MORE SHOULD BE BETTER: INCREMENTAL SELF-DISTILLATION FOR MULTIMODAL BRAIN TUMOR SEGMENTATION WITH MISSING MODALITIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal magnetic resonance imaging provides complementary evidence for brain tumor segmentation, yet one or more modalities are frequently unavailable in clinical practice. A practical model must therefore remain effective for any available modality subset. Existing missing-modality methods largely optimize robustness to modality removal, but overlook the converse question: does the model reliably benefit when an additional modality becomes available? We argue that robustness alone is insufficient and introduce the principle of non-negative modality utility, i.e., enriching a valid modality subset should not worsen its information content for segmentation. Based on this principle, we propose a complementarity-guided incremental self-distillation framework. Specifically, it reformulates robust multimodal representation learning with missing modalities as a progressive multimodal self-distillation problem. Based on the non-negative modality-utility hypothesis, we organize modality subsets into a distillation path and encourage non-negative utility gains along its transitions, allowing representations learned from fewer modalities to approach the representational capacity of the complete-modality setting. Compared with directly learning a joint representation from a restricted modality subset, our incremental self-distillation strategy better preserves modality-specific discriminative information and improves adaptation to diverse missing-modality configurations. Extensive experiments across diverse missing-modality settings demonstrate the effectiveness and robustness of the proposed framework.

1 INTRODUCTION

Multimodal magnetic resonance imaging (MRI) is central to brain tumor analysis because different sequences reveal distinct and complementary characteristics of tumor tissue. For example, fluid-attenuated inversion recovery (FLAIR) makes edema and diffuse tumor-related abnormalities more visible; contrast-enhanced T1-weighted imaging is particularly informative for enhancing tumor; and T1- and T2-weighted images provide complementary anatomical and pathological context. Their joint use has substantially advanced automatic segmentation of clinically important tumor regions such as whole tumor, tumor core, and enhancing tumor (Menze et al., 2015). In practice, however, a complete MRI protocol cannot always be assumed. Modalities may be absent because of acquisition cost, varying scanner protocols, motion artifacts, contraindications to contrast agents, or retrospective data collection. A model that requires complete input is therefore difficult to deploy, motivating a single model that can handle any non-empty subset of the available modalities.

Existing approaches address missing modalities through image or feature completion, modality-invariant representation learning, modality dropout, and knowledge distillation. Early heterogeneous-input models aggregate statistics from the available modalities into a shared representation (Havaei et al., 2016), while more recent methods learn robust representations through masked multimodal pretraining (Liu et al., 2023). Cross-modal compositional self-distillation further transfers information from complete inputs to incomplete modality subsets and progressively simulates modality removal during training (Xie et al., 2026). Despite their differences, the prevailing objective remains removal-centric: *how much discriminative information can be preserved after modalities are removed?* This objective is essential, but it leaves a complementary question

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

unresolved: can the model reliably exploit the information provided by a modality when it becomes available? A model may tolerate missing inputs yet fail to benefit consistently from different clinically informative modalities, and may even produce worse predictions from a richer input.

Rather than viewing missing modalities solely as a problem of information preservation, we further ask whether a multimodal learner can consistently convert newly available information into task utility, as reflected by a reduction in task risk. Inspired by the confidence monotonicity view of multimodal calibration (Ma et al., 2023), we first note an ideal ordering: for two nested modality subsets $\mathcal{T} \subset \mathcal{S}$, the optimal task risk attainable from $\mathcal{S}$ can be no higher than that attainable from $\mathcal{T}$. The intuition is that $\mathcal{S}$ retains every input in $\mathcal{T}$, so any prediction mapping using only the modalities in $\mathcal{T}$ remains feasible when $\mathcal{S}$ is observed, while additional complementary modalities may further reduce task uncertainty. Motivated by this ideal ordering, our non-negative modality-utility hypothesis requires a learned model not to incur a higher task risk after modality enrichment. This property need not arise automatically when a finite-capacity model is jointly optimized over many modality configurations. The model may fit some configurations without learning to incorporate newly available information effectively in others. Consequently, a richer modality subset may yield lower task utility than a smaller subset contained within it. Reliable missing-modality learning therefore requires both preserving task-relevant information under modality absence and realizing non-negative utility gains under modality enrichment.

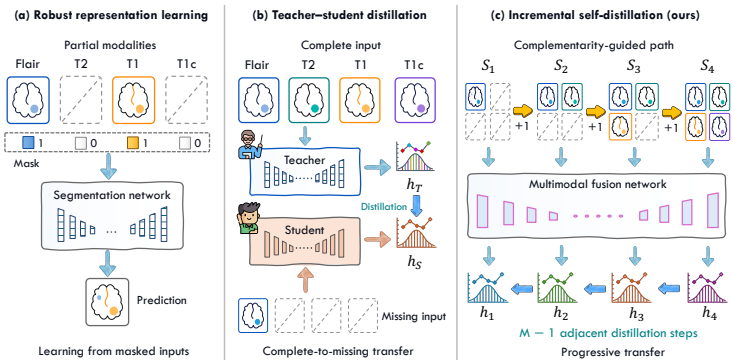

Fig. 1: Three missing multimodal learning paradigms.

Motivated by this dual requirement, we propose complementarity-guided incremental self-distillation (CISD). Specifically, we regard different modality subsets as distinct information states and organize them into a sample-adaptive path of nested subsets, from a single modality to the complete input. At each step, we use feature diversity as a tractable proxy for complementarity to determine which modality is added next, making the enrichment order responsive to the observed sample. Each state with more available modalities then provides an internal teaching target for its immediate predecessor, so that knowledge is transferred through intermediate states that narrow the modality-availability gap. In this way, task-relevant knowledge acquired through modality enrichment is propagated progressively toward states with fewer available modalities. The resulting path contains only $M - 1$ adjacent transfers for M modalities and requires no separate teacher network.

Incremental self-distillation (ISD) promotes information retention in states with fewer modalities, but it does not by itself ensure that an enriched state uses the modality just added. We therefore introduce modality-utility regularization to encourage the hypothesized information ordering. It uses sample-wise supervised losses as surrogates for task risk and penalizes an adjacent transition when the enriched state incurs a larger loss, without allowing the penalty to be reduced by degrading the state with fewer modalities. The two objectives play complementary roles along the same path: incremental self-distillation helps states with fewer modalities retain knowledge available in more complete observations, whereas modality-utility regularization encourages each enriched state to exploit its newly available information. Instantiated with a shared–specific encoder and a unified decoder, the framework accommodates all non-empty modality combinations within a single model and introduces no additional distillation branch at inference time. As shown in Fig. 1, we draw a schematic comparison of three representative imputation-free paradigms for missing-modality segmentation.

Our contributions are summarized as follows:

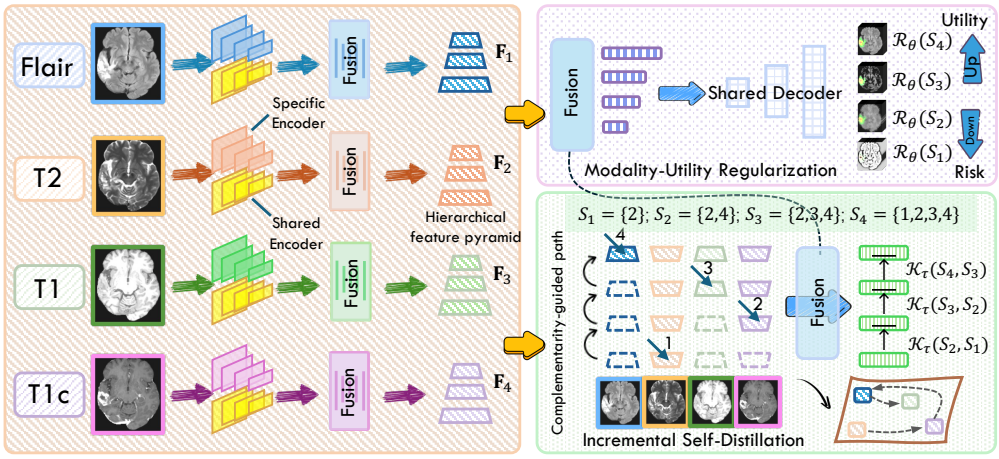

Fig. 2: Overview of the proposed complementarity-guided incremental self-distillation framework. Modality features are encoded and fused for segmentation with a shared decoder. Along a complementarity-guided path of nested subsets, richer states transfer knowledge to their immediate predecessors through adjacent self-distillation. Modality-utility regularization encourages each added modality to reduce the segmentation loss. The incremental path is actually constructed based on modal complementarity.

- We formulate missing-modality learning from a modality-utility perspective and establish the non-negative modality-utility hypothesis, complementing information preservation under modality absence with the requirement that modality enrichment should not reduce task utility.
- We propose CISD, which progressively transfers knowledge along a sample-adaptive path of nested modality subsets. A modality-utility regularizer further encourages each enriched state to exploit newly available information.
- Our unified framework accommodates arbitrary missing-modality configurations without introducing an auxiliary teacher branch. Extensive experiments on BraTS 2024 demonstrate state-of-the-art performance across diverse modality subsets.

2 Related Work

2.1 Learning with Missing Imaging Modalities

Existing methods for missing-modality medical image analysis broadly follow two paradigms. *Modality-completion methods* synthesize unavailable images from the observed modalities using generative adversarial networks or variational models (Sharma & Hamarneh, 2020; Hamghalam et al., 2021). Although the completed inputs allow a conventional multimodal model to operate unchanged, synthesis errors may propagate to the downstream task, and separate completion processes can become costly as missing patterns vary. *Representation-based methods* instead learn directly from the available inputs. HeMIS aggregates first- and second-order statistics of modality-specific latent features (Havaei et al., 2016); RFNet performs region-aware fusion to exploit cross-modal correlations (Ding et al., 2021); and mmFormer models both intra-modal features and inter-modal dependencies with modality-specific transformers (Zhang et al., 2022). More recent approaches focus on disentangling shared and modality-specific factors (Wang et al., 2023a) or using masked multimodal reconstruction to regularize representations across input configurations (Liu et al., 2023). These designs substantially improve robustness to incomplete observations, but they mostly treat different modality subsets as parallel input conditions rather than explicitly modeling how their information utility should change under modality enrichment.

2.2 KNOWLEDGE DISTILLATION FOR MISSING MODALITIES

As we know, knowledge distillation transfers predictive or representational knowledge from a capable teacher to a constrained student (Hinton et al., 2015). In missing-modality learning, a common strategy trains a full-modality network as the teacher and distills its knowledge into a unimodal or partial-modality student (Hu et al., 2020). Such teacher–student designs provide rich supervision but require an additional teacher and usually specialize the student to a predefined missing pattern. Further, self-distillation reduces this dependency by letting different input configurations of a shared network provide supervision to one another (Liu et al., 2023; Xie et al., 2026), thereby supporting a broader set of modality subsets without maintaining separate models.

Both paradigms implicitly assume that the configuration supplying the teaching signal contains more usable task information. This assumption is reasonable for an independently optimized full-modality teacher, but it need not hold among configurations of a jointly optimized shared model. Our CISD makes this relation explicit: it performs self-distillation between adjacent states on a complementarity-guided path, while modality-utility regularization encourages each enriched state to provide non-negative task utility over the state from which it is derived.

3 METHOD

3.1 PROBLEM FORMULATION AND MODALITY UTILITY

Let $\mathcal{M} = \{1, \ldots, M\}$ denote the complete modality set, where M is the number of modalities. A multimodal sample is written as (X, Y), where $X = \{x_m\}_{m \in \mathcal{M}}$ contains the modality images and Y is the segmentation target. At inference time, only a non-empty subset $\mathcal{S} \subseteq \mathcal{M}$ may be observed; we write $X_{\mathcal{S}} = \{x_m : m \in \mathcal{S}\}$ for its available inputs. Our goal is to learn a single model f_θ, parameterized by θ, that produces segmentation logits $o_{\mathcal{S}} = f_\theta(X_{\mathcal{S}})$ for every such subset. The corresponding foreground probabilities are $\hat{Y}_{\mathcal{S}} = \sigma(o_{\mathcal{S}})$, where σ denotes the element-wise sigmoid function. We measure the population task risk under $\mathcal{S}$ by

$$\mathcal{R}_\theta(\mathcal{S}) = \mathbb{E}_{(X,Y)}\left[\mathcal{L}_{\mathrm{SEG}}\left(f_\theta(X_{\mathcal{S}}), Y\right)\right]. \tag{1}$$

Here, $\mathbb{E}_{(X,Y)}$ denotes expectation over the data distribution and $\mathcal{L}_{\mathrm{SEG}}$ is a differentiable supervised segmentation loss, instantiated by soft Dice and voxel-wise binary cross-entropy in Equation 9.

Non-negative modality utility. Consider two nested subsets $\mathcal{T} \subset \mathcal{S}$ and write the enriched observation as $X_{\mathcal{S}} = (X_{\mathcal{T}}, X_{\mathcal{S}\setminus\mathcal{T}})$. Because $X_{\mathcal{T}}$ is contained in $X_{\mathcal{S}}$, any segmentation mapping that uses $X_{\mathcal{T}}$ remains feasible when $X_{\mathcal{S}}$ is observed: the same mapping can act on the $X_{\mathcal{T}}$ component while disregarding the added modalities $X_{\mathcal{S}\setminus\mathcal{T}}$. Let $\mathcal{H}_{\mathcal{A}}$ denote the set of admissible segmentation mappings from an observation $X_{\mathcal{A}}$ to a segmentation prediction, for any non-empty $\mathcal{A} \subseteq \mathcal{M}$, and let h denote an element of this set. The minimum risk achievable with the enriched observation therefore satisfies

$$\mathcal{R}^*(\mathcal{S}) = \inf_{h \in \mathcal{H}_{\mathcal{S}}} \mathbb{E}[\mathcal{L}_{\mathrm{seg}}(h(X_{\mathcal{S}}), Y)] \leq \inf_{h \in \mathcal{H}_{\mathcal{T}}} \mathbb{E}[\mathcal{L}_{\mathrm{seg}}(h(X_{\mathcal{T}}), Y)] = \mathcal{R}^*(\mathcal{T}). \tag{2}$$

Eq. (2) states that additional information has non-negative value at the optimum. Motivated by this relation, we introduce the following hypothesis for a learned multimodal model.

Hypothesis 3.1 (Non-negative modality utility)**.** *For any two valid modality subsets $\mathcal{T} \subset \mathcal{S}$, enriching the observation from $X_{\mathcal{T}}$ to $X_{\mathcal{S}}$ should not increase the task risk. We define the resulting modality utility as*

$$\mathcal{U}_\theta(\mathcal{S} \mid \mathcal{T}) \triangleq \mathcal{R}_\theta(\mathcal{T}) - \mathcal{R}_\theta(\mathcal{S}) \geq 0.$$

Positive utility indicates that the added modalities improve the prediction, whereas zero utility indicates that they are harmless. We therefore encourage the empirical risks of selected nested subsets to follow it during training.

3.2 Unified Representation for Arbitrary Modality Subsets

Unlike multimodal settings involving heterogeneous sources, brain MRI modalities observe the same anatomy while highlighting different tissue characteristics. A shared–specific encoding architecture captures this anatomical commonality while retaining contrast-dependent information. For modality m, we represent the shared representation $\mathbf{a}_m$, modality-specific representation $\mathbf{s}_m$, and fused feature pyramids $\mathbf{F}_m$ respectively as follows:

$$\mathbf{a}_m = E_{\mathrm{sh}}(x_m), \qquad \mathbf{s}_m = E_m^{\mathrm{sp}}(x_m), \qquad \mathbf{F}_m = \Phi(\mathrm{Concat}(\mathbf{a}_m, \mathbf{s}_m)), \tag{3}$$

where $\mathbf{F}_m = \{\mathbf{f}_m^{(l)} \in \mathbb{R}^{C_l \times D_l \times H_l \times W_l}\}_{l=1}^{L}$ contain L resolution levels. C_l denotes the channel and (D_l, H_l, W_l) is the spatial resolution at level l, which decreases as l increases. Concat performs channel-wise concatenation at corresponding pyramid levels, and Φ is a level-wise convolutional fusion operator. The shared encoder E_{sh} and Φ are reused across modalities, whereas E_m^{sp} retains modality-dependent parameters. For the convenience of parallel computing, each modality occupies a fixed feature slot, with unavailable input set to zero. Multimodal fusion and shared decoding then yield the prediction for subset $\mathcal{S}$:

$$\mathbf{Z}_{\mathcal{S}} = \underset{m=1}{\overset{M}{\mathrm{Concat}}} \left(\mathbb{I}[m \in \mathcal{S}]\mathbf{F}_m\right), \qquad o_{\mathcal{S}} = D(\mathbf{Z}_{\mathcal{S}}). \tag{4}$$

Here, $\mathbf{Z}_{\mathcal{S}} = \{\mathbf{z}_{\mathcal{S}}^{(l)} \in \mathbb{R}^{(MC_l)\times D_l \times H_l \times W_l}\}_{l=1}^{L}$ is the feature pyramid representing $\mathcal{S}$. $\mathbb{I}[\cdot]$ is the indicator function, concatenation is applied at corresponding pyramid levels, and D is the shared decoder. Within D, we define the deepest-level convolutional layer as the bottleneck layer B_θ to get cross-modal semantic representation $\mathbf{h}_{\mathcal{S}} = B_\theta(\mathbf{z}_{\mathcal{S}}^{(L)}) \in \mathbb{R}^{C \times D_L \times H_L \times W_L}$, reducing MC_L channels to $C = C_L$ without changing spatial resolution. We argue that this bottleneck feature is the semantic representative of subsequent hierarchical decoding, and thus it is regarded as the main object of the proposed self-distillation strategy.

3.3 Complementarity-Guided Incremental Self-Distillation

As we know, typical robust representation learning for missing-modality segmentation generally focuses on extracting discriminative evidence from each observed configuration. This treatment regards different missing patterns as parallel training conditions, but overlooks their nested structure: for $\mathcal{T} \subset \mathcal{S}$, the sparse observation $X_{\mathcal{T}}$ is obtained by removing modalities from the richer observation $X_{\mathcal{S}}$. Consequently, the progressive information degradation induced by modality removal is not explicitly modeled. Besides, teacher–student distillation provides a possible connection between these configurations, but conventional designs typically fix the complete-modality branch as the teacher and require every partial configuration to imitate it. For M modalities, this involves as many as $2^M - 1$ non-empty configurations in total, causing huge computational cost. More importantly, we argue that direct transfer across a severe modality-availability gap is unnecessarily difficult. A unimodal state, for example, is required to match the complete-modality state in a single step, despite the large gap in the discriminative information available to the two states. Such a large supervision gap may hinder knowledge transfer under extreme modality absence.

Incremental self-distillation (ISD). To bridge this gap, we model modality availability as a progressive enrichment process and introduce a transfer path with $M - 1$ adjacent steps. Specifically, we organize M subset states into a nested path $\mathcal{P} = (\mathcal{S}_1, \ldots, \mathcal{S}_M)$, where $|\mathcal{S}_k| = k$ and $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_M = \mathcal{M}$. Along each transition, the enriched state $\mathcal{S}_{k+1}$ teaches its predecessor $\mathcal{S}_k$ through their fused representations $\mathbf{h}_{\mathcal{S}_{k+1}}$ and $\mathbf{h}_{\mathcal{S}_k}$. These internal teaching states make the transfer a form of self-distillation, requiring neither an external teacher nor subset-specific branches.

To be specific, we propose to conduct self-distillation on these low-resolution fused responses to preserve their spatial structure. Let $V = D_L H_L W_L$ denote the number of spatial positions. For channel $c \in \{1, \ldots, C\}$ and position $u \in \{1, \ldots, V\}$, we define voxel-wise response as $q_{\mathcal{S}}^{(c,u)} = \sigma(h_{\mathcal{S}}^{(c,u)}/\tau)$, where $h_{\mathcal{S}}^{(c,u)}$ is an element of $\mathbf{h}_{\mathcal{S}}$ and $\tau > 0$ is the distillation temperature. Each response parameterizes a Bernoulli distribution independently, avoiding normalization across channels or spatial positions. For a richer subset $\mathcal{S}$ and a student subset $\mathcal{T} \subset \mathcal{S}$, the distillation discrepancy is

$$\mathcal{K}_\tau(\mathcal{S}, \mathcal{T}) = \frac{\tau^2}{CV} \sum_{c=1}^{C} \sum_{u=1}^{V} D_{\mathrm{KL}}\Big(\mathrm{Bern}(\mathrm{sg}[q_{\mathcal{S}}^{(c,u)}]) \,\|\, \mathrm{Bern}(q_{\mathcal{T}}^{(c,u)})\Big). \tag{5}$$

Here, $\mathrm{sg}[\cdot]$ stops gradients through the teaching response, $\mathrm{Bern}(a)$ denotes the binary distribution $(a, 1-a)$, and D_{KL} is Kullback–Leibler divergence; for probabilities $a, b \in (0,1)$, its binary form is $a\log(a/b) + (1-a)\log[(1-a)/(1-b)]$. By introducing the temperature scaling factor τ^2, our incremental self-distillation (ISD) loss is formulated as follows:

$$\mathcal{L}_{\mathrm{ISD}} = \frac{1}{M-1}\sum_{k=1}^{M-1} \mathcal{K}_\tau(\mathcal{S}_{k+1}, \mathcal{S}_k). \tag{6}$$

Noted that each transition updates its student toward a detached richer target and an intermediate state still receives gradients when serving as the student of the next transition. The path $\mathcal{P}$ thus provides a progressive bridge from richer to restricted observations through $M-1$ adjacent terms.

Complementarity-guided path. For the incremental path $\mathcal{P}$ defined above, a natural question remains: should its progression from $\mathcal{S}_1$ to $\mathcal{S}_M$ be determined randomly, or can it simply follow a fixed, predefined modality order? We argue that neither choice is appropriate. MRI modalities provide different evidence, and their complementarity also depends on the observed content. Both random and fixed orderings ignore this sample-dependent relation and may add a modality that is largely redundant with the current state, providing little new information for the adjacent transfer. We therefore determine the path from the encoded content of each sample. At each transition, we select the remaining modality with the largest average feature distance from those already included, as it is expected to contribute the least redundant information. Specifically, for each modality $i \in \mathcal{M}$, let $\mathbf{v}_i = \mathrm{vec}(\mathbf{f}_i^{(L)})$ denote its flattened deepest-level feature. The cosine distance between modalities i and j is $d(i,j) = 1 - \frac{\mathbf{v}_i^\top \mathbf{v}_j}{\|\mathbf{v}_i\|_2 \|\mathbf{v}_j\|_2}$, where $(\cdot)^\top$ denotes transpose and $\|\cdot\|_2$ is the Euclidean norm. Then, we initialize the path with the modality m_1 whose representation has the largest average distance from the others:

$$m_1 = \arg\max_{m\in\mathcal{M}} \frac{1}{M-1}\sum_{j\in\mathcal{M}\setminus\{m\}} d(m,j), \qquad \mathcal{S}_1 = \{m_1\}. \tag{7}$$

Given state $\mathcal{S}_k$, for $k \in \{1,\dots,M-1\}$, we select the next modality by its average distance from those already included:

$$m_{k+1} = \arg\max_{m\in\mathcal{M}\setminus\mathcal{S}_k} \frac{1}{|\mathcal{S}_k|}\sum_{j\in\mathcal{S}_k} d(m,j), \qquad \mathcal{S}_{k+1} = \mathcal{S}_k \cup \{m_{k+1}\}. \tag{8}$$

Repeating this greedy update yields a sample-adaptive nested path $\mathcal{P}$, where each transition adds the modality that is least redundant with the current subset under the feature-distance criterion. However, merely establishing a path is not enough. According to Hypothesis 3.1, the modality addition is encouraged to maintain or improve task utility, i.e., *Modality-Utility Regularization*.

3.4 Modality-Utility Regularization

Although $\mathcal{P}$ prioritizes minimally redundant modality additions, this criterion alone does not guarantee that the shared model can translate newly available information into a lower task risk. Revisiting Hypothesis 3.1, we therefore impose a margin-based utility constraint on every adjacent transition along $\mathcal{P}$. Specifically, for each subset $\mathcal{S}$, we define its sample-wise task loss as

$$r_\theta(\mathcal{S}) = \mathcal{L}_{\mathrm{SEG}}\big(f_\theta(X_\mathcal{S}), Y\big) = \mathcal{L}_{\mathrm{Dice}}(\sigma(o_\mathcal{S}), Y) + \mathcal{L}_{\mathrm{BCE}}(o_\mathcal{S}, Y), \tag{9}$$

where $\mathcal{L}_{\mathrm{Dice}}$ is the soft Dice loss evaluated on sigmoid probabilities and $\mathcal{L}_{\mathrm{BCE}}$ is voxel-wise binary cross-entropy with logits. Thus, $r_\theta(\mathcal{S})$ is the sample-wise realization of $\mathcal{L}_{\mathrm{SEG}}$ and provides a differentiable surrogate for the task risk $\mathcal{R}_\theta(\mathcal{S})$ in Eq. (1). For adjacent states along $\mathcal{P}$, the modality-utility regularizer is

$$\mathcal{L}_{\mathrm{MU}} = \frac{1}{M-1}\sum_{k=1}^{M-1} \left[r_\theta(\mathcal{S}_{k+1}) - \mathrm{sg}[r_\theta(\mathcal{S}_k)] + \delta\right]_+, \tag{10}$$

where $[z]_+ = \max(0,z)$ and $\delta \geq 0$ is a utility margin. We set $\delta = 0$, penalizing a transition whenever the enriched state has a larger task loss than its preceding state. The risk of $\mathcal{S}_k$ is held fixed in this regularization, preventing the constraint from being satisfied through directly degrading the utility of fewer modalities subset. We therefore formulate $\mathcal{L}_{\mathrm{ISD}}$ and $\mathcal{L}_{\mathrm{MU}}$ with two complementary goals: to progressively bring $\mathbf{h}_{\mathcal{S}_k}$ closer to $\mathbf{h}_{\mathcal{S}_{k+1}}$, and to encourage modality enrichment from $\mathcal{S}_k$ to $\mathcal{S}_{k+1}$ to yield non-negative task utility, respectively.

3.5 Overall Objective and Inference

Alongside the incremental path, we retain the full-to-partial supervision pattern of hierarchical modality self-distillation (HMSD), inspired by Xie et al. (2026), adapting it to the fused representations and Bernoulli discrepancy in Eq. (5). At each iteration, we sample a subset size r uniformly from $\{1, \ldots, M-1\}$ and define $\mathcal{B}_r = \{\mathcal{S} \subset \mathcal{M} : |\mathcal{S}| = r\}$ as the collection of all subsets of that size. As a result, we let the complete-modality state teach each of these subsets as follows:

$$\mathcal{L}_{\mathrm{HMSD}} = \frac{1}{|\mathcal{B}_r|} \sum_{\mathcal{S} \in \mathcal{B}_r} \mathcal{K}_\tau(\mathcal{M}, \mathcal{S}). \tag{11}$$

The HMSD broadens subset coverage beyond a single path, while ISD adds intermediate teaching states rather than relying exclusively on direct full-to-partial alignment. During the training, all samples provide the complete modality set, which are encoded once to construct $\mathcal{P}$ and its internal teaching states. In each iteration, we uniformly sample a non-empty availability subset $\mathcal{A} \subseteq \mathcal{M}$ and set $\mathcal{L}_{\mathrm{SEG}} = r_\theta(\mathcal{A})$ as the main segmentation loss. The complete objective $\mathcal{L}$ combines this supervised loss with global full-to-partial alignment, incremental self-distillation, and modality utility:

$$\mathcal{L} = \mathcal{L}_{\mathrm{SEG}} + \lambda_{\mathrm{H}} \mathcal{L}_{\mathrm{HMSD}} + \lambda_{\mathrm{C}} \mathcal{L}_{\mathrm{ISD}} + \lambda_{\mathrm{U}} \mathcal{L}_{\mathrm{MU}}. \tag{12}$$

Here, λ_{H}, λ_{C}, and λ_{U} are non-negative coefficients weighting HMSD, ISD, and modality-utility regularization, respectively. All auxiliary objectives are used only during training, and an arbitrary non-empty modality subset is encoded once and decoded by the shared network at inference, without external teacher, auxiliary distillation branch, or subset-specific model.

4 Experiments

In this section, we introduce the experimental setup, analysis of experimental results, and ablation study. For more supplementary experimental results, please refer to Appendix B and C.

4.1 Experimental Setup

Datasets, preprocessing, and evaluation. In experiments, we use the 1,350-case BraTS 2024 glioma MRI collection (Menze et al., 2015; de Verdier et al., 2024) and BraTS 2018 dataset (Menze et al., 2015; Bakas et al., 2017) with 285 labeled cases, with four modalities per case: FLAIR, T1, T1c (T1ce), and T2. Following existing methods (Zhu et al., 2026), we adopt an 80%/20% training–evaluation ratio to partition dataset. Following the public preprocessing protocol of existing methods (Zhu et al., 2026), for BraTS 2024, labels 1, 2, and 3 represent non-enhancing tumor core, surrounding non-enhancing FLAIR hyperintensity, and enhancing tissue, respectively; resection cavity (label 4) is merged with background. The reported regions are WT (labels 1–3), TC (labels 1 and 3), and ET (label 3), rather than the full four-region challenge target. For BraTS 2018, labels 1, 2, and 4 denote necrotic/non-enhancing tumor core, edema, and enhancing tumor, respectively; WT, TC, and ET correspond to labels $\{1, 2, 4\}$, $\{1, 4\}$, and $\{4\}$. We report Dice (%, higher is better) for regional overlap and 95th-percentile Hausdorff distance (HD95, mm, lower is better) for boundary accuracy on WT, TC, and ET across all 15 non-empty modality subsets. Overall scores average equally across subsets and regions. Implementation details are provided in Appendix A.

4.2 Comparison under Missing Modalities

To fairly evaluate the performance of our method, we introduce 14 advanced methods as competitors, i.e., LCKD (Wang et al., 2023b), ShaSpec (Wang et al., 2023a), MriSeg (Ting & Liu, 2024), MouseGAN++ (Yu et al., 2023), Hyper-GAE (Yang et al., 2023), SRMNet (Li et al., 2024), RFNet (Ding et al., 2021), mmFormer (Zhang et al., 2022), M^{2}FTrans (Shi et al., 2024), RMBTS (Chen et al., 2019), RA-HVED (Jeong et al., 2022), ACN (Wang et al., 2021), SMUNet (Azad et al., 2022), and MSTKDNet (Zhu et al., 2025). These methods span modality completion, robust representation learning, multi-task learning, and knowledge distillation. Table 1 reports Dice across the 15 modality combinations, with separate blocks for WT, TC, and ET. The corresponding HD95 results are provided in Table 3 in Appendix B.

From Table 1, our CISD achieves mean Dice scores of 83.32%, 62.94%, and 62.51% for WT, TC, and ET, respectively, giving a three-region average of 69.59%. With all four modalities, Dice reaches

Table 1: Dice (%) on BraTS 2024 under 15 modality combinations. ● / ○: available / missing. Avg. and Std. are recomputed from the 15 displayed entries. Bold denotes the highest Dice in each region.

Region	FLAIR	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	Avg.	Std.
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●		
	T1c	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●		
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●		
WT	LCKD	**78.99**	79.96	**75.62**	77.98	78.68	78.45	79.74	79.97	80.78	81.43	82.03	82.55	82.95	82.91	83.28	80.35	2.13
	ShaSpec	67.13	49.16	41.58	58.18	72.89	72.17	74.82	52.13	64.11	61.14	74.29	76.57	76.03	64.01	76.67	65.39	10.70
	MriSeg	78.82	55.90	64.28	64.81	81.81	70.84	72.80	**81.44**	82.16	66.84	73.93	83.04	83.46	83.38	83.74	75.15	8.60
	MouseGAN++	64.38	66.95	57.79	61.45	75.35	61.22	70.58	78.98	77.67	66.43	79.47	63.63	65.49	68.77	80.77	69.26	7.22
	Hyper-GAE	75.93	60.91	70.05	66.94	74.82	72.64	75.03	70.91	76.87	60.73	78.07	78.55	77.68	74.77	76.98	72.73	5.62
	SRMNet	64.75	49.74	43.81	73.66	72.04	58.25	75.50	67.39	79.52	78.60	78.50	79.26	81.72	72.69	81.12	70.44	11.27
	RFNet	73.38	67.57	68.12	81.22	76.83	70.83	83.09	76.27	83.33	83.87	84.22	83.83	84.50	77.23	84.56	78.59	5.98
	mmFormer	64.77	46.71	57.19	76.68	67.65	55.83	78.05	68.17	78.83	77.57	77.27	78.86	78.87	67.77	77.96	70.15	9.89
	M^{2}FTrans	69.87	58.81	59.42	79.56	73.32	65.57	81.01	73.61	81.71	81.76	82.09	82.17	82.62	74.76	82.70	75.27	8.13
	RMBTS	68.63	50.67	51.96	63.38	73.62	59.93	76.06	73.62	75.75	75.00	78.82	78.65	78.89	74.60	79.64	70.61	9.34
	RA-HVED	50.96	**80.75**	52.13	65.55	**83.79**	**82.75**	69.37	58.11	70.42	83.31	83.62	70.05	84.85	**84.30**	84.86	73.65	11.89
	ACN	68.95	61.95	57.93	79.49	73.23	64.34	81.99	74.27	81.10	81.94	82.08	82.81	81.64	74.69	82.45	75.26	8.08
	SMUNet	75.20	70.56	69.19	83.64	77.74	72.95	85.17	78.31	84.96	84.81	85.46	85.64	85.40	78.34	85.81	80.21	5.78
	MSTKDNet	77.29	72.88	73.11	84.76	80.32	74.89	85.35	79.12	86.07	87.17	87.02	85.92	87.45	80.26	86.98	81.91	5.20
	Ours	78.40	73.11	74.10	**85.61**	81.36	76.43	**87.32**	80.73	**87.74**	**88.40**	**88.56**	**88.21**	**89.04**	81.75	**89.07**	**83.32**	5.52
TC	LCKD	27.59	27.51	35.91	33.82	33.18	37.08	39.69	41.77	40.57	39.61	41.20	42.47	41.67	42.79	43.73	37.91	5.10
	ShaSpec	28.52	43.62	32.13	38.08	58.28	58.09	60.74	56.98	52.59	50.17	61.00	63.61	61.32	55.50	62.67	52.22	11.01
	MriSeg	18.65	50.41	17.08	25.89	55.10	52.45	27.45	23.52	31.36	55.04	62.10	31.27	61.50	60.09	62.72	42.31	16.87
	MouseGAN++	24.63	48.28	29.13	21.71	64.70	58.67	50.59	34.63	35.45	60.65	57.16	48.80	59.20	61.42	63.53	47.90	14.42
	Hyper-GAE	24.87	33.61	39.72	28.28	57.94	57.16	54.87	54.63	**59.73**	60.99	60.85	59.77	60.85	59.77	62.95	51.73	12.65
	SRMNet	24.12	41.21	24.84	24.62	52.35	47.04	31.23	27.99	28.29	48.73	57.38	31.65	57.38	54.40	62.92	40.94	13.54
	RFNet	29.66	57.12	31.02	24.87	62.21	61.54	33.42	34.47	32.53	60.36	63.52	35.32	63.52	64.76	64.01	47.89	15.48
	mmFormer	26.52	41.08	22.74	22.01	49.27	47.23	26.92	28.95	29.34	51.39	54.81	31.03	53.26	50.63	55.71	39.39	12.41
	M^{2}FTrans	22.85	47.06	25.31	23.66	57.36	55.71	30.72	29.90	27.87	56.97	63.21	31.07	62.53	60.47	60.28	43.66	15.82
	RMBTS	23.38	41.51	42.49	24.63	41.11	49.56	57.38	60.64	33.26	33.89	57.10	65.99	46.62	62.09	64.26	46.93	13.65
	RA-HVED	**54.57**	20.38	10.40	15.86	55.55	21.22	13.19	48.99	57.62	23.03	21.41	58.48	59.78	61.56	63.23	39.02	20.19
	ACN	27.91	24.66	58.74	22.19	31.17	64.46	64.56	64.50	31.18	30.18	67.89	67.72	30.78	65.60	67.79	47.96	18.64
	SMUNet	26.24	30.32	**64.26**	28.81	30.43	69.31	**67.70**	**67.98**	31.65	31.73	71.14	**69.79**	32.03	70.97	71.91	50.95	19.55
	MSTKDNet	36.96	65.76	34.84	34.50	68.33	68.75	37.63	39.22	40.30	65.23	70.75	40.23	67.21	70.27	70.57	54.04	15.46
	Ours	42.97	**71.83**	42.32	**46.02**	**74.82**	**77.76**	48.16	48.40	51.59	**76.65**	**78.87**	51.57	**76.60**	**78.20**	**78.34**	**62.94**	14.91
ET	LCKD	24.44	24.39	33.82	31.36	30.59	34.93	37.83	40.09	38.69	37.60	39.37	40.77	39.85	41.09	42.12	35.80	5.58
	ShaSpec	27.93	31.76	35.99	26.63	48.70	53.90	48.10	61.39	39.44	55.27	66.43	51.30	57.27	59.16	68.46	48.78	13.11
	MriSeg	14.93	50.98	13.94	21.82	56.37	53.08	24.07	19.66	28.33	56.81	64.23	28.25	64.20	62.11	65.55	41.62	19.49
	MouseGAN++	18.48	49.70	27.37	22.29	58.19	59.03	30.81	45.14	57.37	58.64	59.70	50.81	61.01	61.18	62.04	48.12	15.00
	Hyper-GAE	12.65	28.06	28.39	20.46	34.58	43.21	62.65	63.21	24.52	43.24	66.17	63.04	53.44	54.46	62.95	44.07	17.57
	SRMNet	10.91	42.23	11.71	11.10	53.52	48.39	16.71	15.57	14.75	49.49	53.86	17.76	51.82	54.40	54.59	33.79	18.75
	RFNet	26.30	58.05	27.92	21.93	63.19	62.95	29.92	31.20	29.47	61.73	65.28	32.15	64.45	66.14	65.80	47.10	17.71
	mmFormer	22.40	41.31	21.79	19.12	50.27	48.06	23.51	25.28	23.86	52.40	56.08	26.76	55.24	52.49	57.53	38.41	14.71
	M^{2}FTrans	22.85	48.20	22.60	23.66	57.36	57.23	26.52	26.98	24.43	58.87	63.21	28.75	62.53	62.53	63.51	43.28	17.42
	RMBTS	23.38	26.19	46.41	24.68	30.37	53.93	57.38	49.29	28.26	27.33	57.54	58.97	59.68	61.23	63.80	44.56	15.22
	RA-HVED	**53.54**	15.79	8.83	12.64	57.87	18.06	10.72	49.97	**58.07**	18.64	17.46	59.75	61.67	61.94	61.86	37.79	22.04
	ACN	25.61	21.23	59.51	19.45	27.27	65.14	65.35	65.14	28.01	26.61	69.42	67.72	27.90	66.33	68.18	46.86	20.52
	SMUNet	29.75	34.22	**63.93**	33.36	35.01	67.82	**67.70**	**67.98**	35.25	35.59	70.34	**68.89**	35.46	70.26	71.09	52.44	17.29
	MSTKDNet	33.98	65.76	32.43	31.87	68.23	69.35	35.16	36.43	37.51	65.23	70.75	38.03	71.16	71.35	71.16	53.23	17.15
	Ours	40.45	**72.97**	40.17	**44.08**	**75.83**	**78.80**	46.14	46.07	49.88	**77.71**	**79.89**	49.62	**77.56**	**79.14**	**79.31**	**62.51**	16.45

89.07%/78.34%/79.31%, averaging 82.24%. Compared with MSTKDNet, the strongest baseline in terms of mean Dice for all three regions, our method improves the mean Dice score by 1.41, 8.90, and 9.28 percentage points for WT, TC, and ET, respectively. The larger gains on TC and ET highlight its advantage in tumor-subregion segmentation under missing modalities. Beyond the overall gains, the results reveal region-specific dependence on modality composition: FLAIR alone yields higher WT Dice than T1c alone, whereas T1c is more informative for TC and ET. Thus, the value of an additional modality depends on the target region and the evidence already available, rather than simply on the number of inputs. This observation supports the motivation for complementarity-guided transfer, while the remaining variation across configurations indicates that missing informative modalities is still challenging.

4.3 Ablation Study

Table 2 evaluates five component configurations under the same 200-epoch training budget, using segmentation-only training (SEG) as the baseline. Both HMSD and ISD increase the mean Dice score and reduce HD95 compared with SEG, supporting the value of transferring informa-

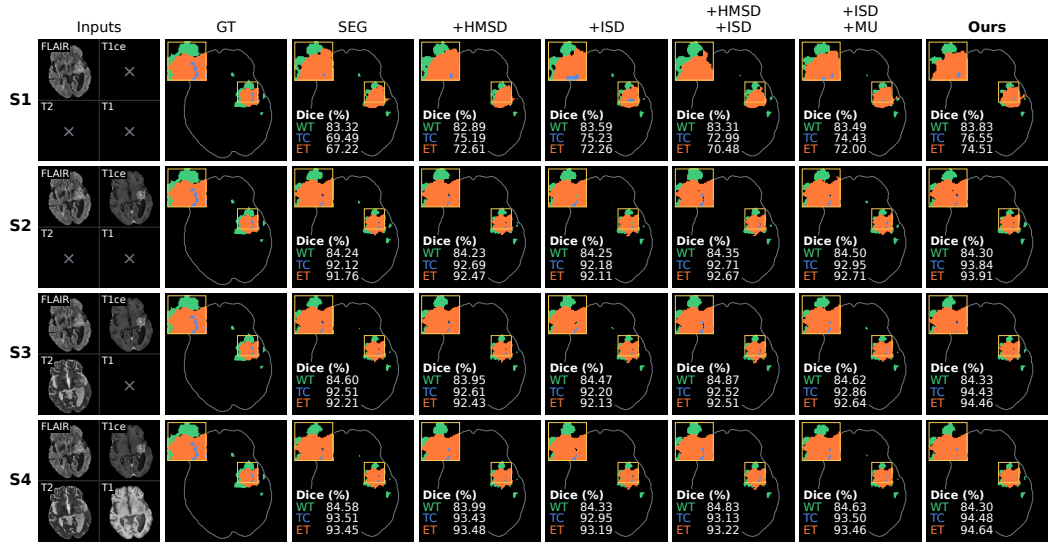

Fig. 3: Qualitative comparison on a BraTS 2024 case. S1–S4 progressively add FLAIR, T1c, T2, and T1 and crosses indicate missing modalities. Green/blue/orange denote (WT,TC), (TC, ET), and ET; yellow boxes mark the zoomed regions.

tion across modality subsets. Adding MU to ISD further improves both metrics, yielding the best configuration among the five ablations. Compared with SEG, ISD+MU improves the mean Dice score by 1.45 percentage points and reduces HD95 by 3.33 mm. This result supports coupling incremental representation transfer with a task-level utility constraint, so that alignment between neighboring modality states is accompanied by supervision on their segmentation utility. Combining HMSD and ISD alone does not outperform either term individually, indicating that the benefits do not follow simply from adding more distillation objectives. In Fig. 3, we complement this quantitative analysis with a case containing all three tissue classes. The shared slice and local zoom enable direct comparison of tumor extent and internal boundaries across S1–S4. Our CISD achieves the highest three-region mean Dice in each displayed subset, and the ablation ordering broadly follows the aggregate trend. The example illustrates the final model's potential across different input subsets, rather than estimating typical case-level performance.

Table 2: Component ablation on BraTS 2024. Results are averaged over 15 modality combinations and three tumor regions.

Configuration	Dice (%)	HD95 (mm)
SEG	59.65	64.66
SEG + HMSD	60.44	61.74
SEG + ISD	60.50	61.71
SEG + HMSD + ISD	60.20	62.95
SEG + ISD + MU	61.10	61.33

5 Conclusion

In this paper, we present a complementarity-guided incremental self-distillation framework for brain tumor segmentation with missing MRI modalities. By organizing modality subsets into a nested transfer path, the framework introduces intermediate teaching states between restricted and richer inputs, while modality-utility regularization encourages beneficial information transfer along adjacent transitions. A shared network supports all non-empty input subsets, and the final model combines this framework with lesion-presence supervision and inference-time gating. Evaluation across 15 modality configurations on BraTS 2024, complemented by BraTS 2018 results, demonstrates its ability to handle diverse input availability. On BraTS 2024, the method achieves the highest average Dice among the compared methods for WT, TC, and ET, with particularly pronounced gains in the tumor subregions. It needs to be admitted that the performance remains sensitive to which modalities are missing. Future work will focus on reducing this dependence and validating generalization on independent cohorts.

AI USE STATEMENT

We used generative AI tools for language polishing of the manuscript, including improving grammar, clarity, and readability. This use was limited to editing the presentation of the paper. The authors take full responsibility for the final content of this work, including all text, claims, and conclusions.

REFERENCES

Reza Azad, Nika Khosravi, and Dorit Merhof. SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities. In *Proceedings of the 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research*, pp. 48–62, 2022. URL https://proceedings.mlr.press/v172/azad22a.html.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4 (1):170117, 2017.

Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 447–456, 2019. doi: 10.1007/978-3-030-32248-9_50.

Maria Correia de Verdier, Rachit Saluja, Louis Gagnon, Dominic LaBella, Ujjwall Baid, Nourel Hoda Tahon, Martha Foltyn-Dumitru, Jikai Zhang, Maram Alafif, Saif Baig, et al. The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI. *arXiv preprint arXiv:2405.18368*, 2024.

Yuhang Ding, Xin Yu, and Yi Yang. RFNet: Region-aware fusion network for incomplete multimodal brain tumor segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3975–3984, 2021.

Mohammad Hamghalam, Alejandro F. Frangi, Baiying Lei, and Amber L. Simpson. Modality completion via gaussian process prior variational autoencoders for multi-modal glioma segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 442–452, 2021.

Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. HeMIS: Hetero-modal image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 469–477, 2016.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Minhao Hu, Matthis Maillard, Ya Zhang, Tommaso Ciceri, Giammarco La Barbera, Isabelle Bloch, and Pietro Gori. Knowledge distillation from multi-modal to mono-modal segmentation networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 772–781. Springer, 2020.

Seungwan Jeong, Hwanho Cho, Junmo Kwon, and Hyunjin Park. Region-of-interest attentive heteromodal variational encoder-decoder for segmentation with missing modalities. In *Proceedings of the Asian Conference on Computer Vision*, pp. 3707–3723, 2022. URL https://openaccess.thecvf.com/content/ACCV2022/html/Jeong_Region-of-interest_Attentive_Heteromodal_Variational_Encoder-Decoder_for_Segmentation_with_Missing_Modalities_ACCV_2022_paper.html.

Zhiyuan Li, Yafei Zhang, Huafeng Li, Yi Chai, and Yushi Yang. Deformation-aware and reconstruction-driven multimodal representation learning for brain tumor segmentation with missing modalities. *Biomedical Signal Processing and Control*, 91:106012, 2024. doi: 10.1016/j.bspc.2024.106012.

Hong Liu, Dong Wei, Donghuan Lu, Jinghan Sun, Liansheng Wang, and Yefeng Zheng. M3ae: Multimodal representation learning for brain tumor segmentation with missing modalities. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 1657–1665, 2023.

Huan Ma, Qingyang Zhang, Changqing Zhang, Bingzhe Wu, Huazhu Fu, Joey Tianyi Zhou, and Qinghua Hu. Calibrating multimodal learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23429–23450, 2023.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34 (10):1993–2024, 2015.

Anmol Sharma and Ghassan Hamarneh. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *IEEE Transactions on Medical Imaging*, 39(4):1170–1183, 2020.

Junjie Shi, Li Yu, Qimin Cheng, Xin Yang, Kwang-Ting Cheng, and Zengqiang Yan. M^2FTrans: Modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, 28(1):379–390, 2024. doi: 10.1109/JBHI.2023.3326151.

Hsienchih Ting and Manhua Liu. Multimodal transformer of incomplete MRI data for brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, 28(1):89–99, 2024. doi: 10.1109/JBHI.2023.3286689.

Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, and Gustavo Carneiro. Multi-modal learning with missing modality via shared-specific feature modelling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15878–15887, 2023a.

Hu Wang, Congbo Ma, Jianpeng Zhang, Yuan Zhang, Jodie Avery, Louise Hull, and Gustavo Carneiro. Learnable cross-modal knowledge distillation for multi-modal learning with missing modality. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 216–226, 2023b. doi: 10.1007/978-3-031-43901-8_21.

Yixin Wang, Yang Zhang, Yang Liu, Zihao Lin, Jiang Tian, Cheng Zhong, Zhongchao Shi, Jianping Fan, and Zhiqiang He. ACN: Adversarial co-training network for brain tumor segmentation with missing modalities. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 410–420, 2021. URL https://arxiv.org/abs/2106.14591.

Dongqing Xie, Yonghuang Wu, Zisheng Ai, Jun Min, Zhencun Jiang, Shaojin Geng, and Lei Wang. CCSD: Cross-modal compositional self-distillation for robust brain tumor segmentation with missing modalities. *Knowledge-Based Systems*, 346:116136, 2026.

Heran Yang, Jian Sun, and Zongben Xu. Learning unified hyper-network for multi-modal MR image synthesis and tumor segmentation with missing modalities. *IEEE Transactions on Medical Imaging*, 42(12):3678–3689, 2023. doi: 10.1109/TMI.2023.3301934.

Ziqi Yu, Xiaoyang Han, Shengjie Zhang, Jianfeng Feng, Tingying Peng, and Xiao-Yong Zhang. MouseGAN++: Unsupervised disentanglement and contrastive representation for multiple MRI modalities synthesis and structural segmentation of mouse brain. *IEEE Transactions on Medical Imaging*, 42(4):1197–1209, 2023. doi: 10.1109/TMI.2022.3225528.

Yao Zhang, Nanjun He, Jiawei Yang, Yuexiang Li, Dong Wei, Yawen Huang, Yang Zhang, Zhiqiang He, and Yefeng Zheng. mmFormer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 107–117, 2022.

Shenghao Zhu, Yifei Chen, Weihong Chen, Yuanhan Wang, Chang Liu, Shuo Jiang, Feiwei Qin, and Changmiao Wang. Bridging the gap in missing modalities: Leveraging knowledge distillation and style matching for brain tumor segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 95–106, 2025. doi: 10.1007/978-3-032-04984-1_10.

Shenghao Zhu, Yifei Chen, Weihong Chen, Shuo Jiang, Guanyu Zhou, Yuanhan Wang, Feiwei Qin, Changmiao Wang, and Qiyuan Tian. No modality left behind: Adapting to missing modalities via knowledge distillation for brain tumor segmentation. *Medical Image Analysis*, 112:104108, 2026. doi: 10.1016/j.media.2026.104108.

A Implementation Details

We implement our framework in PyTorch and train each model on a single NVIDIA RTX 4090 GPU with mixed precision and a batch size of 1. We use Adam with an initial learning rate of 10^{-4} and weight decay of 10^{-5}, and decrease the learning rate with a polynomial schedule of power 0.9. The distillation temperature is set to $\tau = 4$, and the loss weights are $\lambda_{\mathrm{H}} = 1$, $\lambda_{\mathrm{C}} = 0.1$, and $\lambda_{\mathrm{U}} = 0.5$, with utility margin $\delta = 0$. Although all four modalities are available during training, we uniformly sample one of the 15 non-empty subsets for the main supervised branch.

During training, we train from scratch for 400 epochs with all supervision. At inference, we obtain binary tissue masks by thresholding sigmoid outputs at > 0.5. In the final model, a presence probability of ≤ 0.5 additionally suppresses the entire corresponding tissue mask before the WT, TC, and ET regions are constructed. Gating relies solely on the model's predictions, with no ground-truth information or presence-based reweighting of Dice. Besides, we use full-precision inference and keep the thresholds fixed without tuning and encoded features are reused across modality subsets to avoid redundant computation.

B Additional HD95 Results on BraTS 2024

Table 3 reports HD95 under the same 15 modality combinations as the main Dice comparison. Our CISD achieves regional averages of 5.32, 60.19, and 59.09 mm for WT, TC, and ET, respectively, achieving the lowest mean HD95 among all compared methods in each region. Compared with the strongest baseline for each regional average, our method reduces HD95 by 0.91 mm for WT (LCKD), 10.36 mm for TC (RFNet), and 12.73 mm for ET (RFNet). It also attains the lowest HD95 in 12 of the 15 modality configurations for WT and 10 for each of TC and ET, showing that the advantage extends across a broad range of missing-input settings rather than relying on a few favorable combinations. Together with the Dice results, these findings highlight improvements in both regional overlap and distance-based segmentation quality, with particularly pronounced average reductions for TC and ET. The higher TC/ET values reflect both boundary errors and region-level failures: when exactly one mask is empty, our evaluation assigns a fixed penalty of 373.13 mm.

C Additional Results on BraTS 2018

In Tables 4 and 5, we report experimental results regarding Dice and HD95 under 15 modality combinations. Our model is trained on 228 cases and evaluated on 57 grade-stratified holdout cases, using a $160 \times 192 \times 128$ center crop. WT, TC, and ET correspond to labels $\{1, 2, 4\}$, $\{1, 4\}$, and $\{4\}$, respectively. Our method achieves mean Dice of 74.54% and HD95 of 18.76 mm across the three regions. Regional Dice scores are 85.34/75.43/62.86%, and HD95 values are 8.54/8.76/38.98 mm for WT/TC/ET. With all four modalities, Dice reaches 90.21/84.25/77.87%; As a result, our method has also achieved competitive performance on the small-scale BraTS 2018 dataset.

D Hyperparameter Sensitivity Analysis

Analysis design. We examine λ_{H}, λ_{C}, and λ_{U}, which control hierarchical self-distillation, incremental self-distillation (ISD), and modality-utility regularization, respectively. Starting from the reference configuration $(\lambda_{\mathrm{H}}, \lambda_{\mathrm{C}}, \lambda_{\mathrm{U}}) = (1, 0.1, 0.5)$, the hyperparameter sensitivity analysis varies each coefficient over $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$ while keeping the other two fixed. For each setting, Dice and HD95 will be averaged over all 15 non-empty modality subsets and reported separately for WT, TC, and ET. The comparisons will use the same training budget, data split, random seed, and checkpoint-selection rule.

Sensitivity to λ_{C}. The analysis of λ_{C} focuses on the balance between cross-subset transfer and segmentation supervision (Fig. 4). A small value may provide too little guidance between neighboring modality states, whereas a large value may overemphasize representation alignment at the expense of the segmentation objective. The illustrative curves in Fig. 4 are constructed to peak in Dice and reach their lowest HD95 at the configured value $\lambda_{\mathrm{C}} = 0.1$ across WT, TC, and ET.

Table 3: HD95 (mm) on BraTS 2024 under 15 modality combinations (lower is better). ● / ○: available / missing. Avg. and Std. denote the mean and population standard deviation across combinations. Bold marks the lowest HD95 in each region and column.

Region	FLAIR	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	Avg.	Std.
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●		
	T1c	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●		
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●		
WT	LCKD	7.07	**6.89**	**7.28**	6.89	6.55	**6.70**	6.41	6.29	6.11	5.88	5.69	5.54	5.40	5.43	5.32	6.23	0.64
	ShaSpec	28.84	40.88	47.63	25.32	23.85	22.57	19.72	44.79	23.95	30.51	20.97	16.57	16.76	29.69	14.29	27.09	9.88
	MriSeg	10.61	23.63	12.72	12.27	8.40	12.97	8.54	8.26	6.48	11.81	9.88	6.23	7.80	6.37	7.22	10.21	4.24
	MouseGAN++	27.50	27.02	13.69	21.72	15.61	14.98	13.85	19.90	15.46	12.21	14.62	18.09	16.67	13.94	11.25	17.10	4.78
	Hyper-GAE	19.00	17.64	24.78	27.01	22.46	22.61	17.70	24.71	13.92	20.19	13.10	13.04	17.92	16.34	14.31	18.98	4.37
	SRMNet	17.76	34.60	50.23	24.57	13.00	29.53	22.59	18.01	14.20	18.23	17.32	15.49	17.32	14.35	13.37	21.37	9.74
	RFNet	7.50	12.58	11.69	10.31	7.75	10.59	7.35	6.76	6.68	8.40	7.82	6.22	7.44	7.83	7.47	8.43	1.86
	mmFormer	16.39	55.86	30.99	18.20	18.31	38.98	14.09	14.40	10.89	19.54	15.97	11.21	11.88	17.62	12.10	20.43	12.00
	M^{2}FTrans	17.85	27.96	20.37	14.18	12.76	15.70	10.75	10.42	12.15	10.98	10.06	9.36	10.63	10.25	9.06	13.50	4.99
	RMBTS	26.33	52.62	53.01	23.53	24.43	46.46	19.95	22.67	18.04	25.41	24.51	17.97	24.51	23.34	18.34	28.07	11.68
	RA-HVED	27.82	12.28	21.76	15.84	6.79	8.39	11.34	17.21	10.99	8.03	6.16	9.51	5.69	5.76	5.45	11.53	6.34
	ACN	17.60	18.79	18.80	10.91	13.97	18.65	10.94	11.85	10.43	9.62	9.62	8.45	11.24	11.01	9.92	12.79	3.63
	SMUNet	8.64	15.65	11.81	8.78	10.17	11.37	6.00	8.48	6.71	7.53	7.30	5.63	6.57	8.53	5.55	8.58	2.65
	MSTKDNet	8.25	12.95	11.80	7.39	8.36	11.09	5.31	8.54	5.33	6.17	6.26	5.06	4.80	8.32	6.27	7.73	2.46
	Ours	**6.13**	9.06	8.74	**4.94**	**5.58**	6.93	**3.97**	**5.43**	**3.79**	**4.17**	**4.22**	**3.62**	**4.06**	**5.38**	**3.82**	**5.32**	1.68
TC	LCKD	86.47	91.46	85.20	88.98	89.18	85.27	82.48	80.19	81.88	82.76	81.20	79.52	80.41	79.82	78.47	83.55	3.87
	ShaSpec	104.95	109.58	97.65	102.02	98.64	94.58	86.54	84.72	82.38	89.67	82.97	78.47	87.15	74.60	58.85	88.85	12.60
	MriSeg	109.43	89.07	109.16	99.65	85.00	91.95	98.96	103.91	84.12	88.24	72.82	83.67	72.37	70.80	68.53	88.51	13.22
	MouseGAN++	115.43	109.02	100.73	99.22	79.41	80.71	83.90	95.58	81.13	76.00	67.53	78.87	66.28	68.87	64.56	84.48	15.45
	Hyper-GAE	110.12	92.11	93.48	95.02	96.70	93.53	88.06	94.03	99.06	88.76	78.19	81.62	87.76	78.02	63.35	89.32	10.55
	SRMNet	111.16	107.32	102.08	105.76	89.02	90.51	93.11	90.75	89.22	82.31	92.51	88.14	85.19	80.78	68.50	91.76	10.79
	RFNet	89.83	74.04	85.41	97.55	64.45	57.31	73.25	63.85	73.57	66.89	62.02	74.30	60.51	55.98	59.30	70.55	11.96
	mmFormer	97.35	125.43	110.61	99.33	96.52	109.14	92.58	92.18	85.20	89.49	82.44	85.42	79.79	94.48	73.00	94.20	12.89
	M^{2}FTrans	88.02	104.19	106.17	95.05	76.97	83.12	85.23	81.53	91.07	72.26	63.80	71.86	64.31	64.34	62.28	80.68	13.95
	RMBTS	115.96	113.92	138.67	116.04	106.35	117.74	115.17	111.18	109.46	110.02	106.26	111.69	106.87	104.43	103.09	112.46	8.27
	RA-HVED	**86.06**	117.79	142.20	132.76	83.36	117.95	139.68	116.13	71.00	124.41	124.46	76.75	70.27	69.91	63.51	102.42	27.58
	ACN	94.26	103.15	70.69	102.24	93.69	66.48	71.62	61.48	95.67	95.05	57.54	**53.88**	101.33	64.17	57.44	79.25	18.20
	SMUNet	105.65	83.34	**65.37**	92.80	89.14	65.48	**62.26**	**59.96**	93.31	90.76	59.22	60.72	92.94	58.19	56.45	75.71	16.48
	MSTKDNet	94.05	66.43	90.75	91.87	68.75	63.75	81.12	67.25	76.98	73.94	60.20	84.77	68.31	56.10	60.38	73.64	11.94
	Ours	89.55	**52.66**	96.45	**66.68**	**48.94**	**39.53**	80.59	83.53	**67.05**	**40.58**	**39.98**	70.45	**42.80**	**41.65**	**42.40**	**60.19**	19.49
ET	LCKD	95.13	98.02	90.47	93.98	94.13	89.16	86.39	83.80	85.38	86.19	84.46	82.75	83.61	82.51	80.88	87.79	5.18
	ShaSpec	106.58	113.93	94.17	95.87	82.21	85.51	87.14	90.73	89.95	78.1	71.17	76.11	75.29	71.86	56.64	85.02	14.06
	MriSeg	111.66	91.51	112.03	103.45	84.43	94.23	99.16	104.04	84.34	85.39	69.84	83.71	69.32	67.96	65.41	88.43	15.16
	MouseGAN++	97.14	94.31	100.54	99.22	99.78	90.22	96.62	95.58	83.26	85.62	88.12	85.23	73.17	74.65	70.23	88.91	9.73
	Hyper-GAE	98.06	109.4	100.92	94.22	96.7	93.53	88.06	94.03	99.06	88.76	87.76	74.01	72.46	73.19	69.38	89.30	11.57
	SRMNet	110.23	115.98	123.98	114.41	87.34	109.16	112.08	110.06	107.69	100.97	90.65	107.51	93.89	88.94	86.86	103.98	11.35
	RFNet	89.83	74.04	85.41	97.55	64.45	57.31	73.25	63.85	85.41	66.89	62.02	81.54	60.51	55.98	59.3	71.82	12.76
	mmFormer	97.35	125.76	107.56	99.33	98.32	109.94	92.34	91.93	84.97	91.04	83.72	84.75	79.79	93.77	74.75	94.35	12.55
	M^{2}FTrans	**88.02**	104.19	106.17	95.05	76.97	83.12	85.23	81.53	91.07	72.26	63.8	74.78	64.31	64.34	62.28	80.87	13.85
	RMBTS	117.18	114.76	138.71	118.13	107.63	117.93	116.7	112.81	110.94	110.02	106.26	111.69	106.87	104.43	104.32	113.23	8.25
	RA-HVED	92.28	124.48	142.72	132.88	76.46	117.88	134.33	116.05	70.57	124.23	124.91	69.70	63.43	69.46	69.56	101.93	28.29
	ACN	94.07	103.15	70.69	102.28	98.66	66.48	**71.62**	**61.48**	95.67	97.43	54.27	**53.88**	99.46	64.17	58.36	79.44	18.70
	SMUNet	107.62	87.29	**67.15**	95.67	97.09	59.49	76.06	66.23	75.53	92.98	58.62	57.55	95.55	56.33	56.0	76.61	17.32
	MSTKDNet	92.54	66.43	90.75	91.87	68.75	63.75	81.12	82.24	88.26	73.94	60.2	84.77	68.31	56.1	58.69	75.18	12.46
	Ours	91.77	**49.42**	96.04	**63.76**	**45.79**	**39.00**	80.45	86.07	**64.16**	**37.23**	**39.51**	70.37	**39.67**	**41.18**	**41.95**	**59.09**	20.52

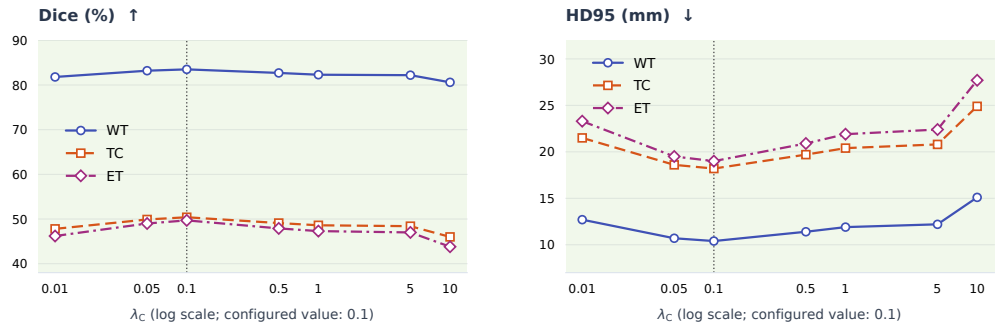

Fig. 4: Sensitivity to λ_C. Dice (left) and HD95 (right) for WT, TC, and ET.

Sensitivity to λ_U. The analysis of λ_U examines how strongly modality utility should constrain learning (Fig. 5). A weak constraint may leave harmful modality transitions insufficiently penalized, while an overly strong constraint may dominate the other training objectives. The analysis examines

Table 4: Dice (%) on BraTS 2018 under 15 modality combinations (higher is better). ● / ○: available / missing. Avg. and Std. denote the mean and population standard deviation across combinations. Bold marks the highest Dice in each region and column.

Region	FLAIR	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	Avg.	Std.
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●		
	T1c	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●		
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●		
WT	LCKD	76.18	71.32	70.72	74.31	75.28	74.93	76.73	77.26	78.31	79.16	79.94	80.59	81.08	81.12	81.56	77.23	3.35
	ShaSpec	78.89	42.74	43.99	66.55	81.97	82.15	82.34	56.70	69.91	71.98	83.19	83.53	83.75	73.32	84.41	72.36	13.70
	MriSeg	82.90	68.58	78.07	47.89	85.90	82.33	80.67	85.82	85.60	66.20	80.71	86.66	87.86	87.85	87.69	79.65	10.63
	MouseGAN++	66.67	75.83	60.28	60.51	79.07	79.71	71.36	70.63	68.60	79.88	79.10	73.00	81.20	80.14	82.81	73.92	7.13
	Hyper-GAE	73.99	65.36	67.59	74.44	72.81	60.99	73.15	74.44	76.17	75.43	78.80	78.27	79.24	77.75	80.62	73.94	5.30
	SRMNet	70.84	69.09	69.51	76.08	75.29	76.46	76.81	73.72	77.53	78.60	78.67	77.66	78.76	80.86	84.73	76.31	4.07
	RFNet	68.28	55.71	56.04	71.45	69.62	60.33	73.62	71.32	75.24	73.48	73.99	75.94	75.77	70.72	75.92	69.83	6.69
	mmFormer	81.13	73.15	69.75	87.15	85.26	76.74	88.25	83.21	88.19	88.73	88.71	88.43	89.18	85.81	89.12	84.19	6.05
	M^{2}FTrans	83.49	74.67	73.21	87.45	86.16	78.64	88.28	85.06	88.30	88.46	88.81	88.58	88.93	86.40	89.12	85.04	5.11
	RMBTS	78.76	75.57	67.98	61.35	77.87	76.11	72.55	73.75	72.19	79.28	80.14	84.40	82.92	80.59	83.56	76.47	6.02
	RA-HVED	73.33	**87.00**	71.13	81.29	**89.42**	**87.30**	82.85	77.16	84.02	87.55	87.85	84.26	89.23	**89.31**	88.98	84.05	5.73
	ACN	82.52	71.80	73.00	87.45	84.28	75.22	88.24	84.27	87.60	87.47	88.21	88.34	87.67	84.37	88.29	83.92	5.61
	SMUNet	83.75	72.00	76.04	87.67	84.94	77.20	**89.35**	85.29	88.25	88.14	89.26	89.39	88.31	85.59	89.26	84.96	5.32
	MSTKDNet	79.81	75.82	**78.68**	69.19	83.62	80.26	78.66	85.58	75.51	77.05	79.38	83.99	79.73	85.81	84.84	79.86	4.35
	Ours	**83.92**	72.87	70.41	**88.07**	86.24	76.17	89.29	**86.72**	**90.14**	**89.14**	**89.28**	**90.41**	**90.15**	87.03	**90.21**	**85.34**	6.42
TC	LCKD	61.97	60.33	66.64	66.16	66.23	68.71	70.65	72.02	71.74	71.62	72.73	73.54	73.37	74.01	74.64	69.62	4.30
	ShaSpec	73.95	45.03	66.21	65.42	76.53	74.90	75.83	75.31	65.60	77.48	80.20	77.12	81.08	79.61	82.79	73.14	9.18
	MriSeg	59.39	75.37	61.07	32.76	79.48	79.84	61.83	61.69	62.82	74.85	80.16	66.67	79.90	80.43	81.16	69.16	12.76
	MouseGAN++	61.92	67.87	60.25	61.79	68.47	60.86	66.28	67.02	67.71	70.33	72.58	67.85	70.97	73.02	74.75	67.44	4.41
	Hyper-GAE	57.42	62.28	54.38	61.06	72.97	67.42	60.48	61.06	63.09	73.51	79.76	72.58	76.61	80.63	82.42	68.38	8.80
	SRMNet	62.90	74.62	54.96	61.63	71.95	80.55	67.76	66.03	67.23	71.79	83.20	68.37	72.64	82.65	83.14	71.29	8.19
	RFNet	53.33	65.47	46.65	55.54	72.07	68.74	59.62	57.83	59.84	73.85	73.90	61.95	75.67	73.08	75.67	64.88	8.92
	mmFormer	62.76	81.63	59.72	64.00	82.48	81.14	71.64	68.10	69.74	82.58	82.48	72.34	82.66	82.09	82.51	75.06	8.22
	M^{2}FTrans	67.71	**82.25**	64.82	67.47	**83.75**	83.46	72.60	72.88	72.67	**84.41**	**84.75**	74.26	**84.67**	84.29	**84.67**	**77.64**	7.22
	RMBTS	61.17	56.77	75.69	44.45	68.19	82.37	78.28	81.92	65.51	63.21	82.71	82.19	68.70	82.83	82.99	71.80	11.45
	RA-HVED	**80.57**	64.12	60.31	67.11	81.95	67.42	70.23	81.81	**82.20**	68.61	70.81	82.99	82.58	82.82	82.93	75.10	7.99
	ACN	69.28	62.80	82.03	68.44	68.69	81.13	83.38	83.24	69.69	69.46	83.32	83.65	70.64	83.11	82.83	76.11	7.39
	SMUNet	68.98	61.96	81.68	**69.91**	70.28	82.89	**84.25**	84.02	71.35	70.95	84.34	**84.22**	71.69	84.12	84.07	76.98	7.52
	MSTKDNet	69.28	68.32	**83.94**	55.34	73.44	**84.18**	81.92	**84.76**	61.68	62.83	81.78	83.69	66.95	**85.00**	83.78	75.13	9.87
	Ours	64.44	77.96	57.44	66.15	82.76	81.20	70.56	70.04	70.16	83.06	83.71	72.03	83.91	83.80	84.25	75.43	8.42
ET	LCKD	38.36	35.38	46.97	44.73	43.91	48.44	51.73	54.20	**52.96**	51.89	53.71	55.18	54.35	55.59	56.65	49.60	6.28
	ShaSpec	30.97	21.24	46.20	25.80	31.29	62.90	53.01	56.71	45.44	62.57	63.05	33.34	61.23	64.26	66.37	48.29	15.36
	MriSeg	31.34	66.58	35.25	14.88	68.93	68.61	36.30	37.33	34.95	67.35	70.37	39.10	68.99	69.67	70.34	52.00	18.78
	MouseGAN++	38.33	41.27	27.28	36.08	41.72	51.38	40.72	41.03	48.91	50.10	58.22	43.25	58.53	56.65	59.46	46.20	9.18
	Hyper-GAE	34.52	30.72	36.55	38.12	36.79	54.52	48.20	53.11	40.47	52.74	53.36	38.81	53.26	57.50	60.47	45.94	9.33
	SRMNet	35.19	44.55	26.96	33.24	60.80	58.61	35.91	36.54	39.95	50.55	61.81	38.57	63.15	62.63	63.89	47.49	12.74
	RFNet	28.13	56.69	17.30	26.53	61.93	59.78	29.44	32.29	33.60	62.32	62.48	35.35	64.47	62.87	64.86	46.54	16.98
	mmFormer	36.97	68.75	31.48	37.02	70.37	69.46	39.14	39.47	42.57	69.69	69.81	42.29	69.77	70.59	69.82	55.15	15.83
	M^{2}FTrans	40.13	71.18	34.43	35.55	73.19	72.37	41.36	45.63	43.03	72.60	74.85	44.27	74.91	73.46	75.00	58.13	16.62
	RMBTS	34.66	33.71	46.38	17.35	40.85	50.41	48.31	49.75	37.39	34.82	51.08	59.82	40.74	65.39	64.38	45.00	12.41
	RA-HVED	39.28	45.62	56.18	**46.11**	36.65	43.58	43.80	38.44	39.05	44.42	43.49	36.35	36.60	36.24	36.23	41.47	5.33
	ACN	42.25	35.24	69.84	40.27	41.74	70.06	72.06	72.70	42.53	38.42	71.21	71.94	42.75	73.69	71.51	57.08	15.67
	SMUNet	42.73	34.21	71.21	41.49	42.22	72.38	72.98	74.04	43.59	42.05	72.94	72.53	43.35	72.89	72.49	58.07	15.76
	MSTKDNet	32.61	40.67	**73.82**	35.48	41.71	73.09	**73.53**	**74.99**	28.72	40.93	73.28	**73.85**	36.48	75.26	74.07	56.57	18.88
	Ours	**44.90**	**74.35**	36.62	45.83	**77.62**	**79.25**	47.95	47.53	50.72	**75.22**	**79.73**	50.03	**77.29**	**77.97**	**77.87**	**62.86**	15.89

the balance between enforcing modality utility and preserving segmentation accuracy as λ_{U} varies. In the illustrative curves in Fig. 5, $\lambda_{\mathrm{U}} = 0.5$ yields the highest Dice and lowest HD95 across WT, TC, and ET.

Sensitivity to λ_{H}. The analysis of λ_{H} considers the balance between guidance from the complete-modality representation and learning from the available inputs (Fig. 6). Increasing this coefficient strengthens full-to-partial alignment, but excessive alignment may limit adaptation to subset-specific evidence. In the illustrative curves in Fig. 6, $\lambda_{\mathrm{H}} = 1$ yields the highest Dice and lowest HD95 across WT, TC, and ET.

Table 5: HD95 (mm) on BraTS 2018 under 15 modality combinations (lower is better). ● / ○: available / missing. Avg. and Std. denote the mean and population standard deviation across combinations. Bold marks the lowest HD95 in each region and column.

Region	FLAIR	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	Avg.	Std.
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●		
	T1c	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●		
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●		
WT	LCKD	9.42	10.62	9.46	8.10	7.69	7.34	6.74	6.67	6.32	6.01	5.72	5.52	5.36	5.31	5.16	7.03	1.66
	ShaSpec	30.29	67.62	74.98	35.17	24.01	17.89	24.43	52.28	26.95	23.77	11.55	21.43	13.67	16.91	10.21	30.08	19.09
	MriSeg	**4.00**	12.70	9.17	30.53	**2.57**	**4.67**	8.12	**4.25**	**3.78**	12.70	8.20	**3.55**	**2.46**	**2.23**	**2.21**	7.41	7.09
	MouseGAN++	15.51	13.60	14.04	19.14	15.30	16.85	18.71	15.11	16.35	16.54	15.09	14.17	14.82	14.69	9.14	15.27	2.25
	Hyper-GAE	13.63	**8.71**	16.39	18.48	11.22	9.63	18.85	10.48	11.17	10.32	12.17	10.67	9.64	10.60	8.16	12.01	3.24
	SRMNet	13.90	23.16	20.10	10.87	10.01	14.22	8.20	12.28	7.39	8.29	5.48	7.36	6.62	7.65	7.50	10.87	4.94
	RFNet	38.64	47.90	39.14	34.50	37.47	38.12	27.08	33.85	25.91	29.65	28.56	24.48	23.89	35.36	24.84	32.63	6.75
	mmFormer	12.26	13.78	31.55	9.15	9.45	10.14	7.37	8.20	6.90	7.52	7.29	6.88	5.73	7.13	5.77	9.94	6.18
	M^2FTrans	9.29	10.60	16.06	8.81	7.29	8.97	6.57	7.73	8.15	8.45	6.18	6.36	8.20	7.03	5.81	8.37	2.42
	RMBTS	13.97	12.43	22.19	39.66	11.35	9.62	20.40	10.74	10.71	8.28	7.72	8.02	7.75	7.14	6.52	13.10	8.38
	RA-HVED	17.02	11.74	23.23	20.21	6.34	11.21	17.04	14.51	14.87	9.80	8.70	14.70	7.74	6.48	7.69	12.75	4.99
	ACN	10.01	13.01	11.60	6.14	8.12	10.59	5.33	8.24	5.39	5.97	5.01	6.00	5.72	7.95	5.99	7.67	2.46
	SMUNet	9.76	12.78	10.92	**5.41**	7.66	9.49	**4.75**	6.86	5.33	**5.36**	**4.72**	4.83	5.74	6.57	4.77	**7.00**	2.49
	MSTKDNet	10.77	10.36	**9.10**	17.49	6.60	8.98	9.05	5.96	16.82	11.12	8.84	7.63	12.42	6.42	7.40	9.93	3.34
	Ours	8.87	14.52	19.24	8.87	7.59	12.32	6.43	8.47	5.19	7.37	5.67	5.00	5.78	6.85	5.90	8.54	3.84
TC	LCKD	19.44	16.14	12.50	11.79	12.52	11.08	9.98	9.20	8.99	8.78	8.27	7.90	7.79	7.46	7.20	10.60	3.35
	ShaSpec	36.45	60.32	69.55	45.56	26.04	30.52	25.49	59.48	33.35	37.36	23.00	22.02	24.15	32.86	20.04	36.41	15.00
	MriSeg	12.30	13.79	13.10	32.66	**5.26**	**5.04**	12.90	10.14	10.67	12.69	6.09	9.15	5.28	5.27	4.60	10.60	6.78
	MouseGAN++	13.67	13.81	14.69	11.62	9.80	9.58	11.59	11.69	12.62	9.23	10.58	11.03	10.12	9.89	8.25	11.21	1.79
	Hyper-GAE	19.83	21.59	27.28	19.06	17.15	19.83	17.89	19.06	18.10	17.29	14.48	10.07	8.00	8.77	7.80	16.41	5.39
	SRMNet	17.89	22.67	21.27	18.22	12.66	11.44	16.49	14.26	13.94	9.26	8.14	10.12	8.70	7.95	7.40	13.36	4.83
	RFNet	18.43	49.95	31.18	23.73	32.75	42.03	19.04	15.48	16.52	25.61	24.63	13.58	17.87	32.23	17.47	25.37	10.14
	mmFormer	19.71	14.53	30.84	19.25	6.61	13.10	10.84	22.71	10.11	7.23	6.09	15.54	5.50	6.65	5.95	12.98	7.24
	M^2FTrans	17.08	**9.97**	25.57	13.10	6.37	6.19	15.43	15.66	8.74	6.09	5.73	8.38	**4.53**	6.32	**4.49**	10.24	5.79
	RMBTS	24.56	13.98	27.14	47.79	9.82	6.36	17.08	9.16	12.98	10.94	5.45	6.68	9.83	6.66	5.48	14.26	11.00
	RA-HVED	12.79	17.16	21.89	17.80	7.39	13.91	12.54	11.56	10.03	12.45	10.85	7.53	6.91	6.71	6.68	11.75	4.40
	ACN	10.95	12.95	7.62	11.36	10.92	8.01	5.55	5.34	10.99	11.42	5.16	5.09	10.83	5.44	5.48	8.47	2.84
	SMUNet	11.67	12.31	8.63	**10.59**	10.56	6.23	**5.09**	5.33	10.47	10.67	4.99	**5.04**	10.65	5.20	5.03	**8.16**	2.81
	MSTKDNet	**10.79**	16.12	**6.42**	21.77	8.49	5.19	6.00	**4.97**	19.98	18.28	6.15	5.72	11.23	5.35	5.80	10.15	5.76
	Ours	12.08	12.77	18.58	11.60	6.69	7.01	10.33	10.89	**8.25**	**5.23**	**4.88**	7.86	5.22	**4.97**	5.00	8.76	3.79
ET	LCKD	50.79	54.74	48.82	47.27	48.82	46.86	45.23	44.01	43.69	44.04	43.27	42.65	42.95	42.51	42.07	45.85	3.53
	ShaSpec	57.54	74.89	65.64	54.05	53.23	48.57	51.70	69.00	51.29	47.63	42.09	48.22	40.41	43.30	38.73	52.42	10.21
	MriSeg	45.90	42.78	47.01	64.44	36.79	36.49	46.87	43.83	45.03	41.45	38.00	43.07	36.31	36.69	35.93	42.71	7.05
	MouseGAN++	97.54	98.70	97.92	87.46	80.73	89.93	93.44	84.26	93.73	88.77	87.89	80.09	77.21	76.85	68.65	86.88	8.55
	Hyper-GAE	107.79	109.44	93.89	99.23	96.76	87.79	87.95	99.25	88.73	97.36	76.60	77.38	75.17	75.70	73.04	89.74	11.66
	SRMNet	108.89	99.33	101.64	98.62	89.63	90.92	90.68	96.47	85.27	78.55	68.31	66.15	66.50	67.61	64.65	84.88	14.59
	RFNet	50.67	76.68	55.02	52.13	62.38	68.48	54.89	46.30	45.56	57.59	54.62	48.37	47.90	61.27	47.16	55.27	8.56
	mmFormer	81.11	44.76	70.48	44.48	36.58	49.50	43.94	80.14	55.51	36.91	37.03	62.16	36.74	37.51	37.03	50.26	15.47
	M^2FTrans	62.48	**37.08**	64.16	76.07	36.42	35.93	74.95	60.49	79.16	36.36	29.61	56.98	**29.30**	36.25	29.24	49.63	18.01
	RMBTS	103.70	121.46	95.57	104.91	102.45	96.74	99.07	116.84	83.46	98.03	86.70	76.84	72.56	76.68	66.72	93.45	15.43
	RA-HVED	71.22	40.49	33.34	**43.51**	72.42	42.33	44.21	71.65	72.35	45.46	46.56	72.48	72.58	72.72	72.95	58.28	15.24
	ACN	43.99	45.44	38.52	44.89	43.63	38.18	36.52	36.33	43.96	44.94	36.22	36.15	43.91	29.85	36.62	39.94	4.56
	SMUNet	43.80	51.78	38.21	44.02	43.77	42.24	36.10	30.17	43.81	43.94	36.23	36.12	43.82	36.66	36.36	40.47	5.16
	MSTKDNet	44.25	61.77	**31.17**	46.97	49.63	36.54	**30.96**	**29.67**	51.05	50.12	30.95	**30.65**	56.02	30.25	30.79	40.72	10.75
	Ours	**42.76**	37.61	71.95	47.99	**30.14**	**21.99**	54.30	54.36	**33.23**	**35.90**	**22.17**	45.21	29.59	**28.63**	**28.93**	**38.98**	13.37

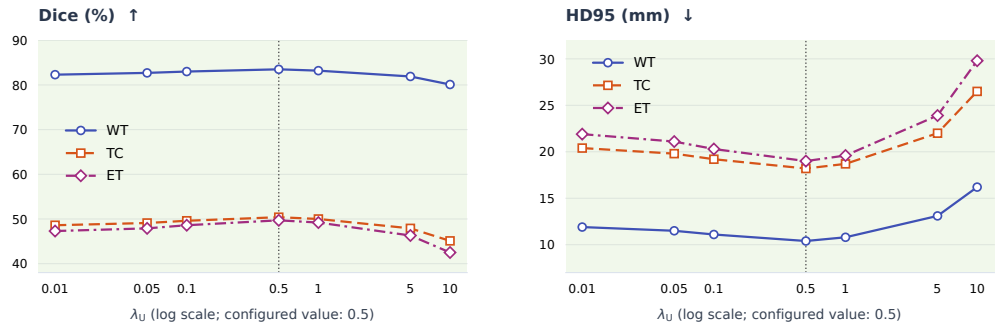

Fig. 5: Sensitivity to λ_U. Dice (left) and HD95 (right) for WT, TC, and ET.

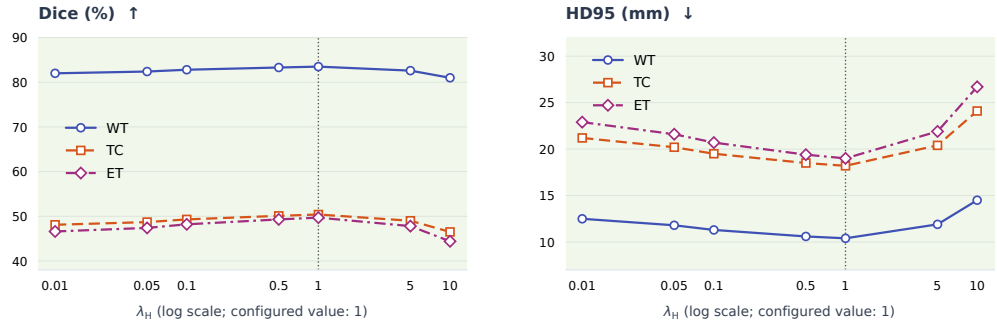

Fig. 6: Sensitivity to λ_{H}. Dice (left) and HD95 (right) for WT, TC, and ET.