Semi-Supervised Multi-Modal Medical Image Segmentation for Complex Situations

Dongdong Meng[1], Sheng Li[2,*], Hao Wu[3], Guoping Wang[2], and Xueqing Yan[1,*]

[1] School of Physics, Peking University, Beijing, China
[2] School of Computer Science, Peking University, Beijing, China
[3] Department of Radiotherapy, Peking University Cancer Hospital, Beijing, China
* Corresponding author: {lisheng, x.yan}@pku.edu.cn

Abstract. Semi-supervised learning addresses the issue of limited annotations in medical images effectively, but its performance is often inadequate for complex backgrounds and challenging tasks. Multi-modal fusion methods can significantly improve the accuracy of medical image segmentation by providing complementary information. However, they face challenges in achieving significant improvements under semi-supervised conditions due to the challenge of effectively leveraging unlabeled data. There is a significant need to create an effective and reliable multi-modal learning strategy for leveraging unlabeled data in semi-supervised segmentation. To address these issues, we propose a novel semi-supervised multi-modal medical image segmentation approach, which leverages complementary multi-modal information to enhance performance with limited labeled data. Our approach employs a multi-stage multi-modal fusion and enhancement strategy to fully utilize complementary multi-modal information, while reducing feature discrepancies and enhancing feature sharing and alignment. Furthermore, we effectively introduce contrastive mutual learning to constrain prediction consistency across modalities, thereby facilitating the robustness of segmentation results in semi-supervised tasks. Experimental results on two multi-modal datasets demonstrate the superior performance and robustness of the proposed framework, establishing its valuable potential for solving medical image segmentation tasks in complex scenarios. The code is available at: https://github.com/DongdongMeng/SMMS.

Keywords: Semi-supervised learning · Multi-modal segmentation · Contrastive learning.

1 Introduction

Fully supervised segmentation methods play an essential part in the field of medical image analysis. However, their progress is impeded due to the limited availability of large, high-quality labeled training datasets. This challenge makes semi-supervised segmentation a cost-effective alternative for training robust models with limited carefully labeled data and extensive unlabeled data

[10]. Many semi-supervised techniques have been effectively applied to medical image segmentation tasks [18]. These methods employ either self-training or co-training strategies to enhance pseudo-labels [2,13], thereby expanding the labeled dataset, or incorporate consistency-based mutual training to ensure consistency across data [3], models [15,19], or tasks [8]. However, due to individual differences among patients and limitations in image quality, most existing methods find it difficult to segment complex targets, particularly when dealing with irregular lesion shapes, complex adjacent tissue structures, and edge blurring, resulting in limited segmentation performance.

A common strategy for improving segmentation accuracy involves utilizing multi-modal learning methods [5]. These methods can effectively leverage complementary information from multiple modalities, thereby reducing prediction uncertainty and enhancing the accuracy of clinical diagnosis and analysis [22]. However, most existing multi-modal approaches are designed for fully supervised tasks, failing to fully leverage the advantages of multi-modal data in semi-supervised segmentation, resulting in a significant performance gap compared to fully supervised methods.

Recently, a few semi-supervised multi-modal segmentation methods have successfully demonstrated that multi-modal data can effectively mitigate the accuracy degradation resulting from limited labeled data. Zhang et al. [20] proposed to apply multi-modal information in semi-supervised contrastive mutual learning. However, this approach ignores the feature discrepancies between modalities at earlier stages and focuses solely on the explicit consistency constraints among multi-modal prediction results, which may exacerbate these differences and even lead to consistent yet incorrect predictions [9]. To address this challenge, Chen et al. [1] and Zhou et al. [23] developed a cross-modal collaboration strategy for feature fusion and alignment, thereby enabling more effective feature sharing across various modalities. However, using separate independent encoders to extract modality-specific features may still enlarge the discrepancies between features, leading to noticeable differences in segmentation accuracy across modalities.

To effectively reduce the feature discrepancies extracted by different encoders, it is beneficial to perform feature fusion and alignment during the early encoding stage. This can be effectively accomplished through multi-stage feature fusion, where high-resolution low-level features are shared [6]. Furthermore, incorporating a modality-aware enhancement strategy enables dynamic adjustment of the contributions from various modalities, thereby guaranteeing the efficient utilization of multi-modal data [21,11]. Moreover, the consistency constraints of predictions also promote multi-modal mutual learning processes [20]. However, these studies fail to achieve effective multi-modal fusion and mutual learning supervision, which are key elements essential for enhancing the accuracy of semi-supervised segmentation.

In this paper, we propose a novel semi-supervised multi-modal medical image segmentation approach. Our method achieves high accuracy for complex segmentation targets with limited labeled data. To reduce the disparity between modalities, we introduce a multi-stage feature fusion strategy to adequately align and fuse low-level visual features. Additionally, we introduce a modality-aware feature enhancement module to emphasize important modality-specific features while ignoring irrelevant information. Furthermore, we design a collaborative mutual learning objective to facilitate mutual learning across different modalities, ensuring the consistency and robustness of the cross-modal segmentation results. We conducted a series of experiments on two complex tumor segmentation datasets, and the results show that the proposed method achieves remarkable performance compared with the state-of-the-art segmentation methods in semi-supervised tasks.

2 Methdology

The proposed framework for semi-supervised multi-modal medical image segmentation is depicted in Fig.1. First, to prevent accuracy degradation caused by limited labeled data, we adopt a multi-modal learning strategy to incorporate finer details and enhance segmentation performance. The input multi-modal data is initially fed into a dual-branch segmentation network for feature extraction, followed by multi-stage feature fusion to achieve feature alignment and minimize discrepancies. Subsequently, the features pass through the modality-aware enhancement module to adaptively choose effective multi-modal information and generate enhanced feature representations. Finally, the model is trained by a contrastive mutual learning strategy composed of supervised and unsupervised consistency losses.

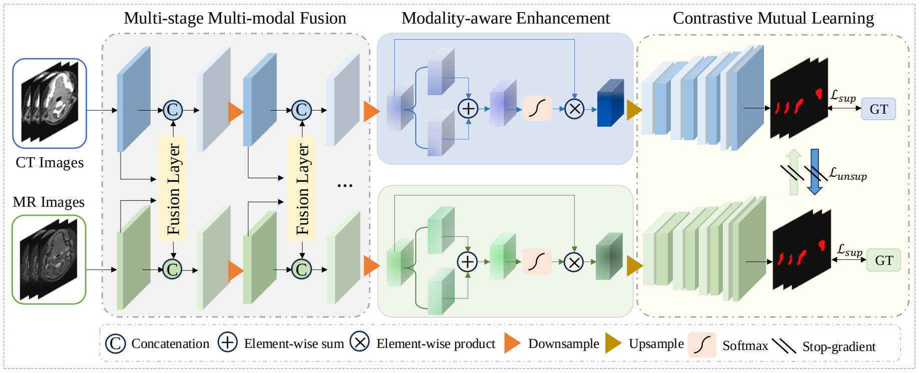

Fig. 1. Overview of our semi-supervised multi-modal segmentation method.

2.1 Multi-stage Multi-modal Feature Fusion and Enhancement

We denoted the multi-modal dataset D^a and D^b following [20,23] as:

$$\begin{aligned} D_l^a &= \{(x_i^a, y_i)\}_{i=1}^M, D_u^a = \{(x_i^a)\}_{j=1}^N, \\ D_l^b &= \{(x_i^b, y_i)\}_{i=1}^M, D_u^b = \{(x_i^b)\}_{j=1}^N, \end{aligned} \tag{1}$$

where D_l and D_u denoted the labeled and unlabeled datasets. The x^a, x^b $\in \mathbb{R}^{H \times W \times D}$ are input image modalities with the size of $H \times W \times D$, and $y_i \in \mathbb{R}^{C \times H \times W \times D}$ is the ground truth with C classes. The M and N denote the number of labeled and unlabeled samples, and $M \ll N$. Our semi-supervised segmentation framework adopts an end-to-end design, taking two different image modalities as input, with the same annotation in the training stage, and obtaining their prediction results at the output of the network simultaneously. We use $F^a = \{f_s^a\}_{s=1}^4$ and $F^b = \{f_s^b\}_{s=1}^4$ represent the low-level visual features of the encoder for each modality at different stages, respectively. To fully integrate the two modalities and reduce the differences between them, we introduce a multi-modal fusion strategy that captures and aligns multi-modal features $F^{ab} = \{f_s^{ab}\}_{s=1}^4$ at four encoding stages, and then transmits them to the next stage of each encoder:

$$f_s^{ab} = F_{sigmoid}(F_{conv2}(F_{conv1}(F_{cat}(f_s^a, f_s^b)))), \tag{2}$$

where $F_{sigmoid}$ denotes the sigmoid activation, F_{conv1} and F_{conv2} represent two 3×3×3 convolutional layers with padding of 1 and stride of 1, followed by PReLU activation and batch normalization, F_{cat} represents the concatenation operation.

After the multi-stage feature fusion process, which aligns the features of different modalities and gradually enhances semantic information. We introduce a novel modality-aware enhancement module to adaptively adjust the weights of various modalities, thereby enhancing the importance of multi-modal features. Attention mechanisms have been proven to effectively enhance feature representation and improve model robustness [21,4], which also applies to our method. We employ multiple convolutional layers with different receptive fields to model both local and global information. By learning the channel-wise dependencies, we weigh the fused features to enhance crucial feature representations and suppress redundant information. The modality-aware attention weight for one of the modalities, as illustrated in Fig.1, is indirectly optimized by updating the learnable parameters defined in the following equation:

$$W^a = F_{softmax}(F_{fc}(F_{gap}(\psi_1(F^a) + \psi_2(F^a)))), \tag{3}$$

where $F_{softmax}$ denotes the softmax function, F_{fc} represents the fully-connected layer, F_{gap} represents the global average pooling operation, and ψ_1 and ψ_2 map the input feature from $\mathbb{R}^{H \times W \times D}$ to a transformed space $\mathbb{R}^{H' \times W' \times D'}$ through a sequence of convolution, batch normalization, and ReLU activations. Similarly, the same weight learning process is applied to the other modality. Subsequently, the emphasized features F^a and F^b will pass through the decoder path to generate the final output.

2.2 Multi-modal Contrastive Mutual Learning

Contrastive learning can constrain neural networks to produce consistent segmentation results, effectively alleviating the accuracy degradation caused by limited labeled data [20]. This method has been proven effective in multiple semi-supervised segmentation tasks and is also applicable to our scenario [10]. We define the segmentation networks $f_\phi(\cdot)$ and $g_\phi(\cdot)$ for each respective modality. We first apply a supervised loss to constrain the predictions, exclusively targeting the labeled data with the ground truth:

$$\min_{f_\phi, g_\phi} L_{sup}(f_\phi, g_\phi) = \mathbb{E}_{x^a, x^b, y}[L_{CE}(f_\phi(x^a), g_\phi(x^b), y) + L_{DICE}(f_\phi(x^a), g_\phi(x^b), y)], \tag{4}$$

where L_{CE} and L_{DICE} represent cross-entropy loss and dice coefficient loss function. In addition, to further obtain high-quality segmentation results, we introduce the contrastive mutual learning loss to constrain cross-modal consistency for unlabeled data. Given the distinct attributes of features from different modalities, the prediction results produced by the f_ϕ and g_ϕ networks are different, enabling the generation of respective pseudo-labels:

$$pl^a = f_\phi(x^a), pl^b = g_\phi(x^b), \tag{5}$$

Then, the contrastive mutual learning across modalities is defined as:

$$\min_{f_\phi, g_\phi} L_{unsup}(f_\phi, g_\phi) = \mathbb{E}_{x^a, x^b}[\|f_\phi(x^a) - pl^b\|^2 + \|g_\phi(x^b) - pl^a\|^2]. \tag{6}$$

Significantly, to prevent the model from overfitting to self-generated pseudo-labels and avoid erroneous convergence, gradient back-propagation is not performed between pl^a and $f_\phi(x^a)$, nor between pl^b and $g_\phi(x^b)$. Overall, the total learning strategy for training the semi-supervised multi-modal segmentation model is summarized as $L_{total} = L_{sup}(f_\phi, g_\phi) + \alpha L_{unsup}(f_\phi, g_\phi)$, where α represents the constraint weight between modalities.

3 Experiments and Results

3.1 Experimental Details

We evaluated our method using two types of multi-modal tumor datasets: the publicly accessible BraTS 2019 Challenge dataset and a private nasopharyngeal carcinoma (NPC) dataset.

BraTS 2019 The dataset [12] contains 335 multi-modal MRI scans of brain tumor patients with four modalities: FLAIR, T1, T1ce, and T2. The MRI scans are 155 × 240 × 240, with the pixel size of 1.0 mm^3. In our study, we investigate the semi-supervised segmentation of whole tumors using T1ce and T2 images. We randomly select 250, 25, and 60 cases for training, validation, and testing, respectively. For pre-processing, we crop zero-intensity regions and apply min-max normalization to each scan.

NPC Dataset The dataset contains 161 patients who received radiotherapy treatment at a Cancer Hospital. The CT images were reconstructed using a matrix size of 512 × 512, thickness of 3.0 mm, and pixel size of 1.27 × 1.27 mm^2. The MR T2 images were reconstructed using a matrix size of 384 × 384, thickness of 3.0 mm, and pixel size of 1.30 × 1.30 mm^2. The manual segmentation of NPC was contoured by a radiation oncologist and verified by an experienced oncologist. In our study, we randomly selected 112, 17, and 32 cases for training, validation, and testing, respectively. For pre-processing, we rigidly register CT and MR T2 images, convert CT intensities to Hounsfield units (HU), and normalize CT images using window width/level. MR T2 images are normalized with the min-max method.

Implementation Details The model was implemented with the PyTorch framework on a NVIDIA A6000 GPU. All models were trained from scratch with the same experimental settings. The training used the SGD optimizer with an initial learning rate of 1×10^{-2}, a batch size of 4, a maximum of 60k iterations, and a dropout rate of 0.5. The Hyper-parameter α is set to 1.0. In the training procedure, the input images were randomly cropped to a 3D volume with sizes of $112 \times 112 \times 96$ for the NPC dataset and $96 \times 96 \times 96$ for the BraTS2019 dataset. To formally assess the segmentation performance, we utilize two extensively recognized metrics: the Dice Coefficient (DSC) and the Average Surface Distance (ASD) for quantitative evaluation.

3.2 Results

Comparison with State-of-the-art Methods We first compared our method with five semi-supervised learning approaches, such as mean teacher (MT) [15], interpolation consistency training (ICT) [16], entropy minimization (EM) [17], uncertainty rectified pyramid consistency (URPC) [10] and mutual learning with reliable pseudo label (MLRPL) [14]. Furthermore, we compared our method with semi-supervised multi-modal learning approaches, such as multi-modal contrastive mutual learning (MMCML) [20] and cross-modality collaboration (CMC) [23]. For fair comparison, for single-modal semi-supervised segmentation methods, we concatenate multi-modal images along the channel dimension prior to inputting them into the segmentation model.

Table 1 presents the results of our model and other semi-supervised methods for tumor segmentation in terms of DSC and ASD metrics. The results demonstrate that our method outperforms the comparison methods in both multi-modal MR images and multi-modal CT-MR images, achieving high-accuracy segmentation results with both 5% and 10% labeled data ratios. Due to the highly variable spatial location distribution, irregular shapes and edges, as well as the low contrast between the tumor and the background, tumor segmentation is considered much more challenging than organ segmentation. The limited labeled data further complicates the task of tumor segmentation, thereby affecting clinical diagnosis and treatment evaluation. Our approach exploits multi-modal information and reduces feature discrepancies through effective multi-stage feature fusion and enhancement, whereas existing methods typically rely on consistency constraints applied to prediction results [20] or bottleneck features [23]. Moreover, our contrastive mutual learning loss helps reduce overfitting and error accumulation during training. As a result, our method overcomes these challenges and achieves the highest DSC scores for both brain tumor and NPC tumor segmentation, demonstrating its effectiveness in complex scenarios. Fig. 2 visualizes the results, demonstrating that our method closely aligns with the ground truth. It can accurately segment brain tumors with complex shapes and capture more edge details. Moreover, our method excels in identifying and segmenting dispersed NPC tumors, particularly outperforming other approaches when using only 5% labeled data. A DSC score above 80% is often used as a practical benchmark for whole tumor segmentation in clinical practice [12], while a 3 *mm* margin of error is widely accepted as sufficient for head and neck radiotherapy [7]. Our semi-supervised method achieves segmentation accuracy within this clinically acceptable range. Therefore, our method successfully models the spatial distributions of pathological structures and is able to fully utilize consistent multi-modal information, ultimately achieving high-accuracy segmentation results in complex scenarios.

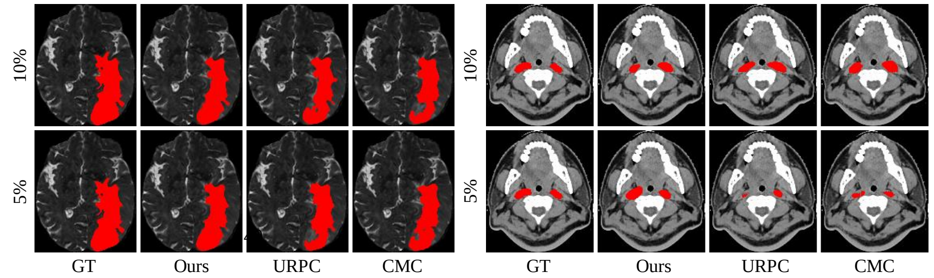

Fig. 2. Qualitative comparison between our method and SOTA semi-supervised methods on the BraTS 2019 and NPC datasets. The first row used 10% labeled data, and the second row used 5% labeled data.

Ablation Study We conduct ablation studies on the multi-scale multi-modal fusion (MMF), modality-aware enhancement (MAE), and multi-modal contrastive mutual learning (MCML) components of our network, assessing their impact on performance. Initially, we establish a baseline network without these components. Then, we incrementally integrate each of the three key components into the baseline network to systematically evaluate their individual contributions. Table 2 presents the quantitative evaluation results of our ablation study. The results demonstrate that these three components can effectively enhance segmentation accuracy. Specifically, the MMF strategy promotes feature fusion and

Table 1. Quantitative comparison of our method with other state-of-the-art (SOTA) methods on the BraTS 2019 and NPC datasets using 5% and 10% labeled data.

Labeled (%)	Method	BraTS 2019 Dataset		NPC Dataset	
		DSC (%) ↑	ASD (mm) ↓	DSC (%) ↑	ASD (mm) ↓
5	MT [15]	82.59±10.09	3.11±3.96	67.59±8.56	3.13±2.59
	ICT [16]	79.24±12.14	3.10±3.44	68.27±8.43	2.45±1.38
	EM [17]	81.38±10.58	3.76±4.55	68.31±9.11	2.72±2.02
	URPC [10]	83.14±8.93	3.19±3.70	69.03±8.57	2.29±1.69
	MLRPL [14]	81.01±12.47	2.52±3.48	60.05±18.18	4.70±10.00
	MMCML [20]	75.83±17.53	8.39±9.36	52.54±8.54	7.47±4.11
	CMC [23]	80.71±7.70	2.80±4.68	67.90±8.26	2.42±1.39
	Ours	**85.16±7.10**	**2.38±2.83**	**69.33±9.16**	**2.19±1.45**
10	MT [15]	84.02±8.66	3.45±3.96	70.73±6.47	1.80±0.84
	ICT [16]	84.16±8.58	3.08±3.44	71.48±6.40	2.17±2.33
	EM [17]	83.84±9.42	3.21±3.76	71.30±7.12	**1.73±0.84**
	URPC [10]	84.82±9.35	2.33±2.57	71.47±7.64	1.88±0.99
	MLRPL [14]	82.68±13.69	2.52±2.87	70.81±8.07	2.53±1.61
	MMCML [20]	79.97±13.81	5.56±5.21	53.61±8.69	6.04±3.18
	CMC [23]	83.27±7.94	2.29±3.07	70.26±7.00	2.11±1.03
	Ours	**85.82±7.97**	**2.19±2.73**	**72.67±7.56**	1.74±0.89
100	FullySup	88.35±6.3	1.48±1.53	77.03±6.35	1.53±0.98

alignment in the early encoding stage, thereby reducing the discrepancies between T1ce and T2 modalities and achieving an absolute improvement in the dual-branch segmentation network. Notably, the branch of T1ce scans shows a greater increase, achieving a 13.5% improvement in DSC score, because it benefits significantly from shared features with the T2 branch. The MAE and MCML strategies highlight the modality-aware features and cross-modal mutual learning, respectively, thereby further improving the segmentation accuracy. Specifically, these strategies improve the DSC for the T1ce modality by 13.5%, 1.9%, and 2.0%, respectively, and for the T2 modality by 3.7%, 0.8%, and 1.5%. Therefore, by combining these three components within our semi-supervised framework, we achieve the best performance in complex tumor segmentation with limited labeled data.

Table 2. Ablation study of our method on the BraTS dataset using 5% labeled data.

Method				DSC (%) ↑		ASD (mm) ↓	
Baseline	MMF	MAE	MCML	T1ce	T2	T1ce	T2
✓	×	×	×	68.44±14.46	79.28±10.82	4.28±3.07	2.93±2.73
✓	✓	×	×	81.94±10.48	82.98±9.64	3.10±3.69	3.21±3.87
✓	✓	✓	×	83.84±8.74	83.74±8.63	3.66±4.25	3.91±4.44
✓	✓	×	✓	83.96±8.02	84.43±7.56	2.62±3.54	2.70±3.53
✓	✓	✓	✓	85.16±7.10	84.95±7.25	2.38±2.83	2.67±3.09

4 Conclusion

In this paper, we propose a novel approach for semi-supervised multi-modal medical image segmentation. We introduce a multi-stage multi-modal feature fusion and enhancement strategy to promote feature sharing and reduce feature discrepancies among modalities. Additionally, this strategy emphasizes important modality-aware features. Furthermore, we introduce multi-modal contrastive mutual learning to achieve cross-modal consistency across different modalities. Extensive experimental results on the BraTS and NPC datasets demonstrate that we outperform the state-of-the-art approaches and achieve highly accurate segmentation performance in complex situations. Future work will aim to extend and evaluate the approach on more challenging medical image segmentation tasks and across diverse modality combinations.

Acknowledgments. This work was supported by the National Key Research and Development Program of China (No.2019YFF01014404) and Natural Science Foundation of China (No.62172013).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, X., Zhou, H.Y., Liu, F., Guo, J., Wang, L., Yu, Y.: Mass: Modality-collaborative semi-supervised segmentation by exploiting cross-modal consistency from unpaired ct and mri images. Medical Image Analysis **80**, 102506 (2022)
2. Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S.: Correlation-aware mutual learning for semi-supervised medical image segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 98–108. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-43907-0_10
3. Han, K., Sheng, V.S., Song, Y., Liu, Y., Qiu, C., Ma, S., Liu, Z.: Deep semi-supervised learning for medical image segmentation: A review. Expert Systems with Applications **245**, 123052 (2024)
4. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7132–7141 (2018). https://doi.org/10.1109/CVPR.2018.00745
5. Li, D., Yang, B., Zhan, W., He, X.: Multi-category graph reasoning for multi-modal brain tumor segmentation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 445–455. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72111-3_42
6. Li, Z., Huang, C., Xie, S.: Multimodality-assisted semi-supervised brain tumor segmentation in nondominant modality based on consistency learning. IEEE Transactions on Instrumentation and Measurement **73**, 1–11 (2024)

7. Lin, L., Dou, Q., Jin, Y.M., Zhou, G.Q., Tang, Y.Q., Chen, W.L., Su, B.A., Liu, F., Tao, C.J., Jiang, N., et al.: Deep learning for automated contouring of primary tumor volumes by mri for nasopharyngeal carcinoma. Radiology **291**(3), 677–686 (2019)
8. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 8801–8809 (2021)
9. Luo, X., Hu, M., Song, T., Wang, G., Zhang, S.: Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In: International conference on medical imaging with deep learning. pp. 820–833. PMLR (2022)
10. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Medical Image Analysis **80**, 102517 (2022)
11. Meng, D., Li, S., Sheng, B., Wu, H., Tian, S., Ma, W., Wang, G., Yan, X.: 3d reconstruction-oriented fully automatic multi-modal tumor segmentation by dual attention-guided vnet. The Visual Computer **39**(8), 3183–3196 (2023)
12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2015)
13. Shen, W., Peng, Z., Wang, X., Wang, H., Cen, J., Jiang, D., Xie, L., Yang, X., Tian, Q.: A survey on label-efficient deep image segmentation: Bridging the gap between weak supervision and dense prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 9284–9305 (2023)
14. Su, J., Luo, Z., Lian, S., Lin, D., Li, S.: Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. Medical Image Analysis **94**, 103111 (2024)
15. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
16. Verma, V., Kawaguchi, K., Lamb, A., Kannala, J., Solin, A., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. Neural Networks **145**, 90–106 (2022)
17. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2512–2521 (2019)
18. Weng, Y., Zhang, Y., Wang, W., Dening, T.: Semi-supervised information fusion for medical image analysis: Recent progress and future perspectives. Information Fusion **106**, 102263 (2024)
19. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 605–613. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_67
20. Zhang, S., Zhang, J., Tian, B., Lukasiewicz, T., Xu, Z.: Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation. Medical Image Analysis **83**, 102656 (2023)

21. Zhang, Y., Yang, J., Tian, J., Shi, Z., Zhong, C., Zhang, Y., He, Z.: Modality-aware mutual learning for multi-modal medical image segmentation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 589–599. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_56
22. Zhou, T., Ruan, S., Canu, S.: A review: Deep learning for medical image segmentation using multi-modality fusion. Array **3-4**, 100004 (2019)
23. Zhou, X., Sun, Y., Deng, M., Chu, W.C.W., Dou, Q.: Robust semi-supervised multimodal medical image segmentation via cross modality collaboration. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 57–67. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72378-0_6