Learning Unified Hyper-Network for Multi-Modal MR Image Synthesis and Tumor Segmentation With Missing Modalities

Heran Yang, Jian Sun, *Member, IEEE*, and Zongben Xu

***Abstract*— Accurate segmentation of brain tumors is of critical importance in clinical assessment and treatment planning, which requires multiple MR modalities providing complementary information. However, due to practical limits, one or more modalities may be missing in real scenarios. To tackle this problem, existing methods need to train multiple networks or a unified but fixed network for various possible missing modality cases, which leads to high computational burdens or sub-optimal performance. In this paper, we propose a unified and adaptive multi-modal MR image synthesis method, and further apply it to tumor segmentation with missing modalities. Based on the decomposition of multi-modal MR images into common and modality-specific features, we design a shared hyper-encoder for embedding each available modality into the feature space, a graph-attention-based fusion block to aggregate the features of available modalities to the fused features, and a shared hyper-decoder for image reconstruction. We also propose an adversarial common feature constraint to enforce the fused features to be in a common space. As for missing modality segmentation, we first conduct the feature-level and image-level completion using our synthesis method and then segment the tumors based on the completed MR images together with the extracted common features. Moreover, we design a hypernet-based modulation module to adaptively utilize the real and synthetic modalities. Experimental results suggest that our method can not only synthesize reasonable multi-modal MR images, but also achieve state-of-the-art performance on brain tumor segmentation with missing modalities.**

***Index Terms*— Multi-modal MR images, missing modality synthesis, tumor segmentation with missing modalities.**

I. Introduction

MALIGNANT brain tumors have become an aggressive and dangerous disease that leads to death worldwide. Accurate segmentation of brain tumors is of great importance for quantitative assessment of tumor progression and surgery treatment planning. Magnetic Resonance (MR) imaging is a commonly-used imaging technique with good soft-tissue contrast to visualize the brain tissue, and the captured multiple contrasts (refer to as modalities in this paper) can provide complementary information for measuring the tumor-induced tissue changes. The multiple modalities in MR imaging include T1-weighted (T1), contrast enhanced T1-weighted (T1c), T2-weighted (T2), Fluid Attenuation Inversion Recovery (FLAIR) images, etc. For example, T2 and FLAIR images highlight the differences between tumor and normal regions, while T1c image emphasizes tumor core boundary [1]. Due to practical limits, e.g., various scanning protocols, motion artifacts, etc., one or more modalities may be missing or unusable.

To tackle this problem, some image synthesis methods [1], [2] have been proposed to impute the missing modalities based on the available modality images of one subject, and then the completed data can be used for segmentation. Moreover, some segmentation methods [3], [4] segment the brain tumors directly from incomplete multi-modal MR images by learning modality-invariant image representations. However, as there are various possible missing modality cases, these methods generally need to train multiple networks or a unified but fixed network for these missing cases, leading to high computational burdens or sub-optimal performance for each case.

In this work, we propose a unified and adaptive multi-modal MR image synthesis network, dubbed *hyper-GAE*, and apply it to tumor segmentation with missing modalities. Our hyper-GAE consists of a shared encoder for embedding each input modality into the feature space, a fusion block for feature aggregation, and a shared decoder for reconstructing output images. Our method is based on the decomposition of multi-modal MR images of a subject into the common features and modality-specific features, respectively modeling the common anatomical structures of subject and the imaging parameters of scanner. Specifically, we first define modality-specific information as one-hot code indicating MR modality, and construct two modality modulators as hypernetwork [5] (i.e., a small network to tune the weights for a larger network) to respectively modulate the encoder and decoder adaptive to input and output modalities. We also design a graph-attention-based fusion block to adaptively interact and fuse the extracted features based on input modalities, and propose an adversarial common feature constraint to constrain these features only contain the modality-common information.

Manuscript received 20 March 2023; revised 8 June 2023; accepted 31 July 2023. Date of publication 4 August 2023; date of current version 30 November 2023. This work was supported in part by the National Key Research and Development Program under Grant 2022YFA1004201; and in part by NSFC under Grant 62206216, Grant 12125104, Grant 12090021, and Grant U20B2075. *(Corresponding author: Jian Sun.)*

The authors are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: hryang@xjtu.edu.cn; jiansun@xjtu.edu.cn; zbxu@xjtu.edu.cn).

Digital Object Identifier 10.1109/TMI.2023.3301934

As for the tumor segmentation with missing modality(ies), instead of directly using completed MR images for segmentation, we utilize the image completion as a regularization constraint, and take the completed MR images together with extracted common features (for missing modality completion) as the inputs of segmentation network. Moreover, we propose a hypernet-based modulation (HBM) module to adaptively use the input real and synthesized modalities, which is a plug-in block and tunes full-modality tumor segmentation network adaptive to the input modality combination. In this work, we plug the HBM module into the two-stage cascaded U-Net (TC-UNet) [6] and construct the cascaded hyper-segmentors (i.e., a segmentor tuned by a modulator) as our segmentation network.

We applied our method to multi-modal brain MR image synthesis and segmentation. Experimental results show that our method can not only impute reasonable multi-modal MR images, but also achieve state-of-the-art performance on brain tumor segmentation with missing modalities. We also conduct the ablation study and verify the effectiveness of our designed modules. Additional experiments evaluate the performance of different segmentation strategies. The results show that our proposed segmentation strategy of using both feature-level and image-level completed information improves the performance, compared with the strategies only using either of them.

A preliminary version of this work was published in [7]. This journal version presents the following extensions. (1) We extend the methodology from one-to-one to many-to-many image synthesis setting, and design a graph-attention-based fusion block and an adversarial common feature constraint. The network is improved from 2D to 3D. (2) We extend the synthesis method to multi-modal brain tumor segmentation with missing modality(ies), and design a segmentation strategy and a hypernet-based modulation module. (3) We conducted comprehensive experiments on both multi-modal MR image synthesis and tumor segmentation with missing modalities, and evaluated the influence of different segmentation strategies.

A. Related Work

1) Missing MR Modality Synthesis: Synthesizing missing modalities from the available ones has attracted increasing interests in recent researches, which can be classified into the following three categories. The *one-to-one synthesis methods* synthesize one modality from another modality [7], [8], [9], [10], [11], [12], [13]. For example, Dar et al. [9] proposed a conditional generative adversarial network (GAN) to synthesize T2w image from a T1w image. However, when there are two or more existing modalities, these methods can not make full use of all available modality information. To overcome this limitation, some *many-to-one synthesis methods* estimate the missing modality by combining information from all available modalities [2], [14], [15], [16]. For instance, Lee et al. [2] proposed a collaborative GAN for missing modality synthesis. Generally, these methods train specific network for each case of missing modality, and they need to train multiple networks for multi-modal image completion with different missing modality cases. Some *many-to-many synthesis methods* leverage the available modalities to impute all missing modalities by a single model [1], [17], [18], [19]. For example, Sharma et al. [18] proposed a multi-input multi-output GAN for generating missing modalities, where the missing inputs are imputed with zero. However, these methods use a fixed network for image synthesis in various possible missing modality cases, which might be insufficient to handle the complexity of diverse missing modality cases. In fact, for N MR modalities, there would be $2^N - 2$ possible missing modality cases in total, which omits the two cases that all N modalities are scanned or missing.

In this work, we propose a unified and adaptive many-to-many synthesis method. Compared with many-to-one methods, our method only needs to train a unified network for imputing multi-modal MR images at arbitrary missing modality cases. Compared with many-to-many methods, our method utilizes a unified and adaptive network with learned parameters adaptive to different missing modality cases.

2) Brain Tumor Segmentation: Accurately segmenting brain tumors from multi-modal MR images is useful for preoperative treatment planning. Various brain tumor segmentation methods were proposed [20], [21]. However, it is often hard to fully collect all high-quality MR modalities, and there are researches focusing on brain tumor segmentation with missing modalities in recent years. Some *feature-based methods* directly segment brain tumors from incomplete multi-modal MR images by extracting modality-invariant image representations, and an extra image reconstruction path is usually used as the regularization [3], [4], [22], [23], [24], [25], [26], [27]. For example, Yang et al. [27] proposed a dual disentanglement network to decompose modality-specific and tumor-specific features for segmentation. Some *synthesis-based methods* first synthesize the missing modalities to construct the complete multi-modal data, which is further used for segmenting the tumors [1], [19]. For example, Shen et al. [1] proposed a multi-domain image completion method, and then extended it for missing modality segmentation. In addition, some *knowledge distillation methods* train the independent model for specific missing case via the knowledge distilled from full modalities [28], [29], [30]. For example, Azad et al. [30] proposed a co-training network with a content and style-matching mechanism to distill the informative features from full modalities. However, these methods need to train a specific pair of student and teacher networks for each missing case (e.g., requiring $2^N - 2$ pairs of networks for N modalities), which is highly computational inefficient.

In this work, we propose a novel multi-modal MR image synthesis method, and then extend it to a tumor segmentation method with missing modalities. Our segmentation method utilizes image completion as an extra constraint, and takes both completed multi-modal MR images and the extracted common features as the inputs of segmentation network, compared with synthesis-based and feature-based methods only using either image-level or feature-level completed information.

II. Problem Formulation

We aim at constructing a multi-modal MR image synthesis model, and extend it to a tumor segmentation model with missing modalities. For T1, T1c, T2 and FLAIR modalities

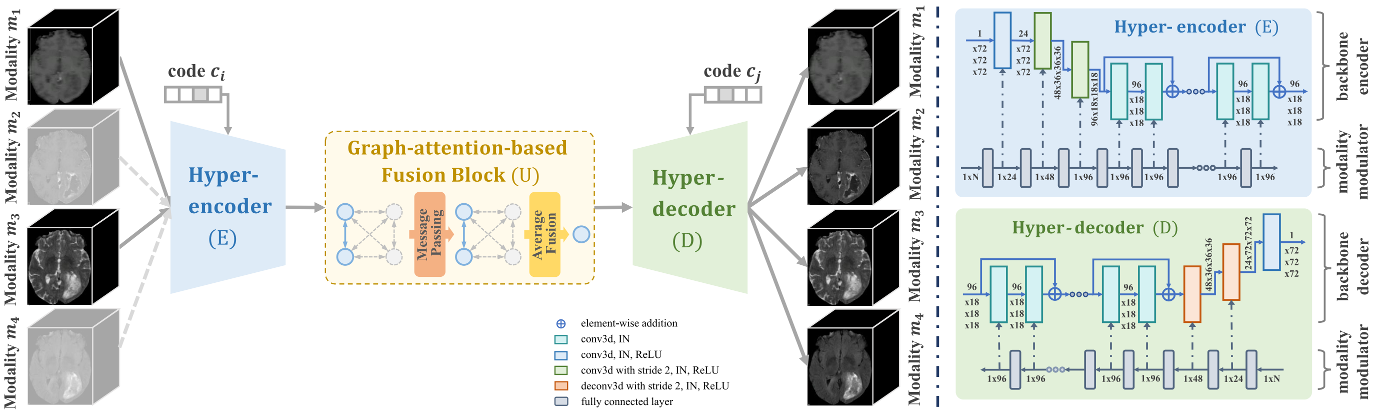


Fig. 1. Illustration of our proposed multi-modal MR image synthesis network (left), and the detailed architectures of our designed hyper-encoder (right top) and hyper-decoder (right bottom). The inputs and dotted lines/nodes in gray represent that these components are missing for this case.

(denoted as $\mathcal{M} = \{m_i\}_{i=1}^{N}$, $N = 4$), we denote $\mathbf{I} = \{I_{m_i}\}_{i=1}^{N}$ as the multi-modal MR images of one subject, where $I_{m_i} \in \mathbb{R}^{H\times W\times D}$ is MR scan of modality m_i. We represent each modality m_i by a one-hot-like *modality code* c_i, with the value of 1 representing corresponding modality in a given list of multiple modalities. Let $y \in \{0,1\}^{H\times W\times D\times K}$ denote segmentation label, with 1 for foreground and 0 for background. K denotes the number of semantic categories. A segmentation model with complete multi-modal MRI data can be defined as

$$\hat{y} = S\big(F(\mathbf{I})\big) = S\big(F(I_{m_1}, I_{m_2}, \cdots, I_{m_N})\big), \tag{1}$$

where $\hat{y}$ denotes the estimated segmentation label. S and F represent the segmentation and feature extraction operators.

When there exists missing modality(ies), one solution is to impute the missing image (e.g., I_{m_i}) by estimating a synthesized $\hat{I}_{m_i}$ based on available modalities (i.e., $\mathcal{M}_a = \{m_j \in \mathcal{M} | I_{m_j} \text{is available}\}$). We define a *modality-availability vector* $v_a \in \{0,1\}^N$ representing the available modality information, with the value of 1 and 0 corresponding to available and missing modalities respectively. After image synthesis process, we can obtain a completed multi-modal MRI data (i.e., $\hat{\mathbf{I}} = (\cdots, I_{m_{i-1}}, \hat{I}_{m_i}, I_{m_{i+1}}, \cdots)$). Then, for a subject with incomplete multi-modal MRI data, our segmentation model employs both the image-level (i.e., the completed MR images $\hat{\mathbf{I}}$) and feature-level (i.e., extracted common features $F(\mathbf{I}_{\mathcal{M}_a})$ from incomplete data $\mathcal{M}_a$) information, formulated as

$$\hat{y} = S\big(F(\mathbf{I}_{\mathcal{M}_a}), \hat{\mathbf{I}}\big). \tag{2}$$

Considering missing modality(ies), there are $2^N - 1$ possible cases if at least one modality is scanned, and it is impractical to train a specific model for each missing modality case. In this work, we propose a multi-modal MR synthesis method for adaptively synthesizing arbitrary missing modality(ies) by a unified network, and further apply it to tumor segmentation with missing modalities. The reminder of this paper is organized as follows. Sections III and IV present our proposed multi-modal MR image synthesis method and its application to missing modality segmentation. Section V presents the experimental results, and conclusions are presented in Section VI.

III. Unified Multi-Modal MR Image Synthesis

Multiple MR imaging modalities provide complementary diagnostic information. Our synthesis method is based on the decomposition of multi-modal MR images of a subject into the common features and modality-specific features. Specifically, we implicitly model the common features using a common feature space, and define the modality-specific information using modality code (i.e., one-hot code indicating MR modality). The encoding process in hyper-encoder computes the common features (i.e., extracted common features) of an input modal image based on this image and its modality-specific features (i.e., input modality code), while the decoding process in hyper-decoder estimates the target modal image based on its common and modality-specific features, i.e., the extracted common features and target modality code.

General Pipeline for Multi-Modal MR Image Synthesis: As in left sub-figure of Fig. 1, given a group of existing modalities of one subject, we first use a shared hyper-encoder to individually extract deep features from each input modality, and then these features are interacted and aggregated to be fused features by a graph-attention-based fusion block. An *adversarial common feature constraint* is designed to constrain the fused features extracted from different input modality groups within a common feature space modeling the common anatomical structures. Finally, these common features are fed into a shared hyper-decoder to reconstruct the output modalities. By this strategy, our network for multi-modal MR image synthesis can adaptively synthesize the missing modality(ies) based on the available modalities using a unified network.

In the remainder of this section, we introduce the designs of hyper-encoder and hyper-decoder, graph-attention-based fusion block, adversarial common feature constraint, and the training loss for multi-modal MR image synthesis.

A. Hyper-Encoder and Hyper-Decoder

We utilize a shared pair of hyper-encoder E and hyper-decoder D to extract deep features from each available modality and reconstruct the multi-modal MR images respectively. To do this, each modality is represented by a modality code, which is used to adaptively tune the encoding and decoding processes. Specifically, our hyper-encoder/hyper-decoder is respectively an encoder/decoder with the parameters modulated by a modality modulator with the modality code as input. In this way, we design a unified encoder-decoder framework flexibly adaptive to input and output modalities, instead of training individual encoder and decoder for each missing case.

1) Architecture of Hyper-en(de)Coder: As shown in right sub-figures of Fig. 1, our hyper-encoder E and hyper-decoder D respectively consist of two subnets, i.e., a backbone encoder/decoder and a modality modulator. The encoder extracts deep features from input modality, while the decoder estimates target modal image from deep features. The backbone encoder and decoder are respectively paired with a modality modulator, achieving modality-adaptive tuning of the parameters of encoder and decoder. The modality modulator is a multilayer perceptron (MLP) with the modality code as input, and each layer in modulator is successively corresponded to an instance normalization (IN) layer in backbone encoder/decoder and tunes its parameters. That is, the output of each layer in modality modulator is fed into the corresponding IN layer in encoder/decoder and works as this layer's affine parameters γ and β. Suppose that $\mu(x)$ and $\sigma(x)$ are the mean and standard deviation of input features x, the output of this IN layer, which is tuned by the modulator, can be written as $\tilde{x} = \gamma\frac{x-\mu(x)}{\sigma(x)} + \beta$.

2) Backbone of Hyper-en(de)Coder: We utilize [31] with several modifications to extend to 3D network. Specifically, the encoder contains two stride-2 3D convolutional layers and six residual blocks to extract features, while the decoder consists of six residual blocks and two stride-2 deconvolutional layers to reconstruct images. The modality modulators tune instance normalization layers in backbone encoder and decoder. Please refer to the right sub-figures in Fig. 1 for more details.

B. Graph-Attention-Based Fusion Block

The graph-attention-based fusion block U adaptively interacts and fuses multi-modal features extracted from different input modalities by hyper-encoder E. Specifically, each graph node corresponds to the features of each modality, and node number is dynamically adapted to the number of available modalities. The graph edges model the connection between each pair of available modalities. Each node borrows the complementary information from the other nodes via an attention-based message passing operation modulated by the modality code and modality-availability vector, and then these updated node features are aggregated by average fusion.

Modality-Modulated Message Passing: Given a graph G with N_0 nodes (i.e., N_0 available modalities $\mathcal{M}_a$), as in Fig. 2, the i-th node features $f_i \in \mathbb{R}^d$ are the deep features extracted from modality m_i per voxel. For each pair of modalities $m_i, m_j \in \mathcal{M}_a$, we define the message passing from modality m_j to m_i as the production of node features f_j and an attention weight w_{ij}. This weight w_{ij} aims at learning to attend on the complementary node features and can be computed as

$$w_{ij} = \Psi([f_i', f_j', v_a]; \theta_w), \tag{3}$$

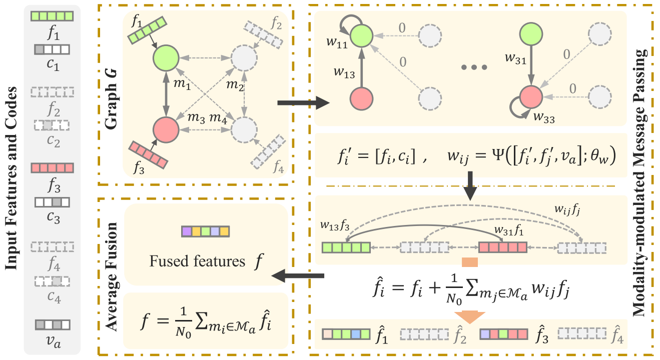


Fig. 2. Illustration of our graph-attention-based fusion block. The dotted lines/nodes in gray represent that these components are missing.

where $f_i' \in \mathbb{R}^{d+N}$ is deep features f_i expanded with modality code c_i, i.e., $f_i' = [f_i, c_i]$, and $[\cdot, \cdot]$ denotes the concatenation of two vectors. $v_a \in \{0, 1\}^N$ is the modality-availability vector representing the available modality information. Ψ is a function with parameters θ_w aiming to estimate the attention weight between each pair of modalities in G, and we design it as a two-layer MLP with d units and tanh function in the final layer. We further collect the messages of modality m_i borrowed from all available modalities $\mathcal{M}_a$, and then the updated features $\hat{f}_i$ of modality m_i can be defined as

$$\hat{f}_i = f_i + \frac{1}{N_0}\sum_{m_j \in \mathcal{M}_a} w_{ij} f_j. \tag{4}$$

Finally, we aggregate these features into the fused features f by average fusion, i.e., $f = \frac{1}{N_0}\sum_{m_i \in \mathcal{M}_a} \hat{f}_i$. These fused features would be utilized to reconstruct the target modal images by hyper-decoder D, as shown in Fig. 1.

C. Adversarial Common Feature Constraint

We propose an adversarial common feature (ACF) constraint to enforce the extracted fused features (by hyper-encoder E and graph-attention-based fusion block U) to be in a common space shared by different missing modality cases for each subject, implicitly modeling the modality-common anatomical structures of the subject. To do this, we introduce an extra classifier C for predicting each modality is available or not when extracting the fused features, and our ACF constraint would adversarially force the hyper-encoder E and fusion block U producing the fused features that can not be correctly classified by classifier C, i.e., within a common feature space.

The classifier C consists of a gradient reversal layer (GRL) [32] and four 1 × 1 × 1 convolutional layers to voxel-wisely predict the modality probabilities. With the fused features (by E and U) as input, the classifier C predicts a N-length vector per voxel, and each element in this vector represents the probability that the corresponding input modality is available or not when extracting these fused features. The GRL in C is a layer acting as an identity transform during forward propagation but flipping the gradient sign during backpropagation, which achieves the adversarial training between C and E, U. Our ACF constraint is defined over the fused features $F_{\mathcal{M}}$ and $F_{\mathcal{M}_a}$ respectively extracted from the full modalities $\mathcal{M}$ and available modalities $\mathcal{M}_a$, and constrains them to be correctly

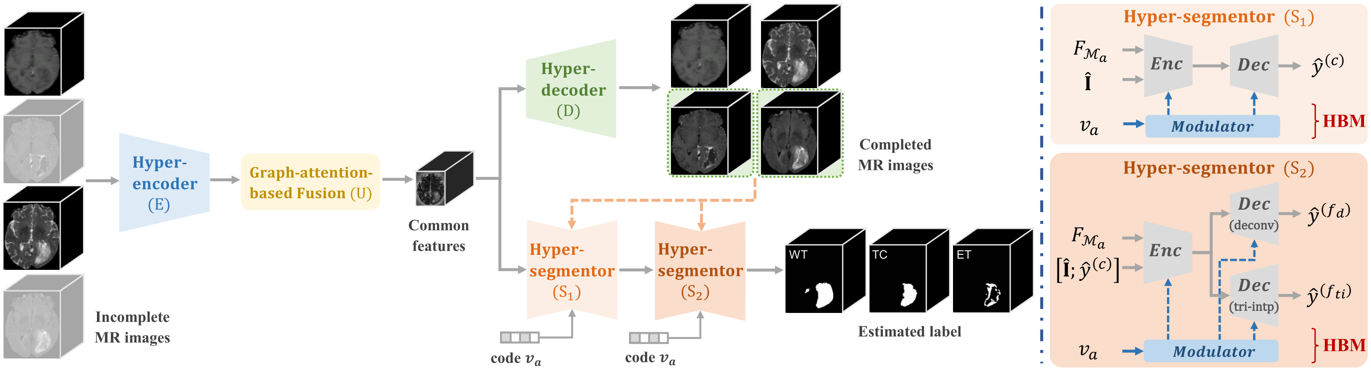


Fig. 3. Illustration of our segmentation method based on hyper-GAE (left), and our hyper-segmentors S_1 (right top) and S_2 (right bottom).

classified by the classifier C, which can be defined as

$$\mathcal{L}_{ACF}(E,U,C) = L_{BCE}(C(F_{\mathcal{M}}),\mathbf{1}) + L_{BCE}(C(F_{\mathcal{M}_a}), v_a), \tag{5}$$

where L_{BCE} denote the binary cross entropy loss. $\mathbf{1}$ and v_a are the all-one vector and modality-availability vector. Due to the GRL in C, this constraint adversarially enforces classifier C is not able to distinguish the modalities of these fuse features, and thus forces hyper-encoder E and fusion block U to produce the fused features within a common feature space.

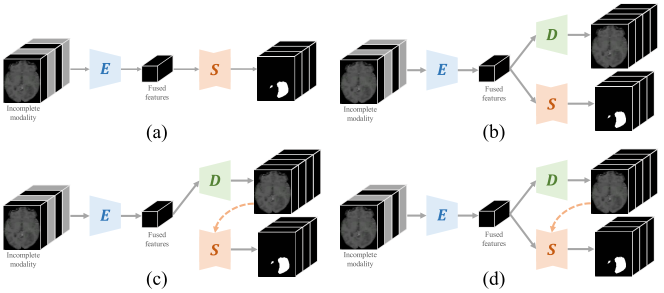


Fig. 4. Illustration of different segmentation strategies, including three possible strategies in (a)-(c) and our proposed strategy in (d).

D. Training Loss for MR Image Synthesis

Our training loss for multi-modal MR image synthesis includes the reconstruction loss and adversarial common feature (ACF) constraint. As our ACF constraint was presented in Section III-C, the reconstruction loss is introduced below.

1) Reconstruction Loss: A reconstruction loss is to enforce the synthetic multi-modal MR images $\{\hat{I}_{m_i}\}_{i=1}^{N}$ to be identical to the ground-truth images $\{I_{m_i}\}_{i=1}^{N}$, and it is defined as

$$\mathcal{L}_R(E,U,D) = \frac{1}{N}\sum_{i=1}^{N} \|\hat{I}_{m_i} - I_{m_i}\|_1, \tag{6}$$

where $\|\cdot\|_1$ denotes the L_1 norm.

2) Total Training Loss: The total training loss is defined as

$$\mathcal{L}_{syn}(E,U,D,C) = \mathcal{L}_R + \lambda_1 \mathcal{L}_{ACF}, \tag{7}$$

where λ_1 controls the relative importance of the loss terms.

IV. Missing-Modality MR Image Segmentation

Our multi-modal MR image synthesis method in Section III can be applied to tumor segmentation with missing modalities. In our approach, we not only utilize image completion task as a regularization and the completed MR images as inputs of segmentation, but also employ the extracted common features during completion as extra inputs for segmentation.

General Pipeline for MR Image Segmentation With Missing Modalities: As shown in left sub-figure of Fig. 3, we utilize our hyper-GAE to extract the fused common features (by E and U) and reconstruct multi-modal MR images (by D), and then both the features and completed multi-modal MR images (consisting of available and reconstructed images) are fed into the segmentation network to estimate tumor labels. As there are two types of input images, i.e., real or synthetic, we design a novel hypernet-based modulation (HBM) module to adaptively utilize these images, which can be plugged in full-modality tumor segmentation network and tune its parameters adaptive to the input real and synthetic modality combinations. In this work, we plug our HBM module into the two-stage cascaded U-Net (TC-UNet) [6] as the segmentation network, resulting in the two-stage cascaded *hyper-segmentors* (i.e., a backbone segmentor modulated by a modulator).

Comparison With Variants of Segmentation Strategies: In Fig. 4, we show the possible different segmentation strategies with missing modalities. Figures 4(a) and (b) show the strategies that extract features and segment from incomplete modalities respectively without and with image completion task as a constraint. In Fig. 4(c), the strategy completes the multi-modal images first and then segments using the completed images. Compared with these strategies, our proposed method, as in Fig. 4(d), employs the image completion as an extra constraint and utilizes both the extracted features and completed images for segmentation. We experimentally verify the effectiveness of our proposed strategy in Section V-E. Next, we introduce the designs of cascaded hyper-segmentors and training loss.

A. Cascaded Hyper-Segmentors

We propose the cascaded hyper-segmentors as the segmentation network in our segmentation framework, as shown in the left sub-figure of Fig. 3. Specifically, we design a novel hypernet-based modulation (HBM) module, which adds an extra modulator (i.e., a small MLP) to a tumor segmentation network (dubbed backbone segmentor) for modulating its parameters. The cascaded hyper-segmentors are constructed by plugging the HBM module into the TC-UNet.

1) Architecture of Hyper-Segmentors: Our cascaded hyper-segmentors, based on segmentors in [6], contain two successive hyper-segmentors S_1 and S_2, each of which is a backbone segmentor with parameters modulated by a modulator, as in right sub-figures of Fig. 3. The first hyper-segmentor S_1 (i.e., a backbone U-Net modulated by a modulator) is used to predict a coarse segmentation map $\hat{y}^{(c)} \in [0, 1]^{H\times W\times D\times K}$:

$$\hat{y}^{(c)} = S_1(F_{\mathcal{M}_a}, \hat{\mathbf{I}}, v_a), \tag{8}$$

where $F_{\mathcal{M}_a}$ is the common features extracted from available modalities $\mathcal{M}_a$, and $\hat{\mathbf{I}} \in \mathbb{R}^{H\times W\times D\times N}$ is completed multi-modal images. Features $F_{\mathcal{M}_a}$ are fed into backbone segmentor via embedding by a convolutional layer and concatenated with deep features at same resolution level in encoding process. v_a is modality-availability vector and fed into the modulator. Then, both common features $F_{\mathcal{M}_a}$ and the concatenation of completed images $\hat{\mathbf{I}}$ and estimated coarse label $\hat{y}^{(c)}$ are fed into the second hyper-segmentor S_2 for predicting a more accurate segmentation map. As shown in right bottom sub-figure of Fig. 3, the backbone segmentor in S_2 is a modified U-Net containing two decoding paths, the structures of which are the same except for one using deconvolution and the other using trilinear interpolation. The estimated fine segmentation maps $\hat{y}^{(fd)}$ and $\hat{y}^{(fti)}$ (from two decoding paths) are written as

$$\hat{y}^{(fd)}, \hat{y}^{(fti)} = S_2(F_{\mathcal{M}_a}, [\hat{\mathbf{I}}; \hat{y}^{(c)}], v_a). \tag{9}$$

As in [6], $\hat{y}^{(fti)}$ is used only during training for regularization, and $\hat{y}^{(fd)}$ is utilized to produce the final segmentation with a threshold of 0.5 during testing. Please refer to [6] for more details on backbone segmentors.

2) Hypernet-Based Modulation Module: This paragraph presents the architecture of modulators in hyper-segmentors S_1 and S_2. Our HBM module, as a plug-in block, is built on top of a backbone segmentor, and adds an extra modulator to the segmentor, as shown in right sub-figures of Fig. 3. The modulator is a MLP with modality-availability vector v_a as input, and each layer in modulator links to a specific layer in backbone segmentor and tunes its parameters. Specifically, two modulators in hyper-segmentors S_1 and S_2 respectively consist of 8 and 11 fully connected (FC) layers, which successively modify the convolutional layer for embedding common features and the last convolutional layers at each resolution level in two backbone U-Net of S_1 and S_2 (i.e., modifying 8 layers for S_1 and 11 layers for S_2). Note that the modulator in S_2 has a "Y" shape, with three brunches respectively corresponding to an encoding path and two decoding paths in backbone U-Net of S_2. The first half of this modulator contains 5 FC layers tuning the encoding path, and then it is bifurcated into two brunches, each of which contains 3 FC layers and tunes a decoding path. The filter scaling strategy [33] is used for parameter modulation. That is, for a convolutional layer with d filters $\{\kappa_i\}_{i=1}^{d}$ in backbone segmentor, the corresponding layer in modulator contains d hidden nodes outputting d scalars $\{\alpha_i\}_{i=1}^{d}$, and modifies each filter κ_i as $\kappa_i' = \alpha_i\kappa_i$, i.e., the scalar multiplication between α_i and κ_i.

B. Training Loss for MR Image Segmentation

Our training loss for segmentation includes adversarial common feature constraint, reconstruction loss, and segmentation loss. The former two terms were introduced in Sections III-C and III-D, and the segmentation loss is introduced below.

1) Segmentation Loss: As introduced in Section IV-A, our method produces three different segmentation maps, including $\hat{y}^{(c)}$ of hyper-segmentor S_1 and $\hat{y}^{(fd)}, \hat{y}^{(fti)}$ of hyper-segmentor S_2. Therefore, the segmentation loss is defined between the ground-truth label y and these three estimated segmentation maps respectively, which can be defined as

$$\mathcal{L}_S(E, U, D, S_1, S_2) = \frac{1}{3K}\sum_{k=1}^{K}\Big(L_{SD}(\hat{y}_k^{(c)}, y_k) + L_{SD}(\hat{y}_k^{(fd)}, y_k) + L_{SD}(\hat{y}_k^{(fi)}, y_k)\Big), \tag{10}$$

where L_{SD} is the soft Dice loss [34]. K denotes the number of semantic categories, which is equal to 3 in this work. $y_k \in \{0, 1\}^{H\times W\times D}$ denotes the ground-truth label of k-th category.

2) Total Training Loss: The total training loss for brain tumor segmentation with missing modalities can be defined as

$$\mathcal{L}_{seg}(E, U, D, C, S_1, S_2) = \mathcal{L}_S + \lambda_1\mathcal{L}_{ACF} + \lambda_2\mathcal{L}_R, \tag{11}$$

where the weights λ_1 and λ_2 control the relative importance of the loss terms. The network parameters are learned by an end-to-end training procedure by minimizing the loss $\mathcal{L}_{seg}$.

V. Experiments

In this section, we evaluate our proposed method on the multi-modal MR image synthesis and brain tumor segmentation with missing modalities. In particular, we compare our method with the other state-of-the-art synthesis and segmentation methods, and evaluate the impacts of the segmentation strategy and our hypernet-based modulation module on performance. A paired two-sided Wilcoxon signed-rank test without false discovery rate correction is conducted to compare the performance, and we report the p-values where the performance difference is statistically significant ($p < .05$).

A. Dataset

1) BraTS 2019 Dataset: The MICCAI 2019 Multimodal Brain Tumor Segmentation Challenge (BraTS 2019) dataset is used to evaluate our method. The samples in this dataset are ample multi-institutional routine clinically-acquired preoperative multi-modal MR scans of subjects with glioblastoma or lower grade glioma. Each subject contains four modality images (i.e., T1, T1c, T2 and FLAIR), and ground-truth tumor labels annotated by experts. Annotations comprise the enhancing tumor (ET), peritumoral edema (ED), necrotic and non-enhancing tumor core (NCR/NET), which are grouped into three sub-regions for evaluation in this challenge: (1) the whole tumor (WT), containing all tumor tissues; (2) the tumor core (TC), consisting of ET and NCR/NET; (3) the enhancing tumor (ET).

This dataset includes 335 training subjects, 125 validation subjects, and 166 test subjects, and the tumor labels of training subjects are provided. The experiments are performed over 335 training subjects, which are randomly divided into a training set of 218 subjects, a validation set of 6 subjects for model selection, and a test set of 111 subjects for evaluation.

2) BraTS 2018 Dataset: To compare with more state-of-the-art segmentation methods, we also test our method on MICCAI 2018 Brain Tumor Segmentation (BraTS 2018) dataset. This dataset contains 285 training subjects with ground-truth labels, which are split into 199, 29 and 57 subjects for training, validation and test using the same split list as in [3], and we conduct 3-fold cross validation to evaluate the performance.

3) Pre-Processing: The data has been pre-processed by organizers, i.e., co-registered to the same anatomical template, interpolated to the same resolution ($1mm^3$) and skull-stripped. Additionally, we conduct extra pre-processing steps, including N4 correction [35], normalizing white matter peak by fuzzy C-means [36], and cutting out black background area outside brain (as in [3]). The maximal intensities of four modalities are 4000, 6000, 10000 and 7000 (arbitrary units), all of which were uniformly linearly normalized to $[-1, 1]$. The pre-processed dataset is utilized for all the compared methods.

4) Evaluation Metrics: To evaluate the *synthesis performance*, we use the mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) [37] between ground-truth MR volume and the synthetic one. For $N = 4$ modalities, there are $C_4^1 + C_4^2 + C_4^3 = 14$ missing modality cases in total. For each missing case, we impute all missing modalities based on the available ones, and the final accuracies for this case are averaged over all these synthetic volumes and all test subjects. As for evaluating *segmentation performance*, for each input modality combination out of all 15 possible cases, we compute the 3D Dice scores and 95% Hausdorff distance (HD95) [38] between the ground-truth and estimated segmentation labels respectively of whole tumor (WT), tumor core (TC), and enhancing tumor (ET). The final segmentation accuracies for each missing case are defined as the averaged Dice and HD95 scores over all subjects in the test set.

B. Implementation Details

Our implementation is based on TensorFlow, and all experiments are performed on a GeForce RTX 3090 GPU (24GB). The networks are optimized in 1200 epochs using an Adam optimizer solver [39] with a learning rate of 0.0002 and a batch size of 2. The weights λ_1 and λ_2 are set to 0.001 and 0.8 empirically. No learning rate decay or weight decay strategy is utilized during training. Both our synthesis and segmentation models are trained in an end-to-end manner. During training, the input volumes are randomly cropped to $72\times72\times72$ voxels, which are then augmented with random flipping and intensity shifts as in [6]. At test phase, following [24], we feed the $72\times72\times72$ patches sliding on test volumes with 50% overlaps into the networks, and we additionally flip these patches in 8 directions for segmentation network. The final estimation is obtained by fusing the predictions of these patches. As in [24], when the number of voxels predicted as ET is too small (i.e., less than 300), we employ a post-processing step and replace the enhancing tumor with necrosis to reduce the false alarm of ET, which is conducted for all compared methods for fairness.

In practice, our synthesis method in Section III requires about 8 GB of GPU memory for training with 6.49M trainable parameters and 338.80G FLOPs (estimated using Tensorflow), and each step of parameter update takes about 0.54s. At test phase, it takes about 4.2s to synthesize all four MR modality volumes of one subject. Besides, our segmentation method in Section IV requires about 16 GB of GPU memory for training with 32.49M parameters and 675.81G FLOPs, and each update takes about 1.1s. At test phase, it takes about 51.5s for our method with 8-direction flips to segment the brain tumors for a subject in each missing modality case. Codes are available at https://github.com/HeranYang/hyper-GAE.

C. Multi-Modal MR Image Synthesis

We conduct comparisons between our method and the other state-of-the-art synthesis methods on the BraTS 2019 dataset, including the hypernet-based generative adversarial network (HyperGAN) [7], collaborative generative adversarial network (CollaGAN) [2], and multi-modal generative adversarial network (MM-GAN) [18]. HyperGAN designs a pair of encoder and decoder for unpaired multi-contrast MR image translation. CollaGAN generates one modality based on the remaining modalities by a single generator and discriminator. MM-GAN utilizes a multi-input, multi-output network to synthesize the missing modalities. As HyperGAN and CollaGAN are respectively one-to-one and many-to-one methods, we are only able to report the accuracies in 4 out of 14 missing cases. In addition, as these two methods are unsupervised, we modify them into the supervised versions by replacing the adversarial loss by a L_1 reconstruction loss for fairness of comparison.

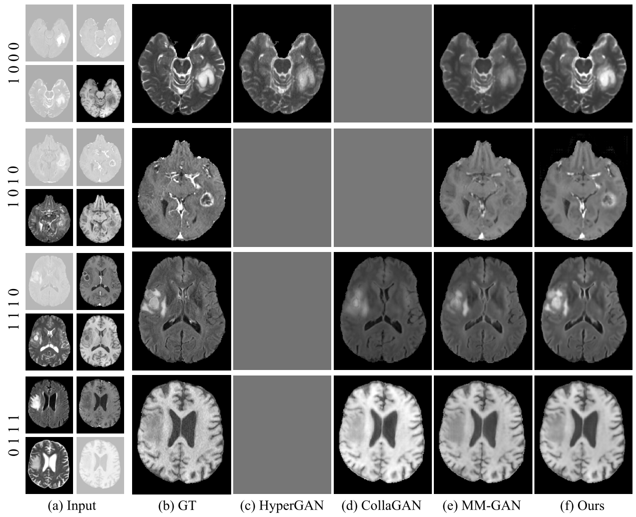
The results of HyperGAN, CollaGAN and MM-GAN are generated by the published codes[1],[2],[3]. To verify the effectiveness of each module in our method, two ablated versions, i.e., our method without adversarial common feature constraint ("ours (w/o ACF)") and our method without graph-attention-based fusion block and adversarial common feature constraint ("ours (w/o GF&ACF)"), are also included in comparison.

Table I reports the accuracies of different synthesis methods. The results show that our method ("Ours") works better than the methods of HyperGAN, CollaGAN and MM-GAN on synthesizing multi-modal MR images in averaged MAE, PSNR and SSIM ($p < .001$). Specifically, our method achieves the best MAE scores in 11 out of 14 cases, and the highest PSNR and SSIM scores in 10 out of 14 cases. In addition, our baseline ("ours (w/o GF&ACF)") performs better than compared methods ($p < .001$) and produces 0.0090/34.25/0.921 in averaged MAE, PSNR and SSIM, justifying the effectiveness of our proposed synthesis frame-

[1] https://github.com/HeranYang/Hyper-GAN
[2] https://github.com/jongcye/CollaGAN_MRI
[3] https://github.com/trane293/mm-gan

TABLE I
MULTI-MODAL MR IMAGE SYNTHESIS ACCURACIES OF DIFFERENT METHODS IN DIFFERENT MISSING CASES ON BRATS 2019 DATASET

Modality	**HyperGAN** [7]	**CollaGAN** [2]	**MM-GAN** [18]	**Ours**	**Ours (w/o ACF)**	**Ours (w/o GF&ACF)**
[T1 T1c T2 F]	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM
[0 0 1 0]	0.0101 / 33.19 / 0.909	- / - / -	0.0095 / **34.43** / **0.921**	**0.0094** / 33.78 / 0.915	0.0096 / 33.59 / 0.911	0.0100 / 33.27 / 0.910
[0 1 0 0]	0.0101 / 33.55 / 0.905	- / - / -	0.0136 / 31.79 / 0.899	**0.0094** / **34.40** / **0.918**	0.0096 / 34.13 / 0.916	0.0102 / 33.66 / 0.907
[1 0 0 0]	0.0096 / 33.26 / 0.910	- / - / -	0.0116 / 33.11 / 0.912	**0.0090** / **34.04** / **0.920**	0.0092 / 33.79 / 0.916	0.0094 / 33.61 / 0.914
[0 0 0 1]	0.0109 / 32.56 / 0.894	- / - / -	0.0130 / 32.43 / 0.898	**0.0102** / **33.32** / **0.904**	**0.0102** / 33.08 / **0.904**	0.0109 / 32.71 / 0.896
[0 1 1 0]	- / - / -	- / - / -	0.0102 / 34.27 / 0.926	**0.0083** / **35.58** / **0.934**	0.0084 / 35.39 / 0.930	0.0088 / 35.01 / 0.928
[1 1 0 0]	- / - / -	- / - / -	0.0144 / 30.96 / 0.897	**0.0090** / **34.41** / **0.925**	0.0093 / 34.17 / 0.922	0.0094 / 33.97 / 0.918
[1 0 0 1]	- / - / -	- / - / -	0.0116 / 33.47 / 0.924	**0.0079** / **34.71** / **0.932**	0.0082 / 34.41 / 0.929	0.0084 / 34.18 / 0.926
[1 0 1 0]	- / - / -	- / - / -	**0.0078** / **36.07** / **0.932**	0.0081 / 34.88 / 0.926	0.0083 / 34.69 / 0.922	0.0084 / 34.48 / 0.920
[0 0 1 1]	- / - / -	- / - / -	**0.0083** / **35.79** / **0.940**	0.0087 / 34.51 / 0.927	0.0089 / 34.28 / 0.925	0.0092 / 34.02 / 0.923
[0 1 0 1]	- / - / -	- / - / -	0.0141 / 31.71 / 0.913	**0.0082** / **35.51** / **0.939**	0.0084 / 35.18 / 0.936	0.0088 / 34.81 / 0.931
[1 1 0 1]	- / - / -	0.0132 / 32.25 / 0.907	0.0173 / 29.08 / 0.901	**0.0079** / **35.18** / **0.941**	0.0082 / 34.83 / 0.939	0.0084 / 34.53 / 0.934
[1 0 1 1]	- / - / -	0.0129 / 31.44 / 0.905	**0.0054** / **38.38** / **0.953**	0.0074 / 34.92 / 0.930	0.0076 / 34.62 / 0.927	0.0078 / 34.43 / 0.924
[0 1 1 1]	- / - / -	0.0118 / 33.60 / 0.932	0.0098 / 35.10 / 0.947	**0.0076** / **36.63** / **0.950**	0.0078 / 36.32 / 0.947	0.0081 / 36.02 / 0.946
[1 1 1 0]	- / - / -	0.0147 / 31.64 / 0.881	0.0102 / 33.86 / 0.912	**0.0085** / **35.10** / **0.924**	0.0087 / 35.01 / 0.920	0.0089 / 34.81 / 0.918
Average	0.0102 / 33.14 / 0.905	0.0131 / 32.23 / 0.906	0.0112 / 33.60 / 0.920	**0.0085** / **34.78** / **0.927**	0.0087 / 34.53 / 0.925	0.0090 / 34.25 / 0.921



(a) Input (b) GT (c) HyperGAN (d) CollaGAN (e) MM-GAN (f) Ours

Fig. 5. Visual comparison of multi-modal MR image synthesis results by different methods in different missing cases on BraTS 2019 dataset. The empty image denotes the method is not applicable to this case.

work. Moreover, an extra graph-attention-based fusion block ("ours (w/o ACF)") further improves the results ($p < .001$) and obtains 0.0087/34.53/0.925. Our whole model performs best ($p < .001$) and achieves 0.0085/34.78/0.927. These results indicate the effectiveness of our graph-attention-based fusion block and adversarial common feature constraint. Figure 5 shows the imputed MR images in four different missing cases. The synthetic MR images by our method not only have better contrast but also maintain the tumor regions better than compared methods.

Synthesis Performance Analysis: As shown in Table I, our method works consistently well on multi-modal MR image synthesis in different missing modality cases. The compared GAN-based method MM-GAN achieves 0.0054/38.38/0.953 in MAE, PSNR and SSIM for imputing T1c image based on the other three modalities, but obtains unsatisfactory results of 0.0173/29.08/0.901 for imputing T2, compared with 0.0074/34.92/0.930 and 0.0079/35.18/0.941 of ours. Moreover, our method achieves 0.0095/33.89/0.914 in MAE, PSNR and SSIM (averaged over 4 missing cases) when only one modality exists, and the results are improved to 0.0084/34.93/0.931 and 0.0078/35.46/0.936 (averaged over 6 and 4 cases) when two and three modalities are available. The effectiveness of our method maybe because it is based on the hyper-encoder/hyper-decoder adaptive to different missing modality cases, and also models the common feature space of multiple modalities.

D. Missing-Modality MR Image Segmentation

We compare different methods for brain tumor segmentation with missing modalities on the BraTS 2019 dataset, including the UNet-based hetero-modal variational encoder-decoder (U-HVED) [3], robust multi-modal segmentation method (RobustSeg) [24], as well as the MM-GAN [18]. U-HVED uses multi-modal variational auto-encoders to embed all modalities into a shared representation for segmentation. RobustSeg designs a multi-modal segmentation network with feature disentanglement and gated fusion. MM-GAN imputes the missing modalities and then utilizes a segmentation network to segment the completed images, which is a TC-UNet in this experiment.

The results of U-HVED, RobustSeg and MM-GAN are generated by the published codes[3,4,5]. As our proposed hypernet-based modulation (HBM) module in Section IV-A can extend full-modality tumor segmentation network to missing-modality version, we also compare with a modified TC-UNet using HBM module ("TC-UNet (w/ HBM)") to verify the effectiveness of our completion and segmentation framework. Two ablated versions of our method (i.e., "ours (w/o ACF)" and "ours (w/o GF&ACF)") are also included in this comparison.

Table II reports the test accuracies on BraTS 2019 dataset. The results show that our method ("Ours") works better than

[4] https://github.com/ReubenDo/U-HVED
[5] https://github.com/cchen-cc/Robust-Msegn

TABLE II

BRAIN TUMOR SEGMENTATION ACCURACIES OF DIFFERENT METHODS IN DIFFERENT MISSING MODALITY CASES ON BRATS 2019 DATASET

Modality	MM-GAN		U-HVED		RobustSeg		TC-UNet (w/ HBM)		Ours		Ours (w/o ACF)		Ours (w/o GF&ACF)	
	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95
[T1 T1c T2 F]	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET
[0 0 1 0]	72.90 / 33.87 / 8.30	14.05 / 22.29 / 23.11	79.34 / 57.16 / 30.21	20.06 / 21.45 / 18.10	80.99 / 62.13 / 35.79	20.21 / 22.09 / 17.25	84.47 / 73.57 / 49.49	10.50 / 11.99 / 9.27	**86.17** / **74.40** / **50.97**	9.08 / **9.60** / **9.00**	86.03 / 72.01 / 50.62	**8.68** / 10.63 / 10.10	84.59 / 69.05 / 48.27	9.09 / 10.73 / 10.04
[0 1 0 0]	65.61 / 75.47 / 73.17	24.41 / 13.65 / 8.85	67.00 / 73.15 / 65.13	21.71 / 23.69 / 15.78	70.19 / 76.90 / 68.57	28.81 / 29.39 / 22.56	76.23 / **85.14** / 78.68	13.89 / 9.84 / 4.96	**77.51** / 84.21 / 78.99	12.38 / 7.43 / 4.50	75.96 / 82.42 / **80.22**	**12.12** / **7.10** / **4.15**	76.53 / 80.83 / 78.40	14.51 / 10.65 / 5.93
[1 0 0 0]	56.64 / 31.96 / 8.33	22.38 / 21.14 / 23.99	61.07 / 43.67 / 14.88	26.02 / 29.65 / 24.86	73.20 / 58.00 / 29.77	34.71 / 34.97 / 29.14	76.63 / 70.04 / 45.04	12.90 / 12.27 / 11.44	**79.25** / **73.30** / **50.29**	12.55 / **10.43** / **9.40**	78.32 / 71.15 / 48.13	**11.66** / 10.75 / 9.91	77.54 / 71.20 / 46.78	12.20 / 11.20 / 9.69
[0 0 0 1]	78.44 / 27.75 / 7.55	14.15 / 23.80 / 23.47	82.51 / 46.55 / 22.14	16.83 / 20.66 / 17.24	85.27 / 62.41 / 37.29	14.90 / 21.87 / 16.08	85.79 / 72.12 / 48.87	11.06 / 13.56 / 12.01	**88.23** / **75.38** / **49.62**	9.77 / 11.08 / 10.71	**88.23** / 73.53 / 48.91	**6.83** / 10.59 / 10.01	88.10 / 73.22 / 49.01	7.67 / **10.43** / **9.69**
[0 1 1 0]	84.05 / 84.18 / 78.83	11.69 / 9.83 / 6.30	81.48 / 76.36 / 68.24	15.66 / 13.33 / 8.37	84.23 / 83.49 / 72.80	12.37 / 12.58 / 9.39	86.49 / 86.15 / 79.76	8.31 / 6.71 / 4.42	**87.62** / **86.36** / 79.70	7.23 / 6.07 / 4.05	86.99 / 85.10 / **79.99**	**6.70** / 5.97 / 4.25	87.20 / 84.13 / 79.73	7.00 / **5.94** / **3.76**
[1 1 0 0]	68.16 / 77.39 / 75.72	18.82 / 10.59 / 7.75	70.15 / 73.92 / 66.51	20.59 / 17.76 / 12.36	76.73 / 81.85 / 72.29	20.30 / 18.43 / 11.04	78.06 / 85.17 / 80.80	12.06 / 8.40 / 5.34	**80.00** / **85.50** / **81.10**	11.10 / 6.21 / 4.11	79.67 / 85.35 / 79.86	**10.23** / **5.83** / **3.35**	79.21 / 83.83 / 80.60	11.01 / 6.64 / 4.43
[1 0 0 1]	83.20 / 46.30 / 11.07	9.77 / 17.28 / 22.72	83.71 / 52.17 / 17.33	13.98 / 18.56 / 16.02	87.93 / 69.06 / 42.63	11.97 / 16.49 / 13.22	87.80 / 75.10 / 51.39	7.09 / 10.57 / 9.58	**89.71** / **76.64** / 52.84	**5.90** / 8.69 / 8.60	89.66 / 75.96 / 53.54	6.10 / **8.63** / 8.90	89.11 / 76.15 / **54.30**	6.88 / 8.76 / **8.55**
[1 0 1 0]	81.11 / 46.27 / 13.04	12.16 / 18.18 / 22.26	80.21 / 58.72 / 27.50	16.65 / 20.37 / 15.55	84.91 / 68.17 / 41.44	13.99 / 19.39 / 16.37	85.48 / 74.24 / 51.38	9.25 / 11.16 / 8.63	**87.51** / **76.13** / 54.62	7.26 / **9.13** / **8.54**	86.95 / 73.35 / **54.85**	7.47 / 9.55 / 9.36	86.34 / 72.81 / 51.84	**6.94** / 9.65 / 8.87
[0 0 1 1]	81.15 / 39.30 / 9.56	10.61 / 18.96 / 22.76	86.85 / 60.12 / 30.20	12.14 / 14.45 / 13.40	87.98 / 69.54 / 43.44	11.15 / 14.88 / 11.76	88.75 / 75.52 / 52.64	6.45 / 9.94 / 9.32	**89.78** / **77.29** / 54.30	5.87 / **8.39** / **8.39**	89.65 / 76.08 / 53.43	**5.67** / 8.61 / 9.03	89.17 / 74.90 / 55.18	6.85 / 9.17 / 8.67
[0 1 0 1]	88.90 / 84.56 / 78.25	9.34 / 8.93 / 5.58	85.26 / 77.14 / 68.27	12.78 / 12.97 / 9.21	87.94 / 84.31 / 72.88	11.83 / 13.70 / 9.59	87.74 / 84.78 / 78.40	7.81 / 7.66 / 5.18	**90.02** / **86.46** / 79.40	**6.03** / 6.48 / 4.57	89.50 / 86.01 / **79.81**	6.55 / 6.61 / **3.66**	89.79 / 85.33 / 79.71	6.91 / **6.43** / 4.27
[1 1 0 1]	88.67 / 84.88 / 77.89	8.16 / 7.61 / 4.43	86.20 / 76.51 / 68.17	12.63 / 12.30 / 8.39	88.17 / 84.94 / 73.96	11.35 / 11.41 / 7.56	87.94 / 84.81 / 79.07	7.59 / 8.40 / 5.43	**90.01** / **86.39** / 81.10	**5.69** / 6.32 / 3.83	89.89 / 85.94 / 79.93	5.93 / **5.98** / **3.62**	89.93 / 85.50 / **82.21**	7.20 / 6.26 / 4.08
[1 0 1 1]	84.29 / 47.78 / 11.60	9.31 / 16.55 / 21.69	86.88 / 61.58 / 27.62	11.22 / 15.09 / 13.37	88.67 / 71.52 / 46.34	11.00 / 14.49 / 12.14	88.76 / 75.59 / 53.51	7.10 / 9.81 / 8.86	**89.99** / **77.62** / 55.43	5.66 / **8.31** / 8.33	89.87 / 76.84 / 56.42	**5.64** / 8.54 / 8.74	89.36 / 75.79 / **56.58**	7.10 / 8.77 / **7.96**
[0 1 1 1]	89.76 / 85.39 / 79.12	7.80 / 7.77 / 4.52	87.48 / 77.69 / 68.99	10.82 / 11.04 / 8.17	88.96 / 85.42 / 73.50	9.67 / 9.62 / 6.79	89.08 / 85.41 / 78.62	7.04 / 7.30 / 4.43	**90.45** / **86.48** / 80.08	5.53 / 5.92 / 4.06	90.15 / 86.08 / 79.26	**5.50** / 5.80 / 3.86	90.05 / 85.41 / **80.68**	7.39 / **5.77** / **3.82**
[1 1 1 0]	84.05 / 84.66 / 78.28	10.50 / 8.96 / 5.24	82.29 / 76.29 / 68.78	14.51 / 14.00 / 7.14	85.27 / 84.09 / 74.66	12.25 / 11.44 / 8.13	85.95 / 85.41 / 79.77	8.24 / 7.69 / 4.33	87.01 / **85.97** / 81.08	7.26 / 6.22 / 3.88	87.07 / 85.62 / 79.20	**6.58** / 5.92 / 3.93	**87.46** / 84.27 / **81.48**	7.28 / **5.91** / **3.74**
[1 1 1 1]	89.68 / 85.62 / 78.29	6.82 / 6.86 / 3.81	87.66 / 77.55 / 68.91	10.11 / 11.31 / 8.07	89.05 / 85.29 / 74.06	8.78 / 8.91 / 6.67	88.88 / 85.18 / 78.80	7.41 / 7.33 / 4.42	**90.28** / **86.43** / 80.18	**5.57** / 5.84 / 3.82	90.17 / 86.05 / 79.41	**5.57** / 5.90 / 3.91	90.01 / 85.36 / **81.15**	7.59 / **5.61** / **3.68**
Average	79.77 / 62.36 / 45.93	12.67 / 14.16 / 13.77	80.54 / 65.91 / 47.52	15.71 / 17.11 / 13.07	83.97 / 75.14 / 57.30	15.55 / 17.31 / 13.18	85.20 / 79.88 / 65.75	9.11 / 9.51 / 7.17	**86.90** / **81.24** / **67.31**	7.79 / **7.74** / **6.39**	86.54 / 80.10 / 66.91	**7.42** / 7.76 / 6.45	86.29 / 79.19 / 67.06	8.37 / 8.13 / 6.48

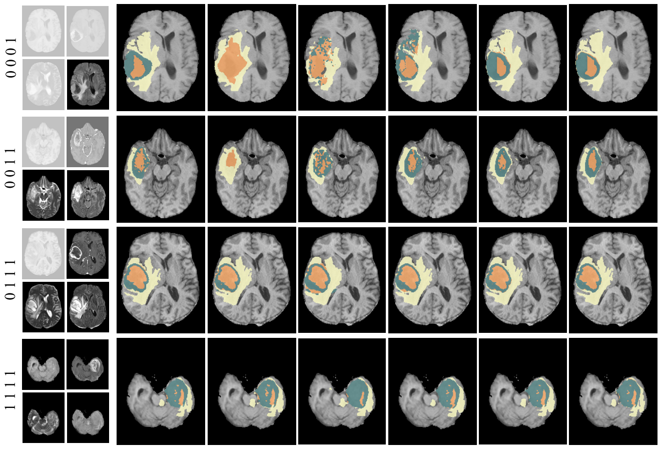
the methods of U-HVED, RobustSeg, and MM-GAN on all three tumor classes in terms of averaged Dice and HD95 scores over 15 missing cases ($p < .001$). Specifically, our method performs best on Dice/HD95 scores in 15/15 out of 15 missing cases for whole tumor (WT), 14/15 out of 15 cases for tumor core (TC), and 14/14 out of 15 cases for enhancing tumor (ET). Interestingly, even our baseline ("ours (w/o GF&ACF)") performs better than compared methods in average Dice and HD95 scores ($p < .001$), and obtains 86.29/79.19/67.06 (Dice) and 8.37/8.13/6.48 (HD95) in WT/TC/ET, justifying the effectiveness of our segmentation framework. Starting from this baseline, an extra graph-attention-based fusion block ("ours (w/o ACF)") improves the accuracies to 86.54/80.10/66.91 (Dice) and 7.42/7.76/6.45 (HD95) ($p < .001$, except the Dice scores on ET), and the average accuracies are further improved to 86.90/81.24/67.31 (Dice) and 7.79/7.74/6.39 (HD95) ($p <$ .05, except the HD95 scores on WT) when adversarial common feature constraint is also used (i.e., "Ours"). These results indicate the effectiveness of each module in our method.

In Table II, the modified TC-UNet with our HBM module (i.e., "TC-UNet (w/ HBM)") outperforms compared methods on all three tumor classes in averaged Dice and HD95 scores ($p < .001$). However, our baseline also works better than this modified TC-UNet, which could be regarded our baseline without using hyper-GAE, verifying the effectiveness of our proposed segmentation strategy in Fig. 4(d). Figure 6 shows the brain tumor segmentation results using different methods in four different missing cases, and our method produces more accurate brain tumor segmentations than compared methods.

Results on BraTS 2018 dataset. For a thorough comparison, we also compare our method with more state-of-the-art methods using 3-fold cross validation on BraTS 2018 dataset, including U-HVED [3], RobustSeg [24], MCA-Net [25], FMC-Net [26], RFNet [4], and SMU-Net [30]. Table III reports the results on 171 subjects in three test sets, and the accuracies of compared methods are taken from their papers, which were obtained by using the same cross validation as ours. The results show that our method works better than

TABLE III
BRAIN TUMOR SEGMENTATION ACCURACIES OF DIFFERENT METHODS IN DIFFERENT MISSING MODALITY CASES ON BRATS 2018 DATASET

Modality	**U-HVED** [3]	**RobustSeg** [24]	**MCA-Net** [25]	**FMC-Net** [26]	**RFNet** [4]	**SMU-Net** [30]	**Ours**
[T1 T1c T2 F]	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET	WT / TC / ET
[0 0 1 0]	80.90 / 54.10 / 30.80	82.24 / 57.49 / 28.97	32.20 / 15.70 / 7.20	80.40 / 59.50 / 35.20	84.30 / 67.62 / 40.71	85.70 / 67.20 / 43.10	**86.46** / **71.51** / **46.23**
[0 1 0 0]	62.40 / 66.70 / 65.50	73.31 / 76.83 / 67.07	33.50 / 55.90 / 53.50	72.00 / 83.10 / 75.00	74.93 / 80.99 / 69.43	**80.30** / **84.10** / **78.30**	77.80 / 83.62 / 77.46
[1 0 0 0]	52.40 / 37.20 / 13.70	70.11 / 47.90 / 17.29	5.30 / 6.30 / 5.30	74.90 / 55.50 / 29.20	74.68 / 64.42 / 34.43	**78.60** / **69.50** / 42.80	76.34 / 68.03 / **43.04**
[0 0 0 1]	82.10 / 50.40 / 24.80	85.69 / 53.57 / 25.69	73.70 / 48.60 / 25.80	85.90 / 64.60 / 39.50	86.46 / 64.89 / 33.92	87.50 / **71.80** / **46.10**	**88.35** / 69.27 / 43.16
[0 1 1 0]	82.70 / 73.70 / 70.20	85.19 / 80.20 / 69.71	48.30 / 50.40 / 52.40	81.70 / 84.80 / 75.50	86.39 / 83.27 / 73.01	86.10 / **85.00** / 75.70	**87.75** / 84.63 / **77.09**
[1 1 0 0]	66.80 / 69.70 / 67.00	77.18 / 78.72 / 69.06	29.20 / 54.80 / 53.80	76.20 / 84.20 / 75.80	78.59 / 82.22 / 70.73	**80.30** / **84.40** / 75.10	80.10 / 84.26 / **77.79**
[1 0 0 1]	84.30 / 55.30 / 24.20	88.24 / 60.68 / 32.13	80.40 / 51.50 / 10.20	86.50 / 67.00 / 43.20	88.78 / 71.59 / 39.68	87.30 / 71.20 / 44.00	**89.33** / **75.91** / **50.46**
[1 0 1 0]	82.20 / 57.20 / 30.70	84.78 / 62.19 / 32.01	35.50 / 14.30 / 6.10	83.40 / 62.60 / 38.00	86.15 / 70.89 / 41.42	85.60 / 73.50 / 47.70	**87.33** / **74.98** / **51.42**
[0 0 1 1]	87.50 / 59.70 / 34.60	88.28 / 61.16 / 33.84	81.30 / 25.00 / 10.00	86.50 / 66.60 / 45.50	89.12 / 70.82 / 43.77	87.90 / 71.20 / 46.00	**89.78** / **74.51** / **51.83**
[0 1 0 1]	85.50 / 72.90 / 70.30	88.51 / 80.62 / 70.30	81.50 / 73.40 / 67.50	86.20 / 85.00 / 77.10	89.17 / 82.94 / 72.84	88.40 / 84.10 / **77.30**	**89.83** / **85.45** / 76.91
[1 1 0 1]	86.20 / 74.20 / 71.10	88.73 / 81.06 / 70.78	82.70 / 75.80 / 68.40	86.60 / **85.60** / **77.20**	89.71 / 83.77 / 73.17	88.20 / 84.20 / 76.20	**89.85** / 85.46 / 77.11
[1 0 1 1]	88.00 / 61.50 / 34.10	88.81 / 64.38 / 36.41	85.40 / 44.40 / 12.90	86.80 / 68.00 / 45.60	89.68 / 73.09 / 44.79	88.30 / 67.90 / 43.10	**89.95** / **76.57** / **53.80**
[0 1 1 1]	88.60 / 75.60 / 71.20	89.27 / 80.72 / 70.88	87.70 / 77.40 / 67.20	86.40 / **86.00** / 76.90	90.09 / 83.54 / 73.13	88.20 / 82.50 / 75.40	**90.32** / 85.37 / **77.10**
[1 1 1 0]	83.30 / 75.30 / 71.10	86.01 / 80.33 / 70.10	50.10 / 52.10 / 54.80	82.90 / **85.20** / 76.20	86.78 / 83.97 / 72.56	86.50 / 84.40 / 76.20	**87.87** / 84.88 / **77.34**
[1 1 1 1]	88.80 / 76.40 / 71.70	89.45 / 80.86 / 71.13	88.10 / 78.80 / 69.10	86.60 / 85.80 / 76.90	90.26 / 84.02 / 73.21	88.90 / **87.30** / **79.30**	**90.29** / 85.46 / 77.29
Average	80.10 / 64.00 / 50.00	84.39 / 69.78 / 51.02	59.67 / 48.23 / 37.61	82.87 / 74.90 / 59.12	85.67 / 76.53 / 57.12	85.90 / 77.90 / 61.80	**86.76** / **79.33** / **63.87**



(a) Input (b) GT (c) MM-GAN (d) U-HVED (e) RobustSeg (f) TC-UNet (g) Ours

Fig. 6. Visual comparison of brain tumor segmentation results using different methods in different missing cases on BraTS 2019 dataset.

compared methods on all three tumor classes in averaged Dice scores. Compared with the current SoTA method, i.e., SMU-Net, our method improves the average Dice scores by 1.00%, 1.84% and 3.35% for the WT, TC and ET respectively. Moreover, our method achieves the highest accuracies in 11 out of 15 cases for WT, 6 out of 15 cases for TC, and 10 out of 15 cases for ET.

E. Impact of Segmentation Strategy

We evaluate the performance of our method using different segmentation strategies, as introduced in Section IV and Fig. 4, on BraTS 2019 dataset. We compare our whole method (refer to "Ours", as shown in Fig. 4(d)) with three ablated versions, including our method without using image reconstruction path (refer to "Ours (w/o Rec&Img)", as in Fig. 4(a)), our method without using completed multi-modal images (refer to "Ours (w/o Img)", as in Fig. 4(b)), and our method without using common features (refer to "Ours (w/o Feat)", as in Fig. 4(c)). All the other modules proposed in Sections III and IV (except the segmentation strategy) are utilized in all ablated versions. Table IV reports the averaged Dice and HD95 scores over 15 missing modality cases. The results show that, all three ablated versions outperform the methods of MM-GAN, U-HVED and RobustSeg in Table II ($p < .001$), which justifies the effectiveness of the other modules we designed in Sections III and IV. However, our method ("Ours") still performs better than three ablated versions in all metrics ($p < .05$) except the comparable TC and ET scores of our method without using common features ("ours (w/o Feat)"). These results indicate the superiority of our proposed strategy of combining feature-level and image-level completed information for segmentation (i.e., "Ours"), compared with the segmentation strategies only using either of them (i.e., three ablated versions).

TABLE IV
IMPACT OF SEGMENTATION STRATEGY AND HYPERNET-BASED MODULATION MODULE

Method	**Averaged Dice Score**	**Averaged HD95 Score**
	WT / TC / ET | AUG	WT / TC / ET | AUG
Ours (w/o Rec&Img)	85.98 / 80.43 / 65.49 | 77.30	8.35 / 8.16 / 7.21 | 7.91
Ours (w/o Img)	86.26 / 80.34 / 66.61 | 77.74	9.28 / 7.85 / 6.84 | 7.99
Ours (w/o Feat)	86.68 / 80.26 / **67.37** | 78.10	9.29 / **7.64** / 6.48 | 7.81
Ours (w/o HBM)	86.34 / 80.83 / 66.94 | 78.03	8.61 / 8.28 / 6.65 | 7.84
Ours	**86.90** / **81.24** / 67.31 | **78.48**	**7.79** / 7.74 / **6.39** | **7.31**

F. Impact of Hypernet-Based Modulation Module

This experiment compares with our method without using hypernet-based modulation (HBM) module in the segmentation network (i.e., "Ours (w/o HBM)"). The test accuracies are reported in Table IV. The results show that our method using a standard TC-UNet as the segmentation network ("Ours (w/o HBM)") works well and outperforms the methods of MM-GAN, U-HVED and RobustSeg in Table II in all metrics ($p < .001$), indicating that our hyper-GAE can conduct relatively accurate missing-modality completion in feature and image levels for tumor segmentation. However, our method using a modified TC-UNet with HBM module (i.e., the backbone TC-UNet is tuned by a modulator) as the segmentation network ("Ours") still works better than that using a standard TC-UNet on all three tumor regions in averaged Dice and HD95 scores ($p < .001$, except the Dice scores on TC), demonstrating the effectiveness of our proposed HBM module.

VI. Conclusion

In this work, we propose a unified and adaptive multi-modal MR image synthesis network, and apply it to tumor segmentation with missing modalities. We design a unified network consisting of a hyper-encoder, a graph-attention-based fusion block and a hyper-decoder accomplishing the flexible missing modality completion, and propose an adversarial common feature constraint to enforce the common feature space. To apply our method to tumor segmentation with missing modalities, we utilize both the feature-level and image-level completed information for segmentation, and propose a hypernet-based modulation module to adaptively utilize the synthesized and real modalities. Experimental results demonstrated that our method can achieve state-of-the-art performance on both multi-modal MR image synthesis and brain tumor segmentation with missing modalities. We also conducted the ablation study and investigated the impact of different segmentation strategies on performance.

Although our approach qualitatively and quantitatively outperforms all the other competing methods, we note that there are still several failure cases of our method. For multi-modal MR image synthesis, our method failed on 2 out of 111 test subjects (e.g., PSNR $<$ 25), which contain residual skull regions and strong artifacts in some modalities. For missing-modality MR image segmentation, our proposed method failed in several missing modality cases on 4 out of 111 test subjects (e.g., Dice $<$ 0.05). This is probably because some modalities in these subjects have poor contrast over tumor and normal tissues, resulting in unsatisfactory tumor segmentation results when only these modalities are available. One possible solution to these failure cases might be introducing uncertainty qualification to predict the reliability of synthesis and segmentation results. For future work, as our method is a general framework for missing-modality MR image synthesis and segmentation, it would be interesting to apply it to other multi-modal MR image datasets. In addition, we are also interested in attempting other attention modeling in fusion block and extending our methodology to the multi-institutional setting [40] for further improving the performance.

References

[1] L. Shen et al., "Multi-domain image completion for random missing input data," *IEEE Trans. Med. Imag.*, vol. 40, no. 4, pp. 1113–1122, Apr. 2021.

[2] D. Lee, W.-J. Moon, and J. C. Ye, "Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 34–42, Jan. 2020.

[3] R. Dorent, S. Joutard, M. Modat, S. Ourselin, and T. Vercauteren, "Hetero-modal variational encoder–decoder for joint modality completion and segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Shenzhen, China: Springer, 2019, pp. 74–82.

[4] Y. Ding, X. Yu, and Y. Yang, "RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3955–3964.

[5] D. Ha, A. M. Dai, and Q. V. Le, "Hypernetworks," in *Proc. ICLR*, 2017, pp. 1–11.

[6] Z. Jiang, C. Ding, M. Liu, and D. Tao, "Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task," in *Proc. Int. MICCAI Brainlesion Workshop*, 2019, pp. 231–241.

[7] H. Yang, J. Sun, L. Yang, and Z. Xu, "A unified hyper-GAN model for unpaired multi-contrast MR image translation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Strasbourg, France: Springer, 2021, pp. 127–137.

[8] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Med. Image Anal.*, vol. 35, pp. 475–488, Jan. 2017.

[9] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, "Image synthesis in multi-contrast MRI with conditional generative adversarial networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2375–2388, Oct. 2019.

[10] T. D. Bui, M. Nguyen, N. Le, and K. Luu, "Flow-based deformation guidance for unpaired multi-contrast MRI image-to-image translation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Lima, Peru: Springer, 2020, pp. 728–737.

[11] Y. Huang, F. Zheng, R. Cong, W. Huang, M. R. Scott, and L. Shao, "MCMT-GAN: Multi-task coherent modality transferable GAN for 3D brain image synthesis," *IEEE Trans. Image Process.*, vol. 29, pp. 8187–8198, 2020.

[12] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Sample-adaptive GANs: Linking global and local mappings for cross-modality MR image synthesis," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2339–2350, Jul. 2020.

[13] M. Sohail, M. N. Riaz, J. Wu, C. Long, and S. Li, "Unpaired multi-contrast MR image synthesis using generative adversarial networks," in *Simulation and Synthesis in Medical Imaging*. Shenzhen, China: Springer, 2019, pp. 22–31.

[14] H. Li et al., "DiamondGAN: Unified multi-modal generative adversarial networks for MRI sequences synthesis," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Shenzhen, China: Springer, 2019, pp. 795–803.

[15] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao, "Hi-Net: Hybrid-fusion network for multi-modal MR image synthesis," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2772–2781, Sep. 2020.

[16] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, "MustGAN: Multi-stream generative adversarial networks for MR image synthesis," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101944.

[17] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris, "Multimodal MR synthesis via modality-invariant latent representation," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 803–814, Mar. 2018.

[18] A. Sharma and G. Hamarneh, "Missing MRI pulse sequence synthesis using multi-modal generative adversarial network," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1170–1183, Apr. 2020.

[19] M. Hamghalam, A. F. Frangi, B. Lei, and A. L. Simpson, "Modality completion via Gaussian process prior variational autoencoders for multi-modal glioma segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Strasbourg, France: Springer, 2021, pp. 442–452.

[20] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.

[21] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[22] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "HeMIS: Hetero-modal image segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Athens, Greece: Springer, 2016, pp. 469–477.

[23] Y. Shen and M. Gao, "Brain tumor segmentation on MRI with missing modalities," in *Information Processing in Medical Imaging*. Hong Kong: Springer, 2019, pp. 417–428.

[24] 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 *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Shenzhen, China: Springer, 2019, pp. 447–456.

[25] T. Zhou, S. Canu, P. Vera, and S. Ruan, “Brain tumor segmentation with missing modalities via latent multi-source correlation representation,” in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Lima, Peru: Springer, 2020, pp. 533–541.

[26] 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.

[27] Q. Yang, X. Guo, Z. Chen, P. Y. M. Woo, and Y. Yuan, “D^2-Net: Dual disentanglement network for brain tumor segmentation with missing modalities,” *IEEE Trans. Med. Imag.*, vol. 41, no. 10, pp. 2953–2964, Oct. 2022.

[28] Y. Zhang et al., “Modality-aware mutual learning for multi-modal medical image segmentation,” in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Strasbourg, France: Springer, 2021, pp. 589–599.

[29] Y. Wang et al., “ACN: Adversarial co-training network for brain tumor segmentation with missing modalities,” in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Strasbourg, France: Springer, 2021, pp. 410–420.

[30] R. Azad, N. Khosravi, and D. Merhof, “SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities,” in *MIDL*. Zürich, Switzerland: PMLR, 2022, pp. 48–62.

[31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[32] Y. Ganin and V. Lempitsky, “Unsupervised domain adaptation by backpropagation,” in *Proc. ICML*, 2015, pp. 1180–1189.

[33] Y. Alharbi, N. Smith, and P. Wonka, “Latent filter scaling for multimodal unsupervised image-to-image translation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1458–1466.

[34] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[35] N. J. Tustison et al., “N4ITK: Improved N3 bias correction,” *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[36] J. C. Reinhold, B. E. Dewey, A. Carass, and J. L. Prince, “Evaluating the impact of intensity normalization on MR image synthesis,” *Proc. SPIE*, vol. 10949, pp. 890–898, Mar. 2019.

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[38] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, 1993.

[39] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. ICLR*, 2015, pp. 1–15.

[40] L. Zuo et al., “Information-based disentangled representation learning for unsupervised MR harmonization,” in *Proc. IPMI*, 2021, pp. 346–359.