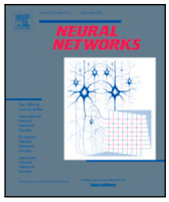
Neural Networks 180 (2024) 106657

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

Reconstruct incomplete relation for incomplete modality brain tumor segmentation

Jiawei Su [a,b], Zhiming Luo [b,*], Chengji Wang [c], Sheng Lian [d], Xuejuan Lin [e], Shaozi Li [b]

[a] School of Computer Engineering, Jimei University, Xiamen, China
[b] The Department of Artificial Intelligence, Xiamen University, Fujian, China
[c] The School of Computer Science, Central China Normal University, Wuhan, China
[d] The College of Computer and Data Science, Fuzhou University, Fujian, China
[e] The Department of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fujian, China

ARTICLE INFO

Keywords:
Brain tumor segmentation
Incomplete modalities
Structural relation knowledge
Knowledge distillation

ABSTRACT

Different brain tumor magnetic resonance imaging (MRI) modalities provide diverse tumor-specific information. Previous works have enhanced brain tumor segmentation performance by integrating multiple MRI modalities. However, multi-modal MRI data are often unavailable in clinical practice. An incomplete modality leads to missing tumor-specific information, which degrades the performance of existing models. Various strategies have been proposed to transfer knowledge from a full modality network (teacher) to an incomplete modality one (student) to address this issue. However, they neglect the fact that brain tumor segmentation is a structural prediction problem that requires voxel semantic relations. In this paper, we propose a Reconstruct Incomplete Relation Network (RIRN) that transfers voxel semantic relational knowledge from the teacher to the student. Specifically, we propose two types of voxel relations to incorporate structural knowledge: Class-relative relations (CRR) and Class-agnostic relations (CAR). The CRR groups voxels into different tumor regions and constructs a relation between them. The CAR builds a global relation between all voxel features, complementing the local inter-region relation. Moreover, we use adversarial learning to align the holistic structural prediction between the teacher and the student. Extensive experimentation on both the BraTS 2018 and BraTS 2020 datasets establishes that our method outperforms all state-of-the-art approaches.

1. Introduction

Accurate medical image segmentation in magnetic resonance imaging (MRI) plays an essential role during the diagnosis (Gridach, 2021; Malik & Bzdok, 2022; Wang & Chung, 2022). MRI images of brain tumors possess a variety of different modality data (i.e., T1, T1ce, T2, and Flair). The different modality data provide each other with complementary information about the brain structure and physiopathology (Bakas et al., 2017; Pang, Du, Orgun, Wang, & Yu, 2021; Zhuang, Liu, Song, & Hung, 2022). By comparing them, tumor regions are manually segmented into heterogeneous subregions (i.e., GD-enhancing tumor (ET), peritumoral edema (ED), and the necrotic and non-enhancing tumor core (NCR/NET)) (Menze et al., 2014). These subregions are combined into three nested subregions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET). Thus, existing deep learning-based methods (Wu et al., 2021) improve the accuracy of segmentation by fusing MRI from different modalities (Chen, Liu, Ding, Zheng, & Li, 2019; Fidon et al., 2017; Isensee, Jaeger, Kohl, Petersen, & Maier-Hein, 2021; Wang et al., 2020; Zhang et al., 2020). In a clinical scenario, however, it is often difficult to obtain data for all modalities simultaneously due to the limited number of physicians and scanners (Dorent, Joutard, Modat, Ourselin, & Vercauteren, 2019; Havaei, Guizard, Chapados, & Bengio, 2016). In particular, only the T1 MRI, which is the most commonly obtained, can be taken to make the important clinical evaluation (Huang et al., 2019; Orbes-Arteaga et al., 2018; Yu, Zhou, Wang, Fripp, & Bourgeat, 2018). In this case, the well-trained model performs poorly under clinical conditions. It is of great practical importance to study the problem of brain tumor segmentation under incomplete modality conditions.

To deal with those problems where multiple modalities are available at training time but some or most of them are missing at inference time, many works have been proposed (Ben-Cohen et al., 2019; Dorent et al., 2019; Havaei et al., 2016; van Tulder & de Bruijne, 2018). In general, it can be classified into three main categories: **Feature projection, Feature synthesis, and Knowledge transfer**. The **Feature projection** methods usually learn common representations from the

* Corresponding author.
E-mail address: zhiming.luo@xmu.edu.cn (Z. Luo).

https://doi.org/10.1016/j.neunet.2024.106657
Received 28 November 2023; Received in revised form 14 August 2024; Accepted 19 August 2024
Available online 22 August 2024

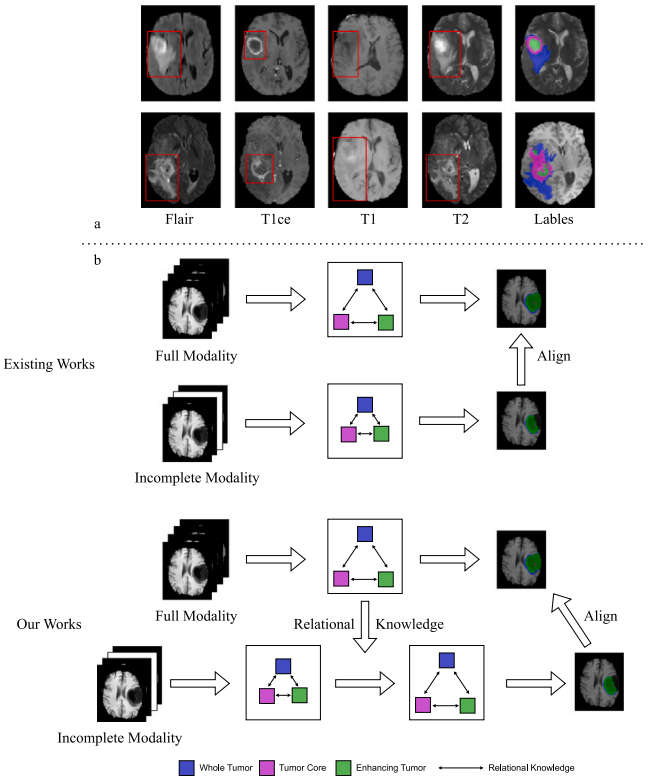


Fig. 1. (a) Distinct brain tumor modalities data contain diverse tumor-specific information. Due to the lack of specific modal data, the model fails to obtain specific information about the brain tumor regions. (b) Compared with the full modality, the incomplete modality cannot provide the model with sufficient information about the tumor region, making it difficult to establish appropriate voxel relations. This results in insufficient clarity regarding the distinctions between different categories. Different colors represent different brain tumor regions (Whole Tumor, Tumor Core, and Enhancing Tumor). Considering that semantic segmentation is a structured prediction problem, the poor voxel relation will degrade the segmentation performance. So, to build a robust structural prediction in an incomplete modality situation, we attempt to reconstruct the voxel relation by learning the voxel relational knowledge from the model trained with full-modality data.

full modalities (Dorent et al., 2019; Havaei et al., 2016; van Tulder & de Bruijne, 2018). The **Feature synthesis** attempts to train a generative model to synthesize the missing modalities (Ben-Cohen et al., 2019; Tulder & Bruijne, 2015). In general, the first two works perform poorly when more than one modality is missing, which is more common in clinical practice. Therefore, instead of training one general model to deal with all missing situations, the **Knowledge transfer** attempt to build a series of independent models dedicated to each missing situation (Azad, Khosravi, & Merhof, 2022; Chen, Dou, Jin, Liu, & Heng, 2021; Hu et al., 2020; Wang et al., 2021). However, the preceding work ignores the fact that semantic segmentation is a structural prediction problem and instead focuses on aligning the distinct voxel features separately in the intermediate or output layers. To obtain reliable structural prediction, the model should figure out the relation between the different voxels. It motivated us to fully exploited voxel relation to address the brain tumor segmentation under the incomplete modality situation.

As shown in Fig. 1(a), it is obvious that different modal data have a distinct appearance on diverse tumor regions. In the case of incomplete modality, a part of the tumor-specific information is not available to the model. Without sufficient information, the model cannot accurately distinguish between voxels in different tumor regions. This means that the similarity relation between voxels is ambiguous. However, semantic segmentation is a structured prediction task, which requires establishing clear voxel relations. Therefore, we assume that in the case of incomplete modality, the failure of the model to construct a reliable relation between voxels degrades the segmentation performance. Motivated by this assumption, in this paper, we attempt to address the problem of brain tumor segmentation under the incomplete modality condition by reconstructing reliable voxel relations.

In contrast to the previous approaches, we introduce Reconstruct Incomplete Relation Network (RIRN), a novel framework to explore the voxel similarity relationship. We transfer the relational knowledge from a full modality network (teacher) to an incomplete modality (student) towards more accurate incomplete modality segmentation during inference. With the additional guidance of the teacher model, the student model can reconstruct the voxel relation and produce more accurate structural predictions in incomplete modality situations. Specifically, we divide the relations between voxels into two categories: Class-relative relations (CRR) and Class-agnostic relations (CAR).

The CRR groups voxels into the three tumor regions (WT, ET, and TC) and establishes a relation between them. Under the supervision of the teacher model, the student model can utilize CRR to produce reliable inter-region relations and classify voxels into the correct tumor region. However, the CRR constructs a local tumor region relation solely, which ignores the relationship between the tumor region and the non-tumor region. Complementary to CRR, we also propose CAR to establish the global relation between all pairs of voxels on the intermediate features. Overall, by learning class-relative and class-agnostic relations in the teacher model, the student model can reconstruct the voxel relation to obtain improved structural prediction results. Moreover, we integrate the adversarial training strategy (Xu, Hsu, & Huang, 2017) into our model to align holistic structural prediction. It aligns the output distribution in the teacher model and the student model, which further improves segmentation accuracy. Inspired by Qiao et al. (2018), the RIRN is trained in a co-training approach. On the BraTS 2018 dataset and BraTS 2020 dataset, we obtain greater accuracy using the suggested framework than the state-of-the-art approaches. This illustrates our method's superiority.

Overall, our contributions are threefold:

- We propose RIRN for brain tumor segmentation under the incomplete modality situation. Instead of aligning voxel features separately as before, we explicitly propose to improve the performance of brain tumor segmentation by reconstructing the voxel relation.
- We propose two types of voxel relation: CRR and CAR. They distill the local class-relative relation and global class-agnostic relation from the teacher to the student model, respectively. Both of them have a significant effect on improving brain tumor segmentation in incomplete modality situations.
- Taking advantage of the proposed framework, we achieve state-of-the-art performance on the BraTS 2018 dataset and BraTS 2020 dataset.

2. Related work

Multimodal medical images present varying contrasts and complementary information for anatomies or lesion areas (Menze et al., 2014). To utilize the valuable multimodal information, previous approaches have attempted to fuse this multimodal information at different levels, typically by concatenating multiple modalities as inputs (Kamnitsas et al., 2017; Pereira, Pinto, Alves, & Silva, 2016; Zhou, Ding, Lu, Wang, & Tao, 2018) or fusing intermediate multimodal representations (Ding, Zheng, et al., 2021; Ding et al., 2018; Xie et al., 2023). However, these methods require the use of full modality data during testing, which may not be met under clinical conditions. So, it is valuable to study the solution of incomplete modality segmentation. In this section, we briefly review the most related works on incomplete modality brain tumor segmentation and knowledge distillation.

2.1. Incomplete modality brain tumor segmentation

The previous work can be divided into three main types roughly: **Feature projection**, **Feature synthesis**, and **Knowledge transfer**.

Feature projection methods first extracted common representations from the full modalities. The features are then projected into a uniform feature space and a uniform model is built for brain tumor segmentation. Havaei et al. (2016) proposed HeMIS, which trains a different feature extractor for each modality. And then the decoder is trained to predict the segmentation map through project those features to the latent feature maps. Inspired by HeMIS, Dorent et al. (2019) proposed U-HVED, which introduces a hetero-modal variational encoder–decoder for tumor segmentation and missing modality completion. van Tulder and de Bruijne (2018) minimizing the L1 distance of each modality feature to learn modality-invariant representation in latent space.

The goal of **Feature synthesis** is to train a generative model to synthesize missing modalities. And then build the segmentation model based on multi-modalities. Tulder and Bruijne (2015) have shown that the accuracy of classification of brain tumors can be promoted by using the synthesized modality data. Ben-Cohen et al. (2019) present a novel system for the generation of virtual PET images by using CT scans. However, this strategy is very demanding and computationally complex for data generation models. Overall, it is hard for the first two works to generate common representations or synthesize the missing modalities when more than one modality is missing, which is more common in clinical. It limits their performance and applicability in clinical settings.

Therefore, instead of training one general model to deal with all missing situation, the **Knowledge transfer** attempts to construct a set of separate models for each missing circumstance. Hu et al. (2020) propose a framework to distill knowledge from a multi-modal network (teacher) to a mono-modal one (student). Based on this work, Wang et al. (2021) further propose an adversarial knowledge distillation network. Azad et al. (2022) decompose the representation space into content and style representations and utilize the matching modules to reconstruct the missing information. Chen et al. (2021) introduce contrastive learning to fully exploit the privileged knowledge of multimodal training data towards more accurate unimodal segmentation. In general, the above work only focuses on aligning the separate voxel features in the intermediate or output layers to narrow the gap between the teacher and student model. They ignored the fact that semantic segmentation is a structural prediction task and treated it as a pixel classification problem. In contrast to previous work, we focus on developing more precise structural predictions by fully exploring the relation between different voxels.

2.2. Knowledge distillation

Hinton et al. (2015) firstly introduced knowledge distillation to transfer the dark knowledge from the teacher model to the student model by minimizing the KL-divergence between them at the output layer. Recently, KD has been applied to a variety of tasks, such as image classification (Xu, Rui, Li, & Gu, 2020; Zhao, Cui, Song, Qiu, & Liang, 2022), object detection (Chen, Choi, Yu, Han, & Chandraker, 2017; Wang, Yuan, Zhang, & Feng, 2019), and semantic segmentation (Dou, Liu, Heng, & Glocker, 2020; Liu et al., 2019; Liu, Desrosiers, Ayed, & Dolz, 2023). In general, most current KD methods are mainly divided into response-based, feature-based, and relation-based approaches (Gou, Yu, Maybank, & Tao, 2021). Response-based KD usually directly mimics the class probabilities produced from the teacher model. The main idea of Feature-based method is to match the intermediate feature map or the refined information of the teacher and the student. Relation-based KD further aligns corrections between different layers or data samples. Different from previous work (Azad et al., 2022; Hu et al., 2020; Wang et al., 2021) that separately align each voxel to narrow the gap between the teacher and the student, we consider knowledge distillation as a structured prediction problem in incomplete modality brain tumor segmentation. We explore the relational knowledge among all pairs of voxels from the feature maps of the teacher model, which significantly improves learning better voxel-wise representations.

3. Method

In this work, we designed RIRN for brain tumor segmentation under the incomplete modality situation, which works in a teacher-student model. Specifically, the RIRN contains three main components, including CAR, CRR, and ADV. The CAR and CRR attempt to reconstruct the relation between different voxels in order to make a more reliable structural prediction. Additionally, the ADV aligns comprehensive structural prediction between the teacher and student to enhance segmentation performance. In this section, we first give the definition of our problem and then introduce our designed RIRN in detail.

3.1. Problem definition

The purpose of incomplete modality brain tumor segmentation is to segment three brain tumor regions (i.e., WT, TC, and ET) in the situation that some modality is missing at inference. Concretely, we denote the full modality data as $X_{full} \in \mathbb{R}^{H\times W\times D\times M}$. And the incomplete modality data is denoted as $X_{incomplete} \in \mathbb{R}^{H\times W\times D\times N}$. The corresponding ground truth mask denoted as $Y_{GT} \in \{WT, TC, ET\}^{H\times W\times D}$. In our setting, the X_{full} and $X_{incomplete}$ denotes 3D modality image, with spatial dimension(H, W, D). M shows the set of four modalities (T1, T2, T1ce, and Flair). N represents the number of available modal data, ranging from 1 to 4. To demonstrate the superiority of our method, we evaluate the segmentation results on all fifteen combinations of the four modalities.

3.2. Co-training network

The pipeline of our proposed framework is illustrated in Fig. 2. Our approach considers the co-training strategy's ability to extract relational information from the teacher model and transfer it to the student model. In particular, we proposed two types of voxel relation: CRR and CAR. The CRR captures the relation between different local tumor regions. Furthermore, in addition to CRR, the CAR established a relationship based on global voxel features. To further improve segmentation performance, we also attempt to align the holistic structural prediction between teacher and student by using adversarial training. The teacher model and the student model share the same 3D U-Net (Çiçek, Lienkamp, Brox, & Ronneberger, 2016) architecture and are trained independently and simultaneously. Both of them are trained with Dice loss, which is commonly used in medical segmentation tasks. We defined the basic segmentation loss as:

$$L_{SEG} = L^{Dice}_{full} + L^{Dice}_{incomplete}. \tag{1}$$

In the subsection, we will elaborate on each module in detail.

3.3. Class-relative relation (CRR)

Because of the lack of a distinct modality, the model cannot obtain enough information to distinguish across tumor areas, resulting in an unclear inter-region relationship. It results in poor structural prediction and decreases the segmentation performance. Therefore, we attempt to reconstruct the class-relative relation under the supervision of the teacher model. As shown in Fig. 3, the CRR establishes relations between different tumor regions. We first divide the voxels into three tumor regions based on ground truth labels. Then, to acquire strong relationship information from the teacher model, we filtered out features that the teacher model mistakenly predicted. We established the class center by averaging the correctly predicted voxel features in the same

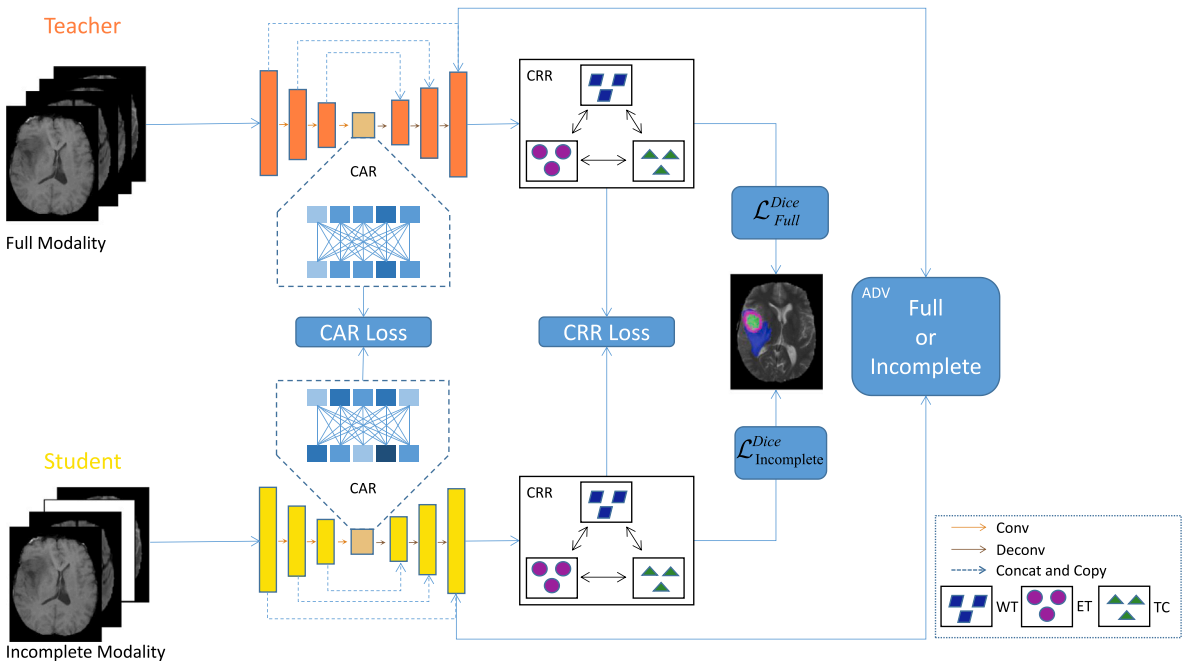


Fig. 2. Overview of our proposed RIRN. The CRR builds relationships between local tumor regions. As a complement to CRR, the CAR distills the global voxel relational knowledge from the teacher. Furthermore, the ADV aligns the holistic structural predictions of the teacher model and the student model. WT, TC, and ET mean whole tumor, tumor core, and enhancing tumor, respectively.

tumor location. We utilize the class center to represent different tumor regions. And then the similarity is counted between them to obtain the class-relative relationships. Finally, we transfer relational knowledge by aligning the CRR of the teacher model and the student model.

Let f^T_{region} and f^S_{region} denote the regional feature of the teacher and student, respectively. We get f^T_{region} and f^S_{region} at the penultimate layer of the model. The $region \in \{WT, TC, ET\}$. As illustrated in Fig. 3, we construct a graph structure $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ to encode the CRR, where $\mathcal{V} = \{v_i | i = 1, 2, 3\}$ is a group of nodes and $\mathcal{E} = \{e_{i,j} \mid i = 1, 2, 3; j = 1, 2, 3; i \neq j\}$ represents a group of edges. v_i denotes the regional features of ith region, $e_{i,j}$ is the regional relation, which is constructed between ith and jth region. Intuitively, we can directly get nodes f^T_{region} according to regional labels and define edges by computing the relation with a distance function. However, a larger tumor area corresponds to a greater number of pixels. When calculating similarity and loss across all pixels, the larger tumor area disproportionately influences the overall loss, leading to higher cumulative loss values. Consequently, during the optimization process, the model focuses more on the larger tumor region, potentially neglecting the smaller tumor regions. To mitigate this issue, we calculate the mean value to represent the category features based on the number of pixels in the tumor region. Hence, we denote the f^T_{region} as:

$$f^T_{region} = \frac{f^T_p * M_{region}}{\sum M_{region}}, \tag{2}$$

where f^T_p is the feature of the penultimate layer of the model and M_{region} is the mask of f^T_p. The f^S_{region} is proposed in the same approach. Intuitively, M_{region} can be obtained by upsampling ground truth to the same size as f^T_p. However, to ensure that student models learn more effectively from the teacher, we only apply the CRR on the voxels that the teacher predicted correctly. We filter out features with wrong predictions. So, we denote M_{region} as:

$$M_{region} = \mathbb{I}(P_{region} > \lambda) \cap GT_{region}, \tag{3}$$

where the $\mathbb{I}(\cdot)$ is indicated function, the P_{region} means the prediction result of the teacher and GT_{region} denotes ground truth. GT_{region} is 1 if the feature belongs to the region, and 0 otherwise. λ represents the confidence of the teacher model prediction. We empirically set the λ as 0.5.

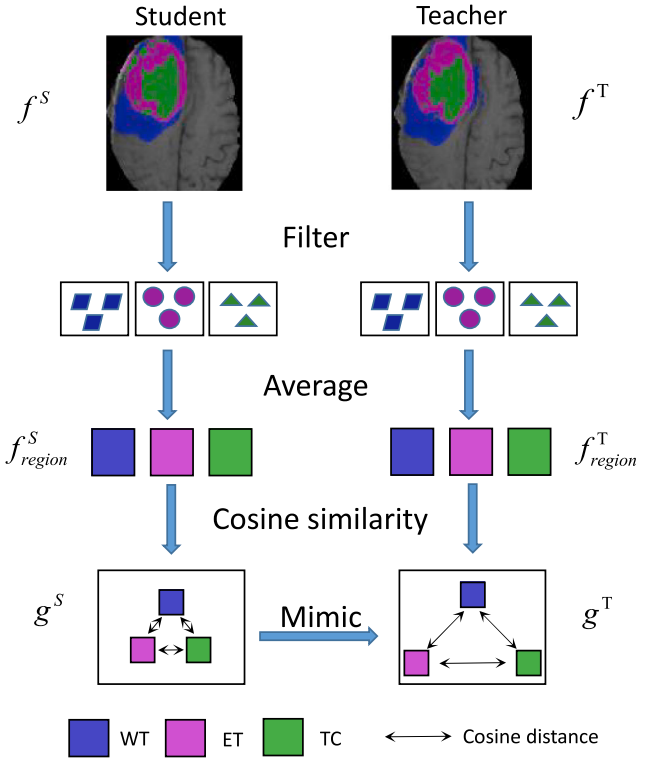


Fig. 3. Illustration of the CRR process. We construct a class-relative relationship between different tumor regional features. Different colors represent different brain tumor regions (WT, TC, and ET). f^T_{region} and f^S_{region} denote class feature from teacher and student. g^T and g^S denote the CRR in the teacher and the student.

Then, we obtained the relation between tumor regional features from the teacher model and the student model separately. We use the

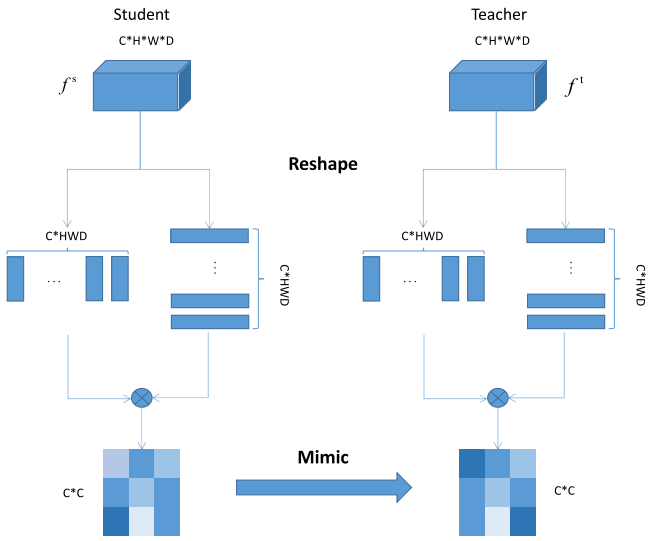


Fig. 4. The architecture or CAR. We adopt a non-local module to build class-agnostic relations between different voxels.

cosine function to calculate the similarity between them. As shown in Fig. 3, the CRR denoted as:

$$CRR\left(f^T_{region_i}, f^T_{region_j}\right) = \cos\left(f^T_{region_i}, f^T_{region_j}\right), i \neq j. \tag{4}$$

For convenience, we denote $CRR\left(f^T_{region_i}, f^T_{region_j}\right)$ as $CRR\left(i^T, j^T\right)$. Finally, the student model mimics the regional relation graph of the teacher. We construct the CRR loss as:

$$L_{CRR} = \sum_{i=1}^{M}\sum_{j=1}^{M}\left|CRR\left(i^T, j^T\right) - CRR\left(i^S, j^S\right)\right|, i \neq j, \tag{5}$$

where the M is the number of brain tumor regions (i.e., WT, TC and ET), which is 3 in our task.

3.4. Class-agnostic relation (CAR)

The absence of modality not only leads to a weakened ability to delineate tumor regions, but also makes the model less capable of judging tumor and non-tumor regions. The CRR built a local relation between different tumor regions, however, ignoring the relationship between tumor and non-tumor regions. It also contains critical relational knowledge that can help the model identify tumor regions, particularly at the boundary between tumor and non-tumor regions. With complement to CRR, we further propose CAR, which is performed on global voxels features in the intermediate feature layer. The CAR can build long-range spatial semantic dependency to capture global contextual information, which can make more precise segmentations at the tumor boundary. As shown in Fig. 4, we adapt the Non-local module (Wang, Girshick, Gupta, & He, 2018) to capture the global relation knowledge. We build the class-agnostic relational matrix by counting the similarity between every pair of voxels. We argue that the class-agnostic relational matrix contains extensive structural information. The student can learn more about the tumor structure by mimicking the relation matrix from the teacher.

Assume that $f^t \in \mathbb{R}^{C\times H\times W\times D}$ and $f^s \in \mathbb{R}^{C\times H\times W\times D}$ denote as the intermediate feature from teacher and student, respectively. Then we construct the relation matrix as:

$$CAR(f^t) = softmax(e^{\theta(f^t_i)^T\phi(f^t_j)}), \tag{6}$$

where $\theta(f^t_i) = W_\theta f^t_i$ and $\phi(f^t_j) = W_\phi f^t_j$ are two embeddings. Both W_θ and W_ϕ are $1\times1\times1$ convolutions. And we build $CAR(f^s_i, f^s_j)$ in the same manner. Finally, we use the CAR loss to align the relation matrix, which is denoted as:

$$L_{CAR} = \frac{1}{CHWD}\left\|(CAR(f^t) - CAR(f^s))\right\|_2^2. \tag{7}$$

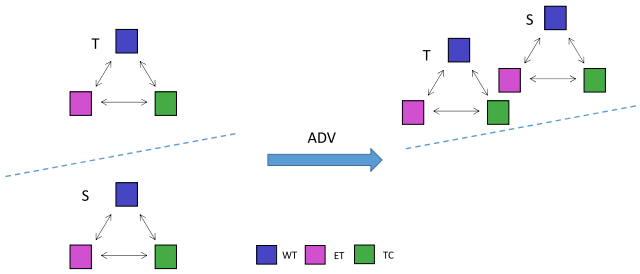


Fig. 5. We utilize ADV to ensure that the features of the student model and the teacher model can be kept in the same feature space. The dotted lines represent the division of different spaces. Different colors represent different brain tumor regions (WT, TC, and ET).

3.5. Adversarial training

Both CRR and CAR focus on restoring relational knowledge through teacher guidance. However, still lacks the ability to capture holistic feature distribution from the teacher model. As shown in Fig. 5, applying CAR and CRR allows the student model to learn relational knowledge from the teacher, but it does not guarantee that the features of both are in the same feature space, which may lead to inconsistent predictions. Adversarial learning, a paradigm rooted in game theory, has emerged as a potent tool in machine learning for capturing intricate data distributions. Motivated by the efficacy of adversarial learning in discerning and approximating complex data distributions, we leverage this technique to address the challenge of aligning holistic structural predictions. In our framework, we devise an adversarial training strategy wherein we construct a discriminator network, denoted as D, tasked with discriminating between complete and incomplete structural predictions. This discriminator serves as a critical component in guiding the learning process of our primary model, the student network denoted as S. During training iterations, the student network strives to generate structural predictions that are coherent and comprehensive, while the discriminator network endeavors to differentiate between these high-quality predictions and incomplete or erroneous outputs. This alternating training regime fosters a symbiotic relationship between the student and discriminator networks, driving continual improvement in the quality of the structural predictions generated by our model. The training of discriminator D is under the control of loss function L_{ADV}, which is formulated as:

$$L_{ADV}(T_{out}, S_{out}) = \log\left(1 - D(T_{out})\right) + \log\left(D(S_{out})\right). \tag{8}$$

We apply the L_{ADV} loss on the output layer to align the distribution of the teacher and the student in a holistic approach.

3.6. Joint objective

We trained the teacher model and student model at the same time. The teacher model is only trained with Dice loss. In addition to the conventional segmentation loss, the student model also employs CRR and CAR loss to reconstruct local and global relations. Additionally, we also adopt an adversarial learning strategy to align holistic feature distribution between the teacher and student. The overall loss function optimized during the training consists of three terms:

$$L_{joint} = L_{SEG} + L_{CRR} + L_{CAR} + L_{ADV}. \tag{9}$$

Table 1
Comparison with the state-of-art methods based on Feature Projection and Feature Synthesis. Available modalities are marked by •, and missing one is marked by ∘.

Modalities				WT							TC							ET						
Fl	T1	T1c	T2	HEMIS	HVED	RSeg	D2Net	mmFm	ShaSpec	Ours	UHEMIS	HVED	RSeg	D2Net	mmFm	ShaSpec	Ours	UHEMIS	HVED	RSeg	D2Net	mmFm	ShaSpec	Ours
•	∘	∘	∘	52.5	84.4	85.7	85.1	86.1	**88.7**	87.6	26.1	57.9	53.6	50.8	61.2	69.4	**70.1**	11.8	23.8	25.7	13.0	39.3	43.5	**43.8**
∘	•	∘	∘	57.6	49.5	70.1	19.4	67.5	73.4	**78.3**	37.4	33.9	47.9	20.3	56.6	63.2	**71.3**	10.2	8.6	17.3	15.3	32.5	41.0	**45.6**
∘	∘	•	∘	61.5	53.6	73.3	43.8	72.2	73.8	**80.7**	65.3	59.6	76.8	64.3	75.4	78.7	**86.0**	62.0	57.6	67.1	69.5	72.6	73.3	**77.9**
∘	∘	∘	•	81.0	79.8	82.2	76.6	81.2	84.0	**86.6**	57.2	54.7	57.5	59.9	64.2	69.0	**73.1**	25.6	22.8	29.0	22.8	43.1	**46.3**	45.6
•	•	∘	∘	64.6	85.7	88.2	86.2	87.1	**89.8**	87.8	41.1	61.1	60.7	62.5	65.9	**72.7**	68.7	10.7	28.0	32.1	14.1	43.0	**44.8**	43.5
•	∘	•	∘	69.0	85.9	88.5	87.9	87.3	**90.0**	88.5	71.5	75.1	80.6	82.5	77.9	84.5	**85.5**	66.1	68.4	70.3	69.3	75.1	75.6	**76.3**
•	∘	∘	•	83.0	87.6	88.3	86.4	87.6	**90.0**	88.0	57.7	62.7	61.2	61.7	69.8	72.9	**73.6**	30.2	32.3	33.8	17.8	47.5	47.2	**48.8**
∘	•	•	∘	68.5	64.2	77.2	65.4	74.4	78.7	**80.0**	72.5	67.6	78.7	78.0	78.6	82.1	**85.1**	66.2	61.1	69.1	71.4	74.0	75.8	**76.0**
∘	•	∘	•	82.4	81.6	84.8	82.6	82.2	86.0	**86.4**	60.9	56.3	62.2	64.2	69.4	71.4	**73.1**	32.4	24.3	32.0	19.3	45.0	46.8	**47.0**
∘	∘	•	•	82.5	81.3	85.2	84.6	83.0	85.4	**86.6**	76.6	73.9	80.2	81.1	78.6	83.8	**85.2**	67.8	67.8	69.7	70.5	74.5	**76.0**	75.1
•	•	•	∘	72.3	86.7	88.7	87.3	87.3	**90.3**	89.0	76.0	77.1	81.1	79.4	79.8	**85.2**	84.9	68.5	68.6	70.8	68.3	75.5	**76.4**	75.4
•	•	∘	•	83.4	88.1	88.8	88.6	87.8	**90.4**	88.0	60.3	63.1	64.4	67.5	71.5	**74.0**	72.4	31.1	32.3	36.4	23.7	47.7	46.6	**48.9**
•	∘	•	•	83.9	88.1	89.3	89.1	88.1	**90.8**	88.6	77.5	76.8	80.7	81.5	79.6	85.3	**85.4**	68.7	68.9	70.9	67.1	75.7	**76.0**	75.3
∘	•	•	•	83.9	82.3	86.0	83.8	82.7	**86.5**	85.7	79.0	75.3	80.3	80.2	80.4	84.2	**85.3**	69.9	67.8	70.1	68.6	74.8	**76.4**	76.0
•	•	•	•	84.7	88.5	89.5	88.9	89.6	**90.9**	89.6	79.5	77.7	80.9	80.9	85.8	85.5	**86.4**	70.2	69.0	71.1	67.7	77.6	**78.0**	76.8
Mean				74.0	79.2	84.4	77.0	82.9	85.9	**86.1**	62.6	64.8	69.8	67.6	73.0	77.5	**79.1**	46.1	46.8	51.0	45.2	59.9	61.6	**62.1**

4. Experimental results

In this section, to validate the superiority of our RIRN framework, we conduct experiments on brain tumor segmentation in incomplete modality conditions. With compare to the state-of-the-art methods, our RIRN framework achieves superior segmentation results in most incomplete modality cases. Furthermore, we also build ablation experiments to validate the effectiveness of the component of our RIRN framework, e.g., CRR, CAR, and ADV. We will elaborate on the details in the subsequent content

4.1. Implementation details

RIRN adopts 3D-Unet (Çiçek et al., 2016) with four-stage encoders and decoders, which is the same as in Wang et al. (2021). Following Wang et al. (2021), each encoder progressively halves the image dimensions and doubles the feature channels across four stages. All convolutions use a kernel size of $3 \times 3 \times 3$ and the initial channel number is 16. We also apply spatial dropout with a rate of 0.2 after the initial encoder convolution. In addition, the discriminator is a two-layer fully-convolutional network.

Our framework is developed in PyTorch (Paszke et al., 2019) and runs on a single RTX 3090 GPU. We train the model in an end-to-end manner, using Adam (Kingma & Ba, 2014) optimization with an initial learning rate 1e−4 and progressively decreases according to a poly policy $\left(1 - epoch/epoch_{max}\right)^{0.9}$. We trained the model with the batch size 1 and the $epoch_{max}$ is set as 300. For a fair comparison, we followed Wang et al. (2021) to utilize the same random seed and randomly split the dataset into training and validation set by a ratio of 2:1. For the image pre-processing, we normalize all input images to have zero mean and unit std. We also apply a random axis mirror flip (for all 3 axes) with a probability of 0.5. During training, we perform the random crop to generate patches of size $160 \times 192 \times 128$ as input to the network. For testing, we use center crop to generate the input data.

4.2. Dataset and evaluation metrics

Dataset: We conducted an evaluation of our model using two publicly available datasets, namely BraTS 2018 and BraTS 2020, sourced from the Multi-modal Brain Tumor Segmentation Challenge (Bakas et al., 2017, 2018; Menze et al., 2014). Both BraTS 2018 and BraTS 2020 datasets comprise multi-contrast MRI scans encompassing the four modalities mentioned earlier (T1, T1ce, T2, FLAIR). Specifically, BraTS 2018 encompasses 285 training samples, whereas BraTS 2020 comprises 369 training samples. The ground truth segmentation encompasses one class for normal tissue and four classes for tumors, including necrosis, edema, non-enhancing tumor, and enhancing core. For evaluation purposes, the brain tumor regions are categorized into three subregions: the whole tumor (WT), the tumor core (TC), and the enhancing tumor (ET).

Evaluation Metric: For the evaluation, we use the Dice coefficient to measure the segmentation performance of the proposed method. A higher Dice means that predictions are more similar to the ground truth, and indicate a better segmentation performance, defined as:

$$Dice_k\left(\hat{y}, y\right) = \frac{2 \cdot \left\|\hat{y}_k \cap y_k\right\|_1}{\left\|\hat{y}_k\right\|_1 + \left\|y_k\right\|_1}, \tag{10}$$

where k denotes different tumor classes. The $\hat{y}_k$ and y_k denote prediction and ground truth of k classes, respectively. The $Dice$ denotes the Dice scores of the tumor class k.

4.3. Comparisons with the state-of-the-art

To demonstrate the effectiveness of the proposed framework, we compared our proposed RIRN with all three types of SOTA methods on the BraTS 2018 dataset, e.g., Feature projection, Feature synthesis, and knowledge transfer. First, as shown in Table 1, we compared the methods of types of Feature projection and Feature synthesis. And then, in Table 2, we have made a separate comparison with the methods of Knowledge transfer, which belong to the same class as our methods. Finally, in Table 3, we present a comparative analysis with SOTA methods on BraTS 2020, showcasing the generalization capabilities of our approach. Overall, we achieved the best results in most cases.

4.3.1. Comparisons with feature projection based methods and feature synthesis-based methods

As shown in Table 1, we compare our model with six SOTA methods, including HeMIS (Havaei et al., 2016), HVED (Dorent et al., 2019), RSeg (Chen, Dou, et al., 2019), D2Net (Yang, Guo, Chen, Woo, & Yuan, 2022), mmFm (Zhang et al., 2022) and ShaSpec (Wang et al., 2023). These approaches attempt to build unified models to solve all modal incompleteness cases. However, it is hard for them to get impressive results when more than one modality is missing. Particularly under unimodal conditions, these methods lack sufficient data to generate the missing modality data or map the features to an appropriate feature space. In comparison to them, we significantly improved all unimodal results, especially ET and TC. Benefiting from the relational knowledge transferred from the teacher model, the student model can improve the structural output based on the available modal and voxel relationships, even in the unimodal case. Specifically, compared with the second-best method, i.e., ShaSpec (Wang et al., 2023), our method improves the average Dice scores on four unimodal by 3.3%, 5.1%, and 2.2% in the WT, the TC, and the ET, respectively. Along with the increase in the number of available modal data, our method still maintains respectable results. In particular, our method achieves a significant improvement in ET and TC regions. It indicates that our method is able to achieve better segmentation results in smaller tumor regions. Overall, our model achieves better results in most cases. Compared with ShaSpec (Wang et al., 2023), our method improves the average Dice scores on all cases by 0.2%, 1.6%, and 0.5% in the WT, the TC, and the ET, respectively.

Table 2
Comparison with the state-of-art methods based on Knowledge Transfer. Available modalities are marked by •, and missing one is marked by ∘.

Modalities				WT			TC			ET		
Fl	T1	T1c	T2	ACN	SMU	Ours	ACN	SMU	Ours	ACN	SMU	Ours
∘	∘	∘	•	85.6	85.7	**86.6**	67.9	67.2	**73.1**	43.0	43.1	**45.6**
∘	∘	•	∘	80.5	80.3	**80.7**	84.2	84.1	**86.0**	78.1	**78.3**	77.9
∘	•	∘	∘	**79.3**	78.6	78.3	71.2	69.5	**71.3**	41.5	42.8	**45.6**
•	∘	∘	∘	87.3	87.5	**87.6**	67.7	**71.8**	70.1	42.8	**46.1**	43.8
∘	∘	•	•	86.4	86.1	**86.6**	84.4	85.0	**85.2**	**75.7**	**75.7**	75.1
∘	•	•	∘	80.1	**80.3**	80.0	84.6	84.4	**85.1**	75.2	75.1	**75.9**
•	•	∘	∘	87.5	87.3	**87.8**	**71.3**	71.2	68.7	43.7	**44.0**	43.5
∘	•	∘	•	85.5	85.6	**86.4**	73.3	**73.5**	73.1	47.4	**47.7**	47.0
•	∘	∘	•	87.8	87.9	**88.0**	71.6	71.2	**73.6**	46.0	46.0	**48.8**
•	∘	•	∘	88.3	88.4	**88.5**	83.4	84.1	**85.5**	**77.5**	77.3	76.3
•	•	•	∘	89.0	88.2	**89.0**	84.3	84.2	**84.9**	76.2	**76.2**	75.4
•	•	∘	•	**88.4**	88.3	88.0	67.9	67.9	**72.4**	42.1	43.1	**48.8**
•	∘	•	•	88.3	88.2	**88.6**	82.9	82.5	**85.2**	**76.0**	75.4	75.3
∘	•	•	•	**86.9**	86.5	85.7	84.7	84.4	**85.3**	76.1	**76.2**	76.0
•	•	•	•	89.2	88.9	**89.6**	85.2	**87.3**	86.4	77.1	**79.3**	76.8
Mean				85.9	85.9	**86.1**	77.6	77.9	**79.1**	61.2	61.8	**62.1**

Table 3
Comparison with the state-of-art methods on BraTS 2020. Available modalities are marked by •, and missing one is marked by ∘.

Modalities				WT							TC							ET						
Fl	T1	T1c	T2	HEMIS	RSeg	mmFm	RFNet	M3AE	TMFm	Ours	HEMIS	RSeg	mmFm	RFNet	M3AE	TMFm	Ours	HEMIS	RSeg	mmFm	RFNet	M3AE	TMFm	Ours
•	∘	∘	∘	52.29	82.87	86.27	87.32	86.53	87.45	**88.28**	24.97	60.72	67.80	69.19	68.04	70.19	**71.54**	9.00	34.68	38.39	38.15	40.49	42.28	**45.24**
∘	•	∘	∘	63.01	71.41	74.91	77.16	76.71	78.94	**80.27**	42.42	54.30	64.24	66.02	66.00	67.12	**72.14**	16.53	28.99	34.78	37.30	39.93	38.21	**46.54**
∘	∘	•	∘	64.58	71.39	74.86	76.77	73.85	78.53	**80.03**	69.41	76.68	80.74	81.51	81.39	**82.59**	82.25	63.24	67.91	76.45	74.85	72.43	76.21	**78.03**
∘	∘	∘	•	79.85	82.20	85.37	86.05	86.09	86.46	**86.61**	54.22	61.88	70.21	71.02	70.27	71.84	**74.69**	31.43	36.46	46.12	46.29	45.97	46.94	**49.49**
•	•	∘	∘	65.29	88.10	88.91	89.73	89.38	89.64	**89.76**	41.58	68.18	71.83	73.07	73.82	**74.65**	73.26	13.99	39.67	41.17	40.98	43.20	**48.20**	47.01
•	∘	•	∘	69.37	87.33	89.03	89.89	89.48	89.67	**90.81**	70.86	81.85	84.44	84.65	82.01	84.62	**87.92**	68.31	70.78	77.22	76.67	74.66	76.37	**79.53**
•	∘	∘	•	81.56	88.09	89.04	89.87	89.25	**89.98**	89.23	55.89	68.20	72.82	74.14	74.95	74.76	**75.57**	28.91	42.19	47.98	49.32	47.30	**51.67**	50.94
∘	•	•	∘	72.50	76.84	79.61	81.12	78.11	81.97	**83.00**	75.59	80.28	82.00	83.40	82.39	83.13	**87.81**	70.71	70.11	77.12	78.01	75.42	**78.68**	78.13
∘	•	∘	•	82.31	85.53	87.19	87.73	87.20	87.75	**88.22**	56.38	66.46	72.61	73.13	72.54	73.33	**74.30**	28.58	39.92	48.07	45.65	46.63	48.81	**49.76**
∘	∘	•	•	84.45	85.97	87.35	87.74	87.37	88.01	**88.82**	77.60	82.44	84.30	83.45	83.01	84.24	**87.81**	70.30	71.42	75.29	75.93	76.81	78.25	**79.16**
•	•	∘	•	73.31	88.87	89.61	90.69	89.99	90.23	**91.15**	75.07	82.76	84.59	85.07	82.44	84.69	**88.05**	70.80	71.77	77.13	76.81	75.94	78.64	**79.23**
•	•	•	∘	83.03	89.24	89.96	**90.60**	90.42	90.51	90.48	57.40	70.46	73.90	**75.19**	75.09	75.17	74.29	29.53	43.90	48.77	49.92	48.19	**51.23**	50.26
•	∘	•	•	84.64	88.68	89.82	90.68	90.18	90.53	**91.74**	77.69	81.89	84.63	84.97	83.06	84.64	**87.68**	71.36	71.17	76.36	77.12	77.08	78.45	**79.48**
∘	•	•	•	85.19	86.63	87.85	88.25	88.61	88.54	**89.25**	79.05	82.85	84.28	83.47	84.06	84.00	**88.33**	71.67	71.87	76.69	76.99	77.40	**78.51**	78.39
•	•	•	•	85.19	89.47	90.31	91.11	90.56	90.83	**91.30**	78.58	82.87	84.49	85.21	84.40	84.88	**87.90**	71.49	71.52	78.11	78.00	78.00	78.98	**79.75**
Mean				75.10	84.17	86.01	86.98	86.25	87.27	**87.93**	62.45	73.45	77.53	78.23	77.56	78.66	**80.90**	47.72	55.49	61.31	61.47	61.30	63.43	**64.73**

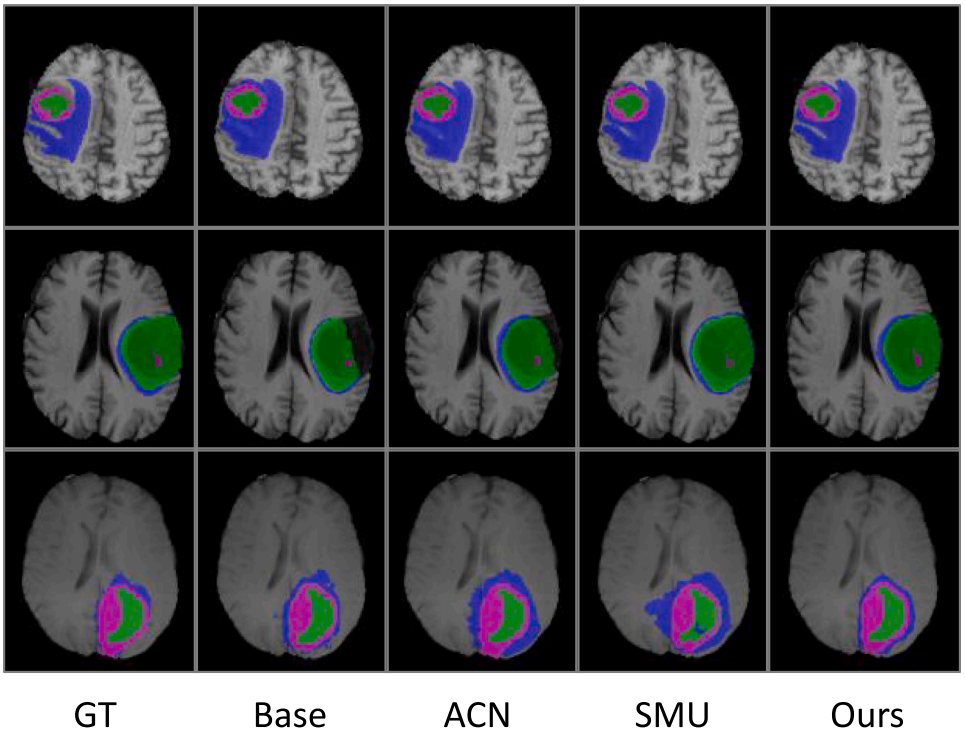


Fig. 6. Visual comparison of brain tumor segmentation produced by the different frameworks. The whole tumor, tumor core, and enhancing core are indicated in blue, green, and purple colors respectively.

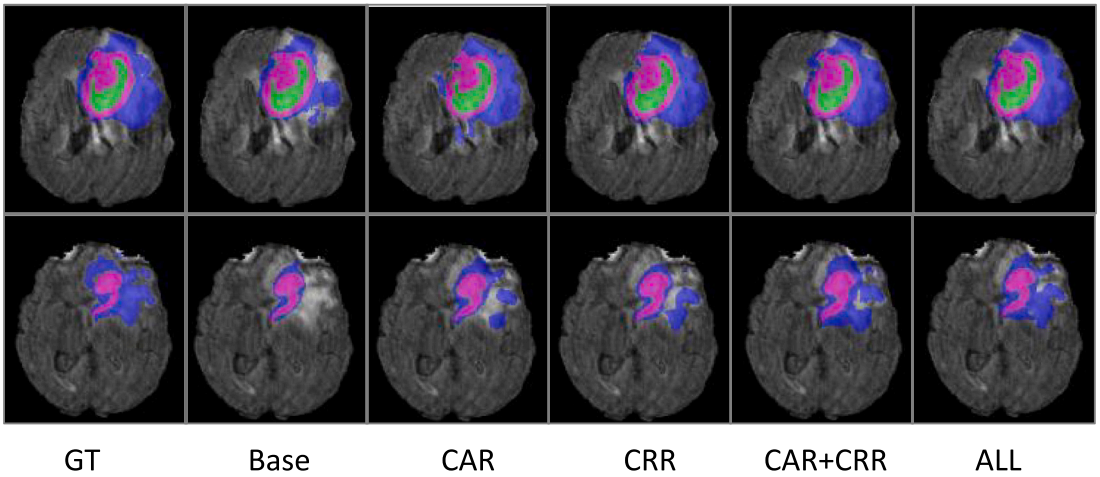


Fig. 7. Visual comparison of brain tumor segmentation produced by a different component of our methods. The whole tumor, tumor core, and enhancing core are indicated in blue, green, and purple colors respectively.

Table 4
Comparison of different methods for the test case of T1ce modality on brain tumor segmentation. (DSC%). MF denotes the manual filtering.

	CRR	MF	CAR	ADV	WT	TC	ET	AVG
T1ce					78.83	84.92	74.80	79.52
	✓				79.55	85.33	74.36	79.75
	✓	✓			79.98	85.71	75.84	80.51
			✓		80.07	85.39	76.11	80.53
	✓	✓	✓		80.38	86.03	75.98	81.11
	✓	✓	✓	✓	80.69	86.04	77.93	81.55

4.3.2. Comparisons with knowledge transfer-based methods

To demonstrate the superiority of our method, we also compare it with two knowledge transfer SOTA methods: ACN (Wang et al., 2021) and SMU (Azad et al., 2022), which belong to the same category as our method. Compare to other knowledge transfer methods, rather than just aligning the outputs of the teacher and student models at different layers, we pay more attention to transferring relational knowledge from the teacher model to the student model. Given that brain tumor segmentation is a structural prediction task, we contend that reconstructing voxel relationships leads to better segmentation results than aligning pixels individually in an incomplete modality situation. As shown in Table 2, we achieve superior segmentation performance in most cases. Specifically, compare with SMU, the second-best method, our method improves the average Dice scores on all cases by 0.2%, 1.2%, and 0.3% in the WT, the TC, and the ET, respectively. It is clear that our method achieves greater improvement in smaller tumor regions, especially in the TC. Overall, our method has significant improvement over other methods, which demonstrates the superiority of our method.

4.3.3. Comparison result on BraTS 2020 dataset

To evaluate the model's generalization performance across diverse datasets, we conducted supplementary comparative experiments utilizing the BraTS 2020 dataset. We compare with six SOTA approaches, including HeMIS (Havaei et al., 2016), RSeg (Chen, Dou, et al., 2019), mmFm (Zhang et al., 2022), RFNet (Ding, Yu, & Yang, 2021), M3AE (Liu, Wei, et al., 2023) and TMFm (Zhang et al., 2024). As illustrated in Table 3, our approach consistently demonstrates superior segmentation performance across most cases. As the number of available modality increases, the performance of the model generally improves. However, our model consistently maintains leading performance. For example, when the number of available modalities is three, our method outperforms the second-best model, TMFm, with enhancements in average Dice scores across all cases by 0.7%, 2.46%, and 0.13% in the WT, TC and ET regions, respectively. Additionally, compared with TMFm, our method exhibits enhancements in average Dice scores across all cases by 0.66%, 2.24%, and 1.3% in the WT, the TC, and the ET, respectively. Collectively, these findings underscore the substantial advancement achieved by our method over alternative approaches, thus affirming its superiority.

Table 5
Comparison of parameters and FLOPs between different methods based on knowledge transfer.

	ACN	SMU	Ours
Prams (M)	9.62	4.77	4.73
FLOPs (G)	381.83	273.13	272.75

4.3.4. Qualitative results

The visual comparison results of different methods are presented in Fig. 6. All the experiments are performed with only the T1ce modal available. Compare to other methods, we can see that RIRN produces segmentation masks with more regular boundaries and compact shapes.

4.3.5. Comparison of network parameters and computational efficiency

We compared the number of learnable parameters and floating-point operations per second (FLOPs) across two knowledge transfer-based methods to highlight the efficiency of our proposed RIRN. As shown in Table 5, RIRN achieved lower parameters and FLOPs compared to other methods. Specifically, compared to ACN, RIRN significantly reduces the number of learnable parameters from 9.62 million to 4.73 million. Additionally, RIRN maintains a lower FLOPs count of 272.75 G, indicating its computational efficiency.

4.4. Ablation result

In this section, we conduct the ablation study to verify the effectiveness of our methods. First, we conduct experiments using only the T1ce modal to demonstrate the effectiveness of each component of our method. Then, we show the effectiveness of our method in other unimodal cases. Finally, we visualize the segmentation results of our framework.

4.4.1. Ablation result on T1ce only

First, we train the baseline model along with the T1ce modal only. And then we add CRR, MF, CAR, and ADV gradually. The MF stands for manual filtering in the CRR module. It is observed in Table 4 that both of them improve the accuracy of brain tumor segmentation over three regions, which proves their effectiveness. For instance, using CRR without MF enhances the average Dice scores of the three tumor regions by 0.23% in terms of average DSC. Additionally, incorporating MF further improves the average Dice scores of the three tumor regions by 0.76% in average DSC. And employing CAR also increases the average Dice scores by 1.01% on average DSC. It demonstrates that reconstructing the voxel relationship, either CRR or CAR, can assist the model in developing a robust structural prediction under an incomplete modality situation. In addition, combining CAR and CRR jointly can improve the average Dice scores of three tumor regions by 1.53%. It shows that CAR and CRR can work in a complementary form and further improve the segmentation performance. Furthermore, adding adversarial learning gives rise to further improvement by 0.44%. In general, our method improves by 1.86%, 1.12%, and 3.13% in WT, TC, and ET, respectively. And improve the average Dice scores of three tumor regions by 2.03% in all.

4.4.2. Ablation result on other unimodal cases

As shown in Table 6, we performed the ablation experiment on other unimodal cases to demonstrate the advantages of our method. Regardless of which unimodal situation, our approach has made considerable progress. For example, our method improves by 3.16%, 2.46%, and 1.16% on average DSC on T1, T2, and Flair modality, respectively. The model can get improved structural prediction based on the available modality data and relational knowledge. It demonstrates the generalizability and superiority of our method.

4.4.3. Qualitative results

The visual comparison results of different components of our methods are presented in Fig. 7. It is obvious to see that our method can achieve better results compared to the benchmark method. Each component of our framework can enhance the segmentation performance very well. Specifically, as can be seen from the second row of Fig. 7, the CRR achieved more accurate segmentation results at the edge of the tumor compared to CAR. It confirmed that CRR can help the model to achieve more accurate segment result in the boundary of the tumor. The fifth column shows that CAR and CRR can work well in a complementary form to obtain better segmentation results. The last column indicates that the best result of our method is achieved by using all components.

4.4.4. Comparison with different aggregation methods

In the CRR module, we employ an average operation to generate the class feature representation, denoted as f^{T}_{region}. To validate its efficiency, we conducted experiments comparing different aggregation methods, including maximum, minimum, and their average. As presented in Table 7, we observe that the average operation consistently achieves the best performance. We hypothesize that this superiority arises because the average operation more accurately captures class-specific features. In contrast, the maximum, minimum, or their average can be influenced by outliers and fail to adequately represent the underlying class characteristics.

Table 6
Comparing the performance of the method on different modalities (DSC%).

Modality	Method	WT	TC	ET	AVG
T1	Baseline	77.10	67.49	41.06	61.89
	Ours	78.29	71.26	45.62	65.05
T2	Baseline	85.68	69.33	42.93	65.98
	Ours	86.58	73.12	45.60	68.44
Flair	Baseline	87.86	67.75	42.33	65.99
	Ours	87.57	70.10	43.79	67.15

Table 7
Comparison with different aggregation methods.

Methods	WT	TC	ET	AVG
Max	79.85	85.49	77.69	81.01
Min	79.76	86.06	76.34	80.72
(Max+Min)/2	79.88	85.57	77.49	80.98
Ours(Avg)	80.69	86.04	77.93	81.55

5. Conclusion

In this study, we suggest a brand-new methodology for brain tumor segmentation under an incomplete modality situation by transferring the relational knowledge from the teacher to the student. We note that different modalities exhibit distinct appearances of brain tumor regions. The absence of specific modality data will lead to a lack of tumor-specific information. The model cannot get sufficient information to build a reliable voxel relation. The poor voxel relation can result in inaccurate structural predictions in segmentation tasks, which lowers segmentation performance. Therefore, we propose the CAR and CRR to reconstruct relations between different voxels in the teacher–student model. Moreover, we add an adversarial training strategy to align the holistic structural prediction and further improve the segmentation performance. Extensive experiments demonstrate that our method significantly outperforms the state-of-the-art. In this study, we focused exclusively on a small brain tumor dataset, which may limit the performance of the model in practical applications. For example, data imbalance significantly impacted the model's performance. In future work, we plan to incorporate federated learning to leverage data from different centers, thereby improving the model's generalization ability.

CRediT authorship contribution statement

Jiawei Su: Writing – original draft, Methodology, Conceptualization. **Zhiming Luo:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Chengji Wang:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Sheng Lian:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Xuejuan Lin:** Supervision, Resources, Project administration, Funding acquisition. **Shaozi Li:** Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62276221, 62376232, 81973752); the Natural Science Foundation of Fujian Province of China (No. 2022J01002, 2023J05117); Education and Scientific Research Project for Middle-aged and Young Teachers in Fujian Province (No. JAT220017).

References

Azad, R., Khosravi, N., & Merhof, D. (2022). Smu-net: Style matching u-net for brain tumor segmentation with missing modalities. arXiv preprint arXiv:2204.02961.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., et al. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, *4*, 1–13.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629.

Ben-Cohen, A., Klang, E., Raskin, S. P., Soffer, S., Ben-Haim, S., Konen, E., et al. (2019). Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection. *Engineering Applications of Artificial Intelligence*, *78*, 186–194.

Çiçek, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424–432). Springer.

Chen, G., Choi, W., Yu, X., Han, T., & Chandraker, M. (2017). Learning efficient object detection models with knowledge distillation. *Advances in Neural Information Processing Systems*, *30*.

Chen, C., Dou, Q., Jin, Y., Chen, H., Qin, J., & Heng, P. A. (2019). Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In *International conference on medical image computing and computer-assisted intervention* (pp. 447–456). Springer.

Chen, C., Dou, Q., Jin, Y., Liu, Q., & Heng, P. A. (2021). Learning with privileged multimodal knowledge for unimodal segmentation. *IEEE Transactions on Medical Imaging*, *41*, 621–632.

Chen, C., Liu, X., Ding, M., Zheng, J., & Li, J. (2019). 3D dilated multi-fiber network for real-time brain tumor segmentation in mri. In *International conference on medical image computing and computer-assisted intervention* (pp. 184–192). Springer.

Ding, Y., Yu, X., & Yang, Y. (2021). Rfnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3975–3984).

Ding, Y., Zheng, W., Geng, J., Qin, Z., Choo, K. K. R., Qin, Z., et al. (2021). Mvfusfra: a multi-view dynamic fusion framework for multimodal brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, *26*, 1570–1581.

Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., & Ayed, I. B. (2018). Hyperdense-net: a hyper-densely connected cnn for multi-modal image segmentation. *IEEE Transactions on Medical Imaging*, *38*, 1116–1126.

Dorent, R., Joutard, S., Modat, M., Ourselin, S., & Vercauteren, T. (2019). Hetero-modal variational encoder–decoder for joint modality completion and segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 74–82). Springer.

Dou, Q., Liu, Q., Heng, P. A., & Glocker, B. (2020). Unpaired multi-modal segmentation via knowledge distillation. *IEEE Transactions on Medical Imaging*, *39*, 2415–2425.

Fidon, L., Li, W., Garcia-Peraza-Herrera, L. C., Ekanayake, J., Kitchen, N., Ourselin, S., et al. (2017). Scalable multimodal convolutional networks for brain tumour segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 285–293). Springer.

Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge distillation: A survey. *International Journal of Computer Vision*, *129*, 1789–1819.

Gridach, M. (2021). Pydinet: Pyramid dilated network for medical image segmentation. *Neural Networks*, *140*, 274–281.

Havaei, M., Guizard, N., Chapados, N., & Bengio, Y. (2016). Hemis: Hetero-modal image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 469–477). Springer.

Hinton, G., Vinyals, O., Dean, J., et al. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Hu, M., Maillard, M., Zhang, Y., Ciceri, T., Barbera, G. La., Bloch, I., et al. (2020). *Knowledge distillation from multi-modal to mono-modal segmentation networks*, *International conference on medical image computing and computer-assisted intervention* (pp. 772–781). Springer.

Huang, P., Li, D., Jiao, Z., Wei, D., Li, G., Wang, Q., et al. (2019). Coca-gan: common-feature-learning-based context-aware generative adversarial network for glioma grading. In *International conference on medical image computing and computer-assisted intervention* (pp. 155–163). Springer.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *18*, 203–211.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, *36*, 61–78.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Liu, Y., Chen, K., Liu, C., Qin, Z., Luo, Z., & Wang, J. (2019). Structured knowledge distillation for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2604–2613).

Liu, B., Desrosiers, C., Ayed, I. B., & Dolz, J. (2023). Segmentation with mixed supervision: Confidence maximization helps knowledge distillation. *Medical Image Analysis*, *83*, Article 102670.

Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., & Zheng, Y. (2023). M3ae: Multimodal representation learning for brain tumor segmentation with missing modalities. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 1657–1665).

Malik, N., & Bzdok, D. (2022). From youtube to the bra Transfer learning can improve brain-imaging predictions with deep learning. *Neural Networks*, *153*, 325–338.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, *34*, 1993–2024.

Orbes-Arteaga, M., Cardoso, M. J., Sørensen, M., Ourselin, S., Nielsen, M., & Pai, A. (2018). Simultaneous synthesis of flair and segmentation of white matter hypointensities from t1 mris. arXiv preprint arXiv:1808.06519.

Pang, S., Du, A., Orgun, M. A., Wang, Y., & Yu, Z. (2021). Tumor attention networks: Better feature selection, better tumor segmentation. *Neural Networks*, *140*, 203–222.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*.

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging*, *35*, 1240–1251.

Qiao, S., Shen, W., Zhang, Z., Wang, B., & Yuille, A. (2018). Deep co-training for semi-supervised image recognition. In *Proceedings of the European conference on computer vision* (pp. 135–152).

Tulder, G. v., & Bruijne, M. d. (2015). Why does synthesized data improve multi-sequence classification?. In *International conference on medical image computing and computer-assisted intervention* (pp. 531–538). Springer.

van Tulder, G., & de Bruijne, M. (2018). Learning cross-modality representations from multi-modal images. *IEEE Transactions on Medical Imaging*, *38*, 638–648.

Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., & Carneiro, G. (2023). 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).

Wang, P., & Chung, A. C. (2022). Relax and focus on brain tumor segmentation. *Medical Image Analysis*, *75*, Article 102259.

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7794–7803).

Wang, T., Yuan, L., Zhang, X., & Feng, J. (2019). Distilling object detectors with fine-grained feature imitation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4933–4942).

Wang, Y., Zhang, Y., Hou, F., Liu, Y., Tian, J., Zhong, C., et al. (2020). Modality-pairing learning for brain tumor segmentation. In *International MICCAI brainlesion workshop* (pp. 230–240). Springer.

Wang, Y., Zhang, Y., Liu, Y., Lin, Z., Tian, J., Zhong, C., et al. (2021). Acn: Adversarial co-training network for brain tumor segmentation with missing modalities. In *International conference on medical image computing and computer-assisted intervention* (pp. 410–420). Springer.

Wu, W., Hu, D., Niu, C., Broeke, L. V., Butler, A. P., Cao, P., et al. (2021). Deep learning based spectral ct imaging. *Neural Networks*, *144*, 342–358.

Xie, L., Huang, J., Yu, J., Zeng, Q., Hu, Q., Chen, Z., et al. (2023). Cntseg: A multimodal deep-learning-based network for cranial nerves tract segmentation. *Medical Image Analysis*, *86*, Article 102766.

Xu, Z., Hsu, Y. C., & Huang, J. (2017). Training shallow and thin networks for acceleration via knowledge distillation with conditional adversarial networks. arXiv preprint arXiv:1709.00513.

Xu, K., Rui, L., Li, Y., & Gu, L. (2020). Feature normalized knowledge distillation for image classification. In *European conference on computer vision* (pp. 664–680). Springer.

Yang, Q., Guo, X., Chen, Z., Woo, P. Y., & Yuan, Y. (2022). D 2-net: Dual disentanglement network for brain tumor segmentation with missing modalities. *IEEE Transactions on Medical Imaging*, *41*, 2953–2964.

Yu, B., Zhou, L., Wang, L., Fripp, J., & Bourgeat, P. (2018). 3D cgan based cross-modality mr image synthesis for brain tumor segmentation. In *2018 IEEE 15th international symposium on biomedical imaging* (pp. 626–630). IEEE.

Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., et al. (2022). mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 107–117). Springer.

Zhang, D., Huang, G., Zhang, Q., Han, J., Han, J., Wang, Y., et al. (2020). Exploring task structure for brain tumor segmentation from multi-modality mr images. *IEEE Transactions on Image Processing*, *29*, 9032–9043.

Zhang, Z., Yang, G., Zhang, Y., Yue, H., Liu, A., Ou, Y., et al. (2024). Tmformer: Token merging transformer for brain tumor segmentation with missing modalities. *Proceedings of the AAAI Conference on Artificial Intelligence*, *741*, 4–7422.

Zhao, B., Cui, Q., Song, R., Qiu, Y., & Liang, J. (2022). Decoupled knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11953–11962).

Zhou, C., Ding, C., Lu, Z., Wang, X., & Tao, D. (2018). One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 637–645). Springer.

Zhuang, Y., Liu, H., Song, E., & Hung, C. C. (2022). A 3d cross-modality feature interaction network with volumetric feature alignment for brain tumor and tissue segmentation. *IEEE Journal of Biomedical and Health Informatics*.