Article

A Foreground Prototype-Based One-Shot Segmentation of Brain Tumors

Ananthakrishnan Balasundaram [1], Muthu Subash Kavitha [2,*], Yogarajah Pratheepan [3], Dhamale Akshat [1] and Maddirala Venkata Kaushik [1]

1   School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, Tamil Nadu, India; balasundaram.a@vit.ac.in (A.B.)
2   School of Information and Data Sciences, Nagasaki University, Nagasaki 852-8521, Japan
3   School of Computing, Engineering and Intelligent System, Ulster University, Londonderry BT48 7JL, UK; p.yogarajah@ulster.ac.uk
*   Correspondence: kavitha@nagasaki-u.ac.jp

Abstract: The potential for enhancing brain tumor segmentation with few-shot learning is enormous. While several deep learning networks (DNNs) show promising segmentation results, they all take a substantial amount of training data in order to yield appropriate results. Moreover, a prominent problem for most of these models is to perform well in unseen classes. To overcome these challenges, we propose a one-shot learning model to segment brain tumors on brain magnetic resonance images (MRI) based on a single prototype similarity score. With the use of recently developed few-shot learning techniques, where training and testing are carried out utilizing support and query sets of images, we attempt to acquire a definitive tumor region by focusing on slices containing foreground classes. It is unlike other recent DNNs that employed the entire set of images. The training of this model is carried out in an iterative manner where in each iteration, random slices containing foreground classes of randomly sampled data are selected as the query set, along with a different random slice from the same sample as the support set. In order to differentiate query images from class prototypes, we used a metric learning-based approach based on non-parametric thresholds. We employed the multimodal Brain Tumor Image Segmentation (BraTS) 2021 dataset with 60 training images and 350 testing images. The effectiveness of the model is evaluated using the mean dice score and mean IoU score. The experimental results provided a dice score of 83.42 which was greater than other works in the literature. Additionally, the proposed one-shot segmentation model outperforms the conventional methods in terms of computational time, memory usage, and the number of data.

Keywords: magnetic resonance imaging; few-shot learning; foreground prototypes

1. Introduction

Brain tumors are one of the most dangerous tumors. The most frequent primary brain tumor is gliomas, and it comprises about 2% of all malignancies [1,2]. Over the course of a year, registrations in the neuro-oncology clinic at a tertiary care facility were thoroughly analyzed from hospital-based databases tracking CNS (central nervous system) cancers. For people aged 15 to 39, the 5-year survival rate nears 72%. In India, the prevalence of brain tumors is from 5 to 10 per 100,000 people and is on the rise. Hence, brain tumors are needed to be identified as early as possible for early treatment and recovery. In several applications of computer-aided diagnosis, medical image segmentation is crucial. The region of interest is extracted automatically without involving any manual segmentation procedures and is regarded as the most crucial step in medical imaging procedures.

The conventional structural MRI image segmentation and analysis of brain tumors is an onerous and time-consuming task.

In general, three theoretical categories are used for automated brain segmentation techniques utilizing MR images: intensity-based, machine learning, and hybrid. The highly

heterogeneous appearance of brain tumors, especially sub-regions, makes segmentation difficult [3]. Furthermore, aberrations such as motion and/or field in homogeneity can make the problem even worse. There are several deep learning techniques that have already been implemented in automatic segmentation. However, these processes call for an extensive amount of training data to detect tumors thus make the process computationally expensive. CNN-based models [4,5], being the most common in deep learning architectures, use three-dimensional filters or custom filters to optimize themselves for tumor foreground regions.

State-of-the-art architectures in this field tend to be U-Net and PANet, which show the likelihood of high accuracy scores but on the downside, they use more training data. The use of multi-task networks [6] utilizes the model cascade strategy to develop a solution for class imbalances. Despite its outstanding performance, this method leads to undesired complexity in the model. The use of shared parameters to learn joint features is effective but requires a good size of training data and task-specific parameters which makes it computationally heavy to train. Ref. [6] has shown that they perform much better than other architectures (including a few-shot based) but use ensemble models and post-processing approaches, which adds time required to evaluate each image. The use of transformer-based learning is also one of the most widely used methods to extract and fuse multi-scale features. Ref. [7] has adopted a similar squeeze-and-expansion transformer at the core as a segmentation framework. Having an enormous number of parameters and loading three-dimensional images makes it very computationally inefficient to train and test. Such networks required extensive training in the model. The winner of the BraTS 2018 challenge has performed segmentation using autoencoder regularizations [8]. Although that network proved considerable results, due to its large network width, it required a heavy GPU to assess the large number of parameters.

In order to reduce the complexity of the model, we propose using a few-shot learning technique to segment brain tumors. Generally, unique and distinct classes are chosen for validation and the base data are set in the few-shot model so that the generalization performance is assessed on new classes. We are partially inspired by the state-of-the-art model by Hansen [9] to segment brain tumors. Few-shot learning is a part of meta-learning, which is making the model learn to identify the similarities between the image and support set and label the query image. As it uses only a few labeled samples to segment the query image, the training data required to perform the computation decreases significantly; thus, it reduces the number of network parameters and computational costs.

In this work, we used prototype-based few-shot learning, i.e., the support images are selected and their features are extracted and embedded in a vector space. The features of the query image are then extracted and the similarity score is calculated using the cosine score. In contrast to the conventional frameworks, our segmentation method utilizes a one-shot learning method. As far as we know, this is the first study to implement one-shot learning that uses a single prototype for modeling the foreground class for the segmentation of tumor regions from the MRI images of the brain.

The contribution of this study can be summarized as follows:

1. Propose a one-shot learning segmentation model by considering the foreground prototypes for brain tumor detection on brain MRI that transcends the current deep learning-based tumor segmentation models in regard to the size of the training data.
2. Adopt the VGG-16-based encoder with pre-trained weights from Resnet101 trained on MS COCO datasets as our few-shot learning model.
3. Experiment with different N-shot K-way methods to show the effectiveness of the proposed single prototype for modeling the foreground classes based on its uniformity.
4. Compare the performance of the proposed approach to architectures that are used for brain tumor segmentation.
5. Evaluate the performance of our methods by using the dice similarity coefficient and intersection over union.

2. Related Work

2.1. Deep Learning-Based Brain Tumor Segmentation

Deep learning-based brain tumor segmentation methods utilize various CNN-like neural networks to segment the tumor regions, as they can occur anywhere in any shape or size and so as to automate the process. Ref. [10] has implemented a CNN-based network for glioblastoma. The method in [11] proposed to work only a small part of the image containing the tumor using a C-CNN that utilized two different paths for extracting local and global features. Ref. [12] used nested structures to tackle segmenting different image modalities. Ref. [13] has made use of BCM-CNN and utilized a sine–cosine grey fitness algorithm for tuning. The method in [14] applied U-Net to segment tumor images. It utilized an ensemble of various models with different hyper-parameters to obtain better results and reduce random errors. The method in [15] developed a 3D-Dense-UNet model to simplify complex multi-class segmentation problems. These methods do not work very well on unseen tumor classes and also take a lot of processing time as well. The method in [16] proposed a unique CNN and probabilistic neural network architecture to segment the tumor effectively. Cascaded CNNs were implemented by [17] on three tumor regions, thus implementing multiclass segmentation. However, almost all of these DNN learning models are classification-based models. A comparative analysis of existing tumor segmentation methods [18] shows that they require a large amount of training data to accurately segment the tumor images effectively, and its supervised learning performance on unseen classes is not effective. However, our one-shot-based learning can understand the similarities and differences that work well in unseen classes.

2.2. Self-Supervised Learning

Self-supervised learning learns to identify one part of the input through another by eliminating the need for labels. The few-shot learning (FSL) model is a fully unsupervised model which performs semantic segmentation, and it improved the mIoU score of the previous model. The method in [19] carried out self-supervised depth estimation by utilizing diverse in-depth features and the difficulty in learning the depths. Ref. [20] predicted the position by analyzing the anatomy of cardiac MR images. Context restoration has also been used for medical image segmentation [21,22]. Our proposed model also fits the criteria of the self-supervised learning paradigm. Our model extracts the features of the query prototype and embeds them into the feature vector space where the similarity scores of the feature vector and support vector are compared. Subsequently, the query image is segmented, making the process semi-self-supervised. The support images in our model are introduced in batches and thus the model learns to find the similarity scores between the query and support images in a semi-supervised manner.

2.3. Few-Shot Learning

In medical-based image segmentation, fewer datasets are available. Therefore, we are using few-shot learning, a method in which the model learns to segment the query images using a small set of images which reduces the training data. It performs well on unseen classes as it works on similarity scores between support and query images. The method in [19] has stated that few-shot classification learns classifiers from just a few samples of each class that have been tagged. Multi-feature comparison utilizing two branches is also an effective mechanism for the few-shot learning model [23]. The method in [24] stated that there are two types of few-shot learning methods: inductive, in which only the support images are introduced to the classifier, and transudative, in which both support and query sets are utilized by the classifier. Model agnostic meta-learning is a technique for discovering any standard model's parameters through meta-learning in order to prepare that model for quick adaptation [25]. Matching networks introduced by [26] have an embedded space in which the weighted nearest neighbors classifiers are applied. Prototypical neural networks revolve around developing an embedding space

that points the clusters around a single prototype [27]. The method in [28] has applied few-shot learning on natural language processing very efficiently.

2.4. Few-Shot Segmentation

Semantic segmentation forms clusters of images which belong to the same object class at the pixel level. The method in [4] used a model which comprises two commanding modules. The first module determines if the support and query image visually correspond, while the second dictates the network to focus on the targeted query objects. The method in [29] proposed a novel strategy that automated the process of matching query prototype and query image which are obtained by high-profile query image predictions. The tumor intensity, shape, and location may differ from image to image. However, identifying them requires pixel-level classification. Utilizing both foreground and background features at the pixel level provided optimum results in implementation [30,31]. Ref. [32] proposed a one-shot image segmentation technique similar to ours. The method in [29] utilized a Gaussian mixture model to evaluate prototypical clusters of different classes. The method in [33] has implemented a few-shot-based U-Net architectures for detecting radiographic patterns and obtained an improvement in accuracy. Path aggregation network (PANet) is a unique metric learning-based model that utilizes a prototype alignment network. In the embedding space, PANet learns by extracting specific prototypes of each class from a small number of support images and segments the query images by matching each pixel to the prototypes it has learned. PANet provides high-quality prototypes that are discriminative for many classes and representative of each semantic class via non-parametric metric learning [34]. ADnet main [9] used a model and utilized the homogeneity of the foreground prototype by computing anomaly scores. It is reported to use one foreground prototype and, therefore, outperforms state-of-the-art models for abdominal organ and heart segmentation. Our model utilizes prototype-based learning using one-shot learning that is different from the aforementioned models.

3. Proposed Methodology

In this work, we propose a similarity-measured prototypical one-shot segmentation model to classify tumor regions in MRI images. Using a common feature extractor between query and support sets of images, metric learning-based segmentation is performed in the target domain. We solely focus on slices containing foreground classes, unlike many other recent deep learning algorithms that employ the entire set of images.

3.1. Problem Foundation

We aim to obtain a segmentation model trained using few-shot learning methods which can learn faster than traditional learning methods and require only a few labeled images from the same classes. In this model, images containing distinct training classes C_{train} (e.g.,: C_{train} = NCR/NET—label 1, ED—label 2, ET—label 4) and testing C_{test} are obtained from the training and testing dataset D_{train} and D_{test}, respectively. The segmentation model M is trained on C_{train} and tested on C_{test}.

Taking C_{train} as the training and C_{test} as the testing set, the model is trained in an episodic way. Each episode consists of a group $\{S, Q\}$ (a set of support images and a set of query images). These pairs of $\{S_i, Q_i\}_{i=1}^{N}$, where N denotes the total number of images in the dataset, sum up the whole dataset. Considering the C-shot K-way segmentation learning task for each episode pair $\{S_i, Q_i\}$, we obtain the output as the predicted query mask. The support set $S = \{(x_1, y_1), .., (x_{C \times K}, y_{C \times K})\}$ consists of $C \times K$ number of image mask pairs {x, y}, where $x, y \in R^{X \times Y}$ are slices of each image having $X \times Y$ image dimensions. The query set Q = {$\hat{x}, \hat{y}$} contains an image mask pair with a single slice.

3.2. Few-Shot Learning Model

Given a set of labeled images, our few-shot learning-based model aims to learn adaptation to new classes when trained under very few labeled samples. The training of the

model is carried out in an iterative manner where in each iteration, random slices containing foreground classes of randomly sampled data are selected as the query set along with a different random slice from the same sample as the support set. Unlike previous methods for obtaining prototypes for classes in both the foreground and background [35,36], we are only considering the foreground prototypes. This eliminates the possibility of overtraining the large black background class. The process summarizing the model training is shown in Figure 1.
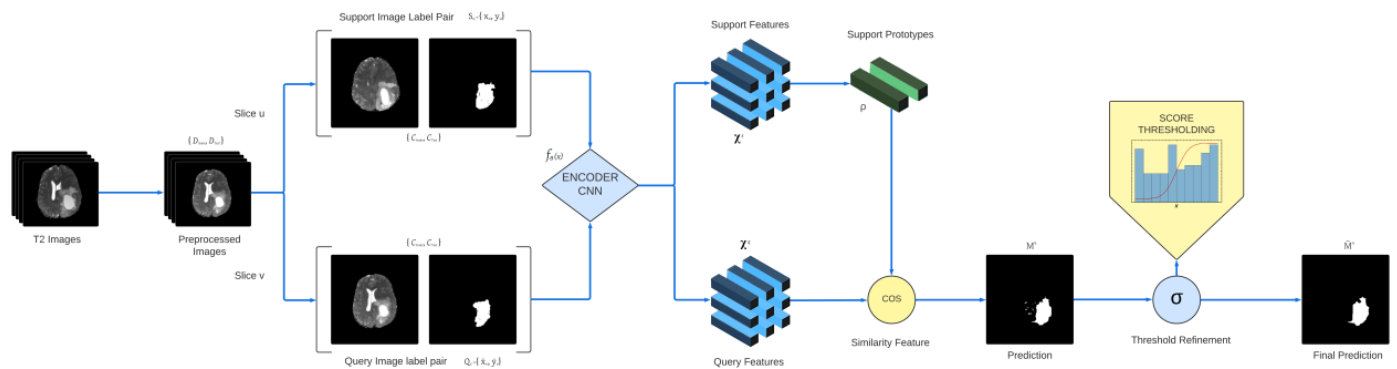

Figure 1. A pipeline of the proposed prototypical one-shot segmentation model to separate brain tumor regions.

3.2.1. Preprocessing and the Extraction of Prototype Features

A brain tumor is highly heterogeneous in respect of pixel intensity, size, and location. It results in highly imbalanced data, as the number of health voxels constitutes 98% of all voxels [10]. To overcome this challenge, our model uses prototype representations of labeled classes. Before that, each T2 multi-parametric MRI (mpMRI image) and its corresponding label is first pre-processed. The bright ends of each image are cut off to alleviate the off-resonance issue. Images and labels are then resampled to unify spacing. The ROIs of each image are then cropped out according to the label to unify image sizes.

Once the dimensionality of images is reduced, the feature encoder f_θ extract features for both query $f_\theta(x) : Q \to \chi^q$ and support images $f_\theta(x) : S \to \chi^s$. As our aim is to model only the foreground class in each episode, we considered only the prototypical features of foreground class s to be extracted. The feature map χ^q is then resized to mask dimensions (X, Y). Each foreground prototype ρ in embedding space is calculated using masked average pooling (MAP) for each foreground class:

$$\rho = \sum_c \frac{\sum_{(i,j)} \chi^q(i,j).[\eta(i,j) = c]}{\sum_{(i,j)}[\eta(i,j) = c]} \tag{1}$$

where (i,j) indexes the spatial locations of query slice and $\eta(i,j)$ denotes the Dirac measure which represents the almost sure probability outcome in the sample space. The Dirac measure is computed by:

$$\eta_{(i,j)}(Z) = l_Z(i,j) = \begin{cases} 0, \forall (i,j) \notin Z \\ 1, otherwise \end{cases} \tag{2}$$

where $Z_{(i,j)} \leftarrow \left[Q^{(i,j)} = c\right]$ is a sampling function that extracts given spatial coordinates (i,j) from query image Q having c foreground class. Figure 2 shows samples of multi-parametric MRI images from the dataset depicting Flair, T2, T1ce, T1 modalities, and label. The yellow, green, and red regions on the label denote the tumor regions of different levels.
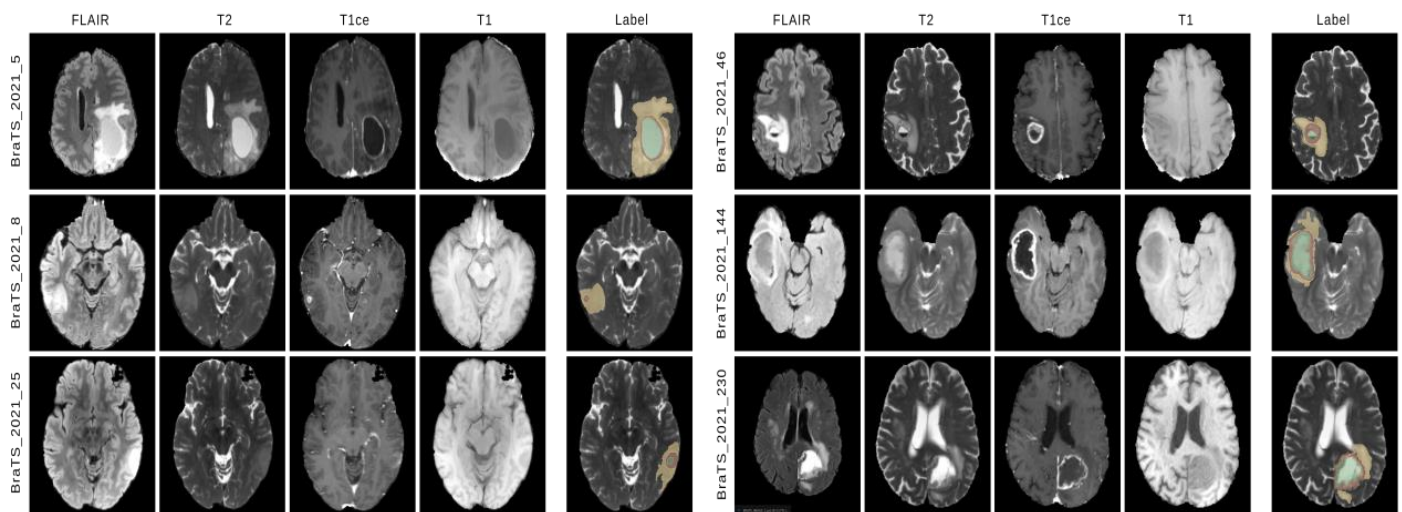
Figure 2. Illustration of six samples of multi-parametric MRI images (slice containing tumor region) from the BraTS 2021 dataset depicting Flair, T2, T1ce, T1 modalities, and label. The yellow, green, and red regions on the label indicate the tumor regions of different levels.

3.2.2. Similarity Feature-Based Segmentation

In order to differentiate query images c from class prototypes, we used a metric learning-based approach based on non-parametric thresholds. This is performed using a similarity measure R (Tucker's congruence coefficient) to compute similarities between foreground prototype feature class ρ and query feature χ^q, which is computed by:

$$R_{(i,j)} = -\alpha \frac{\chi^q_{(i,j)} \cdot \rho}{\sqrt{\left(\chi^q_{(i,j)}\right)^2 \cdot \rho^2}} \tag{3}$$

where α is a scaling factor whose value is set to 20. Studies [9,34] indicated that multiplying Tucker's congruence coefficient by α attains considerable performance compared to using Euclidean distance for similarity measures. It enables incongruent query feature vectors to receive a relative score of α. To obtain the final dense predicted mask, we performed soft thresholding with sigmoid function $\sigma(.)$ along with an evaluated parameter β:

$$M^q_{(i,j)} = 1 - \sigma\left(R_{(i,j)} - \beta\right) \tag{4}$$

This enables the query feature vectors with a similarity score below β to achieve a minimum foreground probability which is set to 0.5, whereas feature vectors above 0.5 obtain a foreground probability below β.

Having an end-to-end network to be trained in an iterative manner, where each episode Q_i, S_i is taken as input. We calculated the binary cross-entropy segmentation loss as:

$$L_{seg} = \frac{-1}{N} \sum_{(i,j)} \sum_{p_c \in \rho} [\eta(i,j) = c].log(M)^q_{(i,j)} \tag{5}$$

where p_c represents prototype features of class c. Following [34], we have also added a prototype alignment regularization loss which reverses the query and supports set predictions enabling the information to flow both ways. This greatly helps the model to align queries and support prototypes. It is computed by:

$$L_{PAR} = \frac{-1}{XY} \sum_{(i,j)} \sum_{p_c^q \in \rho} [\eta(i,j) = c].log(M)^q_{(i,j)} \tag{6}$$

where p_c^q represents query features of class *c*. PAR helps the model to learn in the embedding space consistently and to align the prototypes generated from the support set.

Thus, the total loss for our model can be formulated as:

$$L_{Total} = L_{seg} + \lambda L_{PAR} \tag{7}$$

where λ works as a controlling parameter for the effect of PAR loss. In our experiments, different values of lambda did not lead to significant improvements in score, and hence λ was kept at 1.

The Algorithm 1 pertaining to foreground prototypes based on few-shot segmentation is shown below in Figure 3.

Algorithm 1. Foreground prototypes based on few-shot segmentation

---

Require: Pre-processing of images, such as identifying and cropping out ROI regions of the image.

Require: Base feature encoder model f_θ.

1.   Sample D_{train} and D_{test} from pre-processed images
2.   Initialize $\{C_{train}, C_{test}\}$ empty
3.   **for** Image set $\{I_{train}, I_{test}\}$ in $\{D_{train}, D_{test}\}$ **do**
4.   cur $\leftarrow$ 0
5.   **while** i = 0 to C x K **do**
6.   index $\leftarrow$ i + curr
7.   Sample slice u, v from $\{I_{train}[index], I_{test}[index]\}$
8. $S_u \leftarrow \{x_u, y_u\}$ //where x_u is T2 image and y_u is its corresponding label
9. $Q_v \leftarrow \{x_v, y_v\}$
10. $\{C_{train}, C_{test}\} \leftarrow \{\{S_{train}, Q_{train}\}, \{S_{test}, Q_{test}\}\} \in \{S_u, Q_v\}$
11.   **end while**
12.   **end for**
13.   **for** $\{C_{i\_train}, C_{i\_test}\}$ in $\{C_{train}, C_{test}\}$ **do**
14. $Q \leftarrow \{C_{i\_train}, C_{i\_test}\} \in$ Query set
15. $S \leftarrow \{C_{i\_train}, C_{i\_test}\} \in$ Support set
16. $\chi^q \leftarrow f_\theta(Q)$
17. $\chi^s \leftarrow f_\theta(S)$
18.   **end for**
19.   **for** k = 0 to len(χ^s) **do**
20. $\mathbb{P}[k] \leftarrow \rho(\chi_k^s)$ // ρ = masked average pooling function (MAP)
21.   **end for**
22.   **for** n = 0 to len(χ^q) **do**
23.   Predictions[n] $\leftarrow$ R($\mathbb{P}$[n], χ_n^s)
24.   Result[n] $\leftarrow$ σ (Predictions[n]) // σ = Threshold function to eliminate small artifacts
25.   **end for**

Here, C x K refers to C-shot K-way based segmentation. C stands for the number of classes and K for the number of samples from each class to train on. S_u and Q_v represents support and query sets for slice u and slice v, respectively. χ^q and χ^s are the features extracted from the query and support set, respectively. R (x,y) represents the cosine similarity between x and y objects.

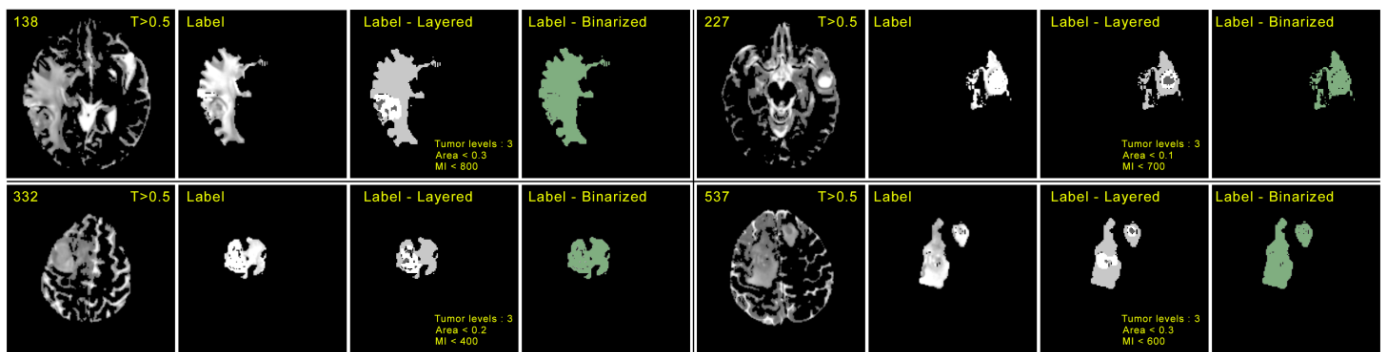The pseudocode logic for training our few-shot segmentation model is shown in Algorithm 2.

Figure 3. Demonstration of the procedure for determining the portion of the tumor in each slice. Here, the first column of each image contains the serial number according to the BraTS dataset and the corresponding random slice with an intensity threshold > 0.5. The second column contains the corresponding label of each scan. The third column depicts the layered structure of labels highlighting tumor levels, area, and mean intensities (MI). The fourth column contains the binarized form of label data.

Algorithm 2.

Pseudocode for training one-shot segmentation model.
Define the size and dimension of images in the dataset.
Initialization:
Support set $S = (x, y)$, Query set $Q = \{\hat{x}, \hat{y}\}$ contains image mask pair having a single slice.
Choose a backbone architecture as a feature encoder model f_θ with multi-class identification having θ parameters.
For epoch in S steps / I iterations.
    For each image label pair in the train dataset.
        Divide image label pairs into query and support sets containing random slices from the image. Both the query and support set contain one slice and its corresponding label image.
        Pass both sets through the feature encoder model to extract their features.
        Calculate support prototypes using masked average pooling of the support features set.
        Perform segmentation based on similarities found between query features and support prototypes.
        Update total loss consisting of segmentation and PAR loss.
    End for
End for

---

4. Experimental Settings

4.1. Dataset

We employed the multimodal Brain Tumor Image Segmentation 2021 (BraTS) dataset [37–39] as a benchmark to evaluate our model. It contains 1470 NIfTI files of 3D mpMRI scans with an average of 150 slices each. Out of these, 1250 were labeled and 220 were unlabeled validation sets. Only the T2-weighted NIfTI files of each scan were used for evaluation. In pre-processing, the following steps were carried out: (1) eliminate the 0.5% intensities as an effort to reduce the sharpness in the high end of the intensity curve, (2) re-sample the number of slices to around 20, preserving the same spatial resolution, and (3) resize the image to 256 × 256 in order to standardize the dimensions. Furthermore, each slice is replicated thrice in order to accommodate the RGB channels in our network. To increase the variability of the data, image transformations (shearing, geometrical changes, rotations, translations) were applied randomly on slices before feeding them to the network without augmenting the data.

4.2. Parameter Settings

We initialized the VGG-16-based encoder with pre-trained weights from Resnet101 trained on MS COCO [40]. The model is trained in an end-to-end manner with a stochastic gradient descent with a momentum of 0.9, a batch size of 1, a learning rate of 10^{-3}, and a decay rate of 0.97 per 1000 epochs. Additionally, another weight decay of 5×10^{-4} is used over 50 k iterations. Weights are applied on foreground classes (1.0) and background regions (0.2) with cross-entropy loss to address the class imbalances. Out of 1250 MRI scans, only 60 MRI scans were used for training purposes. The training process as a whole takes 2 h on a Nvidia RTX 3060 GPU. During testing, results are averaged based on a 5-fold cross-validation of over 1000 iterations per epoch. Foreground classes are produced as a form of mask over the background region in a binarized format. as shown in Figures 3 and 4.
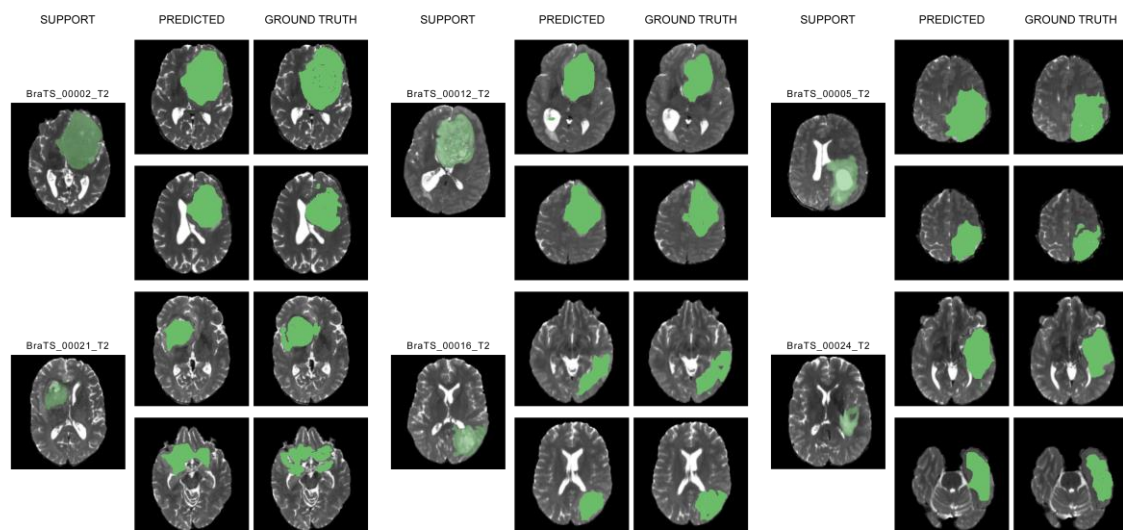

Figure 4. Qualitative results of our model with different support sets on two slices of different query images along with their ground truths. In these figures, tumor regions are depicted in green. The proposed learning method achieves desirable segmentation results which can be verified with their corresponding ground truth images.

4.3. Evaluation Metrics

We have utilized dice score and IOU score as metrics to evaluate our model. Dice similarity coefficient or dice score D is a reproducibility validation metric between segmentations X and Y, and is defined as:

$$D(X, Y) = \frac{2 * |X \cap Y|}{|X| + |Y|} \tag{8}$$

We used intersection over union (IoU) score to understand the ratio of the predicted segmentation area to the underlying ground truth. This helps us to understand the amount of overlap and mean average precision of our model. Precision is another valuable metric to identify the correctness of our predicted segmentation mask for foreground classes. It is calculated by dividing the number of correctly predicted tumor region pixels by the sum of correctly and incorrectly classified tumor region pixels.

4.4. Evaluation Protocols

In the inference stage, query scans are segmented episode-wise based on annotated support data. The position of the foreground class varies largely. Therefore, an evaluation protocol that does not depend on the target volume is required. Here, we sampled each slice from the support foreground volume and used this data to segment the whole query scan. To effectively carry out this technique, we selected a random slice from the middle of the support volume which contains a high amount of information regarding the foreground

class. For effective classification between foreground and background classes, we have clustered the different levels of tumors in a single class and it is fed forward into the neural network as a singular target feature. Scores are calculated based on the accuracy of the mask generated by our model for the target region.

5. Results and Discussion

The experimental results obtained from our FSS model based on training on the BraTS 2021 dataset (60 training images and 350 testing images) for different N-shot K-way methods are shown in Table 1. It is observed that in practicality, increasing N and K values results in a slight-to-no increase in score metrics. This could be due to the random extraction of slices for the support set which may or may not contain many similarities with the query images. Moreover, an increasing number of sample images (N-shot) increases the potential of sampling drastically different sizes, shapes, and intensity tumor regions of different scans which can affect the score in a negative way.

Table 1. Encoding shape (spatial) and memory usage statistics along with a feature encoding shape of (1,5,10)—shot (1,5)—way learning models.

Method	Input Size (MB)	Forward/Backward Pass Size (MB)	Estimated Total Size (MB)	Encoding Shape
1-shot 1-way	17.56	21,038.63	21,228.29	(22,256,32,32)
1-shot 5-way	21.76	24,863.83	25,057.69	(26,256,32,32)
5-shot 1-way	21.76	24,863.83	25,057.69	(26,256,32,32)
5-shot 5-way	42.73	43,989.76	44,204.69	(46,256,32,32)
10-shot 1-way	27.00	29,645.34	29,844.44	(31,256,32,32)
10-shot 5-way	68.95	67,897.39	68,138.44	(71,256,32,32)

The similarity between ground truth and predicted segmentation is assessed by two major comparison metrics. We used the mean and maximum dice score and mean and maximum intersection over union (IOU) score to assess our findings. The maximum metric of these scores is considered in order to quantify the extent to which our proposed one-shot segmentation proves to be effective. The performance metrics with 1-shot 1-way show a high average dice core of 83% and an average IOU of 80.9% compared to the other increasing number of samples carried out in the conventional studies. Model predictions along with their ground truths are depicted in Figure 4.

In a typical CNN, the number of features in each feature map is the constant times the number of input pixels n (typically the constant is <1). Convolving a fixed size filter across an image with n pixels takes O(n) time since each output is just the sum product between k pixels in the image and k weights in the filter, and k does not vary with n. Similarly, any max. or avg. pooling operation does not take more than linear time in the input size. Therefore, the overall runtime is still linear.

The total trainable parameters in our few-shot model are estimated to be 59,215,000, with a size of 172 MB. This is subject to change based on the feature encoder model used. The encoding shape represents the shape of χ^q and χ^s after query set (Q) and support set (S) is passed through f_θ, respectively. Table 1 shows the encoding shape and memory usage statistics of the proposed model.

The experimental results obtained from several deep learning segmentation methods for brain tumor segmentation are listed in Table 2. We compared our model with these frameworks, as there are no other few-shot models for brain tumor segmentation. The OM-Net [6] resolves the imbalance problem of the tumor segmentation class but utilized more training data. Segtran [7] generated effective receptive fields but had a more complex architecture. The use of autoencoders in NVDLMED [8] for segmentation makes it more computationally expensive. In comparison to these models [41], our proposed approach contains fewer parameters and is not much computationally expensive. Furthermore, our one-shot segmentation model took a very low inference time for each image, which makes it very suitable to test unseen images in clinical settings.

Table 2. Comparative table showing dice scores (in percentage) of different deep learning methods with the amount of training data (number of MRI scans) used along with dataset names.

Method	Training Data	Dice (%)	Dataset
OM-Net + CGap	559	91.59	BraTS 2015 + BraTS 2018
3DCNN + CRF	338	90.10	BraTS 2015 + ISLES 2015
Segtran	335	81.70	BraTS 2019
NVDLMED	285	87.04	BraTS 2018
AFN-6	274	89.30	BraTS 2015
CNN + 3D filters	274	80.00	BraTS 2015
Few-shot-based	**60**	**83.42**	**BraTS 2021**

Additionally, the proposed approach allows extracting the data required to classify the definition of borders between foreground and background classes, which can be trained using very few training images. This unlocks the possibilities of obtaining effective results in medical imaging applications where a sizable amount of data may not be present. Few-shot learning models have been proven to be effective in various other domains including mask aggregation [30], semantic scene segmentation [42], image classification [43], and text generation [44]. Using minimal data, the few-shot learning technique is able to achieve significant results in several scenarios. Table 3 shows the dice score results of 1 shot, 5-shot, and 10-shot for 1-way and 5-way segmentation.

Table 3. Results of (1, 5, 10)-shot (1, 5)-way segmentation on the BraTS 2021 dataset using mean and maximum dice and IoU metrics (in percentage). Scores are obtained from the validation of 250 randomly sampled volume datasets.

Method	Score Metrics				
	Avg. Dice	Max. Dice	Avg. IOU	Max. IOU	Precision
1-shot 1-way	**83.42 $\pm$ 0.35**	**83.85 $\pm$ 1.14**	80.97 $\pm$ 1.12	**81.57 $\pm$ 1.23**	**61.03 $\pm$ 0.38**
1-shot 5-way	83.78 $\pm$ 1.14	82.67 $\pm$ 2.81	**80.99 $\pm$ 1.16**	81.39 $\pm$ 1.61	60.78 $\pm$0.82
5-shot 1-way	83.28 $\pm$ 2.27	82.50 $\pm$ 3.13	79.44 $\pm$ 2.29	80.97 $\pm$ 0.25	60.45 $\pm$ 1.05
5-shot 5-way	83.41 $\pm$ 2.30	81.23 $\pm$ 3.43	79.64 $\pm$ 2.44	80.22 $\pm$ 1.68	60.61 $\pm$ 1.75
10-shot 1-way	82.89 $\pm$ 2.78	81.45 $\pm$ 2.25	79.59 $\pm$ 2.54	80.98 $\pm$ 3.02	79.62 $\pm$ 1.68
10-shot 5-way	83.40 $\pm$ 3.28	80.23 $\pm$ 4.17	80.02 $\pm$ 2.69	80.82 $\pm$ 2.93	59.88 $\pm$ 2.19

The critical aspect of this study is to identify the largely homogeneous foreground class in terms of size, shape, and intensity values. This sporadically leads to arbitrary or inaccurate identification of foreground region in volumes where tumor pixel intensity (MI < 200) and the background brain region have almost similar values. Our approach of a single prototype for modeling the foreground classes is based on its uniformity. However, if this uniformity is not met, and the foreground consists of multiple distinct regions with strong edges, and then one prototype may not suffice. One option to capture multiple foreground regions during inference is to take inspiration from [45] and cluster the features into consolidate multiple foreground classes into one region.

6. Conclusions

In this work, we proposed a prototypical one-shot learning framework to segment brain tumors on MRI scans. The proposed method is able to extract robust prototypes and performed segmentation of the foreground class using non-parametric distance calculation. It is observed that despite using a small fraction of data, we are able to generate almost equal results compared to the other deep learning approaches. As it is a metric learning-based approach, it is not bound to be used on a particular type of MRI-scanned image. Furthermore, it is not required to set a large number of support images as, it is observed that a single support image (1-shot, 1-way) proves to be effective for segmentation. Moreover, increasing the support set size reduces the performance slightly. This may be due to

differences in similarity scores obtained across various heterogeneous-sized foreground classes. Therefore, we conclude that the foreground prototype-based few-shot learning approach proposed reduces the amount of training and testing time significantly compared to conventional deep learning methods. With this model, we prove that significant results can be obtained in clinical settings where a high amount of data may not be present. Future work will be in the direction of developing a generic framework that is not bound to any specific dataset across the healthcare domain.

References

1. Dasgupta, A.; Gupta, T.; Jalali, R. Indian data on central nervous tumors: A summary of published work. *S. Asian J. Cancer* **2016**, *5*, 147–153.
2. Wrensch, M.; Minn, Y.; Chew, T.; Bondy, M.; Berger, M.S. Epidemiology of primary brain tumors: Current concepts and review of the literature. *Neuro-Oncology* **2002**, *4*, 278–299. [CrossRef] [PubMed]
3. Hicham, M.; Bouchaib, C.; Lhoussain, B. Convolutional Neural Networks for Multimodal Brain MRI Images Segmentation: A Comparative Study. In *Proceedings of the Smart Applications and Data Analysis: Third International Conference, SADASC 2020, Marrakesh, Morocco, 25–26 June 2020*; Springer International Publishing: New York, NY, USA, 2020; Volume 1207, pp. 329–338.
4. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.-M.; Larochelle, H. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **2017**, *35*, 18–31. [CrossRef] [PubMed]
5. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2016**, *36*, 61–78. [CrossRef]
6. Zhou, C.; Ding, C.; Wang, X.; Lu, Z.; Tao, D. One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. *IEEE Trans. Image Process.* **2020**, *29*, 4516–4529. [CrossRef]
7. Li, S.; Sui, X.; Luo, X.; Xu, X.; Liu, Y.; Goh, R. Medical image segmentation using squeeze-and-expansion transformers. *arXiv* **2021**, arXiv:2105.09511.
8. Myronenko, A. 3D MRI brain tumor segmentation using Autoencoder regularization. In *Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, Granada, Spain, 16 September 2018*; Springer International Publishing: New York, NY, USA, 2019; pp. 311–320.
9. Hansen, S.; Gautam, S.; Jenssen, R.; Kampffmeyer, M. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. *Med. Image Anal.* **2022**, *78*, 102385. [CrossRef]
10. Bendou, Y.; Hu, Y.; Lafargue, R.; Lioi, G.; Pasdeloup, B.; Pateux, S.; Gripon, V. Easy—Ensemble Augmented-Shot-Y-Shaped Learning: State-of-The-Art Few-Shot Classification with Simple Components. *J. Imaging* **2022**, *8*, 179. [CrossRef]
11. Jie, C.; Luming, Z.; Naijie, G.; Xiaoci, Z.; Minquan, Y.; Rongzhang, Y.; Meng, Q. A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *J. Vis. Commun. Image Represent.* **2019**, *58*, 316–322.
12. Lucas, F.; Wenqi, L.; Luis, C.G.P.H.; Jinendra, E.; Neil, K.; Sebastian, O.; Tom, V. *Scalable Multimodal Convolutional Networks for Brain Tumour Segmentation; Medical Image Computing and Computer Assisted Inter-vention—MICCAI 2017*; Springer: Cham, Switzerland, 2017; Volume 10435.

13. ZainEldin, H.; Gamel, S.A.; El-Kenawy, E.-S.M.; Alharbi, A.H.; Khafaga, D.S.; Ibrahim, A.; Talaat, F.M. Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization. *Bioengineering* **2023**, *10*, 18.

14. Feng, X.; Tustison, N.J.; Patel, S.H.; Meyer, C.H. Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. *Front. Comput. Neurosci.* **2020**, *14*, 25. [CrossRef]

15. Yogananda, C.G.B.; Shah, B.R.; Vejdani-Jahromi, M.; Nalawade, S.S.; Murugesan, G.K.; Yu, F.F.; Pinho, M.C.; Wagner, B.C.; Emblem, K.E.; Bjørnerud, A.; et al. A fully automated deep learning network for brain tumor segmentation. *Tomography* **2020**, *6*, 186–193. [CrossRef]

16. Madhupriya, G.; Guru, N.M.; Praveen, S.; Nivetha, B. Brain tumor segmentation with deep learning technique. In Proceedings of the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 23–25 April 2019; pp. 758–763.

17. Wang, G.; Li, W.; Ourselin, S.; Vercauteren, T. Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Proceedings of the Conjunction with MICCAI 2017, Quebec City, QC, Canada, 14 September 2017, Revised Selected Papers 3*; Springer International Publishing: New York, NY, USA, 2018; pp. 178–190.

18. Liu, J.; Li, M.; Wang, J.; Wu, F.; Liu, T.; Pan, Y. A survey of MRI-based brain tumor segmentation methods. *Tsinghua Sci. Technol.* **2014**, *19*, 578–595.

19. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv. (Csur)* **2020**, *53*, 1–34. [CrossRef]

20. Bai, W.; Chen, C.; Tarroni, G.; Duan, J.; Guitton, F.; Petersen, S.E.; Guo, Y.; Matthews, P.M.; Rueckert, D. Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, Proceedings of the 22nd International Conference, Shenzhen, China, 13–17 October 2019; Proceedings, Part II 22 2019*; Springer International Publishing: New York, NY, USA, 2019; pp. 541–549.

21. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. Self-supervised learning for medical image analysis using image context restoration. *Med. Image Anal.* **2019**, *58*, 101539. [CrossRef]

22. Tong, T.; Wolz, R.; Wang, Z.; Gao, Q.; Misawa, K.; Fujiwara, M.; Mori, K.; Hajnal, J.V.; Rueckert, D. Discriminative dictionary learning for abdominal multi-organ segmentation. *Med. Image Anal.* **2015**, *23*, 92–104. [CrossRef]

23. Zhang, C.; Lin, G.; Liu, F.; Yao, R.; Shen, C. Canet: Class-Agnostic Segmentation Networks with Iterative Refinement and Attentive Few-Shot Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 5217–5226.

24. Yu, M.; Liang, L.; Biqing, H.; Xiu, L. Few-shot RUL estimation based on model-agnostic meta-learning. *J. Intell. Manuf.* **2022**, 1572–8145.

25. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.

26. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. In *Proceedings of the Advances in Neural Information PROCESSING systems 29, Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain, 5–10 December 2016*; Morgan Kaufmann Publishers Inc.: Burlington, MA, USA, 2016.

27. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In *Proceedings of the Advances in Neural Information Processing Systems 30, Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017*; MIT Press: Cambridge, MA, USA, 2017.

28. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.

29. Fan, Q.; Pei, W.; Tai, Y.-W.; Tang, C.-K. Self-Support Few-Shot Semantic Segmentation. *arXiv* **2022**, arXiv:2207.11549.

30. Shi, X.; Wei, D.; Zhang, Y.; Lu, D.; Ning, M.; Chen, J.; Ma, K.; Zheng, Y. Dense cross-query-and-support attention weighted mask aggregation for few-shot segmentation. In *Proceedings of the Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022*; Springer Nature: Cham, Switzerland, 2022; pp. 151–168.

31. Boudiaf, M.; Kervadec, H.; Masud, Z.I.; Piantanida, P.; Ben Ayed, I.; Dolz, J. Few-shot segmentation without meta-learning: A good transductive inference is all you need? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13979–13988.

32. Cui, H.; Wei, D.; Ma, K.; Gu, S.; Zheng, Y. A unified framework for generalized low-shot medical image segmentation with scarce data. *IEEE Trans. Med. Imaging* **2020**, *40*, 2656–2671. [CrossRef]

33. Voulodimos, A.; Protopapadakis, E.; Katsamenis, I.; Doulamis, A.; Doulamis, N. A few-shot U-net deep learning model for COVID-19 infected area segmentation in CT images. *Sensors* **2021**, *21*, 2215. [CrossRef]

34. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9197–9206.

35. Li, S.; Liu, J.; Song, Z. Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 2435–2445. [CrossRef]

36. Ouyang, C.; Biffi, C.; Chen, C.; Kart, T.; Qiu, H.; Rueckert, D. Self-Supervision with Superpixels: Training Few-shot Medical Image Segmentation without Annotation. *arXiv* **2020**, arXiv:2007.09886.

37. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 1–13. [CrossRef]

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

39. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef]

40. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollar, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer International Publishing: New York, NY, USA, 2014; pp. 740–755.

41. Han, Z.; Hao, Y.; Dong, L.; Sun, Y.; Wei, F. Prototypical Calibration for Few-shot Learning of Language Models. *arXiv* **2022**, arXiv:2205.10183.

42. Lang, C.; Cheng, G.; Tu, B.; Han, J. Learning what not to segment: A new perspective on few-shot segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.

43. Hu, S.X.; Li, D.; Stuhmer, J.; Kim, M.; Hospedales, T.M. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.

44. Schick, T.; Schütze, H. Few-shot text generation with pattern-exploiting training. *arXiv* **2020**, arXiv:2012.11926.

45. Huo, Y.; Xu, Z.; Xiong, Y.; Aboud, K.; Parvathaneni, P.; Bao, S.; Bermudez, C.; Resnick, S.M.; Cutting, L.E.; Landman, B.A. 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage* **2019**, *94*, 105–119. [CrossRef]