Few-shot learning for brain tumor segmentation from MRI images

Abdelouahad Achmamad
LITIS - Quantif,
University of Rouen
Rouen, France
abdelouahad.achmamad@univ-rouen.fr

Fethi Ghazouani
LITIS - Quantif,
University of Rouen
Rouen, France
fethi.ghazouani@univ-rouen.fr

Su Ruan
LITIS - Quantif,
University of Rouen
Rouen, France
su.ruan@univ-rouen.fr

***Abstract*—Automated brain tumor segmentation from magnetic resonance imaging (MRI) scans is of great significant for brain diagnosis. However, how to segment tumor regions accurately with very limited labeled MRI images is still an appealing challenge. To tackle this issue, the present paper proposes to use few-shot under meta-learning setup. The idea is to exploit knowledge derived from a handful annotated support images during episodic learning to guide the segmentation of query images. Specifically, for each episode, the encoder extract feature maps for the both support and query images. Then, a masked average pooling is performed with the support mask to get the guidance features by only considering the target categories belonging to the support image. We use convolution operation to construct the relationship between the guidance features and query feature maps. With an aim to achieve better generalization on few-shot semantic segmentation, decoder based U-Net architecture is used. The proposed method is applied on the benchmark BraTS2021 dataset for brain tumor segmentation. The experimental results in terms of quantitative and qualitative are satisfactory in generating brain tumor segments. The presented segmentation method might be useful to help doctors perform a subtle diagnosis so that the life expectancy of patients becomes longer.**

***Index Terms*—Few-shot learning, brain tumor segmentation, meta-learning, U-Net architecture, Masked average pooling.**

I. Introduction

A brain tumor is a primitive of abnormal cells growing rapidly inside the brain [1]. There are a wide variety of brain tumors, all ranging from non-cancerous (benign) to cancerous (malignant) in nature depending on the severity of tumorous cell [2]. According to World Health Organisation(WHO) statistics, brain tumor is considered as one of the most prevalence cancers [3]. World-wild, an approximately 308,102 people were diagnosed and followed up with primary brain tumor in 2020 [4]. Therefore, an early diagnosis based on medical imaging modalities including magnetic resonance image (MRI) scan is mandatory to prevent progression and even more importantly to reduce mortality rate due to brain tumors. For this purpose, multiple automated brain tumor segmentation techniques using convolution neural networks (CNNs) have been established by investigators for a few years now [1], [5]–[8]. As result, their sophisticated methods can effectively assist the radiologist to timely diagnose and perform consistent treatment of brain tumors during radiotherapy or surgery.

To date, CNNs have shown a great performance in computer vision tasks especially in the field of semantic segmentation [9], [10]. However, their extreme robustness is regularly depend on large sets of labeled data, which are often not available in the medical image area [11], [12]. Furthermore, when the model is trained with insufficient labeled data, there is a high probability of an over-fitting problem to be occurred [13]. Few-shot learning (FSL) is an excellent alternative to classical learning frameworks because it may keep models away from the aforementioned shortcoming while improving their generalization performance [14], [15]. Taking inspiration from how human beings can quickly adapt and learn new concepts with only a few labeled data, FSL attempts to mimic this human-level performance. In this regard, researches toward meta-learning paradigm is attractive since it has proved its simplicity and effectiveness to deal with the FSL challenge [16]. The basic idea behind this solution is learning to learn throughout a large number of training episodes. At each episode level, k labeled samples denoted as support set are used to predict unlabeled samples usually referred to as query set [11]. FSL under episodic setting has been mostly explored in the natural images for the classification [17], [18], and also for the segmentation tasks [19], [20].

However, the investigation of such a new learning paradigm in medical images remains relatively scarce despite of its valuable application potential. In the current state of the art for medical few-shot segmentation, there exist limited works. For instance, Arnab *et al.* [21] first addressed the issue of multi-modal medical data segmentation from a FSL perspective. Following the success of adversarial learning in semi-supervised classification, the authors used generative adversarial networks (GANs) to train a segmentation model with annotated and non-annotated images. They showed well how their proposed method can avoid over-fitting by learning to distinguish true and false patches achieved by a generative network. Ouyang *et al.* [22] leverage few-shot segmentation together with superpixel-based self-supervised learning in order to remove the necessity of having annotations during training. Moreover, they used an adaptive local prototype pooling module for prototypical networks to solve the foreground-background imbalance problem in medical image segmentation. In the work introduced by Roy *et al*. [23], a novel architecture for few-shot segmentation incorporating 'channel squeeze and spatial excitation' modules, has been applied to tackle the absence of pre-trained networks to start from, and 3D medical scans. Zhao *et al*. [24] outlined an another automated approach to data augmentation leverages unlabeled MRI brain scans in a semi-supervised learning framework. This work demonstrate the utility of data augmentation and learned transformation on one-shot segmentation by learning to synthesize and realistic labeled medical images. Inspired by the latter study, Fayjie *et al*. [25] investigated the role of unlabeled data trough surrogate tasks, in the task of performing skin lesion segmentation by using semi-supervised FSL scenario. Using similar medical dataset, Khadka *et al*. [26] evaluted successfully an optimization-based implicit model agnostic meta-learning (iMAML) algorithm, where the goal is to improve model parameter estimation in the context of few-shot segmentation. Compared to others FSL methods, their approach enhanced results across all the dataset in terms of generalization ability with a very few labeled samples.

In the present study, we formulate the FSL under meta-learning setup for brains tumor segmentation (BraTS). The main objective is to segment precisely tumors regions to help predict cancer outcomes after the treatment. Note that, the proposed method inspires the recent work on semi-supervised for medical image segmentation [25].

The reminder of this manuscript is organised as follow: section 2 provides detailed information about the proposed method. Section 3 gives a brief description of the chosen medical imaging dataset, evaluation metrics, experimental setup used. Section 4 presents quantitative and qualitative results. Lastly, conclusion is drawn in section 5.

II. Proposed Method

A. Problem Setting

Generally, we have a training set $X \in \mathbb{R}^{(H \times W \times Ch)}$ with corresponding class segmentation labels $Y \in \mathbb{R}^{(H \times W \times Ch)}$ taken from a dataset D. In case of few-shot semantic segmentation, the model should be trained on dataset D_{train} containing samples with training classes C_{train}, and then evaluated on a dataset D_{test} containing samples with unseen testing classes C_{test}. Where, the class set $C_{train} \in D_{train}$ has no overlap with the class set $C_{test} \in D_{test}$, which are mutually disjoint; i.e., $C_{train} \cap C_{test} = \emptyset$. Few-shot semantic segmentation focuses on both training the model with reference to a few labeled data and fastly generating it over unknown classes during testing phase. To achieve this purpose, we adopt the standard N-way K-shot meta-learning strategy which is an effective approach for few-shot semantic segmentation to better exploit the training and testing sets. In this setting, each support set is composed of N classes with K-shot learning samples from which query learn. Following this setup, both D_{train} and D_{test} are sampled into several of randomly episodes, and each episode contain support set S and a query set Q with $N_{Support}$ and N_{Test} samples, respectively. Considering meta-training setup, S and Q can be expressed as follows:

$$S = \{X_i^s, Y_i^s\}_{i=1}^{N_{Support}} \subset D_{train} \tag{1}$$

$$Q = \{X_i^q, Y_i^q\}_{i=1}^{N_{Test}} \subset D_{train} \tag{2}$$

B. Network Architecture

The proposed few-shot segmentation framework is depicted in Fig. 1. It consists of two branches, i.e., segmentation branch (top) and reconstruction branch (bottom). Firstly, the encoder VGGnet-16 (pretrained on ImageNet) takes the support image and query image as input to extract the corresponding feature embeddings $f_s \in \mathbb{R}^{(H' \times W' \times D)}$ for support and $f_q \in \mathbb{R}^{(H' \times W' \times D)}$ for query. The variables (H', W') and D denote the spatial size of the feature maps and the dimension of the embedding space, respectively. Afterwards, by passing the support feature maps via masked average pooling (MAP) operation along with support mask $\tilde{Y}_s \in \{0,1\}^{(H' \times W')}$, we obtain the foreground prototypes-related to the object regions of the labeled support image while excluding the background content. These prototypes will further be used as representative feature of reference objects for segmenting unseeen query image. More specifically, before applying MAP, we need to resize the support mask to the same size as the support feature maps by using down-sampling procedure. Then, the kth element p_k of the prototype p is calculated by averaging the pexels within the regions of interest on the kth feature map with respect to the following mathematical expression bellow:

$$p_k = \frac{\sum_{x,y=1}^{W' \times H' \times D} f_{s,k} \bigodot \tilde{Y}_s}{\sum_{x,y=1}^{W' \times H' \times D} \tilde{Y}_s} \tag{3}$$

where $\bigodot$ refers to the Hadamard product and $f_{s,k}$ is the kth feature embeddings for support. As an advantage, the MAP does not harm the statistic distribution of input image, allowing us to apply unified network within image support and query without losing spatial resolution. The problem with existing methods is that the learned features from image support can biased towards the background content [27].

After extracting prototypes, we expand each single prototype to the identical spacial resolution of the query feature maps by using unpooling operation. Hereby, instead of adopting similarity guidance method used in the most previous methods [11], [27], we use the convolution as an alternative approach to match the up-sampled prototypes with the query feature maps. The convolution serves as guidance attention to localize the desired object regions. Finally, we forward the generated guidance maps through decoder based U-Net for precisely predicting the segmentation mask $\hat{Y}_q \in \{0,1\}^{H' \times W'}$.

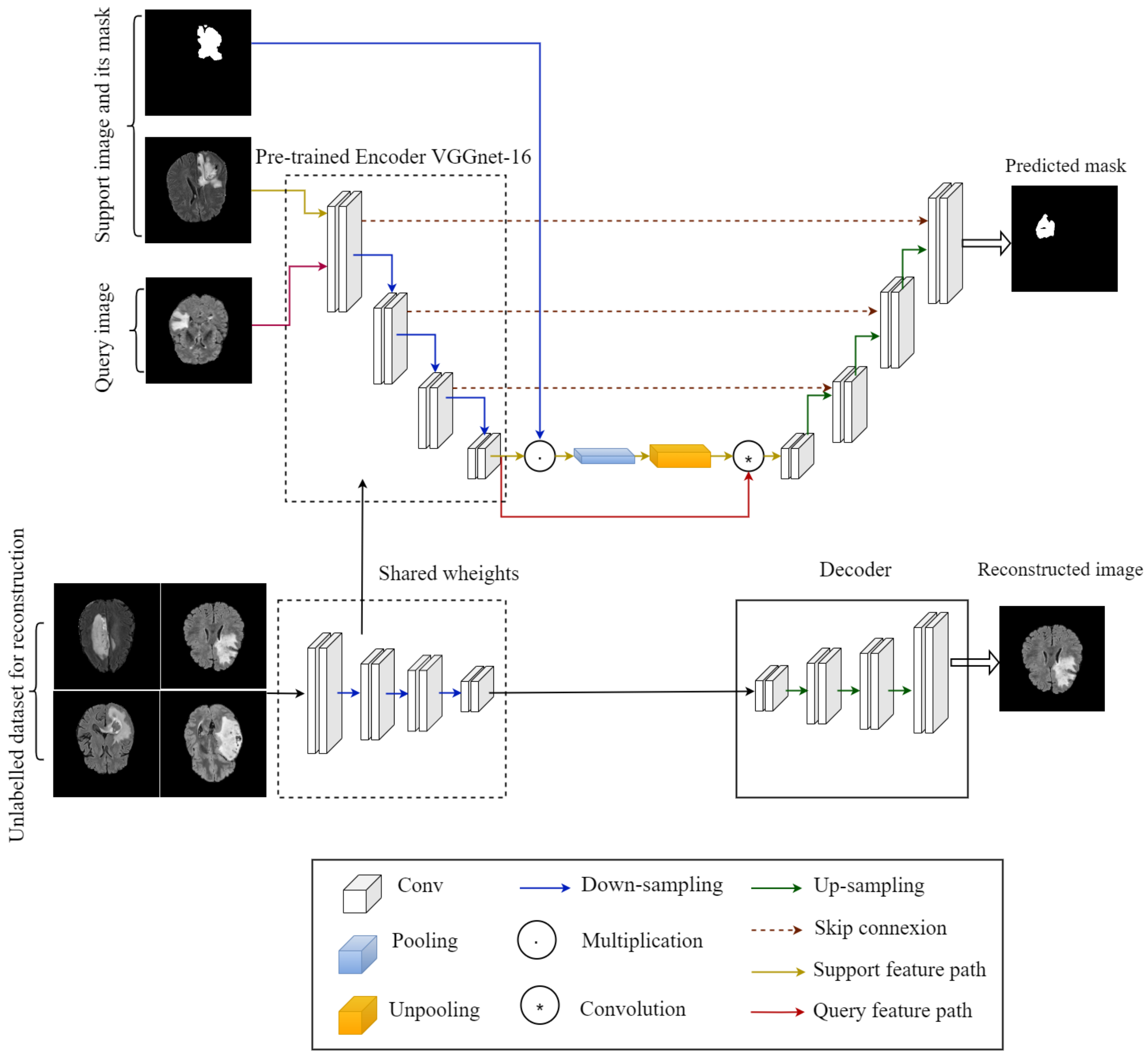

Fig. 1. Illustration of the proposed framework for few-shot semantic segmentation.

C. Auxiliary task

Major benefits of this auxiliary task resulting from the generalization of a trained model in weightsharing include compensation for the lack of labeled support images on new categories and boosting the base model for better adaptability to small training dataset. To further quantify these advantages, we evaluate the encoder-decoder architecture by incorporating unlabeled images. This unsupervised learning method focuses on training a network to reconstruct an input image.

D. Loss function

The loss function plays an important role in solving the optimization problem. It directly affects the convergence of the model during learning. Thus, If it does not penalize incorrect output in a manner appropriate to its magnitude, it can delay convergence and affect learning. In the literature, we can find several loss functions, among them, we distinguish the cross-entropy loss which is very widely employed in deep learning [28]. In fact, the cross entropy loss measures the error between two probability distributions for a given random variable. For example, when we have only two classes, the binary cross entropy loss ($\mathfrak{L}_{BCE}$) is defined as the following:

$$\mathfrak{L}_{BCE}(y, \hat{y}) = -(ylog(\hat{y})) + (1-y)log(1-\hat{y}) \quad (4)$$

where $y, \hat{y} \in \{0, 1\}^{H \times W}$, y refers to the ground truth label and $\hat{y}$ refers to predicted output value. In this study, $\mathfrak{L}_{BCE}$ is applied to evaluate the training model performance for both image segmentation and image reconstruction.

III. Materials

A. Dataset description

The benchmark dataset used in this work comes from BraTS 2021 which is publicly available as NIfTI files (.nii.gz) and delivered by the Medical Image Computing and Computer-Assisted Intervention Society (MICCAI) [29]. It contains four sequence modalities for each patient namely, T1-weighted (T1w), T1-weighted contrast-enhance (T1CE), T2-weighted (T2w) and fluid attenuated inversion recovery (Flair) with corresponding ground-truth mask hand-annotated by experienced neurologists. The size of all 3D MRI scans is 240 × 240 × 155 voxels, which were acquired from different scanners and protocols across many medical institutions. Here, only Flair modality is used to test our method. As our method work on 2D image, all NIfTI files related to Flair are converted to size of (224×224) pixels. It is worth noting that some slices are completely black (i.e. they do not contain any information), and these are excluded in the first step of the pre-possessing. Moreover, it is necessary to eliminate the parts where the tumor is missing from the MRI slices enabling the model to detect only the features of the tumor region. Furthermore, MRI slices may contain artifacts such as bias field distortion leading to intensity in-homogeneity. Therefore, data must be normalized before it is given to the model for training to eradicate the effects of artifact as well as to further enhance the segmentation results.

B. Evaluation metrics

Following previous works [25], [29], we use the widely adopted dice similarity coefficient (DSC) as our major evaluation metric for the proposed few-shot semantic segmentation framework. DSC is a metric used to find how well a model predicts the positive class correctly. Given a reference segmentation S_{ref} , the DSC of a predicted segmentation S_{pred} can be calculated as given in equation (5).

$$DSC(S_{pred}, S_{ref}) = \frac{2\,|S_{pred} \cap S_{ref}|}{|S_{pred} + S_{ref}|} \times 100\% \tag{5}$$

The range of the DSC is between 0 and 1, with 1 corresponding to a perfect intersection between the predicted segmentation and ground-truth mask.

C. Implementation details

The medical imaging community has witnessed the tremendous success of the U-Net architecture especially in brain tumor segmentation [30]. In our implementation, we use the standard U-Net for few-shot semantic segmentation. U-Net consists of encoder and decoder paths. The pretrained VGGnet-16 encoder on the large-scale ImageNET dataset is employed as the backbone for feature extraction. This helps the built model to converge quickly. Our pretrained VGGnet-16 encoder consists of four encoding layers. Each layer consists of two 3 × 3 2D convolutions, followed by a ReLU activation, and a 2 × 2 max pooling operation. The decoder path also composed of four decoding layers, each of which has a residual block preceded by a 2 × 2 convolutional transpose layer. Every output feature map of an encoder layer is concatenated with the corresponding input decoder layer using a skip connection with the same filter number. This connectivity aids to retain lost of information after a series of down-sampling operation in the encoder path. On the other hand, the encoder-decoder architecture of the auxiliary model is similar to that made up for the base model, without skip connection. Both designed models are trained under NVIDIA GPU with 8 GB RAM. A total number of 40 epochs is set and Adam optimizer is used for optimization with initial learning rate of $1 \times e^{-5}$. All simulations are implemented using Keras/Tensorflow bakend in Linux/Ubuntu 20.04.4 LTS operating system.

IV. Results

Throughout the following section, we will report the obtained qualitative and quantitative results. We tested our proposed method on BraTS2021 dataset, 75 % for meta-training and the rest reserved for meta-testing. The results of the segmentation based U-Net architecture are provided in Table I for one-way with different K-shot setups. At each experiment, the base model is trained and evaluated on setups of 1-shot, 5-shot and 10-shot. For each setup, the mean and standard deviation of DSC are computed. DSC increased significantly when varying the K-shot values. As can be seen from table, the performance improvement compared to other settings is more important for 1-way 10-shot setting. This clearly implies that adding images to the support set leads to improved segmentation performance.

TABLE I
Quantitative results on BraTS dataset with different K-shot setting

Method	N-way	K-shot	N-episode	DSC($\mu \pm std$)%*
		1		0.57 ± 0.19
FSL with U-Net	1	5	50	0.63 ± 0.16
		10		0.65 ± 0.17

* μ= mean value, std = standard deviation value.

To demonstrate the effectiveness of few-shot semantic segmentation based U-Net, a comparison against baseline segmentation framework using auto-encoder is performed. The Table II reports the comparison results for 1-way 10-shot setting. We can observe that our proposed method outperforms FSL based auto-encoder architecture with DSC improvement of 4%. One possible explanation for this could be due to the addition of the skip connection which contributes positively to more useful feature extraction with better representation. Thereby, helping the model to learn more categories from support image for few-shot semantic segmentation.

Fig. 2 presents the qualitative examples of 1-shot semantic segmentation on unseen query image. From left to right, we show the support slice with manual annotation, query input with its manual annotation, and predicted query mask. We

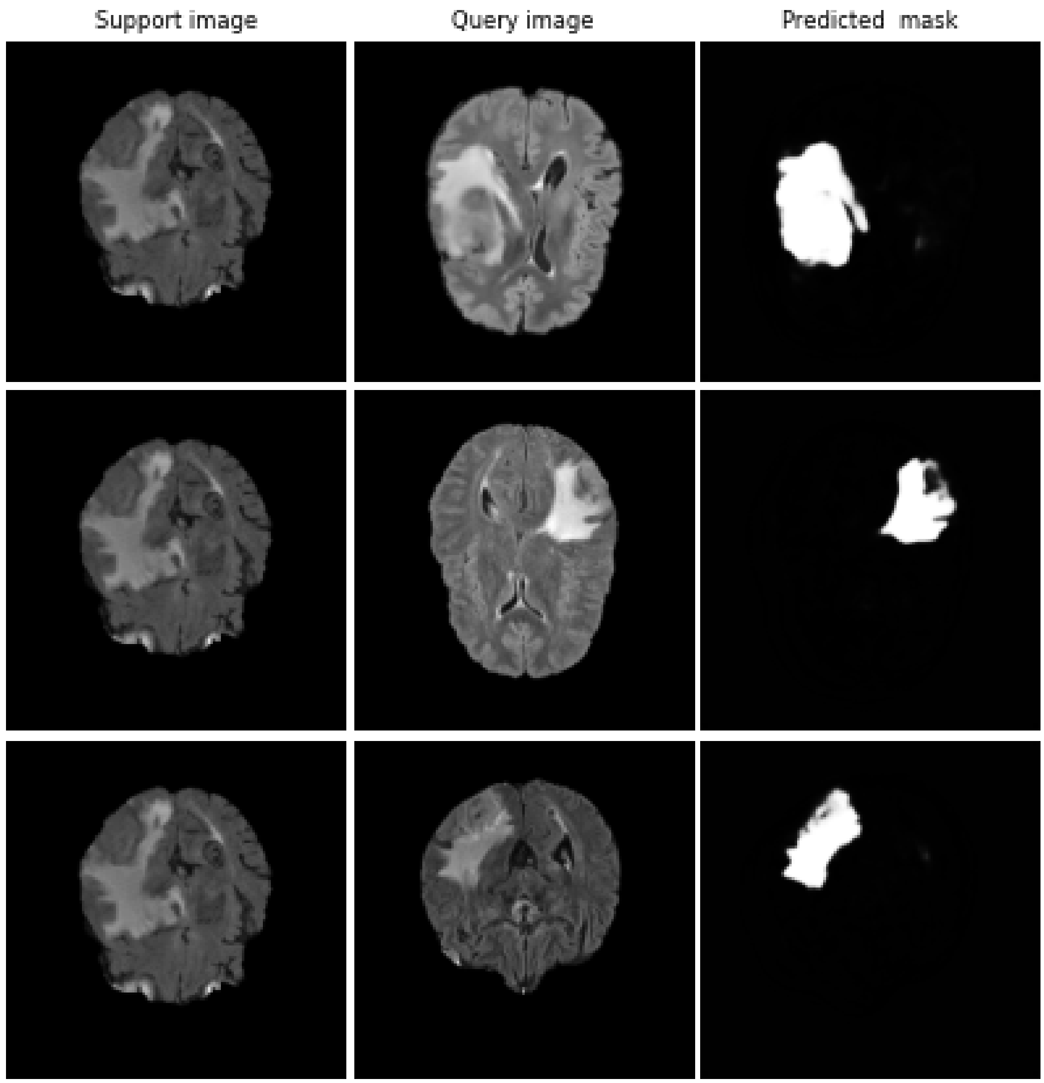

Fig. 2. Qualitative results for 1-shot semantic segmentation. Left to right: Support images, Query images and segmentation results of a query slice

TABLE II
COMPARISON AGAINST ORIGINAL METHOD

	1-way 10-shot
Method	DSC($\mu \pm std$)%
FSL with U-Net	0.65 ± 0.17
FSL with autoencoder	0.61 ± 0.14

can first see that one-shot semantic segmentation based U-Net can accurately differentiate the tumor regions from the background with the guidance of the support slice, even though some support and query slices do not share many appearance similarities. We also observe that the segmentation results are promising despite the variation in the tumor shape and location in both support and the query slices.

V. CONCLUSIONS

In this paper, we outlined a few-shot under meta-learning setup for brain tumor segmentation. The proposed framework used the advantage of combining U-Net architecture, which showed enhanced segmentation performance and generalization capability compared to the original method. Our framework is simple, effective and can be implemented to further benefit many upcoming few-shot semantic segmentation methods. In the future work, brain tumor volumetric segmentation based on FSL will be investigated.

ACKNOWLEDGMENT

In the context of investing in research, technological development and innovation, promoting the development of information and communication technologies. This work was supported by the European Regional Development Fund (FEDER), which funded by the European Project (Reference: FEDER PREGLIO N°21E03175

REFERENCES

[1] Chinnam, S. K. R., Sistla, V., Kolli, V. K. K. (2022). Multimodal attention-gated cascaded U-Net model for automatic brain tumor detection and segmentation. Biomedical Signal Processing and Control, 78, 103907.

[2] Balwant, M. K. "A Review on Convolutional Neural Networks for Brain Tumor Segmentation: Methods, Datasets, Libraries, and Future Directions." IRBM (2022).

[3] WEN, Patrick Y. et PACKER, Roger J. The 2021 WHO classification of tumors of the central nervous system: clinical implications. Neuro-oncology, 2021, vol. 23, no 8, p. 1215-1217.

[4] International agency for research cancer https://gco.iarc.fr/ [accessed 27 June 2022]

[5] AKBAR, Agus Subhan, FATICHAH, Chastine, et SUCIATI, Nanik. Single level UNet3D with multipath residual attention block for brain tumor segmentation. Journal of King Saud University-Computer and Information Sciences, 2022.

[6] XU, Weijin, YANG, Huihua, ZHANG, Mingying, et al. Brain tumor segmentation with corner attention and high-dimensional perceptual loss. Biomedical Signal Processing and Control, 2022, vol. 73, p. 103438.

[7] HUANG, Zheng, ZHAO, Yiwen, LIU, Yunhui, et al. GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation. Biomedical Signal Processing and Control, 2021, vol. 70, p. 102958.

[8] HUANG, Ling, RUAN, Su, et DENOEUX, Thierry. Belief function-based semi-supervised learning for brain tumor segmentation. In : 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021. p. 160-164.

[9] GONCALVES, Juliano P., PINTO, Francisco AC, QUEIROZ, Daniel M., et al. Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests. Biosystems Engineering, 2021, vol. 210, p. 129-142.

[10] Asgari Taghanaki, S., Abhishek, K., Cohen, J. P., Cohen-Adad, J., Hamarneh, G. (2021). Deep semantic segmentation of natural and medical images: a review. Artificial Intelligence Review, 54(1), 137-178.

[11] HANSEN, Stine, GAUTAM, Srishti, JENSSEN, Robert, et al. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Medical Image Analysis, 2022, vol. 78, p. 102385.

[12] OUYANG, Cheng, BIFFI, Carlo, CHEN, Chen, et al. Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In : European Conference on Computer Vision. Springer, Cham, 2020. p. 762-780.

[13] Dong, Nanqing, and Eric P. Xing. "Few-shot semantic segmentation with prototype learning." BMVC. Vol. 3. No. 4. 2018.

[14] Ravi, Sachin, and Hugo Larochelle. "Optimization as a model for few-shot learning." (2016).

[15] WANG, Yong, WU, Xiao-Ming, LI, Qimai, et al. Large margin few-shot learning. arXiv preprint arXiv:1807.02872, 2018.

[16] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 1199–1208

[17] JAMAL, Muhammad Abdullah et QI, Guo-Jun. Task agnostic meta-learning for few-shot learning. In : Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 11719-11727.

[18] CHEN, Yinbo, LIU, Zhuang, XU, Huijuan, et al. Meta-baseline: Exploring simple meta-learning for few-shot learning. In : Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 9062-9071.

[19] PAMBALA, Ayyappa Kumar, DUTTA, Titir, et BISWAS, Soma. SML: Semantic meta-learning for few-shot semantic segmentation. Pattern Recognition Letters, 2021, vol. 147, p. 93-99.

[20] TIAN, Pinzhuo, WU, Zhangkai, QI, Lei, et al. Differentiable meta-learning model for few-shot semantic segmentation. In : Proceedings of the AAAI Conference on Artificial Intelligence. 2020. p. 12087-12094.

[21] MONDAL, Arnab Kumar, DOLZ, Jose, et DESROSIERS, Christian. Few-shot 3d multi-modal medical image segmentation using generative adversarial learning. arXiv preprint arXiv:1810.12241, 2018.

[22] OUYANG, Cheng, BIFFI, Carlo, CHEN, Chen, et al. Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In : European Conference on Computer Vision. Springer, Cham, 2020. p. 762-780.

[23] ROY, Abhijit Guha, SIDDIQUI, Shayan, PÖLSTERL, Sebastian, et al. 'Squeeze excite'guided few-shot segmentation of volumetric images. Medical image analysis, 2020, vol. 59, p. 101587.

[24] ZHAO, Amy, BALAKRISHNAN, Guha, DURAND, Fredo, et al. Data augmentation using learned transformations for one-shot medical image segmentation. In : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 8543-8553

[25] FEYJIE, Abdur R., AZAD, Reza, PEDERSOLI, Marco, et al. Semi-supervised few-shot learning for medical image segmentation. arXiv preprint arXiv:2003.08462, 2020.

[26] KHADGA, Rabindra, JHA, Debesh, ALI, Sharib, et al. Few-shot segmentation of medical images based on meta-learning with implicit gradients. arXiv preprint arXiv:2106.03223, 2021.

[27] ZHANG, Xiaolin, WEI, Yunchao, YANG, Yi, et al. Sg-one: Similarity guidance network for one-shot semantic segmentation. IEEE transactions on cybernetics, 2020, vol. 50, no 9, p. 3855-3865.

[28] YEUNG, Michael, SALA, Evis, SCHÖNLIEB, Carola-Bibiane, et al. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics, 2022, vol. 95, p. 102026.

[29] MICCAI Brain Tumor Segmentation (BraTS) Challenge 2021 http://braintumorsegmentation.org/.[accessed 07 Julie 2022]

[30] SUN, Liyan, LI, Chenxin, DING, Xinghao, et al. Few-shot medical image segmentation using a global correlation network with discriminative embedding. Computers in biology and medicine, 2022, vol. 140, p. 105067.