Pattern Recognition 88 (2019) 90–100

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Dual-force convolutional neural networks for accurate brain tumor segmentation

Shengcong Chen [a], Changxing Ding [a,*], Minfeng Liu [b]

[a] School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong 510000, China
[b] Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China

ARTICLE INFO

Article history:
Received 28 February 2018
Revised 13 September 2018
Accepted 10 November 2018
Available online 14 November 2018

Keywords:
Brain tumor segmentation
Dual-force network
Convolutional neural network
Label distribution
Post-processing

ABSTRACT

Brain tumor segmentation from Magnetic Resonance Imaging scans is vital for both the diagnosis and treatment of brain cancers. It is widely accepted that accurate segmentation depends on multi-level information. However, exiting deep architectures for brain tumor segmentation fail to explicitly encourage the models to learn high-quality hierarchical features. In this paper, we propose a series of approaches to enhance the quality of the learnt hierarchical features. Our contributions incorporate four aspects. First, we extend the popular DeepMedic model to Multi-Level DeepMedic to make use of multi-level information for more accurate segmentation. Second, we propose a novel dual-force training scheme to promote the quality of multi-level features learnt from deep models. It is a general training scheme and can be applied to many exiting architectures, e.g., DeepMedic and U-Net. Third, we design a label distribution-based loss function as an auxiliary classifier to encourage the high-level layers of deep models to learn more abstract information. Finally, we propose a novel Multi-Layer Perceptron-based post-processing approach to refine the prediction results of deep models. Extensive experiments are conducted on two most recent brain tumor segmentation datasets, i.e., BRATS 2017 and BRATS 2015 datasets. Results on the two databases indicate that the proposed approaches consistently promote the segmentation performance of the two popular deep models.

1. Introduction

Gliomas refer to tumors developed from glial cells, and they are the most common primary tumors in adult human brains [1,2]. They can be roughly classified into two groups according to their grade: High-Grade Gliomas (HGG) and Low-Grade Gliomas (LGG). Gliomas are huge threats to human's health due to both the high incidence and lethality rates. For example, the Median Survival Time (MST) of HGG patients is less than 2 years [2,3] and the MST of patients with glioblastoma, the most aggressive tumors, is only 4.9 months [4]. Nowadays, Magnetic Resonance Imaging (MRI) is widely used for diagnosis and treatment of brain tumors. MRI can provide doctors with multi-modality 3D scans of the brain. Segmentation of brain tumors and surrounding abnormal tissues based on MRI images can provide doctors with direct understanding of tumors and assistance for analysis and treatment. Therefore, brain tumor segmentation is regarded as an important step in MRI analysis.

MRIs are multi-modality and large-volume 3D scans; thus manual segmentation is time-consuming and tedious. Additionally, manual segmentation is likely to be affected by raters' personal experience; thus it is error-prone. Therefore, fully automatic and accurate systems for brain tumor segmentation are highly desired in practice. However, designing an accurate brain tumor segmentation system still remains a challenging problem. This is for a number of reasons. First, the shape and internal structures of tumors are varied and complicated. Second, surrounding normal tissues are also in a wide variety of appearance due to variable locations of tumors and the so-called tumor mass effect [5]. Third, the boundaries between normal tissues and tumor tissues tend to be obscure and therefore they are difficult to differentiate.

Various convolutional neural networks (CNNs) have been designed for brain tumor segmentation and other medical image segmentation tasks. These networks have quickly developed from single-label prediction schemes to dense prediction schemes. State-of-the-art dense prediction networks include DeepMedic [6] and U-Net [7], etc. Kamnitsas et al. [6] proposed a plain 3D network, named DeepMedic, which utilizes feature maps extracted from the final convolutional layer to predict the labels of voxels within the central regions of input patches efficiently.

* Corresponding author.
E-mail addresses: c.shengcong@mail.scut.edu.cn (S. Chen), chxding@scut.edu.cn (C. Ding), matthew1@smu.edu.cn (M. Liu).

https://doi.org/10.1016/j.patcog.2018.11.009
0031-3203/

However, the final convolutional layer in CNN usually focuses on extracting high-level semantic information. The lack of low- and middle-level information restricts the performance of DeepMedic. Hariharan et al. [8] proposed the popular Hypercolumn architecture that extracts multi-level information from a plain CNN model for segmentation. This architecture has been widely adopted in various computer vision tasks [9,10]. However, the quality of the extracted multi-level features may not be optimal as they are employed jointly and there is only one single loss function for penalty; therefore the explicit guarantee of feature hierarchy is required [11]. Besides, U-Net and its variants [7,12] are an important series of networks in medical image segmentation. They consist of an encoding network with down-sampling operations and a decoding network with up-sampling or deconvolutional operations. U-Nets introduce feature maps in the encoding network to the decoding network through successive concatenation operations in order to introduce multi-level information gradually and finally promote the prediction accuracy. Similar to the Hypercolumn network, the quality of the multi-level features may not be optimal as there is only one loss function for penalty.

To solve the above problems, we propose a novel dual-force training strategy to explicitly encourage CNNs to learn high-quality hierarchical features. This training strategy is general and can be applied to many popular networks, e.g., DeepMedic and U-Net. CNNs adopting this training strategy are named Dual-Force Networks (DFNs) in this paper. Our contributions include four aspects:

(1) We extend DeepMedic to Multi-Level DeepMedic (MLDeepMedic) to utilize multi-level information jointly for more accurate segmentation.
(2) We propose the dual-force training strategy and apply it to both MLDeepMedic and U-Net. This training strategy is realized by attaching an auxiliary classifier to both MLDeepMedic and U-Net, and encourages their high-level convolutional layers to learn more abstract semantic information.
(3) We propose a label distribution-based loss function for the auxiliary classifier. Compared with existing popular loss functions for brain tumor segmentation, e.g., softmax loss function, the proposed loss function can describe high-level semantic information more accurately.
(4) We propose an Multi-Layer Perceptron (MLP) based method for post-processing, which can refine the prediction results of CNNs and further promote the segmentation performance. Compared with existing post-processing methods, e.g., Conditional Random Field (CRF), the proposed method is light and easy to use in practice.

We conduct extensive experiments on two most recent brain tumor segmentation datasets, i.e., BRATS 2017 and BRATS 2015. Experimentation results show that the proposed dual-force training strategy consistently improves the Dice scores of the segmentation results by both MLDeepMedic and U-Net. Besides, the proposed MLP-based post-processing method can further improve the segmentation performance with very limited computational cost. Moreover, to further verify the effectiveness of the proposed methods on ensemble systems, we also apply the proposed methods to Model-Cascade Networks (MC-Net for short) [13].

The remainder of this paper is organized as follows: Section 2 reviews recent related works in brain tumor segmentation. Section 3 first briefly introduces the baseline networks, i.e., DeepMedic and U-Net. Then we propose the dual-force training strategy, including network architecture and the label distribution-based loss function. Section 4 presents and analyzes the experimental results, leading to conclusions in Section 5.

2. Related works

We review the literature in two parts: (1) CNN architectures for brain tumor segmentation, and (2) post-processing methods that refine the prediction results of CNNs.

2.1. CNN architectures

Recent automatic and semi-automatic methods for medical image segmentation can be roughly classified into generative and discriminative methods. Generative methods estimate the joint probability distribution of labels and features of voxels [14–19], and most of them are designed based on atlas [15,17]. The others are discriminative methods, which rely less on prior knowledge and make predictions based on the intensity of voxels. Classic machine learning algorithms make use of hand-crafted features, and then the extracted features are fed into classification algorithms such as Support Vector Machines (SVM) [20], Random Forest [21,22], Fuzzy C-Means algorithm [23], Level Set [24] or Particle Swarm Optimization [25]. Among the wide range of classic methods, Random Forest is one of the most commonly adopted methods because of its efficiency and its capability of handling high dimensional features. For example, Pinto et al. [22] proposed a hierarchical Extremely-Randomized-Forest-based method for glioma segmentation and achieved competitive performance on BRATS 2013 Leaderboard dataset. As the rise of deep learning [26–34], recent works applied CNNs of various structures for medical image segmentation, and achieved state-of-the-art performance. In the following, we only review CNN-based methods for the brain tumor segmentation task.

According to the dimension of input data for CNNs, existing approaches can be classified into 2D [12,35–37], 2.5D [38], and 3D networks [6,7,39–43]. Most early approaches [12,35,36] are based on 2D networks as they are advantageous in efficiency. However, they fail to make full use of 3D contextual information from 3D MRI scans. To relieve this problem, Zhu et al., utilize recursive neural networks [44] to treat different slices as a data sequence. In addition, 2.5D networks are proposed [38], which take a set of 2D slices extracted from planes perpendicular to the axial, coronal, and sagittal axis as input together. However, most 3D contextual information is still ignored. Therefore, to take full advantage of 3D information in MRI scans, 3D networks with 3D convolutional filters are proposed. 3D networks can process 3D MRI data directly. However, the produced 3D feature maps require dramatically more GPU memory and computational cost, which limits the depth of 3D networks and usually requires elaborate design. Despite their limitations, 3D networks are becoming increasingly popular due to their outstanding performance in medical image segmentation.

According to the size of predictions, CNNs for brain tumor segmentation can be roughly grouped into single-label prediction and dense prediction architectures. Single-label prediction architectures take cropped patches as input and only predict the label of the central voxel in the patch [35–37,45,46]. For instance, Pereira et al. [36] proposed a 2D convolutional neural network that makes single-label prediction and achieves the best performance in the BRATS 2015 challenge. In order to introduce more contextual information for the central voxel, Havaei et al. [37] proposed a two-pathway architecture and integrated features of two-scales. In addition, a two-phase training scheme was proposed in [37] to deal with the class-imbalance problem. Since only the label of the central voxel is predicted each time, the single-label prediction networks are very slow during the inference stage.

Compared with single-label prediction networks, dense prediction networks [6,7,22,39,43,47] are more efficient, as they predict

the labels of voxels within the input patch simultaneously. Fully Convolutional Network (FCN) [47] is a typical instance of dense prediction networks. FCN replaces the fully connected layers in common CNNs with convolutional layers to generate full-resolution prediction maps. In order to better utilize multi-level features, U-Net [12] gradually introduces feature maps by concatenation from its encoding layers to its decoding layers of the same resolution. Due to the excellent performance, U-Nets are also widely adopted in other computer vision tasks. For example, a U-Net-style architecture is adopted for feature extraction in [48]. To fully utilize 3D contextual information, 3D variants of FCN and U-Net are also proposed for medical image segmentation [7,49]. Another important dense prediction network is named DeepMedic [6] that predicts the labels of voxels lying in the central region of the input patch. In DeepMedic, there is no pooling operations and the reduction of feature map size is realized by cancelling padding operations in convolutional layers.

Moreover, recent works propose various strategies to improve the performance of CNN models for medical image segmentation. For example, different from the above networks that are constructed by simply stacking convolutional layers, Chen et al. [39] proposed VoxResNet that is inspired by the idea of residual learning [31]. Besides, most existing methods stack multi-modality MRI data together as input, while [42] argued that the correlations between different modalities should be explicitly considered before concatenation, and proposed a novel cross-modality network to aggregate information from different modalities. Additionally, to fill the gap between training data and testing data, Kamnitsas et al. [50] built an adversarial network for unsupervised domain adaptation based on DeepMedic [6].

In this paper, our proposed dual-force training strategy is mainly based on DeepMedic [6] and U-Net [7], which are two most popular baselines for brain tumor segmentation. Networks adopting the dual-force training strategy are named Dual-Force Networks (DFNs). Different from the existing works, e.g., Deeply Supervised Network (DSN) [51], GoogleNet [29], and VoxResNet [39] that introduce auxiliary classifiers to solve the vanishing gradient problem, our proposed DFN apply an auxiliary loss function to promote the quality of hierarchical features. Besides, instead of using the common softmax loss functions as Lee and co-workers [29,39,51], we introduce the label distribution-based loss function for the auxiliary loss. Label distribution loss is usually adopted in age estimation task [52] to describe the correlation between nearby ages. In this paper, it is employed to describe high-level semantic information more accurately.

2.2. Post-processing methods

In addition to network architectures, various post-processing methods have also been proposed to refine the prediction results of CNNs. For example, Kamnitsas et al. [6] adopted 3D-CRF for post-processing, which refines segmentation results through iteratively minimizing the Gibbs energy of each voxel. Besides, Havaei et al. [37] proposed to remove the abnormal predictions in regions close to the skull according to the intensities of voxels, and the volume of the tumor area. Moreover, Zhao et al. [53] proposed a more complicated post-processing pipeline including empirically determined rules, which are also based on the voxel intensity, volume of predicted area, etc. One shortage of this method is that the rules rely on prior knowledge summarized by humans and they may not be readily transformed to other segmentation tasks. Our proposed post-processing method is more related to Zhao et al. [53]. Compared with the original method in [53], we propose an MLP-based method to automatically learn the set of rules. Therefore, it can be easily applied to many segmentation tasks.

3. Dual-force networks

In this section, we first briefly review two popular baselines for brain tumor segmentation. Then, we introduce our contributions in detail. First, we extend DeepMedic model by utilizing multi-level features for more accurate segmentation. Second, we introduce the dual-force training strategy to help CNNs to learn high quality multi-level features. This strategy can be applied to many existing networks, e.g., DeepMedic and U-Net. Third, we propose a light, simple but effective post-processing method to refine the prediction results of CNNs.

3.1. Brief review of the baselines

The proposed dual-force training strategy is based on DeepMedic and U-Net. Therefore, we first briefly review their architectures in the following. We also illustrate their structure in Fig. 1. Each 'Conv' block in Fig. 1 refers to the combination of one convolutional layer, one batch normalization layer, and one ReLU layer. Important parameters for the two models adopted in this paper are tabulated in Tables 1 and 2, respectively.

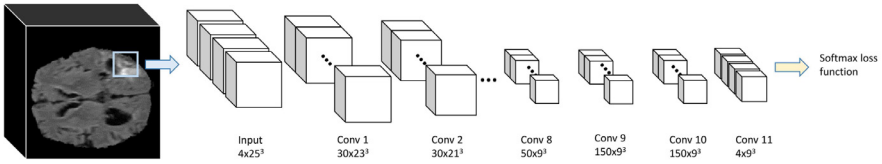
DeepMedic [6] achieved the best performance in BRATS 2016 challenge. It consists of two pathways, separately designed for extracting features of two scales. Each of the two pathways includes eight $3 \times 3 \times 3$ convolutional layers and there is no pooling layer. The reduction of feature map size is realized through omitting the marginal voxels in feature maps during convolution. After concatenating feature maps extracted from the two pathways, three $1 \times 1 \times 1$ convolutional layers are applied to fuse their information and make final predictions. For simplicity, only the full-resolution pathway of DeepMedic is shown in Fig. 1(a). The two-pathway scheme is illustrated in Fig. 2.

U-Net [7,12] is another popular architecture for brain tumor segmentation. Model structure of U-net is illustrated in Fig. 1(b). Inspired by FCN [47], U-Net efficiently predicts the full-resolution label maps for the input patch in one pass. U-Net comprises an encoding network and a decoding network. In the encoding network, stacked convolutional layers are followed by pooling layers to compress the size of feature maps and extract high-level semantic information. The decoding network incorporates a set of deconvolutional layers for resolution recovery. Moreover, feature maps generated by deconvolutional layers are concatenated with those produced by convolutional layers of the encoding network with the same resolution. The combination of feature maps enables U-Net to gradually fuse multi-level information for accurate brain tumor segmentation.

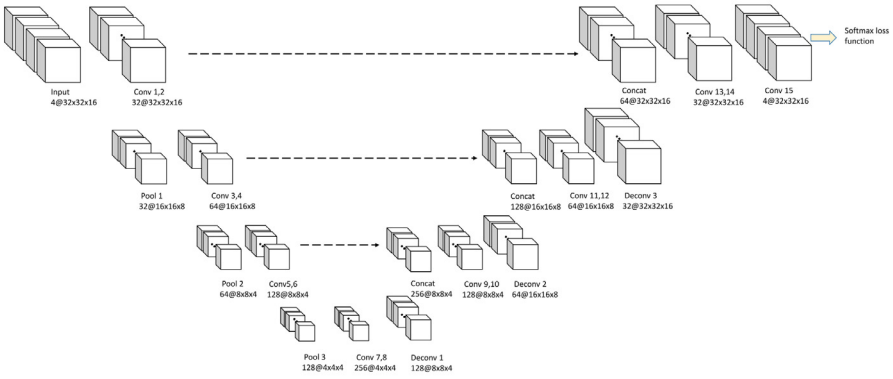
3.2. Multi-level deepmedic

It is widely accepted that multi-level information is vital for segmentation tasks. In particular, they are helpful to discriminate semantically obscure voxels that are usually lies in the boundary area between tumor and normal tissues. However, the DeepMedic model only utilizes high-level features extracted from the last convolutional layer for prediction, ignoring low- and middle-level features. In this subsection, we extend the DeepMedic model to include multi-level information for more accurate segmentation, and name the new model as Multi-Level DeepMedic (MLDeepMedic). Considering that the low-resolution pathway of DeepMedic is designed for extracting contextual information in larger scales, we remain the structure of the low-resolution pathway and only introduce multi-level information to the full-resolution pathway of DeepMedic.

As illustrated in Fig. 3, we equip DeepMedic with several crop layers to extract multi-level features of the same size, i.e., $9 \times 9 \times 9$, from five selected layers of DeepMedic. Each crop layer crops a



(a) DeepMedic



(b) U-Net

Fig. 1. Model architectures for DeepMedic and U-Net. (a) DeepMedic. (b) U-Net. For simplicity, only the full-resolution pathway of DeepMedic is shown in (a). Each Conv block in the figure refers to the combination of one convolutional layer, one batch normalization layer and one ReLU layer.

Table 1
Important model parameters of DeepMedic and MLDeepMedic adopted in this paper. Filter number in the last convolutional layer depends on the number of normal and tumor classes, i.e., 4 (BRATS 2017) or 5 (BRATS 2015) in this paper.

Layer	Type	Kernel size	Filter number	Padding	Stride	Pathway
conv1,2	convolution+BN+relu	$3\times3\times3$	30	0	1	L,F
conv3,4,5,6	convolution+BN+relu	$3\times3\times3$	40	0	1	L,F
conv7,8	convolution+BN+relu	$3\times3\times3$	50	0	1	L,F
deconv	deconvolution+BN+relu	$2\times2\times3$	50	0	2	L
concat	concat	–	–	–	–	–
conv9	convolution+BN+relu	$1\times1\times1$	150	0	1	–
dropout	dropout (P = 0.5)	–	–	–	–	–
conv10	convolution+BN+relu	$1\times1\times1$	150	0	1	–
dropout	dropout (P = 0.5)	–	–	–	–	–
conv11	convolution	$1\times1\times1$	4(5)	0	1	–
softmax	softmax	–	–	–	–	

Table 2
Important model parameters of 3D U-Net adopted in this paper. Filter number in the last convolutional layer depends on the number of normal and tumor classes, i.e., 4 (BRATS 2017) or 5 (BRATS 2015) in this paper.

Layer	Type	Kernel size	Filter number	Padding	Stride
conv1,2	convolution+BN+relu	$3\times3\times3$	32	1	$1\times1\times1$
pool1	max pooling	$2\times2\times2$	–	0	$2\times2\times2$
conv3,4	convolution+BN+relu	$3\times3\times3$	64	1	$1\times1\times1$
pool2	max pooling	$2\times2\times2$	–	0	$2\times2\times2$
conv5,6	convolution+BN+relu	$3\times3\times3$	128	1	$1\times1\times1$
pool3	max pooling	$2\times2\times1$	–	0	$2\times2\times1$
conv7,8	convolution+BN+relu	$3\times3\times3$	256	1	$1\times1\times1$
deconv1	deconvolution+BN+relu	$2\times2\times2$	128	0	$2\times2\times1$
conv9,10	convolution+BN+relu	$3\times3\times3$	128	1	$1\times1\times1$
deconv2	deconvolution+BN+relu	$2\times2\times2$	64	0	$2\times2\times1$
conv11,12	convolution+BN+relu	$3\times3\times3$	64	1	$1\times1\times1$
deconv3	deconvolution+BN+relu	$2\times2\times2$	32	0	$2\times2\times1$
conv13,14	convolution+BN+relu	$3\times3\times3$	32	1	$1\times1\times1$
conv15	convolution	$1\times1\times1$	4(5)	0	$1\times1\times1$
softmax	softmax	–	–	–	–

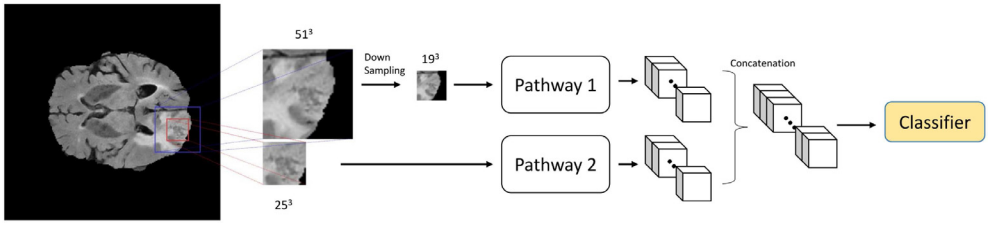


Fig. 2. Model architecture of two-pathway DeepMedic [6]. Feature maps generated from the low-resolution pathway are upsampled to the same size as that in the full-resolution pathway. In this paper, we replace the up-pooling layer adopted in [6] with a deconvolutional layer.

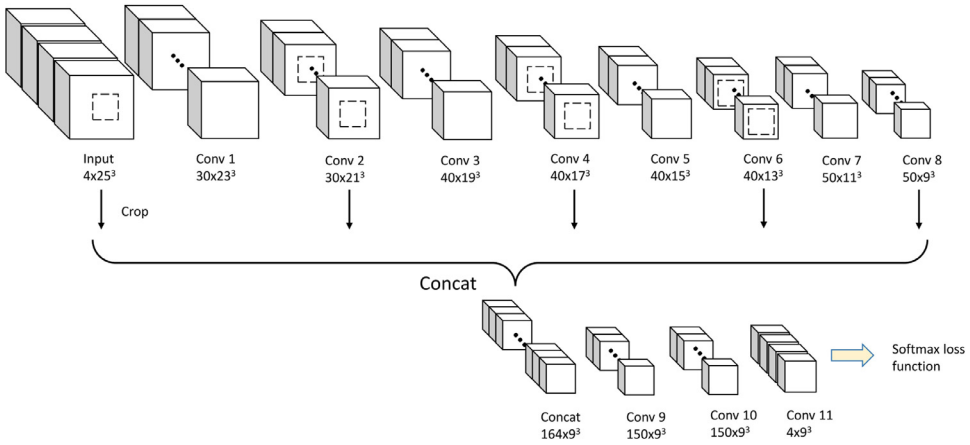


Fig. 3. Model architecture of MLDeepMedic. Only the full-resolution pathway of DeepMedic is shown in this figure. Feature map patches are cropped from five selected layers and then concatenated to compose multi-level features.

3D patch from the center region of the feature map. Since there is no pooling operation in DeepMedic, elements in the cropped 3D patches maintain spatial correspondence. Consequently, we obtain multi-level information for the voxels within the cropped 3D patches. To fuse these multi-level information, the cropped feature map patches are first concatenated and then fed into three successive $1 \times 1 \times 1$ convolutional layers for final prediction.

3.3. Dual-force networks

Both MLDeepMedic and U-Net employ multi-level features for accurate segmentation. However, there is only one loss function to train MLDeepMedic and U-Net, respectively; therefore, they may not learn high-quality multi-level features without explicit guarantee in loss functions. In this subsection, we propose a dual-force (DF) training strategy that attach an auxiliary loss function to the network to explicitly encourage the network to learn high-quality multi-level features. Networks using the DF training strategy are named Dual-Forced networks (DFN) in this paper. In the following, we take MLDeepMedic and U-Net as examples to describe the idea of DFN. However, it is worth noting that the DF training strategy is general and can be applied to more advanced architectures that utilize multi-level features for dense prediction, e.g., FusionNet [13].

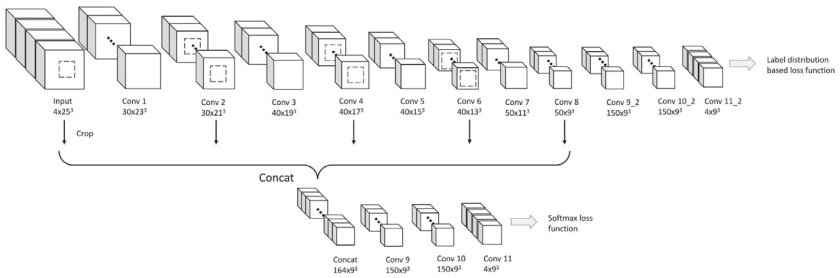
As illustrated in Fig. 4, we attach a complementary loss function to the original networks to describe the high-level semantic information contained in the input patch. The two loss functions work together to explicitly encourage MLDeepMedic and U-Net to learn high-quality multi-level features. The two networks using this training strategy are denoted as DF-MLDeepMedic and DF-U-Net, respectively. In testing, the auxiliary classifiers can be omitted.

Model structures of both DF-MLDeepMedic and DF-U-Net are presented in Fig. 4. Similar to Fig. 1, only the full-resolution pathway of DF-MLDeepMedic is illustrated. For DF-MLDeepMedic, two auxiliary classifiers are introduced for the two pathways, respectively, as illustrated in Fig. 4(a). One of them is attached to the 'Conv8' layer in the full-resolution pathway and the other is attached to the deconvolutional layer in the low-resolution pathway. For DF-U-Net, the new loss function is attached after the last convolutional layer of the encoding network, as illustrated in Fig. 4(b). This is because feature maps from this layer have the minimum spatial size and therefore are expected to contain the most abstract semantic information.

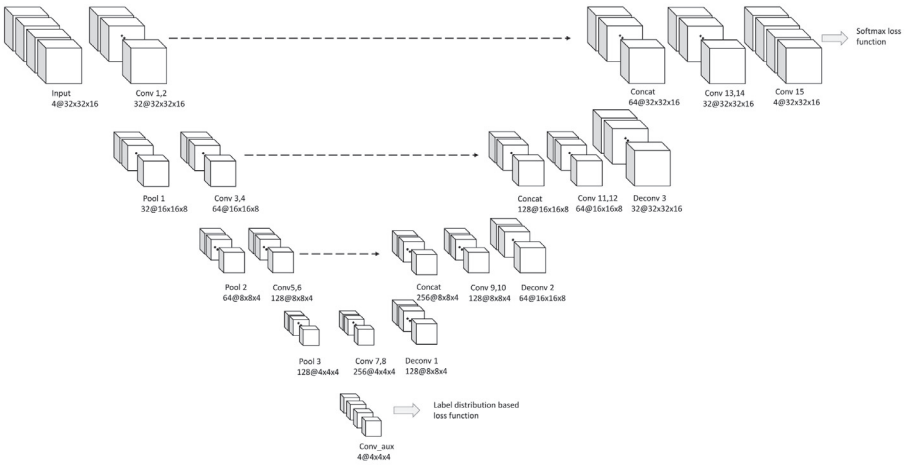
3.4. Label distribution-based loss function

The inputs for the original loss function in DeepMedic and U-Net are voxel-wise classification probability; therefore, the popular softmax loss function is adopted. However, elements of the inputs for the newly introduced loss function in DFN stand for high-level semantic information, which cannot be described by discrete labels. Therefore, we propose label distribution-based loss function and present different solutions for DFN variants in the following. Compared with discrete labels, label distribution vectors can describe high-level semantic information more accurately.

In MLDeepMedic, spatial correspondence is maintained as there is no pooling layer, which means elements in the 'Conv11' layer keep the spatial relationship as the voxels in the input patch. Therefore, we construct the ground-truth label distribution vectors using the 3D Gaussian model to highlight the contribution of each focal voxel in the 'Conv11' layer. In detail, the ground-truth label map is first transformed into a series of hard-coding binary probability maps, denoted as $\mathbf{B}$, whose channel number is dependent on the number of classes. Then, each of the binary hard-coding maps is convolved with one 3D Gaussian filter, which is



(a) DF-MLDeepMedic



(b) DF-U-Net

Fig. 4. Model structure of DFNs. (a) DF-MLDeepMedic. For simplicity, only the full-resolution pathway is presented. The auxiliary loss function is attached to the 'Conv8' layer of the full-resolution pathway. For the two-pathway DF-MLDeepMedic, there is another auxiliary loss function attached to the deconvolutional layer of the low-resolution pathway. (b) DF-U-Net. A new convolutional classification layer and the auxiliary loss function is attached to the 'Conv8' layer.

formulated as

$$D_k(x,y,z) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^3} \sum_{(x_l,y_l,z_l)\in \mathbf{S}_{x,y,z}} B_k(x_l,y_l,z_l) e^{-\frac{(x_l-x)^2+(y_l-y)^2+(z_l-z)^2}{2\sigma^2}}, \tag{1}$$

where $\mathbf{S}_{x,y,z}$ stands for the neighborhood of the target voxel (x, y, z). Size of the neighborhood is $17 \times 17 \times 17$ in our experiments. k indicates the kth class and σ is the standard deviation of the Gaussian kernel. The value of σ can control the ratio of contextual information in the label distribution vector. Obviously, hard-coding labels can be also considered as a special form of the label distribution vector which has zero stand deviation.

For U-Net, voxels in different convolutional layers do not maintain spatial correspondence due to the pooling operations in the encoding network; therefore, we adopt average pooling to construct label distribution vectors. Specifically, the ground-truth label map is divided to $4 \times 4 \times 4$ non-overlapped regions with size of $8 \times 8 \times 4$. Similar to DF-MLDeepMedic, the ground-truth label map for each region is first transformed into a series of hard-coding binary probability maps, denoted as $\mathbf{B}$, whose channel number is dependent on the number of classes. The binary labels in the kth channel of $\mathbf{B}$ are fused by average pooling to construct label distribution maps:

$$D_k(x,y,z) = \frac{1}{n} \sum_{(x_l,y_l,z_l)\in \mathbf{S}_{x,y,z}} B_k(x_l,y_l,z_l), \tag{2}$$

where n denotes the number of voxels involved in the region $\mathbf{S}_{x,y,z}$ that is, $8 \times 8 \times 4$ in this paper.

We take KL divergence as the loss function to measure the disparity between the ground-truth and predicted label distribution vectors:

$$L = -\frac{1}{n} \sum_i \sum_k d_k^i log\left(\hat{p_k^i}\right), \tag{3}$$

where n denotes the number of label distribution vectors in a batch. For the ith training sample, d_k^i is the probability value in the ground truth distribution of class k, and $\hat{p_k^i}$ is the predicted probability value. It is worth noting that the label distribution-based loss function aims to encourage the network to learn high-quality semantic information in the training stage. In the testing stage, this loss function can be omitted.

3.5. Post-processing

Post-processing is an indispensable step for brain tumor segmentation as it refines the prediction results of CNNs and promotes segmentation performance. Recently, Zhao et al. [53] proposed a high-performance post-processing pipeline including a set of empirically summarized rules and manually determined thresholds, which are based on the probability values predicted by CNNs, the voxel intensity, and the volume of predicted tumors, etc. Since this approach is highly handcrafted, it may not be readily transfered to other related segmentation tasks. In this subsection, we

propose a learning-based approach to automatically learn the post-processing rules from data; therefore, it is easier to be utilized for the other segmentation task as well.

The proposed post-processing method is based on MLP. The input of MLP is a feature vector for each voxel in the predicted tumor area by CNNs, denoted as $\mathbf{f} \in \mathbb{R}^{257\times 1}$ in BRATS 2015 or $\mathbf{f} \in \mathbb{R}^{229\times 1}$ in BRATS 2017. Similar to Zhao et al. [53], the feature vector includes three groups of information, i.e., the probability values predicted by CNNs for voxels, the voxel intensity, and the volume of predicted tumors:

$$\mathbf{f} = [\mathbf{p}; \mathbf{i}; v], \tag{4}$$

where $\mathbf{p} \in \mathbb{R}^{140\times 1}$ in BRATS 2015 or $\mathbf{p} \in \mathbb{R}^{112\times 1}$ in BRATS 2017 denotes the probability values predicted by CNNs for voxels. It consists of the predicted probability values of all classes for the target voxel and its $3\times 3\times 3$ neighbors, as well as the label distribution vector of the target voxel constructed via (1) with σ set to be 5. $\mathbf{i} \in \mathbb{R}^{116\times 1}$ denotes the intensity feature vector, composed of the intensity values of the target voxel and its $3\times 3\times 3$ neighbors, as well as the mean value and standard deviation of the connected tumor area where the target voxel locates in. Finally, v is a scaler that stands for the ratio between the volume of the tumor region that the target voxel locates in and the volume of the maximum connected tumor area in the MRI image. **p, i** and v are concatenated together, constructing the input feature vector of MLP.

The employed MLP model is composed of three fully connected layers, with filter numbers of 128, 128, and c respectively, where c denotes the number of classes. In testing, a feature vector $\mathbf{f}$ is constructed for each voxel in the predicted tumor area by CNNs. The feature vector is fed into MLP and obtains the final prediction result for the voxel.

4. Experiments

4.1. Datasets

We evaluate the proposed methods on two most recent publicly available brain tumor datasets, i.e., BRATS 2017 and BRATS 2015. The BRATS 2017 database includes 285 3D MRI scans, which are composed of 210 HGG scans and 75 LGG scans. The BRATS 2015 database includes 220 HGG scans and 54 LGG scans. For each scan, there are 4 available modalities, i.e., T1, T1C, T2, and Flair. Image Size for each modality is $240\times 240\times 155$. In the BRATS 2015 database, the tumor tissues are divided into four classes, i.e., necrotic, edema, non-enhancing tumor, and enhancing tumor. In the BRATS 2017 database, the necrotic and non-enhancing tumors are grouped into one category; therefore, there are only three tumor classes in this database. To quickly justify the effectiveness of the proposed methods, we randomly select 30 scans from all the 285 cases in the BRATS 2017 database to construct a local validation subset. The remaining 255 scans are utilized as the training set. At the end of this section, we compare the performance of the proposed approaches with state-of-the-art methods on the online validation sets of the two databases.

4.2. Implementation details

For pre-processing, intensities of voxels inside the brain are normalized to have zero mean and unit variance for each modality image. To compensate for the class imbalance problem, we adopt the same patch sampling strategy as proposed in [6]. Training patches are uniformly sampled (50% vs 50%) from the normalized images according to the label of central voxels for tumor and normal tissues, respectively.

The open-source deep learning package C3D [54–56] is utilized to train the deep models. We adopt the Stochastic Gradient Descent (SGD) scheme to optimize all networks. Batch size is set to be 64 and the initial learning rate is set as 0.001, decreasing by one half after each 5 epochs. The maximum number of iterations is 40 epochs. For fair comparison, we test all saved snapshots and report the best performance for each model on the local validation subset.

Besides, the size of input patches for DeepMedic and U-Net is $25\times 25\times 25$ and $32\times 32\times 16$, respectively. Due to the lack of contextual information, the labels predicted by U-Net for the marginal voxels in the patch may not be reliable. Therefore, we only keep the predicted labels of voxels within the central $20\times 20\times 10$ region in each patch in testing stage.

4.3. Evaluation criteria

According to the protocols in the two BRATS datasets, different tumor classes are grouped into three tumor regions, i.e., whole tumor, core tumor and enhancing tumor. The whole tumor region comprises all tumor classes. The core tumor region comprises all tumor classes but edema. The enhancing tumor region comprises the enhancing tumor class only. To quantitatively evaluate the segmentation results, three indices are usually adopted, i.e., Dice Score, Positive Predictive Value (PPV), and Sensitivity. Each of them is formulated as bellow:

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{5}$$

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{7}$$

where TP represents the number of true positive voxels, FP represents the number of false positive voxels, and FN represents the number of false negative voxels. It is worth noting that Dice score can be considered as an overall evaluation criterion of PPV and Sensitivity. Therefore, we will mainly compare the Dice scores of different methods in the following.

4.4. Performance comparison between DeepMedic and MLDeepMedic

We first compare the performance of DeepMedic and the proposed MLDeepMedic. Experiments are conducted on both the single-pathway and two-pathway variants of DeepMedic. Experimentation results are tabulated in Table 3.[1] It is shown that the two-pathway architecture indeed improves the performance of DeepMedic, which is consistent with the conclusion in [6]. It is also shown that the two-pathway architecture also improves the performance of MLDeepMedic with considerable margin.

It is clear that MLDeepMedic consistently outperforms DeepMedic, indicating that multi-level feature is helpful for the segmentation task. However, the improvement is not significant and there is even subtle decline in Dice score for the enhancing tumor. We argue that this is because the quality of the learnt multi-level features is not optimal. In MLDeepMedic, there is only one loss function for classification and the quality of multi-level features is not explicit guaranteed.

4.5. Effectiveness of DFN and MLP-based post-processing method

The proposed dual-force training scheme is evaluated on both MLDeepMedic and U-Net. Since two-pathway MLDeepMedic outperforms its single-pathway counterpart, we consistently employ the two-pathway architecture in the following. Experimentation results on MLDeepMedic are tabulated in Table 4. DF-

[1] Dice score is the overall evaluation criterion. For clarity, only the best results of Dice score are highlighted in all tables.

Table 3
Performance comparison between DeepMedic and MLDeepMedic on the local validation subset of BRATS 2017 (%).

Method	Dice			PPV			Sensitivity		
	Com.	Core	Enh.	Com.	Core	Enh.	Com.	Core	Enh.
Single-Path DeepMedic	78.26	69.76	68.82	73.00	68.54	68.94	88.42	79.04	75.58
Single-Path MLDeepMedic	79.73	71.59	68.14	76.07	72.14	68.58	87.30	0.7854	75.94
Double-Path DeepMedic	**83.34**	72.24	69.34	82.50	73.33	70.38	85.65	75.89	75.49
Double-Path MLDeepMedic	83.25	**73.25**	**69.96**	82.53	77.23	69.24	85.72	74.67	76.96

Table 4
Performance comparison between DeepMedic, MLDeepMedic, and DF-MLDeepMedic on the local validation subet of BRATS 2017 (%).

Method	Dice			PPV			Sensitivity		
	Com.	Core	Enh.	Com.	Core	Enh.	Com.	Core	Enh.
DeepMedic	83.34	72.24	69.34	82.50	73.33	70.38	85.65	75.89	75.49
MLDeepMedic	83.25	73.25	69.96	82.53	77.23	69.24	85.72	74.67	76.96
DF-MLDeepMedic (S)	83.51	76.05	**70.85**	83.03	77.83	74.89	85.59	79.43	7399
DF-MLDeepMedic (L)	83.98	76.81	70.27	80.98	83.45	77.38	88.76	78.67	74.20
DF-MLDeepMedic $(L)^p$	**85.03**	**77.46**	70.77	84.74	84.11	77.39	86.29	79.11	74.91

Table 5
Performance of DF-MLDeepMedic (L) with different σ values on the local validation subset of BRATS 2017 (%).

σ	Dice			PPV			Sensitivity		
	Com.	Core	Enh.	Com.	Core	Enh.	Com.	Core	Enh.
(2, 1)	83.66	75.90	**71.10**	81.48	73.99	70.40	87.59	80.15	77.99
(5, 1)	83.95	75.70	69.23	80.96	76.15	72.30	88.62	78.96	73.91
(10, 1)	83.98	**76.81**	70.27	80.98	83.45	77.38	88.76	78.67	74.20
(5, 2)	83.93	74.57	69.87	83.05	75.49	73.40	86.34	79.21	73.42
(10, 2)	**84.44**	74.96	69.00	80.99	76.64	72.14	89.33	78.67	74.05

MLDeepMedic (S) stands for the DF-MLDeepMedic model using softmax loss as the auxiliary loss function, while DF-MLDeepMedic (L) utilizes the designed label distribution-based loss function. DF-MLDeepMedic $(L)^p$ denotes the performance of DF-MLDeepMedic (L) after the MLP-based post-processing step.

It is shown that DF-MLDeepMedic consistently outperforms MLDeepMedic, regardless of the detailed form of the auxiliary loss function. In particular, DF-MLDeepMedic (L) outperforms MLDeepMedic by as much as 3.56% on Dice score of the core tumor. These experimentation results indicate that higher-quality of multi-level features are learnt with the help of the auxiliary loss function.

The performance of DF-MLDeepMedic (L) is affected by the parameter σ of the two pathways. In Table 5, we evaluate different combinations of σ values, and select the best set of parameters according to the average Dice score on the three tumor regions. Finally, we set σ for the low-resolution pathway to be 10 and that for the full-resolution pathway to be 1. We then compare the performance of DF-MLDeepMedic (L) and DF-MLDeepMedic (S) in Table 4. On the one hand, it is clear that the label distribution-based loss function improves the performance of DF-MLDeepMedic on the complete tumor and tumor core regions. This is because the label distribution-based loss function encourages the network to learn more contextual information, which is important for the detection of complete tumor and tumor core regions. On the other hand, there is slight performance degradation of Dice score on the enhancing tumor region. We infer that the enhancing tumor region is usually very small in volume; therefore, introducing contextual information may weaken the original features for the enhancing tumor detection. This phenomenon can also be reflected by the performance of DF-MLDeepMedic (L) with different σ values, as shown in Table 5. In general, the label distribution-based loss function can improve the performance of DF-MLDeepMedic.

To further illustrate the advantage and mechanism of DFNs, we select several examples from the local validation subset and compare the prediction results by the above models in Fig. 5. As shown in Fig. 5, both DF-MLDeepMedic (S) and DF-MLDeepMedic (L) can make less false positive predictions. We owe this advantage to the improvement in robustness of features. Moreover, introducing the label distribution-based loss function can further improve the segmentation accuracy by taking more contextual information into account.

We also compare the performance between DF-U-Net and U-Net. Results are tabulated in Table 6. It is clear that DF-U-Net consistently outperforms U-Net for all three Dice scores. The above experimentation results justify the effectiveness of the dual-force training scheme.

Besides, as indicated in Tables 4 and 6, the MLP-based post-processing method can consistently improve the segmentation performance for both DF-MLDeepMedic and DF-U-Net by about 1% for the Dice scores of the three regions. In order to clearly show the advantage of the MLP-based post-processing method, we compare it with the RF-based (Random Forest-based) and 3D CRF-based (Conditional Random Fields) post-processing methods, and show their performance based on DF-U-Net in Table 7. For fair comparison, the feature vectors used by the RF-based method are exactly the same as those for our MLP-based method. For 3D CRF, we directly employ the method in [6,57], where the four normalized modalities of MRI scans and the prediction results are used as the input. Overall, the MLP-based post-processing method achieves the best performance among the three methods. Additionally, tt is worth noting that MLP can be computed very fast and therefore it is advantageous in efficiency. Furthermore, it can be readily applied to many other related segmentation tasks since it learns the post-processing rules automatically from data.

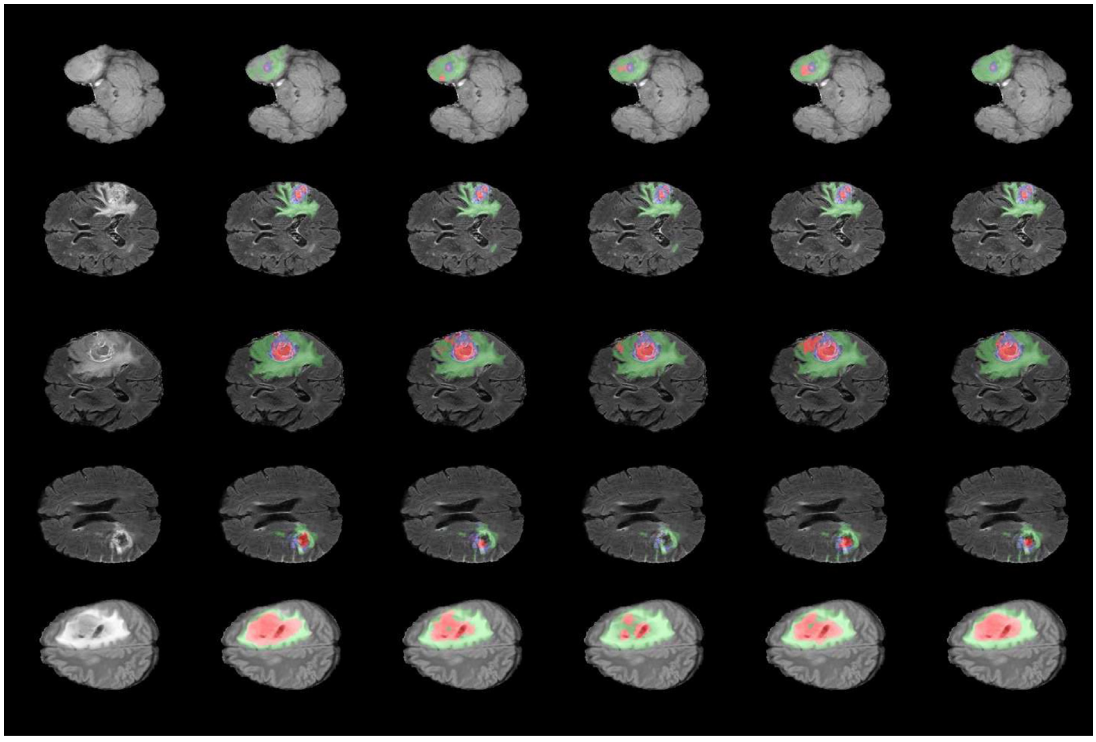


Fig. 5. Segmentation examples that demonstrate the effectiveness of DFN. The first column shows the original MRI slices. For simplicity, only the Flair modality is plotted. The second column presents the ground-truth segmentation results. The third to sixth columns illustrate the predictions of the DeepMedic, MLDeepMedic, DF-MLDeepMedic (S), and DF-MLDeepMedic (L), respectively. Colors in the figure represent different tumor classes, i.e., Edema (green), enhancing tumor (red), and necrosis and non-enhancing tumor (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Performance comparison between U-Net and DF-U-Net on the local validation subset of BRATS 2017 (%).

Method	Dice			PPV			Sensitivity		
	Com.	Core	Enh.	Com.	Core	Enh.	Com.	Core	Enh.
U-Net	84.11	73.80	69.15	84.32	73.03	71.66	85.15	79.88	74.30
DF-U-Net	84.14	75.32	70.72	84.85	76.27	69.93	84.74	78.73	78.05
DF-U-Net^p	**85.30**	**76.28**	**71.54**	86.54	78.09	74.08	84.98	78.78	75.58

Table 7
Performance Comparison of Different Post-Processing Methods on Local Validation Dataset of BRATS 2017 (%).

Method	Dice			PPV			Sensitivity		
	Com.	Core	Enh.	Com.	Core	Enh.	Com.	Core	Enh.
None	84.14	75.32	70.72	84.85	76.27	69.93	84.74	78.73	78.05
3D CRF	84.93	**77.33**	70.85	89.69	81.45	73.23	81.65	76.68	75.67
RF	85.06	75.88	70.26	86.35	77.74	74.11	84.81	78.65	73.94
MLP	**85.30**	76.28	**71.54**	86.54	78.09	74.08	84.98	78.78	75.58

4.6. Comparison with state-of-the-art methods

We compare the performance between DFNs and state-of-the-art methods on both BRATS 2017 online validation dataset and BRATS 2015 online test dataset. Comparison results are shown in Tables 8 and 9, respectively. In Table 8, it is shown that DFNs consistently outperform their baselines, further justifying the effectiveness of the proposed approaches. We cite the performance of the other approaches from their respective papers, including the state-of-the-art ones. DFNs outperform [58,59], but its performance is lower than Zhou et al. [13]. This is because approaches in [13] are complex cascade systems. For example, Zhou et al. [13] utilize as many as three models to promote the performance of their system. In comparison, the main purpose of this

Table 8
Comparison of Dice scores between DFNs and the state-of-the-art methods on BRATS 2017 Online Validation Dataset (%).

Method	Com.	Core	Enh.
DeepMedic	87.07	73.92	69.76
DF-MLDeepMedic (L)^p	89.30	73.88	73.46
U-Net	88.06	76.39	69.65
DF-U-Net^p	89.08	78.22	70.29
MC-Net	88.93	78.59	71.04
DF-MC-Net^p	89.68	79.84	72.05
Zhou et al. [13]	89	80	75
Jesson et al. [58]	89.9	75.1	71.3
Islam et al. [59]	87.6	76.1	68.9

Table 9
Comparison of Dice scores between DFNs and the state-of-the-art methods on the BRATS 2015 online test dataset (%).

Method	Com.	Core	Enh.
DeepMedic	83	67	63
DF-MLDeepMedic (L)^p	**85**	70	63
3D U-Net	83	68	63
DF-U-Net ^{p}	84	69	**64**
Kamnitsas et al. [6]	**85**	67	63
Zhao et al. [53]	84	73	62
Isensee et al. [60]	**85**	**74**	**64**

Table 10
Complexity analysis.

Method	Average time cost (per MRI scan)	#Params
DeepMedic	168.84s	84K
MLDeepMedic	186.93s	91K
DF-MLDeepMedic	186.93s	100K
U-Net	15.25s	6.3M
DF-U-Net	15.25s	6.3M

paper is to justify the effectiveness of the proposed approaches, rather than beating the performance of complex segmentation systems. It is also worth noting that the proposed dual-force training strategy can be applied to many existing models and therefore it can be employed to further improve the performance of existing complex systems. e.g., [13]. In order to verify the effectiveness of the proposed methods on cascade or ensemble systems, we choose Model-Cascade Networks (MC-Net for short) as a stronger baseline. MC-Net is a network-cascade system, that employs separate FusionNets [13] for each kind of tumor region. We apply the proposed methods to MC-Net, and report the results in Table 8. The improvement of applying the proposed methods to MC-Net clearly verifies that the proposed methods are applicable to ensemble or cascade systems.

We conduct similar comparison on the BRATS 2015 online test database in Table 9. It is also shown that the proposed methods consistently improve the segmentation performance of the baselines. We directly cite the performance of the state-of-the-art approaches [53,60] from their respective papers. DFNs achieve better performance than Kamnitsas et al. [6] and get comparable Dice scores on the complete tumor and enhancing tumor with [53,60]. The superior performance achieved in [53] partly depends on a series of handcrafted post-processing rules and manually determined parameters. In comparison, the proposed approach can automatically learn the post-processing rules and thus is easier to use in practice. Besides, the approach designed by Isensee et al. [60] is also based on U-Net and its contribution is parallel to DFN, which means that they can be combined together to further promote the performance of U-Net.

4.7. Complexity analysis

In Table 10, we list the average running time cost for segmenting one MRI scan for time complexity comparison and the number of learnable parameters in different networks for space complexity comparison. The running time costs of different networks are compared on an NVIDIA GTX Titan Xp GPU. It is worth noting that the proposed DF-Nets only predict label distributions in the training stage; therefore, they don't increase the time cost in the testing stage. It is clear that MLDeepMedic only increases the time and space complexity of DeepMedic by about 10%. Therefore, we can conclude that the complexity of the proposed models is acceptable.

5. Conclusion

Accurate brain tumor segmentation depends on multi-level information. However, existing deep models do not explicitly guarantee the quality of the learnt hierarchical features. In this paper, we propose a dual-force training strategy to explicitly encourage deep models to learn high-quality multi-level features. This is realized by a label distribution-based loss function to learn the abstract semantic information and a softmax loss function for segmentation using multi-level features. The dual-force training strategy can be applied to many popular networks, e.g., DeepMedic and U-Net. Applying the proposed strategy to deep models only slightly increases the time and space complexity while training. Besides, we also propose an MLP-based post-processing method that can automatically learn post-processing rules from data rather than manual summarization. Extensive experiments on two most recent brain tumor segmentation databases justify the efficiency and effectiveness of the proposed approaches. One shortage of the MLP-based post processing method is that its training process is separated from that of DFN; therefore, the entire framework is not completely end-to-end. In the future, we will further enhance the segmentation capability of the deep architecture, so that a separate post-processing stage can be skipped.

Acknowledgment

Changxing Ding was supported in part by the National Natural Science Foundation of China (Grant No.: 61702193), Science and Technology Program of Guangzhou (Grant No.: 201804010272), and the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (Grant No.: 2017ZT07X183).

References

[1] Q.T. Ostrom, H. Gittleman, P. Liao, T. Vecchione-Koval, Y. Wolinsky, C. Kruchko, J.S. Barnholtz-Sloan, Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2010–2014, Neuro-oncology 19 (suppl_5) (2017) v1–v88.
[2] H. Ohgaki, P. Kleihues, Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas, J. Neuropathol. Exp.Neurol. 64 (6) (2005) 479–489.
[3] D.N. Louis, H. Ohgaki, O.D. Wiestler, W.K. Cavenee, WHO Classification of Tumours of the Central Nervous System. World Health Organization Classification of Tumours, World Health Organization, 2007.
[4] N. Hattori, Y. Hirose, H. Sasaki, S. Nakae, S. Hayashi, S. Ohba, K. Adachi, T. Hayashi, Y. Nishiyama, M. Hasegawa, et al., World health organization grade ii–iii astrocytomas consist of genetically distinct tumor lineages, Cancer Sci. 107 (8) (2016) 1159–1164.
[5] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (brats), IEEE Trans. Med. Imaging 34 (10) (2015) 1993–2024.
[6] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78.
[7] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.
[8] B. Hariharan, P. Arbeláez, R. Girshick, J. Malik, Hypercolumns for object segmentation and fine-grained localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 447–456.
[9] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1395–1403.
[10] C. Ding, D. Tao, Trunk-branch ensemble convolutional neural networks for video-based face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 1002–1014.
[11] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, D. Tao, Stacked convolutional denoising auto-encoders for feature representation, IEEE Trans. Cybern. 47 (4) (2017) 1017–1027.
[12] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[13] C. Zhou, C. Ding, Z. Lu, T. Zhang, Brain tumor segmentation with cascaded convolutional neural networks, in: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS), 2017, pp. 328–333.

[14] B.H. Menze, K. Van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, P. Golland, A generative model for brain tumor segmentation in multi-modal images, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2010, pp. 151–159.
[15] S. Bauer, L.-P. Nolte, M. Reyes, Segmentation of brain tumor images based on atlas-registration combined with a markov-random-field lesion growth model, in: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, IEEE, 2011, pp. 2018–2021.
[16] L. Weizman, L.B. Sira, L. Joskowicz, S. Constantini, R. Precel, B. Shofty, D.B. Bashat, Automatic segmentation, internal classification, and follow-up of optic pathway gliomas in mri, Med. Image Anal. 16 (2012) 177–188.
[17] B.H. Menze, K. Van Leemput, D. Lashkari, T. Riklin-Raviv, E. Geremia, E. Alberts, P. Gruber, S. Wegener, M.-A. Weber, G. Szekely, et al., A generative probabilistic model and discriminative extensions for brain lesion segmentation with application to tumor and stroke, IEEE Trans. Med. Imaging 35 (4) (2016) 933–946.
[18] T. Wang, Z. Ji, Q. Sun, Q. Chen, Q. Ge, J. Yang, Diffusive likelihood for interactive image segmentation, Pattern Recognit. 79 (2018) 440–451.
[19] O. Gloger, K. Tönnies, Subject-specific prior shape knowledge in feature-oriented probability maps for fully automatized liver segmentation in mr volume data, Pattern Recognit. 84 (2018) 288–300.
[20] S. Bauer, L.-P. Nolte, M. Reyes, Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2011, pp. 354–361.
[21] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O.M. Thomas, T. Das, R. Jena, S.J. Price, Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2012, pp. 369–376.
[22] A. Pinto, S. Pereira, D. Rasteiro, C.A. Silva, Hierarchical brain tumour segmentation using extremely randomized trees, Pattern Recognit. 82 (2018) 105–117.
[23] B. Caldairou, N. Passat, P.A. Habas, C. Studholme, F. Rousseau, A non-local fuzzy segmentation method: application to brain mri, Pattern Recognit. 44 (9) (2011) 1916–1927.
[24] A. Gharipour, A.W.-C. Liew, Segmentation of cell nuclei in fluorescence microscopy images: an integrated framework using level set segmentation and touching-cell splitting, Pattern Recognit. 58 (2016) 1–11.
[25] Q. Zhu, B. Du, P. Yan, H. Lu, L. Zhang, Shape prior constrained pso model for bladder wall mri segmentation, Neurocomputing 294 (2018) 19–28.
[26] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556v1 (2014).
[27] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of International Conference on Machine Learning, 2015, pp. 448–456.
[28] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting., J.Mach.Learn.Res. 15 (1) (2014) 1929–1958.
[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[32] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 630–645.
[33] G. Huang, Z. Liu, K.Q. Weinberger, L. van der Maaten, Densely connected convolutional networks, arXiv:1608.06993v1 (2016).
[34] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, et al., Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377.
[35] D. Zikic, Y. Ioannou, M. Brown, A. Criminisi, Segmentation of brain tumor tissues with convolutional neural networks, in: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS), 2014, pp. 36–39.
[36] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in mri images, IEEE Trans. Med. Imaging 35 (5) (2016) 1240–1251.
[37] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31.
[38] S.S.M. Salehi, D. Erdogmus, A. Gholipour, Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging, IEEE Trans. Med. Imaging 36 (11) (2017) 2319–2330.
[39] H. Chen, Q. Dou, L. Yu, J. Qin, P.-A. Heng, Voxresnet: deep voxelwise residual networks for brain segmentation from 3d mr images, NeuroImage (2017). https://doi.org/10.1016/j.neuroimage.2017.04.041.
[40] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J.C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, X. Lladó, Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach, NeuroImage 155 (2017) 159–168.
[41] T. Brosch, L.Y. Tang, Y. Yoo, D.K. Li, A. Traboulsee, R. Tam, Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation, IEEE Trans. Med. Imaging 35 (5) (2016) 1229–1239.
[42] K.-L. Tseng, Y.-L. Lin, W. Hsu, C.-Y. Huang, Joint sequence learning and cross-modality convolution for 3d biomedical segmentation, arXiv:1704.07754v1 (2017).
[43] P.F. Christ, F. Ettlinger, F. Grün, M.E.A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, et al., Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks, arXiv:1702.05970v1 (2017).
[44] Q. Zhu, B. Du, B. Turkbey, P. Choyke, P. Yan, Exploiting interslice correlation for mri prostate image segmentation, from recursive neural networks aspect, Complexity 2018 (2018).
[45] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, D. Shen, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, NeuroImage 108 (2015) 214–224.
[46] P. Moeskops, J.M. Wolterink, B.H. van der Velden, K.G. Gilhuijs, T. Leiner, M.A. Viergever, I. Išgum, Deep learning for multi-task medical image segmentation in multiple modalities, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 478–486.
[47] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[48] F. Li, H. Qiao, B. Zhang, Discriminatively boosted image clustering with fully convolutional auto-encoders, Pattern Recognit. 83 (2018) 161–173.
[49] H. Chen, Q. Dou, X. Wang, J. Qin, J.C. Cheng, P.-A. Heng, 3d fully convolutional networks for intervertebral disc localization and segmentation, in: Proceedings of International Conference on Medical Imaging and Virtual Reality, Springer, 2016, pp. 375–382.
[50] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, et al., Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, in: Proceedings of International Conference on Information Processing in Medical Imaging, Springer, 2017, pp. 597–609.
[51] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, in: Artificial Intelligence and Statistics, 2015, pp. 562–570.
[52] X. Geng, C. Yin, Z.-H. Zhou, Facial age estimation by learning from label distributions, IEEE Trans. Pattern Anal. Mach. Intell. 35 (10) (2013) 2401–2412.
[53] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, Y. Fan, A deep learning model integrating fcnns and crfs for brain tumor segmentation, Med. Image Anal. 43 (2018) 98–111.
[54] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM International Conference on Multimedia, ACM, 2014, pp. 675–678.
[55] D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3d convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2015, pp. 4489–4497.
[56] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale video classification with convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1725–1732.
[57] P. Krähenbühl, V. Koltun, Efficient inference in fully connected crfs with gaussian edge potentials, in: Advances in Neural Information Processing Systems, 2011, pp. 109–117.
[58] A. Jesson, T. Arbel, Brain tumor segmentation using a 3d fcn with multi-scale loss, in: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS), 2017, pp. 116–126.
[59] M. Islam, H. Ren, Fully convolutional network with hypercolumn features for brain tumor segmentation, in: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS), 2017, pp. 108–115.
[60] F. Isensee, P. Kickingereder, D. Bonekamp, M. Bendszus, W. Wick, H.-P. Schlemmer, K. Maier-Hein, Brain Tumor Segmentation Using Large Receptive Field Deep Convolutional Neural Networks, in: Bildverarbeitung für die Medizin 2017, Springer, 2017, pp. 86–91.

Shengcong Chen is a Ph.D. candidate in South China University of Technology, where he received his B.S. degree in Electrical Engineering in 2017. His research focuses on medical image analysis and deep learning.

Changxing Ding received the Ph.D. degree from the University of Technology Sydney, Australia, in 2016. He is now with the School of Electronic and Information Engineering, South China University of Technology. His research interests include computer vision, deep learning, and medical image analysis.

Minfeng Liu obtained the Ph.D. degree in Tongji Medical College, Huazhong University of Science and Technology in 2006. He is now with the Nanfang Hospital, Southern Medical University. His research interests focus on cancer treatment and related medical image analysis topics.