University
of Dundee

University of Dundee

Expectation-Maximization Regularized Deep Learning for Weakly Supervised Tumor Segmentation for Glioblastoma

Li, Chao; Huang, Wenjian; Chen, Xi; Wei, Yiran; Price, Stephen J.; Schönlieb, Carola-Bibiane

Expectation-Maximization Regularized Deep Learning for Weakly Supervised Tumor Segmentation for Glioblastoma

Chao Li[*]
Department of Clinical Neurosciences
University of Cambridge

Wenjian Huang[*†]
Department of Computer Science and Engineering
Southern University of Science and Technology

Xi Chen
Department of Computer Science
University of Bath

Yiran Wei
Department of Clinical Neurosciences
University of Cambridge

Stephen J. Price
Department of Clinical Neurosciences
University of Cambridge

Carola-Bibiane Schönlieb
Department of Applied Mathematics and Theoretical Physics
University of Cambridge

Abstract

We present an Expectation-Maximization (EM) Regularized Deep Learning (EM-ReDL) model for the weakly supervised tumor segmentation. The proposed framework was tailored to glioblastoma, a type of malignant tumor characterized by its diffuse infiltration into the surrounding brain tissue, which poses significant challenge to treatment target and tumor burden estimation based on conventional structural MRI. Although physiological MRI can provide more specific information regarding tumor infiltration, the relatively low resolution hinders a precise full annotation. This has motivated us to develop a weakly supervised deep learning solution that exploits the partial labelled tumor regions.

EMReDL contains two components: a physiological prior prediction model and EM-regularized segmentation model. The physiological prior prediction model exploits the physiological MRI by training a classifier to generate a physiological prior map. This map was passed to the segmentation model for regularization using the EM algorithm. We evaluated the model on a glioblastoma dataset with the available pre-operative multiparametric MRI and recurrence MRI. EMReDL was shown to effectively segment the infiltrated tumor from the partially labelled region of potential infiltration. The segmented core and infiltrated tumor showed high consistency with the tumor burden labelled by experts. The performance comparison showed that EMReDL achieved higher accuracy than published state-of-the-art models. On MR spectroscopy, the segmented region showed more aggressive features than other partial labelled region. The proposed model can be generalized to other segmentation tasks with partial labels, with the CNN architecture flexible in the framework.

---

[*]Equal contribution
[†]Current affiliation

Preprint. Under review.

1 Introduction

Glioblastoma is the most common malignant primary brain tumor, characterized by poor outcomes (Wen et al., 2020). The first-line treatment includes maximal safe resection followed by chemoradiotherapy (Stupp et al., 2005), which requires an accurate tumor delineation to enhance the treatment efficacy and reduce the neurological deficits of patients (Mazzara et al., 2004; Stupp et al., 2005). As the manual delineation is often subjective and laborious, an automated tumor segmentation model is crucial in aiding clinical practice. Currently, Magnetic Resonance Imaging (MRI) is the mainstay for diagnosis, treatment planning, and disease monitoring of glioblastoma (Weller et al., 2014, 2017; Wen et al., 2020) . It however remains a challenge to accurately segment the glioblastoma based on MRI (Wadhwa et al., 2019), mainly due to several reasons. Firstly, glioblastoma is characterized by diffuse infiltration into the surrounding brain, leading to a poorly demarcated tumor margin. Secondly, glioblastoma is highly heterogeneous with regard to the tumor location, morphology and intensity values. Thirdly, glioblastoma may demonstrate similar appearance with neurodegenerative or white matter pathologies. All of the above may pose significant challenges to a robust segmentation model.

Incorporating multiple MRI modalities is considered beneficial for tumor segmentation (Ghaffari et al., 2020). Clinically, the most commonly used sequences include T1-weighted, T2-weighted, post-contrast T1-weighted (T1C), and fluid attenuation inversion recovery (FLAIR) sequences. A multimodal brain tumor image segmentation (BraTS) challenge represents the collective efforts to develop segmentation models using a large glioblastoma dataset with multiple MRI sequences available (Bakas et al., 2018). A wide spectrum of models has since been proposed with dramatic success in performance (Ghaffari et al., 2020). Among these models, deep learning shows unique advantages in using multiple MRI sequences for tumor segmentation, compared to the traditional methods of using hand-crafted features. However, the BraTS dataset only includes the most widely used structural sequences, which was shown to be prone to the low specificity in targeting actual tumor infiltration (Verburg et al., 2020). Particularly, for the non-enhancing lesion beyond the contrast-enhancing margin, it remains challenging to differentiate the infiltrated tumor from edema, even combining all the structural sequences (Verburg et al., 2020). An effective imaging model with higher specificity in segmenting the infiltrated tumor is of crucial value for clinical decision making.

An increasing amount of literature provides evidence that physiological MRI can facilitate the characterization of tumor infiltration (Li et al., 2019a; Yan et al., 2019). In particular, diffusion and perfusion MRI can identify the infiltrated tumor beyond the contrast enhancement by offering parametric measures describing tumor physiology, which may complement the non-specificity of the structural sequences. Specifically, The diffusion MRI is the only imaging method of describing brain microstructure by measuring water molecule mobility (Jellison et al., 2004), which can detect the subtle infiltration (Li et al., 2019b), characterize tumor invasiveness (Li et al., 2019d) and predict tumor progression (Yan et al., 2020). On the other hand, as a widely used perfusion technique, dynamic susceptibility contrast (DSC) imaging can derive the relative cerebral blood volume (rCBV), mean transit time (MTT) and relative cerebral blood flow (rCBF), reflecting the aberrant tumor vascularization (Lupo et al., 2005). Therefore, integrating physiological MRI into the tumor segmentation model shows potential to more accurately identify tumor infiltration.

Here we proposed a deep learning model to automatically segment the core and infiltrated tumor based on both structural and physiological multiparametric MRI. We hypothesized that the physiological MRI information of the core tumor could be used to guide the deep learning model to segment the infiltrated tumor beyond the core tumor. In the next section, we summarize the related work of tumor segmentation, including both supervised and weakly supervised models.

2 Related work

Tumor segmentation is an active research field with a growing number of models proposed. These models can be generally classified into generative or discriminative models (Ghaffari et al., 2020). Typically, generative models rely on the prior knowledge of the voxel distributions of the brain tissue, which is derived from the probabilistic atlas (Prastawa et al., 2004), whereas the discriminative models rely on the extracted image features that could be mapped to the classification labels. In general, discriminative models show superior performance than generative models. Most successful discriminative approaches in the BraTS challenge (Menze et al., 2015) are based on fully supervised convolutional neural networks (CNN).

In BraTS 2014, a CNN-based model was firstly introduced. The top-ranked algorithm employed a 3D CNN model trained on small image patches, which consisted of four convolutional layers with six filters in the last layer corresponding to six labels (Urban et al., 2014). In BraTs 2015, a 2D CNN model with a cascaded architecture was proposed. Two parallel CNNs were employed to extract local and global features which were then concatenated and fed into a fully connected layer for classification (Dutil et al., 2015). In BraTS 2016, DeepMedic, a 3D CNN model of eleven layers with residual connections was proposed. Two pathways were employed to process the inputs in parallel, to increase the receptive field of the classification layer (Kamnitsas et al., 2016). In BraTS 2017, the Ensembles of Multiple Models and Architectures (EMMA) separately trained several models (DeepMedics, 3D FCN, and 3D U-net) using different optimization approaches, while the output was defined as the average to reduce bias from individual models (Kamnitsas et al., 2017). The top-ranked model in BraTS 2018 proposed an asymmetric U-net architecture, where an additional variational auto-encoder branch was added to the shared encoder, providing additional regularization (Myronenko, 2018; Warrington et al., 2020). In BraTS 2019, the top-ranked model proposed a two-stage cascaded U-Net (Jiang et al., 2019). The first stage used a U-Net variant for preliminary prediction, whereas the second stage concatenated the preliminary prediction map with the original input images to refine the prediction.

In summary, the above top-ranked models from the BraTS depict the advantages of CNN-based segmentation model, which highlights the capacity of feature extraction of CNN. Further, to enhance the model performance or reduce the computational cost, various techniques were employed to improve the backbone CNN by a series of procedures, e.g., increasing network depth or width, optimizing the loss function, increasing receptive fields, or adopting an ensemble model. For more details of the BraTS models, please refer to (Bakas et al., 2018; Ghaffari et al., 2020). All these state-of-the-art models heavily rely on the full classification labels to train a model that could approximate the accuracy of experts. The infiltrative nature of glioblastoma, however, poses significant challenges to accurate delineation of the interface between tumor and healthy tissue. Although the binary contrast-enhancement provided a reference for "core tumor", the surrounding non-enhancing region, regarded as the edema in BraTS labels, has established as diffusely infiltrated with tumor.

As outlined in the previous section, multiparametric MRI allows more accurate identification of the non-enhancing infiltrated tumor. Nevertheless, the low resolution of physiological MRI hinders the precise annotation based on these images. A full annotation based on physiological MRI therefore is prone to the subjective errors, even by experienced clinical experts. As a result, those models with high reliance on the full labels may not be suitable for segmented the infiltrated tumor.

Other studies investigated the feasibility of delineating tumor infiltration based on the weak labels of cancerous and healthy tissues. (Akbari et al., 2016) proposed a tumor infiltration inference model using the physiological and structural MRI (Akbari et al., 2016). Two types of weak labels were used, i.e., one scribble immediately adjacent to the enhancing tumor and another scribble near the distal margin of the edema. These two scribble regions, representing the tissue near and far from the core tumor respectively, were hypothesized to correspondingly have higher and lower tumor infiltration. The classifier was trained based on the weak labels using the support vector machine (SVM) which yielded a voxelwise infiltration probability. The model achieved excellent performance and was subsequently validated by another cohort and the tumor recurrence on the follow-up scans.

Although in relatively small sample size, this study underpinned the advantage of physiological MRI in identifying tumor infiltration and supported the feasibility of weakly supervised learning models to tackle the challenge of lacking precise full annotations. The proposed model, however, ignored the spatial continuity of tumor infiltration. The CNN model could empower the weakly supervised learning model (Chan et al., 2020) by effectively extracting multiparametric MRI features with spatial information.

Training a weakly supervised CNN model using a partial cross-entropy loss may lead to poor boundary localization of saliency maps (Zhang et al., 2020). To mitigate this limitation, additional regularization is often employed. For instance, (Tang et al., 2018)introduced a normalized cut loss as a regularizer with a partial cross-entropy loss. (Kervadec et al., 2019) introduced a regularization term constraining the size of the target region that was combined with a partial cross-entropy loss. (Roth et al., 2019) used the random walker algorithm to generate the pseudo full label from the partial labels and then constructed the regularized loss by enforcing the CNN outputs to match the pseudo labels. The results of above studies supported the usefulness of additional regularizers in the weakly

supervised models. Due to the advantages of physiological MRI in detecting tumor infiltration, here we hypothesized that a regularizer from the physiological MRI could enhance the weakly supervised model for segmenting the infiltrated tumor by incorporating domain-specific information.

We sought to propose a CNN-based weakly supervised model, in which a regularization term was constructed by incorporating the prior information obtained from the physiological MRI by an prediction model through an expectation-maximization (EM) framework. We evaluated the model validation using tumor recurrence on follow-up scans and MR spectroscopy that non-invasively measures the metabolic alternation. The remainder of this paper is organized as follows: Section 3 will describe the overall study design, main components of the proposed framework and the performance evaluation of the model. Section 4 gives details of the dataset and the implementation of the experiments. Section 5 will provide the results and discussion followed by the conclusions in Section 6.

3 Methods

3.1 Notation

Consider the multiparametric MRI from N (patients) training samples $\mathbf{X} = \{X_1, X_2, \cdots, X_N\}$, including both structural sequences (T1-weighted, T2-weighted, T1C and FLAIR) and physiological sequences (diffusion and perfusion MRI), denoted as $\mathbf{X}_s$ and $\mathbf{X}_p$, respectively. From a clinical perspective, three regions of interest (ROI) can be delineated (Figure 1):

- ○ ROI1: core tumor, which is the contrast-enhancing tumor region on T1C images and the surgery target for clinical practice;
- ○ ROI2: potential infiltrated region, which is the hyperintensities in FLAIR images outside of ROI1. We are specifically interested in this region as it represents the clinically extendable treatment target;
- ○ ROI3: normal-appearing region on both T1C and FLAIR sequences.

All MRI sequences have been co-registered. The voxel labels can be classified into *observed labels* $\mathbf{y}_o$ and *unobserved labels* $\mathbf{y}_u$. A voxel label y is a value either 0 or 1, and $\mathbf{y}_o$ indicates the labels of ROI1 and ROI3, where $y = 1$ indicates a confirmed tumor voxel and $y = 0$ represents a voxel from the normal-appearing brain region. The $\mathbf{y}_u$ indicates label of ROI2. Given $\mathbf{X}$, we aimed to simultaneously segment the core tumor (ROI1) and the peritumoral infiltrated tumor in ROI2.

3.2 Overview of the proposed method

Our goal was to segment the core and infiltrated tumor using the model trained by the existing MRI data $\mathbf{X}$ and its corresponding observed labels $\mathbf{y}_o$. For the standard supervised CNN models, full training labels are necessary to be used as the 'ground-truth' to train the weights $\boldsymbol{\theta}$ of the CNN. In our proposed application, however, as it is not possible to obtain a full annotation for the unobserved labels $\mathbf{y}_u$, which renders a supervised CNN training inappropriate. In this paper, we cast the underlying problem into a weakly supervised learning problem by leveraging the EM algorithm, which can recursively estimate both the unknown parameters (M-step) and the unobserved labels (E-step) in the proposed segmentation problem. The problem can now be treated as a CNN model training task using partial labels.

As shown in Figure 1, the proposed method consists of two main components: physiological prior prediction model (left panel) and EM-regularized segmentation model (right panel). The left panel takes in physiological MRI information to train a classifier and generate voxelwise estimate of the unobserved labels in ROI2. The estimated label information is then passed into the right panel to improve the prediction performance of the segmentation model. Specifically, the label information is used to initialize ROI2 labels in the CNN model training in M-step, and is also integrated into E-step to recursively update the estimation of the unobserved label $\mathbf{y}_u$. The expected outcome of the right panel is a trained CNN segmentation model that can effectively distinguish the infiltrated tumor from the non-cancerous abnormalities, e.g., edema.

The pipeline introduced in Figure 1 can be further generalized to other similar segmentation problems with partially unobserved labels. Both the classifier in the left panel and the CNN segmentation model
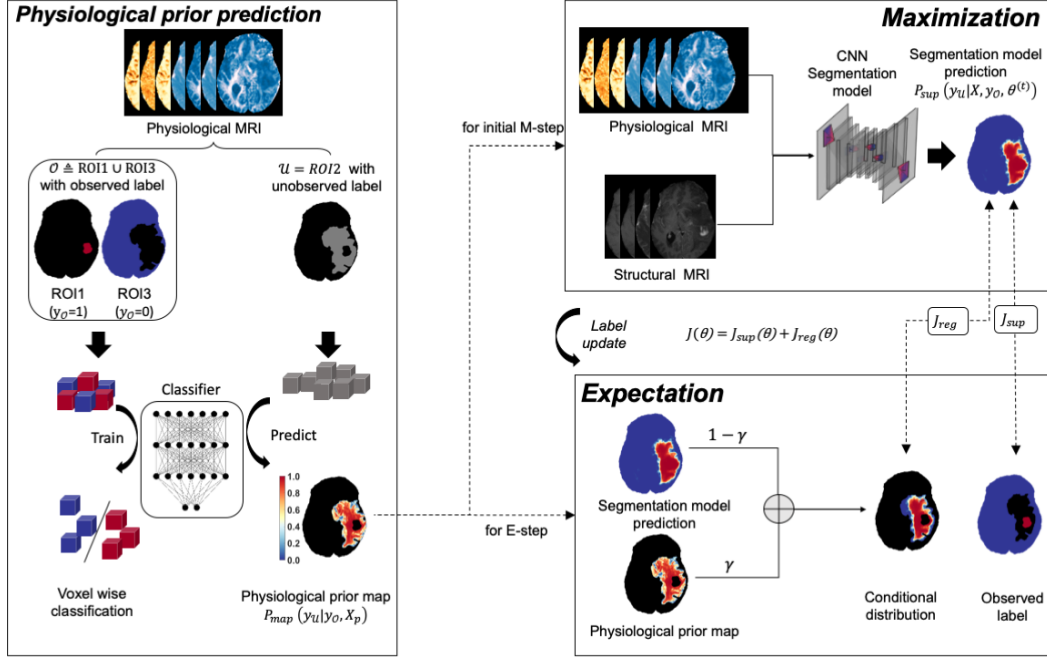
Figure 1: Diagram of the proposed method. The left panel describes the physiological prior prediction process. A classifier is trained to generate physiological prior map. The right panel depicts the EM regularized CNN model training process. The Expectation-Maximization (EM) framework is used to fulfill and optimize the weakly supervised model, where a CNN model is trained in the M-step and the distribution of unobserved ROI2 are estimated in the E-step. $J(\boldsymbol{\theta})$ is the loss term of CNN model, and is calculated using scaled summation of both $J_{\mathrm{reg}}(\boldsymbol{\theta})$ and $J_{\mathrm{sup}}(\boldsymbol{\theta})$. The term $J_{\mathrm{reg}}(\boldsymbol{\theta})$ denotes the regularized loss generated by the conditional distribution computed by Equation (5), and $J_{\mathrm{sup}}(\boldsymbol{\theta})$ denotes the supervised loss from the observed labels $\mathbf{y}_o$.

in the right panel are flexible to be replaced by other feed-forward deep learning models or CNN models with architectures other than the ones used in this paper. Given this, we will not explicitly describe detailed architecture of the CNN models used in the proposed method.

3.3 Physiological prior prediction

As discussed above, physiological MRI is more specific for tumor infiltration but in lower resolution than structural MRI. Treating physiological MRI and structural MRI equally may not able to effectively leverage the specific information from physiological MRI. Therefore, a *physiological prior map* which incorporates only the information of physiological MRI is generated to describe the extracted knowledge of ROI2. In particular, we constructed the underpin component to approximate the unobserved labels $\mathbf{y}_u$ of ROI2, using a classifier trained by both the physiological MRI $\mathbf{X}_p$ and the observed labels $\mathbf{y}_o$.

Since the labels $\mathbf{y}_o$ in ROI1 and ROI3 only contain binary values 1 and 0, we used a binary classifier constructed by a fully connected neural network with two hidden layers. The number of hidden neurons is set equal to the number of input features from $\mathbf{X}_p$. The model produces probabilisitic prodicton for the distribuion of unoberserved labels $\mathbf{y}_u$ in ROI2 with predicted value between $[0, 1]$. The predicted physiological prior map can then be formulated as $p_{map}(\mathbf{y}_u|\mathbf{X}_p, \mathbf{y}_o)$, which was used in the EM-regularized weakly supervised learning segmentation component.

5

3.4 Segmentation with EM-regularized weakly supervised learning

In this component, a segmentation model constructed by a typical U-Net CNN architecture is trained for tumor segmentation. Different from the physiological prior prediction model, the segmentation model is trained using both physiological MRI $\mathbf{X}_p$ and structural MRI $\mathbf{X}_s$. The EM algorithm is leveraged in this component to estimate the unobserved label $\mathbf{y}_u$ and recursively optimize both the model accuracy and label accuracy of the partial labels potential infiltrated region. To perform this weakly supervised learning segmentation task, we firstly define the likelihood function as:

$$L(\boldsymbol{\theta}) = p(\mathbf{y}_u, \mathbf{y}_o | \mathbf{X}, \boldsymbol{\theta}), \tag{1}$$

for which the maximum likelihood estimate with respect to the weights θ (of CNN) can be computed by integrating out the unknown term $\mathbf{y}_u$ and maximizing the marginal distribution:

$$p(\mathbf{y}_o | \mathbf{X}, \theta) = \int p(\mathbf{y}_u, \mathbf{y}_o | \mathbf{X}, \boldsymbol{\theta}) d\mathbf{y}_u. \tag{2}$$

Nevertheless, the integral is often intractable and exact integration over all possible $\mathbf{y}_u$ values is challenging.

EM algorithm solves the problem by iteratively estimating the unknown term $\mathbf{y}_u$ in the expectation step (*E-step*) and $\boldsymbol{\theta}$ in the maximization step (*M-step*). See (McLachlan and Krishnan, 2007) for details of the standard EM algorithm.

In this work, EM performs E-step by defining

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) = \mathbb{E}_{p(\mathbf{y}_u | \mathbf{y}_o, \mathbf{X}, \boldsymbol{\theta}^{(t)})}[\log L(\boldsymbol{\theta})], \tag{3}$$

$$= \int p(\mathbf{y}_u | \mathbf{y}_o, \mathbf{X}, \boldsymbol{\theta}^{(t)}) \log p(\mathbf{y}_u, \mathbf{y}_o | \mathbf{X}, \boldsymbol{\theta}), \tag{4}$$

where $\boldsymbol{\theta}^{(t)}$ denotes the estimated CNN weights in iteration t. $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$ computes the expectation of the log-likelihood of function $L(\boldsymbol{\theta})$ with respect to the conditional distribution $p(\mathbf{y}_u | \mathbf{y}_o, \mathbf{X}, \boldsymbol{\theta}^{(t)})$, which can be defined as:

$$p(\mathbf{y}_u | \mathbf{y}_o, \mathbf{X}, \boldsymbol{\theta}^{(t)}) = \gamma p_{map}(\mathbf{y}_u | \mathbf{X}_p, \mathbf{y}_o) + (1 - \gamma) p_{seg}(\mathbf{y}_u | \mathbf{X}, \mathbf{y}_o, \boldsymbol{\theta}^{(t)}). \tag{5}$$

The former term on the RHS is the physiological prior map generated by the binary classifier and the latter term is the predicted labels in the current tth iteration of EM. γ denotes a voxelwise coefficient, which will be used to integrate the physiological prior map and the prediction of segmentation model.

M-step is to maximize the above quantity to derive new estimate $\boldsymbol{\theta}^{(t+1)}$:

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) \tag{6}$$

The conditional distribution $p(\mathbf{y}_u | \mathbf{y}_o, \mathbf{X}, \boldsymbol{\theta}^{(t)})$ can be obtained by the designed CNN model, where its weight is given by $\theta^{(t)}$.

From the perspective of loss function in CNN model training, Equation (6) can also be treated as the regularization terms to minimize the training loss of the segmentation model in M-step. In practice, the training loss $J(\boldsymbol{\theta})$ is defined as:

$$J(\boldsymbol{\theta}) = J_{\text{sup}}(\boldsymbol{\theta}) + J_{\text{reg}}(\boldsymbol{\theta}), \tag{7}$$

which is a summation of both the supervised loss $J_{\text{reg}}(\boldsymbol{\theta})$ from the fixed observed labels $\mathbf{y}_o$ and the regularised loss $J_{\text{sup}}(\boldsymbol{\theta})$ from pseudo labels $\hat{\mathbf{y}}_u$ calculated using the conditional distribution in Equation (5).

3.5 Model evaluation

We validated the proposed model using tumor burden, tumor recurrence and MRS. To examine the usefulness of the regularizer, we compared our model performance with the baseline model which employed the U-net with a partial cross-entropy loss without the additional regularizer from the physiological prior. We also compared our model with other methodsAkbari et al. (2016); Tang et al. (2018); Kervadec et al. (2019); Roth et al. (2019).

1) Tumor burden estimation

The finally segmented tumor volume was calculated as the core tumor burden (the delineated tumor in ROI1) and infiltrated tumor burden (the delineated tumor in ROI2). A linear regression was used to test the consistency of the segmented volumes from different models with the ground truth. Forthe core tumor (ROI1), the ground truth was used as the volume of the manual label. For the infiltrated tumor, the ground truth was used as the volumme of the recurrence within the potential infiltrated region (ROI2).

2) Tumor burden and recurrence prediction

The finally segmented tumor region was examined in the prediction of complete tumor burden and tumor recurrence region in the follow-up MRI of 68 patients who received the complete resection, which is defined clinically as a complete resection of contrast-enhancing tumor (ROI1). The potential infiltrated region (ROI2) on the pre-operative images was divided into recurrence region R_{recur} and non-recurrence region R_{recur}^C, according to the manual label, where C represents the complementary operation.

For each patient, the pre-operative contrast-enhancing core tumor (ROI1) on T1C image was denoted as R_{CE}, therefore the total tumor burden was defined as $R_{total} = R_{recur} \cup R_{CE}$, whereas the normal-appearing area R_{normal} was defined as R_{total}^C. The segmented tumor area $\hat{R}_{seg}$ and normal-appearing area $\hat{R}_{seg}^C$ can be derived automatically by thresholding the tumor infiltration probability that was finally produced by EMReDL. Finally, The sensitivity and specificity of predicting tumor burden were defined as:

$$Sensitivity = \frac{Volume\left(R_{total} \cap \hat{R}_{seg}\right)}{Volume\left(R_{total}\right)} \tag{8}$$

$$Specificity = \frac{Volume\left(R_{recur}^C \cap \hat{R}_{seg}^C\right)}{Volume\left(R_{recur}^C\right)} \tag{9}$$

After calculating the sensitive and specificity, the optimum threshold T for discriminating predicted infiltration mask was chosen by maximizing the Youden Index of the ROC curves.

3) Magnetic resonance spectroscopy validation

The metabolic signature was compared for the infiltrated region and non-infiltrated region segmented by our model in the potential infiltrated region (ROI2). The metabolic measures, including Choline, N-acetylaspartate (NAA) and Cho/NAA were calculated for the infiltrated region and non-infiltrated region, respectively. To account for the resolution difference between T2 and MRS space, all co-registered data were projected to MRS space, according to their coordinates using MATLAB. The proportion of T2-space tumor pixels occupying each MRS voxel was calculated. Paired t-test was used to compare the metabolic measures of the infiltration and non-infiltration regions.

4 Experiments

4.1 Data description

This study was approved by the local institutional review board and informed consent was obtained from all patients. A total of 115 glioblastoma patients was prospectively recruited for maximal safe resection. Each patient underwent pre-operative multiparametric MRI, using a 3-Tesla MRI system (Magnetron Trio; Siemens Healthcare, Erlangen, Germany) with a standard 12-channel receive-head coil. The sequences included T1, T1C, T2, T2-FLAIR, diffusion imaging, DSC and multivoxel 2D 1H-MRS chemical shift imaging.

4.2 Image pre-processing

1) Multiparametric MRI processing

Diffusion MRI was processed using the diffusion toolbox (FDT) in FSL v5.0.8 (FMRIB Software Library, Centre for Functional MRI of the Brain, Oxford, UK). After normalization and eddy current

correction, parametric maps of fractional anisotropy (FA), mean diffusivity (MD), p (isotropy) and q (anisotropy) were calculated as previously described (Li et al., 2019e,c). DSC was processed using the NordicICE (NordicNeuroLab, Bergen, Norway), with arterial input function automatically defined and leakage corrected. The parametric maps of rCBV, MTT and rCBF maps were calculated. The MRS data were processed using LCModel (Provencher, Oakville, Ontario) as previously described. All metabolites were calculated as a ratio to creatine (Cr).

2) Image co-registration

All pre-operative parametric maps were co-registered to the T2 space using FSL linear image registration tool (FLIRT) with an affine transformation. For the co-registration of the recurrence image to the pre-operative images, the recurrence T1C images were non-linearly co-registered to the pre-operative T2 images using the Advanced Normalization Tools (ANTs), with the pre-operative lesion masked out.

3) Image normalization

All MRI from different patients were normalized using the histogram matching method. Specifically, for each sequence, the image histograms for all patients were calculated, where the histogram closest to the averaged histogram was determined as the reference and normalized to [0, 1]. Finally, other image were matched to the reference histogram.

4.3   Labelling of pre-operative and recurrence tumor

Preoperative tumor and recurrence regions were manually delineated on the T1C and FLAIR images using the 3D slicer v4.6.2 (https://www.slicer.org/). The delineation was independently performed by a neurosurgeon (XX) and reviewed by a neuroradiologist (XX). Each rater used consistent criteria in each patient and was blinded to patient outcomes. The contrast-enhancing (CE) core tumor was defined as the regions within the contrast-enhancing margin on T1C images. The FLAIR ROI was defined as the hyperintensities on FLAIR images. Finally, the peritumoral ROIs were defined as the non-enhancing regions outside of contrast-enhancing regions, obtained by a Boolean subtraction of CE and FLAIR ROIs in MATLAB.

4.4   Treatments

Patient was treated and followed up by the multidisciplinary team (MDT) according to the clinical guidelines. The extent of resection was assessed according to post-operative MRI within 72 hours. During the follow up of patients, clinical and radiological data were incorporated according to the Response Assessment in Neuro-oncology criteria.

4.5   Implementation details

We divided the complete dataset into two sets randomly: 50% as the training set (images of 57 patients) and 50% as the testing set (images of 58 patients). For the training set, 75% of the data was used for model training and the remaining 25% was used for model validation.

For the training of physiological prior prediction model, the multiparametric MRI feature vector for of the voxels in the ROI1 and ROI3 were used as the input of the empirical fully connected network. The model was trained to minimize the losss function. Adam optimizer was applied to train the model with initial learning rate set to 10^{-4}, and the model was trained for 1000 epochs using mini-batches of size $5x10^4$. To tackle the class imbalance problem, equal numbers of majority- and minority-class samples were randomly selected for each mini-batch. Finally, the model with smallest validation error was adopted.

After the training of the physiological prior prediction model, a physiological prior map with the tumor infiltration probability was obtained. The EM-regularized weakly supervised segmentation model was trained for 200 epochs using Adam optimizer with initial learning rate of 10^{-4}, and mini-batch size of 8. For the training of the first epoch, the prior infiltration probability was used as the probabilistic training labels in ROI2, the potential infiltration regions. Afterwards, the probabilistic training labels were updated for each epoch. The model with lowest validation error was finally chosen.

5   Results and Discussion

The experiment results showed that the proposed weakly supervised model achieved high accuracy in segmenting the core and infiltrated tumor area, which could be validated by the tumor burden estimation, tumor recurrence prediction and identification of invasive areas in MRS. The results are presented in below.

5.1   Tumor burden estimation

Tumor burden is crucial for patient risk stratification and treatment planning. We calculated the tumor burden estimated from the different models as the volume of the segmented regions (Table 1). For the core tumor, the results showed that all CNN models achieved comparable volumes with the grund truth, highlighting the capacity of CNN in core tumor segmentation. For the infiltrated tumor, our results showed EMReDL achieved most similar results with the recurrence volume.

Table 1: Tumor burden estimation of different models

		Ground truth	Baseline	Comparison model 1	Comparison model 2	Comparison model 3	Comparison model 4	EMReDL
Core tumor	Training	45.4±29.4	44.8±29.2	33.7±18.9	45.0±29.2	43.6±27.7	43.4±27.7	44.3±28.7
	Testing	48.8±29.7	46.8±29.0	36.4±20.5	46.7±28.8	45.7±27.7	43.8±26	45.0±27.1
Infiltrated tumor	Training	17.9±16.2	9.4±6.2	31.4±22.1	9.1±5.2	20.9±10.5	16.0±10.8	17.5±17.5
	Testing	24.0±19.3	13.2±18	34.8±26.5	12.5±18.4	22.4±15.1	20.2±18.5	24.2±22.4

Unit: cm^3; Comparison model 1: SVM. Comparison model 2: Normalized cut loss. Comparison model 3: Size-constrained loss; Comparison model 4: Random walker regularized loss

We also performed the regression analysis between the tumor burden estimated from the models with the ground truth (Table 2). The results showed that for the core tumor, all tested models showed consistency in core tumor burden estimation. However, for the infiltrated tumor, EMReDL achived better consistency over other tested models.

Table 2: Correlations of tumor burden from ground truth and segmentation

		Baseline	Comparison model 1	Comparison model 2	Comparison model 3	Comparison model 4	EMReDL
Core tumor	R	0.998	0.939	0.998	0.989	0.990	0.995
	P	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Infiltrated tumor	R	0.641	0.839	0.549	0.842	0.859	0.978
	P	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Comparison model 1: SVM. Comparison model 2: Normalized cut loss. Comparison model 3: Size-constrained loss; Comparison model 4: Random walker regularized loss

5.2   Recurrence prediction

Firstly, we compared the performance of the baseline model and EMReDL. The ablation experiment showed that EMReDL achieved superior accuracy in predicting tumor recurrence compared to the baseline model which employed the U-net with a partial cross-entropy loss. The results suggest the usefulness of incorporating the additional regularizer constructed from the physiological MRI. Of note, the baseline model achieved higher higher sensitivity, but lower specificity than EMReDL, which is mainly due to the much smaller segmentation regions. The quantitative comparison results of the EMReDL and baseline model are in Table 3.

Figure 2 presents two examples of infiltration area predicted by the EMReDL and baseline model. The pre-operative structural MRIs, including FLAIR, T1C (Figure 2A,B), recurrence T1C (Figure 2C), and physiological MRI including DTI-q, DTI-p, FA, MD, MTT, rCBV and rCBF (Figure 2H-N), as well as the overlaid labels (red: contrast-enhancing core tumor, ROI1; blue: non-enhancing peritumoral region, ROI2). The prediction of two models is overlaid on pre-operative (Figure 2D: bassline, Figure 2E: EMReDL) and recurrence (Figure 2F: bassline, Figure 2G: EMReDL) T1C images. Note the recurrence area is well beyond the contrast-enhancing tumor core on the pre-operative MRI, which showed high correspondence with the infiltrated area identified by EMReDL. This improvement could possibly be explained by the tumor invasion area revealed by the physiological MRI shown

Table 3: Comparisons of baseline and EMReDL

		EMReDL	Baseline
AUC	Train	0.971	0.897
	Test	0.965	0.890
Sensitivity	Train	0.906	0.789
	Test	0.898	0.772
Specificity	Train	0.918	0.929
	Test	0.916	0.926
Youden index	Train	0.825	0.697
	Test	0.813	0.718
Dice	Train	0.849	0.745
	Test	0.846	0.733
MCC	Train	0.823	0.716
	Test	0.808	0.689

AUC: area under the curve. MCC: Matthews correlation coefficient

underneath. Note the ground truth (the red region) of the complete tumor burden was taken as the combination of the core tumor and the recurrence tumor, with the assumption that the infiltrated tumor in the FLAIR is more responsible for the recurrence outside of the core tumor than other regions.
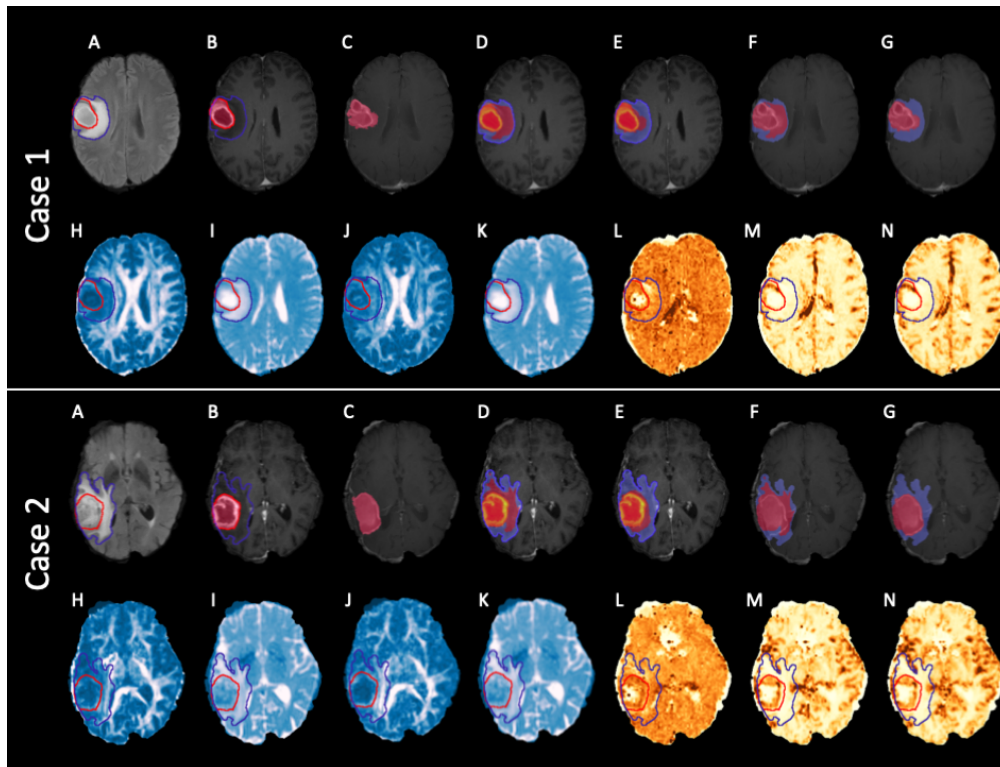

Figure 2: Two case examples with the segmentation results of baseline model and EMReDL. For both cases, A: FLAIR, B: T1C, C: recurrence T1C (red: ROI1, contrast-enhancing core tumor; blue: ROI2, peritumoral non-enhancing region); D-G: model results (red) with the ROI2 (blue) overlaid. D: baseline result on pre-operative T1C image; E: EMReDL result on pre-operative T1C image; F: baseline result on recurrence T1C images; G: EMReDL result on recurrence T1C images; H-N: pre-opearive DTI-q, DTI-p, FA, MD, MTT, rCBV and rCBF images in sequence.

Next, we compared our results of the segmented infiltration area with other weakly-supervised models proposed in (Akbari et al., 2016; Kervadec et al., 2019; Roth et al., 2019; Tang et al., 2018). The results (Table 4) showed that all the models with additional loss achieved better accuracy than the SVM model, suggesting the usefulness of considering the spatial information through CNN in the

prediction. Further, the EMReDL obtained higher accuracy than other weakly supervised models, which again supports the value of incorporating the physiological information through the separate physiological prior prediction model from the main segmentation model. As mentioned, physiological MRI has higher specificity in reflecting tumor biology but lower resolution than structural MRI. Benefiting from the separately designed model, the physiological information could be effectively employed and less affected by the structural MRI, which hence could improve the model performance. In comparison, the pseudo labels generated through the normalized cut loss in (Tang et al., 2018) and the random walker loss in (Roth et al., 2019) were obtained by treating the structural and physiological MRI equally, therefore may not effectively leverage the information from physiological MRI.

Table 4: Comparison of weakly supervised models

		Comparison model 1	Comparison model 2	Comparison model 3	Comparison model 4	EMReDL
AUC	Train	0.764	0.901	0.855	0.923	0.971
	Test	0.788	0.888	0.866	0.919	0.965
Sensitivity	Train	0.757	0.790	0.845	0.838	0.906
	Test	0.765	0.764	0.824	0.815	0.898
Specificity	Train	0.664	0.934	0.799	0.882	0.918
	Test	0.679	0.930	0.841	0.891	0.916
Youden index	Train	0.422	0.724	0.644	0.720	0.825
	Test	0.444	0.693	0.664	0.706	0.813
Dice	Train	0.593	0.749	0.725	0.764	0.849
	Test	0.621	0.727	0.739	0.755	0.846
MCC	Train	0.423	0.722	0.645	0.717	0.823
	Test	0.444	0.685	0.658	0.697	0.808

AUC: area under the curve. MCC: Matthews correlation coefficient. Comparison model 1: SVM. Comparison model 2: Normalized cut loss. Comparison model 3: Size-constrained loss; Comparison model 4: Random walker regularized loss

Figure 3 presents an example with the comparison of different models. Figure 3a-d show the structural images including T1C, FALIR, T1 and T2. Figure 3e and 3f show the FLAIR abnormality and contrast-enhancing tumor respectively, while Figure 3g indicates the recurrence regions on the follow up scans. The physiological MRI, including DTI-q, DTI-p, FA, MD, MTT, rCBV and rCBF, are shown in Figure 2H-N. Indeed, the EMReDL shows the highest performance, whereas the SVM model shows lower accuracy than all other models.

Lastly, we compared the performance of the different models in segmenting the infiltrated area in Table 5. As expected, all models obtained lower performance than segmenting the complete tumor burden including the core tumor, as we only take the recurrence region as the ground truth, while some non-recurrence area may also display invasive imaging features in the pretreatment MRI. For the model comparison, however, EMReDL achieved higher performance than other models, which may imply the value of the additionally constructed regularizer.

To summarize, the model comparisons may validate the performance of the proposed weakly supervised model. Also, our model showed comparable performance in both training and testing sets, which could suggest the robustness of the model.

Table 5: Comparison of infiltrated tumor segmentation

		Baseline	Comparison model 1	Comparison model 2	Comparison model 3	Comparison model 4	EMReDL
AUC	Train	0.674	0.781	0.680	0.778	0.804	0.915
	Test	0.707	0.807	0.701	0.807	0.837	0.938
Sensitivity	Train	0.463	0.787	0.480	0.771	0.736	0.809
	Test	0.523	0.801	0.517	0.790	0.779	0.876
Specificity	Train	0.868	0.664	0.860	0.676	0.757	0.890
	Test	0.866	0.679	0.858	0.711	0.774	0.889
Dice	Train	0.339	0.408	0.346	0.407	0.441	0.621
	Test	0.408	0.478	0.398	0.492	0.528	0.711
Youden index	Train	0.331	0.451	0.340	0.448	0.493	0.699
	Test	0.389	0.480	0.375	0.501	0.553	0.765
MCC	Train	0.353	0.400	0.356	0.398	0.450	0.677
	Test	0.414	0.449	0.398	0.471	0.527	0.746

AUC: area under the curve. MCC: Matthews correlation coefficient. Comparison model 1: SVM. Comparison model 2: Normalized cut loss. Comparison model 3: Size-constrained loss; Comparison model 4: Random walker regularized loss
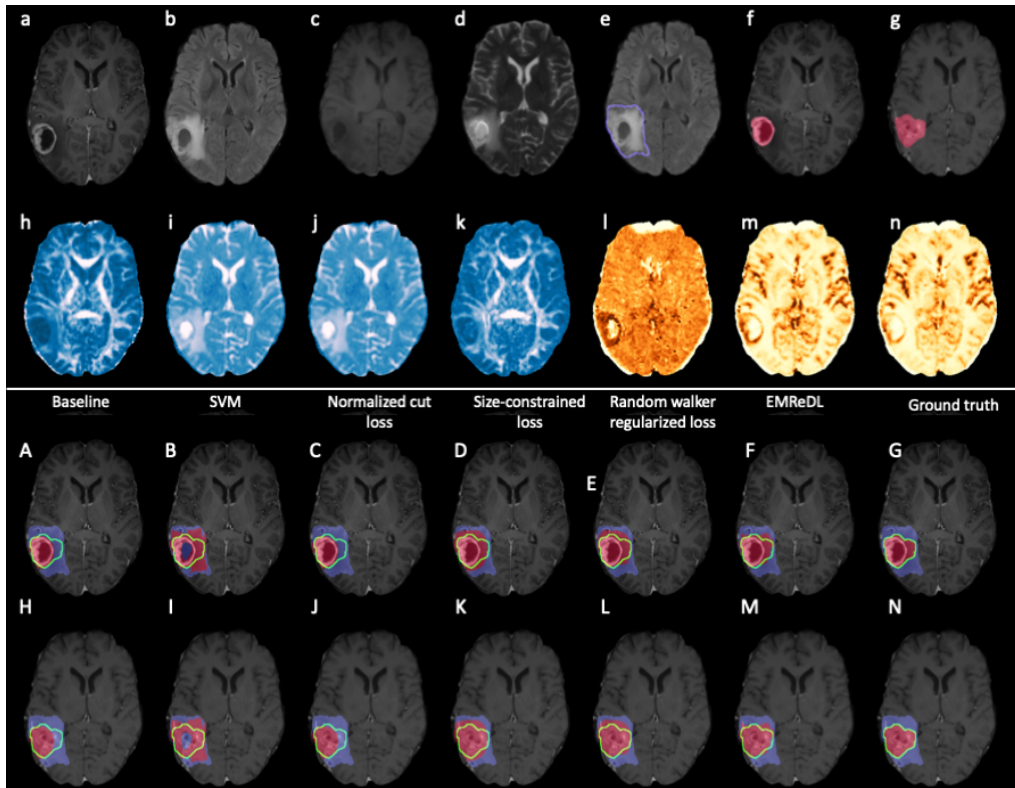

Figure 3: Case examples of model comparison. Top panel: pre-operative and recurrence images. a-d: pre-operative T1C, FLAIR, T1Pre, T2 in sequence. e-g: labelled pre-operative FLAIR, T1C, and recurrence T1C; h-n: pre-operative DTI-q, DTI-p, FA, MD, MTT, rCBV and rCBF images in sequence. Bottom panel. Model results (red) with the ROI2 (blue) overlaid. The green lines indicate the ground truth. A-G: segmentation of different models overlaid on pre-operative T1C images. H-N: segmentation of different models overlaid on recurrence T1C images.

5.3 MRS results

The MRS results showed that the predicted infiltrated region showed significantly more aggressive signature than the non-infiltrated region, which suggests the infiltration prediction could have signifi-

cance regarding the tumor-induced metabolic change. Specifically, choline is a marker of cellular turnover and membrane integrity, which is correlated with tumor proliferation. NAA is a maker of neuron structure, which may be destructed by the tumor infiltration. In previous studies, the choline/NAA ratio was frequently used an imaging marker to indicate tumor invasiveness, which was shown to correlate with patient outcomes. The detailed comparison of MRS data from the predicted infiltrated ad non-inlfiltrated regions are detailed in Table 6.
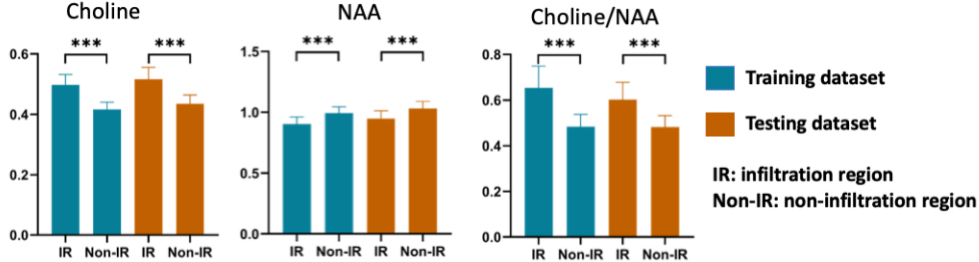

Figure 4: MRS comparison of the infiltrated and non-infiltrated regions

Table 6: MRS comparison of the segmented infiltration and non-infiltration

		IR	Non-IR	p-value
Choline	Training	0.50±0.13	0.42±0.09	$3.1{\times}10^{-12}$
	Testing	0.52±0.14	0.44±0.11	$4.0{\times}10^{-10}$
Cho/NAA	Training	0.65±0.35	0.48±0.20	$1.4{\times}10^{-8}$
	Testing	0.60±0.27	0.48±0.18	$4.1{\times}10^{-7}$
NAA	Training	0.90±0.22	0.99±0.20	$9.3{\times}10^{-8}$
	Testing	0.95±0.24	1.03±0.21	$5.9{\times}10^{-6}$

IR: infiltration region; NAA: N-acetylaspartate

Our study has limitations. Firstly, our manual labels were delineated by human experts. Therefore, different from the synthetic images, any analysis performed on this dataset may be biased and subjective compared to the synthetic images. Secondly, the other weakly supervised models that we compared with our models are not developed based on MRI. Therefore the performance may be affected when applied to our images. Lastly, due to the nature of tumor infiltration and ethics issue, some infiltrated tumor may not be directed observed and measured, as some tumor regions are more sensitive to treatment, Therefore, incorporating longitudinal MRI into the model could yield a more accurate infiltrated tumor estimation, which we are improving in our current study.

6 Conclusions

In this paper, we presented an expectation-maximization regularized weakly supervised tumor segmentation model based on the deep convolutional neural networks. The proposed method was developed to segment both the core and peritumoral infiltrated tumor based on the multiparametric MRI. This weakly supervised model was developed to tackle the challenge of obtaining the full accurate labels for the infiltrated tumour. To effectively leverage the physiological MRI that has higher specificity but lower resolution than structural MRI, we constructed a physiological prior map generated from a fully connected neural network, for the iterative optimization of the CNN segmentation model. Using the tumor burden, tumor recurrence and MRS, the model evaluation confirms that our proposed model achieved higher accuracy than the published state-of-the-art weakly supervised methods, using the regularizer constructed from physiological MRI.

References

Akbari, H., Macyszyn, L., Da, X., Bilello, M., Wolf, R.L., Martinez-Lage, M., Biros, G., Alonso-Basanta, M., O'Rourke, D.M., Davatzikos, C., 2016. Imaging surrogates of infiltration obtained via

multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. Neurosurgery 78, 572–580.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 .

Chan, L., Hosseini, M.S., Plataniotis, K.N., 2020. A comprehensive analysis of weakly-supervised semantic segmentation in different image domains. International Journal of Computer Vision URL: `https://doi.org/10.1007/s11263-020-01373-4`, doi:10.1007/s11263-020-01373-4.

Dutil, F., Havaei, M., Pal, C., Larochelle, H., Jodoin, P., 2015. A convolutional neural network approach to brain lesion segmentation. Ischemic Stroke Lesion Segment , 51–6.

Ghaffari, M., Sowmya, A., Oliver, R., 2020. Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the brats 2012-2018 challenges. IEEE Rev Biomed Eng 13, 156–168. URL: `https://www.ncbi.nlm.nih.gov/pubmed/31613783`, doi:10.1109/RBME.2019.2946868.

Jellison, B.J., Field, A.S., Medow, J., Lazar, M., Salamat, M.S., Alexander, A.L., 2004. Diffusion tensor imaging of cerebral white matter: a pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. American Journal of Neuroradiology 25, 356–369.

Jiang, Z., Ding, C., Liu, M., Tao, D., 2019. Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task, in: International MICCAI Brainlesion Workshop, Springer. pp. 231–241.

Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al., 2017. Ensembles of multiple models and architectures for robust brain tumour segmentation, in: International MICCAI Brainlesion Workshop, Springer. pp. 450–462.

Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B., 2016. Deepmedic for brain tumor segmentation, in: International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries, Springer. pp. 138–149.

Kervadec, H., Dolz, J., Tang, M., Granger, E., Boykov, Y., Ben Ayed, I., 2019. Constrained-cnn losses for weakly supervised segmentation. Med Image Anal 54, 88–99. URL: `https://www.ncbi.nlm.nih.gov/pubmed/30851541`, doi:10.1016/j.media.2019.02.009.

Li, C., Wang, S., Serra, A., Torheim, T., Yan, J.L., Boonzaier, N.R., Huang, Y., Matys, T., McLean, M.A., Markowetz, F., Price, S.J., 2019a. Multi-parametric and multi-regional histogram analysis of mri: modality integration reveals imaging phenotypes of glioblastoma. Eur Radiol 29, 4718–4729. URL: `https://www.ncbi.nlm.nih.gov/pubmed/30707277`, doi:10.1007/s00330-018-5984-z.

Li, C., Wang, S., Yan, J.L., Piper, R.J., Liu, H., Torheim, T., Kim, H., Zou, J., Boonzaier, N.R., Sinha, R., Matys, T., Markowetz, F., Price, S.J., 2019b. Intratumoral heterogeneity of glioblastoma infiltration revealed by joint histogram analysis of diffusion tensor imaging. Neurosurgery 85, 524–534. URL: `https://www.ncbi.nlm.nih.gov/pubmed/30239840`, doi:10.1093/neuros/nyy388.

Li, C., Wang, S., Yan, J.L., Piper, R.J., Liu, H., Torheim, T., Kim, H., Zou, J., Boonzaier, N.R., Sinha, R., et al., 2019c. Intratumoral heterogeneity of glioblastoma infiltration revealed by joint histogram analysis of diffusion tensor imaging. Neurosurgery 85, 524–534.

Li, C., Wang, S., Yan, J.L., Torheim, T., Boonzaier, N.R., Sinha, R., Matys, T., Markowetz, F., Price, S.J., 2019d. Characterizing tumor invasiveness of glioblastoma using multiparametric magnetic resonance imaging. J Neurosurg , 1–8URL: `https://www.ncbi.nlm.nih.gov/pubmed/31026822`, doi:10.3171/2018.12.JNS182926.

Li, C., Wang, S., Yan, J.L., Torheim, T., Boonzaier, N.R., Sinha, R., Matys, T., Markowetz, F., Price, S.J., 2019e. Characterizing tumor invasiveness of glioblastoma using multiparametric magnetic resonance imaging. Journal of Neurosurgery 1, 1–8.

Lupo, J.M., Cha, S., Chang, S.M., Nelson, S.J., 2005. Dynamic susceptibility-weighted perfusion imaging of high-grade gliomas: characterization of spatial heterogeneity. American Journal of Neuroradiology 26, 1446–1454.

Mazzara, G.P., Velthuizen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H., 2004. Brain tumor target volume determination for radiation treatment planning through automated mri segmentation. Int J Radiat Oncol Biol Phys 59, 300–12. URL: https://www.ncbi.nlm.nih.gov/pubmed/15093927, doi:10.1016/j.ijrobp.2004.01.026.

McLachlan, G.J., Krishnan, T., 2007. The EM algorithm and extensions. volume 382. John Wiley & Sons.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K., 2015. The multimodal brain tumor image segmentation benchmark (brats). IEEE Trans Med Imaging 34, 1993–2024. URL: https://www.ncbi.nlm.nih.gov/pubmed/25494501, doi:10.1109/TMI.2014.2377694.

Myronenko, A., 2018. 3d mri brain tumor segmentation using autoencoder regularization, in: International MICCAI Brainlesion Workshop, Springer. pp. 311–320.

Prastawa, M., Bullitt, E., Ho, S., Gerig, G., 2004. A brain tumor segmentation framework based on outlier detection. Med Image Anal 8, 275–83. URL: https://www.ncbi.nlm.nih.gov/pubmed/15450222, doi:10.1016/j.media.2004.06.007.

Roth, H., Zhang, L., Yang, D., Milletari, F., Xu, Z., Wang, X., Xu, D., 2019. Weakly supervised segmentation from extreme points, in: Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention. Springer, pp. 42–50.

Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., Curschmann, J., Janzer, R.C., Ludwin, S.K., Gorlia, T., Allgeier, A., Lacombe, D., Cairncross, J.G., Eisenhauer, E., Mirimanoff, R.O., European Organisation for, R., Treatment of Cancer Brain, T., Radiotherapy, G., National Cancer Institute of Canada Clinical Trials, G., 2005. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352, 987–96. URL: https://www.ncbi.nlm.nih.gov/pubmed/15758009, doi:10.1056/NEJMoa043330.

Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., Schroers, C., 2018. Normalized cut loss for weakly-supervised cnn segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1818–1827.

Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J., 2014. Multi-modal brain tumor segmentation using deep convolutional neural networks. MICCAI BraTS (brain tumor segmentation) challenge. Proceedings, winning contribution , 31–35.

Verburg, N., Koopman, T., Yaqub, M.M., Hoekstra, O.S., Lammertsma, A.A., Barkhof, F., Pouwels, P.J.W., Reijneveld, J.C., Heimans, J.J., Rozemuller, A.J.M., Bruynzeel, A.M.E., Lagerwaard, F., Vandertop, W.P., Boellaard, R., Wesseling, P., de Witt Hamer, P.C., 2020. Improved detection of diffuse glioma infiltration with imaging combinations: a diagnostic accuracy study. Neuro Oncol 22, 412–422. URL: https://www.ncbi.nlm.nih.gov/pubmed/31550353, doi:10.1093/neuonc/noz180.

Wadhwa, A., Bhardwaj, A., Singh Verma, V., 2019. A review on brain tumor segmentation of mri images. Magn Reson Imaging 61, 247–259. URL: `https://www.ncbi.nlm.nih.gov/pubmed/31200024`, doi:`10.1016/j.mri.2019.05.043`.

Warrington, S., Bryant, K.L., Khrapitchev, A.A., Sallet, J., Charquero-Ballester, M., Douaud, G., Jbabdi, S., Mars, R.B., Sotiropoulos, S.N., 2020. Xtract-standardised protocols for automated tractography in the human and macaque brain. NeuroImage , 116923.

Weller, M., van den Bent, M., Hopkins, K., Tonn, J.C., Stupp, R., Falini, A., Cohen-Jonathan-Moyal, E., Frappaz, D., Henriksson, R., Balana, C., Chinot, O., Ram, Z., Reifenberger, G., Soffietti, R., Wick, W., European Association for Neuro-Oncology Task Force on Malignant, G., 2014. Eano guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma. Lancet Oncol 15, e395–403. URL: `https://www.ncbi.nlm.nih.gov/pubmed/25079102`, doi:`10.1016/S1470-2045(14)70011-7`.

Weller, M., van den Bent, M., Tonn, J.C., Stupp, R., Preusser, M., Cohen-Jonathan-Moyal, E., Henriksson, R., Le Rhun, E., Balana, C., Chinot, O., Bendszus, M., Reijneveld, J.C., Dhermain, F., French, P., Marosi, C., Watts, C., Oberg, I., Pilkington, G., Baumert, B.G., Taphoorn, M.J.B., Hegi, M., Westphal, M., Reifenberger, G., Soffietti, R., Wick, W., European Association for Neuro-Oncology Task Force on, G., 2017. European association for neuro-oncology (eano) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol 18, e315–e329. URL: `https://www.ncbi.nlm.nih.gov/pubmed/28483413`, doi:`10.1016/S1470-2045(17)30194-8`.

Wen, P.Y., Weller, M., Lee, E.Q., Alexander, B.M., Barnholtz-Sloan, J.S., Barthel, F.P., Batchelor, T.T., Bindra, R.S., Chang, S.M., Chiocca, E.A., Cloughesy, T.F., DeGroot, J.F., Galanis, E., Gilbert, M.R., Hegi, M.E., Horbinski, C., Huang, R.Y., Lassman, A.B., Le Rhun, E., Lim, M., Mehta, M.P., Mellinghoff, I.K., Minniti, G., Nathanson, D., Platten, M., Preusser, M., Roth, P., Sanson, M., Schiff, D., Short, S.C., Taphoorn, M.J.B., Tonn, J.C., Tsang, J., Verhaak, R.G.W., von Deimling, A., Wick, W., Zadeh, G., Reardon, D.A., Aldape, K.D., van den Bent, M.J., 2020. Glioblastoma in adults: a society for neuro-oncology (sno) and european society of neuro-oncology (eano) consensus review on current management and future directions. Neuro Oncol 22, 1073–1113. URL: `https://www.ncbi.nlm.nih.gov/pubmed/32328653`, doi:`10.1093/neuonc/noaa106`.

Yan, J.L., Li, C., Boonzaier, N.R., Fountain, D.M., Larkin, T.J., Matys, T., van der Hoorn, A., Price, S.J., 2019. Multimodal mri characteristics of the glioblastoma infiltration beyond contrast enhancement. Ther Adv Neurol Disord 12, 1756286419844664. URL: `https://www.ncbi.nlm.nih.gov/pubmed/31205490`, doi:`10.1177/1756286419844664`.

Yan, J.L., Li, C., van der Hoorn, A., Boonzaier, N.R., Matys, T., Price, S.J., 2020. A neural network approach to identify the peritumoral invasive areas in glioblastoma patients by using mr radiomics. Sci Rep 10, 9748. URL: `https://www.ncbi.nlm.nih.gov/pubmed/32546790`, doi:`10.1038/s41598-020-66691-6`.

Zhang, J., Yu, X., Li, A., Song, P., Liu, B., Dai, Y., 2020. Weakly-supervised salient object detection via scribble annotations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12546–12555.