Missing as Masking: Arbitrary Cross-modal Feature Reconstruction for Incomplete Multimodal Brain Tumor Segmentation

Zhilin Zeng[1], Zelin Peng[1], Xiaokang Yang[1], and Wei Shen[1(✉)]

[1] MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, China.
wei.shen@sjtu.edu.cn

Abstract. Automatic brain tumor segmentation using multimodal MRI images is a critical task in medical imaging. A complete set of multimodal MRI images for a subject offers comprehensive views of brain tumors, and thus providing ideal tumor segmentation performance. However, acquiring such modality-complete data for every subject is frequently impractical in clinical practice, which requires a segmentation model to be able to 1) flexibly leverage both modality-complete and modality-incomplete data for model training, and 2) prevent significant performance degradation in inference if certain modalities are missing. To meet these two demands, in this paper, we propose M^{3}FeCon (**M**issing as **M**asking: arbitrary cross-**M**odal **Fe**ature Re**Con**struction) for incomplete multimodal brain tumor segmentation, which can learn approximate modality-complete feature representations from modality-incomplete data. Specifically, we treat missing modalities also as masked modalities, and employ a strategy similar to Masked Autoencoder (MAE) to learn feature-to-feature reconstruction across arbitrary modality combinations. The reconstructed features for missing modalities act as supplements to form approximate modality-complete feature representations. Extensive evaluations on the BraTS18 dataset demonstrate that our method achieves state-of-the-art performance in brain tumor segmentation with incomplete modalities, especicall in enhancing tumor with 4.61% improvement in terms of Dice score.

Keywords: Incomplete Multimodal Segmentaion · Brain Tumor Segmentaion

1 Introduction

Utilizing multiple medical image modalities to jointly enhance diagnostic accuracy is widely adopted in many clinical applications. For instance, in brain tumor segmentation, multi-modal Magnetic Resonance Imaging (MRI) is used to obtain comprehensive views of brain tumors. This includes a variety of image modalities such as T1-weighted (T1), contrast-enhanced T1-weighted (T1c), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR) images,

each offering distinct tissue contrasts to help identify different tumor areas in the brain. However, clinical practice often present scenarios where some image modalities are unavailable due to various constraints such as scanning cost, time, and patient condition. These limitations can lead to significant model performance degradation when certain image modalities are missing during inference and also intensify the scarcity of modality-complete training datasets. Consequently, tackling incomplete multimodal learning for brain tumor segmentation is of great clinical value.

To deal with this problem, a critical challenge is how to learn a modality-complete feature representation for each subject from modality-incomplete data, so that the ideal tumor segmentation performance can be achieved. Existing works can be mainly divided into two classes: (1) feature-fusion based methods and (2) knowledge distillation based methods. For the feature-fusion based methods, one of the pioneering work is HeMIS [5] that learns a fused feature representation for each subject by computing statistics from the available modalities. Following works [1,4,15,14,13,11] propose more advanced fusion strategies. For instance, RFNet [4] proposes a region-aware fusion strategy to differentiate the contribution of the available modalities to different regions. MMFormer [14] uses inter-modal and intra-modal transformer architecture to learn fused feature representations from the available modalities. However, the fused feature representation for each subject is not the modality-complete feature representation, which still lacks the information from the missing modalities. The knowledge distillation based methods [6,12,2,8,9] attempt to learn modality-complete feature representations by distilling knowledge from modality-complete data to modality-incomplete data. However, their success relies on a strict precondition: the modality-complete data for each subject should be provided for training. M^{3}AE [9] adopts a self-distillation strategy to distill knowledge between features of different modality combinations, yet the feature representations it learns are compromised modality-shared representations rather than modality-complete feature representations.

In this paper, we propose M^{3}FeCon (**M**issing as **M**asking: arbitrary cross-**M**odal **Fe**ature Re**Con**struction), a novel approach for incomplete multimodal brain tumor segmentation, which can learn approximate modality-complete feature representations from modality-incomplete data. Unlike knowledge distillation based methods, which depend on modality-complete data for training, M^{3}FeCon learns to approximate modality-complete features through cross-modal feature-to-feature reconstruction. Specifically, we employ a strategy similar to Masked Autoencoder (MAE) to realize this cross-modal feature-to-feature reconstruction, where we random mask some modalities of each subject and treat the missing modalities also as the masked modalities, then reconstruct feature representations of masked modalities from the remaining unmasked ones. The reconstructed features for missing modalities act as supplements to form the approximate modality- complete feature representation for each subject. M^{3}FeCon enables a feature-to-feature translation between arbitrary modality combinations, and thus allows for a flexible utilization of both modality-complete and

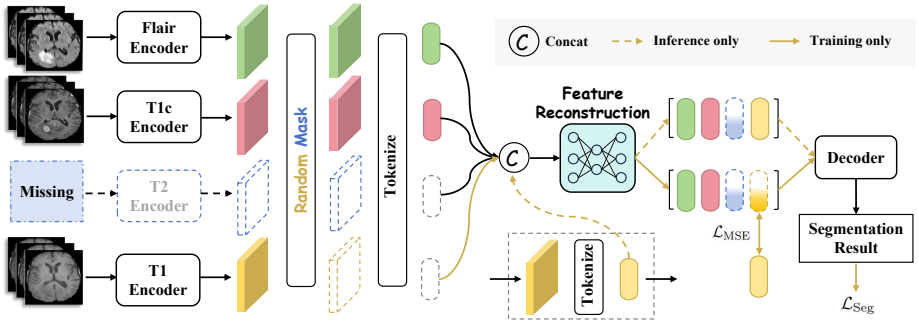

Fig. 1. Our M^{3}FeCon architecture. During training, some modalities of a subject are missing. The encoders are applied to the remaining available modalities. We randomly mask some modalities and treat the missing modalities also as the masked modalities. Then, all modalities are tokenized and processed by a multi-layer transformer block to reconstruct the features of the masked ones. After training, the random mask strategy is discarded and all available modalities are applied to reconstruct the features of missing modalities, which act as supplements to form approximate modality-complete feature representations for segmentation.

modality-incomplete data during model training. Our novel contributions are as follows:

- We propose M^{3}FeCon, an arbitrary cross-modal feature reconstruction model for incomplete multimodal brain tumor segmentation by treating missing modalities as masked modalities. As far as we know, it is the first attempt to learn modality-complete feature representations from the training with modality-incomplete data.
- Experiment results on BraTS18 dataset demonstrate that M^{3}FeCon achieve state-of-the-art results in incomplete multimodal brain tumor segmentation, especially in enhancing tumor with 4.61% Dice score improvement.

2 Method

We denote the input data as $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M\}$, representing M modalities per sample, with $\mathbf{X}_m \in \mathbb{R}^{H\times W\times D}$ being the image of the m-th modality and $H \times W \times D$ being the size of the 3D medical images. The presence of modalities is indicated by a binary vector $\mathbf{p} = [p_1, p_2, \ldots, p_M]$, where $p_m = 1$ indicates the m-th modality is present and $p_m = 0$ indicates it is absent. The training dataset comprises a collection of tuples $\{(\mathcal{X}^{(i)}, \mathbf{p}^{(i)}, \mathbf{Y}^{(i)})\}_{i=1}^{N}$, with N denoting the total number of training samples. Each tuple includes the multimodal inputs $\mathcal{X}^{(i)}$, the modality presence vector $\mathbf{p}^{(i)}$, and the corresponding ground truth of segmentation map $\mathbf{Y}^{(i)} \in \{0, 1, \ldots, c\}^{H\times W\times D}$, where $\{0, 1, \ldots, c\}$ is a set of pre-defined semantic classes.

Fig. 1 illustrates the overall architecture of our proposed M^{3}FeCon. Our framework is based on an encoder-decoder structure. Each modality image undergoes initial encoding to extract high-level features, denoted as $\mathbf{F}_m = \mathbf{p}[m] \cdot f_m(\mathbf{X}_m)$, where f_m is the encoder corresponding to the m-th modality and $\mathbf{p}[m]$ ensures encoding is performed only for available modalities. Following this is a multi-layer transformer block, which aims to learn feature-to-feature reconstruction across arbitrary modality combinations through a random modality masking strategy. The decoder is then fed with a combined features from each modality, selectively using either the originally encoded or the reconstructed features based on their availability and the random masking applied during the current training iteration. Finally, The training process is supervised by both a reconstruction loss and a segmentation loss, ensuring features generated as supplements can most effectively enhance segmentation results.

2.1 Arbitrary Cross-modal Feature Reconstruction

Given the initial presence vector $\mathbf{p}$, we apply a random modality masking strategy through binary vector $\mathbf{r}$ to selectively mask among the available modalities, while ensuring at least one remains unmasked for each training iteration. $\mathbf{r}[m] = 0$ indicates that the m-th modality is masked. The final presence state of each modality in current training iteration is represented by the binary indicator $\mathbf{g}[m]$: $\mathbf{g}[m] = \mathbf{p}[m] \wedge \mathbf{r}[m]$.

We then transform each unmasked modality feature map $\mathbf{F}_m$ into modality tokens $\mathbf{T}_m$ through a flattening and projection process, defined as $\mathbf{T}_m = f_p(\text{flatten}(\mathbf{F}_m))$, where f_p represents a projection function that maps the flattened feature map into a token space of dimension $N \times C$, where N is length of the tokens and C is the hidden dimension.

For each modality that is masked or originally missing, we introduce a corresponding learnable replacement tokens $\mathbf{T}_m^r$ to fill in its position. These replacement tokens inform transformer about the specific modality features to reconstruct and can also learn prior knowledge from each modality that can help to better reconstruct corresponding modality features. The final input sequence $\mathbf{T}^{\text{in}} \in \mathbb{R}^{MN \times C}$ is a concatenation of available modality tokens and learnable replacement tokens:

$$\mathbf{T}^{\text{in}} = [\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_M], \quad \mathcal{I}_m = \begin{cases} \mathbf{T}_m & \text{if } \mathbf{g}[m] = 1 \\ \mathbf{T}_m^r & \text{if } \mathbf{g}[m] = 0 \end{cases} \text{ for } 1 \leq m \leq M \tag{1}$$

Finally, we add positional encoding to $\mathbf{T}^{\text{in}}$ and feed it to a multi-layer transformer block f_r for feature reconstruction $\mathbf{T}^{\text{recon}} = f_r(\mathbf{T}^{\text{in}})$.

2.2 Training and Inference

Given the output tokens sequence $\mathbf{T}^{\text{recon}}$, we split it to get the reconstructed tokens of each modality $\{\mathbf{T}_m^{\text{recon}}\}_{m=1}^{M}$. Then we use Mean Square Error (MSE)

to calculate the reconstruction loss, which is only computed for a modality m if it is both present in the original dataset ($\mathbf{p}[m] = 1$) and has been randomly masked for the current iteration ($\mathbf{r}[m] = 0$). The reconstruction loss is given by:

$$\mathcal{L}_{\text{recon},m} = \begin{cases} \text{MSE}(\hat{\mathbf{T}}_m, \mathbf{T}_m^{\text{recon}}), & \text{if } \mathbf{p}[m] = 1 \text{ and } \mathbf{r}[m] = 0, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $\hat{\mathbf{T}}_m$ is a detached copy of $\mathbf{T}_m$. The overall reconstruction loss is averaged over all modalities: $\mathcal{L}_{\text{recon}} = \sum_{m=1}^{M} \mathcal{L}_{\text{recon},m}$.

To perform segmentation, we first unfold the reconstructed modality tokens to get the corresponding reconstructed feature map $\mathbf{F}_m^{\text{recon}} \in \mathbb{R}^{h \times w \times d \times C}$, where $h \times w \times d$ is the size of the feature map. The decoder is fed with a concatenated feature map $\mathbf{F}_{\text{concat}} \in \mathbb{R}^{h \times w \times d \times MC}$ along the channel dimension, composed of either the original or reconstructed features for each modality, depending on their availability and the masking applied for the current training iteration. Specifically, for each modality m, the original feature map $\mathbf{F}_m$ from the encoder is used if the modality is available in the input ($\mathbf{p}[m] = 1$) and not masked in the current iteration ($\mathbf{r}[m] = 1$). Otherwise, the reconstructed feature map $\mathbf{F}_m^{\text{recon}}$ is employed. The final concatenated feature map $\mathbf{F}_{\text{concat}}$ can be represented as:

$$\mathbf{F}_{\text{concat}} = \bigoplus_{m=1}^{M} \begin{cases} \mathbf{F}_m & \text{if } \mathbf{p}[m] = 1 \wedge \mathbf{r}[m] = 1 \\ \mathbf{F}_m^{\text{recon}} & \text{otherwise} \end{cases} \tag{3}$$

where $\bigoplus$ denotes the concatenation operation along the channel dimension. $\mathbf{F}_{\text{concat}}$ is then fed to the decoder and produce the segmentation map. The segmentation loss is calculated as:

$$\mathcal{L}_{\text{seg}} = \text{Dice}(f_d(\mathbf{F}_{\text{concat}}), \mathbf{Y}) + \text{CrossEntropy}(f_d(\mathbf{F}_{\text{concat}}), \mathbf{Y}) \tag{4}$$

where f_d denotes the decoder. The total loss $\mathcal{L}_{\text{total}}$ is a combination of the reconstruction loss and the segmentation with α being a hyper-parameter to adjust the weight of reconstruction loss.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \alpha \cdot \mathcal{L}_{\text{recon}} \tag{5}$$

In general, the segmentation task ensures the reconstructed features contribute positively to segmentation outcomes, meanwhile the reconstruction task guides feature learning from missing modalities, that can most effectively enhance segmentation performance. Given the diversity of modality combinations throughout the training process, it can be ensured that all types of feature-to-feature reconstruction process receive comprehensive guidance from both tasks.

During inference, given modality-incomplete data as input, our model reconstructs the features of missing modalities based on the features of available modalities. By concatenating the originally encoded and the reconstructed features, we form an approximate modality-complete feature representation and feed it to the decoder to deliver segmentation results.

Table 1. Comparison of model performance in terms of the Dice score (%) on the BraTS2018 dataset. The highest scores are highlighted in bold. "•" and "○" denote the presence and absence of the modality for testing, respectively.

Modalities				Enhancing tumor				Tumor Core				Whole Tumor			
F1	T1	T1c	T2	mmF	MD	M^{3}AE	Ours	mmF	MD	M^{3}AE	Ours	mmF	MD	M^{3}AE	Ours
•	○	○	○	37.98	35.32	34.96	**46.59**	61.61	63.24	66.02	**69.42**	86.37	86.68	88.04	**89.05**
○	•	○	○	31.38	32.22	36.54	**44.35**	56.96	65.83	**65.77**	65.72	67.92	75.05	73.83	**75.24**
○	○	•	○	71.37	67.49	73.04	**75.88**	75.93	80.94	**82.53**	82.31	72.62	**77.25**	75.16	75.26
○	○	○	•	41.92	42.88	46.93	**49.32**	64.61	67.49	69.14	**72.48**	81.43	84.67	84.22	**85.13**
•	•	○	○	41.75	38.41	48.16	**48.47**	66.47	71.73	70.53	**72.82**	87.46	85.14	88.49	**90.17**
•	○	•	○	73.93	72.45	74.08	**76.23**	78.34	82.28	84.03	**84.14**	87.61	86.96	88.85	**90.35**
•	○	○	•	46.47	45.27	40.57	**51.12**	70.18	72.42	70.59	**72.81**	87.99	86.42	89.31	**90.46**
○	•	•	○	72.77	68.44	74.58	**76.78**	79.01	81.97	**83.11**	82.88	74.75	77.97	76.50	**79.16**
○	•	○	•	43.85	42.68	44.73	**50.17**	69.89	71.92	71.46	**72.01**	82.46	85.43	**86.34**	86.03
○	○	•	•	73.13	70.47	74.39	**76.55**	79.07	83.01	83.85	**84.75**	83.25	85.82	85.58	**86.61**
•	•	•	○	74.31	70.86	73.42	**76.74**	80.22	83.08	83.72	**84.81**	87.77	84.71	88.07	**89.62**
•	•	○	•	46.51	45.93	44.15	**52.63**	71.97	74.69	72.41	**75.98**	88.08	83.98	89.24	**90.24**
•	○	•	•	74.53	72.91	74.68	**77.64**	79.94	83.22	84.25	**84.64**	88.47	84.57	89.46	**90.07**
○	•	•	•	73.49	70.83	73.37	**77.33**	80.89	83.65	84.13	**84.39**	83.08	83.32	85.06	**86.79**
•	•	•	•	76.36	71.25	74.85	**77.81**	**86.23**	83.27	84.24	85.03	89.93	85.64	89.56	**90.69**
Average				58.65	56.49	59.23	**63.84**	73.42	76.58	77.05	**78.28**	83.28	83.57	85.18	**86.32**
p-value				3.8e-6	7.3e-6	1.9e-5	-	9.5e-6	6.0e-6	1.3e-5	-	1.7e-4	2.5e-5	2.2e-3	-

3 Experiments

Dataset and Implementation Details. The evaluation is conducted on the BraTS2018 [10] dataset, which includes 285 multimodal MRI scans across four modalities: T1, T1c, T2, and FLAIR. The dataset focuses on segmenting three key brain tumor sub-regions: enhancing tumor (ET), tumor core (TC), and whole tumor (WT), with expert-annotated ground truths provided for each case. We randomly split the dataset and use 200 cases for training and 85 cases for validation.

Our method was implemented in PyTorch 2.0 on a NVIDIA GTX 3090 GPU. The backbone network is a 3D UNet [3] architecture. At the bottleneck of the network, we incorporated a 4 layer transformer block with hidden dimension of size 512 to learn feature-to-feature reconstruction. The input size of each image modality is $128 \times 128 \times 128$ voxels and batch size is set to 2. We followed the data preprocessing and augmentation steps from nnunet [7]. The hyperparamater α for the reconstruction loss was set to 1. Our model was trained with Adam optimizer with an initial learning rate of 0.01 for 500 epochs.

Comparisons with the State-of-the-Arts. We compared our model with several other state-of-the-art methods for incomplete multimodal brain tumor segmentation, including feature-fusion based method (mmFormer(mmF)[14]) and knowledge distillation based methods (ModDrop++(MD) [8], M^{3}AE [9]). For

a fair comparison, all models listed in Table 1 were trained under the same settings as described in [8], where entire modality-complete datasets are used and different partial modalities are randomly dropped in each training iteration to simulate situations with missing modalities. We use Dice score as the evaluation metric. Table 1 shows the segmentation performance on the three types of tumors evaluated on all 15 possible cases of modality availability. It can be observed that our model achieves the highest scores on most of the modality-incomplete cases across all three types of tumors. On average, our model improves the dice scores upon previous state-of-the-art methods with 4.61% on enhancing tumor, 1.23% on tumor core and 1.14% on whole tumor. Our model's advantage is particularly evident in enhancing tumor, due to the outstanding ability of reconstructing missing modality information, thus improving segmentation accuracy greatly in enhancing tumor segmentation when T1c modality is absent. Besides, on the harder cases with more modalities or important modality (T1c) missing, our method has even larger improvements. For example, we achieve over 10% gains with only {Flair} for enhancing tumor, over 3% gains with {Flair} or {T2} for tumor core. The significant improvements on more challenging tasks and cases demonstrate our method's superiority.

It is also worth noting that our model has fewer parameters compared to ModDrop++, a method that is also designed to learn approximate modality-complete features. By learning feature-to-feature reconstruction within the network's bottleneck, our model eliminates the need for an additional cumbersome teacher branch. Fig. 2 shows the visualization results of our model given various modality-incomplete data. More visualizations can be found in the supplementary material.

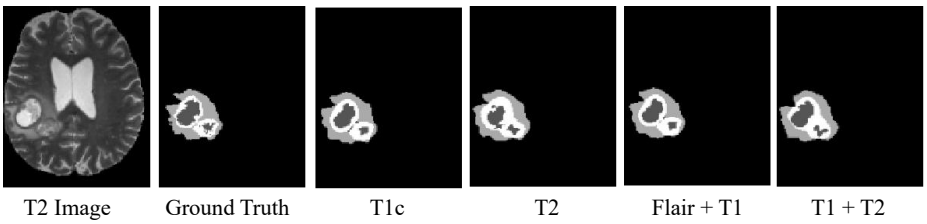

Fig. 2. Segmentation results of M3FeCon given various modality combinations.

Training with different missing rates. To evaluate the models' effectiveness of including modality-incomplete data with different missing rates for training, we pre-processed and partioned the training set to simulate diverse completeness levels, by randomly dropping 1-3 modalities from certain samples to create predefined sets of modality-incomplete data. Fig. 3 shows the segmentation performance by training with different ratio of modality-complete to modality-incomplete data. ModDrop++ is excluded as it cannot utilize modality-incomplete

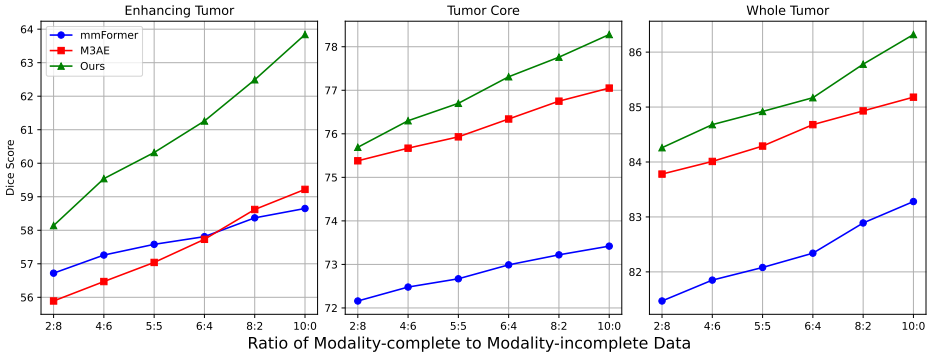

Fig. 3. Performance comparison by training with different ratios of modality-complete to modality-incomplete data.

data for training. The results shows that our method consistently outperforms other methods given different proportion of modality-incomplete data in training set. This demonstrates that our model can better fully utilize the modality-incomplete data by learning to approximate modality-complete feature representations, thus better improving segmentation results, rather than learning a compromised shared or fused feature representations. It can also be seen that as the proportion of modality-complete data in the training set increases, our model has a growing performance lead over others, primarily attribute to a better feature reconstruction ability if given more complete data during training.

Ablation Study. We further conduct ablation study on the contribution of feature-to-feature reconstruction objective to segmentation performance and the effectiveness of the learnable replacement tokens. As shown in Table 2, without the guidance from reconstruction target ($\alpha = 0$), the segmentation performance drops greatly. Results in Table 3 illustrate that introducing learnable replacement tokens for cross-modal feature reconstruction can improve performance.

Table 2. Ablation study on different value of α for reconstruction loss.

α	Average Dice [%]		
	Enhancing	Core	Whole
0	54.68	72.16	81.75
1	**63.84**	**78.28**	**86.32**
1.5	63.05	77.45	85.53

Table 3. Ablation study on learnable replacement tokens.

Config	Average Dice [%]		
	Enhancing	Core	Whole
w/ $\mathbf{T}_m^r$	**63.84**	**78.28**	**86.32**
w/o $\mathbf{T}_m^r$	63.27	77.69	85.45

4 Conclusion

In this paper, we proposed M^{3}FeCon, a novel approach for segmentation using incomplete multimodal data. Specifically, we first masked some modalities and treated missing modalities also as the masked modalities. Then, M^{3}FeCon introduced an MAE-like strategy for reconstructing masked modalities, which effectively learns to approximate modality-complete feature representations for segmentation. Experimental results on the BraTS18 dataset validated the effectiveness of our method.

Acknowledgments. This work was supported by the NSFC under Grant 62322604 and 62176159, and in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0102.

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

References

1. Chen, C., Dou, Q., Jin, Y., Chen, H., Qin, J., Heng, P.A.: Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22. pp. 447–456. Springer (2019)
2. Chen, C., Dou, Q., Jin, Y., Liu, Q., Heng, P.A.: Learning with privileged multimodal knowledge for unimodal segmentation. IEEE transactions on medical imaging **41**(3), 621–632 (2021)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 424–432. Springer (2016)
4. Ding, Y., Yu, X., Yang, Y.: 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 (2021)
5. Havaei, M., Guizard, N., Chapados, N., Bengio, Y.: Hemis: Hetero-modal image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 469–477. Springer (2016)
6. Hu, M., Maillard, M., Zhang, Y., Ciceri, T., La Barbera, G., Bloch, I., Gori, P.: Knowledge distillation from multi-modal to mono-modal segmentation networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23. pp. 772–781. Springer (2020)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)

8. Liu, H., Fan, Y., Li, H., Wang, J., Hu, D., Cui, C., Lee, H.H., Zhang, H., Oguz, I.: Moddrop++: A dynamic filter network with intra-subject co-training for multiple sclerosis lesion segmentation with missing modalities. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 444–453. Springer (2022)
9. Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y.: M3ae: Multimodal representation learning for brain tumor segmentation with missing modalities. arXiv preprint arXiv:2303.05302 (2023)
10. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
11. Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., Carneiro, G.: 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 (2023)
12. Wang, Y., Zhang, Y., Liu, Y., Lin, Z., Tian, J., Zhong, C., Shi, Z., Fan, J., He, Z.: Acn: Adversarial co-training network for brain tumor segmentation with missing modalities. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 410–420. Springer (2021)
13. Wang, Z., Hong, Y.: A2fseg: Adaptive multi-modal fusion network for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 673–681. Springer (2023)
14. Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., Zhang, Y., He, Z., Zheng, Y.: 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 (2022)
15. Zhao, Z., Yang, H., Sun, J.: Modality-adaptive feature interaction for brain tumor segmentation with missing modalities. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 183–192. Springer (2022)