



## From Semantic Drift to Objective Miscalibration: Domain Incremental Brain Tumor Segmentation with Missing Modality

Junze Wang<sup>a,b</sup>, Lei Fan<sup>c</sup>, Dezheng Zhang<sup>a,b</sup>, Donglin Di<sup>d</sup>, Yang Song<sup>e</sup>, Sidong Liu<sup>f</sup>, Cong Cong<sup>f,\*</sup>

<sup>a</sup>School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China

<sup>b</sup>Beijing Key Laboratory of Integrated Traditional Chinese and Western Medicine for Organ Function Protection in Cancer, Beijing, China

<sup>c</sup>Centre for Healthy Brain Ageing, Discipline of Psychiatry and Mental Health, School of Clinical Medicine, Faculty of Medicine and Health, UNSW Sydney

<sup>d</sup>School of Software, Tsinghua University, Beijing, China

<sup>e</sup>School of Computer Science and Engineering, University of New South Wales, Kensington, NSW 2033, Australia

<sup>f</sup>Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2113, Australia

### ARTICLE INFO

#### Article history:

#### Keywords:

Brain tumor segmentation

Missing modality

Domain incremental learning

Hypergraph

### ABSTRACT

Multimodal MRI brain tumor segmentation methods are often developed under the assumption that all modalities are available, yet in real clinical practice one or more modalities are frequently missing due to acquisition order or incomplete scanning protocols, leading to degraded segmentation performance. In this work, we propose a novel training scheme PMD for missing MRI modality segmentation based on Domain Incremental Learning (DIL), where each modality is treated as an independent domain. We identify two key challenges in this setting, *i.e.*, **Semantic Drift (SD)** and **Objective Miscalibration (OM)**. **SD** arises when stage-wise representation shifts make previously learned semantic references progressively unreliable, yielding order-sensitivity under different incremental orders. While **OM** occurs when fixed loss weighting fails to account for modality dependent error characteristics, resulting in unstable optimization and degraded calibration. To address **SD**, we introduce a prototype Tversky-Aware contrastive (pTAC) loss that stabilizes tumor subregion features across stages by providing consistent global semantic supervision, together with a Prototype Anchored Representation Memory (PRM) that stores class specific prototypes as cross stage semantic anchors and filters by cross-patient hypergraph aware gate. To address **OM**, we develop a Meta Controller that adaptively reweights multiple training objectives, including overlap supervision, less-represented tumor classes, and cross modality semantic alignment, throughout incremental learning. Extensive experiments on BraTS 2019, FeTS 2022, and MU Glioma Post demonstrate that the proposed method consistently outperforms state of the art approaches, achieving over 2% improvement in Dice Scores, while also improving generalization across all tumor subregions.

© 2026 Elsevier B. V. All rights reserved.

### 1. Introduction

Accurate multimodal brain tumor segmentation using MRI enables computer-aided systems to quantify tumor burden and

delineate clinically significant subregions for assessment and treatment planning (Rees, 2003; Weller et al., 2021). To achieve this, segmentation methods typically integrate multiple MRI modalities, including T1-weighted, contrast-enhanced T1 (T1CE), T2-weighted, and FLAIR imaging. Each modality provides complementary diagnostic information that supports the

\*Corresponding author.

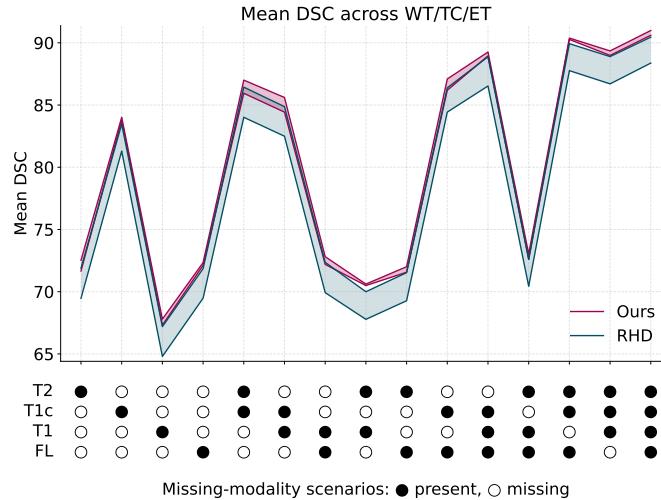


Fig. 1: Order sensitivity in DIL-based missing-modality segmentation. We train each model with identical settings, differing only in the modality arrival order. Each model is evaluated on 15 missing-modality scenarios using the mean Dice averaged over WT/TC/ET. To isolate order-induced variability, we plot an envelope across the three orders: for each scenario, the shaded band spans the minimum to maximum mean Dice across orders. Previous state-of-the-art method, RHD (Wang et al., 2025), exhibits a visibly wider band, indicating stronger dependence on modality order. In contrast, our method produces a consistently narrower band, showing more stable and superior performance under different arrival orders.

identification of clinically relevant tumor subregions, such as the whole tumor (WT), tumor core (TC), and enhancing tumor (ET) (Sawlani et al., 2020). However, most existing segmentation methods implicitly assume the availability of all modalities at inference time (Jing et al., 2025; Tang et al., 2025). This assumption is frequently violated in real-world clinical practice: despite standardized recommendations for multimodal MRI acquisition (Wen et al., 2010), practical constraints such as acquisition order, and incomplete scanning protocols often result in missing modalities (Zhong et al., 2025). As a consequence, segmentation performance can degrade substantially, limiting the robustness and clinical applicability of current methods.

Recently, the missing-modality brain tumor segmentation methods have been proposed (Havaei et al., 2016; Zhang et al., 2022; Liu et al., 2023; Shi et al., 2023). However, these approaches are typically trained under the assumption that all imaging modalities are jointly available during training, which does not fully reflect real clinical deployment. In practice, MRI modalities are often acquired sequentially over time, and adapting a segmentation model to newly available modalities commonly requires retraining from scratch or extensive fine-tuning. Such updates are inefficient and may cause the model to overfit to the most recent modality configuration while degrading previously learned knowledge (Zhou et al., 2021). Motivated by this, we formulate missing modality segmentation from a Domain Incremental Learning (DIL) perspective (Van de Ven et al., 2022; Yuan and Zhao, 2024), where training proceeds stage by stage with newly arrived modality.

Applying DIL to missing-modality segmentation poses several unique challenges. First, DIL is highly sensitive to the order in which modalities are introduced. As shown in Fig. 1,

simply altering the modality arrival order results in performance differences. We term this phenomenon as **Semantic Drift (SD)**, whereby different arrival orders induce inconsistent representation shifts and ultimately lead to divergent performance. Second, DIL is inherently susceptible to catastrophic forgetting, a limitation that is further exacerbated in MRI segmentation due to the small size of tumors relative to the background. Recent work (Wang et al., 2025) attempts to mitigate this issue by introducing a Tversky-enhanced three-way loss formulation, *i.e.*, segmentation-focused, region-imbalance-focused, and cross-modality-focused losses, where the Tversky similarity alleviates tumor–background imbalance (Tversky, 1977). Nonetheless, the use of fixed loss weights across training stages introduces additional limitations. In practice, the relative importance of these objectives evolves as new modalities are introduced, *e.g.*, cross-modality-focused losses should receive increasing emphasis as more modalities become available, while region-imbalance-focused losses should be adaptively adjusted to reflect the imbalance characteristics of the current modality. Fixed loss weighting fails to capture these modality-dependent requirements, leading to a misalignment between the optimization objective and the evolving learning dynamics. We refer to this mismatch as **Objective Miscalibration (OM)**.

Accordingly, we propose PMD, a Prototype-guided Meta DIL framework for missing-modality segmentation. PMD trains the model modality by modality and maintains a replay buffer to preserve previously learned knowledge. To address **SD**, we propose a prototype-based Tversky-Aware Contrastive (pTAC) loss for cross-modality alignment via semantic anchors. These anchors are implemented as class-wise global prototypes (one per tumor subregion) in the feature space, providing consistent tumor subregion representations under different incremental orders. To maintain reliable anchors, we further propose a Prototype Anchored Representation Memory (PRM), which selects representative patient semantics via a cross-patient hypergraph gate and updates prototypes with a weighted EMA to suppress outliers.

To address **OM**, PMD introduces a Meta Controller that adaptively regulates loss weights to keep the training objective calibrated under modality changes. Specifically, we optimize the model with a weighted sum of three losses: Dice–Tversky and Focal–Tversky for overlap supervision and rare-structure emphasis, and the proposed pTAC loss for cross-modality alignment. By reweighting these objectives online, the Meta Controller prevents any single objective from being persistently over- or under-emphasized under modality shifts, thereby keeping the optimization objective calibrated during incremental training. Furthermore, we conduct a comprehensive evaluation under diverse missing-modality scenarios at inference, including both in-distribution and out-of-distribution settings across multiple datasets, and explicitly quantify sensitivity to different incremental orders. The contributions are summarized:

- To the best of our knowledge, PMD is the first work to explicitly mitigate sensitivity on modality arrival orders, and achieve superior segmentation performance across diverse missing-modality scenarios.

- We introduce a prototype Tversky-Aware Contrastive (pTAC) loss with Prototype Anchored Representation Memory (PRM). The pTAC loss uses global prototypes to align features of tumor subregion, effectively reducing **SD**.
- We propose a Meta Controller that learns stage-adaptive loss reweighting to match modality-dependent optimization needs, effectively addressing **OM** and improving segmentation performance.
- Extensive experiments on BraTS 2019, FeTS 2022, and MU-Glioma-Post demonstrate superior performance over prior state-of-the-art methods under missing-modality settings, with consistent robustness across different modality arrival orders.

This paper extends our prior MICCAI work (Wang et al., 2025) with three improvements. First, we propose a prototype-based Tversky-Aware Contrastive (pTAC) loss using class-wise global prototypes as semantic anchors to reduce **SD**, and enhance prototype reliability via cross-patient high-order relationships and EMA updates. Then, we introduce an online Meta Controller that adaptively reweights multiple losses across stages to mitigate **OM**. Furthermore, we expand the evaluation to quantify order sensitivity and generalization under missing-modality and out-of-distribution settings. The code available at <https://github.com/reeive/PMD>.

## 2. Related Work

### 2.1. Missing Modality Brain Tumor Segmentation

Multimodal MRI brain tumor segmentation has advanced with deep learning, but performance often degrades under missing modality. Prior work mainly follows three directions. First, feature fusion methods aggregate the available modalities into a unified representation. HeMIS (Havaei et al., 2016) summarizes modality-specific embeddings by statistics, while mmFormer (Zhang et al., 2022) and MFI (Zhao et al., 2022) further align and reweight modality features via attention and adaptive interactions. Recent variants reduce modality bias with factorized or disentangled representations, *i.e.*, KMD (Liu et al., 2025), DC-Seg (Li et al., 2025); and IM-Fuse (Pipoli et al., 2025) improves efficiency under sparse modalities. However, fusion will be unstable when the missing modality shifts. Then, distillation-based methods transfer knowledge from full modality to missing modality. M3AE (Liu et al., 2023) pretrains a multimodal masked autoencoder and applies self-distillation for segmentation, while PASSION (Shi et al., 2024) explicitly handles imbalanced missing rates with distillation; MST-KDNet (Zhu et al., 2025) further distills multi-scale features and matches global style to narrow the complete–incomplete gap. However, distillation assumes a reliable teacher learned supervision from complete modalities, which may be mismatched in realistic missing modality streams. Finally, masking reconstruction strategies treat missing modalities as masked data and suppress the absent information. A2FSeg (Wang and Hong, 2023) gates missing channels, and M<sup>3</sup>FeCon (Zeng et al., 2024) and M2FTrans (Shi et al., 2023) reconstruct missing

modality features via masked attention and learnable tokens. However, their robustness is tightly coupled with reconstruction fidelity and the priors encoded in the masking mechanism, which may not hold under severe modality loss.

Despite this progress, most methods assume to train jointly with full-modality, necessitating retraining for new modalities. This conflicts with the sequential nature of clinical acquisition. In contrast, we adopt a DIL framework to mirror this reality, enabling efficient, incremental adaptation to arriving modalities without retraining from scratch.

### 2.2. Domain Incremental Learning

Domain Incremental learning (DIL), also known as continual or lifelong learning, enables models to learn from non-stationary data streams without forgetting previous knowledge. It is typically categorized into three scenarios: task-incremental learning addresses a sequence of distinct tasks, while class-incremental learning recognizes an expanding set of classes within a unified task (Van de Ven et al., 2022; Yuan and Zhao, 2024). DIL lies between these extremes: the model faces the same overall task in each stage, but the input domain or data distribution shifts at each incremental step (Van de Ven et al., 2022). Notably, the domain-incremental scenario is highly pertinent to medical imaging applications, which are frequently characterized by distribution shifts and data heterogeneity across clinical sites, imaging devices, and patient populations (Varoquaux and Cheplygina, 2022; Pati et al., 2022). Consequently, incremental learning offers a framework to gradually adapt medical image analysis models to new domains as data becomes available, thereby avoiding the high cost of retraining from scratch on all data (Ye et al., 2024; Kumari et al., 2025). For instance, in brain tumor segmentation, algorithms are typically developed on complete multi-modal MRI datasets like BraTS (Bakas et al., 2018). However, clinical practice often suffers from missing modalities due to sequential acquisition. ReHyDIL (Wang et al., 2025) employs a DIL pipeline that incorporates CHSNet to capture high-order patient correlations and utilizes a TAC loss to alleviate both inter- and intra-modality class imbalance.

However, applying DIL to missing modalities introduces a unique challenge: sensitivity to modality arrival order. Unlike joint training scenarios, altering the modality sequence induces varying degrees of **SD**, leading to inconsistent outcomes. We specifically address this order sensitivity by employing stable semantic anchors and dynamic calibration to ensure robust performance across different incremental sequences.

## 3. Methodology

### 3.1. Overview

The Prototype-guided Meta DIL framework (PMD) formulates missing-modality brain tumor segmentation as a Domain Incremental Learning (DIL) problem, in which MRI modalities are introduced sequentially during training. In this formulation, PMD incrementally updates the CHSNet (Wang et al., 2025)  $F_\theta$  to integrate the newly available modalities while retaining the knowledge acquired from earlier stages, enabling direct inference in various missing modality scenarios after training.

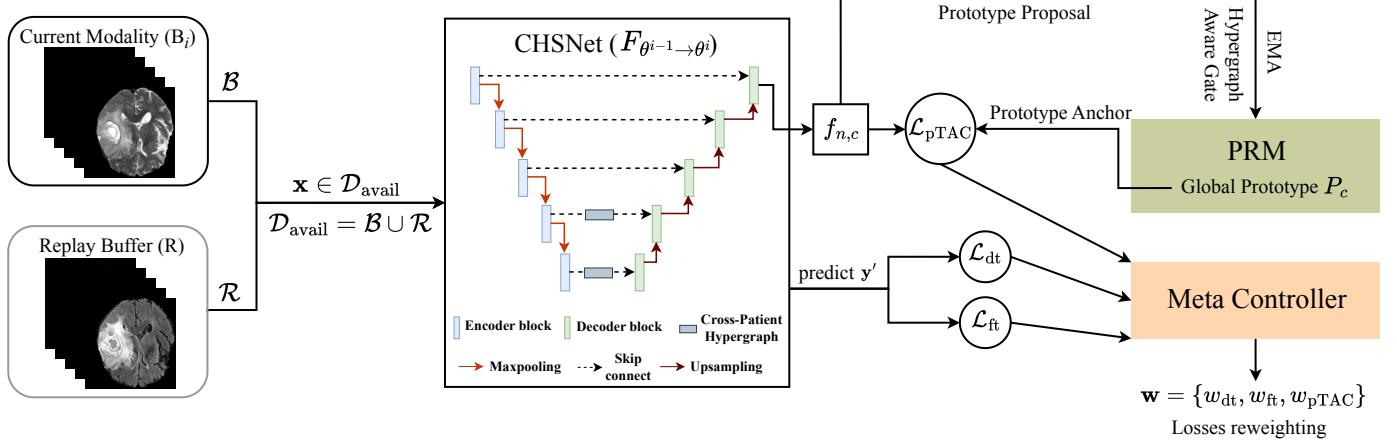


Fig. 2: Overview of stage  $i$  training in PMD. At each stage, CHSNet  $F_{\theta^{(i)}}$  is optimized using  $\mathcal{D}_{\text{avail}}$ . From the final decoder feature map, we extract region-level features  $f_{n,c}$  for each patient  $n$  and tumor subregion  $c$  as prototype proposals. These proposals are passed through a cross-patient hypergraph-aware gate in the Prototype Anchored Representation Memory (PRM), which filters out outliers and updates global class prototypes  $P_c$  via weighted EMA. The resulting prototypes serve as semantic anchors for the prototype-based Tversky-Aware Contrastive loss  $\mathcal{L}_{pTAC}$ , which aligns region features across modalities to mitigate **SD**. Finally, a Meta Controller dynamically reweights  $\mathcal{L}_{dt}$ ,  $\mathcal{L}_{ft}$ , and  $\mathcal{L}_{pTAC}$  through bi-level optimization, addressing **OM** by adapting the training objective to stage-specific modality characteristics.

### 3.2. Problem Formulation

Let  $I = \{\text{T1, T1CE, T2, FLAIR}\}$  denote the modality set and  $\{\mathcal{B}_i\}_{i=1}^I$  denote the stage-wise data stream, where  $\mathcal{B}_i$  contains samples from the modality introduced at stage  $i$ . Each input is an image-mask pair  $(\mathbf{x}, \mathbf{y})$ , where  $\mathbf{y} = \{\mathbf{y}_c\}_{c \in C}$  and  $C = \{\text{WT, TC, ET}\}$ . To enable the cross-patient hypergraph construction in CHSNet, we organize each mini-batch at the patient level, *i.e.*, each input  $(\mathbf{x}, \mathbf{y})$  corresponds to a distinct patient  $n$ . CHSNet outputs prediction  $\mathbf{y}' = F_{\theta}(\mathbf{x})$ , where  $\mathbf{y}' = \{\mathbf{y}'_c\}_{c \in C}$ . Training follows a domain incremental learning paradigm, where modalities are introduced sequentially across stages. At the first stage, the model  $F_{\theta^1}$  is trained using samples from the initial modality, and a replay buffer  $R$  is initialized to mitigate catastrophic forgetting. Following (Wang et al., 2025),  $R$  retains 10% of samples from each stage. At subsequent stages, the model is initialized with the parameters from the previous stage and incrementally updated using both newly introduced modality data and samples from the replay buffer.

### 3.3. Cross-Modality Incremental Training

At stage  $i$ , we jointly optimize  $F_{\theta^{(i)}}$  using an input batch  $\mathcal{D}_{\text{avail}} = \mathcal{B} \cup \mathcal{R}$ , where  $\mathcal{B} \in \mathcal{B}_i$  is a current-modality mini-batch and  $\mathcal{R} \in R$  is a replay mini-batch sampled from the buffer. As illustrated in Fig. 2, we forward  $\mathcal{D}_{\text{avail}}$  through the network and extract tumor subregion features  $f_{n,c}$  as prototype proposals. These proposals are selectively written into a Prototype Anchored Representation Memory (PRM; Sec. 3.3.1) via a cross-patient hypergraph-aware gate, yielding class-wise global prototypes that act as stable semantic anchors. We then compute a prototype-based Tversky-Aware Contrastive loss  $\mathcal{L}_{pTAC}$  to align tumor subregion features across stages and mitigate **Semantic Drift (SD)** (Sec. 3.3.2). Finally, we optimize the network with segmentation supervision and cross-stage alignment, where the relative weights of objectives are adaptively recalibrated to address **Objective Miscalibration (OM)** (Sec. 3.4).

#### 3.3.1. Prototype Anchored Representation Memory (PRM)

PRM maintains class-wise global prototypes  $P_c$  as cross-stage semantic anchors. To keep these anchors reliable across stages, we avoid naively updating each prototype. In contrast, we only update globally representative and reliable proposals to prevent stage-shifted or outlier features from corrupting the anchors. At each iteration, PRM forms global prototype proposals from  $\mathcal{D}_{\text{avail}}$ , then selects globally representative prototype proposals through a cross-patient hypergraph gate, and performs a weighted EMA update.

**Global Prototype Proposal.** Given the last decoder feature map  $\mathbf{z}_n \in \mathbb{R}^{D \times H \times W}$  for patient  $n$ , where  $D$  denotes the channel dimension and  $H \times W$  denotes the image resolution. We extract the tumor subregion feature  $f_{n,c}$  for class  $c$  using masked average pooling:

$$f_{n,c} = \frac{\sum_{i=1}^H \sum_{j=1}^W \mathbf{z}_n(i, j) \mathbf{y}_{n,c}(i, j)}{\sum_{i=1}^H \sum_{j=1}^W \mathbf{y}_{n,c}(i, j)}. \quad (1)$$

If  $\sum_{i,j} \mathbf{y}_{n,c}(i, j) = 0$ , we ignore it, since subregion  $c$  is absent and the masked pooling is undefined; thus no proposal is generated for  $\{n, c\}$ . Then,  $f_{n,c} \in \mathbb{R}^D$  serves as a prototype proposal  $\tilde{P}_{n,c}$ .

**Hypergraph-Aware Gate.** To ensure that prototype updates reflect globally consistent semantics, we select representative prototype proposals using a cross-patient hypergraph-aware gate. As shown in Fig. 3, we extract patient embeddings from fused 4th and 5th encoder features and apply  $\ell_2$ -normalization. We treat each patient embedding as a vertex and construct a cross-patient hypergraph over these nodes. Following the adaptive hypergraph construction (Wang et al., 2025), we build an incidence matrix  $\mathbf{H} \in \mathbb{R}^{N \times E}$  and hyperedge weight matrix  $\mathbf{W} \in \mathbb{R}^{E \times E}$ ; we set  $\mathbf{W} = \mathbf{I}_E$ , *i.e.*, all hyperedges are equally weighted. Then we inject higher-order relationships into the patient topology (Feng et al., 2019):

$$\mathbf{G} = \mathbf{D}_V^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^\top \mathbf{D}_V^{-\frac{1}{2}}, \quad (2)$$

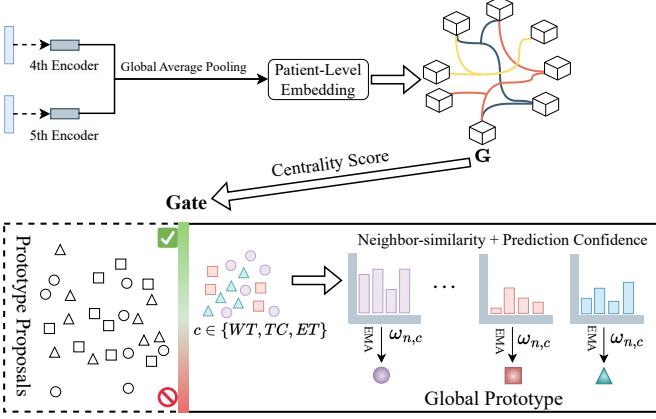


Fig. 3: PRM maintains class-wise global prototypes  $P_c$  and updates them by selectively writing reliable proposals. Specifically, we build a hypergraph using patient embeddings. We then compute a centrality score and apply gate to filter out outlier proposals before memory updates. We further weight the remaining proposals by neighbor consistency and prediction confidence and use them to update  $P_c$  via EMA, yielding semantic anchors for  $\mathcal{L}_{\text{pTAC}}$ . Moreover, we update only representative and reliable proposals to prevent stage-shifted or outlier features from corrupting the anchors.

where  $\mathbf{D}_V$  and  $\mathbf{D}_E$  are the vertex and hyperedge degree matrices. We then compute a topology-aware centrality score:

$$\mathbf{v}^{(t+1)} = \frac{\mathbf{G}\mathbf{v}^{(t)}}{\mathbf{1}^\top \mathbf{G}\mathbf{v}^{(t)}}, \quad s_n = \mathbf{v}_n^{(T)}, \quad (3)$$

where  $t$  denotes the iteration index,  $T$  is the total number of iterations,  $\mathbf{v}^{(1)}$  is initialized randomly and normalized to sum to 1. Intuitively, in the cross-patient hypergraph, proposals with higher centrality are more consistent and representative across patients. Moreover, under DIL, such cross-patient relationships better capture modality-agnostic global regularities. Therefore, we define the gate  $g_n = \mathbb{I}[n \in s_q]$ , where  $s_q$  denotes the set of patients with the top- $q$  centrality values. We then apply  $g_n$  to filter out unstable and outlier proposals and prevent them from being written into the PRM.

**Weighted EMA Update.** For each gated proposal, we compute a per-class reliability weight  $\omega_{n,c}$  by combining two complementary cues: neighbor-consistency and prediction confidence. The neighbor-consistency term promotes cross-patient semantic agreement, whereas the confidence term downweights ambiguous regions, e.g., boundary or class-confusing areas. Specifically, neighbor-consistency is computed as the average cosine similarity between  $\bar{P}_{n,c}$  and the class- $c$  proposals of its neighbors in the hypergraph, while confidence is computed as the mean predicted probability within the ground-truth mask  $\mathbf{y}_{n,c}$ . Then we ignore invalid proposals with  $\sum \mathbf{y}_{n,c} = 0$ , and obtain  $\omega_{n,c}$  by applying a class-wise softmax function, and update the global prototype:

$$\tilde{P}_c = \sum_n \omega_{n,c} g_n \bar{P}_{n,c}, \quad P_c \leftarrow \eta P_c + (1 - \eta) \frac{\tilde{P}_c}{\|\tilde{P}_c\|_2}, \quad (4)$$

where  $\eta$  is the EMA rate. Overall, PRM maintains global prototype  $P_c$  that serve as semantic anchors for  $\mathcal{L}_{\text{pTAC}}$ .

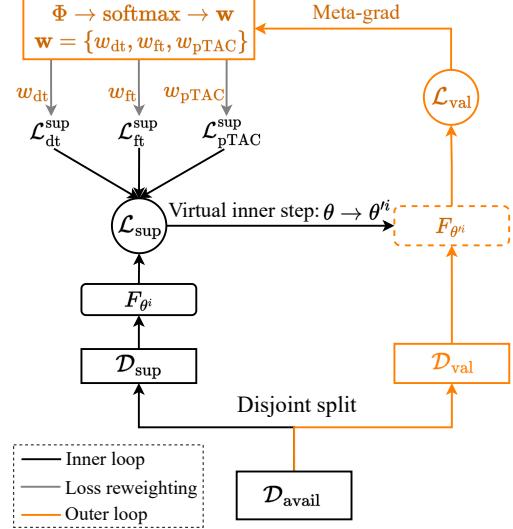


Fig. 4: Bi-level Meta Controller for stage-adaptive loss reweighting. Given  $\mathcal{D}_{\text{avail}}$ , we perform a disjoint split into a support batch  $\mathcal{D}_{\text{sup}}$  and a validation batch  $\mathcal{D}_{\text{val}}$ . We compute the weighted support loss  $\mathcal{L}_{\text{sup}}$  using three losses, i.e.,  $\mathcal{L}_{\text{dt}}$ ,  $\mathcal{L}_{\text{ft}}$ ,  $\mathcal{L}_{\text{pTAC}}$ , with weights  $\mathbf{w} = \text{softmax}(\Phi) = \{w_{\text{dt}}, w_{\text{ft}}, w_{\text{pTAC}}\}$ . A one-step virtual inner update  $\theta \rightarrow \theta'$  is performed on  $\mathcal{D}_{\text{sup}}$  to obtain  $F_{\theta'}$ . The controller logits  $\Phi$  are then updated by meta-gradients from a validation loss  $\mathcal{L}_{\text{val}}(\theta')$  evaluated on  $\mathcal{D}_{\text{val}}$ , then the weights are refreshed.

### 3.3.2. Prototype-based Tversky Contrastive Loss

We propose a prototype-based Tversky-Aware Contrastive (pTAC) loss to mitigate **SD** by aligning feature of tumor sub-regions extracted from  $\mathcal{D}_{\text{avail}}$  to class-wise global prototypes  $\{P_c\}_{c \in C}$  maintained by PRM.

**Generalized Tversky similarity.** Compared with cosine similarity, Tversky similarity explicitly reweights mismatch dimensions, making the alignment less sensitive to order-induced shifts. We therefore adopt Tversky similarity as our matching function, defined for  $\mathbf{a}, \mathbf{b} \in [0, 1]^D$  as:

$$S_{\text{tve}}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{b} + \alpha \mathbf{1}^\top (\mathbf{a}(1 - \mathbf{b})) + \beta \mathbf{1}^\top ((1 - \mathbf{a})\mathbf{b})}, \quad (5)$$

where  $\alpha$  and  $\beta$  control the penalty on false positives and false negatives, respectively. Furthermore,  $S_{\text{tve}}$  is input-agnostic because it measures element-wise soft overlap between two vectors. This means setting  $(\mathbf{a}, \mathbf{b})$  to prediction–mask pairs yields segmentation supervision, while setting them to features and prototypes enables prototype-based contrastive alignment.

**Prototype-based TAC loss.** To mitigate **SD** caused by order-induced shifts, we use class-wise global prototypes  $P_c$  as cross-stage anchors instead of queue-based instance pairs, and compute the feature-level Tversky similarity between each  $f_{n,c}$  and  $P_c$  as the matching score. Specifically, we optimize PMD by a prototype-based contrastive loss:

$$\mathcal{L}_{\text{pTAC}} = -\log \frac{\exp(S_{\text{tve}}(f_{n,c}, \mathbf{P}_c)/\tau)}{\exp(S_{\text{tve}}(f_{n,c}, \mathbf{P}_c)/\tau) + \sum_{k \neq c} \exp(S_{\text{tve}}(f_{n,c}, \mathbf{P}_k)/\tau)}, \quad (6)$$

where  $\tau$  is the temperature parameter. Prototypes are updated only through PRM, and treated as stable semantic anchors for tumor subregion representations.

### 3.4. Stage-Adaptive Objective Recalibration

We propose an online Meta Controller to adaptively rebalance three objectives during training: overlap-based supervision ( $\mathcal{L}_{dt}$ ), rare-structure emphasis ( $\mathcal{L}_{ft}$ ), and cross-stage semantic alignment ( $\mathcal{L}_{pTAC}$ ). This stage-adaptive reweighting is needed because the effective supervision signal can change across stages, making fixed loss weights prone to **OM**. Specifically, both segmentation losses are built on the Tversky similarity, which can be configured to penalize false negatives more than false positives. Concretely,  $\mathcal{L}_{dt}$  directly minimizes the Tversky overlap error, while  $\mathcal{L}_{ft}$  applies a focal-style modulation to emphasize hard or rare pixels:

$$\mathcal{L}_{dt}(\mathbf{y}', \mathbf{y}) = 1 - S_{tve}(\mathbf{y}', \mathbf{y}), \quad \mathcal{L}_{ft}(\mathbf{y}', \mathbf{y}) = (1 - S_{tve}(\mathbf{y}', \mathbf{y}))^\gamma, \quad (7)$$

where  $\mathbf{y}'$  denotes the predicted per-class probability maps,  $\mathbf{y}$  is the corresponding ground-truth mask, and  $\gamma$  controls the strength of the focusing effect. The controller architecture in Fig. 4 maintains learnable logits  $\Phi \in \mathbb{R}^3$  and outputs normalized loss weights  $\mathbf{w} = \{w_{dt}, w_{ft}, w_{pTAC}\}$  via  $\mathbf{w} = \text{softmax}(\Phi)$ .

**Weighted training objective.** At each meta update, we split the available batch  $\mathcal{D}_{\text{avail}}$  into two disjoint mini-batches: a support batch  $\mathcal{D}_{\text{sup}}$  for the inner (virtual) update and a validation batch  $\mathcal{D}_{\text{val}}$  for the outer evaluation, with  $\mathcal{D}_{\text{sup}} \cap \mathcal{D}_{\text{val}} = \emptyset$ . We compute these losses on  $\mathcal{D}_{\text{sup}}$ :

$$\mathcal{L}_{\text{sup}} = w_{dt} \mathcal{L}_{dt}^{\text{sup}} + w_{ft} \mathcal{L}_{ft}^{\text{sup}} + w_{pTAC} \mathcal{L}_{pTAC}^{\text{sup}}. \quad (8)$$

**Bi-level meta update.** We use a bi-level optimization to learn stage-adaptive loss weights. In the inner loop, we take a virtual gradient step on  $\mathcal{D}_{\text{sup}}$  to obtain updated network parameters:

$$\theta'^i = \theta^i - \eta_{\text{inner}} \nabla_{\theta^i} \mathcal{L}_{\text{sup}}, \quad (9)$$

where  $\eta_{\text{inner}}$  is the inner-loop learning rate. In the outer loop, we evaluate the virtually updated model on  $\mathcal{D}_{\text{val}}$  and update the controller logits  $\Phi$  through the virtual step:

$$\mathcal{L}_{\text{val}} = \mathcal{L}_{\text{val}}(\theta'^i), \quad \Phi \leftarrow \Phi - \eta_{\text{meta}} \nabla_{\Phi} \mathcal{L}_{\text{val}}(\theta'^i(\Phi)). \quad (10)$$

where  $\eta_{\text{meta}}$  is the meta learning rate. Finally, we refresh the loss weights by  $\mathbf{w} \leftarrow \text{softmax}(\Phi)$  for subsequent iterations. Thus, the total loss is defined:

$$\mathcal{L} = w_{dt} \mathcal{L}_{dt} + w_{ft} \mathcal{L}_{ft} + w_{pTAC} \mathcal{L}_{pTAC}. \quad (11)$$

This bi-level update lets the controller adapt the objective across stages, keeping optimization calibrated when the modality changes. The complete training procedure of PMD is summarized in Algorithm 1.

## 4. Experiments

### 4.1. Datasets

**BraTS 2019** (Bakas et al., 2018) contains 335 subject-level 3D MRI volumes with four aligned modalities (T1, T1CE, T2, and FLAIR) and expert-annotated tumor subregions (ET/TC/WT). Each volume is  $240 \times 240 \times 155$ . We split subjects into

---

### Algorithm 1: PMD Incremental Training Loop

---

**Input:** Current modality  $\{\mathbf{B}_i\}_{i=1}^I$ , replay buffer  $\mathbf{R}$ , global prototype  $P_c$ , controller logits  $\Phi$ , model  $F_{\theta^{i-1}}$ .

**Output:** Updated  $(\mathbf{R}, P_c, \Phi, F_{\theta^i})$

#### Step 0: Stage initialization

When  $i=1$ , model randomly initialize.

When  $i > 1$ ,  $F_{\theta^{i-1}}$  warm-starts from stage  $i-1$ .

#### Step 1: Build available batch

Sample  $\mathcal{B} \sim \mathbf{B}_i$  and  $\mathcal{R} \sim \mathbf{R}$ ; set  $\mathcal{D}_{\text{avail}} \leftarrow \mathcal{B} \cup \mathcal{R}$ ;  $\mathcal{D}_{\text{avail}}$  splits into  $\mathcal{D}_{\text{sup}}$  and  $\mathcal{D}_{\text{val}}$ .

#### Step 2: PRM update with hypergraph gate

Image-mask pair  $\{\mathbf{x}, \mathbf{y}\}$  is as input,  $\mathbf{x} \in \mathcal{D}_{\text{avail}}$ .

Forward  $\mathbf{x}$  to obtain decoder feature  $\mathbf{z}_n$ , and prediction  $\mathbf{y}'$  for  $\mathcal{L}_{dt}$  and  $\mathcal{L}_{ft}$ .

Generate proposals  $\tilde{P}_c = f_{n,c}$ ; build a hypergraph gate  $g_n$  and update prototypes:  $\tilde{P}_c = \sum_n \omega_{n,c} g_n \tilde{P}_{n,c}$ ,

$$P_c \leftarrow \eta P_c + (1 - \eta) \sum_n \omega_{n,c} \tilde{P}_c / \|\tilde{P}_c\|_2.$$

$P_c$  serves as a semantic anchor for  $\mathcal{L}_{pTAC}$ .

#### Step 3: Support loss and virtual inner step

Compute  $\mathcal{L}_{\text{sup}}$  on  $\mathcal{D}_{\text{sup}}$ ;  $\mathbf{w} \leftarrow \text{softmax}(\Phi)$ ;

$$\mathcal{L}_{\text{sup}} \leftarrow w_{dt} \mathcal{L}_{dt} + w_{ft} \mathcal{L}_{ft} + w_{pTAC} \mathcal{L}_{pTAC},$$

$$\theta'^i \leftarrow \theta^i - \eta_{\text{inner}} \nabla_{\theta^i} \mathcal{L}_{\text{sup}}.$$

#### Step 4: Update segmentation model at stage $i$

Recompute  $(w_{dt}, w_{ft}, w_{pTAC})$  and update  $\theta^i$  using

$$\mathcal{L} = w_{dt} \mathcal{L}_{dt} + w_{ft} \mathcal{L}_{ft} + w_{pTAC} \mathcal{L}_{pTAC}.$$

#### Step 5: Update replay buffer

Compute per-sample loss on  $\mathbf{B}_i$ , select the 10% samples, and add them to  $\mathbf{R}$ .

#### Step 6: Proceed to the next stage

Pass  $(\mathbf{R}, P_c, \Phi, F_{\theta^i})$  to stage  $i+1$  and repeat.

---

train/val/test with an 8:1:1 ratio, and then extract 2D slices with 41,540/5,115/5,270 train/val/test slices from each split.

**FeTS 2022** (Pati et al., 2021) is used to evaluate OOD generalization under dataset- and protocol-level shifts. It provides 1,251 MRI scans collected from 33 institutions, with the same four modalities and a canonical volume size of  $240 \times 240 \times 155$ . We apply the same slicing and cropping protocol as BraTS to obtain  $224 \times 224$  2D slices for evaluation.

**MU-Glioma-Post** (Mahmoud et al., 2025) is employed for OOD evaluation in a real missing modality setting. It includes 203 patients and 596 longitudinal timepoints, with T1/T1CE/T2/FLAIR acquisitions and expert-validated segmentation masks. We apply the same preprocessing to obtain 2D slices and exclude 149 timepoints without segmentation masks. In the released package, only 307/596 (51.51%) timepoints are complete; 70/596 (11.74%), 210/596 (35.23%), and 9/596 (1.51%) miss 1/2/3 modalities, respectively, with similar per-modality missing rates (~21–22%).

### 4.2. Implementation Details

All models were trained for 200 epochs (batch size 128) using Adam (learning rate  $1 \times 10^{-4}$ , weight decay  $4 \times 10^{-4}$ ). For the

Table 1: Segmentation results of WT, TC, and ET in terms of DSC(%) on the BraTS2019 dataset. Missing and available modalities are denoted by  $\circ$  and  $\bullet$ , respectively. “ $\dagger$ ” denote significance levels as p-value  $< 0.05$ .

Modalities				WT						TC						ET											
T2	T1c	T1	FL	mmF	MFI	$M^3$	M+P	MST	DC	RHD	Ours	mmF	MFI	$M^3$	M+P	MST	DC	RHD	Ours	mmF	MFI	$M^3$	M+P	MST	DC	RHD	Ours
•	◦	◦	◦	80.66	81.87	84.49	82.74	86.58	87.03	86.71	86.66	69.45	58.26	71.35	55.74	72.46	73.01	73.68	76.14 <sup>†</sup>	37.93	39.82	47.61	36.29	49.98	50.65	49.75	52.19 <sup>†</sup>
◦	•	◦	◦	73.64	74.39	78.62	75.88	79.05	79.71	79.48	81.83 <sup>†</sup>	78.62	81.32	83.95	82.83	83.55	84.08	84.65	86.93 <sup>†</sup>	70.64	76.53	79.33	79.29	80.12	80.74	80.17	82.56 <sup>†</sup>
◦	◦	◦	◦	73.76	73.28	79.32	78.32	80.23	80.87	80.77	83.16 <sup>†</sup>	61.36	65.14	65.98	62.13	68.55	69.12	69.34	71.62 <sup>†</sup>	33.05	33.82	44.43	38.40	45.41	46.07	45.22	47.64 <sup>†</sup>
◦	◦	◦	◦	87.31	86.96	86.38	87.05	87.80	88.34	88.25	90.62 <sup>†</sup>	66.85	57.31	70.37	64.53	71.78	72.35	72.87	75.31 <sup>†</sup>	37.37	40.67	45.62	38.23	48.84	49.53	48.61	51.08 <sup>†</sup>
•	◦	◦	◦	84.55	85.81	85.74	84.47	86.71	87.29	87.13	89.41 <sup>†</sup>	79.98	85.20	83.25	82.94	84.93	85.44	85.96	85.83	71.92	80.04	78.76	79.75	80.55	81.18	80.51	82.93 <sup>†</sup>
◦	◦	◦	◦	77.60	74.21	80.31	79.95	81.43	81.92	81.48	83.92 <sup>†</sup>	79.98	83.09	85.43	86.59	85.18	85.75	86.32	86.06	71.38	77.47	80.31	82.35	81.47	82.06	80.97	83.31 <sup>†</sup>
◦	◦	◦	◦	88.90	88.09	90.17	89.12	89.39	89.93	89.63	92.08 <sup>†</sup>	70.44	65.96	66.32	66.59	70.92	71.49	71.75	74.18 <sup>†</sup>	41.93	41.47	48.83	42.49	49.88	50.52	49.55	51.97 <sup>†</sup>
◦	◦	◦	◦	83.17	84.52	86.29	85.30	87.21	87.78	87.58	89.93 <sup>†</sup>	68.73	60.63	65.52	62.16	68.47	69.03	69.42	71.83 <sup>†</sup>	42.46	43.36	44.65	38.98	47.62	48.28	47.36	49.72 <sup>†</sup>
◦	◦	◦	◦	88.68	89.53	90.21	88.67	90.03	90.63	90.39	92.74 <sup>†</sup>	70.72	65.43	70.21	68.48	70.08	70.65	70.96	73.41 <sup>†</sup>	43.54	47.29	46.69	46.75	46.95	47.62	46.93	49.38 <sup>†</sup>
◦	◦	◦	◦	88.50	89.91	90.75	89.56	91.27	91.86	91.42	91.83	80.84	82.26	83.02	82.37	82.31	82.81	83.22	85.67 <sup>†</sup>	72.03	77.52	78.17	75.58	78.93	79.56	78.65	81.06 <sup>†</sup>
◦	◦	◦	◦	89.02	89.78	89.89	89.91	90.48	91.13	90.75	93.14 <sup>†</sup>	82.33	85.96	86.13	86.06	85.47	86.02	86.46	88.82 <sup>†</sup>	72.82	83.23	84.17	83.19	83.72	84.31	83.26	85.69 <sup>†</sup>
◦	◦	◦	◦	89.08	90.06	90.44	89.99	90.87	91.39	91.19	93.63 <sup>†</sup>	71.90	66.37	71.87	59.87	71.58	72.14	72.50	74.98 <sup>†</sup>	45.34	47.89	50.26	37.08	48.35	49.02	48.11	49.27 <sup>†</sup>
◦	◦	◦	◦	88.94	91.05	91.21	89.83	91.26	91.83	91.46	93.87 <sup>†</sup>	81.64	86.99	86.53	86.20	86.69	87.24	87.79	90.24 <sup>†</sup>	72.25	83.32	82.48	82.10	85.06	85.72	84.62	87.03 <sup>†</sup>
◦	◦	◦	◦	84.95	86.47	87.89	86.59	87.68	88.33	87.98	90.46 <sup>†</sup>	81.27	85.58	86.41	88.12	88.02	88.58	89.24	91.68 <sup>†</sup>	72.98	81.11	82.35	82.88	83.86	84.45	83.46	85.91 <sup>†</sup>
◦	◦	◦	◦	89.21	90.81	91.36	90.35	92.37	92.86	92.43	94.72 <sup>†</sup>	82.47	86.79	86.51	87.73	86.74	87.31	88.06	90.47 <sup>†</sup>	72.74	83.34	83.88	82.76	85.92	86.56	85.35	87.78 <sup>†</sup>
Means				84.53	85.12	86.87	85.91	87.50	88.05	87.78	90.15	75.09	74.42	77.53	74.82	78.45	79.01	79.49	81.89	57.23	62.46	65.17	61.74	66.44	67.08	66.16	68.58

Table 2: The OOD segmentation results of WT, TC, and ET in terms of DSC(%) on FeTS2022 (simulated missing-modality) and MU-Glioma-Post (real missing-modality). Missing and available modalities are denoted by  $\circ$  and  $\bullet$ , respectively. “-” indicates the modality combination does not occur in the dataset. “ $\dagger$ ” denotes significance level is reached as p-value  $< 0.05$ .

Modalities				FeTS						MU-Glioma-Post																			
				WT			ET			WT			TC			ET													
T2	T1c	T1	FL	MFI	M+P	DC	RHD	Ours	MFI	M+P	DC	RHD	Ours	MFI	M+P	DC	RHD	Ours	MFI	M+P	DC	RHD	Ours						
•	◦	◦	◦	71.20	79.93	82.07	83.76	83.67	48.16	65.61	65.82	71.01	72.65 <sup>†</sup>	25.85	41.40	46.85	49.81	52.45 <sup>†</sup>	58.78	60.87	63.79	64.54	67.68 <sup>†</sup>	13.48	32.43	31.38	16.48	19.31	
◦	•	◦	◦	62.14	70.11	70.49	76.28	75.41	69.98	82.22	82.76	88.64	88.06	68.50	80.79	81.45	84.45	83.89	-	-	-	-	-	-	-	-	-		
◦	◦	◦	◦	54.04	61.48	71.44	74.10	74.53 <sup>†</sup>	31.81	52.59	61.32	68.16	65.59	6.98	33.86	38.96	44.22	43.55	-	-	-	-	-	-	-	-	-		
◦	◦	◦	◦	76.73	83.68	84.36	87.97	88.98 <sup>†</sup>	48.87	64.41	64.83	73.86	72.96	10.47	41.81	39.16	52.00	52.46 <sup>†</sup>	-	-	-	-	-	-	-	-	-		
◦	◦	◦	◦	78.45	84.64	86.14	87.12	87.64 <sup>†</sup>	77.45	85.56	86.40	89.95	90.72 <sup>†</sup>	76.47	84.22	84.21	85.47	85.91 <sup>†</sup>	76.31	77.99	77.99	78.86	81.51 <sup>†</sup>	34.91	35.32	38.78	39.95	43.35 <sup>†</sup>	
◦	◦	◦	◦	68.64	72.80	77.10	79.23	79.31 <sup>†</sup>	76.01	84.87	85.69	90.22	90.33 <sup>†</sup>	75.01	82.23	83.06	85.59	85.64 <sup>†</sup>	59.56	60.57	61.87	62.20	66.74 <sup>†</sup>	52.73	53.62	54.93	55.66	58.49 <sup>†</sup>	
◦	◦	◦	◦	79.06	86.98	87.92	88.05	89.78 <sup>†</sup>	49.21	71.11	71.76	77.21	79.70 <sup>†</sup>	13.13	51.07	47.55	55.69	59.80 <sup>†</sup>	66.36	67.32	78.34	74.41	77.18	29.44	29.06	40.93	41.08	44.43 <sup>†</sup>	
◦	◦	◦	◦	78.12	82.97	84.29	85.72	86.40 <sup>†</sup>	44.34	70.79	72.06	75.23	77.58 <sup>†</sup>	17.07	49.06	51.76	53.12	57.24 <sup>†</sup>	65.80	67.13	81.46	79.67	82.94 <sup>†</sup>	28.76	32.59	32.82	33.84	37.53 <sup>†</sup>	
◦	◦	◦	◦	80.14	87.99	88.35	89.05	90.56 <sup>†</sup>	57.94	74.42	72.07	77.07	80.04 <sup>†</sup>	29.43	51.92	51.45	57.66	61.72 <sup>†</sup>	-	-	-	-	-	-	-	-	-		
◦	◦	◦	◦	79.69	88.18	88.96	89.95	90.70 <sup>†</sup>	78.45	87.58	87.69	90.30	91.18 <sup>†</sup>	78.92	84.90	85.20	85.62	86.43 <sup>†</sup>	-	-	-	-	-	-	-	-	-		
◦	◦	◦	◦	81.48	88.73	89.25	89.75	90.71 <sup>†</sup>	80.53	88.49	88.04	90.93	91.73 <sup>†</sup>	79.91	85.35	85.39	85.92	86.60 <sup>†</sup>	79.69	80.61	81.85	82.33	85.92 <sup>†</sup>	41.31	43.93	42.96	43.03	46.71 <sup>†</sup>	
◦	◦	◦	◦	80.57	88.25	89.01	88.63	90.18 <sup>†</sup>	55.14	76.60	74.64	78.02	82.22 <sup>†</sup>	25.93	56.05	53.84	58.45	63.83 <sup>†</sup>	83.45	83.27	86.77	84.77	90.93 <sup>†</sup>	40.59	39.18	43.46	43.70	47.92 <sup>†</sup>	
◦	◦	◦	◦	81.30	88.93	90.09	90.41	91.24 <sup>†</sup>	78.79	88.11	87.41	90.59	91.45 <sup>†</sup>	79.09	85.20	84.86	85.64	86.38 <sup>†</sup>	60.21	61.35	64.62	64.44	67.61 <sup>†</sup>	40.91	34.57	44.83	45.00	48.84 <sup>†</sup>	
◦	◦	◦	◦	79.73	84.86	86.64	87.25	87.71 <sup>†</sup>	79.50	87.26	86.94	90.71	91.13 <sup>†</sup>	77.89	84.79	84.58	85.93	86.12 <sup>†</sup>	70.42	71.05	77.74	74.39	77.17	38.70	39.99	41.92	42.13	45.52 <sup>†</sup>	
◦	◦	◦	◦	81.86	89.06	90.21	90.18	91.25 <sup>†</sup>	79.94	88.54	87.83	90.91	91.79 <sup>†</sup>	79.50	85.41	85.17	85.87	86.55 <sup>†</sup>	80.64	81.52	83.38	83.68	86.19	46.75	46.45	50.14	51.67	55.54 <sup>†</sup>	
◦	◦	◦	◦	Means	75.54	82.57	84.42	85.83	86.54	63.74	77.88	77.68	82.85	83.81	49.58	66.54	66.23	70.36	71.90	70.81	71.86	76.00	75.58	78.61	35.11	36.72	40.28	39.49	43.11

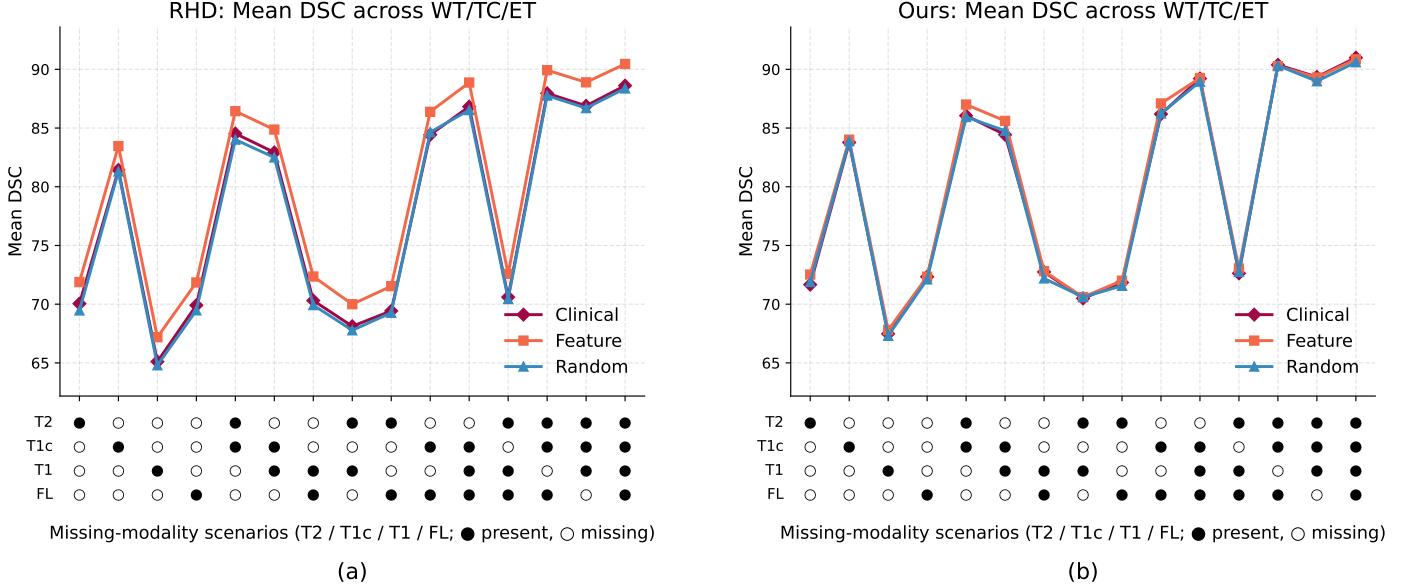


Fig. 5: Order sensitivity on BraTS2019 under three incremental orders: clinical, feature-driven, and random. We evaluate each model on 15 missing modality scenarios using the mean Dice score (averaged over WT/TC/ET). The resulting performance differs markedly across orders, indicating strong sensitivity to the arrival order. (a) RHD (Wang et al., 2025) exhibits pronounced order-induced fluctuations across scenarios. (b) Our method substantially reduces the discrepancy across orders and yields more consistent performance.

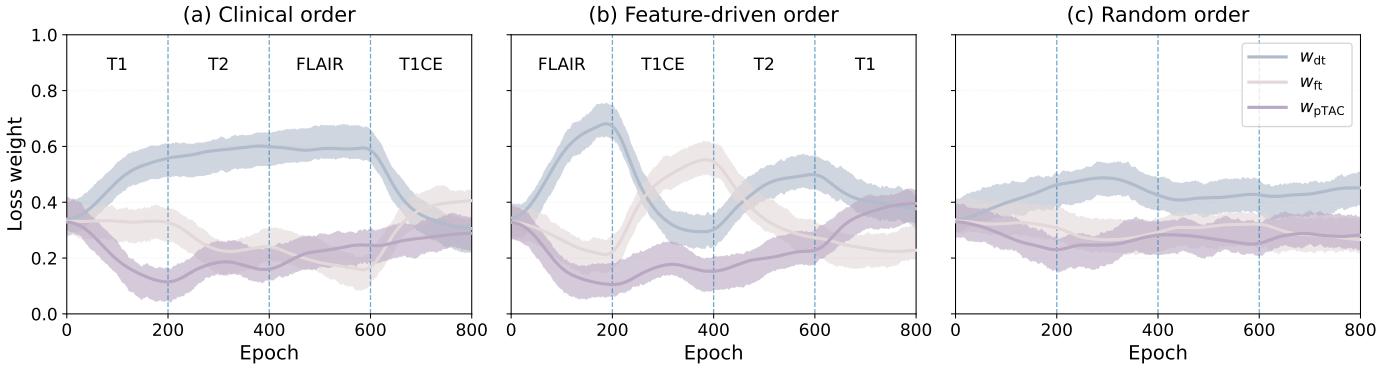


Fig. 6: Dynamic loss-weight evolution under different modality orders. We visualize the weights for the three objectives  $w_{dt}$ ,  $w_{ft}$ ,  $w_{pTAC}$  across incremental training under three modality orders (clinical, feature-driven, and random). Each stage lasts 200 epochs; vertical dashed lines indicate stage boundaries. Solid curves show the mean weights and the shaded regions denote the mean  $\pm$  std.

ducted on an NVIDIA H100 GPU.

#### 4.4. Quantitative Results

**ID Results.** We first performed ID testing and compared with other SOTA methods. As shown in Table 1, our proposed method achieves the best performance in 40 out of 45 entries, ranking first in 13, 13, and 14 out of 15 settings for WT, TC, and ET, respectively. Averaged over all configurations, our method reaches 90.15% DSC in WT, 81.89% in TC, and 68.58% in ET, improving upon the incremental baseline RHD by a mean of 2.39%. Compared with the strongest non-incremental baseline DC, our method further gains 2.10% in WT, 2.88% in TC, and 1.50% in ET. Moreover, the improvements are statistically significant ( $p$ -value  $< 0.05$ ), highlighting consistent advantages across missing-modality scenarios. Notably, while most prior approaches assume full-modality training with static missing settings, our method is trained under a

modality-incremental protocol that better matches realistic clinical acquisition, and the strong results validate its robustness under progressive modality availability.

**OOD Results.** We further compared our method with others on two OOD datasets: FeTS 2022 and MU-Glioma-Post (Table 2). Our method consistently achieves the highest mean DSC across all tumor regions on both benchmarks. On FeTS 2022, we obtain 71.90% for ET, outperforming the second-best baseline by 1.54%. On MU-Glioma-Post, which exhibits greater domain shift, our method reaches 78.61% in WT, 43.11% in TC, and 44.71% in ET, improving over DC (Li et al., 2025) by 2.61% in WT, 2.83% in TC, and 2.76% in ET. The majority of results achieve statistical significance. Nevertheless, the relatively lower absolute performance on MU-Glioma-Post reflects the compounded challenge of domain shift and clinically correlated missingness. Under these conditions, pTAC with PRM provides stable cross-patient semantic anchors, while the Meta

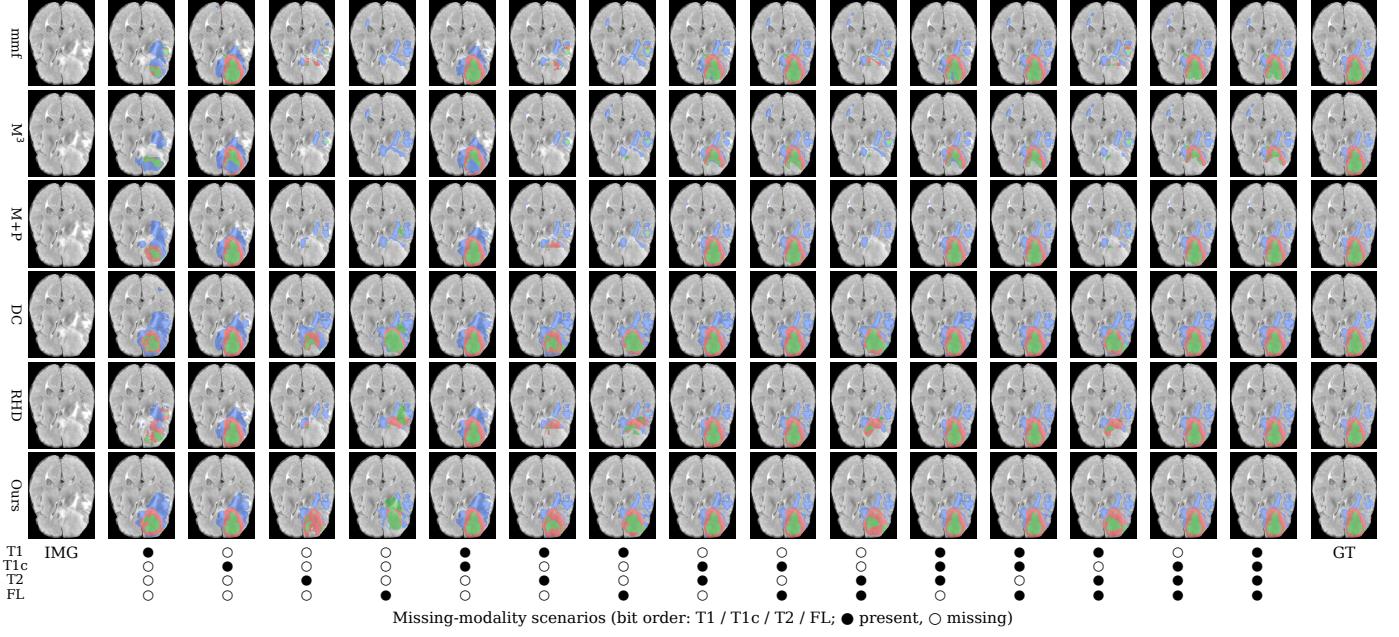


Fig. 7: The segmentation results across various missing-modality scenarios are shown. The rows represent different methods: mmf, M3, M+P, DC, RHD, and Ours. The columns show the different missing-modality combinations, with the corresponding symbols at the bottom indicating the presence or absence of T1, T1c, T2, and FL modalities. The ground truth (GT) is shown in the last column for reference. In all scenarios, our method achieves nearly the best segmentation performance, particularly in T1+T2, T2+FL and T1+T2+FL missing-modality scenarios. Additionally, in the full-modality scenario, our method demonstrates superior segmentation of small tumor regions, especially for the ET region, outperforming other methods.

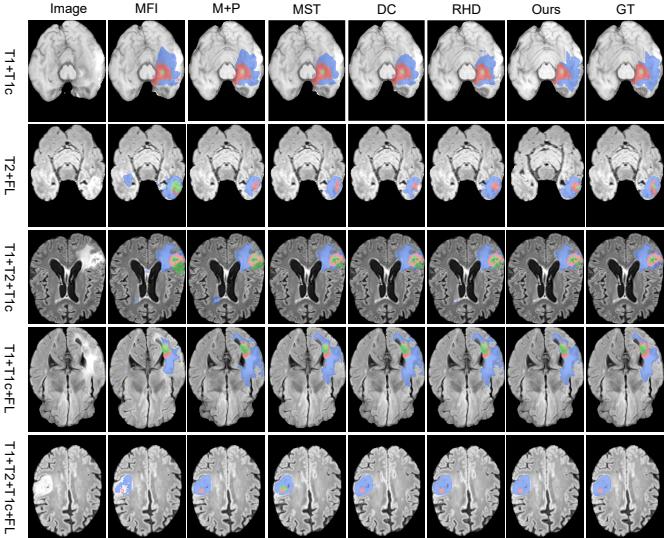


Fig. 8: OOD segmentation results on MU-Glioma-Post. Columns show the input image, predictions by MFI, M+P, MST, DC, RHD, and Ours, and the ground truth (GT). Rows correspond to diverse real missing modality scenarios. In these scenarios, several baselines exhibit inaccurate WT contours or fragmented regions, while under full modalities the OOD shift may still induce false positive ET predictions; in contrast, our method yields more consistent boundaries and fewer spurious subregion predictions.

Controller adaptively reweights losses to mitigate **OM** when missing modalities vary.

**Order Sensitivity Evaluation.** We investigate sensitivity to modality training order under three incremental schedules: clinical, feature-driven, and random, as shown in Fig. 5. Across all 15 missing modality scenarios, our method consistently out-

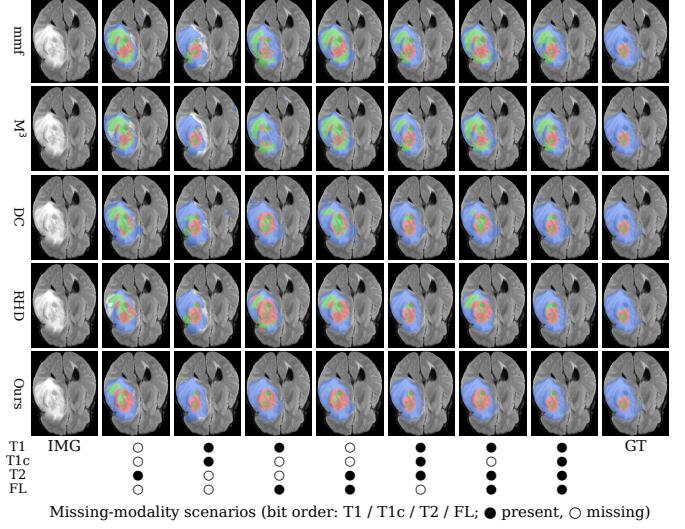


Fig. 9: Failure case analysis. A light-gray circular region within the WT area is misclassified as TC/ET by all methods. Available and missing modalities are denoted by ● and ○, respectively.

performs RHD regardless of the training order. Furthermore, the three curves in our method exhibit notably smaller variance compared to RHD, indicating reduced sensitivity to the order in which modalities are introduced during training. For instance, under the clinical order, the mean DSC improves from 87.78 to 90.15 for WT and from 66.16 to 68.58 for ET, with comparable gains observed under feature-driven and random schedules. This order-invariant behavior suggests that the PRM effectively decouples semantic learning from modality-specific training dynamics, enabling robust feature aggregation.

Table 3: Incremental ablation study of WT, TC, and ET in terms of DSC(%) on the BraTS2019 dataset. Missing and available modalities are denoted by  $\circ$  and  $\bullet$ , respectively. The columns +proto, +meta represent the incremental components: Base (DIL with replay), +proto (adding PRM and pTAC), +meta (further adding Meta Controller), Ours (full method).

Modalities				WT			TC			ET					
T2	T1c	T1	FL	Base	+proto	+meta	Ours	Base	+proto	+meta	Ours	Base	+proto	+meta	Ours
•	◦	◦	◦	86.71	88.23	88.67	88.97	73.68	75.29	75.83	76.14	49.75	51.32	51.87	52.19
◦	•	◦	◦	79.48	81.09	81.46	81.83	84.65	86.21	86.59	86.93	80.17	81.74	82.18	82.56
◦	◦	•	◦	80.77	82.31	82.74	83.16	69.34	70.86	71.27	71.62	45.22	46.81	47.25	47.64
◦	◦	◦	•	88.25	89.84	90.26	90.62	72.87	74.49	74.93	75.31	48.61	50.19	50.67	51.08
•	•	◦	◦	87.13	88.69	89.08	89.41	85.96	87.57	87.98	88.34	80.51	82.07	82.53	82.93
◦	•	•	◦	81.48	83.04	83.51	83.92	86.32	87.93	88.39	88.79	80.97	82.58	82.97	83.31
◦	◦	•	•	89.63	91.21	91.68	92.08	71.75	73.32	73.78	74.18	49.55	51.13	51.58	51.97
•	◦	•	◦	87.58	89.17	89.58	89.93	69.42	70.94	71.41	71.83	47.36	48.91	49.34	49.72
•	◦	◦	◦	90.39	91.96	92.38	92.74	70.96	72.58	73.02	73.41	46.93	48.54	48.97	49.38
◦	•	◦	◦	91.42	93.03	93.45	93.81	83.22	84.81	85.27	85.67	78.65	80.29	80.71	81.06
◦	•	•	•	90.75	92.27	92.73	93.14	86.46	88.09	88.48	88.82	83.26	84.83	85.29	85.69
•	◦	•	•	91.19	92.78	93.23	93.63	72.50	74.14	74.59	74.98	48.11	49.67	50.06	50.41
•	•	◦	•	91.46	93.09	93.51	93.87	87.79	89.36	89.83	90.24	84.62	86.26	86.67	87.03
•	•	◦	◦	87.98	89.53	90.02	90.46	89.24	90.83	91.28	91.68	83.46	85.08	85.52	85.91
•	•	•	•	92.43	93.98	94.38	94.72	88.06	89.69	90.11	90.47	85.35	86.96	87.41	87.78

**Meta Controller Weight Dynamics.** We study the learned loss weights across training epochs under all three incremental orders, as shown in Fig. 6. A consistent trend emerges:  $w_{dt}$  gradually increases as more modalities arrive, reflecting the growing need for precise overlap supervision;  $w_{ft}$  decreases over stages, as complementary modalities partially alleviate class imbalance; and  $w_{pTAC}$  remains stable with a mild increase when more cross-modal pairs become available. The weight trajectories shift visibly at each stage boundary, confirming that the controller responds to newly arrived modalities. Despite different arrival orders, the final weight distributions converge to comparable ranges, explaining the order-robust performance.

#### 4.5. Qualitative Results.

**ID qualitative comparisons on BraTS2019.** Qualitative comparisons further validate the effectiveness of our method across diverse missing modality scenarios, as illustrated in Fig. 7. Our predictions most closely approximate the ground truth under various missing configurations. For instance, in the single-modality case with only T1 available, competing methods exhibit substantial under-segmentation of TC (red) and ET (green), whereas our approach preserves the spatial extent of both regions. In the full-modality scenario, our method demonstrates superior delineation of the small ET region compared to other approaches, which tend to either under-segment or produce fragmented boundaries. This improved fine-grained segmentation can be attributed to the prototype-guided semantic alignment, which provides consistent structural priors across tumor subregions regardless of modality availability, thereby enhancing boundary precision for small but critical structures.

**OOD qualitative comparisons on MU-Glioma-Post.** Qualitative results on MU-Glioma-Post are shown in Fig. 8 under real missing-modality evaluation, where some modalities are genuinely unavailable because they have not arrived in clinical acquisition. Each case is therefore tested with its naturally observed modality set. Under this OOD shift, many baselines become less stable, showing WT boundary drift and unreliable TC/ET localization when key modalities are missing. With only T1+T1CE, MFI and M+P often over-segment WT without FLAIR cues, whereas our predictions better preserve the

ground-truth boundary and remain more spatially coherent. In contrast, our predictions remain more coherent and closer to the ground truth. Even with full modalities, MST hallucinates ET, whereas our method suppresses spurious ET, indicating robustness to missing scenarios and distribution shift.

**Failure case analysis.** Despite the overall improvements, failure cases reveal remaining challenges, as shown in Fig. 9. Specifically, a light-gray circular region within the WT area is consistently misclassified as TC/ET by all methods, including ours. This false positive likely arises from the region’s intensity profile resembling enhancing tissue on T1c, leading to ambiguous feature representations that even prototype-based alignment cannot fully resolve. Such cases highlight the inherent difficulty in distinguishing necrotic or cystic components from active tumor when imaging characteristics overlap, suggesting that incorporating domain knowledge about atypical tumor presentations may further improve segmentation.

#### 4.6. Ablation Study

##### 4.6.1. Overall Ablation Study

We conduct an overall ablation on BraTS 2019 across all 15 missing-modality scenarios in Table 3, progressively adding pTAC with PRM (+proto) and the Meta Controller (+meta) to a replay-based DIL baseline (Base). In single-modality cases, +proto contributes the majority of the gain. For example, under T1-only, +proto raises WT from 80.77% to 82.31%, while +meta adds a further 0.43%, indicating that stabilizing region features via prototype anchoring is more critical than loss recalibration when semantic cues are scarce. In multi-modality cases, the contribution of +meta becomes more pronounced as the objective must adapt to diverse modality compositions; under T1c+T2+FLAIR, +proto improves WT by 1.55% over Base, and +meta adds another 0.49%. Furthermore, the improvements persist even with full modalities, where +proto and +meta together yield gains of 2.29% in WT, 2.41% in TC, and 2.43% in ET, confirming that the two components strengthen overall representations beyond merely compensating for missing modalities. In summary, +proto stabilizes features against semantic drift, while +meta calibrates the objective to the current modality composition.

Table 4: Component-wise ablation study: mean DSC(%) over 15 missing-modality scenarios. Variants are grouped by prototype reliability and alignment, loss calibration, and hypergraph configuration.

Variant	BraTS 2019			FeTS 2022			MU-Glioma-Post		
	WT	TC	ET	WT	TC	ET	WT	TC	ET
<i>Prototype reliability and semantic alignment</i>									
w/o weighted EMA	89.81	81.55	68.31	86.28	83.64	71.72	78.44	42.96	44.58
w/o PRM gate	89.67	81.38	68.13	86.09	83.29	71.41	78.22	42.59	44.35
w/o pTAC	89.04	80.91	67.62	85.41	82.89	71.03	77.52	42.18	43.84
Cosine (in pTAC)	89.61	81.43	67.82	86.03	83.38	71.21	78.14	42.72	44.08
<i>Loss calibration</i>									
Heuristic schedule	89.48	81.27	68.03	86.18	83.51	71.58	78.31	42.64	44.42
Fixed weights	88.94	80.78	67.51	85.89	82.95	70.86	77.88	41.92	43.62
<i>Hypergraph encoder layers</i>									
5th only	89.74	81.51	68.03	86.16	83.47	71.42	78.26	42.80	44.25
3rd + 4th + 5th	89.91	81.74	68.39	86.33	83.68	71.73	78.42	42.99	44.55
Ours (full)	90.15	81.89	68.58	86.54	83.81	71.90	78.61	43.11	44.71

#### 4.6.2. Component-Wise Ablation Study

We ablate each key module and report mean DSC (%) over all 15 missing-modality scenarios on BraTS 2019, FeTS 2022, and MU-Glioma-Post (Table 4). We test five variants that isolate prototype reliability, loss calibration, and semantic alignment: (i) *w/o weighted EMA*, (ii) *w/o PRM gate*, (iii) *Heuristic schedule*, (iv) *w/o pTAC*, and (v) *Fixed weights*.

**Prototype Reliability and Semantic Alignment.** Removing weighted EMA yields a small drop of about 0.3% on BraTS, showing that weighting mainly refines prototype quality. Removing the PRM gate hurts more consistently across all datasets, indicating the gate is the primary protection against outliers and stage-shifted samples corrupting global prototypes. The largest degradation comes from removing pTAC, which causes around 1.0% decline on all three datasets, confirming that prototype-anchored contrastive alignment is the main driver for reducing Semantic Drift by keeping region features consistent across stages.

**Adaptive Loss Reweighting.** The heuristic schedule linearly increases  $w_{pTAC}$  from 0.1 to 0.35 across stages while proportionally decreasing  $w_{dt}$  and  $w_{ft}$ , reflecting the intuition that cross-modality alignment should receive greater emphasis as more modalities become available. Moreover, using fixed weights degrades the most among calibration variants, with about 1.1% decline on BraTS. The gap becomes more visible under OOD settings: on MU-Glioma-Post, fixed weights drop TC by about 1.2%, supporting that meta-learned reweighting improves both optimization stability and OOD robustness by keeping the objective calibrated as modality composition evolves.

**Similarity Metric and Hypergraph Layers.** We further study the similarity metric in  $\mathcal{L}_{pTAC}$  and the encoder layers for hypergraph construction (Table 4). Replacing Tversky similarity with cosine similarity leads to consistent degradation across all three datasets, with ET suffering the largest drop (0.76% on BraTS, 0.69% on FeTS, 0.63% on MU-Glioma-Post). This is expected, as cosine similarity treats all dimensions equally and fails to suppress spurious activations, whereas the asymmetric penalties in Tversky similarity provide stronger alignment for small, under-represented structures that are most vulnerable to order-induced shifts. For hypergraph construction, using only the 5th encoder reduces ET by 0.55% on BraTS and 0.48% on FeTS, indicating that high-level features alone lack sufficient spatial granularity for reliable patient similarity estimation. Adding

a lower layer (3rd+4th+5th) yields only a modest decline of about 0.2% across datasets, suggesting that early-stage features introduce texture noise that weakens the hypergraph topology. The 4th+5th combination therefore strikes the best trade-off between semantic abstraction and spatial detail.

## 5. Conclusion

We presented PMD, a Prototype-guided Meta Domain Incremental Learning framework for missing-modality brain tumor segmentation under progressive modality availability. PMD mitigates **Semantic Drift (SD)** by maintaining cross-stage prototype anchors via a Prototype Anchored Representation Memory (PRM) and aligning region features with a prototype-based Tversky-Aware Contrastive loss (pTAC). To address **Objective Miscalibration (OM)** induced by modality switches and varying missing-modality patterns, we further introduced a bi-level Meta Controller that learns stage-adaptive reweighting over  $\mathcal{L}_{dt}$ ,  $\mathcal{L}_{ft}$ , and  $\mathcal{L}_{pTAC}$ . Extensive experiments on BraTS 2019 and two OOD datasets (FeTS 2022 and MU-Glioma-Post) show the improvements under diverse missing-modality scenarios, reduced sensitivity to modality arrival orders, and stronger robustness under real missing and OOD scenarios.

## 6. Limitation

While PMD improves robustness under DIL, several challenges remain. First, the bi-level Meta Controller introduces additional computation and memory overhead due to the virtual inner step and meta-gradient updates. Moreover, failure cases on OOD data suggest that ambiguous intensity patterns can still cause systematic confusion, thus needing stronger priors.

## References

- Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinozaki, R.T., Berger, C., Ha, S.M., Rozycki, M., et al., 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 .
- Ellingson, B.M., Bendszus, M., Boxerman, J., Barboriak, D., Erickson, B.J., Smits, M., Nelson, S.J., Gerstner, E., Alexander, B., Goldmacher, G., et al., 2015. Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. Neuro-oncology 17, 1188–1198.
- Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y., 2019. Hypergraph neural networks, in: Proceedings of the AAAI conference on artificial intelligence, pp. 3558–3565.
- Havaei, M., Guizard, N., Chapados, N., Bengio, Y., 2016. Hemis: Hetero-modal image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19, Springer, pp. 469–477.
- Jing, W., Wang, J., Di, D., Li, D., Song, Y., Fan, L., 2025. Multi-modal hypergraph contrastive learning for medical image segmentation. Pattern Recognition 165, 111544.
- Kalb, T., Mauthe, B., Beyerer, J., 2022. Improving replay-based continual semantic segmentation with smart data selection, in: 2022 IEEE 25th international conference on intelligent transportation systems (ITSC), IEEE, pp. 1114–1121.
- Kumari, P., Chauhan, J., Bozorgpour, A., Huang, B., Azad, R., Merhof, D., 2025. Continual learning in medical image analysis: A comprehensive review of recent advancements and future prospects. Medical Image Analysis , 103730.

- Li, H., Li, Z., Mao, Y., Ding, Z., Huang, Z., 2025. Dc-seg: Disentangled contrastive learning for brain tumor segmentation with missing modalities, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 138–148.
- Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y., 2023. M3ae: multimodal representation learning for brain tumor segmentation with missing modalities, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1657–1665.
- Liu, T., Jiang, H., Huang, K., 2025. Kmd: Koopman multi-modality decomposition for generalized brain tumor segmentation under incomplete modalities, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 15663–15671.
- Mahmoud, E., Gass, J., Dhemesh, Y., Greaser, J., Pogorzelski, K., Isufi, E., Garrett, F., Thacker, J., Tahon, N.h., Sinclair, J., et al., 2025. Mu-glioma post: A comprehensive dataset of automated mr multi-sequence segmentation and clinical features. *Scientific data* 12, 1847.
- Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S.H., Reina, G.A., Foley, P., Gruzdev, A., Karkada, D., Davatzikos, C., et al., 2022. Federated learning enables big data for rare cancer boundary detection. *Nature communications* 13, 7346.
- Pati, S., Baid, U., Zenk, M., Edwards, B., Sheller, M., Reina, G.A., Foley, P., Gruzdev, A., Martin, J., Albarqouni, S., et al., 2021. The federated tumor segmentation (fets) challenge. *arXiv preprint arXiv:2105.05874*.
- Pipoli, V., Saporita, A., Marchesini, K., Grana, C., Ficarra, E., Bolelli, F., 2025. Im-fuse: A mamba-based fusion block for brain tumor segmentation with incomplete modalities, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 225–235.
- Rees, J., 2003. Advances in magnetic resonance imaging of brain tumours. *Current opinion in neurology* 16, 643–650.
- Sawlani, V., Patel, M.D., Davies, N., Flintham, R., Wesolowski, R., Ughratdar, I., Pohl, U., Nagaraju, S., Petrik, V., Kay, A., et al., 2020. Multiparametric mri: practical approach and pictorial review of a useful tool in the evaluation of brain tumours and tumour-like lesions. *Insights into imaging* 11, 1–19.
- Shi, J., Shang, C., Sun, Z., Yu, L., Yang, X., Yan, Z., 2024. Passion: Towards effective incomplete multi-modal medical image segmentation with imbalanced missing rates, in: Proceedings of the 32nd ACM International Conference on Multimedia, pp. 456–465.
- Shi, J., Yu, L., Cheng, Q., Yang, X., Cheng, K.T., Yan, Z., 2023. M2ftrans: Modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*
- Tang, Q., Fan, L., Pagnucco, M., Song, Y., 2025. Prototype-based image prompting for weakly supervised histopathological image segmentation, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 30271–30280.
- Tversky, A., 1977. Features of similarity. *Psychological review* 84, 327.
- Varoquaux, G., Cheplygina, V., 2022. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ digital medicine* 5, 48.
- Van de Ven, G.M., Tuytelaars, T., Tolias, A.S., 2022. Three types of incremental learning. *Nature Machine Intelligence* 4, 1185–1197.
- Wang, J., Fan, L., Jing, W., Di, D., Song, Y., Liu, S., Cong, C., 2025. Hypergraph tversky-aware domain incremental learning for brain tumor segmentation with missing modalities, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 283–293.
- Wang, Z., Hong, Y., 2023. A2fseg: Adaptive multi-modal fusion network for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 673–681.
- Weller, M., van den Bent, M., Preusser, M., Le Rhun, E., Tonn, J.C., Minniti, G., Bendszus, M., Balana, C., Chinot, O., Dirven, L., et al., 2021. Eano guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nature reviews Clinical oncology* 18, 170–186.
- Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorenson, A.G., Galanis, E., DeGroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., et al., 2010. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of clinical oncology* 28, 1963–1972.
- Ye, Y., Xie, Y., Zhang, J., Chen, Z., Wu, Q., Xia, Y., 2024. Continual self-supervised learning: Towards universal multi-modal medical data representation learning, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11114–11124.
- Yuan, B., Zhao, D., 2024. A survey on continual semantic segmentation: Theory, challenge, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Zeng, Z., Peng, Z., Yang, X., Shen, W., 2024. Missing as masking: Arbitrary cross-modal feature reconstruction for incomplete multimodal brain tumor segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 424–433.
- Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., Zhang, Y., He, Z., Zheng, Y., 2022. mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 107–117.
- Zhao, Z., Yang, H., Sun, J., 2022. Modality-adaptive feature interaction for brain tumor segmentation with missing modalities, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 183–192.
- Zhong, W., Cong, C., Wang, Z., Yan, Z., Di Ieva, A., Liu, S., 2025. Fmm-diff: A feature mapping and merging diffusion model for mri generation with missing modality, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 227–236.
- Zhou, T., Canu, S., Vera, P., Ruan, S., 2021. Latent correlation representation learning for brain tumor segmentation with missing mri modalities. *IEEE Transactions on Image Processing* 30, 4263–4274.
- Zhu, S., Chen, Y., Chen, W., Wang, Y., Liu, C., Jiang, S., Qin, F., Wang, C., 2025. Bridging the gap in missing modalities: Leveraging knowledge distillation and style matching for brain tumor segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 95–106.