A Fully Automated Multimodal MRI-Based Multi-Task Learning for Glioma Segmentation and IDH Genotyping

Jianhong Cheng, Jin Liu, *Member, IEEE*, Hulin Kuang, and Jianxin Wang, *Senior Member, IEEE*

***Abstract*— The accurate prediction of isocitrate dehydrogenase (IDH) mutation and glioma segmentation are important tasks for computer-aided diagnosis using preoperative multimodal magnetic resonance imaging (MRI). The two tasks are ongoing challenges due to the significant inter-tumor and intra-tumor heterogeneity. The existing methods to address them are mostly based on single-task approaches without considering the correlation between the two tasks. In addition, the acquisition of IDH genetic labels is expensive and costly, resulting in a limited number of IDH mutation data for modeling. To comprehensively address these problems, we propose a fully automated multimodal MRI-based multi-task learning framework for simultaneous glioma segmentation and IDH genotyping. Specifically, the task correlation and heterogeneity are tackled with a hybrid CNN-Transformer encoder that consists of a convolutional neural network and a transformer to extract the shared spatial and global information learned from a decoder for glioma segmentation and a multi-scale classifier for IDH genotyping. Then, a multi-task learning loss is designed to balance the two tasks by combining the segmentation and classification loss functions with uncertain weights. Finally, an uncertainty-aware pseudo-label selection is proposed to generate IDH pseudo-labels from larger unlabeled data for improving the accuracy of IDH genotyping by using semi-supervised learning. We evaluate our method on a multi-institutional public dataset. Experimental results show that our proposed multi-task network achieves promising performance and outperforms the single-task learning counterparts and other existing state-of-the-art methods. With the introduction of unlabeled data, the semi-supervised multi-task learning framework further improves the performance of glioma segmentation and IDH genotyping. The source codes of our framework are publicly available at https://github.com/miacsu/MTTU-Net.git.**

***Index Terms*— Multi-task learning, semi-supervised, glioma segmentation, IDH genotyping, transformer.**

Manuscript received December 13, 2021; accepted January 5, 2022. Date of publication January 12, 2022; date of current version June 1, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFF1201200, in part by the National Natural Science Foundation of China under Grant 62172444 and Grant 62102454, in part by the Hunan Provincial Science and Technology Innovation Leading Plan under Grant 2020GK2019, and in part by the High Performance Computing Center of Central South University. *(Corresponding authors: Jin Liu; Jianxin Wang.)*

Jianhong Cheng is with the Hunan Provincial Key Laboratory on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China, and also with the Institute of Guizhou Aerospace Measuring and Testing Technology, Guiyang 550009, China (e-mail: jianhong_cheng@csu.edu.cn).

Jin Liu, Hulin Kuang, and Jianxin Wang are with the Hunan Provincial Key Laboratory on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China (e-mail: liujin06@csu.edu.cn; hulinkuang@csu.edu.cn; jxwang@mail.csu.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2022.3142321, provided by the authors.

Digital Object Identifier 10.1109/TMI.2022.3142321

I. Introduction

GLIOMAS are the most common primary brain tumors that occur in the central nervous system [1]. They are subdivided into low-grade gliomas (LGGs, grade II and III) and high-grade gliomas (HGGs, grade IV) according to the World Health Organization (WHO) classification of tumors in 2016 [2]. It is reported that isocitrate dehydrogenase (IDH) mutation status is one of the most important prognostic markers in gliomas [2]. Clinical studies have found that patients with IDH-mutant have a better prognosis than those with IDH-wildtype in LGGs. Beside, traditional manual segmentation of glioma from medical images is subjective and time-consuming. Thus, accurate IDH genotyping and precise glioma segmentation are of great significance to guide the treatment and assess the prognosis.

Magnetic resonance imaging (MRI) has been considered the most promising candidate due to its non-invasive nature and its role in routine clinical practice. Numerous studies have shown that some quantitative MRI features are associated with histological and genetic information of gliomas [3]–[5]. In the past few years, imaging phenotypes have been prevalently used to characterize tumor grades and genotypes, known as radiomics or radiogenomics [6]–[8]. Typically, those methods rely on rigorous multi-step pipelines including image acquisition, data preprocessing, tumor segmentation, radiomic feature extraction, feature selection, and classification modeling [4], [9]. Preoperative multimodal MRI is routine for diagnosis of gliomas, so we also use it to predict IDH mutations. Besides, the segmentation or location of tumor is the key step before IDH genotyping in radiogenomics. At present, glioma segmentation is still performed manually by experienced neuro-radiologists, which is an extremely tedious and time-consuming procedure. Deep learning-based methods, especially convolutional neural networks (CNNs), have shown the potential to perform automatic segmentation. The performance of the existing segmentation networks such as U-Net [10], V-Net [11], DMFNet [12] still needs to be improved. Furthermore, the radiomic features

extracted from MRI for radiogenomics are highly interpretable, but they are vulnerable to site-specific variations that affect reproducibility and robustness [9]. Thus, automated glioma segmentation and IDH genotyping are extremely urgent tasks for computer-aided diagnosis (CAD) systems using MRI images.

Different IDH genotypes in routine MRI images usually have different morphological characteristics. Specifically, gliomas with IDH-mutant show distinct nonenhancing tumor margins and small cysts, whereas gliomas with IDH-wildtype shows an indistinct margin of a bithalamic and rim enhancement surrounding central necrosis [13]. In clinical practice, these image phenotypes are helpful neuro-radiologists in distinguishing gliomas. However, it is difficult for even neuro-radiologists to accurately distinguish glioma genotypes based on these image phenotypes due to the heterogeneity of gliomas. Moreover, the existing automatic-based approaches are mostly based on single-task approaches without considering the correlation between the two tasks. Thus, training two tasks jointly in the same deep learning network to encourage feature sharing between glioma segmentation and IDH genotyping is a promising direction for CAD systems.

In this study, we propose a multi-task learning network to jointly train glioma segmentation and IDH genotyping in an end-to-end manner from multimodal MRI images. Specifically, the task correlation and heterogeneity are tackled with a hybrid CNN-Transformer encoder that consists of a convolutional neural network and a transformer [14] to extract the shared spatial and global information learned from a decoder for glioma segmentation and a multi-scale classifier for IDH genotyping. The multi-task learning loss is designed to balance the performance of segmentation and classification tasks using uncertain weights. To make full use of unlabeled data, we further propose a semi-supervised multi-task learning framework based on an uncertainty-aware pseudo-label selection (UPS) to improve the performance of IDH genotyping. Experiments on a multi-institutional public dataset demonstrate that our proposed method achieves an encouraging performance and outperforms several existing state-of-the-art methods for glioma segmentation and IDH genotyping. Our main contributions are summarized as follows:

1) We develop an end-to-end multi-task learning network for simultaneous glioma segmentation and IDH genotyping by sharing the spatial and global feature representation extracted from the hybrid CNN-Transformer encoder.
2) A multi-task loss is proposed to balance the performance of glioma segmentation and IDH genotyping, which combines segmentation and classification losses with uncertain weights to accurately infer the two tasks.
3) A semi-supervised multi-task learning framework based on UPS is built to further improve the performance of IDH genotyping with larger unlabeled data.

II. Related Work

A. Glioma Segmentation

Early researches on glioma segmentation mainly rely on traditional machine learning approaches, such as region growing [15], clustering [16], graph cuts [17], and Bayesian models [18], and shape prior [19]. In previous studies, many glioma segmentation methods based on convolutional neural networks (CNNs) have been proposed for multimodal MR images [20]. In particular, fully convolutional networks (FCNs) [21] have always been dominate in medical image segmentation for many years. Among these variants, U-Net, which is composed of a symmetric encoder-decoder architecture with skip connections to improve feature representations [10], has become an obvious choice for medical image segmentation [22]–[24]. Based on the baseline, several networks such as V-Net [11] and nnU-Net [25] are proposed to improve the segmentation performance. Although CNN-based methods have excellent representation power, they generally exhibit limitations such as high computational cost due to the intrinsic locality of convolution operations, especially for 3D medical image segmentation. Chen *et al.* developed a light-weight network architectures, termed as DMFNet, to balance the model efficiency and accuracy for glioma segmentation from 3D MRIs [12]. Recently, transformers have become increasingly popular for sequence-to-sequence prediction in computer vision tasks [26]. Since transformers benefit from global representation modeling using self-attention mechanisms, it is natural to use them for medical image segmentation. Li *et al.* [27] design a squeeze-and-expansion transformer for medical image segmentation. Hatamizadeh *et al.* develop UNETR [28] for medical image segmentation and employ the vision transformer (ViT) [14] to enhance the feature representation.

B. IDH Genotyping

Traditional radiomics-based methods for IDH mutation classification usually require to manual segmentation of tumor volume of interest (VOI), extract local radiomic features from the VOI, and then build a machine learning classifier for prediction. Ren *et al.* [29] employed the support vector machine with a recursive feature elimination algorithm (SVM-RFE) to identify the IDH mutation status. Wu *et al.* [30] also extracted radiomic features and employed random forest as classifier to predict IDH genotype. These radiomic methods rely on multistep pipelines and the radiomic features are limited to prior knowledge. CNN-based methods simplify this pipeline by learning high-level semantic features directly from the MRI images. Chang *et al.* [31] employed residual convolutional neural network to noninvasively predict IDH genotype by learning semantic features from the selected slices. Matsui *et al.* [32] also selected the axial slices including the tumor center as input, and adopted residual network to predict molecular subtypes of low-grade glioma. Both of them are based on slice-wise prediction, without considering the global features of tumor. To overcome this, Liang *et al.* [33] employed 3D DenseNet model to predict IDH genotypes with the whole tumor lesion patches. T2-Net [34] was proposed to automatically predict IDH-mutant volume and IDH-wild volume only from T2w image, and then the IDH mutation status was determined by majority voting from both volumes. Although a single T2w modality is useful to identify IDH

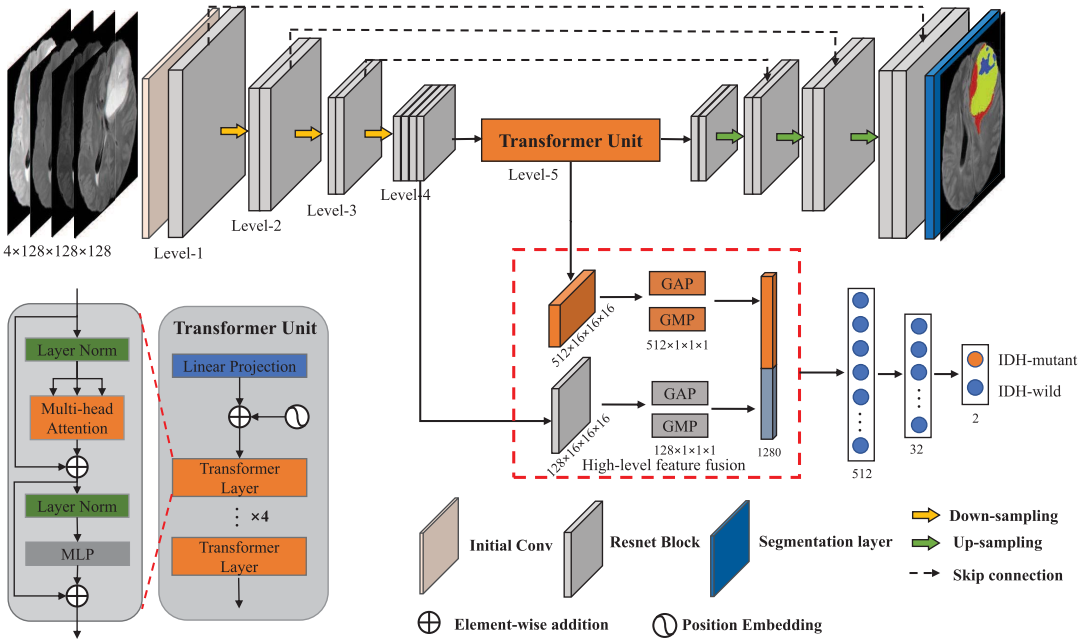

Fig. 1. Overview of the proposed multi-task learning network. GAP and GMP represent global average pooling and global max pooling operation, respectively. Number in subgraph represents the shape of feature map or the number of neurons.

genotype in glioma, multiple MRI volumes can provide more global semantic information for IDH mutation classification.

C. Joint Segmentation and Genotyping

There are few studies on joint prediction of glioma segmentation and IDH genotyping from multimodal MRI images. A recent study found that Wang *et al.* [35] proposed a multi-task network, known as SGPNet, for glioma segmentation and IDH genotype prediction, and it was evaluated on a dataset of 121 patients using 5-fold cross-validation. In addition, the concept of multi-task learning has been applied to other medical applications. Zhou *et al.* [36] developed a multi-task learning network, namely CMSVNet, to simultaneously segment breast tumors and predict benign and malignant tumors from 3D automated breast ultrasound images. Xie *et al.* [37] proposed the mutual bootstrapping deep CNNs for simultaneous skin lesion segmentation and classification. They first employed a coarse segmentation network to generate coarse lesion masks, and then used these coarse masks to provide a prior bootstrapping for a classification network. Zhang *et al.* [38] proposed a multi-attention guided multi-task learning network (MA-MTLN) for simultaneous gastric tumor segmentation and lymph node classification. In addition to the design of the multi-task network architecture, researchers also focus on how to balance tasks by improving the multi-task loss function. In this study, we employ an uncertainty-based method [39] to adaptively weight glioma segmentation and IDH genotyping losses to prevent task bias.

III. Multi-Task Learning for Glioma Segmentation and IDH Genotyping

For clarification, we first define the following notations. Let $D_L = \{(\mathbf{x}^{(i)}, \mathbf{g}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N_L}$ be a labeled dataset with N_L samples. where $\mathbf{x}^{(i)}$ is the input modalities of the ith sample, $\mathbf{g}^{(i)}$ and $\mathbf{y}^{(i)} = \{y^{(i)} | y^{(i)} \in \{0, 1\}\}_{i=1}^{N_L}$ denote the corresponding tumor ground truth and IDH mutation status of the ith sample, respectively. For the ith sample, $y^{(i)} = 1$ denotes that the sample is IDH-mutant, and $y^{(i)} = 0$ represents that the sample is IDH-wild.

A. Multi-Task Learning Network

Fig. 1 shows the proposed 3D multi-task learning network, which consists of three parts: i) a CNN-Transformer encoder $\boldsymbol{E}$ for encoding contextual information; ii) a decoder $\boldsymbol{D}$ for glioma segmentation; and iii) a classifier $\boldsymbol{P}$ for IDH genotyping. The proposed network takes four MRI modalities as input and jointly performs glioma segmentation and IDH genotyping tasks to encourage feature sharing. The encoder and decoder are connected by skip connections to form a U-Net architecture with a transformer unit embedded at the bottom for enhancing feature representation. Considering that the proposed multi-task network integrates transformer and U-Net, we call it MTTU-Net.

1) CNN-Transformer Encoder: To capture the long-range dependency restricted by CNN and reduce the computational complexity, transformer with multi-head self-attention mechanism is introduced into the bottom of the CNN encoder to realize efficient long-range contextual modeling. The encoder employs a series of convolution and transformer operations to extract global semantic features from the input $\mathbf{x} \in \mathbb{R}^{C \times H \times W \times D}$ with a 3D spatial resolution of $H \times W \times D$ and C channels. The input is first operated by an initial convolution with a kernel size of $3 \times 3 \times 3$ to generate feature maps with 16 channels. To capture spatial and depth feature representations, the initial feature maps are then calculated by three down-sampling operations and multiple stacked

residual convolution blocks to generate the feature map $F \in \mathbb{R}^{C' \times H' \times W' \times D'}$. For each down-sampling operation, we use a $3 \times 3 \times 3$ convolution with a stride of 2 to replace the max-pooling operation and the output channel is set to twice the input channel. Each residual convolution block (Resnet Block) consists of two convolution blocks and a residual connection is applied to between the input and output of the residual block. Each convolution block is composed of a batch normalization, a Leaky ReLU function, and a $3 \times 3 \times 3$ convolution layer. The feature map F_λ of CNN-based encoder at the λth level can be denoted by:

$$F_\lambda = E_{\text{cnn}}(\boldsymbol{\theta}; \text{x}; \lambda) \in \mathbb{R}^{C' \times \frac{H}{2^{\lambda-1}} \times \frac{W}{2^{\lambda-1}} \times \frac{D}{2^{\lambda-1}}}, \tag{1}$$

where $\boldsymbol{\theta}$ is the parameters of the CNN-based encoder, λ is the number of the feature level, and $C' = 2^{\lambda+3}$ is the number of channels.

To enhance feature representation, we employ self-attention mechanisms into the encoder via the usage of Transformers. The remarkable transformer, ViT [14], divides a 2D image into fixed-size 16×16 patches and reshapes each patch into a sequence of 16^2 length as a token input. For the 3D volumetric data, we extend the straightforward tokenization of ViT to 3D data by splitting the data into 3D patches. However, embedding a large 3D patch will increase the computational overhead of transformer. To alleviate this, the low-resolution high-level feature map $F_\lambda \in \mathbb{R}^{128 \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}} (\lambda = 4)$ extracted from CNN-based encoder is fed into the transformer-based encoder to further learn global feature representation. To ensure a comprehensive representation of each 3D volume, a linear projection composed of a $3 \times 3 \times 3$ convolution layer is used to increase the channel of the feature map $F_\lambda (\lambda = 4)$ from 128 to $d = 512$. The feature map $F \in \mathbb{R}^{d \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}}$ is then reshaped into $N \times d (N = \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8})$ dimension as input tokens of transformer layer. To learn the positional information, a learnable position embedding $\boldsymbol{E}_{pos} \in \mathbb{R}^{N \times d}$ is fused into the patch embedding $\boldsymbol{E}_{pat} \in \mathbb{R}^{N \times d}$ by an addition operation, which is denoted as follows:

$$\boldsymbol{Z}_0 = \boldsymbol{E}_{pat} + \boldsymbol{E}_{pos}, \tag{2}$$

where $\boldsymbol{Z}_0$ denotes the feature embedding. Afterwards, the feature embedding $\boldsymbol{Z}_0$ is fed into L stacked transformer layers. Each of them consists of multi-head self-attention (MSA) and multilayer perceptron (MLP) blocks. For the ℓth ($\ell \in [1, 2, \ldots, L]$) layer, the input to self-attention is a triplet $(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$ calculated from the input $\boldsymbol{Z}^{\ell-1}$ as:

$$\boldsymbol{Q} = \boldsymbol{Z}^{\ell-1}\boldsymbol{W}_Q, \quad \boldsymbol{K} = \boldsymbol{Z}^{\ell-1}\boldsymbol{W}_K, \quad \boldsymbol{V} = \boldsymbol{Z}^{\ell-1}\boldsymbol{W}_V, \tag{3}$$

where $\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V \in \mathbb{R}^{d \times d}$ are learnable parameters of three linear projection layers and d is the dimension of $(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$. Then, self-attention (SA) is calculated as:

$$\text{SA}(\boldsymbol{Z}^{\ell-1}) = \text{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d}})\boldsymbol{V}. \tag{4}$$

MSA is an important component of transformer layer, which allows the model to simultaneously focus on information from different representation subspaces at different locations. Specifically, it divides the input of transformer layer into n independent parts, processes each part in parallel using SA operation, and then projects these concatenated results using a linear projection layer. Therefore, MSA can be denoted as:

$$\text{MSA}(\boldsymbol{Z}^{\ell-1}) = \text{Concat}(\text{SA}_1(\boldsymbol{Z}^{\ell-1}), \cdots, \text{SA}_n(\boldsymbol{Z}^{\ell-1}))\boldsymbol{W}_O, \tag{5}$$

where $\boldsymbol{W}_O \in \mathbb{R}^{d \times d}$ is the learnable parameters of the linear projection layer and Concat$(\cdot)$ is the concatenation operation. Then, the output of MSA is converted by an MLP block with a residual skip as the layer output $\boldsymbol{Z}^\ell$. They can be denoted as follow:

$$\boldsymbol{Z}'^{\ell}_k = \boldsymbol{Z}^{\ell-1} + \text{MSA}(\boldsymbol{Z}^{\ell-1}), \tag{6}$$

$$\boldsymbol{Z}^\ell = \boldsymbol{Z}'^{\ell}_k + \text{MLP}(\boldsymbol{Z}'^{\ell}_k). \tag{7}$$

It should be noted that layer normalization is used before MSA and MLP blocks, and these blocks are omitted for simplicity. Thus, the feature map of the hybrid CNN-Transformer encoder can be denoted by:

$$F_\lambda = \begin{cases} \boldsymbol{E}_{\text{cnn}}(\boldsymbol{\theta}; \text{x}; \lambda), & \lambda \in \{1, 2, 3, 4\} \\ \boldsymbol{E}_{\text{trans}}(\boldsymbol{\phi}; \text{x}'; \lambda), & \lambda = 5, \end{cases} \tag{8}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are the parameters of CNN-based unit and transformer unit, respectively. x' is the input of transformer unit and should be equal to F_4 in this study.

2) Decoder for Glioma Segmentation: To obtain glioma segmentation from the original 3D multimodal MRI images, we design a 3D CNN decoder to perform voxel-level segmentation. The decoder is composed of multiple cascaded up-sampling operations, which decodes the high-level feature map and outputs the final segmentation result. After reshaping the sequence of high-level feature $\boldsymbol{Z}^\ell \in \mathbb{R}^{N \times d}$ to the shape of $d \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}$, the cascaded up-sampling operations are performed to reach the full resolution from $\frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}$ to $H \times W \times D$. Each of up-sampling operation consists of a $1 \times 1 \times 1$ convolution and a transposed convolution with a stride of 2. To generate finer segmentation with richer spatial details, skip connections are used to fuse the down-sampling feature maps with the up-sampling counterparts via a concatenation operation followed by a $1 \times 1 \times 1$ convolution and a $3 \times 3 \times 3$ residual convolution block. Finally, a $1 \times 1 \times 1$ convolution followed by a softmax function as the segmentation layer of the decoder is applied to generate the segmentation result, which is denoted as:

$$\tilde{\mathbf{g}}(\tilde{F}_\lambda) = \boldsymbol{D}(\boldsymbol{\varphi}; \tilde{F}_\lambda), \tag{9}$$

where $\boldsymbol{\varphi}$ is the parameters of decoder and $\tilde{\mathbf{g}}$ is the output of the decoder. $\tilde{F}_\lambda$ is the input set of decoder and $\tilde{F}_\lambda = \{F_\lambda | \lambda = 1, 2, 3, 5\}$.

3) Classifier for IDH Genotyping: To analyze the features generated by the encoder, we visualize and reconstruct two-dimensional axial view feature maps from different levels of the proposed network to the original input size. As shown in Fig. 2, the first row is IDH-mutant glioma and the second row is IDH-wild glioma. We can see that low-level features (level-1 to level-3) mainly capture the shape and boundary information, high-level features (level-4 and level-5) summarize the target attributes, which are commonly used in classification tasks.

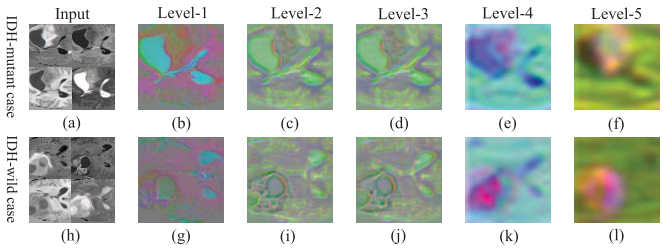

Fig. 2. 2D axial views of feature maps from different levels of the encoder of the proposed network.

Besides, we have performed additional ablation experiments to demonstrate the effectiveness of high-level feature fusion. Therefore, these high-level features at the two different levels are used as multi-scale features to build a multi-scale network $\boldsymbol{P}_\theta$ for IDH genotyping. We use global average pooling (GAP) and global max pooling (GMP) to further transform the multi-scale feature maps to the same size in each channel. We choose GAP and GMP because they do not need to introduce any additional parameters for optimization. Although GAP can benefit from spatial information, it may weaken important features. To remedy this, we use GMP to capture these features to achieve feature complementarity. These pooled multi-scale features are fused by concatenation and then input into three fully connected layers composed of 512, 32, and 2 neurons. Finally, the softmax function will output the probability of IDH genotyping. The probability is denoted by:

$$H_j = \text{Concat}(\text{GAP}(F_j), \text{GMP}(F_j)), \tag{10}$$

$$\tilde{H}_{j-1,j} = \text{Concat}(H_{j-1}, H_j), j = 5, \tag{11}$$

$$p_c^{(i)} = \text{Softmax}(\boldsymbol{P}_\theta(\tilde{H}_{j-1,j})), \tag{12}$$

where $F_j \in \mathbb{R}^{C_j \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}}$ represents the jth level feature map of the encoder. $H_j \in \mathbb{R}^{2C_j \times 1 \times 1 \times 1}$ and $\tilde{H}_j \in \mathbb{R}^{(2C_{j-1}+2C_j) \times 1 \times 1 \times 1}$ are both the fused features. $p_c^{(i)} \in \mathbb{R}^{1 \times 2}$ is the output probability of the sample i in class c.

B. Multi-Task Loss Function

For the glioma segmentation, the imbalance between the foreground and background may result in segmentation bias. To alleviate this problem, Dice coefficient-based loss function is used as the segmentation loss to focus on the shape similarity between the segmentation map and ground truth. The segmentation loss $\mathcal{L}_{\text{seg}}$ is defined as:

$$\mathcal{L}_{\text{seg}}(\tilde{\mathbf{g}}^{(i)}, \mathbf{g}^{(i)}) = 1 - \frac{2}{K} \sum_{k \in K} \frac{\sum \tilde{\mathbf{g}}_k^{(i)} \mathbf{g}_k^{(i)}}{\sum \tilde{\mathbf{g}}_k^{(i)} + \sum \mathbf{g}_k^{(i)} + \epsilon}, \tag{13}$$

where $\tilde{\mathbf{g}}^{(i)}$ is the softmax output of the proposed segmentation network, $\mathbf{g}^{(i)}$ is the corresponding ground truth. Both $\tilde{\mathbf{g}}^{(i)}$ and $\mathbf{g}^{(i)}$ have the same classes K. ϵ is a small constant to prevent division by 0.

Class imbalance and the limited number of samples are common problems in medical image classification. For example, in our glioma training dataset, the number of IDH-wild gliomas is approximately twice that of IDH-mutant gliomas, and the number of available samples is limited. To compensate for these problems, we propose a modified weighted cross-entropy loss as the classification loss function $\mathcal{L}_{\text{idh}}$:

$$\mathcal{L}_{\text{idh}}(\tilde{\mathbf{y}}^{(i)}, \hat{\mathbf{y}}^{(i)}, g^{(i)}) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{c=1}^{C} w_c \tilde{\mathbf{y}}_c^{(i)} \log(\hat{\mathbf{y}}_c^{(i)}), \tag{14}$$

where $\tilde{\mathbf{y}}^{(i)}$ and $\hat{\mathbf{y}}^{(i)}$ are the ground truth of IDH genotype and predicted probability of the proposed network for the sample i, respectively. w_c is class weight for class c ($w_1 = \frac{N_2}{N_1+N_2}$, $w_2 = \frac{N_1}{N_1+N_2}$, N_1 and N_2 are the number of IDH-mutant and IDH-wild samples, respectively).

In this study, the segmentation loss and classification loss are jointly performed to optimize the multi-task learning network. However, the multi-task learning may lead to task bias when task weights are set improperly. To alleviate the negative impact of this multi-task learning, we employ an uncertainty-based method [39] that can adaptively adjust weights to weight the glioma segmentation and IDH genotyping losses. The multi-task loss $\mathcal{L}_{\text{joint}}$ is defined as:

$$\mathcal{L}_{\text{joint}} = \frac{1}{2\sigma_{\text{seg}}^2} \mathcal{L}_{\text{seg}} + \frac{1}{2\sigma_{\text{idh}}^2} \mathcal{L}_{\text{idh}} + \log \sigma_{\text{seg}} \sigma_{\text{idh}} \tag{15}$$

where σ_{seg} and σ_{idh} are uncertain weights and learnable parameters for network learning. In practice, σ_{seg} and σ_{idh} are first initialized to two tensors with values of 1, and then updated adaptively by iteration during the training phase.

IV. Semi-Supervised Multi-Task Learning for IDH Genotyping

Since the acquisition of IDH genetic labels is expensive and costly, there is a large amount of data without IDH mutation information. The limited number of labeled data makes the generalization performance of deep learning network insufficient. In this scenario, we further propose an alternative solution that utilizes semi-supervised multi-task learning to improve the performance of IDH genotyping.

A. Semi-Supervised Multi-Task Learning Framework

The proposed semi-supervised multi-task learning framework utilizes a large amount of unlabeled data to further improve the performance of IDH genotyping. Let $D_U = \{(\mathbf{x}^{(i)}, \mathbf{g}^{(i)})\}_{i=1}^{N_U}$ be an unlabeled dataset with N_U samples, which misses the IDH labels but has the complete multimodal MRI data $\mathbf{x}^{(i)}$ and corresponding segmented ground truth $\mathbf{g}^{(i)}$. For the semi-supervised MTL, a scheme combining self-training with active learning is adopted. Specifically, the MTL network mentioned in III-A is first trained on the labeled dataset D_L, and the MTL parameterized models including an encoder $\boldsymbol{E}_0$, a decoder $\boldsymbol{D}_0$, and classifier $\boldsymbol{P}_0$ are carried out on the unlabeled dataset D_U to generate the corresponding pseudo-labels $\tilde{\mathbf{y}}^{(i)}$ of IDH gene. However, the traditional pseudo-labeling method, even with a high confidence, may be incorrect due to the poor calibration of the model, and then result in noise training. To remedy this, we propose an uncertainty-aware pseudo-label selection to obtain more

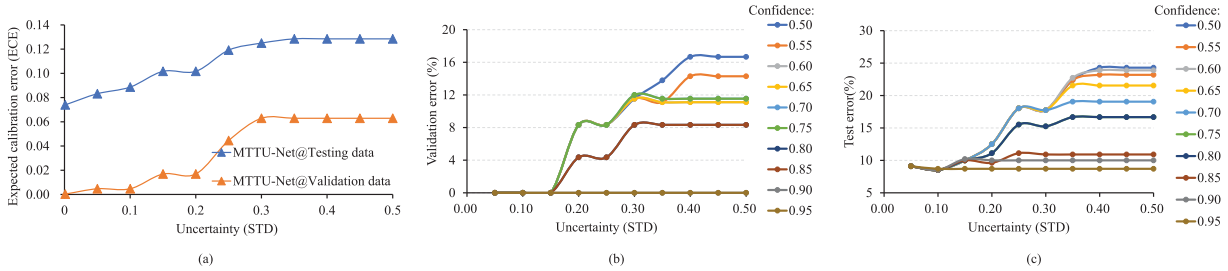

Fig. 3. (a) The relationship between prediction uncertainty and expected calibration error (ECE) on the validation data and testing data, respectively. The validation data has 30 subjects (12 IDH-mutant and 18 IDH-wild) from a part of training data. The testing data has 70 subjects (32 IDH-mutants and 38 IDH-wild). (b) The validation error across different uncertainty and confidence of UPS on the validation data. The curves with confidence levels of 0.6 and 0.65 overlap, as are 0.7 and 0.75, 0.8 and 0.85, and 0.9 and 0.95. (c) The test error across different uncertainty and confidence of UPS on the testing data. The curves with confidence levels of 0.75 and 0.8 overlap.

Algorithm 1 Semi-Supervised Multi-Task Learning

Input: Labeled data: D_L; Data without IDH label: D_U; Max Iterations
1: Train the proposed multi-task network, MTTU-Net, consisting of an encoder $\boldsymbol{E}_0$, a decoder $\boldsymbol{D}_0$, and a classifier $\boldsymbol{P}_0$, using the data from D_L.
2: **for** $i \leftarrow 1$ to Max Iterations **do**
3: Pseudo-labeling D_U using $\boldsymbol{E}_{i-1}$ and $\boldsymbol{P}_{i-1}$;
4: $\tilde{D}_U \leftarrow$ Select pseudo-labels using UPS;
5: $\tilde{D}_M \leftarrow D_L \bigcup \tilde{D}_U$;
6: Load the weights of $\boldsymbol{E}_{i-1}$, $\boldsymbol{D}_{i-1}$, and $\boldsymbol{P}_{i-1}$;
7: Train $\boldsymbol{E}_i$, $\boldsymbol{D}_i$, and $\boldsymbol{P}_i$ using the data from $\tilde{D}_M$.
8: $\boldsymbol{E}, \boldsymbol{D}, \boldsymbol{P} \leftarrow \boldsymbol{E}_i, \boldsymbol{D}_i, \boldsymbol{P}_i$
9: **end for**
10: **return** $\boldsymbol{E}, \boldsymbol{D}, \boldsymbol{P}$.

accurate subset $\tilde{D}_U$ of pseudo-labels and add it into the mixed dataset for further training. The mixed dataset is denoted by $\tilde{D}_M = D_L + \tilde{D}_U = \{(\mathbf{x}^{(i)}, \mathbf{g}^{(i)}, \tilde{\mathbf{y}}^{(i)})\}_{i=1}^{N_L+N_U}$, with $\tilde{\mathbf{y}}^{(i)} = \{y^{(i)} | y^{(i)} \in \{0, 1, -1\}\}_{i=1}^{N_L+N_U}$ for the IDH labels. It is worth noting that $y^{(i)} = -1$ indicates an incomplete sample i without IDH label, which does not perform back propagation in the IDH genotyping branch of MTTU-Net during the training. The training procedure for our proposed semi-supervised MTL framework based on UPS is described in Algorithm 1.

B. Uncertainty-Aware Pseudo-Label Selection

Confidence-based selection can reduce pseudo-label error rates, but the poor calibration of neural networks makes the solution insufficient. In other words, incorrect predictions have high confidence scores for poorly calibrated networks. Calibration can be interpreted as the overall prediction uncertainty of a network [40]. A recent study has also reported that selecting predictions with low uncertainty greatly reduces the impact of poor calibration [41]. To further verify this, we empirically investigate the relationship between prediction uncertainty and expected calibration error (ECE) [42] for IDH genotyping. ECE is a standard measurement for evaluating the network calibration, which is formulated as:

$$ECE = \sum_{l=1}^{L} \frac{|B_l|}{N} |\mathrm{Acc}(B_l) - \mathrm{Conf}(B_l)|, \tag{16}$$

$$\mathrm{Acc}(B_l) = \frac{1}{|B_l|} \sum_{i \in B_l} \mathbf{1}(\arg\max \hat{\mathbf{y}}^{(i)} = \arg\max \tilde{\mathbf{y}}^{(i)}), \tag{17}$$

$$\mathrm{Conf}(B_l) = \frac{1}{|B_l|} \sum_{i \in B_l} \max \hat{\mathbf{y}}^{(i)}, \tag{18}$$

where N is the number of samples on the testing data and the confidence predictions are divided into L equally spaced bins. $\mathrm{Acc}(B_l)$ and $\mathrm{Conf}(B_l)$ are the accuracy and confidence for each bin B_l. As shown in Fig. 3(a), we can observe that as the uncertainty decreases, the ECE score does decrease. Therefore, we leverage both uncertainty and confidence of network predictions to guide the pseudo-label selection procedure, which is called the uncertainty-aware pseudo-label selection (UPS), with the aim of obtaining a more accurate subset of pseudo-labels for training. MC-Dropout method [43] is used to obtain an uncertainty measure by computing the standard deviation of 10 stochastic forward passes. Let $\boldsymbol{\mu}(\cdot)$ be the prediction uncertainty, $g_c^{(i)} \in \{0, 1\}$ be a binary scalar denoting the selected pseudo-labels in sample i, where $g_c^{(i)} = 0$ when $\tilde{y}_c^{(i)}$ is not selected and $g_c^{(i)} = 1$ when $\tilde{y}_c^{(i)}$ is selected. The selection method can be denoted as follows:

$$g_c^{(i)} = \Phi[\boldsymbol{\mu}(p_c^{(i)}) \leq \tau_u] \Phi[p_c^{(i)} \geq \tau_c], \tag{19}$$

where $\Phi[\cdot]$ is a sign function that outputs 1 if $\cdot$ is true and 0 otherwise. τ_u and τ_c are the uncertainty threshold and confidence threshold, respectively. The thresholds are determined by using the proposed MTTU-Net model to evaluate the UPS error of different threshold combinations. As shown in Fig. 3(b) and Fig. 3(c), we can see that the error tends to decrease as the uncertainty threshold τ_u decreases, and the error becomes independent on all confidence thresholds when $\tau_u \leq 0.1$. Considering the tradeoff between the limited sample size and the performance, we chose appropriate thresholds $\tau_c = 0.9$ and $\tau_u = 0.05$.

For the IDH genotyping task, the selected subset of pseudo-labels will be calculated loss to update the model weights through backpropagation. The Eq. (14) is further modified for semi-supervised learning:

$$\mathcal{L}_{\mathrm{idh}} = -\frac{1}{m} \sum_{i=1}^{m} \sum_{c=1}^{C} g_c^{(i)} w_c \tilde{\mathbf{y}}_c^{(i)} \log(\hat{\mathbf{y}}_c^{(i)}), \tag{20}$$

where $g_c^{(i)}$ is mentioned in above, denoting the selected status of the pseudo-label for the ith sample. The usage of UPS to remove noise during training allows improving

TABLE I
SUMMARY OF THE DATASET USED IN THIS STUDY

	Training data	Additional data	Testing data
Subject n	148	221	70
Age median (range)	54 (18-84)	-	55 (23-80)
Sex			
Male	77 [52.03%]	14 [6.33%]	40 [57.14%]
Female	71 [47.97%]	5 [2.26%]	30 [42.86%]
Unknown	0	202 [91.41%]	0
Grade			
LGG	64 [43.24%]	12 [5.43%]	42 [60.00%]
HGG	84 [56.76%]	209 [94.57%]	28 [40.00%]
IDH status			
mutant	57 [38.51%]	0	32 [45.71%]
wild	91 [61.49%]	0	38 [54.29%]
Unknown	0	221 [100.00%]	0

the classification performance compared to the traditional pseudo-labeling. For the semi-supervised multi-task learning, the semi-supervised multi-task loss is adaptively adjusted by the same uncertainty weights as Eq. (15).

V. DATA AND EXPERIMENTS

A. Dataset

The multimodal MRI images of glioma patients are derived from the multimodal brain tumor segmentation (BraTS2020) challenge [44]. Part of BraTS2020 data belongs to The Cancer Imaging Archive (TCIA) [45], which can be available from our public repository. Genomic information is provided from The Cancer Genome Atlas (TCGA) [46]. This study is screened for the availability of IDH mutation status, fluid-attenuated inversion recovery imaging (FLAIR), T1-weighted imaging (T1), T1-weighted contrast-enhanced imaging (T1ce), and T2-weighted imaging (T2) modalities. The final dataset includes 1756 preoperative MRI images from 439 subjects with/without IDH mutation status. Among these data, 148 subjects with IDH mutation status from BraTS2020 training dataset are used as the training data, 70 subjects with IDH mutation status from BraTS2020 validation dataset are used as testing data, and 221 subjects without IDH mutation status from BraTS2020 training dataset are used as additional data for semi-supervised learning. Table I summarizes the data used in this study.

To reduce the heterogeneity among patients, all MRI images are preprocessed, including repositioning to the left-posterior-superior (LPS) coordinate system, co-registering with the T1 anatomical template, resampling to $1mm$ isotropic image resolution, and skull-stripping [44]. The ground truths of all imaging data are provided by BraTS2020 challenge. Three sub-regions including whole tumor (WT), tumor core (TC), and enhancing tumor (ET) are considered for evaluation.

B. Implementation Details

1) Training Process: The proposed network is implemented using Pytorch and is run on two NVIDIA V100 GPUs using a cross-validation training strategy that 80% of training data is used for training the model and 20% for tuning parameters. The proposed model employs an Adam optimizer with a decaying learning rate initialized at 2e-4. For the proposed model and other models, the bath size is set as 2 for each GPU and the maximum number of epochs is set to 1000. Besides, data augmentations including random rotation of -10 to 10 degree, random cropping of $128 \times 128 \times 128$ size, random flipping on three planes, and intensity shift with a factor of 0.1 are used.

2) Benchmark: To evaluate the effectiveness of the proposed method, we compare several benchmark experiments. The single-task models ClsNet and SegNet are trained as the IDH genotyping and glioma segmentation benchmark models, respectively. ClsNet consists of the encoder and classifier parts of the proposed network, while SegNet consists of the encoder and decoder parts of the proposed network. For the multi-task learning, we further compare two weighting method, i.e., equal weights and gradient surgery, known as PCGard [47]. We further compare the proposed method with existing glioma segmentation methods such as U-Net [10], V-Net [11], DMFNet [12], and UNETR [28], IDH genotyping methods such as SENet101 [48], ResNet50 [31] and DenseNet121 [33], and multi-task learning methods such as SGPNet [35], CMSVNet [36], and MA-MTLN [38].

For the semi-supervised multi-task learning, we first directly use the additional data to improve the performance of the network through supervised learning for segmentation branch and semi-supervised learning for classification branch. Further, we adopt the proposed UPS-based pseudo-label to improve classification performance by introducing more IDH label subset from additional data and compare it with traditional pseudo-labeling and confidence-based pseudo-labeling.

3) Evaluation Metrics: To evaluate the performance of the models, we employ Dice similarity score (Dice) and 95% Hausdorff distance (HD95) to quantitatively evaluate the glioma segmentation, and use the area under the curve (AUC) value, accuracy (Acc), sensitivity (Sens), and specificity (Spec) for quantitative evaluation of IDH genotyping. For the statistical analysis, DeLong test is used to compare the AUC of our proposed MTTU-Net with other methods and T-test is used to compare the means of two independent sample sets for Dice and HD95 metrics.

C. Ablation Study

1) Impact of the Number of Transformer Layers and Its Effectiveness: To investigate the impact of the number of transformer layers, we conduct ablation study using the proposed MTTU-Net with various number L of transformer layers. Table II shows the comparison results of transformer layers from 1 to 6. We can see that the proposed network with $L = 4$ transformer layers achieves the best segmentation performance for tumor core and enhancing tumor regions, as well as the best accuracy, sensitivity for IDH genotyping. Although the segmentation performance in whole tumor is moderate, there is not much gap between them. As the number of the transformer layer increases, the model parameters increases. Therefore, $L = 4$ is chosen as the optimal parameter of transformer layer. Furthermore, to demonstrate the effectiveness, we replace the transformer unit with four convolution blocks.

TABLE II
ABLATION STUDIES OF THE PROPOSED MTTU-NET WITH DIFFERENT NUMBER *L* OF TRANSFORMER LAYER

Layers	Glioma segmentation						IDH genotyping			
	Dice (mean±std) (%)			HD95 (mean±std) (mm)			AUC(%)	Acc (%)	Sens (%)	Spec (%)
	WT	TC	ET	WT	TC	ET				
1	88.43±16.23	74.61±25.23	70.40±34.60	5.56±10.19	8.90±12.58	5.51±9.92	90.16	81.43	59.38	**100.00**
2	89.61±10.59	74.88±22.48	70.15±34.64	5.12±7.96	7.68±6.75	5.33±7.43	89.47	82.86	78.13	86.84
3	89.44±9.62	71.06±28.96	72.48±31.73	7.17±13.20	7.69±6.28	5.99±10.33	89.14	82.86	75.00	89.47
4	89.84±9.16	**75.10±22.69**	**72.59±30.51**	4.94±5.67	**7.20±6.03**	**5.31±9.82**	90.37	**85.71**	**81.25**	89.47
5	**89.92±8.05**	74.39±25.88	71.53±31.86	**4.90±7.20**	8.06±7.86	6.73±12.90	90.30	81.43	65.63	94.74
6	89.24±10.74	73.20±27.36	72.14±31.88	5.87±7.93	10.26±14.27	5.65±10.36	**90.46**	80.00	65.63	92.11
Conv*	89.17±9.21	71.95±27.22	71.04±33.33	5.04±7.19	8.93±10.28	5.33±6.93	87.17	82.86	68.75	94.74

* Conv represents that the transformer unit of MTTU-Net is replaced with convolution operations.

TABLE III
ABLATION STUDIES OF SINGLE- AND MULTI-TASK LEARNING ON THE PROPOSED NETWORK

Method	Glioma segmentation						IDH genotyping			
	Dice (mean±std) (%)			HD95 (mean±std) (mm)			AUC(%)	Acc (%)	Sens (%)	Spec (%)
	WT	TC	ET	WT	TC	ET				
ClsNet	–	–	–	–	–	–	88.24	81.43	62.50	97.37
SegNet	88.61±13.12	74.46±24.34	68.65±35.46	6.63±12.84	8.38±10.27	6.86±14.78	–	–	–	–
MTTU-Net	**89.84±9.16**	**75.10±22.69**	**72.59±30.51**	**4.94±5.67**	**7.20±6.03**	**5.31±9.82**	**90.37**	**85.71**	**81.25**	89.47

TABLE IV
PERFORMANCE OF OUR PROPOSED METHOD USING DIFFERENT LEVEL-WISE FEATURE FUSIONS

Schemes		Glioma segmentation						IDH genotyping			
Method	Levels	Dice (mean±std) (%)			HD95 (mean±std) (mm)			AUC(%)	Acc (%)	Sens (%)	Spec (%)
		WT	TC	ET	WT	TC	ET				
MTTU-Net	5	89.07±10.17	67.61±30.28	69.56±35.55	7.33±12.84	9.14±9.03	4.94±5.70	84.29	78.57	56.25	97.37
MTTU-Net	2-5	89.27±9.34	72.41±27.01	73.76±30.14	8.25±13.76	10.83±16.31	6.79±14.66	86.18	77.14	50.00	100.00
MTTU-Net	3-5	89.19±9.71	72.53±27.88	74.13±28.79	6.79±11.55	10.26±12.77	6.87±12.56	88.40	80.00	59.38	97.37
MTTU-Net	4-5	**89.84±9.16**	**75.10±22.69**	**72.59±30.51**	**4.94±5.67**	**7.20±6.03**	**5.31±9.82**	**90.37**	**85.71**	**81.25**	89.47
MTTU-Net	3-4-5	88.16±9.59	71.67±26.68	71.81±32.05	12.96±19.77	10.99±15.13	8.19±16.20	87.25	84.29	68.75	97.37

The experimental results are shown in the last row of Table II, and its performance is inferior to that of the proposed MTTU-Net with 4 transformer layers. This illustrates that transformer can enhance feature representation by extracting the global semantic information.

2) Effectiveness of Multi-Task Learning Network: As shown in Table III, we confirm the superiority of our multi-task network compared with our two single-task networks. The comparison results show that the proposed network can further improve the performance of glioma segmentation and IDH genotyping. More precisely, for the IDH genotyping task, ClsNet achieves the performance with an AUC of 88.24%, an accuracy of 81.43%, a sensitivity of 62.50%, and a specificity of 97.37%. For the glioma segmentation task, SegNet achieves the performance with an average Dice of 88.61% and HD95 of 6.63 *mm* for whole tumor. While the proposed multi-task network, MTTU-Net, improves the Dice and HD95 of whole tumor by 1.23% and 1.69 *mm* for glioma segmentation, and improves the AUC and accuracy of 2.13% and 4.28%, respectively. Besides, the proposed MTTU-Net model also improves the segmentation performance in tumor core and enhancing tumor regions compared to the SegNet model. Therefore, we believe that these improvements benefit from multi-task learning, which can encourage feature sharing for the two tasks.

3) Effectiveness of the High-Level Feature Fusion: To clearly show the types of feature maps learned by the encoder, we visualize the output feature maps of each level of the encoder. As shown in Fig. 2, the high-level (level-4 and level-5) features summarize the tumor attributes, whereas the low-level (level-1 to level-3) features mainly shows the shape and boundary information. The high-level features are used for IDH genptyping by level-wise fusion. To illustrate the effectiveness of the high-level features fusion, we further compare it with different level-wise feature fusions. Table IV shows the comparison results. We can see that the high-level (level-4 and level-5) feature fusion shows better performance compared to other level-wise fusions. This also indicates that high level features can better distinguish the status of IDH mutations.

4) Effectiveness of the Uncertain Weights: To demonstrate the effectiveness of the uncertain weights, we compare it with equal weight and PCGard [47]. The equal weight-based approach is to assign the same weight to the two parts of losses. The PCGard is a multi-objective optimization method designed to alleviate the possible conflicting gradients between

TABLE V
PERFORMANCE OF OUR PROPOSED METHOD USING DIFFERENT WEIGHTING METHODS

Schemes		Glioma segmentation						IDH genotyping			
Method	Weighting	Dice (mean±std) (%)			HD95 (mean±std) (mm)			AUC (%)	Acc (%)	Sens (%)	Spec (%)
		WT	TC	ET	WT	TC	ET				
MTTU-Net	PCGard	88.69±10.49	71.26±23.08	65.54±35.23	5.29±7.35	9.03±11.29	8.63±10.76	85.69	82.86	65.63	**97.37**
MTTU-Net	Equal Weights	88.41±14.10	74.78±23.24	66.58±37.60	5.58±7.76	8.48±10.44	7.04±10.15	88.16	84.29	71.88	94.74
MTTU-Net	Uncert. Weights	**89.84±9.16**	**75.10±22.69**	**72.59±30.51**	**4.94±5.67**	**7.20±6.03**	**5.31±9.82**	**90.37**	**85.71**	**81.25**	89.47

TABLE VI
COMPARISON RESULTS OF OUR PROPOSED METHOD AND OTHER EXISTING STATE-OF-THE-ART METHODS

Schemes		Glioma segmentation						IDH genotyping			
Method	Additional data (w/o)	Dice (mean±std) (%)			HD95 (mean±std) (mm)			AUC (%)	Acc (%)	Sens (%)	Spec (%)
		WT	TC	ET	WT	TC	ET				
U-Net [10]	–	88.79±12.28	71.35±26.45	71.84±32.57	8.52±14.33	9.42±9.01	6.55±11.02	–	–	–	–
V-Net [11]	–	88.84±13.11	70.44±29.26	69.67±34.90	5.07±7.05	8.91±8.65	6.90±10.66	–	–	–	–
DMFNet [12]	–	89.13±12.24	74.62±19.30	**72.82±29.19**	6.91±12.46	8.84±10.45	6.70±11.80	–	–	–	–
UNETR [28]	–	87.42±14.17	68.68±29.39	70.39±32.92	9.56±15.39*	12.47±15.39*	7.20±11.85	–	–	–	–
SENet101 [48]	–	–	–	–	–	–	–	79.61*	77.14	65.63	86.84
DenseNet121 [33]	–	–	–	–	–	–	–	78.21*	77.14	65.63	86.84
ResNet50 [31]	–	–	–	–	–	–	–	77.55*	74.29	62.50	84.21
SGPNet [35]	–	87.04±12.52	69.33±27.55	71.43±32.52	11.17±15.23*	12.87±15.21*	8.89±14.50	81.25*	80.00	62.50	**94.74**
CMSVNet [36]	–	78.29±23.24*	58.43±31.76*	67.50±35.46	15.99±19.69*	17.39±20.22*	9.60±14.98*	75.99*	72.86	56.25	86.84
MA-MTLN [38]	–	88.00±11.68	69.70±28.87	67.71±34.16	7.20±12.07	9.85±11.12	8.02±16.21	82.89*	78.57	65.63	89.47
MTTU-Net	–	**89.84±9.16**	**75.10±22.69**	72.59±30.51	**4.94±5.67**	**7.20±6.03**	**5.31±9.82**	**90.37**	**85.71**	**81.25**	89.47
U-Net [10]	✓	89.57±12.18	77.42±22.07	73.90±30.01	7.50±13.91	8.38±10.64	6.12±10.69	–	–	–	–
V-Net [11]	✓	90.43±8.04	72.83±27.87	70.43±34.36	4.65±6.92	7.89±8.99	5.23±8.60	–	–	–	–
DMFNet [12]	✓	90.41±9.04	74.43±25.33	72.25±30.50	4.70±6.67	8.63±9.51	4.92±8.54	–	–	–	–
UNETR [28]	✓	89.02±12.72	69.08±30.90*	73.14±32.16	6.21±10.56	9.96±12.66	6.33±12.64	–	–	–	–
SGPNet† [35]	✓	89.82±7.96	72.83±26.94	73.99±30.84	10.84±19.32*	11.24±19.03	7.55±18.26	82.65*	80.00	65.63	92.11
CMSVNet† [36]	✓	90.00±8.05	68.10±29.34*	71.61±33.29	6.11±10.75	9.84±11.47	6.25±11.55	77.14*	75.71	65.63	84.21
MA-MTLN† [38]	✓	90.25±7.31	74.96±25.40	72.08±31.77	5.55±9.18	8.33±9.39	5.39±7.66	90.21	82.86	71.88	92.11
MTTU-Net†	✓	90.46±8.91	76.99±22.91	74.31±31.05	4.60±5.94	7.05±6.41	4.84±9.66	90.49	88.57	84.38	92.11
MTTU-Net(UPS‡)	✓	**90.71±7.44**	**78.61±21.80**	**75.89±29.01**	**4.45±5.34**	**6.61±8.52**	**4.49±7.86**	**91.04**	**90.00**	**87.50**	**92.11**

† The method is trained by directly introducing the additional data for semi-supervised IDH genotyping.
‡ The method is trained by pseudo-labels generated by the UPS from additional data for semi-supervised IDH genotyping.
* The marker indicates that the comparison performance between our proposed MTTU-Net and other methods is statistically significant.

two tasks. The comparison results are shown in Table V. We can see that the proposed MTTU-Net using uncertain weight achieves better segmentation performance compared with the network using other weights. Meanwhile, the method also has a better sensitivity to IDH-mutant case and achieve a better AUC and accuracy for IDH genotyping task. Therefore, the uncertain weight is helpful to balance the glioma segmentation and IDH genotyping tasks, which avoids a certain task dominating the training process. Moreover, additional experiments are performed to demonstrate the effectiveness of the adopted losses by utilizing the focal loss to replace either or both of the Dice coefficient loss and weighted cross-entropy loss.

D. Comparison With the State-of-the-Art Methods

1) Glioma Segmentation Methods: To demonstrate the superiority of the proposed network, we compare it with multiple existing segmentation methods including U-Net [10], V-Net [11], DMFNet [12], UNETR [28], SGPNet [35], CMSVNet [36], and MA-MTLN [38]. The segmentation performance of the proposed method and other segmentation methods are presented in the top half of Table VI. Experimental results show that our proposed MTTU-Net network outperforms these existing single- and multi-task segmentation methods in glioma segmentation. Specifically, without using additional data, our proposed MTTU-Net achieves an average Dice of 89.84% and HD95 of 4.94 *mm* for the whole tumor, an average Dice of 75.10% and HD95 of 7.20 *mm* for the tumor core, and an average Dice of 72.59% and HD95 of 5.31 *mm* for the enhancing tumor. Among all the segmentation comparison methods, most of the measurements obtained by our method are the best. It is worth mentioning that the recently proposed UNETR is based on a pure transformer architecture as an encoder, which achieves a relatively inferior segmentation performance compared with our hybrid CNN-Transformer architecture. This also illustrates that our proposed MTTU-Net enhances feature representation by combining convolution operations and transformer.

2) IDH Genotyping Methods: Table VI also shows that our proposed method yields quite remarkable classification performance in the IDH genotyping task. Specifically, without using additional data, our proposed MTTU-Net achieves an AUC of 90.37%, an accuracy of 85.57%, a sensitivity of 81.25%, and a specificity of 89.47%. Compared with the single-task methods such as SENet101 [48], ResNet50 [31] and DenseNet121 [33], our proposed method shows a better classification performance

TABLE VII
PERFORMANCE OF OUR PROPOSED METHOD FOR SEMI-SUPERVISED LEARNING

Schemes		Glioma segmentation						IDH genotyping			
Method	Pseudo-labeling	Dice (mean±std) (%)			HD95 (mean±std) (mm)			AUC (%)	Acc (%)	Sens (%)	Spec (%)
		WT	TC	ET	WT	TC	ET				
MTTU-Net	PL	90.27±7.61	77.56±22.58	74.24±30.30	6.71±7.25	7.48±10.11	5.07±10.56	88.98	85.71	78.13	92.11
MTTU-Net	CPL	89.83±8.87	76.78±23.26	74.63±30.01	8.48±15.24	10.58±16.21	5.33±10.74	90.29	85.71	78.13	92.11
MTTU-Net	UPS	**90.71±7.44**	**78.61±21.80**	**75.89±29.01**	**4.45±5.34**	**6.61±8.52**	**4.49±7.86**	**91.04**	**90.00**	**87.50**	**92.11**

in terms of AUC, accuracy and sensitivity. The architecture of these single-task networks directly learns feature-related IDH genotyping from the input images, which may capture some unreliable features. While our designed MTTU-Net has a tumor decoding branch, so that IDH genotype-related features can be derived from the tumor regions as much as possible. Compared with the multi-task methods such as SGPNet [35], CMSVNet [36], and MA-MTLN [38], our proposed MTTU-Net also shows better predictive performance, demonstrating the ability of the multi-scale feature fusion to predict IDH mutations.

E. Comparison With Semi-Supervised Multi-Task Learning

The additional data without IDH labels are added to the training to allow multi-task networks to enhance the segmentation performance while improving classification performance. A naive approach is to mix the two parts of data and directly train the network by supervising the segmentation task to enhance the coding ability of the network. The half bottom of Table VI shows the comparison results of the networks trained with the mixed data. We can see that the segmentation performance of single-task networks such as U-Net [10], V-Net [11], DMFNet [12], and UNETR [28], has also been significantly improved with the introduction of additional data. More importantly, the segmentation and classification performance of our proposed multi-task network, MTTU-Net and other networks such as SGPNet [35], CMSVNet [36], and MA-MTLN [38] have been improved. Our MTTU-Net achieves a promising segmentation performance with an average Dice of 90.46 and HD95 of 4.84 *mm* for whole tumor, and obtains a better classification performance with an AUC of 90.49%, an accuracy of 88.57%, a sensitivity of 84.38, and a specificity of 92.11%, which outperforms other multi-task networks. Although there is no more IDH genotype labeled data, the multi-task network can enhance feature coding ability by semi-supervised learning, and further promote the improvement of classification performance. When self-training based on UPS is adopted, our proposed MTTU-Net further improves the segmentation and classification performance. The proposed method yields an average Dice of 90.71% and HD95 of 4.45 *mm* for whole tumor, and obtains an AUC of 91.04%, an accuracy of 90.00%, a sensitivity of 87.50, and a specificity of 92.11% for IDH genotyping. Besides, Fig. 4 shows two cases of 3D visualization of the surface distance between segmented surface and ground truth among the segmentation methods. The larger the green area, the closer the segmentation result is to the ground truth.

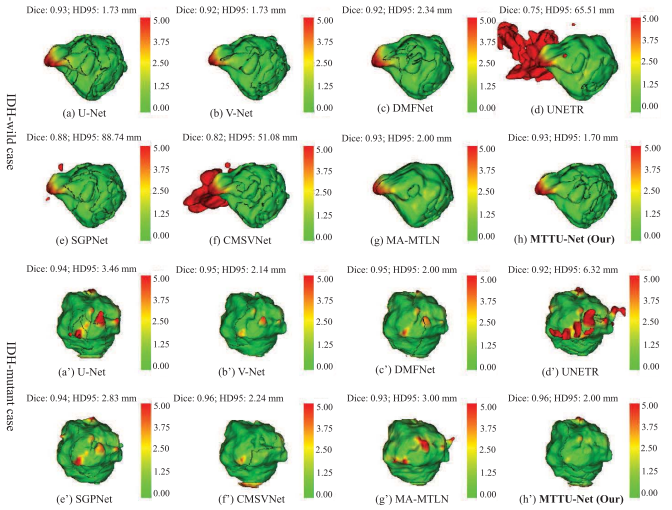

Fig. 4. 3D visualization of the surface distance (in voxel) between segmented surface and ground truth. Different colors represent different surface distances and the distance is compressed in a range of 0 to 5. Two cases (an IDH-wild case and an IDH-mutant case) are randomly selected to present the segmentation results obtained by (a/a') U-Net [10], (b/b') V-net [11], (c/c') DMFNet [12], (d/d') UNETR [28], (e/e') SGPNet [35], (f/f') CMSVNet [36], (g/g') MA-MTLN [38], (h/h') MTTU-Net (Our), respectively. Our method consistently performances well on the whole tumor surface.

F. Comparison With Different Pseudo-Labeling Methods

To demonstrate the superiority of UPS, we compare it with traditional pseudo-labeling (PL) and confidence-based pseudo-labeling (CPL). The comparison results are shown in Table VII. We can see that PL-based and CPL-based methods achieve relatively inferior classification performance, with an accuracy of 85.71%. This may be because the uncertainty of the model using the PL-based method will generate a large number of false labels, resulting in noise training. Although CPL-based method can reduce some false labels via confidence threshold, it may still produce false labels with high confidence. Therefore, the performance of IDH genotyping is not significantly improved, and the sensitivity of IDH mutants is reduced when compared with MTTU-Net without using additional data. Our UPS-based method overcomes this challenge by initially selecting a smaller, more accurate subset, and gradually increasing the number of selected pseudo-labels while maintaining high accuracy.

VI. Discussion

Recently, radiogenomics has emerged in cancer research, focusing on the relationship between imaging phenotypes and genomics [8]. This technique is expected to be an alternative to standard invasive biopsy approaches for the determination of IDH status in gliomas. In this study, we have proposed a multi-task deep learning network called MTTU-Net, which uses conventional multimodal MRI to predict IDH mutations in pathological molecules while automatically performing glioma segmentation. Moreover, we have introduced a semi-supervised multi-task Learning framework to exploit the potential value of unlabeled data for further improving the performance of IDH genotyping and glioma segmentation.

We have demonstrated that the proposed multi-task learning network can improve glioma segmentation and IDH genotyping results by jointly learning two tasks. Compared with the single-task learning networks, the proposed MTTU-Net obtains huge gains in both glioma segmentation and IDH genotyping (Tables III and VI). This illustrates that multi-task learning can make the shared representations achieve more accurate precise localization and accurate classification of tumor. The segmentation task is dedicated to the localization of tumor regions, while the classification task is dedicated to achieving the volume-level classification of IDH genotypes from tumor representations. Therefore, the localization and representation of tumor are strongly related to IDH genotyping task, while homogeneity representations information of IDH genotypes also contributes to tumor segmentation task. As shown in Table III, our proposed multi-task network, MTTU-Net, outperforms the single-task learning counterparts, demonstrating that the performance can be improved by using the sharing feature representation between the two tasks. We have also performed the ablation studies using different modality combinations as inputs to the network, we find that the network using all available modalities performs better than other combinations on the two tasks. Furthermore, our proposed method shows better segmentation and classification performance compared to other existing multi-task networks (Table VI).

We have also demonstrated that the proposed MTTU-Net can further improve segmentation and classification performance by semi-supervised multi-task learning. Semi-supervised learning is one of the most dominant methods to overcome the dependence on huge labeled datasets. In this study, we adopt the ideal of semi-supervised learning and propose a UPS-based method to make full use of a large amount of dataset without IDH genotypes in the real scenario. In general, semi-supervised learning, especially pseudo-labeling, are usually only effective if given a very large unlabeled sample size. Although our unlabeled sample size is not much larger than the labeled sample size, the semi-supervised IDH genotyping can benefit from multi-task learning, which alleviates the requirement due to the complementarity of the two tasks. Fig. 5 demonstrates the effectiveness of the UPS in selecting pseudo-labels. We can see that the UPS has a sensitivity of 100% for selecting IDH-mutant subjects. Furthermore, the performance of IDH genotyping is improved by using the

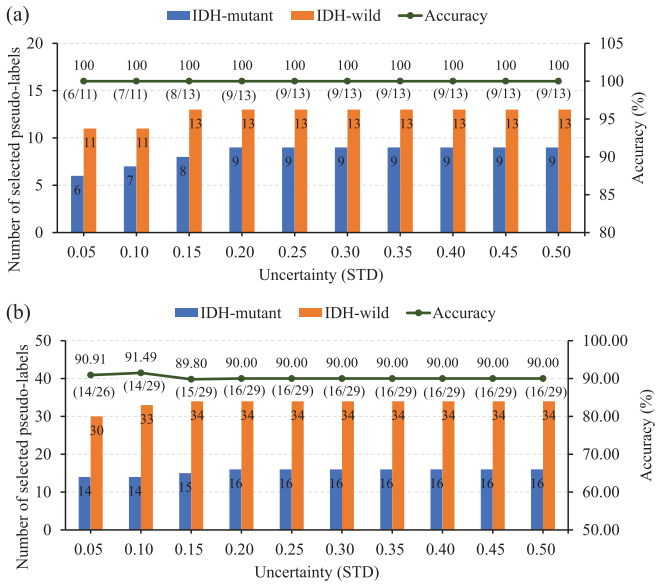

Fig. 5. (a) Comparison of the number of pseudo-label selection and the selection accuracy on the validation data when confidence is 0.9. (b) Comparison of the number of pseudo-label selection and the selection accuracy on the testing data when confidence is 0.9. The (m/n) means the correct number pair to be selected from the IDH-mutant and IDH-wild. In comparison, the accuracy of traditional pseudo-labeling is 83.33% and 85.71% on the validation data and testing data, respectively.

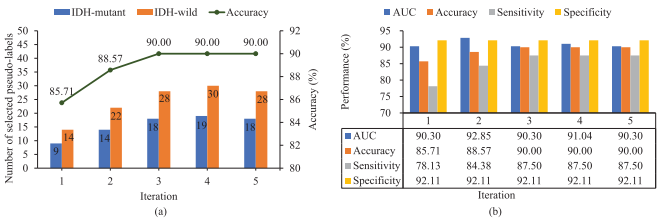

Fig. 6. (a) Comparison of the number of pseudo-label selection and the testing accuracy of IDH genotyping for each iteration during self-training with active learning. (b) Comparison performance of the proposed method for each iteration during self-training with active learning.

proposed UPS with active learning. It is worth noting that the number of IDH-wild subjects selected by UPS in the additional data is much larger than that of IDH-mutant subjects selected, which may cause data imbalance. To alleviate this problem, we balance the number of selected pseudo-labels with a ratio of the same proportions (1:1.6) as the IDH-mutant and IDH-wild subjects in the training data. Fig. 6 shows the number of selected pseudo-labels with testing accuracy and classification performance for each iteration during active learning. We find that the performance of IDH genotyping tends to be stable gradually after three pseudo-labeling iterations, and the best accuracy reaches 90.00%.

Our study has some limitations. First, the data used in this study are a multi-institutional public dataset, but the number of testing samples is small. Future work will focus on validating the performance of the model on larger external independent test data. Second, this study is limited to the prediction of IDH mutations. Next work will be dedicated to the prediction of more pathological molecules such as 1p/19q

co-deletions [5], MGMT promoter status, ATRX status. Third, multiple reports have suggested that T2-FLAIR mismatch sign represents a highly specific imaging biomarker for IDH-mutant 1p19q non-codeleted gliomas in LGG [3], [49]. Inspired by those literatures, we will carry out research for validation in further work.

VII. Conclusion

In conclusion, we propose a multi-task learning network termed as MTTU-Net for simultaneous IDH genotyping and glioma segmentation. To further explore the potential value of unlabeled data, we employ an uncertainty-aware pseudo-label selection for generating IDH pseudo-labels and propose a semi-supervised multi-task learning framework to further improve the performance of IDH genotyping and glioma segmentation. The proposed framework is evaluated and validated on a multi-institutional public dataset. Experimental comparison results show that our proposed method achieves encouraging performance and outperforms the other state-of-the-art methods. It is believed that our proposed framework can be used as a reliable computer-aided genotyping system for the prediction using multimodal MRI data.

References

[1] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomed. Signal Process. Control*, vol. 39, pp. 139–161, Jan. 2018.

[2] D. N. Louis *et al.*, "The 2016 World Health Organization classification of tumors of the central nervous system: A summary," *Acta Neuropathol.*, vol. 131, no. 6, pp. 803–820, Jun. 2016.

[3] S. H. Patel *et al.*, "T2–FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: A TCGA/TCIA project," *Clin. Cancer Res.*, vol. 23, no. 20, pp. 6078–6085, 2017.

[4] J. Cheng *et al.*, "Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading," *IEEE J. Biomed. Health Informat.*, early access, Jul. 8, 2021, doi: 10.1109/JBHI.2021.3095476.

[5] R. L. Delfanti *et al.*, "Imaging correlates for the 2016 update on WHO classification of grade II/III gliomas: Implications for IDH, 1p/19q and ATRX status," *J. Neuro-Oncol.*, vol. 135, no. 3, pp. 601–609, Dec. 2017.

[6] J. Cheng, J. Liu, H. Yue, H. Bai, Y. Pan, and J. Wang, "Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric MRI images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Oct. 26, 2020, doi: 10.1109/TCBB.2020.3033538.

[7] P. Lambin *et al.*, "Radiomics: The bridge between medical imaging and personalized medicine," *Nature Rev. Clin. Oncol.*, vol. 14, no. 12, pp. 749–762, Dec. 2017.

[8] G. Singh *et al.*, "Radiomics and radiogenomics in gliomas: A contemporary update," *Brit. J. Cancer*, vol. 125, pp. 641–657, May 2021.

[9] S. Narang, M. Lehrer, D. Yang, J. Lee, and A. Rao, "Radiomics in glioblastoma: Current status, challenges and potential opportunities," *Transl. Cancer Res.*, vol. 5, no. 4, pp. 383–397, Aug. 2016.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] 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.

[12] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3D dilated multi-fiber network for real-time brain tumor segmentation in MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 184–192.

[13] J. Maynard *et al.*, "World health organization grade II/III glioma molecular status: Prediction by MRI morphologic features and apparent diffusion coefficient," *Radiology*, vol. 296, no. 1, pp. 111–121, Jul. 2020.

[14] A. Dosovitskiy *et al.*, "An image is worth 16 × 16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–22. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[15] M. Gupta *et al.*, "Brain tumor segmentation by integrating symmetric property with region growing approach," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–5.

[16] H.-C. Shin, "Hybrid clustering and logistic regression for multi-modal brain tumor segmentation," in *Proc. Workshops Challanges Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2012, pp. 32–35.

[17] M. Wels *et al.*, "A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-D MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2008, pp. 67–75.

[18] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, "Efficient multilevel brain tumor segmentation with integrated Bayesian model classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 629–640, Apr. 2008.

[19] D. Kwon, R. T. Shinohara, H. Akbari, and C. Davatzikos, "Combining generative models for multifocal glioma segmentation and registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2014, pp. 763–770.

[20] Z. Liu *et al.*, "CANet: Context aware network for brain glioma segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 7, pp. 1763–1777, Jul. 2021.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[22] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: A dimension-fusion U shape network for chronic stroke lesion segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 940–950, May 2021.

[23] K. Qi *et al.*, "X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 247–255.

[24] J. Cheng, J. Liu, L. Liu, Y. Pan, and J. Wang, "Multi-level glioma segmentation using 3D U-Net combined attention mechanism with atrous convolution," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 1031–1036.

[25] 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, Dec. 2020.

[26] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NIPS*, 2017, pp. 6000–6010.

[27] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. Goh, "Medical image segmentation using squeeze-and-expansion transformers," in *Proc. 30th Int. Joint Conf. Artif. Intell. (IJCAI)*, Aug. 2021, pp. 807–815.

[28] A. Hatamizadeh *et al.*, "UNETR: Transformers for 3D medical image segmentation," 2021, *arXiv:2103.10504*.

[29] Y. Ren *et al.*, "Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features," *J. Magn. Reson. Imag.*, vol. 49, no. 3, pp. 808–817, Mar. 2018.

[30] S. Wu, J. Meng, Q. Yu, P. Li, and S. Fu, "Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas," *J. Cancer Res. Clin. Oncol.*, vol. 145, no. 3, pp. 543–550, Mar. 2019.

[31] K. Chang, H. X. Bai, H. Zhou, and C. Su, "Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR imaging," *Clin. Cancer Res.*, vol. 24, no. 5, pp. 1073–1081, 2018.

[32] Y. Matsui *et al.*, "Prediction of lower-grade glioma molecular subtypes using deep learning," *J. Neuro-Oncol.*, vol. 146, no. 2, pp. 321–327, Jan. 2020.

[33] S. Liang *et al.*, "Multimodal 3D DenseNet for IDH genotype prediction in gliomas," *Genes*, vol. 9, no. 8, p. 382, Jul. 2018.

[34] C. G. Bangalore Yogananda *et al.*, "A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas," *Neuro-Oncol.*, vol. 22, no. 3, pp. 402–411, 2020.

[35] Y. Wang, Y. Wang, C. Guo, S. Zhang, and L. Yang, "SGPNet: A three-dimensional multitask residual framework for segmentation and IDH genotype prediction of gliomas," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–9, Apr. 2021.

[36] Y. Zhou *et al.*, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101918.

[37] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2482–2493, Dec. 2020.

[38] Y. Zhang *et al.*, "3D multi-attention guided multi-task learning network for automatic gastric tumor segmentation and lymph node classification," *IEEE Trans. Med. Imag.*, vol. 40, no. 6, pp. 1618–1631, Jun. 2021.

[39] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491.

[40] B. Lakshminarayanan *et al.*, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6405–6416.

[41] M. N. Rizve, K. Duarte, Y. S. Rawat, and M. Shah, "In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–20. [Online]. Available: https://openreview.net/forum?id=-ODN6SbiUU

[42] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1321–1330.

[43] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[44] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, Sep. 2017, Art. no. 170117.

[45] K. Clark *et al.*, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, 2013.

[46] M. Ceccarelli *et al.*, "Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma," *Cell*, vol. 164, no. 3, pp. 550–563, 2016.

[47] T. Yu *et al.*, "Gradient surgery for multi-task learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33. Red Hook, NY, USA: Curran Associates, 2020, pp. 5824–5836.

[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[49] R. Jain *et al.*, "'Real world' use of a highly reliable imaging sign: 'T2-FLAIR mismatch' for identification of IDH mutant astrocytomas," *Neuro-Oncol.*, vol. 22, no. 7, pp. 936–943, Jul. 2020.