2.88	0.872
CMY	0.292	0.391	0.918	2.94	0.860
CMYD	0.244	0.342	0.921	2.41	0.892

The comparative results show that CMYD provides the most effective input representation among the evaluated configurations. It achieves the lowest mean covariance and normalized mutual information, which indicates that the statistical dependence between channels is reduced more effectively than with raw channel concatenation, pseudo-RGB, HSV, LAB, or CMY. This reduction in redundancy suggests that the CMYD representation offers a more organized input space, where the channels provide complementary rather than repetitive information to the segmentation network.

From a segmentation perspective, CMYD also produces the strongest boundary-related performance, with the lowest WT HD95 and the highest WT Surface Dice at 1 mm tolerance. Although HSV and LAB improve over raw concatenation by separating intensity-related information from chromatic components, their results remain less favorable than those obtained with CMYD. This indicates that intensity–chromatic decomposition alone is not sufficient to achieve the same level of boundary conditioning. The additional boundary density channel in CMYD introduces explicit structural information related to tumor margins, which helps improve contour localization and geometric consistency. In this comparison, lower values of mean covariance, normalized mutual information, and WT HD95 reflect better performance, whereas higher values of WT DSC and WT SD@1 indicate better segmentation accuracy.

6.3. Alternative Pseudo-RGB Channel Ordering

To examine the stability of the selected pseudo-RGB configuration, we also evaluated alternative modality-to-channel assignments. This experiment was conducted to determine whether the adopted mapping, in which T1c, T2, and FLAIR are assigned to the R, G, and B channels, respectively, provides a more effective representation than other possible channel orderings, as presented in Table 7. By comparing these configurations under the same experimental setting, we aimed to ensure that the proposed mapping was not an arbitrary choice but a robust arrangement for organizing tumor-relevant MRI contrast information before segmentation learning.

Table 7. Robustness analysis of pseudo-RGB modality-to-channel assignments on the WT region.

Pseudo-RGB Mapping	WT DSC	WT HD95 (mm)	WT SD@1 mm
T1c → R, T2 → G, FLAIR → B	0.915	3.28	0.845
T1c → R, FLAIR → G, T2 → B	0.914	3.36	0.840
T2 → R, T1c → G, FLAIR → B	0.913	3.42	0.836
FLAIR → R, T1c → G, T2 → B	0.914	3.39	0.838

The selected ordering provides the strongest boundary-sensitive performance. This is consistent with the clinical role of the selected modalities, where T1c emphasizes enhancing tumor tissue, T2 captures edema-related abnormalities, and FLAIR improves visualization of infiltrative margins.

6.4. Extended Ablation Study Across WT, TC, and ET

To further clarify the role of each component in the proposed CMYD representation, we conducted an extended ablation study on the BraTS 2023 validation set, considering the three tumor regions: WT, TC, and ET. The analysis started from the baseline setting based on raw channel concatenation, after which the proposed components were added step by step. Specifically, we first introduced the pseudo-RGB organization, followed by the CMY decomposition, and finally the boundary density channel D. This progressive evaluation allowed us to examine how each component contributes to the overall segmentation performance. To ensure a fair comparison, all configurations were evaluated using the same five-fold cross-validation splits.

As presented in Table 8, the progressive introduction of the proposed components leads to consistent improvements across the three tumor regions, with the most noticeable gains observed in HD95. The pseudo-RGB representation slightly improves the results, which suggests that organizing the modalities before feature extraction helps the network learn more stable early representations. Adding the CMY decomposition further reduces

boundary errors, indicating that separating contrast-related information contributes to more reliable tumor delineation. The strongest improvement is obtained with the full CMYD representation, where the boundary density channel is included. This confirms that explicitly encoding structural discontinuities plays an important role in improving contour precision. Overall, the Dice gains remain relatively modest, while the reductions in HD95 are more substantial, showing that the main contribution of the proposed representation lies in improving boundary alignment rather than simply increasing volumetric overlap. Similar trends are observed for TC and ET, demonstrating that the benefit of CMYD is not restricted to the whole tumor region but also extends to the main tumor subregions.

Table 8. Extended ablation study across WT, TC, and ET regions on the BraTS 2023 validation set.

Input Representation	WT DSC	WT HD95 (mm)	TC DSC	TC HD95 (mm)	ET DSC	ET HD95 (mm)
Raw channel concatenation	0.913	3.62	0.869	4.15	0.851	3.98
Pseudo-RGB representation	0.915	3.28	0.870	3.83	0.853	3.66
Pseudo-RGB with CMY decomposition	0.918	2.94	0.872	3.43	0.855	3.25
Full CMYD representation	0.921	2.41	0.873	2.98	0.858	2.74

6.5. Isolation of Cascaded Refinement and CMYD Representation

To clarify the source of the observed performance gains, we evaluated four experimental configurations that separately examine the effect of the input representation and the cascaded refinement strategy, as presented in Table 9. The first configuration used raw channel concatenation with a single nnU-Net and served as the baseline. The second used the same raw input representation within a cascaded nnU-Net framework, allowing us to assess the contribution of the coarse-to-fine refinement alone. The third configuration combined the CMYD representation with a single nnU-Net in order to isolate the effect of representation-level restructuring. Finally, the complete CMYD-SurfaceNet model combined the CMYD representation with the cascaded refinement framework. This comparison allowed us to determine whether the improvement was mainly driven by the proposed representation, the cascaded design, or the interaction between both components.

Table 9. Contribution analysis of CMYD representation and cascaded refinement across WT, TC, and ET regions.

Configuration	WT DSC	WT HD95 (mm)	WT SD@1 mm	TC DSC	TC HD95 (mm)	TC SD@1 mm	ET DSC	ET HD95 (mm)	ET SD@1 mm
Raw concatenation + single nnU-Net	0.913	3.62	0.821	0.869	4.15	0.765	0.851	3.98	0.743
Raw concatenation + cascaded nnU-Net	0.916	3.05	0.855	0.871	3.55	0.792	0.854	3.30	0.780
CMYD + single nnU-Net	0.918	2.85	0.870	0.872	3.35	0.805	0.856	3.05	0.792
CMYD + cascaded nnU-Net/CMYD-SurfaceNet	0.921	2.41	0.892	0.873	2.98	0.828	0.858	2.74	0.816

The comparison indicates that both the CMYD representation and the cascaded refinement strategy contribute to the improvement in boundary-sensitive performance. Using

a cascaded nnU-Net with raw concatenation already reduces boundary errors compared with the single-stage baseline, showing the benefit of coarse-to-fine refinement. Similarly, using CMYD with a single nnU-Net improves HD95 and Surface Dice, which confirms the positive effect of representation-level restructuring. The best results are obtained when CMYD is combined with the cascaded 3D nnU-Net, suggesting that the two components address complementary aspects of the segmentation problem: CMYD improves the quality of the input representation, while the cascade enhances local boundary refinement.

6.6. Boundary Density Channel and Boundary Loss Ablation

Since the proposed framework includes both a boundary density channel and an auxiliary boundary supervision term, we performed an additional ablation experiment to examine the contribution of each component separately, as presented in Table 10. This analysis was designed to determine whether the improvement comes mainly from the explicit boundary information introduced in the input representation, from the boundary-aware supervision during training, or from their combined use.

Table 10. Ablation analysis of boundary density encoding and auxiliary boundary supervision on the WT region.

Configuration	WT DSC	WT HD95 (mm)	WT SD@1 mm
CMY only, without D and without boundary loss	0.918	2.94	0.860
CMY + boundary density channel D only	0.920	2.55	0.882
CMY + auxiliary boundary loss only	0.919	2.70	0.874
CMY + D + auxiliary boundary loss/full CMYD-SurfaceNet	0.921	2.41	0.892

The boundary density channel and the auxiliary boundary loss contribute to the proposed framework in different but complementary ways. The boundary density channel operates at the input level by adding explicit structural information about tumor margins to the CMYD representation. In contrast, the auxiliary boundary loss acts during training by encouraging the model to learn more accurate contour predictions. The ablation results show that each component improves boundary-sensitive performance when used separately, particularly in terms of HD95 and Surface Dice. However, the best performance is achieved when both components are combined, suggesting that input-level boundary encoding and optimization-level boundary guidance work together to improve tumor boundary delineation and enhance geometric consistency.

6.7. T1 Modality Contribution

Although T1-weighted images provide useful anatomical reference, the proposed pseudo-RGB representation focuses on T1c, T2, and FLAIR because these sequences carry more direct information about tumor appearance. T1c is important for identifying enhancing tumor regions, T2 is sensitive to edema and fluid-related changes, and FLAIR helps reveal infiltrative and peritumoral abnormalities. To verify that excluding T1 does not weaken the model performance, we compared the proposed three-modality configuration with an alternative setting that also includes T1, as presented in Table 11. This comparison was used to determine whether the selected modality combination offers a better balance between anatomical context and tumor-specific contrast.

Table 11. Comparison between the proposed three-modality CMYD configuration and a four-channel variant incorporating T1.

Input Scheme	WT DSC	WT HD95 (mm)	WT SD@1 mm
T1c/T2/FLAIR → CMYD	0.921	2.41	0.892
T1/T1c/T2/FLAIR → four-channel variant	0.920	2.50	0.886

The comparison shows that incorporating T1 into the input representation does not lead to an improvement in boundary-sensitive performance. The T1c/T2/FLAIR-based CMYD configuration achieves slightly better HD95 and Surface Dice results than the four-channel variant including T1. This suggests that, in the proposed pseudo-RGB setting, the selected modalities provide more relevant tumor-specific contrast for boundary delineation. Therefore, T1 is kept as part of the preprocessing and anatomical alignment pipeline, while the pseudo-RGB representation focuses on the modalities that contribute more directly to tumor characterization.

6.8. Evaluation on the CHU Fez Clinical Cohort

To evaluate the zero-shot external behavior of the BraTS-trained model under local clinical acquisition conditions, CMYD-SurfaceNet was tested on all 86 cases from the retrospective CHU Fez clinical cohort without additional training or fine-tuning.

Unlike the standardized BraTS dataset, the CHU cohort reflects real-world clinical conditions, including heterogeneous acquisition protocols, scanner variability, and lower signal-to-noise ratios.

As reported in Table 12, the results obtained on the CHU Fez cohort are slightly lower than those observed on BraTS 2023. This difference is expected, given the domain shift introduced by routine clinical acquisition conditions, heterogeneous scanner configurations, and lower signal-to-noise ratios. Despite this domain shift, CMYD-SurfaceNet consistently outperforms nnU-Net across the evaluated tumor regions, with the most evident gains observed in HD95 and Surface Dice. These findings provide preliminary evidence that the proposed representation can support tumor boundary delineation under heterogeneous local clinical acquisition conditions. However, they should be interpreted as zero-shot external evaluation results and confirmed on larger multi-center clinical cohorts before any prospective clinical deployment.

Table 12. Quantitative zero-shot external segmentation performance on the CHU Fez clinical cohort with 95% confidence intervals.

Method	Region	DSC	95% CI DSC	HD95 (mm)	95% CI HD95	SD@1 mm	95% CI SD@1 mm
nnU-Net	WT	0.895 ± 0.020	[0.885, 0.905]	4.25 ± 0.60	[3.95, 4.55]	0.785 ± 0.040	[0.766, 0.804]
nnU-Net	TC	0.842 ± 0.030	[0.828, 0.856]	5.10 ± 0.75	[4.74, 5.46]	0.718 ± 0.050	[0.694, 0.742]
nnU-Net	ET	0.812 ± 0.035	[0.795, 0.829]	4.78 ± 0.70	[4.44, 5.12]	0.690 ± 0.055	[0.664, 0.716]
CMYD-SurfaceNet	WT	0.908 ± 0.018	[0.899, 0.917]	3.21 ± 0.45	[2.99, 3.43]	0.846 ± 0.035	[0.829, 0.863]
CMYD-SurfaceNet	TC	0.856 ± 0.027	[0.843, 0.869]	3.86 ± 0.58	[3.58, 4.14]	0.775 ± 0.044	[0.754, 0.796]
CMYD-SurfaceNet	ET	0.832 ± 0.030	[0.817, 0.847]	3.61 ± 0.55	[3.35, 3.87]	0.760 ± 0.046	[0.738, 0.782]

6.9. Lesion-Size Stratified Analysis

To further evaluate whether the proposed method improves the detection of small tumor components, we performed a lesion-size stratified analysis. The test cases were divided into small, medium, and large lesion groups according to tumor size, and segmentation performance was assessed separately for each group. Lesion size was defined using the maximum tumor extent measured on the reference annotation. For this analysis, DSC and HD95 were reported in order to evaluate both volumetric overlap and boundary accuracy across different lesion scales.

As shown in Table 13, the largest relative reduction in HD95 is observed in the small-lesion group. This finding indicates that the proposed method is particularly beneficial for challenging tumor components with limited spatial extent. The combination of CMYD-based representation and cascaded refinement appears to improve boundary localization in these difficult cases by providing both structured contrast information and focused local refinement. Although performance gains are also observed for medium and large lesions, the stronger improvement in small lesions supports the relevance of the proposed framework for enhancing sensitivity to small tumor regions.

Table 13. Lesion-size stratified zero-shot performance analysis on the CHU Fez clinical cohort.

Lesion Size Group	Number of Cases	nnU-Net DSC	CMYD DSC	nnU-Net HD95 (mm)	CMYD HD95 (mm)	HD95 Reduction (%)
Small lesions < 5 mm	20	0.805	0.832	5.85	4.10	29.9%
Medium lesions 5–10 mm	31	0.862	0.879	4.72	3.42	27.5%
Large lesions > 10 mm	35	0.910	0.918	3.58	2.74	23.5%

6.10. Literature-Based Comparison with Recent BraTS Segmentation Methods

To further contextualize the performance of CMYD-SurfaceNet, we compared the proposed method with several recent multimodal brain tumor segmentation approaches reported in the literature. nnU-Net [3] was included as a strong reference baseline because of its well-established robustness in biomedical image segmentation, while nnU-Net v2 [53] represents a more recent refinement of this framework with improved configuration strategies. Other approaches, including cascaded CNN models [24,25] and transformer-based architectures such as UNETR and Swin-UNETR [16,17], have also shown strong performance by enhancing network design, contextual modeling, and feature fusion. In addition, recent survey studies [54] emphasize the importance of robustness and generalization for clinical deployment. To position the proposed framework within this recent methodological landscape, a literature-based quantitative comparison is presented in Table 14.

Table 14. Comparison with recent state-of-the-art methods on the BraTS dataset (values for competing methods are reported from the literature).

Method	DSC (WT)	HD95 (mm)	Surface Dice @1 mm
nnU-Net [3]	0.913	3.62	0.821
nnU-Net v2 [53]	0.918	3.10	0.850
UNETR [16]	0.918	3.20	0.840
Swin-UNETR [17]	0.920	3.00	0.860
CMYD-SurfaceNet	0.921	2.41	0.892

The values reported for the competing methods were taken from the literature and may have been obtained under different experimental conditions, including variations in datasets, preprocessing procedures, validation strategies, and data splits. For this reason, the comparison should be interpreted as an indicative literature-based overview rather than a strictly controlled head-to-head benchmark.

Overall, the Dice score of CMYD-SurfaceNet remains close to those reported by recent segmentation methods, indicating that the proposed method does not mainly aim to achieve a large volumetric overlap gain. Within this literature-reported comparison, CMYD-SurfaceNet shows favorable boundary-sensitive values, particularly in HD95 and Surface Dice. However, these values should be interpreted as indicative rather than definitive evidence of superiority over other architectures, since the compared methods were not re-evaluated under a unified experimental protocol. Therefore, CMYD-SurfaceNet should be understood as a boundary-focused improvement over strong segmentation baselines rather than as a method that primarily increases Dice performance.

6.11. Qualitative Analysis

In addition to the quantitative evaluation, a qualitative analysis was performed to illustrate how the proposed framework behaves in representative segmentation cases. Three examples were selected to reflect different sources of difficulty: a small-lesion case, a complex-boundary case, and an artifact-affected case. The first case, taken from BraTS 2023, contains a small lesion with limited spatial extent. The second BraTS case presents a large heterogeneous lesion with poorly defined tumor boundaries. The third case, selected from the CHU Fez cohort, includes moderate motion artifacts and lower signal-to-noise ratio, reflecting the more variable acquisition conditions encountered in routine clinical practice.

Figure 2 presents the qualitative results for these three representative cases. For each case, the input MRI, ground truth annotation, baseline nnU-Net prediction, CMYD representation, and final CMYD-SurfaceNet prediction are shown. This visualization helps clarify the role of the proposed representation by showing the additional boundary-aware structural information provided by CMYD, while maintaining a clear relationship between the original MRI image, the reference annotation, and the final segmentation output.

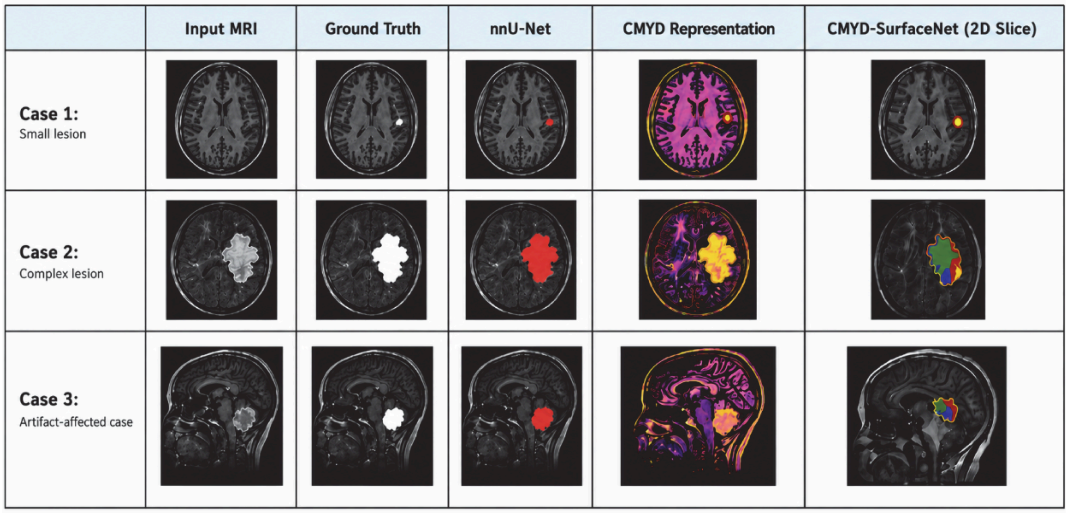
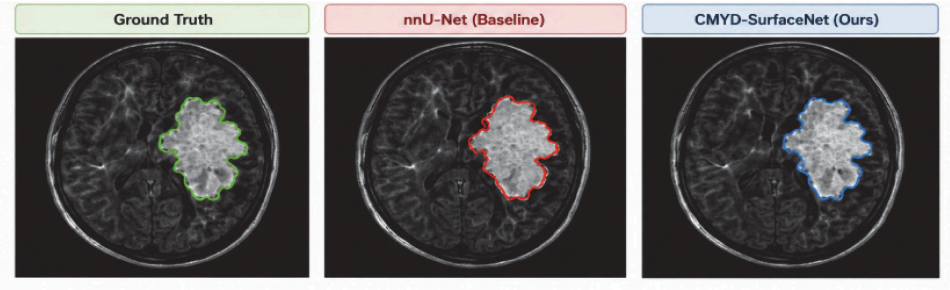
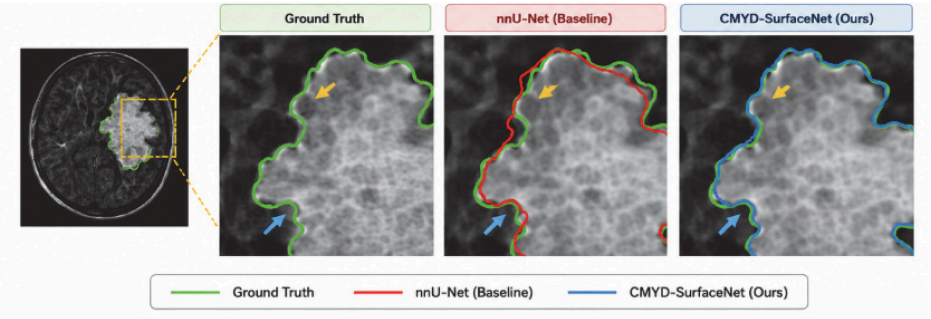


Figure 2. Qualitative segmentation results for three representative cases: a small-lesion case, a complex-boundary case, and an artifact-affected case. For each case, the figure shows the input MRI, the ground truth annotation, the baseline nnU-Net prediction, the CMYD representation, and the final CMYD-SurfaceNet prediction. Color coding is used to distinguish different segmentation outputs, where white represents ground truth annotations, red/yellow indicate predicted tumor regions, and grayscale corresponds to the input MRI.

Figure 3 provides a focused comparison between the baseline nnU-Net and the proposed CMYD-SurfaceNet. For this representative case, a 2D MRI slice showing the maximal tumor area is displayed, with the ground truth annotation, baseline nnU-Net prediction, and CMYD-SurfaceNet prediction shown using color-coded tumor contours overlaid on the grayscale MRI image. As shown in Figure 3, nnU-Net generally identifies the main tumor region, but local boundary deviations remain visible, especially in challenging areas. Compared with the baseline, CMYD-SurfaceNet provides more stable contour delineation and follows the reference annotation more closely, particularly around irregular tumor margins. These visual findings are consistent with the quantitative improvements observed in HD95 and Surface Dice, suggesting that the proposed framework mainly improves boundary regularity and structural consistency rather than producing a major change in volumetric overlap.



(**A**) Qualitative comparison



(**B**) Zoom-in boundary comparison

Figure 3. Qualitative boundary comparison between nnU-Net and CMYD-SurfaceNet. (**A**) Full 2D MRI slice comparison showing the ground truth, nnU-Net baseline prediction, and CMYD-SurfaceNet prediction. (**B**) Zoom-in boundary comparison highlighting local contour differences. Green, red, and blue contours indicate the ground truth, nnU-Net baseline, and CMYD-SurfaceNet, respectively. Arrows indicate regions of boundary discrepancy and zoomed-in areas for detailed visual comparison.

7. Discussion

The experimental results indicate that the proposed CMYD-SurfaceNet framework improves segmentation performance primarily through geometric refinement rather than volumetric expansion. While Dice similarity coefficients show consistent but moderate gains across tumor subregions, the most pronounced improvements were observed in boundary-sensitive metrics, particularly HD95 and Surface Dice. This pattern is important because the improvement in Dice is limited to approximately 0.9% for WT, whereas the HD95 reduction reaches approximately 33.4%, confirming that the proposed method mainly improves boundary precision rather than volumetric overlap. This distinction is essential, as Dice evaluates volumetric overlap, whereas HD95 directly quantifies surface deviation and is therefore more sensitive to contour irregularities. To quantify the magnitude of

geometric improvement, Table 15 reports the absolute and relative gains achieved by CMYD-SurfaceNet compared to the baseline nnU-Net.

Table 15. Relative geometric improvement of CMYD-SurfaceNet compared to baseline nnU-Net on BraTS 2023 validation set.

Metric	Region	Baseline	CMYD-SurfaceNet	Absolute Change	Relative Improvement (%)
HD95 (mm)	WT	3.62	2.41	−1.21	−33.4%
HD95 (mm)	TC	4.15	2.98	−1.17	−28.2%
HD95 (mm)	ET	3.98	2.74	−1.24	−31.2%
Surface Dice @1 mm	WT	0.821	0.892	+0.071	+8.6%
Dice (DSC)	WT	0.913	0.921	+0.008	+0.9%

Note: For HD95, lower values indicate better performance; therefore, relative values are reported as percentage reduction.

As shown in Table 15, HD95 is reduced by approximately 30% across tumor regions, confirming that the proposed framework improves boundary conformity. In contrast, Dice improvements remain moderate, indicating that volumetric agreement was already strong in the baseline while geometric deviations persisted at the tumor margins. These findings suggest that CMYD-SurfaceNet primarily enhances contour stability rather than increasing predicted tumor volume. The observed improvements were particularly pronounced in cases involving thin enhancing regions and complex tumor boundaries. In such scenarios, small geometric inconsistencies can lead to disproportionately large errors in surface-based metrics. The explicit boundary encoding introduced in the CMYD representation helps preserve these fine structures, resulting in more stable delineation of challenging tumor interfaces.

In addition to empirical improvements, the CMYD representation introduces an implicit inductive bias by structuring multimodal inputs into complementary subspaces. The separation of chromatic contrast (C, M, Y) from gradient-based structural information (D) reduces inter-channel redundancy and simplifies early feature extraction. As a result, convolutional filters can specialize in either contrast discrimination or boundary localization, rather than simultaneously resolving multiple entangled representations. This structured decomposition improves optimization stability and facilitates more efficient learning of geometrically consistent features. Unlike many segmentation approaches that rely on architectural expansion, the proposed framework preserves the backbone structure and parameter count of the baseline 3D nnU-Net. Performance gains therefore arise from restructuring the multimodal input representation rather than increasing model capacity. These representational differences explain why the proposed framework improves boundary-sensitive metrics without increasing model complexity.

The novelty of this work does not lie in introducing pseudo-color mapping, gradient-based edge encoding, or cascaded segmentation as isolated components. These concepts have been explored in different forms in previous medical image segmentation studies. Instead, the contribution of CMYD-SurfaceNet lies in integrating these components into a unified representation-level framework that restructures multimodal MRI inputs before feature extraction. This representation-level integration allows contrast-dominant information and boundary-related structural cues to be organized before learning, while preserving the same backbone complexity as nnU-Net.

Importantly, the proposed method does not introduce additional trainable parameters or architectural modifications. From a broader perspective, the proposed approach shifts the focus of segmentation research from model-centric improvements toward representation-centric design. While most existing methods aim to enhance feature extraction through architectural modifications, this work demonstrates that restructuring the input space itself can yield meaningful improvements in boundary-sensitive performance. Conventional concatenation requires early convolutional layers to disentangle luminance redundancy, modality-dominant contrast, and structural transitions simultaneously. In contrast, the CMYD representation separates chromatic contrast components (C, M, Y) from gradient-derived boundary information (D), thereby reducing inter-channel redundancy and conditioning the feature space in a task-oriented manner.

From a clinical perspective, accurate tumor boundary visualization is important for neuro-oncological workflows, including radiological review, preoperative assessment, radiotherapy planning, and surgical margin evaluation. In this study, the CHU Fez zero-shot evaluation was performed on retrospective clinical cases, and the resulting boundary-aware segmentation outputs were provided to the radiological and surgical teams as complementary preoperative decision-support material. The observed improvements in HD95 and Surface Dice suggest that CMYD-SurfaceNet may offer useful boundary-aware segmentation support under local clinical acquisition conditions. However, these outputs should be interpreted as expert-reviewed support information rather than as standalone surgical guidance or evidence of full prospective clinical deployment. Larger multi-center studies and formal clinical workflow evaluations remain necessary before integration into routine clinical practice.

The evaluation on the CHU Fez clinical dataset provides preliminary evidence of robustness under heterogeneous local acquisition conditions. Domain shifts between benchmark datasets and routine clinical imaging can degrade segmentation performance due to scanner variability and intensity distribution differences. By separating luminance structure from contrast-dominant channels and explicitly encoding boundary transitions, the CMYD representation may reduce sensitivity to global intensity variations. Nevertheless, because the CHU Fez evaluation was conducted on a limited retrospective cohort, these findings should be interpreted as supportive but not definitive evidence of external generalization.

From a computational perspective, the backbone architecture and parameter count remain unchanged, approximately 31.2 million parameters. The pseudo-RGB alignment, CMYK decomposition, and boundary density computation are deterministic preprocessing operations introducing negligible overhead relative to convolutional inference. Although the cascaded refinement framework moderately increases inference time due to sequential region-of-interest processing, memory consumption remains comparable to the baseline model.

Several limitations should be acknowledged. First, the CMYD transformation is deterministic and not learnable. While this improves interpretability and preserves model complexity, it may limit adaptability to dataset-specific characteristics. Second, although the CHU Fez zero-shot evaluation included all 86 available retrospective cases, the cohort remains single-center and should therefore be validated further on larger multi-center external datasets. The reported confidence intervals should be interpreted as an estimate of variability within this retrospective clinical cohort.

Third, although HSV and LAB comparisons provide additional support for the proposed representation, broader validation across other anatomical regions and imaging modalities remains necessary. Finally, the cascaded refinement strategy improves boundary precision but may introduce additional inference time compared with a purely single-stage model.

Overall, the findings demonstrate that representation-level structural decoupling can enhance contour fidelity without increasing model complexity. This suggests that careful restructuring of multimodal information prior to learning constitutes a complementary strategy to architectural scaling for improving clinically relevant segmentation performance.

8. Conclusions

This work introduced CMYD-SurfaceNet, a framework that improves multimodal MRI brain tumor segmentation by rethinking how input data is structured rather than modifying the network architecture itself. By explicitly separating contrast-related information from boundary-driven structural cues, the proposed representation allows the model to better capture both tumor appearance and geometric characteristics. The experimental evaluation on the BraTS 2023 benchmark and the CHU Fez clinical dataset shows consistent improvements, particularly in boundary-related metrics. While volumetric overlap remains comparable to strong baseline models, the proposed approach produces more stable and precise contour delineation, which is essential for clinically meaningful segmentation. These results highlight that the way multimodal data is organized before learning can meaningfully influence segmentation performance. By reducing redundancy and introducing explicit structural information, the proposed representation simplifies the learning process and improves geometric consistency without increasing model complexity. Future work will focus on extending this representation to other imaging modalities and anatomical regions, as well as exploring adaptive or learnable variants of the transformation. Overall, this study highlights that representation-level design can serve as a complementary paradigm to architectural innovation in advancing medical image segmentation.

Author Contributions: Conceptualization, C.E.M. and M.M.; methodology, C.E.M. and M.M.; software, C.E.M.; validation, M.M., N.E.A.E.I. and F.Z.A.; formal analysis, C.E.M. and M.M.; data curation, C.E.M.; writing—original draft preparation, C.E.M.; writing—review and editing, M.M., N.E.A.E.I., L.M. and F.Z.A.; visualization, M.M.; supervision, M.M. and N.E.A.E.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the use of publicly available and anonymized datasets.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are publicly available in the BraTS dataset and a locally curated clinical dataset (CHU Fez), subject to institutional data sharing agreements. The code can be made available upon reasonable request.

Acknowledgments: The authors would like to sincerely thank CHU Fez (University Hospital Center of Fez) for providing the clinical dataset used in this study. The authors also acknowledge Fatima Zahra Ammor for her valuable contribution to data acquisition and clinical supervision.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. de Verdier, M.C.; Saluja, R.; Gagnon, L.; LaBella, D.; Baid, U.; Tahon, N.H.; Foltyn-Dumitru, M.; Zhang, J.; Alafif, M.; Baig, S.; et al. The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-Treatment MRI. *arXiv* **2024**, arXiv:2405.18368.
2. D'aNtonoli, T.A.; Berger, L.K.; Indrakanti, A.K.; Vishwanathan, N.; Weiss, J.; Jung, M.; Berkarda, Z.; Rau, A.; Reisert, M.; Küstner, T.; et al. TotalSegmentator MRI: Robust Sequence-Independent Segmentation of Multiple Anatomic Structures in MRI. *arXiv* **2024**, arXiv:2405.19492.

3. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat. Methods* **2021**, *18*, 203–211. [CrossRef] [PubMed]
4. Esmeda, L.A. AI-Driven Brain Tumor Segmentation: Review of the Last Decade. *J. Technol. Res.* **2025**, *51*, 545–557. [CrossRef]
5. Kervadec, H.; Bouchtiba, J.; Desrosiers, C.; Granger, E.; Dolz, J.; Ben Ayed, I. Boundary Loss for Highly Unbalanced Segmentation. *arXiv* **2019**, arXiv:1812.07032.
6. Karimi, D.; Salcudean, S.E. Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks. *IEEE Trans. Med. Imaging* **2020**, *39*, 499–513. [CrossRef] [PubMed]
7. Magadza, T.; Viriri, S. Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art. *J. Imaging* **2021**, *7*, 19. [CrossRef] [PubMed]
8. El Mechal, C.; El Idrissi, N.E.A. Two-Stream nnU-Net: A Novel Architecture for Precise Tumor Segmentation in Medical Imaging. In Proceedings of the 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM), Istanbul, Türkiye, 26–28 October 2023; pp. 1–6.
9. Diana-Albelda, C.; García-Martín, Á.; Bescos, J. A Review on Deep Learning Methods for Glioma Segmentation, Limitations, and Future Perspectives. *J. Imaging* **2025**, *11*, 269. [CrossRef] [PubMed]
10. Tiwari, A.K.; Shukla, N. Brain Tumor Segmentation Using CNN. In *Recent Trends in Communication and Electronics*; CRC Press: Boca Raton, FL, USA, 2021; pp. 411–415.
11. Reinke, A.; Eisenmann, M.; Tizabi, M.D.; Sudre, C.H.; Rädsch, T.; Antonelli, M.; Arbel, T.; Bakas, S.; Cardoso, M.J.; Cheplygina, V.; et al. Common Limitations of Image Processing Metrics: A Picture Story. *arXiv* **2021**, arXiv:2104.05642.
12. Jha, D.; Smedsrud, P.H.; Riegler, M.A.; Halvorsen, P.; Johansen, D.; Johansen, H.D.; de Lange, T. ResUNet++: An Advanced Architecture for Medical Image Segmentation. In Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM), San Diego, CA, USA, 9–11 December 2019; pp. 225–230.
13. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016)*; Springer: Cham, Switzerland, 2016; pp. 424–432. [CrossRef]
14. Allaoui, M.L.; Allili, M.S.; Belaid, A. HA-U^{3}Net: A Modality-Agnostic Framework for 3D Medical Image Segmentation Using Nested V-Net Structure and Hybrid Attention. *Knowl. Based Syst.* **2025**, *327*, 114127.
15. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [CrossRef] [PubMed]
16. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H.R.; Xu, D. UNETR: Transformers for 3D Medical Image Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2022)*; IEEE: Waikoloa, HI, USA, 2022; pp. 574–584. [CrossRef]
17. 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. *arXiv* **2022**, arXiv:2201.01266.
18. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.M.; Larochelle, H. Brain Tumor Segmentation with Deep Neural Networks. *Med. Image Anal.* **2017**, *35*, 18–31. [CrossRef] [PubMed]
19. Chen, C.; Dou, Q.; Jin, Y.; Chen, H.; Heng, P.A. Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Fusion. *Med. Image Anal.* **2019**, *54*, 179–190. [CrossRef]
20. Valindria, V.V.; Pawlowski, N.; Rajchl, M.; Lavdas, I.; Aboagye, E.O.; Rockall, A.; Rueckert, D.; Glocker, B. Multi-Modal Learning from Unpaired Images: Application to Multi-Organ Segmentation in CT and MRI. *IEEE Trans. Med. Imaging* **2018**, *37*, 1049–1060.
21. Kervadec, H.; Dolz, J.; Tang, M.; Granger, E.; Boykov, Y.; Ben Ayed, I. Constrained-CNN Losses for Weakly Supervised Segmentation. *Med. Image Anal.* **2019**, *54*, 88–99. [CrossRef] [PubMed]
22. Mostafa, A.M.; Zakariah, M.; Aldakheel, E.A. Brain tumor segmentation using deep learning on MRI images. *Diagnostics* **2023**, *13*, 1562. [CrossRef] [PubMed]
23. Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.
24. Wang, G.; Li, W.; Ourselin, S.; Vercauteren, T. Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017)*; Springer: Cham, Switzerland, 2017; pp. 178–186. [CrossRef]
25. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. In Proceedings of the 24th International Conference, Strasbourg, France, 27 September–1 October 2021; pp. 36–46.
26. Huang, R.; Li, Y.; Wang, S.; Zhang, Y. Pseudo-Color Image Representation for Brain Tumor Detection. *IEEE Access* **2019**, *7*, 124789–124799. [CrossRef]

27. Tehsin, S.; Nasir, I.M.; Damaševičius, R. NeuroMamba: Efficient State-Space Model for Explainable Brain Tumor Segmentation. In Proceedings of the 2025 International Conference on Computational Intelligence, Security, and Artificial Intelligence (IntelliSecAI), Al Khobar, Saudi Arabia, 17–18 December 2025. [CrossRef]
28. Yu, X.; Liang, M.; Zeng, Q.; Ren, Z.; Kang, C.; Jiang, C.; Chen, G.; Liang, B.; Pu, B.; Zhu, N.; et al. ESE-UNet: An Extremely Simple yet Efficient UNet for Medical Image Segmentation. *Biomed. Signal Process. Control* **2026**, *112*, 108634. [CrossRef]
29. Wu, W.X.; Yan, J.; Zhao, Y.S.; Sun, Q.C.; Zhang, H.L.; Zheng, J.L.; Liang, D.; Chen, Y.S.; Zhang, Z.Y.; Li, Z.C. Multi-task Learning for Concurrent Survival Prediction and Semi-supervised Segmentation of Gliomas in Brain MRI. *Displays* **2023**, *78*, 102402. [CrossRef]
30. Farooq, M.A.; Jan, B.; Farman, H. Learning Deep Feature Representations for Multi-Modal MR Brain Tumor Segmentation. *Neurocomputing* **2025**, *630*, 130162.
31. Praveen, R.C.; Rengen, P.K.; Francis, D. Hybrid fractal ConvXGBNet for brain tumor classification with log-cosh dice loss—Cascaded V-Net based segmentation based on MRI images. *Eur. Spine J.* **2026**. [CrossRef] [PubMed]
32. Fidon, L.; Bulo, S.R.; Valsesia, D.; Bousmalis, K.; Criminisi, A. Generalised Wasserstein Dice Loss for Imbalanced Multi-Class Segmentation Using Holistic Convolutional Networks. *Med. Image Anal.* **2019**, *54*, 28–39.
33. Almansour, A.G.M.; Alshomrani, F. Dynamic expert routing with boundary-aware decoding for accurate brain tumor segmentation from incomplete MRI data. *Sci. Rep.* **2026**, *16*, 48187. [CrossRef] [PubMed]
34. Vamsikrishna, M.; Shieh, C.-S. Cascade CNN: A Two-Stage Segmentation Framework for Efficient and Accurate Brain Tumor Segmentation in Multi-Modal MRI. *Future Technol.* **2025**, *4*, 104–118. [CrossRef]
35. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. Self-Supervised Learning for Medical Image Analysis Using Image Context Restoration. *Med. Image Anal.* **2019**, *58*, 101539. [CrossRef] [PubMed]
36. Kumar, L.; Singh, A.; Sharma, S.; Gupta, R. Hyper-Connections for Adaptive Multi-Modal MRI Brain Tumor Segmentation. *arXiv* **2026**, arXiv:2603.19844.
37. Cai, H.; Wang, L.; Yang, J.; Chen, X.; Li, J. Multi-Stage Cascaded U-Net for Brain Tumor Segmentation. *IEEE Access* **2020**, *8*, 135487–135495. [CrossRef]
38. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Takeshi Shinohara, R.; Berger, C.; Min Ha, S.; Rozycki, M.; et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *Med. Phys.* **2019**, *46*, 2435–2453. [CrossRef]
39. Isensee, F.; Kickingereder, P.; Wick, W.; Bendszus, M.; Maier-Hein, K.H. No new-net. In *Proceedings of the BrainLes Workshop, MICCAI 2018*; Springer: Cham, Switzerland, 2018; pp. 234–244.
40. Baid, U.; Ghodasara, S.; Mohan, S.; Bilello, M.; Calabrese, E.; Colak, E.; Farahani, K.; Kalpathy-Cramer, J.; Kitamura, F.C.; Pati, S.; et al. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. *Med. Phys.* **2021**, *48*, 7437–7449.
41. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2022**, *9*, 23. [CrossRef]
42. Tang, Y.; Yang, D.; Li, W.; Roth, H.R.; Landman, B.; Xu, D.; Nath, V.; Hatamizadeh, A. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 20730–20740.
43. Myronenko, A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In *Transactions on Petri Nets and Other Models of Concurrency XV*; Springer: Cham, Switzerland, 2018; pp. 311–320.
44. El Mechal, C.; El Idrissi, N.E.A.; Mazgouti, L.; Ammor, F.Z. TransUNet++: A Multiscale CNN-Transformer Hybrid Framework for Multi-Modal Brain Tumor Segmentation. In *Proceedings of the 12th International Conference on Wireless Networks and Mobile Communications (WINCOM 2025)*; IEEE: New York, NY, USA, 2025; pp. 1–6.
45. Li, Z.; Yang, X.; Li, J.; Chen, C.; Wang, X.; Zhao, Q.; Cao, S. Accurate Medical Image Segmentation Based on Adaptive Fusion Hierarchical Network with Multi-Scale and Attention Mechanisms. SSRN. 2026. Available online: https://ssrn.com/abstract=5195544 (accessed on 10 June 2026).
46. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]
47. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent Advances in Convolutional Neural Networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
48. Shurrab, S.; Duwairi, R. Self-Supervised Learning Methods and Applications in Medical Imaging Analysis: A Survey. *PeerJ Comput. Sci.* **2022**, *8*, e1045. [CrossRef] [PubMed]
49. Zhang, H.; Zuo, X.; Jiang, J.; Guo, C.; Ma, J. MRFS: Mutually Reinforcing Image Fusion and Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2024; pp. 26974–26983.
50. Karimi, D.; Dou, H.; Warfield, S.K.; Gholipour, A. Deep Learning with Noisy Labels for Medical Image Segmentation. *Med. Image Anal.* **2022**, *80*, 102487.

51. Cai, H.; Qi, L.; Yu, Q.; Shi, Y.; Gao, Y. 3D Medical Image Segmentation with Sparse Annotation via Cross-Teaching between 3D and 2D Networks. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; Springer Nature: Cham, Switzerland, 2023; pp. 614–624.
52. Borse, S.; Wang, Y.; Zhang, Y.; Porikli, F. InverseForm: A Loss Function for Structured Boundary-Aware Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2021; pp. 5901–5911.
53. Isensee, F.; Wald, T.; Ulrich, C.; Baumgartner, M.; Roy, S.; Maier-Hein, K.H.; Jaeger, P.F. nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. *arXiv* **2024**, arXiv:2404.09556.
54. Hamidi, H.; Morovati, M.K.; Zandieh, M. Evaluation of CNN-Based Deep Learning Models for Auto-Contouring in Glioblastoma Radiotherapy: A Review. *Radiat. Oncol.* **2025**, *20*, 169. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.