

A Multimodal Feature Distillation with CNN-Transformer Network for Brain Tumor Segmentation with Incomplete Modalities

Ming Kang , Member, IEEE, Fung Fung Ting , Raphaël C.-W. Phan , Senior Member, IEEE, Zongyuan Ge , Senior Member, IEEE, and Chee-Ming Ting , Senior Member, IEEE

Abstract — Existing brain tumor segmentation methods usually utilize multiple Magnetic Resonance Imaging (MRI) modalities in brain tumor images for segmentation, which can achieve better segmentation performance. However, in clinical applications, some modalities are missing due to resource constraints, leading to severe degradation in the performance of methods applying complete modality segmentation. In this paper, we propose a Multimodal feature distillation with Convolutional Neural Network (CNN)-Transformer hybrid network (MCTSeg) for accurate brain tumor segmentation with missing modalities. We first design a Multimodal Feature Distillation (MFD) module to distill feature-level multimodal knowledge into different unimodality to extract complete modality information. We further develop a Unimodal Feature Enhancement (UFE) module to model the relationship between global and local information semantically. Finally, we build a Cross-Modal Fusion (CMF) module to explicitly align the global correlations among different modalities even when some modalities are missing. Complementary features within and across different modalities are refined via the CNN-Transformer hybrid architectures in both the UFE and CMF modules, where local and global dependencies are both captured. Our ablation study demonstrates the importance of the proposed modules with CNN-Transformer networks and the convolutional blocks in Transformer for improving the performance of brain tumor segmentation with missing modalities. Extensive experiments on the BraTS2018 and BraTS2020 datasets show that the proposed MCTSeg framework outperforms the state-of-the-art methods in missing modalities cases. Our code is available at: <https://github.com/mkang315/MCTSeg>.

Index Terms— Multimodal MRI, missing modalities, medical image segmentation, feature distillation, cross modalities.

Manuscript received April 00, 2024; accepted XXXXXX 00, 2024. Date of publication XXXXXX 00, 2024; date of current version XXXXXX 00, 2024. (Corresponding author: Chee-Ming Ting.)

M. Kang, F. F. Ting, R. C.-W. Phan, and C.-M. Ting are with the School of Information Technology, Monash University, Malaysia Campus, Subang Jaya 47500, Malaysia (e-mails: ming.kang@monash.edu, ting.fung@monash.edu, raphael.phan@monash.edu, ting.cheeming@monash.edu).

Z. Ge is with the Faculty of Information Technology, Monash University, Melbourne, VIC 3800, Australia, and also with the Augmented Intelligence and Multimodal Analytics for Health Lab, and the Monash Medical AI Group, Monash University, Melbourne, VIC 3800, Australia (e-mail: zongyuan.ge@monash.edu).

Digital Object Identifier 00.0000/XXXX.2024.0000000

I. INTRODUCTION

ACCURATE segmentation of brain tumors is crucial for tumor diagnosis and treatment. Multimodal image segmentation has emerged in medical imaging processing [1]–[5]. Multimodal MRI images have different imaging features, including Fluid-attenuated inversion recovery (Flair), T1-weighted (T1), enhanced T1-weighted (T1ce), and T2-weighted (T2) as shown in Fig. 1. Different modalities usually have various information in content, so multimodal complementary information can be used to strengthen the feature representation ability of the model. However, in practical applications, there may be cases where some modalities are missing [6], [7] or modality data are scarce [8], thus, incomplete modalities bring new challenges to multimodality learning tasks. Incomplete multimodal learning methods aim to handle any available modality during model inference. The segmentation of brain tumors in incomplete multimodality MRI involves the problem of segmenting brain tumor regions from some modality-missing brain tumor images. Segmenting tumor regions from incomplete brain tumor data [9]–[12] is more clinically meaningful compared to brain tumor segmentation tasks with complete modalities. However, in some cases, certain modalities may not be available due to technical issues (e.g., image corruption, visual artifacts, imaging protocols), patient allergies to contrast agents, or financial constraints. Segmentation of brain tumors in MRI with missing modalities poses a significant challenge due to the incomplete information available. Researchers and clinicians have developed specialized algorithms and techniques to address this challenge.

Segmentation methods that can adapt to incomplete modalities provide flexibility in clinical practice, ensuring consistent and reliable results across different scenarios. Methods examined in the previous studies include the classical synthesis techniques alongside the more recent strategies that utilize deep learning. The newer approaches include common latent space models, knowledge distillation networks, mutual information maximization, and generative adversarial networks [13], [14]. The issue of missing modalities presents a significant challenge that needs to be addressed when conducting brain tumor segmentation tasks in multimodal MRI. The approaches of brain tumor segmentation in MRI with missing modalities often involve leveraging the available modalities

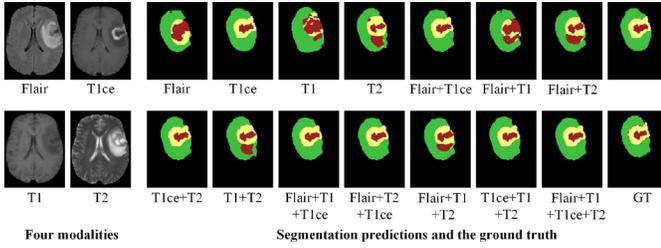


Fig. 1. Examples of four-modality MRI images. From left to right, there are fluid-attenuated inversion recovery images (Flair), T1ce images (T1ce, using contrast agent in T1 images), T1-weighted images (T1), T2-weighted images (T2), and images with ground truth labels that contain four sections: black area(background and healthy tissue), green area (enhancing tumor), red area (gangrene and non-enhancing tumor), and yellow area (peritumoral edema).

effectively, incorporating classical CNN approaches, such as U-Net [15] derived 3D U-Net [16] series, and in recent years, utilizing Transformer [17] or Vision Transformer (ViT) [18] and CNN-Transformer methodologies to compensate for missing imaging modalities.

In this paper, we propose a Multimodal feature distillation with CNN-Transformer hybrid network for incomplete multimodal brain tumor Segmentation (MCTSeg). First, we design a Multimodal Feature Distillation (MFD) module that uses a multimodal encoder network to distill complementary multimodal knowledge into the unimodal encoders to extract modality-specific features that are robust to missing modalities. Secondly, we develop a novel Unimodal Feature Enhancement (UFE) module that incorporates a convolutional block (ConvBlock) into Transformer as an adapter of the network to introduce inductive bias (e.g., translation invariance and locality) without affecting the transformer’s global modeling capability. Thirdly, we build a Cross-Modal Fusion (CMF) module that enables the fusion of feature representations of different modalities even when some modalities are missing. It also uses transformer-based architecture to build inter-modal relationships and incorporates a ConvBlock adapter to learn local features shared among modalities.

The main contributions of this work are summarized as follows:

- 1) To our best knowledge, MCTSeg is the first feature distillation-based framework for incomplete multimodal learning of brain tumor segmentation. It is an end-to-end network that allows for the transferring of multimodal information to unimodal feature extractors or encoders. It can extract representative features for each imaging modality, and learn the relationships between different modality representations to increase robustness against cases of missing modalities.
- 2) By integrating CNNs, which are adept at the locality inductive bias, into Transformers, the unimodal feature extraction and cross-modal fusion modules are designed to capture both local and global dependencies for complementary information within and across different modalities and compensate for the absence of certain modalities.
- 3) We conduct extensive experiments on the BraTS 2018

and BraTS2020 brain tumor datasets and show that our method achieves better performance than some state-of-the-art methods in situations of missing modalities.

II. RELATED WORK

A. CNN-Transformer Methods for Medical Image Segmentation

Several CNN-Transformer methods have been investigated for brain tumor segmentation. TransBTS [9] learned the global correlation by adding a Transformer structure at the bottom of the U-shaped network, and combined the local spatial features extracted by 3D CNN to effectively improve the segmentation accuracy of the model. TransUnet [19] applied Transformer in the field of medical image segmentation for the first time, combining the advantages of U-Net and Transformer, modeling local context information through CNN and obtaining long-range dependencies from Transformer in low-resolution images to improve the feature extraction ability of the encoders. TransFuse [20] combined the complementary nature of the Transformer and CNN to enhance this model’s segmentation capabilities in a parallel way. Swin-UNet [21] employed an encoder based on hierarchical representations and local windows to extract contextual features.

B. Feature Distillation in Knowledge Distillation

Knowledge distillation is to get the student network trained by minimizing the distillation loss, which realizes knowledge transfer between teacher and student models that are isomorphic or non-isomorphic networks. In contrast to transferring label knowledge from teacher network [22], feature distillation as another approach of knowledge distillation utilizes intermediate representations learned by the teacher network as hints to improve the final performance of the student network [23]. Intermediate representation distillation is more effective than label knowledge distillation, which improves the representation ability and information volume of transfer knowledge. The classic methods of isomorphic feature distillation, which does not need size matching of feature maps, include FitNets [23], attention transfer [24], probabilistic knowledge transfer [25], factor transfer [26], overhaul [27], dynamic prior knowledge [28] among others.

With the development of distillation techniques, some advanced feature distillation methods have been applied to medical images. Contrastive representation distillation [29] combines a contrastive objective measuring the mutual information between the representations learned by a teacher and a student networks for feature distillation. Xing et al. [30] employed pathology-genomic knowledge distilled from discrepancy-induced contrastive distillation between the teacher’s and student’s features to improve the accuracy of glioma grading.

C. Brain Tumor Segmentation with Incomplete Modalities

Recently, incomplete multimodal brain tumor segmentation methods focus on end-to-end encoder-decoder network architecture approaches that handle any subset of available

modalities, rather than training multiple networks. The Hetero-Modal Image Segmentation (HeMIS) network [31] learned the embedding of multimodal information by computing the mean and variance of the features of any available modality. Dorent et al. [32] proposed a Hetero-Modal Variational Encoder-Decoder (U-HVED) network for joint modality completion and segmentation tasks, by using information from multiple modality images for joint training. Shen & Gao [33] treated different missing modalities as a specific domain, employed two segmentation networks to segment the complete multimodal and the missing modal images respectively, and used confrontation learning in the segmentation device to project these features into the same common space. However, it is challenging to align the different distributions when a large number of modalities are missing. Zhang et al. [34] proposed an adaptive feature integration method to learn to segment multimodalities using an unimodal model, but this method can only handle cases where pairs or multiple modalities are available. mmFormer [11] employed Transformer to build intra-modal and inter-modal relationships to align global feature representations between different modalities. Ting & Liu [35] presented a multimodal transformer network to learn features of missing modalities for the segmentation of brain tumors. Our approach differs from previous approaches in that we utilize Transformer to not only model the relations within modalities but also consider the interactions between modalities.

In addition to feature completion, recent approaches have explored the use of feature separation [36], correlation representation [37]–[39], attention mechanisms [40], adversarial co-training networks [41], [42], self-supervised multimodal representation [43], and domain-specific classification objective [44] for robust multimodal brain tumor segmentation. Chen et al. [36] used the method of feature disentanglement and gating fusion to fuse the information of multimodality MRI images. The encoder network uses the feature-connected entanglement mechanism to separate the common and individual characteristics of different modal images, and the decoder network uses the gating fusion mechanism to fuse multiple entangled features. It mainly solves the problem of inconsistency and noise in multimodal images; however, when only a few modalities are available, the model performance degrades severely because using only one or two modalities may not be enough to provide reliable missing modality features.

Incomplete multimodal brain tumor segmentation via knowledge distillation is to synthesize missing modalities with a segmentation network with complete modality information [45]–[51]. Hu et al. [45] employed information from multimodal images to train a complex multimodal model to extract multimodal features and used these features together with unimodal images to train a simple unimodal model, making it easy to learn the knowledge of multimodal models. These methods have achieved a certain level of performance, but they need complicated methods to train specific models for each subset of missing modalities, which renders them less suitable for clinical practice. To our knowledge, feature distillation in brain tumor segmentation has not been explored in the existing

literature.

III. METHODS

Fig. 2 illustrates an overview of the proposed MCTSeg network for brain tumor segmentation with missing modalities. We construct our networks based on 3D U-Net, which consists of encoder and decoder stages for incomplete multimodal learning. We first design modality-specific and complete multimodal CNN encoders to extract discriminative features within each modality and over all modalities, respectively. A set of convolutional encoders is also designed to perform progressive up-sampling of the latent feature space to produce robust segmentation. We introduce three novel modules in the encoder-decoder architecture to enhance the model’s robustness to incomplete modalities:

- 1) an MFD for knowledge distillation allowing transfer of feature-level knowledge from the multimodal network (teacher) to unimodal networks (students) to provide complementary information to the latter when dealing with missing modalities.
- 2) a set of intra-modal UFEs—computational blocks that combine a transformer and a ConvBlock, for joint learning of both local and global dependency within each modality, and
- 3) a CMF for multimodal feature fusion to build long-range contextual relationships across different modalities.

A. Encoder-Decoder Architecture

Let $M = \{T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12\}$ be the complete set of modalities. We denote the complete multimodal 3D MRI images by $X_M \in \mathbb{R}^{C \times H \times W \times D}$ where H , W is the size of spatial resolution, C is the number of modalities, and D is the number of slices, and the data for each modality by $X_m \in \mathbb{R}^{1 \times H \times W \times D}$, $m \in M$. Given the complete input X_M , we first design a convolutional encoder E_M to extract compact multimodal features $F_M = E_M(X_M)$. The multimodal encoder E_M consists of five feature extraction layers each composed of 3D convolution with $3 \times 3 \times 3$ convolution kernel, instance normalization, and LeakyReLU. Simultaneously, given the unimodal inputs X_m , a set of modality encoders E_m , $m \in M$ are adopted to extract modality-specific features $F_m = E_m(X_m)$ for each modality m , separately. The unimodal encoders E_m have the same network architecture as E_M . Besides, the convolutional decoders have the symmetric architecture of convolutional encoders, similar to the 3D U-Net.

Then, the MFD module distills prior multimodal knowledge to improve the training of unimodal encoders, by aligning the feature map F_M of the multimodal encoder and F_m of the unimodal encoders, to produce F_m^* for each modality. The resulting output F_m^* is mapped to the same space as F_M . To constrain the feature extraction process of the encoders, a multimodal decoder D_M and a set of modality-specific decoders D_m , $m \in M$ which correspondingly have five layers, takes F_M and F_m^* respectively as inputs to obtain the predicted segmentation maps Y_M and Y_m . The distilled unimodal features F_m^* are then fed into the UFE to obtain feature

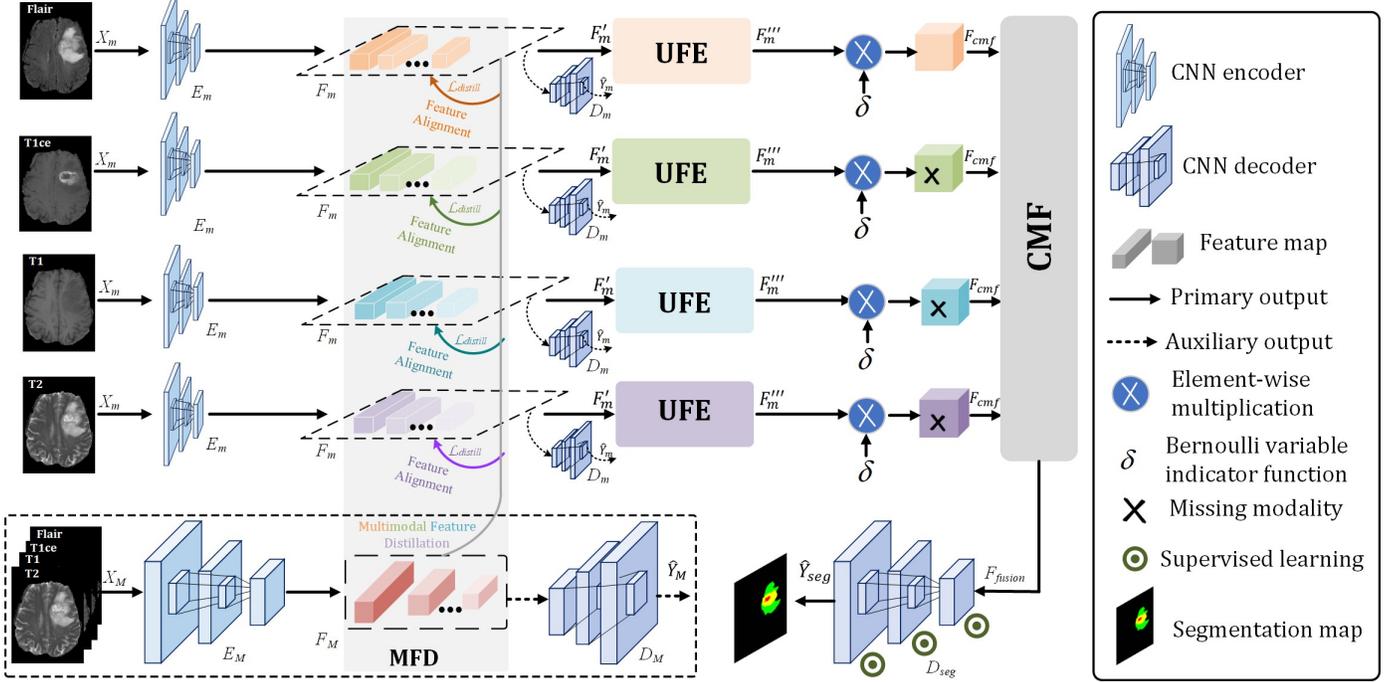


Fig. 2. Overview of the proposed model. The Multimodal Feature Distillation (MFD) module adopts a multimodal encoder network to distill complementary multimodal knowledge into the unimodal encoders. The Unimodal Feature Enhancement (UFE) module extracts intra-modal feature maps. The Cross-Modal Fusion (CMF) Module is then used to aggregate features across modalities. D_{seg} produces the final segmentation result Y_{seg} . The auxiliary outputs Y_m of D_m and Y_M of D_M are used in loss functions along with the segmentation map of the primary output. The description of symbols in the figure is given in the legend box.

maps F_m'' which encodes both the local and global contextual information within each modality. The CMF is then used to aggregate features F_m'' of all available sets of modalities to obtain fused representations F_{fusion} which captures cross-modal long-range dependency even when some modalities are missing. Finally, a decoder D_{seg} is used to obtain the final segmentation result Y_{seg} from the fused features F_{fusion} of incomplete multimodalities.

B. Multimodal Feature Distillation Module

Integration of features learned from multimodal MRIs can provide complementary information for anatomies and lesion areas for accurate brain tumor segmentation. However, features extracted from individual modality encoders are vulnerable to situations of missing modality data, which can degrade the subsequent fused representation and, hence, the segmentation performance. To overcome this issue, we propose an MFD at the encoder stage to transfer complementary knowledge from different modalities from the multimodal teacher encoder to assist the learning of unimodal student encoders in producing features that are robust against missing modalities.

Let F_M^l and F_m^l be the l -th layered feature maps of the multimodal encoder D_M (teacher) and the unimodal encoders D_m (students), respectively. To achieve the multimodal knowledge transfer, the feature maps of the student networks are aligned to the teacher network, by minimizing a feature-level distillation loss (called MFD loss):

$$\mathcal{L}_{MFD} = \sum_{m \in \mathcal{M}} \sum_{l=3}^5 \|F_M^l - F_m^l\|_1 \quad (1)$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm. The unimodal student encoders can learn the multimodal knowledge from layers of the multimodal teacher network, by reducing the discrepancy in the intermediate feature maps with the teacher as in (1). Due to the same architectures of the multimodal and unimodal encoders, the MFD is considered an isomorphic feature distillation. Besides, only the high-level semantic features from the middle layers of the unimodal encoders (i.e., $l = 3, 4, \text{ and } 5$) are involved in the MFD according to (1). This is to maintain the local invariance property of low-level features that are specific to each modality.

C. Unimodal Feature Enhancement Module

For each modality-specific encoder, we further design a UFE module that combines a Transformer and ConvBlocks to extract both local and global contextual information within a modality. The use of Transformers in UFE can overcome the limited receptive field of the generic convolutional encoder and establish a global contextual relationship within a specific modality. Considering that the Transformer has no prior information specific to vision tasks, we further incorporate ConvBlocks as the adapter of the network to offer a better spatial inductive bias into the Transformer. Concretely, the UFE is a hybrid of Transformer and ConvBlocks arranged in a parallel structure. The 3×3 ConvBlock consists of four kinds of layers: convolution, linear, reshape, and GELU [52]. In contrast to ViT which flattens the image into a 1D token sequence, here we retain the 3D structure of the input features before the convolution operation. Given the distilled modality-specific features F_m^i as input, UFE produces feature maps F_m''

for each modality m as follows

$$F_m'' = F_m' \oplus_{m'} \text{MHSA}[\text{LN}(F_m')] \oplus_{m'} \text{ConvBlock}[\text{LN}(F_m')] \quad (2)$$

$$F_m = F_m + \text{FFN}[\text{LN}(F_m)] + \text{ConvBlock}[\text{LN}(F_m)] \quad (3)$$

where $\text{MHSA}(\cdot)$ is the Multi-Head Self-Attention, $\text{FFN}(\cdot)$ is Feed Forward Network, and $\text{LN}(\cdot)$ is Layer Normalization.

In contrast to the prior transformer-based segmentation networks, the proposed CNN-Transformer hybrid network in the unimodal feature extraction and cross-modal fusion can establish long-range dependencies among different modalities via the transformer, balanced with the locality inductive bias from the CNN, which can further improve segmentation.

D. Cross-Modal Fusion Module

We design a CMF module to learn long-range correlations between various modality-specific features, which provide complementary information to improve segmentation. The extracted fused cross-modal features via CMF can represent modality-robust information in the latent space for improving segmentation for settings of missing modalities. The CMF consists of two parallel branches. In contrast to the intra-modal transformers in UFEs, the first branch is an inter-modal Transformer to learn modality-correlated feature representations. The second branch uses the ConvBlock adapter to improve the inductive bias of the inter-modal Transformer in latent space. Specifically, CMF takes as input varying subsets $F_m \otimes \delta_m$ of available modality-specific features $F_m \in \mathbb{R}^{C \times N}$ from UFEs, defined as follows

$$F_{CMF} = \{F_{Flair}''' \otimes \delta_{Flair}, F_{T1ce}'' \otimes \delta_{T1ce}, F_{T1}''' \otimes \delta_{T1}, F_{T2}'' \otimes \delta_{T2}\} \quad (4)$$

where \otimes denotes element-wise multiplication and $\delta_m \in \{0, 1\}$ is the Bernoulli indicator function which represents the presence/absence of a particular modality. Modality features are randomly dropped by setting corresponding $\delta_m = 0$ to stimulate missing modality. This random modality dropout is implemented during model training in order to improve the model's robustness against missing modalities.

The F_{CMF} is then processed by the inter-modal Transformer to build long-range dependency across modalities. For example, the resulting cross-modal feature maps between two modalities α and β are defined as follows

$$F_{\alpha-\beta} = \text{MHSA}(Q_\alpha \cdot K_\beta \cdot V_\beta) = \text{softmax}\left(\frac{Q_\alpha \cdot K_\beta^T}{\sqrt{d_k}}\right) V_\beta \quad (5)$$

where d_k represents queries and keys of dimension, and Q_α , K_β and V_β are formulated as

$$Q_\alpha = \text{LN}(F_\alpha) W^Q \quad (6)$$

$$K_\beta = \text{LN}(F_\beta) W^K \quad (7)$$

$$V_\beta = \text{LN}(F_\beta) W^V \quad (8)$$

The four modality features in F_{CMF} are integrated by the inter-modal Transformer to get modality-correlated features F_{trans} . In parallel, the ConvBlock processes the concatenated features of all modalities $\text{Concat}(F_{CMF})$ to get $F_{conv} =$

$\text{ConvBlock}[\text{Concat}(F_{CMF})]$. The final output of CMF is a fused representation defined by

$$F_{fusion} = F_{trans} \oplus F_{conv} \quad (9)$$

where \oplus denotes element-wise addition. Finally, the fused feature F_{fusion} is fed to the decoder D_{seg} through stepwise upsampling operations to obtain the final segmentation map Y_{seg} .

E. Loss Function

To align the prediction with ground-truth segmentation, the unimodal decoders D_m are optimized based on the Weighted Cross-Entropy (WCE) and Dice loss [53] as follows

$$\mathcal{L}_{um} = \sum_{m \in M} [\mathcal{L}_{WCE}(Y_m, Y_m) + \mathcal{L}_{Dice}(Y_m, Y_m)] \quad (10)$$

where Y_m and Y_m are, respectively, the predicted and ground-truth segmentation labels for each modality m .

The multimodal decoder D_M is optimized by

$$\mathcal{L}_{mm} = \mathcal{L}_{WCE}(Y_M, Y_M) + \mathcal{L}_{Dice}(Y_M, Y_M) \quad (11)$$

where Y_M and Y_M are the predicted and ground truth of complete multimodal images, respectively.

The loss function of the cross-modal decoder D_{seg} is formulated as

$$\mathcal{L}_{seg} = \mathcal{L}_{WCE}(Y_{seg}, Y_M) + \mathcal{L}_{Dice}(Y_{seg}, Y_M) \quad (12)$$

In addition, we adopt a deep supervision strategy to regularize the output features of the first to the fourth layer of the decoder D_{seg} . The layer-wise prediction loss is formulated as

$$\mathcal{L}_{layer} = \sum_{l=1}^4 [\mathcal{L}_{WC}(Y_l, Y_l) + \mathcal{L}_{Dic}(Y_l, Y_l)] \quad (13)$$

where l denotes the index of layer in the decoder D_{seg} .

The overall loss function is defined as

$$\mathcal{L}_{total} = \mathcal{L}_{um} + \mathcal{L}_{mm} + \mathcal{L}_{seg} + \mathcal{L}_{layer} \quad (14)$$

IV. EXPERIMENTS AND RESULTS

We evaluate the effectiveness of our MCTSeg framework for brain tumor segmentation from MRI with missing modalities.

A. Datasets and Evaluation Metrics

1) **Datasets:** We conducted extensive experiments on the 2018 Brain Tumor Segmentation Challenge BraTS2018 and BraTS2020 [54] datasets. The BraTS2018 dataset consists of scan images of 285 patients, and the BraTS2020 dataset contains scan images of 369 patients. Each sample contains MRI scan images of four different modalities: Flair, T1ce, T1, and T2. The ground-truth labels are divided into four categories, including healthy tissue, edema area, necrosis area, and enhancing tumor area. We evaluate segmentation performance on enhanced tumor region (ET), core tumor region (TC), and whole tumor region (WT). WT is composed of an enhancing area, edema area, and necrosis area; TC is composed of an enhancing area and necrosis area; and ET is composed of an enhancing area. The acquired dataset has been preprocessed by registration, skull removal, and resampling to 1 mm^3 resolution.

2) *Evaluation Metrics*: To evaluate the segmentation accuracy quantitatively, we employ the Dice similarity coefficient:

$$Dice_c(\hat{Y}_c, Y_c) = \frac{2\|\hat{Y}_c \cap Y_c\|}{\|\hat{Y}_c\|_1 + \|Y_c\|_1} \quad (15)$$

where the subscript c represents different combinations of complete/incomplete multimodalities. The Dice score aims to evaluate the overlap rate of predicted results and real labels. The range is from 0 to 1, the larger the value of the Dice score, the better the segmentation prediction result.

B. Implementation Details

We used three-fold cross-validation to conduct experiments. During the experiment, we performed Z-score normalization on the input MRI image, randomly cropped the spatial resolution of the input image to a size of $128 \times 128 \times 128$, and then performed random rotation, intensity shift, and mirror flip. Adam optimizer [55] was used to optimize the network parameters, β_1 and β_2 are 0.9 and 0.999 respectively, and the weight decay is $1e^{-5}$. The entire network was trained for 1000 epochs, and the batch size is set to 1. In addition, we adopted a polynomial learning rate strategy similar to [56], with an initial learning rate of $initial_{lr} = 1e^{-4}$, and a current learning rate of $lr = initial_{lr} \times (\frac{1-num}{ma_epoch})^p$ with $p = 0.9$ and num is the epoch number.

C. Quantitative Results

We compare the proposed MCTSeg with state-of-the-art methods on different cases with missing MRI modalities on the BraTS2018 dataset. These competing methods include HeMIS [31], U-HVED [32], RobustSeg [36], D²-Net [7], and mmFormer [11]. For a fair comparison, all experiments use 3-fold cross-validation. As shown in Table I, we use the Dice score as the evaluation index, and the average Dice scores of the proposed model on ET, TC, and WT are 57.44%, 74.37%, and 84.91%, respectively. Compared with mmFormer, our network improves by 1.4% and 1.97% in the TC and WT regions on the mean of 15 different modality deletion cases, and the results show that our model obtains the tumor core and the overall tumor region. Significant improvement, and as can be seen from the Dice scores of WT in Table I, our proposed method achieves the best performance than any other models under 15 different modality missing cases. This can be attributed to the capability of our model to adapt to all tumor regions, and to establish a global semantic relationship well.

To evaluate the robustness of our proposed method, we also conducted experiments on the BraTS2020 dataset. We compared 4 advanced models such as HeMIS, U-HVED, RobustSeg, and RFNet [40]. As shown in Table II, we use Dice Score as the evaluation index, and the average Dice scores of the proposed model on ET, TC, and WT are 64.91%, 81.66%, and 87.24%, respectively. Compared with RFNet, our model improves by 3.44%, 3.43%, and 0.26% on ET, TC, and WT regions, respectively. The results show that our method achieves a significant improvement in tumor core and enhances tumor regions. Compared with pure CNN and Transformer methods, our proposed hybrid architecture can improve the

model’s feature representation fusion and semantic understanding, which leads to more accurate segmentation in MRI brain tumor images with missing modalities.

D. Qualitative Analysis

In Fig. 3, we visualize our model, mmFormer, and RobustSeg in full modalities, missing one modality (Flair + T1 + T2), missing two modalities (T1 + T2), and missing three modalities. The segmentation results in state (T1), and the visualization results show that in most cases, our method can obtain accurate segmentation results. At the same time, we further analyzed the impact of different modes missing. It can be seen from Fig. 3 that the T1 modality plays an important role in the segmentation of the entire tumor. Even if the other three modalities are missing, the T1 modality can obtain more accurate segmentation results. The region is the most sensitive, and when the T1ce modality is missing, the segmentation effect of gangrene and non-enhancing tumor regions is severely degraded, resulting in inaccurate boundary segmentation. When the Flair modality is missing, the segmentation accuracy of peritumoral edema decreases.

In Fig. 4, we show the segmentation results of our model in the absence of 15 different modalities. It is evident that our method can adapt to the absence of different modalities, and can obtain accurate segmentation results even with only one modality.

Our proposed CMF module aims to extract modality-invariant feature representations by establishing and aligning global correlations between modalities. It leverages the attention mechanism of the Transformer to focus on relevant modality-specific features corresponding to tumor regions, compensating for the absence of certain modalities. The comparison results of the qualitative analysis show that different modalities have different sensitivities to different tumor regions, and the CMF module we designed can make full use of the complementary characteristics of different modalities to improve the overall performance of the model.

E. Ablation Study

We conducted a series of ablation studies on our model. As shown in Table III, the proposed MFD, UFE, and CMF modules can significantly improve network performance. For example, after removing the prior knowledge guide module, the performance of the three tumor regions ET, TC, and WT decreased by 1.82%, 1.16%, and 1.73%, respectively, suggesting that the prior knowledge guide module can effectively distill multimodal prior information to unimodality and improve the representativeness of the unimodal features. After removing the UFE module, the performance of the three tumor regions ET, TC, and WT decreased by 2.27%, 1.54%, and 1.89%, respectively, indicating that the UFE module can well model the global relationship within the modality. After removing the CMF module, the performance of the three tumor regions ET, TC, and WT decreased by 1.42%, 0.82%, and 0.84%, respectively. This indicates that the global correlation of different modalities should be effectively aligned through

TABLE I
COMPARISON OF BRAIN TUMOR SEGMENTATION PERFORMANCE BETWEEN MCTSEG AND OTHER STATE-OF-THE-ART METHODS ON THE BRATS2018 DATASET.

M	Flair T1ce T1 T2	◆	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	Avg
		◦	◦	◆	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	
ET	HeMIS [31]	11.78	62.02	10.16	25.63	66.10	32.39	66.22	30.22	67.83	10.71	69.92	68.72	31.07	68.54	70.27	46.10	
	U-HVED [32]	23.80	57.64	8.60	22.82	68.36	24.29	61.11	32.31	67.83	27.96	67.75	68.93	32.34	68.60	69.03	46.76	
	RobustSeg [36]	25.69	67.07	17.29	28.97	70.30	32.01	69.06	33.84	69.71	32.13	70.10	70.88	70.78	70.78	71.13	51.02	
	D ² -Net [7]	8.10	66.30	8.10	16.00	64.80	16.50	70.70	17.40	68.70	9.50	68.30	66.40	19.40	65.70	68.40	42.30	
	mmFormer [11]	39.33	72.60	32.35	43.05	75.05	44.99	74.04	47.52	74.51	42.96	74.75	75.67	47.70	75.47	77.61	59.85	
	MCTSeg (Ours)	34.72	65.74	29.14	37.92	69.81	69.64	68.37	39.86	69.64	39.24	70.09	70.98	40.89	70.24	71.33	57.44	
TC	HeMIS [31]	26.06	65.29	37.39	57.20	71.49	60.92	72.46	57.68	76.64	41.12	78.96	77.53	60.32	76.01	79.48	62.57	
	U-HVED [32]	57.90	59.59	33.90	54.67	75.07	56.26	67.55	62.70	73.92	61.14	75.28	76.75	63.14	77.05	77.71	64.84	
	RobustSeg [36]	53.57	76.83	47.90	57.49	80.62	62.19	78.72	61.16	80.20	60.68	80.33	80.72	81.06	81.06	80.86	69.78	
	D ² -Net [7]	47.30	65.10	16.80	56.70	80.80	63.20	78.20	62.60	80.30	61.60	79.00	80.70	63.70	80.90	80.10	66.50	
	mmFormer [11]	61.21	75.41	56.55	64.20	77.88	69.42	78.59	69.75	78.61	65.91	80.39	79.55	71.52	79.80	85.78	72.97	
	MCTSeg (Ours)	61.05	78.84	59.60	63.97	82.48	82.61	82.31	69.26	82.61	70.83	82.83	82.85	71.87	83.04	82.96	74.37	
WT	HeMIS [31]	52.48	61.53	57.62	80.96	68.99	82.41	68.47	82.95	82.48	64.62	83.94	83.85	83.43	72.31	84.74	74.05	
	U-HVED [32]	84.39	53.62	49.51	79.83	85.93	81.56	64.22	87.58	81.32	85.71	82.32	88.09	88.07	86.72	88.46	79.16	
	RobustSeg [36]	85.69	74.93	70.11	82.24	88.51	84.78	77.18	88.28	85.19	88.24	86.01	89.27	88.73	88.73	89.45	84.39	
	D ² -Net [7]	84.20	42.80	15.50	76.30	87.50	80.10	62.10	87.90	84.10	87.30	80.90	88.80	88.40	87.70	88.80	76.20	
	mmFormer [11]	86.10	72.22	67.52	81.15	87.30	82.20	74.42	87.59	82.99	87.06	82.71	88.14	87.75	87.33	89.64	82.94	
	MCTSeg (Ours)	85.90	75.20	73.92	82.70	88.88	86.69	79.96	89.37	86.69	88.61	87.33	90.14	89.83	89.68	90.31	84.91	

The Dice score (%) is used to evaluate the performance of the model on the three tumor regions of ET, TC, and WT. M represents modalities. Avg indicates the average value of 14 different missing modalities and 1 complete modality. ◦ denotes the modality is missing, ◆ denotes the modality exists. The best results are shown in bold.

TABLE II
COMPARISON OF BRAIN TUMOR SEGMENTATION PERFORMANCE BETWEEN MCTSEG AND OTHER STATE-OF-THE-ART METHODS ON THE BRATS2020 DATASET

Models	Average Dice scores (%)		
	ET	TC	WT
HeMIS [31]	47.73	65.45	75.10
U-HVED [32]	48.55	67.19	81.24
RobustSeg [36]	55.49	73.45	84.17
RFNet [40]	61.47	78.23	86.98
MCTSeg (Ours)	64.91	81.66	87.24

The Dice score (%) is used to evaluate method performance of the three tumor regions of ET, TC, and WT. The best results are shown in bold.

TABLE III
ABLATION STUDY OF MCTSEG ON BRATS2018 DATASET

Method	Average Dice scores (%)		
	ET	TC	WT
MCTSeg	57.44	74.37	84.91
W/O MFD	55.62	73.21	83.18
W/O UFE	55.17	72.83	83.02
W/O CMF	56.02	73.55	84.07

W/O MFD, W/O UFE, and W/O CMF represent removing MFD, UFE, and CMF respectively.

the information interaction between modalities in order to improve the segmentation performance.

We also evaluated the effectiveness of ConvBlock. We examine the effect of the proposed UFE by removing the ConvBlock from UFE and CMF. It can be seen from Table IV that, after removing ConvBlock in UFE, the performance of the three tumor regions of ET, TC, and WT decreased by 0.15%, 0.7%, and 0.82%, respectively. After removing Con-

TABLE IV
ABLATION STUDY ON THE EFFECTIVENESS OF THE CONVBLOCK

Method	Average Dice scores (%)		
	ET	TC	WT
W/O UFE	55.17	72.83	83.02
W/O ConvBlock in UFE	55.02	72.13	82.20
W/O CMF	56.02	73.55	84.07
W/O ConvBlock in CMF	55.79	72.98	83.28

W/O ConvBlock in UFE and W/O ConvBlock in CMF represent removing ConvBlock from UFE and CMF respectively.

vBlock in CMF, the performance of the three tumor regions ET, TC, and WT is reduced by 0.23%, 0.57%, and 0.79%, respectively. This suggests that the introduction of CNN can make up for the insufficient inductive bias of Transformer.

V. CONCLUSION

We developed a novel model that utilizes feature distillation to guide the segmentation of incomplete multimodal brain tumors, resulting in high-precision tumor segmentation. Complementary information distilled across multimodalities and modality-invariant feature representation extracted from each unimodality is effectively fused via novel CNN-Transformer hybrid networks to figure out missing modalities. The experimental evaluation and ablation study demonstrate that the proposed CNN-Transformer hybrid networks play a key role in improving performance on brain tumor segmentation with incomplete MRI modalities by modeling the semantic relationship between global and local information. This overcomes the existing problem in medical image segmentation that it is difficult for convolution to effectively establish global context and the Transformer's inductive bias ability is low.

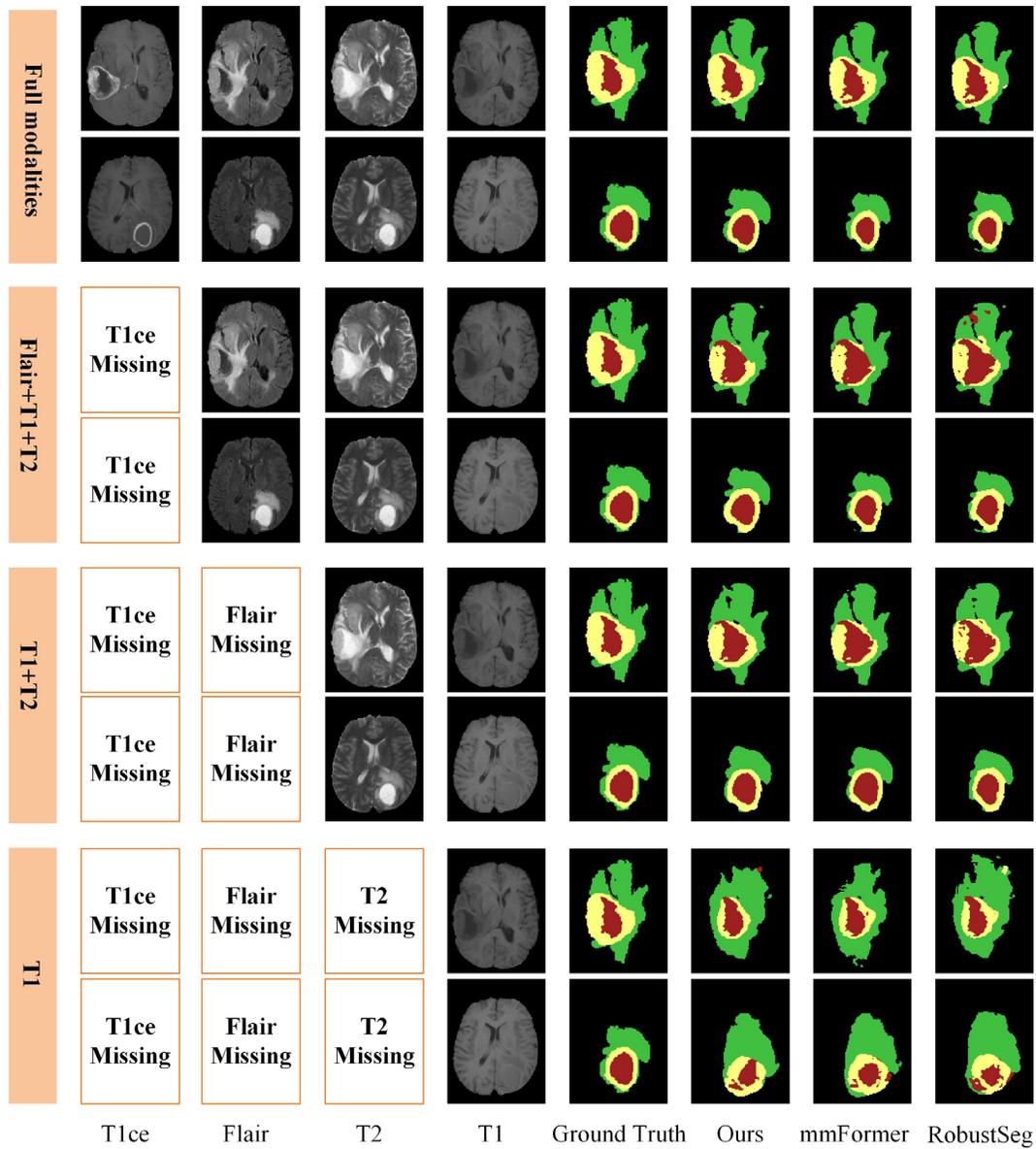


Fig. 3. Visualization of mmFormer [11], RobustSeg [36], and our proposed model on BraTS2018 dataset. There are more segmented areas in images of our proposed model than the other two.

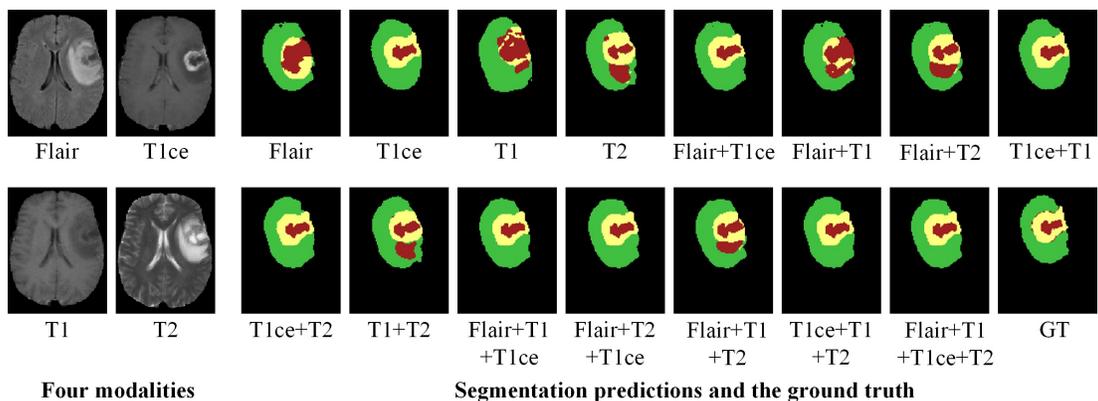


Fig. 4. Qualitative comparison for segmentation results of fifteen different missing modalities. Four different modalities on the left, and segmentation results and ground truth of the proposed model in fifteen different combinations on the right.

REFERENCES

- [1] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes," *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2018, pp. 903–907.
- [2] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. B. Ayed, "Hyperdense-net: A hyper-densely connected cnn for multi-modal image segmentation," *IEEE Trans. Med. Imaging*, vol. 38, no. 5, pp. 1116–1126, May 2019.
- [3] L. Kong and Y. Zhang, "Multi-modal brain tumor segmentation using cascaded 3d u-net," in *Proc. Int. Conf. Image Vis. Comput. (ICIVC)*, 2021, pp. 129–133.
- [4] F. Fang, Y. Yao, T. Zhou, G. Xie, and J. Lu, "Self-supervised multi-modal hybrid fusion network for brain tumor segmentation," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5310–5320, Nov. 2022.
- [5] N. Hashemi, S. Masoudnia, A. Nejad, and M.-R. Nazem-Zadeh, "A memory-efficient deep framework for multi-modal mri-based brain tumor segmentation," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2022, pp. 3749–3752.
- [6] Y. Zhu, S. Wang, R. Lin, and Q. Chen, "Brain tumor segmentation for missing modalities by supplementing missing features," in *Proc. IEEE Int. Conf. Cloud Comput. and Big Data Anal. (ICCCBDA)*, 2022, pp. 652–656.
- [7] Q. Yang, X. Guo, Z. Chen, P. Y. M. Woo, and Y. Yuan, "D²-net: Dual disentanglement network for brain tumor segmentation with missing modalities," *IEEE Trans. Med. Imaging*, vol. 41, no. 10, pp. 2953–2964, Oct. 2022.
- [8] A. Zhao, G. Balakrishnan, F. Durand, J. V. Durand, and A. V. Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 8535–8545.
- [9] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "Transbts: Multimodal brain tumor segmentation using transformer," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2021, pp. 109–119.
- [10] Z. Xing, L. Yu, L. Wan, T. Han, and L. Zhu, "Nestedformer: Nested modality-aware transformer for brain tumor segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2022, pp. 140–150.
- [11] Y. Zhang *et al.*, "Mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2022, pp. 107–117.
- [12] H.-Y. Zhou *et al.*, "Nnformer: Volumetric medical image segmentation via a 3d transformer," *IEEE Trans. Image Process.*, doi: 10.1109/TIP.2023.3293771, Jul. 2023.
- [13] R. Azad, N. Khosravi, M. Dehghanmashadi, J. Cohen-Adad, and D. Merhof, "Medical image segmentation on mri images with missing modalities: A review," arXiv:2203.06217v1 [cs.CV], Mar. 2022.
- [14] D. Shah, A. Barve, B. Vala, and J. Gandhi, "A survey on brain tumor segmentation with missing mri modalities," in *Proc. Int. Conf. Inf. Technol. (InCITE)*, 2023, pp. 209–308.
- [15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
- [16] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2016, pp. 424–432.
- [17] A. Vaswani *et al.*, "Attention is All you Need," in *Proc. Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 6000–6010.
- [18] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [19] J. Chen *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv:2102.04306v1 [cs.CV], Feb. 2021.
- [20] Y. Zhang, H. Liu, and Q. Hu *et al.*, "Transfuse: Fusing Transformers and cnns for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2021, pp. 14–24.
- [21] H. Cao *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 205–218.
- [22] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS) Deep Learn. and Represent. Learn. Workshop*, 2015.
- [23] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [24] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017.
- [25] N. Passalis and A. Tefas, "Learning deep representations with probabilistic knowledge transfer," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 268–284.
- [26] J. Kim, S. Park, and N. Kwak, "Paraphrasing complex network: Network compression via factor transfer," in *Proc. Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 31, 2018, pp. 2765–2774.
- [27] B. Heo, J. Kim, S. Yun, H. Park, N. Kwak, and J. Y. Choi, "A comprehensive overhaul of feature distillation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1921–1930.
- [28] M. Zong *et al.*, "Better teacher better student: Dynamic prior knowledge for knowledge distillation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.
- [29] Y. Tian, D. Krishnan, and P. Isola, "Contrastive representation distillation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2020.
- [30] X. Xing, Z. Chen, M. Zhu, Y. Hou, Z. Gao, and Y. Yuan, "Discrepancy and gradient-guided multi-modal knowledge distillation for pathological glioma grading," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2022, pp. 636–646.
- [31] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "Hemis: Hetero-modal image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2016, pp. 469–477.
- [32] R. Dorent, S. Joutard, M. Modat, S. Ourselin, and T. Vercauteren, "Hetero-modal variational encoder-decoder for joint modality completion and segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2019, pp. 74–82.
- [33] Y. Shen and M. Gao, "Brain tumor segmentation on mri with missing modalities," in *Proc. Int. Conf. Inf. Process. Med. Imaging (IPMI)*, 2019, pp. 417–428.
- [34] Y. Zhang *et al.*, "Modality-aware mutual learning for multi-modal medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2021, pp. 589–599.
- [35] H. Ting and M. Liu, "Multimodal transformer of incomplete mri data for brain tumor segmentation," *IEEE J. Biomed. Health Inform.*, doi: 10.1109/JBHI.2023.3286668, Jun. 2023.
- [36] C. Chen, Q. Dou, Y. Jin, H. Chen, J. Qin, and P.-A. Heng, "Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2019, pp. 447–456.
- [37] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Brain tumor segmentation with missing modalities via latent multi-source correlation representation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2020, pp. 533–541.
- [38] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Latent correlation representation learning for brain tumor segmentation with missing mri modalities," *IEEE Trans. Image Process.*, vol. 30, pp. 4263–4274, Apr. 2021.
- [39] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing mr modalities," *Neurocomputing*, vol. 466, pp. 102–112, Nov. 2021.
- [40] Y. Ding, X. Yu, and Y. Yang, "Rfnnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 3955–3964.
- [41] Y. Wang *et al.*, "Acn: Adversarial co-training network for brain tumor segmentation with missing modalities," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2021, pp. 410–420.
- [42] R. Azad, N. Khosravi, and D. Merhof, "SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities," in *Proc. Med. Imaging Deep Learn. (MIDL)*, PMLR vol. 172, 2022, pp. 48–62.
- [43] H. Liu, D. Wei, D. Lu, J. Sun, L. Wang, and Y. Zheng, "M³AE: Multimodal Representation Learning for Brain Tumor Segmentation

- with Missing Modalities,” in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2023, vol. 37, no. 2, pp. 1657–1665.
- [44] H. Wang, Y. Chen, C. Ma, J. Avery, L. Hull, and G. Carneiro, “Multi-modal learning with missing modality via shared-specific feature Modelling,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 15878–15887.
- [45] M. Hu *et al.*, “Knowledge distillation from multi-modal to mono-modal segmentation networks,” in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2020, pp. 772–781.
- [46] M. Rahimpour, J. Bertels, D. Vandermeulen, F. Maes, K. Goffin, and M. Koole, “Improving t1w mri-based brain tumor segmentation using cross-modal distillation,” in *Proc. SPIE Med. Imaging: Image Process.*, vol. 11596, 2021.
- [47] M. Rahimpour *et al.*, “Cross-modal distillation to improve mri-based brain tumor segmentation with missing mri sequences,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2153–2164, Jul. 2022.
- [48] C. Chen, Q. Dou, Y. Jin, Q. Liu, and P. A. Heng, “Learning with privileged multimodal knowledge for unimodal segmentation,” *IEEE Trans. Med. Imaging*, vol. 41, no. 3, pp. 621–632, Mar. 2022.
- [49] J. Liu, G. Jiao, and Y. Wu, “Siamese self-distillation for brain tumor segmentation with missing modalities,” doi: 10.2139/ssrn.4333683, Feb. 2023.
- [50] Y. Choi, M. A. Al-Masni, K.-J. Jung, R.-E. Yoo, S.-Y. Lee, and D.-H. Kim, “A single stage knowledge distillation network for brain tumor segmentation on limited mr image modalities,” *Comput. Methods Programs. Biomed.*, vol. 240, p. 107644, Oct. 2023.
- [51] S. Karimijafarbigloo, R. Azad, A. Kazerouni, S. Ebadollahi, and D. Merhof, “Mmcformer: Missing modality compensation transformer for brain tumor segmentation,” in *Proc. Med. Imaging Deep Learn. (MIDL)*, 2023.
- [52] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” arXiv:1606.08415v4 [cs.LG], Jul. 2020.
- [53] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecol.*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
- [54] S. Bakas *et al.*, “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge,” arXiv:1811.02629v3 [cs.CV], Apr. 2019.
- [55] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [56] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.