One-Pass Multi-task Convolutional Neural Networks for Efficient Brain Tumor Segmentation

Chenhong Zhou[1], Changxing Ding[1(✉)], Zhentai Lu[2], Xinchao Wang[3], and Dacheng Tao[4]

[1] School of Electronic and Information Engineering,
South China University of Technology, Guangzhou, China
`chxding@scut.edu.cn`
[2] Guangdong Provincial Key Laboratory of Medical Image Processing,
School of Biomedical Engineering, Southern Medical University, Guangzhou, China
[3] Department of Computer Science, Stevens Institute of Technology, Hoboken, USA
[4] UBTECH Sydney AI Centre, SIT, FEIT, University of Sydney, Sydney, Australia

Abstract. The model cascade strategy that runs a series of deep models sequentially for coarse-to-fine medical image segmentation is becoming increasingly popular, as it effectively relieves the class imbalance problem. This strategy has achieved state-of-the-art performance in many segmentation applications but results in undesired system complexity and ignores correlation among deep models. In this paper, we propose a light and clean deep model that conducts brain tumor segmentation in a single-pass and solves the class imbalance problem better than model cascade. First, we decompose brain tumor segmentation into three different but related tasks and propose a multi-task deep model that trains them together to exploit their underlying correlation. Second, we design a curriculum learning-based training strategy that trains the above multi-task model more effectively. Third, we introduce a simple yet effective post-processing method that can further improve the segmentation performance significantly. The proposed methods are extensively evaluated on BRATS 2017 and BRATS 2015 datasets, ranking first on the BRATS 2015 test set and showing top performance among 60+ competing teams on the BRATS 2017 validation set.

1 Introduction

Brain tumors are one of the most fatal cancers worldwide [1]. Timely diagnosis of brain tumors from multimodal Magnetic Resonance Imaging (MRI) is of critical importance for treatment planning [2]. Automatic segmentation methods are highly desired in terms of efficiency and objectivity. However, automatic

brain tumor segmentation is still a challenging task due to large diversity of tumor shape, size, and location. Besides, there are four intra-tumoral classes, i.e., edema, necrosis, non-enhancing, and enhancing tumor. They are grouped into three overlapped regions which are required to be segmented for quantitative evaluation, i.e., complete tumor (all four classes), tumor core (all four classes except edema), and enhancing tumor (the enhancing tumor class only).

In recent years, Convolutional Neural Networks (CNNs) have been widely adopted for MRI-based brain tumor segmentation. CNN model architectures [3–6] have rapidly evolved from single-label prediction (predicting the label of a single voxel of the input patch) to dense-prediction (making predictions for voxels within the input patch simultaneously). To relieve the class imbalance problem, many recent works adopt the Model Cascade (MC) strategy for medical image segmentation [7,8]. For example, Wang et al. [8] decomposed multi-class brain tumor segmentation into a sequence of three successive binary segmentation tasks, each of which is solved by an independent network. MC relieves the class imbalance problem effectively by coarse-to-fine segmentation; therefore, its results are very encouraging. However, it comes with a price of system complexity and ignores the correlation among tasks.

Here we approach the above problems of MC via multi-task learning. We observe that multi-class brain tumor segmentation can be decomposed into three different but related tasks. Instead of training them individually like MC, we propose a One-pass Multi-task Network (OM-Net) that integrates the three tasks in a single model, which not only exploits their correlation in training but also simplifies the inference process by one-pass computation. Moreover, we design an effective training scheme based on curriculum learning, which is helpful to improve the convergence quality of OM-Net. Besides, to further improve performance, we propose a simple yet effective post-processing method to refine the segmentation results of OM-Net. Finally, the proposed approach obtains the first position on BRATS 2015 test set and achieves very competitive performance on BRATS 2017 validation set, respectively.

2   Methods

2.1   Model Cascade: A Strong Baseline

In this section, we first present an MC-based segmentation framework, as a strong baseline for OM-Net. We split multi-class brain tumor segmentation into three different but related tasks and each of them is implemented by an independent network. The three tasks are described as follows.

(1) *Coarse segmentation to detect complete tumor.* A network is trained to locate the complete tumor as a Region of Interest (ROI). Training patches are sampled randomly within the brain. To reduce overfitting, we train the network as a more difficult five-class segmentation problem. In testing, we still employ it as a binary segmentation task by merging the probability of four intra-tumoral classes. (2) *Refined segmentation for complete tumor and its intra-tumoral classes.* The coarse tumor mask obtained above is dilated by 5 voxels
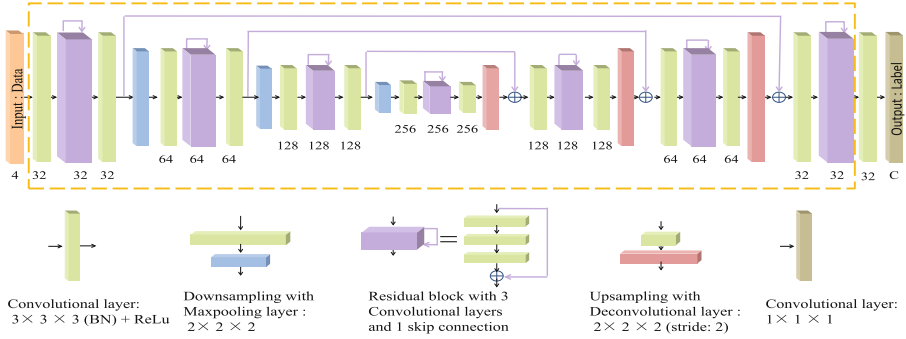
Fig. 1. Network architecture used in each task. The building blocks are represented by colored cubes with numbers below being the number of feature maps. C equals to 5, 5, and 2 for the first, second, and third task, respectively. (Best viewed in color)

to reduce false negatives. Then, the second network predicts labels of all voxels within the dilated region. Training patches are sampled randomly within the dilated ground-truth area of complete tumor. (3) *Precise segmentation for enhancing tumor.* Enhancing tumor is hard to segment due to the very unbalanced training data. We train the third network specially to segment enhancing tumor. Training patches for this network are sampled randomly within the ground-truth area of tumor core which covers all enhancing tumor voxels.

Network architecture for each task is identical except for the final convolutional classification layer. We use a 3D variant of the FusionNet [9], as illustrated in Fig. 1. Size of input patches for the network is $32 \times 32 \times 16 \times 4$, where the number 4 indicates the four MRI modalities. In testing, MC needs to run the three networks successively because the ROI of one network is determined by all its preceding networks. More specifically, the first network produces a coarse mask for complete tumor. The second network classifies all voxels in the dilated mask and obtains the precise region of complete tumor. Finally, we determine the precise enhancing tumor region by scanning all voxels in the complete tumor region using the third network. The tumor core region is meanwhile determined by merging results of the last two networks. Therefore, the entire inference process of MC requires alternate GPU-CPU computations for three times.

2.2   One-Pass Multi-task Network (OM-Net)

The above MC baseline can already achieve promising performance. However, it suffers from system complexity and ignores the correlation among the three tasks. We observe that the networks used for the three tasks are almost identical and their essential difference lies in training data. Inspired by this fact, we propose to transform the MC baseline into a single multi-task learning model. This model includes three tasks with their respective training data being the same as those in MC. Each task owns an independent convolutional layer, one classification layer, and one loss layer. All the other model parameters are shared to utilize

the underlying correlation among the tasks. In this model, predictions of the three classifiers can be obtained simultaneously in a single-pass. Therefore, we name the proposed model as One-pass Multi-task Network (OM-Net).

Observing that the three tasks are of increasing difficulty level, we propose to train OM-Net more effectively based on curriculum learning, which is realized by gradually increasing the difficulty of training tasks and is proved to improve the convergence quality of deep models [10]. Model architecture and training strategy of OM-Net are illustrated in Fig. 2. First, we train OM-Net with the first task only until the loss curve tends to flatten, which enables OM-Net to learn the basic knowledge of differentiating tumor and normal tissues.
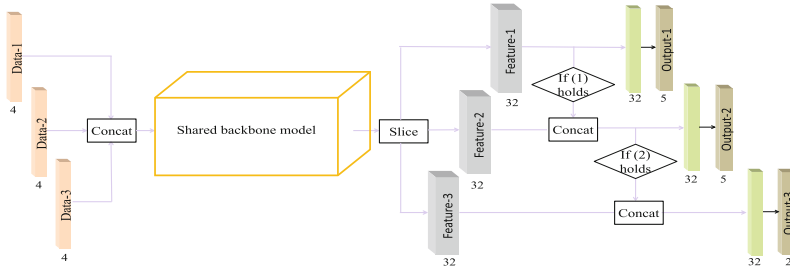

Fig. 2. Architecture of OM-Net. Data-i, Feature-i, and Output-i denote training data, feature, and classification layer for the i-th task, respectively. The shared backbone model refers to the network layers outlined by the yellow dashed line in Fig. 1.

Then, we add the second task to OM-Net. As shown in Fig. 2, Data-1 and Data-2 are concatenated along the batch dimension as the input for OM-Net. Features produced by the shared backbone model are sliced at the same position on the batch dimension to obtain task-specific features and are then used to train task-specific parameters. Moreover, we argue that not only knowledge (model parameters) but also learning material (training data), can be transferred from the easier course (task) to the more difficult course (task) in curriculum learning. Therefore, training patches in Data-1 that conform to the following sampling strategy can be reused in the second task:

$$\frac{\sum_{i=1}^{n} \mathbf{1}\{l_i \in C_{complete}\}}{n} \geq 0.4, \tag{1}$$

where l_i is the label of the i-th voxel in the patch, n is the total number of voxels in the patch, and $C_{complete}$ refers to the all tumor classes. We concatenate the features of patches in Data-1 that satisfy the above sampling condition to Feature-2 and then calculate the loss for the second task. Training process in this step continues until the loss curve of the second task tends to flatten.

Finally, we introduce the third task and its training data to OM-Net. The concatenation and slicing operations are similar to those in the second step.

Training patches from Data-1 and Data-2 that conform to the following sampling strategy can be reused in the third task:

$$\frac{\sum_{i=1}^{n} \mathbf{1} \left\{ l_i \in C_{core} \right\}}{n} \geq 0.5, \qquad (2)$$

where C_{core} refers to the tumor classes that belong to tumor core. The three tasks in OM-Net are trained together until convergence.

During inference, OM-Net obtains the predictions of the three tasks simultaneously. The way that OM-Net utilizes these results for final segmentation is exactly the same as that in MC. It is worth noting that OM-Net is essentially different from one existing multi-task model for brain tumor segmentation [11]. The model in [11] aims to provide more diverse supervision signals for the same training data. In comparison, OM-Net integrates tasks that have respective training data and aims to accomplish coarse-to-fine segmentation by a single model.

2.3   Post-processing

We further propose a novel post-processing method to refine the segmentation results of OM-Net. Our proposed method is mainly inspired by [6], but is more robust and easier to use in practice. It consists of two steps. First, isolated small clusters whose volume is less than one-tenth of the maximum 3D connected tumor area are removed. This step is identical to step 3 in [6]. Second, it is observed that when the volume of predicted enhancing tumor is less than five percent of the volume of the complete tumor, non-enhancing voxels tend to be falsely predicted as edema [6]. We find that this problem also happens in OM-Net and propose a fully-automatic method to relieve this problem. Specifically, we employ the K-means clustering algorithm to cluster the predicted edema voxels into two groups according to their intensity values in MRI images. For each group, we compute the average probability of all its voxels belonging to the non-enhancing class, according to the prediction results of OM-Net. Labels of voxels in the group with the higher averaged probability are converted to non-enhancing, while those in the other group remain unchanged.

Compared with the approach in [6] that depends on manually determined threshold, our proposed approach is automatic and flexible. In the experiment section, we find it promotes the performance of OM-Net significantly.

3   Experiments

We evaluate the performance of the proposed methods on BRATS 2017 and BRATS 2015 datasets, respectively. The brain of each patient is scanned with four modalities, i.e., Flair, T1, T1c, and T2. All the images have been skull-striped and co-registered. For pre-processing, voxel intensities inside the brain are normalized to have zero mean and unit variance for each modality image. We sample around 400,000, 400,000, and 200,000 patches for the first, second,

and third task, respectively. All networks are implemented based on the $C3D$[1] package, a modified version of $Caffe$[12]. We adopt SoftmaxWithLoss as the loss function and use stochastic gradient descent to train all networks. The initial learning rate of all networks is 0.001 and then divided by 2 after every 4 epochs. Each network in MC is trained for 20 epochs. OM-Net is trained for 1 epoch, 1 epoch, and 18 epochs for each of its three steps, respectively.

3.1 Results on BRATS 2017 Dataset

The training set of BRATS 2017 [2,13–15] contains 285 MRI images. The validation set of BRATS 2017 contains 46 MRI images with hidden ground-truth and evaluation on this set is conducted online. For more convenient evaluation, we randomly divide the training set into two subsets, i.e., a training subset including 260 MRI images and a local validation subset including 25 MRI images.

We first carry out a number of experiments on the local validation subset. Quantitative comparison results are tabulated in Table 1[2]. Here MC1, MC2, and MC3 indicate the one-model, two-model, and three-model cascades, respectively. In order to justify the effectiveness of the curriculum learning-based training strategy, we further test OM-Net[0] (a naive multi-task learning model without training data transfer or step-wise training) and OM-Net^d (a multi-task learning model with training data transfer but no step-wise training). OM-Net$^{p^1}$ and OM-Net^p denote OM-Net with the first post-processing step and both post-processing steps, respectively. In addition, we also provide qualitative comparisons between MC3, OM-Net, and OM-Net^p in the *supplementary materials*.

Table 1. Performance on the local validation subset of BRATS 2017 (%)

Method	Parameters	Dice			Positive predictive value			Sensitivity		
		Complete	Core	Enhancing	Complete	Core	Enhancing	Complete	Core	Enhancing
MC1	13.81 M	90.41	78.48	72.91	95.09	87.23	75.72	87.05	77.68	83.44
MC2	27.62 M	91.08	79.11	75.14	91.19	86.67	81.84	91.55	78.03	80.26
MC3	41.43 M	91.08	79.58	79.95	91.19	85.98	84.96	91.55	79.56	81.61
OM-Net	13.86 M	91.10	79.87	**80.87**	92.42	87.60	85.44	90.23	78.31	82.80
OM-Net[0]	13.86 M	90.40	79.42	79.96	91.25	87.57	82.68	90.52	76.87	83.91
OM-Net^d	13.86 M	91.11	79.92	80.24	92.11	87.78	84.19	90.73	78.09	83.28
OM-Net$^{p^1}$	13.86 M	91.28	79.88	80.84	93.38	87.71	85.44	89.87	78.26	82.75
OM-Net^p	13.86 M	**91.28**	**82.50**	80.84	93.38	83.60	85.44	89.87	83.60	82.75

First, Table 1 shows the Dice scores are steadily improved with the increase of model number in MC, which justifies the effectiveness of each model in MC. However, larger number of models leads to system complexity and more storage consumption. Second, with only one-third of the parameters of MC3, OM-Net

---

[1] https://github.com/facebook/C3D.

[2] Dice score is the overall evaluation index, identical to F measure. Therefore, we only highlight the best Dice scores in bold in Tables 1 and 2.

achieves better Dice scores consistently, especially for tumor core and enhancing tumor. Third, OM-Net outperforms both OM-Net^0 and OM-Net^d, demonstrating the effectiveness of the proposed training strategy. Fourth, the first post-processing step slightly improves the Dice score for complete tumor as it removes part of false positives. The proposed second step significantly improves the Dice score of tumor core by as much as 2.62%. The above results justify the effectiveness of the proposed approaches.

Additionally, we evaluate the performance of OM-Net^p on BRATS 2017 validation set and compare it with the other 60+ participants. OM-Net^p achieves Dice scores of 77.841%, 90.386%, and 82.792% for enhanced tumor (ET), whole tumor (WT), and tumor core (TC), respectively, and ranks second on the online leaderboard in terms of the averaged Dice score. The approach proposed in [8] currently ranks first, outperforming OM-Net^p by 0.74%, 0.11%, and 0.99% on the Dice scores for ET, WT, and TC, respectively. However, the approach in [8] is a complicated ensemble system that includes as many as 9 models. In comparison, there is only a single model in our approach.

3.2 Results on BRATS 2015 Dataset

The BRATS 2015 dataset consists of 274 MRI images for training and 110 MRI images for testing. We use all training data to train OM-Net and MC3. Evaluation is conducted on the test set. The results are tabulated in Table 2.

Table 2. Performance on BRATS 2015 test set (%)

Method	Dice			Positive predictive value			Sensitivity		
	Complete	Core	Enhancing	Complete	Core	Enhancing	Complete	Core	Enhancing
MC3	86	70	63	86	82	60	88	67	72
OM-Net	86	71	64	86	83	61	88	68	72
OM-Net$^{p^1}$	87	71	64	87	83	61	89	68	72
OM-Net^p	**87**	**75**	**64**	87	81	61	89	75	72
Isensee et al. [16]	85	74	64	83	80	63	91	73	72
Zhao et al. [6]	84	73	62	89	76	63	82	76	67
Kamnitsas et al. [5]	85	67	63	85	86	63	88	60	67

First, we compare the results of MC3, OM-Net, OM-Net$^{p^1}$, and OM-Net^p. Table 2 shows that OM-Net consistently outperforms MC3, with 1% higher Dice scores on both tumor core and enhancing tumor. Besides, the first post-processing step improves the Dice score of OM-Net by 1% on the complete tumor region; the proposed second post-processing step significantly improves the Dice score of tumor core by 4%. The above results are consistent with the conclusions on the BRATS 2017 data. Second, we compare the performance of OM-Net^p with the other leading approaches on the BRATS 2015 test set. It is observed in Table 2 that OM-Net^p beats the state-of-the-art approaches on Dice scores and ranks first currently on the online leaderboard.

4   Conclusion

We propose the OM-Net model trained with the curriculum learning-based strategy to relieve the class imbalance problem in brain tumor segmentation. Unlike the popular MC framework, OM-Net integrates multiple networks in MC into a single deep model and conducts coarse-to-fine segmentation in a single pass. Therefore, it substantially saves model parameters and reduces system complexity. OM-Net is also advantageous as it effectively utilizes the correlation between the tasks. With a single and light model, the proposed approach ranks first on BRATS 2015 test set and achieves top performance on BRATS 2017 dataset.

References

1. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. Phys. Med. Biol. **58**, R97 (2013)
2. Menze, B.H., Jakab, A., Bauer, S., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE TMI **34**(10), 1993–2024 (2015)
3. Pereira, S., et al.: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imag. **35**(5), 1240–1251 (2016)
4. Havaei, M., Davy, A., Warde-Farley, D., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
5. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
6. Zhao, X., Wu, Y., Song, G., et al.: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med. Image Anal. **43**, 98–111 (2018)
7. Christ, P.F., et al.: Automatic Liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 415–423. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_48
8. Wang, G., et al.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. arXiv preprint arXiv:1709.00382 (2017)
9. Quan, T.M., et al.: Fusionnet: a deep fully residual convolutional neural network for image segmentation in connectomics. arXiv preprint arXiv:1612.05360 (2016)
10. Bengio, Y., et al.: Curriculum learning. In: ICML, pp. 41–48. ACM (2009)
11. Shen, H., Wang, R., Zhang, J., McKenna, S.: Multi-task fully convolutional network for brain tumour segmentation. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 239–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_21
12. Jia, Y., Shelhamer, E., Donahue, J., et al.: Caffe: convolutional architecture for fast feature embedding. In: 22nd ACM MM, pp. 675–678 (2014)

13. Bakas, S.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nat. Sci. Data **4**, 170117 (2017)
14. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. Cancer Imaging Arch. (2017)
15. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. Cancer Imaging Arch. (2017)
16. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 287–297. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_25