DGP-HPCL: DUAL-PRIOR-GUIDED HYPERGRAPH PROTOTYPE CONTRASTIVE LEARNING FOR BRAIN TUMOR SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Geometric supervision can improve segmentation outputs without specifying which intermediate features should participate in relational learning. We propose dual geometric prior-guided hypergraph prototype contrastive learning (DGP-HPCL) for multimodal brain tumor segmentation. A Gaussian center prior and an interior distance prior, both derived from training labels, jointly select decoder locations. Feature similarity then defines Soft-KNN hyperedges and their aggregated representations. Within each subject, the low- and high-center-response tails define two prototypes and the embeddings supervised by the contrastive objective. Each endpoint embedding is contrasted with both prototypes, favoring its own geometric group over the opposite group. Middle-ranked nodes remain available for aggregation without receiving binary prototype labels. Auxiliary prior prediction and cross-gated bidirectional consistency complement this relational objective during training. On BraTS 2019 five-fold validation, DGP-HPCL achieves 82.38% DSC, 73.54% surface Dice, and 6.91 HD95, the best means among the evaluated models. On the BraTS 2021 evaluation split, it achieves 88.87% DSC, 84.52% surface Dice, and 5.14 HD95. Direct cross-year evaluation and controlled design comparisons further support geometry-guided sampling and endpoint supervision. The framework regularizes a shared decoder during training and requires neither labels nor hypergraph construction at inference.

1 INTRODUCTION

Multimodal brain tumor segmentation must distinguish heterogeneous tissue while recovering irregular and sometimes disconnected lesion boundaries. BraTS provides co-registered FLAIR, T1ce, T1, and T2 MRI with annotations for enhancing tumor, tumor core, and whole tumor (Menze et al., 2015; Bakas et al., 2017; Baid et al., 2021). Overlap-based training rewards correct tissue assignment, but small structures and local contour errors can contribute little to the aggregate objective.

Two established directions address different parts of this problem. Convolutional, Transformer, and state-space networks improve contextual feature extraction (Isensee et al., 2021; Wang et al., 2021; Hatamizadeh et al., 2022a; Xing et al., 2024). Boundary, distance, and topology objectives instead impose structural constraints on predictions (Kervadec et al., 2021; Karimi & Salcudean, 2020; Calivá et al., 2019; Hu et al., 2019). These objectives update intermediate features through backpropagation, but do not themselves specify the samples and targets of an additional relational loss. This distinction motivates our question: how can label-derived geometry guide relational learning within a segmentation decoder?

The question has two linked parts: where to sample features and which sampled representations to supervise. Constructing relations over every location in a 3D volume is expensive, so a relational branch requires a restricted candidate set. Random sampling need not concentrate on a lesion, while feature-magnitude ranking does not explicitly encode its geometry. After sampling, assigning every candidate to one of two geometric groups introduces another choice. Locations with intermediate center responses need not form a well-separated group, even though their features may be useful for aggregation.

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

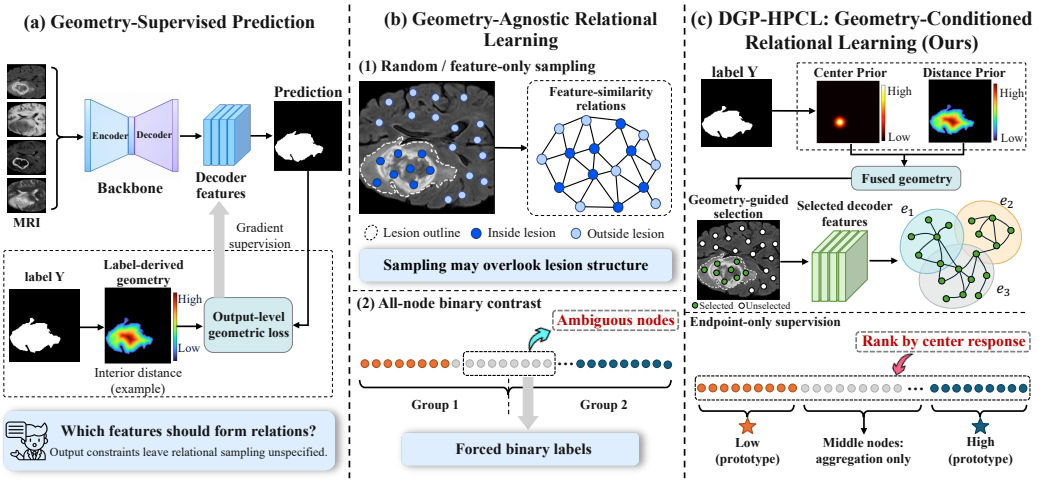

Figure 1: Motivation for DGP-HPCL. Output-level geometric supervision and relational representation learning address different decisions. Our design uses fused label-derived priors to select candidate locations, feature similarity to construct hyperedges, and center-response endpoints to define prototype targets. Middle-ranked candidates remain available for aggregation without direct endpoint labels.

DGP-HPCL addresses both choices using two complementary geometric priors (Figure 1). An unmasked Gaussian encodes proximity to the foreground centroid, and an interior distance field encodes depth within the foreground. Their weighted sum selects decoder locations; feature similarity, rather than spatial proximity, determines the subsequent Soft-KNN hyperedges. Each hyperedge aggregates multiple selected features into one representation. The low- and high-center-response tails define subject-specific prototypes and contrastive targets, while middle-ranked nodes remain available for aggregation.

First, we use dual label-derived priors to guide candidate selection for a sparse hypergraph regularizer. Second, we combine cross-gated prior supervision with hyperedge–prototype contrast that labels only the two center-response tails. Third, we evaluate the resulting framework using within-dataset comparisons, direct cross-year evaluation, component ablations, and separate sampling, aggregation, and prototype controls.

2 Related Work

Segmentation architectures. U-Net and its volumetric extensions combine hierarchical encoding with skip-connected decoding (Ronneberger et al., 2015; Çiçek et al., 2016; Milletari et al., 2016). Self-configuring pipelines and convolutional scaling remain strong alternatives to more complex architectures (Isensee et al., 2021; Roy et al., 2023). UNETR, TransBTS, Swin UNETR, nnFormer, and CoTr extend contextual interaction through volumetric or hybrid Transformers (Hatamizadeh et al., 2022b; Wang et al., 2021; Hatamizadeh et al., 2022a; Zhou et al., 2023; Xie et al., 2021). State-space approaches include SegMamba, U-Mamba, and nnMamba (Xing et al., 2024; Ma et al., 2024; Gong et al., 2025), with recent hybrids including KM-UNet, KMUNet, SAMA-UNet, ACM-UNet, and NexMamba (Zhang, 2025; Zhang et al., 2025; Qamar et al., 2026; Huang et al., 2026; Liang et al., 2026). We use SuperLightNet (Yu et al., 2025) to study training-time regularization of decoder representations, rather than introduce another backbone.

Geometric and relational supervision. Boundary, distance-map, and topology objectives constrain structural properties of segmentation outputs (Kervadec et al., 2021; Karimi & Salcudean, 2020; Calivá et al., 2019; Hu et al., 2019; Shit et al., 2021). They motivate our auxiliary targets but do not specify the candidate pool or contrastive targets of an additional relational branch. Graph networks model pairwise neighborhoods, whereas hypergraphs represent multi-node relations (Kipf & Welling, 2017; Zhou et al., 2006; Feng et al., 2019; Jiang et al., 2019; Bai et al., 2021). Contrastive

learning organizes instance, class, or prototype representations (He et al., 2020; Khosla et al., 2020; Li et al., 2021); structural information has also informed medical contrastive learning (Chaitanya et al., 2020). Our distinction is the combination of fused-prior sampling, feature-based hyperedge aggregation, and within-subject endpoint supervision.

3 METHOD

3.1 PROBLEM SETUP

Let $X \in \mathbb{R}^{B\times 4\times D\times H\times W}$ contain four co-registered MRI modalities: FLAIR, T1, T1ce, and T2. The label map Y uses 0, 1, 2, and 4 for background, necrotic/non-enhancing tumor, edema, and enhancing tumor. Evaluation combines these labels into enhancing tumor ($\mathrm{ET} = \{4\}$), tumor core ($\mathrm{TC} = \{1, 4\}$), and whole tumor ($\mathrm{WT} = \{1, 2, 4\}$).

A 3D U-shaped backbone processes all four modalities jointly and produces $F = f_\theta(X) \in \mathbb{R}^{B\times C_f\times D'\times H'\times W'}$. The segmentation head produces four logits $O_{\mathrm{seg}} = g_{\mathrm{seg}}(F)$. Two sigmoid auxiliary heads predict $P_c = \sigma(g_c(F))$ and $P_d = \sigma(g_d(F))$, the center and interior-distance maps. Each sampled node is the complete C_f-dimensional vector at one decoder location, already containing information from the four modalities.

Figure 2 separates dense prediction from training-time regularization. The segmentation head and the hypergraph branch both read F; hypergraph outputs are not passed into the segmentation head. During training, the relational loss backpropagates through selected features into the shared backbone. During inference, only the backbone and segmentation head are evaluated, without labels, prior generation, node selection, or hypergraph construction.

3.2 DUAL GEOMETRIC PRIORS

For each spatially augmented training label, define the foreground mask $M(p) = \mathbf{1}\{Y(p) > 0\}$ at voxel $p = (x, y, z)$. For a nonempty foreground, its centroid and Gaussian center prior are

$$\boldsymbol{\mu} = \frac{\sum_p M(p)p}{\sum_p M(p) + \epsilon}, \qquad C(p) = \exp\left(-\frac{\|p - \boldsymbol{\mu}\|_2^2}{2\sigma_c^2}\right), \tag{1}$$

where $\sigma_c = 5$ voxels and $\epsilon > 0$ prevents numerical division by zero. The Gaussian is unmasked: it encodes proximity to the foreground centroid and can assign nonzero responses outside the tumor. It is therefore a positional prior, not a foreground probability.

The interior prior normalizes the foreground Euclidean distance transform: $D_0 = \mathrm{EDT}(M)$ and $D(p) = D_0(p)/(\max_q D_0(q) + \epsilon)$. It is zero outside the foreground and increases with distance from the background inside the lesion. Unlike the center prior, this field follows local foreground geometry. Their weighted fusion $S = \alpha C + (1 - \alpha)D$, with $\alpha = 0.7$, defines the selection score. Both priors are computed after spatial augmentation to preserve alignment with the image. They are resized by trilinear interpolation when the decoder grid has lower resolution. Label-derived targets, rather than the predicted maps P_c and P_d, determine node selection and endpoint ranking.

3.3 BIDIRECTIONAL STRUCTURAL CONSISTENCY

The auxiliary heads receive direct supervision through $\mathcal{L}_{\mathrm{center}} = \mathrm{MSE}(P_c, C)$ and $\mathcal{L}_{\mathrm{dist}} = \mathrm{L1}(P_d, D)$, averaged over their aligned grid. Bidirectional structural consistency (BSC) additionally uses each prediction to weight the complementary branch's target error:

$$\begin{aligned} \mathcal{L}_{c\to d} &= \mathrm{MSE}(P_c \odot P_d,\ P_c \odot D), \\ \mathcal{L}_{d\to c} &= \mathrm{MSE}(P_d \odot P_c,\ P_d \odot C), \\ \mathcal{L}_{\mathrm{bsc}} &= \mathcal{L}_{c\to d} + \mathcal{L}_{d\to c}. \end{aligned} \tag{2}$$

Here $\odot$ denotes element-wise multiplication. Thus, the gates emphasize locations with stronger complementary responses, without treating those responses as calibrated confidence estimates. Direct supervision remains necessary because low predicted gates can otherwise reduce the consistency penalty without improving either target prediction.

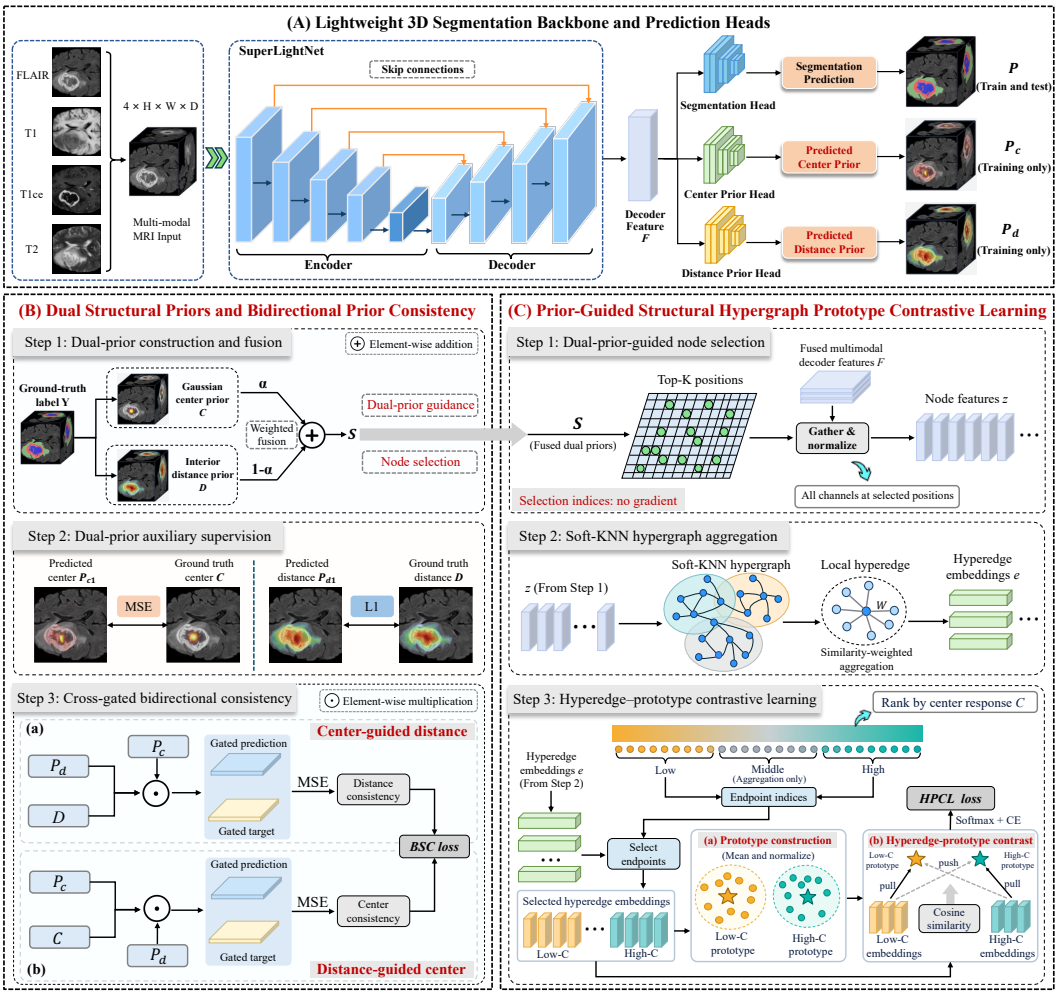

Figure 2: Overview of DGP-HPCL. (A) The shared decoder supplies the segmentation head, auxiliary prior heads, and hypergraph training branch. (B) Label-derived center and normalized interior-distance priors guide selection and auxiliary prediction; cross-gating weights each prediction and its corresponding target. (C) Feature similarity defines Soft-KNN aggregation, and center-response endpoints supervise hyperedge–prototype contrast. Auxiliary and relational losses update the backbone during training; inference uses only the dense segmentation path.

3.4 Hypergraph Prototype Contrastive Learning

Dual-prior-guided node selection. For each subject, the top K values of S identify candidate locations $\Omega_K = \{v_1, \ldots, v_K\}$. We gather and normalize their decoder vectors:

$$\mathbf{z}_i = \frac{F(v_i)}{\|F(v_i)\|_2 + \epsilon}, \qquad i = 1, \ldots, K, \tag{3}$$

with $K = 256$ by default. Selection indices are computed without gradient tracking; gradients still reach the gathered decoder features. This candidate set supports an auxiliary representation loss, while dense segmentation supervision continues to cover the full output volume.

Soft-KNN hypergraph aggregation. Each candidate anchors one hyperedge containing its k most similar candidates under cosine similarity $s_{ij} = \mathbf{z}_i^\top \mathbf{z}_j$. For the neighborhood $\mathcal{N}_k(i)$, the aggregation weights are

$$w_{ij} = \frac{\exp(s_{ij}/\tau)}{\sum_{m \in \mathcal{N}_k(i)} \exp(s_{im}/\tau)}, \qquad j \in \mathcal{N}_k(i), \tag{4}$$

where $k = 8$ and $\tau = 0.10$. The normalized hyperedge embedding is

$$\mathbf{e}_i = \frac{\sum_{j \in \mathcal{N}_k(i)} w_{ij} \mathbf{z}_j}{\left\| \sum_{j \in \mathcal{N}_k(i)} w_{ij} \mathbf{z}_j \right\|_2 + \epsilon}. \tag{5}$$

Soft weighting is differentiable within each selected neighborhood; the top-k membership decision remains discrete. Geometry restricts the candidate pool, whereas feature similarity determines membership and weights within that pool.

Endpoint hyperedge–prototype contrast. We rank candidate anchors by $C(v_i)$, not by the fused score $S(v_i)$ or predicted center response. The lowest and highest $m_\rho = \max(1, \lfloor \rho K \rfloor)$ responses define index sets $\mathcal{I}_{\mathrm{lo}}$ and $\mathcal{I}_{\mathrm{hi}}$. With $\rho = 0.20$ and $K = 256$, each endpoint group contains 51 anchors, leaving 154 middle-ranked anchors. The corresponding *hyperedge embeddings*, rather than the original node vectors, form normalized group prototypes:

$$\overline{\mathbf{e}}_g = \frac{1}{|\mathcal{I}_g|} \sum_{i \in \mathcal{I}_g} \mathbf{e}_i, \qquad \mathbf{p}_g = \frac{\overline{\mathbf{e}}_g}{\|\overline{\mathbf{e}}_g\|_2 + \epsilon}, \quad g \in \{\mathrm{lo}, \mathrm{hi}\}. \tag{6}$$

Each endpoint embedding has two temperature-scaled cosine logits, $\ell_{i,0} = \mathbf{e}_i^\top \mathbf{p}_{\mathrm{lo}}/\tau$ and $\ell_{i,1} = \mathbf{e}_i^\top \mathbf{p}_{\mathrm{hi}}/\tau$. Let $\mathcal{I} = \mathcal{I}_{\mathrm{lo}} \cup \mathcal{I}_{\mathrm{hi}}$ and assign $y_i = 0$ to low-response endpoints and $y_i = 1$ to high-response endpoints. The hypergraph prototype contrastive loss is

$$\mathcal{L}_{\mathrm{HPCL}} = -\frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \log \frac{\exp(\ell_{i,y_i})}{\exp(\ell_{i,0}) + \exp(\ell_{i,1})}. \tag{7}$$

Softmax converts the two scaled similarities into relative group probabilities, and cross-entropy penalizes assignment to the opposite endpoint group. The loss is evaluated within each subject and averaged over the batch. Each endpoint embedding is compared with both prototypes; its own group provides the positive target and the other group the negative target. The pull/push arrows in Figure 2 depict this relative similarity objective; no separate prototype–prototype repulsion term is used. These targets encode relative center response within the selected set, not semantic tissue classes. Middle-ranked nodes can contribute to endpoint hyperedges and receive gradients through aggregation, but their own embeddings have no direct endpoint loss.

3.5 Training objective

The segmentation objective sums binary cross-entropy and Dice losses over the three foreground tissue labels. These exclusive targets differ from overlapping ET, TC, and WT evaluation regions; Appendix A gives the channel mapping and full expression. The full objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{seg}} + \lambda_c \mathcal{L}_{\mathrm{center}} + \lambda_d \mathcal{L}_{\mathrm{dist}} + \lambda_b \mathcal{L}_{\mathrm{bsc}} + \lambda_{\mathrm{HPCL}}(t) \mathcal{L}_{\mathrm{HPCL}}, \tag{8}$$

with $\lambda_c = 0.3$, $\lambda_d = 0.3$, and $\lambda_b = 0.2$. The HPCL weight warms up linearly to $\lambda_{\mathrm{HPCL}} = 0.05$ over 30 epochs.

The five terms serve three roles: dense segmentation, auxiliary prior prediction with cross-gating, and relational contrast.

4 Experiments

4.1 Datasets, metrics, and implementation

Data and evaluation. We evaluated DGP-HPCL on BraTS 2019 and BraTS 2021 (Bakas et al., 2018; Baid et al., 2021). BraTS 2019 used five-fold validation with matched folds across compared methods. BraTS 2021 used a fixed training, validation, and evaluation split, reported separately in the tables. Each modality was z-normalized within the nonzero brain mask. Volumes were padded along the third axis and center-cropped to 128^3 voxels. Training augmentation comprised independent spatial-axis flips and intensity scaling and shifting with magnitude 0.1. Geometric targets were generated online after spatial augmentation.

Table 1: Mean segmentation results on BraTS 2019 (five folds) and BraTS 2021 (evaluation split). Bold and underline mark the best and second-best distinct values, including ties, in each comparison. DSC/SDC: higher is better; HD95: lower is better.

	BraTS 2019			BraTS 2021		
Method	DSC	SDC	HD95	DSC	SDC	HD95
nnU-Net (2021)	81.79	72.16	7.77	<u>88.78</u>	83.86	6.30
TransBTS (2021)	75.01	57.81	9.96	84.51	75.75	6.24
Swin UNETR (2022)	<u>81.85</u>	71.87	<u>7.66</u>	88.55	81.08	7.20
MedNeXt (2023)	80.71	68.81	8.69	80.67	67.70	10.17
SegMamba (2024)	<u>81.85</u>	71.11	<u>7.66</u>	88.56	82.41	5.96
KM-UNet (2025)	79.84	68.41	10.23	87.40	80.53	8.34
KMUNet (2025)	77.50	64.67	15.34	84.18	75.70	15.45
SAMA-UNet (2026)	77.92	65.63	13.50	84.45	76.30	15.50
SuperLightNet (2025)	81.77	<u>72.23</u>	7.80	88.64	<u>84.12</u>	<u>5.84</u>
ACM-UNet (2026)	80.56	70.41	9.39	87.93	81.68	6.76
DGP-HPCL (Ours)	**82.38**	**73.54**	**6.91**	**88.87**	**84.52**	**5.14**

We report Dice similarity coefficient (DSC), surface Dice coefficient (SDC) with one-voxel tolerance, and the 95th-percentile Hausdorff distance (HD95) (Dice, 1945; Taha & Hanbury, 2015; Maier-Hein et al., 2024). Reported region means average ET, TC, and WT; DSC and SDC are expressed as percentages throughout. Differences between these percentages are reported in percentage points (pp).

Models and training controls. The SuperLightNet-based backbone used four input channels, base width 24, and depth $[1, 1, 2, 2, 2]$ (Yu et al., 2025). Default relational settings were $K = 256$, $k = 8$, $\tau = 0.10$, $\rho = 0.20$, and $\alpha = 0.70$. Compared methods span convolutional, Transformer, and state-space architectures. KM-UNet and KMUNet refer to distinct methods, despite their similar names (Zhang, 2025; Zhang et al., 2025). All ablation settings shared preprocessing, augmentation, training budget, and model selection. The full objective is evaluated only during training; deployment retains the backbone and segmentation head.

4.2 Main results

BraTS 2019. Table 1 reports mean results; Appendix B retains all five validation folds. DGP-HPCL achieved the best mean DSC (82.38%), SDC (73.54%), and HD95 (6.91) among the listed methods. Compared with the underlying SuperLightNet baseline, the gains were 0.61 pp DSC and 1.31 pp SDC, with HD95 reduced by 0.89. Against the strongest competing mean in each column, the margins were 0.53 pp DSC, 1.31 pp SDC, and 0.75 HD95. SDC was highest in all five folds, whereas DSC and HD95 were best in three folds each.

BraTS 2021. On the evaluation split, DGP-HPCL achieved 88.87% DSC, 84.52% SDC, and 5.14 HD95 (Table 1). These were the best evaluation values among the listed models. The DSC margin over nnU-Net was 0.09 pp; relative to SuperLightNet, DSC and SDC increased by 0.23 and 0.40 pp. HD95 decreased from 5.84 to 5.14, a reduction of 0.70, or approximately 12%.

Model size and computation. The reported profile for DGP-HPCL was 2.99 M parameters and 0.16 T FLOPs, compared with 2.97 M and 0.09 T for SuperLightNet. Relative to nnU-Net, the reported counts were approximately 73% lower for parameters and 74% lower for FLOPs. Removing the training-only branches eliminates their deployment operations, but the tabulated counts do not demonstrate unchanged inference cost.

Table 2: Cross-year evaluation means. Five source models share one target set.

Method	DSC	SDC	HD95
nnU-Net (2021)	85.67	78.65	8.67
TransBTS (2021)	75.67	58.65	11.38
Swin UNETR (2022)	<u>86.02</u>	77.66	8.13
MedNeXt (2023)	83.68	73.77	9.54
SegMamba (2024)	85.14	77.09	8.57
SuperLightNet (2025)	85.87	<u>79.75</u>	**7.48**
ACM-UNet (2026)	85.50	78.00	9.86
DGP-HPCL (Ours)	**86.18**	**80.60**	<u>7.54</u>

Table 3: Component comparisons on BraTS 2021 evaluation.

Variant	DSC	SDC	HD95
Baseline	88.64	84.12	5.84
Dual priors	88.56	<u>84.38</u>	5.93
GCN relation	88.40	84.10	5.20
HG relation	<u>88.70</u>	84.36	<u>5.18</u>
w/o D	88.34	83.63	5.85
w/o C	88.64	84.36	5.45
DGP-HPCL (Ours)	**88.87**	**84.52**	**5.14**

4.3 Cross-dataset generalization

Models trained and selected on BraTS 2019 were directly evaluated on BraTS 2021 without fine-tuning, target-set selection, or threshold adjustment. Each of the five source-fold models was evaluated on the same target set. Table 2 summarizes the target results; the five model-level measurements in Appendix C are not independent target splits or an ensemble.

DGP-HPCL obtained the highest target mean DSC (86.18%) and SDC (80.60%). It exceeded Swin UNETR by 0.16 pp DSC and SuperLightNet by 0.85 pp SDC. Compared with SuperLightNet directly, its DSC gain was 0.31 pp. Target HD95 was 7.54, the second-best mean in this comparison. DGP-HPCL led DSC for three source-fold models and SDC for four; target HD95 was best for two models. The results support cross-year performance under the reported protocol, subject to a documented audit of overlapping cases between BraTS releases.

4.4 Ablations and mechanism comparisons

Component comparisons. Table 3 reports the BraTS 2021 variants. Adding dual-prior supervision increased evaluation SDC from 84.12% to 84.38%. Auxiliary targets alone therefore did not reproduce the full model's improvement across metrics. The GCN relation variant reduced HD95 to 5.20, while its DSC and SDC remained below the baseline. The HG relation variant reached 88.70% DSC, 84.36% SDC, and 5.18 HD95. Relative to that variant, the full model added 0.17 pp DSC and 0.16 pp SDC, with HD95 reduced by 0.04.

Removing the distance prior reduced evaluation DSC and SDC to 88.34% and 83.63%, with HD95 increasing to 5.85. Removing the center prior gave 88.64% DSC, 84.36% SDC, and 5.45 HD95. The table supports the combined configuration; it does not isolate BSC because no BSC-only removal is reported.

Which locations should form relations? The node-selection block in Table 4 tests the sampling question raised in the introduction. Dual-prior selection improved DSC and SDC over random selection by 0.53 and 0.68 pp, and over feature-norm selection by 1.01 and 1.90 pp. Center-only and distance-only selection also underperformed fusion in DSC and SDC. These controls support combining positional and interior-depth information for overlap and surface agreement.

How should selected features be aggregated? With candidate selection fixed, Soft-KNN improved DSC and SDC over identity features by 1.13 and 1.62 pp, and reduced HD95 by 0.93. It also outperformed the pairwise KNN graph and hard-KNN hyperedge variants on all three means. Relative to hard-KNN aggregation, the gains were 0.39 pp DSC and 0.55 pp SDC, with a 0.67 reduction in HD95. The comparison supports similarity-weighted multi-node aggregation within the tested candidate pool.

Which hyperedges should receive prototype targets? Quantile endpoints outperformed pairwise contrast, all-node hard prototypes, and fixed-threshold endpoints on all three reported means. Compared with all-node supervision, DSC and SDC increased by 0.37 and 0.60 pp, and HD95 decreased

Table 4: Mechanism comparisons on BraTS 2019. Each block changes the indicated design choice. Best and second-best values are bold and underlined within each block.

Variant	DSC	SDC	HD95
Node selection			
Random top-K	<u>81.85</u>	<u>72.86</u>	**6.85**
Feature-norm top-K	81.37	71.64	7.42
Center top-K	81.41	71.80	8.28
Distance top-K	81.13	71.55	7.32
Dual-prior top-K	**82.38**	**73.54**	<u>6.91</u>
Relation			
Identity node	81.25	71.92	7.84
Pairwise KNN graph	81.52	72.12	<u>7.43</u>
Hard-KNN hyperedge	<u>81.99</u>	<u>72.99</u>	7.58
Soft-KNN hyperedge	**82.38**	**73.54**	**6.91**
Prototype			
Pairwise contrast	81.20	72.25	<u>7.10</u>
All-node hard prototype	82.01	72.94	7.25
Fixed-threshold endpoint	<u>82.06</u>	<u>73.37</u>	7.98
Quantile endpoint prototype	**82.38**	**73.54**	**6.91**

by 0.34. This pattern supports selecting a fixed fraction of each subject's ranking tails rather than assigning binary targets to every selected anchor.

The graph, hypergraph, and contrastive controls follow established modeling families (Kipf & Welling, 2017; Feng et al., 2019; Jiang et al., 2019; Li et al., 2021). Random and feature-norm selection are internal controls for the proposed sampling rule, not reproductions of named competing methods.

4.5 Hyperparameter sensitivity

The default configuration performed best on all three metrics among the tested settings in BraTS 2019 fold 1 (Table 10). These one-factor comparisons characterize this fold; they do not demonstrate stability across folds or repeated training seeds.

Candidate and neighborhood size. Reducing K from 256 to 128 decreased DSC by 1.77 pp and SDC by 3.37 pp. Increasing K to 512 recovered some performance but remained below the default, at 81.85% DSC and 72.29% SDC. For hyperedge size, $k = 16$ was closer to the default than $k = 4$ on all three metrics. Neither increasing the candidate count nor enlarging the neighborhood yielded a monotonic improvement.

Temperature and contrastive weight. Both $\tau = 0.05$ and $\tau = 0.20$ reduced DSC and SDC relative to $\tau = 0.10$. The smaller temperature also increased HD95 from 6.06 to 8.14. Likewise, reducing λ_{HPCL} to 0.01 or increasing it to 0.10 worsened all three metrics.

Endpoint fraction and prior fusion. Both smaller ($\rho = 0.10$) and larger ($\rho = 0.30$) endpoint groups underperformed $\rho = 0.20$. Their SDC values were 70.51% and 71.02%, compared with 74.17% for the default. For fusion, $\alpha = 0.50$ remained close in DSC and SDC (82.33% and 73.63%), but HD95 increased to 7.72. Both $\alpha = 0.30$ and $\alpha = 0.90$ were weaker than $\alpha = 0.70$ across the reported metrics. The tested configurations support retaining both priors with a balanced fusion weight.

4.6 Qualitative visualization

Figure 3 compares segmentation overlays for two cases using matched slices across methods. Most methods recover the lesion location, but their outer contours and internal subregions differ. DGP-

HPCL follows the reference extent and internal structure in these examples. These examples complement the aggregate metrics by showing spatial differences in the predictions.

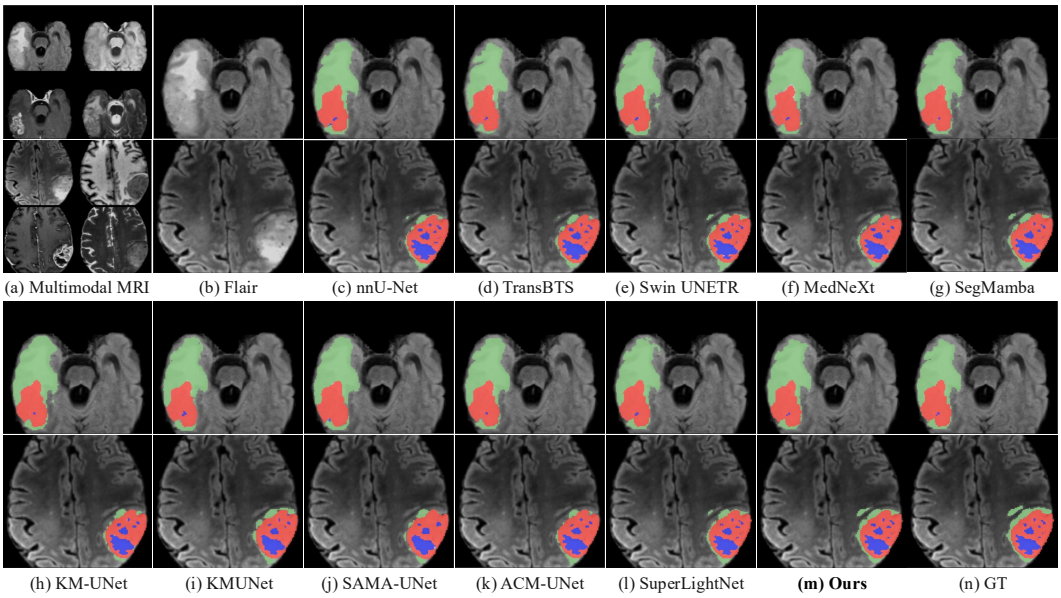

Figure 3: Qualitative BraTS segmentation with FLAIR backgrounds. Rows 1 and 3 show case A; rows 2 and 4 show case B. (a) Multimodal inputs; (b) FLAIR; (c–l) comparison methods; (m) DGP-HPCL (Ours); (n) ground truth. Blue denotes necrotic/non-enhancing tumor core (NCR/NET), green denotes peritumoral edema (ED), and red denotes enhancing tumor (ET). Tumor core (TC) combines red and blue; whole tumor (WT) combines red, blue, and green.

5 Discussion and Conclusion

DGP-HPCL uses dual geometric priors to select decoder locations and center-response endpoints to supervise hyperedge–prototype contrast. The component results show the value of coupling auxiliary geometric supervision with relational learning. Separate controls favored fused-prior selection, weighted aggregation, and endpoint targets, linking the results to the two motivating decisions: where to sample and which representations to supervise.

Endpoint groups describe relative geometry within the selected set, not tissue identity. Middle-ranked nodes remain useful because they can contribute to endpoint hyperedges without receiving direct binary targets. The advantage over all-node supervision is empirical; intermediate responses are not measured uncertainty estimates.

The framework separates geometric candidate selection from feature-based relation construction and restricts direct contrastive supervision to the ranking tails. This design lets intermediate candidates contribute contextual information without imposing binary endpoint targets on them. Further evaluation can examine multi-component lesions, alternative sampling budgets, and an isolated BSC ablation. Within the evaluated backbone and BraTS setting, the findings support dual-prior-guided hypergraph contrast as a useful training regularizer.

The results connect geometric supervision to representation learning at two distinct stages: fused priors determine the candidate pool, while center-response ranking determines the contrastive targets. The gains over identity aggregation and all-node prototype supervision support these complementary roles. Dense segmentation supervision still covers the full volume, so sparse relational sampling supplements rather than replaces voxel-level learning. At inference, predictions depend only on the image and the learned dense segmentation path. Geometry thus shapes decoder representations during training without requiring geometric targets or hypergraph construction for a new scan.

AI USE STATEMENT

Generative AI tools assisted with language editing, LaTeX preparation, reference checking, and code review and debugging. Experimental measurements were not generated by these tools. The authors are responsible for verifying the manuscript, citations, code, and diagrams against the implemented method and original results.

ETHICS STATEMENT

This study evaluates segmentation algorithms using public BraTS research datasets, subject to their licenses and citation requirements. No patient-identifying information is presented, and the experiments do not establish suitability for clinical deployment.

REPRODUCIBILITY STATEMENT

The method specifies prior construction, node selection, Soft-KNN weighting, endpoint targets, and the training objective. Tables report fold-level measurements for BraTS 2019 and separate development and evaluation measurements for BraTS 2021. Complete reproduction additionally requires the split manifests, training configurations, decoding rules, and profiling conventions. The code is provided in the supplementary material.

REFERENCES

Song Bai, Feihu Zhang, and Philip H. S. Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021. doi: 10.1016/j.patcog.2020.107637. URL https://doi.org/10.1016/j.patcog.2020.107637.

Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C. Kitamura, Sarthak Pati, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021. doi: 10.48550/arXiv.2107.02314. URL https://arxiv.org/abs/2107.02314.

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*, 4:170117, 2017. doi: 10.1038/sdata.2017.117. URL https://www.nature.com/articles/sdata2017117.

Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell T. Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. 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*, 2018. doi: 10.48550/arXiv.1811.02629. URL https://arxiv.org/abs/1811.02629.

Francesco Calivá, Claudia Iriondo, Alejandro Morales Martinez, Sharmila Majumdar, and Valentina Pedoia. Distance map loss penalty term for semantic segmentation. In *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track*, 2019. URL https://openreview.net/forum?id=B1eIcvS45V.

Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12546–12558, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/949686ecef4ee20a62d16b4a2d7ccca3-Abstract.html.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432. Springer, 2016. doi: 10.1007/978-3-319-46723-8_49. URL https://link.springer.com/chapter/10.1007/978-3-319-46723-8_49.

Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, 1945. doi: 10.2307/1932409. URL https://doi.org/10.2307/1932409.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3558–3565, 2019. doi: 10.1609/aaai.v33i01.33013558. URL https://ojs.aaai.org/index.php/AAAI/article/view/4235. Issue 1.

Haifan Gong, Luoyao Kang, Yitao Wang, Yihan Wang, Xiang Wan, Xusheng Wu, and Haofeng Li. nnMamba: 3D biomedical image segmentation, classification and landmark detection with state space model. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE, 2025. doi: 10.1109/ISBI60581.2025.10980694.

Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R. Roth, and Daguang Xu. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 272–284. Springer, 2022a. doi: 10.1007/978-3-031-08999-2_22. URL https://link.springer.com/chapter/10.1007/978-3-031-08999-2_22.

Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R. Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 574–584. IEEE, 2022b. URL https://openaccess.thecvf.com/content/WACV2022/html/Hatamizadeh_UNETR_Transformers_for_3D_Medical_Image_Segmentation_WACV_2022_paper.html.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/He_Momentum_Contrast_for_Unsupervised_Visual_Representation_Learning_CVPR_2020_paper.html.

Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In *Advances in Neural Information Processing Systems*, volume 32, pp. 5658–5669, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/2d95666e2649fcfc6e3af75e09f5adb9-Abstract.html.

Jing Huang, Yongkang Zhao, Yuhan Li, Zhitao Dai, Cheng Chen, and Qiying Lai. ACM-UNet: Adaptive integration of CNNs and mamba for efficient medical image segmentation. *Biomedical Signal Processing and Control*, 123:110587, 2026. doi: 10.1016/j.bspc.2026.110587.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021. doi: 10.1038/s41592-020-01008-z. URL https://www.nature.com/articles/s41592-020-01008-z.

Jianwen Jiang, Yuxuan Wei, Yifan Feng, Jingxuan Cao, and Yue Gao. Dynamic hypergraph neural networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 2635–2641, 2019. doi: 10.24963/ijcai.2019/366.

Davood Karimi and Septimiu E. Salcudean. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39 (2):499–513, 2020. doi: 10.1109/TMI.2019.2930068. URL https://arxiv.org/abs/1904.10030.

Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67: 101851, 2021. doi: 10.1016/j.media.2020.101851. URL https://www.sciencedirect.com/science/article/pii/S1361841520302152.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18661–18673, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven C. H. Hoi. Prototypical contrastive learning of unsupervised representations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=KmykpuSrjcq.

Xingzhu Liang, Leilei Lu, Guoao Ge, Jinliang Zhang, Yu-e Lin, Yuhang Yao, and Nianhan Jie. NexMamba: Nexus-driven multi-domain mamba network for medical image segmentation. *Biomedical Signal Processing and Control*, 122:110385, 2026. doi: 10.1016/j.bspc.2026.110385.

Jun Ma, Feifei Li, and Bo Wang. U-Mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024. URL https://arxiv.org/abs/2401.04722.

Lena Maier-Hein, Annika Reinke, Patrick Godau, Minu D. Tizabi, Florian Buettner, Evangelia Christodoulou, Ben Glocker, Fabian Isensee, Jens Kleesiek, Michal Kozubek, Mauricio Reyes, Michael A. Riegler, Manuel Wiesenfarth, A. Emre Kavur, Carole H. Sudre, Michael Baumgartner, Matthias Eisenmann, Doreen Heckmann-Nötzel, Tim Rädsch, Laura Acion, Michela Antonelli, Tal Arbel, Spyridon Bakas, Arriel Benis, Matthew B. Blaschko, M. Jorge Cardoso, Veronika Cheplygina, Beth A. Cimini, Gary S. Collins, Keyvan Farahani, Luciana Ferrer, Adrian Galdran, Bram van Ginneken, Robert Haase, Daniel A. Hashimoto, Michael M. Hoffman, Merel Huisman, Pierre Jannin, Charles E. Kahn, Dagmar Kainmueller, Bernhard Kainz, Alexandros Karargyris, Alan Karthikesalingam, Florian Kofler, Annette Kopp-Schneider, Anna Kreshuk, Tahsin Kurc, Bennett A. Landman, Geert Litjens, Amin Madani, Klaus Maier-Hein, Anne L. Martel, Peter Mattson, Erik Meijering, Bjoern Menze, Karel G. M. Moons, Henning Müller, Brennan Nichyporuk, Felix Nickel, Jens Petersen, Nasir Rajpoot, Nicola Rieke, Julio Saez-Rodriguez, Clara I. Sánchez, Shravya Shetty, Maarten van Smeden, Ronald M. Summers, Abdel A. Taha, Aleksei Tiulpin, Sotirios A. Tsaftaris, Ben Van Calster, Gaël Varoquaux, and Paul F. Jäger. Metrics reloaded: recommendations for image analysis validation. *Nature Methods*, 21(2):195–212, 2024. doi: 10.1038/s41592-023-02151-z. URL https://www.nature.com/articles/s41592-023-02151-z.

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. doi: 10.1109/TMI.2014.2377694. URL https://pubmed.ncbi.nlm.nih.gov/25494501/.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision*, pp. 565–571. IEEE, 2016. doi: 10.1109/3DV.2016.79. URL https://arxiv.org/abs/1606.04797.

Saqib Qamar, Mohd Fazil, Parvez Ahmad, Shakir Khan, and Abu Taha Zamani. UNet with self-adaptive mamba-like attention and causal-resonance learning for medical image segmentation. *Scientific Reports*, 16:135, 2026. doi: 10.1038/s41598-025-28885-8.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28. URL https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28.

Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F. Jäger, and Klaus H. Maier-Hein. MedNeXt: Transformer-driven scaling of ConvNets for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 405–415. Springer, 2023. doi: 10.1007/978-3-031-43901-8_39.

Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. clDice – a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16560–16569, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Shit_clDice_-_A_Novel_Topology-Preserving_Loss_Function_for_Tubular_Structure_CVPR_2021_paper.html.

Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15:29, 2015. doi: 10.1186/s12880-015-0068-x. URL https://doi.org/10.1186/s12880-015-0068-x.

Wenxuan Wang, Chen Chen, Meng Ding, Hong Yu, Sen Zha, and Jiangyun Li. TransBTS: Multimodal brain tumor segmentation using transformer. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 109–119. Springer, 2021. doi: 10.1007/978-3-030-87193-2_11. URL https://link.springer.com/chapter/10.1007/978-3-030-87193-2_11.

Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. CoTr: Efficiently bridging CNN and transformer for 3D medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 171–180. Springer, 2021. doi: 10.1007/978-3-030-87199-4_16. URL https://link.springer.com/chapter/10.1007/978-3-030-87199-4_16.

Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. SegMamba: Long-range sequential modeling mamba for 3d medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 578–588. Springer, 2024. doi: 10.1007/978-3-031-72111-3_54. URL https://link.springer.com/chapter/10.1007/978-3-031-72111-3_54.

Feng Yu, Jiacheng Cao, Li Liu, and Minghua Jiang. SuperLightNet: Lightweight parameter aggregation network for multimodal brain tumor segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Yu_SuperLightNet_Lightweight_Parameter_Aggregation_Network_for_Multimodal_Brain_Tumor_Segmentation_CVPR_2025_paper.html.

Hongsheng Zhang, Yuting Duan, Ting Liu, Weifeng Zhang, and Hongzhong Tang. KMUNet: A novel medical image segmentation model based on KAN and mamba. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15969 of *Lecture Notes in Computer Science*, pp. 298–307. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-032-05127-1_29.

Yibo Zhang. KM-UNet: KAN mamba UNet for medical image segmentation. *arXiv preprint arXiv:2501.02559*, 2025. URL https://arxiv.org/abs/2501.02559.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems*, volume 19, pp. 1601–1608, 2006. URL https://proceedings.neurips.cc/paper_files/paper/2006/hash/dff8e9c2ac33381546d96deea9922999-Abstract.html.

Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Xiaoguang Han, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnFormer: Volumetric medical image segmentation via a 3D transformer. *IEEE Transactions on Image Processing*, 32:4036–4045, 2023. doi: 10.1109/TIP.2023.3293771. URL https://doi.org/10.1109/TIP.2023.3293771.

A SEGMENTATION OBJECTIVE

The segmentation objective applies binary cross-entropy and Dice losses to the three foreground tissue labels. These exclusive tissue labels differ from the overlapping ET, TC, and WT regions used for evaluation. Let $\kappa(1) = 1$, $\kappa(2) = 2$, and $\kappa(4) = 3$ map semantic labels to output channels. For $c \in \{1, 2, 4\}$, define $Y_c(p) = \mathbf{1}\{Y(p) = c\}$ and $Q_c(p) = \sigma(O_{\text{seg},\kappa(c)}(p))$. Then

$$\mathcal{L}_{\text{seg}} = \sum_{c\in\{1,2,4\}} \left[1 - \frac{2\sum_p Q_c(p)Y_c(p) + \epsilon}{\sum_p Q_c(p) + \sum_p Y_c(p) + \epsilon} + \text{BCE}(Q_c, Y_c) \right]. \tag{9}$$

B FULL WITHIN-DATASET RESULTS AND MODEL PROFILES

Table 7 reports all source validation folds. The advantage is most consistent in SDC, where DGP-HPCL leads all five folds.

Tables 5 and 6 separate development measurements from evaluation and profiling. Training Soft Dice is an optimization diagnostic rather than a held-out score. Validation favors nnU-Net in DSC and SDC, whereas the full model leads the evaluation metrics. Relative to nnU-Net, DGP-HPCL has approximately 73% fewer reported parameters and 74% fewer reported FLOPs. The recorded 0.09 T versus 0.16 T values for SuperLightNet and DGP-HPCL cannot be used to claim identical inference cost or measured training overhead.

Table 5: BraTS 2021 development measurements. Training Soft Dice, validation DSC, and validation SDC are percentages.

Method	Train Soft Dice	Valid DSC	Valid SDC	Valid HD95
nnU-Net (2021)	86.25	**89.51**	**85.37**	<u>5.59</u>
TransBTS (2021)	78.58	84.77	75.69	7.10
Swin UNETR (2022)	81.66	88.72	82.24	7.14
MedNeXt (2023)	77.29	80.96	68.48	10.87
SegMamba (2024)	81.89	<u>89.23</u>	83.72	6.29
KM-UNet (2025)	83.74	86.48	80.33	8.29
KMUNet (2025)	80.55	83.00	74.96	17.92
SAMA-UNet (2026)	80.82	84.63	77.10	15.52
SuperLightNet (2025)	<u>87.19</u>	88.96	84.98	6.24
ACM-UNet (2026)	83.84	87.11	81.33	8.03
DGP-HPCL (Ours)	**87.48**	88.79	<u>85.28</u>	**5.47**

Table 6: BraTS 2021 evaluation and reported model profiles. DSC and SDC are percentages; parameters are in millions and FLOPs in trillions.

Method	DSC	SDC	HD95	Params	FLOPs
nnU-Net (2021)	<u>88.78</u>	83.86	6.30	11.06	0.61
TransBTS (2021)	84.51	75.75	6.24	32.99	0.53
Swin UNETR (2022)	88.55	81.08	7.20	62.19	1.59
MedNeXt (2023)	80.67	67.70	10.17	11.65	0.83
SegMamba (2024)	88.56	82.41	5.96	66.90	3.15
KM-UNet (2025)	87.40	80.53	8.34	6.95	0.28
KMUNet (2025)	84.18	75.70	15.45	27.14	**0.07**
SAMA-UNet (2026)	84.45	76.30	15.50	5.84	0.14
SuperLightNet (2025)	88.64	<u>84.12</u>	<u>5.84</u>	**2.97**	<u>0.09</u>
ACM-UNet (2026)	87.93	81.68	6.76	6.30	0.38
DGP-HPCL (Ours)	**88.87**	**84.52**	**5.14**	<u>2.99</u>	0.16

Table 7: Full BraTS 2019 five-fold results. Fold 1–Fold 5 identify source validation folds. DSC and SDC are percentages.

Method	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
nnU-Net (2021)	DSC	82.30	80.95	81.83	85.69	78.18	81.79
	SDC	73.15	72.05	72.16	77.13	66.31	72.16
	HD95	7.93	8.98	7.11	5.99	8.86	7.77
TransBTS (2021)	DSC	75.97	75.81	73.26	82.26	67.75	75.01
	SDC	59.13	58.92	54.12	67.14	49.74	57.81
	HD95	9.60	10.80	9.73	8.47	11.20	9.96
Swin UNETR (2022)	DSC	82.35	81.59	**82.07**	85.39	77.86	81.85
	SDC	73.08	72.46	71.54	76.07	66.18	71.87
	HD95	6.85	10.97	**6.66**	6.57	7.24	7.66
MedNeXt (2023)	DSC	82.18	79.74	80.04	85.12	76.48	80.71
	SDC	71.49	67.92	67.36	73.96	63.31	68.81
	HD95	8.25	10.66	8.26	6.73	9.58	8.69
SegMamba (2024)	DSC	**82.57**	81.80	81.87	85.38	77.63	81.85
	SDC	72.85	71.66	71.66	74.94	64.46	71.11
	HD95	7.80	9.45	6.92	6.28	7.84	7.66
KM-UNet (2025)	DSC	81.60	79.73	78.77	84.16	74.95	79.84
	SDC	70.42	69.36	67.63	73.68	60.95	68.41
	HD95	9.51	10.77	9.10	8.02	13.76	10.23
KMUNet (2025)	DSC	78.20	77.85	77.21	82.11	72.13	77.50
	SDC	65.57	66.25	64.23	69.24	58.05	64.67
	HD95	10.43	17.99	16.87	8.48	22.94	15.34
SAMA-UNet (2026)	DSC	78.87	77.45	77.92	82.13	73.22	77.92
	SDC	67.47	65.73	66.17	69.46	59.32	65.63
	HD95	8.81	20.91	10.08	13.61	14.11	13.50
SuperLightNet (2025)	DSC	82.49	81.41	81.73	85.96	77.27	81.77
	SDC	73.87	72.59	72.82	76.35	65.51	72.23
	HD95	7.68	**8.27**	6.76	6.05	10.22	7.80
ACM-UNet (2026)	DSC	81.80	80.54	79.78	85.24	75.43	80.56
	SDC	71.80	70.73	69.85	75.67	64.02	70.41
	HD95	9.51	10.67	8.85	7.89	10.03	9.39
DGP-HPCL (Ours)	DSC	82.52	**82.61**	81.58	**86.88**	**78.30**	**82.38**
	SDC	**74.17**	**73.71**	**73.16**	**79.23**	**67.41**	**73.54**
	HD95	**6.06**	8.79	8.45	**5.55**	**5.71**	**6.91**

C DIRECT CROSS-YEAR EVALUATION

The five target columns correspond to separately trained BraTS 2019 fold models evaluated on one BraTS 2021 target set. They are not independent target folds and are not ensemble predictions. The target mean is the arithmetic mean of five model-level metrics. DGP-HPCL leads target DSC for three model instances and SDC for four; target HD95 is best for two. Compared directly with SuperLightNet, the mean changes are +0.31 pp DSC, +0.85 pp SDC, and +0.06 HD95. Higher target than source scores do not mean that transfer improves the model, because the underlying subject sets differ. Release-year separation alone does not verify patient-level independence between BraTS cohorts.

Table 8: Direct cross-year results. Source gives the BraTS 2019 validation mean. Model 1–Model 5 give BraTS 2021 scores for the corresponding source-fold models; Mean is their arithmetic mean. DSC and SDC are percentages.

Method	Metric	Source	Model 1	Model 2	Model 3	Model 4	Model 5	Mean
nnU-Net (2021)	DSC	81.79	86.02	85.07	85.64	86.37	85.24	85.67
	SDC	72.16	78.83	78.18	78.50	79.80	77.96	78.65
	HD95	7.77	8.96	9.09	8.65	7.49	9.14	8.67
TransBTS (2021)	DSC	75.01	75.37	78.24	73.17	78.88	72.69	75.67
	SDC	57.81	58.11	61.21	54.68	64.92	54.32	58.65
	HD95	9.96	11.32	10.73	12.15	10.70	11.98	11.38
Swin UNETR (2022)	DSC	81.85	85.61	86.93	**86.37**	86.09	85.12	86.02
	SDC	71.87	76.79	78.56	77.81	77.83	77.33	77.66
	HD95	7.66	8.45	8.62	8.28	8.00	**7.32**	8.13
MedNeXt (2023)	DSC	80.71	84.03	83.84	82.47	84.05	84.01	83.68
	SDC	68.81	74.66	73.52	71.56	74.81	74.28	73.77
	HD95	8.69	9.21	9.13	9.90	9.06	10.42	9.54
SegMamba (2024)	DSC	81.85	85.26	86.54	85.28	84.62	84.02	85.14
	SDC	71.11	77.40	79.66	77.66	75.94	74.77	77.09
	HD95	7.66	8.47	7.84	8.69	8.41	9.43	8.57
SuperLightNet (2025)	DSC	81.77	**86.48**	86.69	85.48	85.36	85.35	85.87
	SDC	72.23	**80.70**	81.06	79.40	78.51	79.06	79.75
	HD95	7.80	6.97	**7.32**	**8.11**	7.57	7.41	**7.48**
ACM-UNet (2026)	DSC	80.56	85.89	85.90	84.46	85.79	85.46	85.50
	SDC	70.41	78.65	78.66	76.48	78.18	78.05	78.00
	HD95	9.39	8.29	9.90	9.44	10.23	11.46	9.86
DGP-HPCL (Ours)	DSC	**82.38**	86.29	**87.02**	85.42	**86.46**	**85.71**	**86.18**
	SDC	**73.54**	80.64	**81.34**	**79.93**	**80.98**	**80.09**	**80.60**
	HD95	**6.91**	**6.47**	7.81	8.78	**7.06**	7.57	7.54

D COMPONENT COMPARISONS

Tables 9 and 3 report the development and evaluation measurements. C and D indicate the corresponding priors, GCN denotes the graph variant, HG denotes hypergraph aggregation, and HPCL denotes endpoint prototype contrast. Auxiliary priors alone improve SDC but do not reproduce the full model's evaluation gains. The GCN relation variant reduces HD95 to 5.20 while its DSC and SDC remain below the baseline. The HG relation variant reaches 88.70% DSC, 84.36% SDC, and 5.18 HD95. The full model adds 0.17 pp DSC and 0.16 pp SDC, with a further 0.04 reduction in HD95. Removing either prior weakens the full evaluation result; removing the distance prior produces the larger decrease in DSC and SDC. The available comparisons evaluate these configurations jointly and do not establish the independent effect of BSC.

Table 9: Component comparison: training and validation measurements on BraTS 2021. Soft Dice, DSC, and SDC are percentages.

Variant	Train Soft Dice	Valid DSC	Valid SDC	Valid HD95
Baseline	<u>87.19</u>	<u>88.96</u>	84.98	6.24
Dual priors	85.72	88.57	84.74	7.43
GCN relation	85.89	88.78	85.24	5.60
HG relation	86.41	<u>88.96</u>	<u>85.25</u>	5.75
w/o D	85.73	88.78	84.69	5.99
w/o C	86.21	**89.01**	85.15	**5.32**
DGP-HPCL (Ours)	**87.48**	88.79	**85.28**	<u>5.47</u>

E FULL SENSITIVITY MEASUREMENTS

Table 10 reports all tested one-factor configurations; the corresponding analysis appears in Section 4.5.

Table 10: Hyperparameter sensitivity on BraTS 2019 fold 1. DSC and SDC are percentages. Ranking is across the listed configurations.

Setting	Configuration	DSC	SDC	HD95
Default	Default	**82.52**	**74.17**	**6.06**
Node count	$K = 128$	80.75	70.80	7.56
Node count	$K = 512$	81.85	72.29	7.33
Hyperedge size	$k = 4$	81.96	71.61	8.14
Hyperedge size	$k = 16$	82.28	73.04	<u>6.72</u>
Temperature	$\tau = 0.05$	82.08	72.46	8.14
Temperature	$\tau = 0.20$	81.78	73.30	7.08
Contrastive weight	$\lambda_{\mathrm{HPCL}} = 0.01$	81.59	71.66	7.39
Contrastive weight	$\lambda_{\mathrm{HPCL}} = 0.10$	81.38	72.21	7.51
Endpoint fraction	$\rho = 0.10$	81.20	70.51	7.40
Endpoint fraction	$\rho = 0.30$	81.30	71.02	8.24
Prior fusion	$\alpha = 0.30$	81.45	72.12	7.15
Prior fusion	$\alpha = 0.50$	<u>82.33</u>	<u>73.63</u>	7.72
Prior fusion	$\alpha = 0.90$	81.53	71.84	7.46