Semi-supervised Brain Lesion Segmentation with an Adapted Mean Teacher Model

Wenhui Cui[1], Yanlin Liu[2], Yuxing Li[2], Menghao Guo[1], Yiming Li[3], Xiuli Li[3,4], Tianle Wang[5], Xiangzhu Zeng[6], and Chuyang Ye[2(✉)]

[1] School of Computer Science and Technology, Xidian University, Xi'an, China
[2] School of Information and Electronics, Beijing Institute of Technology, Beijing, China
chuyang.ye@bit.edu.cn
[3] Deepwise AI Lab, Beijing, China
[4] Peng Cheng Laboratory, Shenzhen, China
[5] Department of Radiology, The People's Hospital of Nantong, Nantong, China
[6] Department of Radiology, Peking University Third Hospital, Beijing, China

Abstract. Automated brain lesion segmentation provides valuable information for the analysis and intervention of patients. In particular, methods that are based on *convolutional neural networks* (CNNs) have achieved state-of-the-art segmentation performance. However, CNNs usually require a decent amount of annotated data, which may be costly and time-consuming to obtain. Since unannotated data is generally abundant, it is desirable to use unannotated data to improve the segmentation performance for CNNs when limited annotated data is available. In this work, we propose a *semi-supervised learning* (SSL) approach to brain lesion segmentation, where unannotated data is incorporated into the training of CNNs. We adapt the mean teacher model, which is originally developed for SSL-based image classification, for brain lesion segmentation. Assuming that the network should produce consistent outputs for similar inputs, a loss of segmentation consistency is designed and integrated into a self-ensembling framework. Self-ensembling exploits the information in the intermediate training steps, and the ensemble prediction based on the information can be closer to the correct result than the single latest model. To exploit such information, we build a student model and a teacher model, which share the same CNN architecture for segmentation. The student and teacher models are updated alternately. At each step, the student model learns from the teacher model by minimizing the weighted sum of the segmentation loss computed from annotated data and the segmentation consistency loss between the teacher and student models computed from unannotated data. Then, the teacher model is updated by combining the updated student model with the historical information of teacher models using an exponential moving average strategy. For demonstration, the proposed approach was evaluated on ischemic stroke lesion segmentation. Results indicate that the proposed method improves stroke lesion segmentation with the incorporation of unannotated data and outperforms competing SSL-based methods.

1   Introduction

Automated segmentation of brain lesions in *magnetic resonance images* (MRIs) provides valuable information for the analysis and intervention of patients [6]. Deep learning based approaches have been developed for the segmentation of different types of brain lesions, such as stroke lesions [6,7] and brain tumors [4, 10,15]. Various architectures of *convolutional neural networks* (CNNs) have been proposed and have achieved state-of-the-art segmentation performance. Deep learning based approaches usually involve a huge number of parameters and thus require a decent amount of annotated data, so that the parameters can be properly learned [14]. However, manual annotation of brain lesions is costly and time-consuming, whereas unannotated data is often abundant. Therefore, it is desirable to exploit the unannotated data when there is limited annotated data for training.

Semi-supervised learning (SSL) techniques have emerged as means to combine the limited annotated data and the abundant unannotated data to improve the training process [16]. Several methods have been proposed for medical image segmentation [1,14]. For example, the consistency of feature embedding between annotated and unannotated data is enforced in [1], where a consistency loss is incorporated into the loss function and provides regularization for training the CNN. A similar idea is developed in [5], where the consistency of feature embedding is ensured with an adversarial learning strategy. Note that although the approach in [5] is originally developed for transfer learning, it can be applied to SSL as well. Another approach in [14] aims to achieve similar quality of segmentation on the annotated and unannotated data. The similarity is encouraged with a deep adversarial network model, which consists of a segmentation network and an evaluation network. These approaches have achieved promising results when limited annotated data is available. However, the development of SSL techniques for CNN-based brain lesion segmentation is still an open problem, where improved segmentation performance is desired.

In this work, we explore the integration of SSL into CNN-based brain lesion segmentation. Inspired by the success of the *mean teacher* (MT) model [11] for SSL-based image classification, we propose an adapted MT model for brain lesion segmentation, where both annotated and unannotated data can be exploited to boost segmentation performance.

We assume that the segmentation should be consistent for similar input data [8], and define a segmentation consistency loss, which is computed for a pair of inputs that are obtained by adding noises to the same unannotated sample. In this way, unannotated data can be incorporated into the learning process and provide regularization information. Note that unlike in previous works [1,14] that measure the consistency between annotated data and unannotated data, here the consistency is computed between two noisy versions of the

same unannotated data. Since it is observed in [8] and [11] that self-ensembling could lead to better classification models, we apply a similar strategy to brain lesion segmentation and integrate the segmentation consistency loss into the self-ensembling framework. Specifically, we build a teacher model and a student model, which share the same network architecture. In this work we select the DeepMedic architecture [6] for the two models, because it has achieved state-of-the-art performance in brain lesion segmentation [9]. Self-ensembling is based on the observation that the ensemble prediction combining the network information after each step is more accurate than the current output [8,11]. Thus, the teacher model records the information at each step, and the student model learns from the teacher model by minimizing the loss of segmentation accuracy for annotated data and the consistency loss with respect to the outputs of the teacher model for unannotated data. Then, the teacher model is updated by combining the historical information of teacher models and the current student model with an *exponential moving average* (EMA) strategy. The student and teacher models are updated alternately, and the final teacher model is used for segmentation on test samples.

For demonstration, the proposed approach was evaluated on ischemic stroke lesion segmentation. Results indicate that the proposed method improves the segmentation quality by incorporating unannotated data and outperforms competing SSL-based segmentation strategies.

2   Methods

In this section, we first introduce the backbone CNN architecture shared by the teacher and student models. Then, we describe how unannotated data is used by the teacher and student models to regularize the model training. Finally, implementation details are given.

2.1   Backbone CNN Architecture

Due to its superior segmentation performance, the DeepMedic model [6] is used as our backbone network structure, which is shared by the teacher and student models. Specifically, DeepMedic is a dual pathway, 11-layer deep, three-dimensional CNN, and it performs multi-scale processing via parallel convolutional pathways. A graphical illustration of DeepMedic is shown in Fig. 1, and the parameters of each layer are summarized in Table 1. Note that the two pathways use the same settings of convolutional layers 1–8 in Table 1. DeepMedic takes image patches at two different resolutions as input. The two patches are centered at the same image location. The upper pathway in Fig. 1 takes normal resolution image patches as input, whereas the bottom pathway in Fig. 1 operates on downsampled patches (by a factor of three). Before the final segmentation, the multi-scale features are concatenated and fed into 1^3 convolutional layers. For more details about DeepMedic, we refer readers to [6].
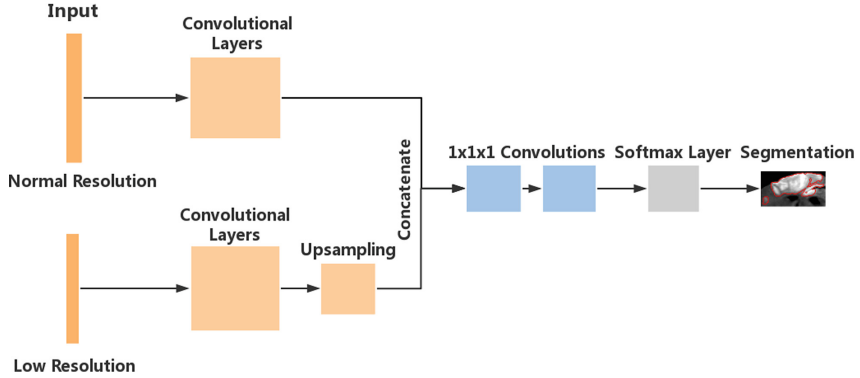
Fig. 1. The DeepMedic model proposed in [6].

Table 1. The specification of layers in DeepMedic.

Layer	Parameters
Convolutional layer 1	30 filters, 3^3
Convolutional layer 2	30 filters, 3^3
Convolutional layer 3	40 filters, 3^3
Convolutional layer 4	40 filters, 3^3
Convolutional layer 5	40 filters, 3^3
Convolutional layer 6	40 filters, 3^3
Convolutional layer 7	50 filters, 3^3
Convolutional layer 8	50 filters, 3^3
1^3 Convolution	250 filters, 1^3
1^3 Convolution	250 filters, 1^3

2.2 Semi-supervised Lesion Segmentation with an Adapted MT Model

To leverage the abundant unannotated data for lesion segmentation, we propose to use an SSL strategy. Our strategy is inspired by the MT model [11], which is developed for SSL-based image classification. Like in the MT model, we assume that CNN models should favor functions that produce consistent outputs for similar inputs. Pairs of similar input samples are generated by adding noises to the same unannotated data. In this way, unannotated data can be used to provide regularization for training the network. Note that unlike MT, we need to measure the consistency of segmentation instead of classification. Thus, we adapt the MT strategy by defining a segmentation consistency loss. The segmentation consistency loss is then integrated into a self-ensembling framework, which is motivated by the observation that the ensemble prediction based on

the combined information after each step can be more accurate than the current output [11]. The detailed description of the proposed approach is given below.

For an unannotated input X_u, we add noises η and η' sampled from the same distribution, and the network is expected to produce similar outputs for the two noisy inputs. Although it is possible to directly incorporate a consistency loss based on the similarity into DeepMedic, which leads to a strategy similar to the Π model in [8] for classification, integration of the consistency loss into a self-ensembling framework can lead to better model training [11]. Therefore, like the original MT approach, we build a teacher model and a student model, where the student model attempts to learn the targets generated by the teacher model.

Both the teacher and student models share the same DeepMedic architecture [6]. Note that the proposed framework is not restricted to a specific segmentation network, and can be applied to other backbone segmentation architectures as well, such as 3D U-Net [2]. The two noisy inputs associated with η and η' are then fed into the student model and the teacher model, respectively. Since the student and teacher models share the same architecture, we denote their output for the noisy input as $f(X_u, \eta, \theta)$ and $f(X_u, \eta', \theta')$, respectively. Here, θ and θ' are the weights in the network of the student model and the teacher model, respectively.

The teacher model is initialized with the DeepMedic network trained with annotated data. Then, the teacher and student models are updated alternately. At each step, the student model learns from the teacher model by minimizing the weighted sum of the consistency loss $\mathcal{L}_c$ of unannotated data and segmentation loss $\mathcal{L}_s$ of annotated data. Specifically, we define $\mathcal{L}_c$ as the soft Dice loss between the predicted probability maps of the student and teacher models based on their corresponding noisy inputs

$$\mathcal{L}_c = 1 - \mathbb{E}_{X_u, \eta, \eta'} \left[\frac{1}{K} \sum_{i=1}^{K} \frac{\sum_{v=1}^{V} 2 f_v^i(X_u, \eta, \theta) f_v^i(X_u, \eta', \theta')}{\sum_{v=1}^{V} f_v^i(X_u, \eta, \theta) + \sum_{v=1}^{V} f_v^i(X_u, \eta', \theta')} \right] \quad (1)$$

where $f_v^i(\cdot)$ represents the $f(\cdot)$ value that is at the v-th voxel and takes the i-th label, K represents the total number of possible labels, and V denotes the total number of voxels in the input. With the loss defined in Eq. (1), the output of the teacher model can also be considered a target label for the student model to learn. As in DeepMedic [6], $\mathcal{L}_s$ is the cross entropy loss between the predictions $f(X_a, \theta)$ (no noise η is applied) of the student model for the annotated input X_a and the corresponding annotation Y

$$\mathcal{L}_s = -\mathbb{E}_{X_a, Y} \left[\frac{1}{V} \sum_{v=1}^{V} \sum_{i=1}^{K} Y_v^i \log \left(f_v^i(X_a, \theta) \right) \right], \quad (2)$$

where Y_v^i represents the value of Y at the v-th voxel with label i.

Then, the total loss function $\mathcal{L}$ of our model is

$$\mathcal{L} = \mathcal{L}_s + \beta \mathcal{L}_c, \quad (3)$$

where β is an adaptive weighting coefficient. As suggested in [11], different β is used at different steps. Specifically, $\beta = \exp\left(-5(1 - \frac{S}{L})^2\right)$ (when $S \le L$), where S is the current training step and L is called the ramp-up length; when $S > L$, β is set to one. In our experiment, we empirically set $L = 400$. Such an adaptive setting of β keeps the effect of consistency down in early steps, because the teacher model may not generate reasonable target labels at the beginning [11].

With the parameters θ_t of the student model at step t, we perform the EMA of weights to aggregate information in training steps as in the original MT model [11]. Specifically, we update the teacher model as follows

$$\theta'_t = \alpha\theta'_{t-1} + (1 - \alpha)\theta_t, \tag{4}$$

where α is the EMA decay. Compared with other ensembling strategies [8], EMA better prevents overfitting, especially when a large number of model parameters are learned from limited training data [11]. Following the MT method [11], we used $\alpha = 0.99$ in the first L steps (the ramp-up phase), and $\alpha = 0.999$ for the rest of the training. This strategy facilitates the teacher model to (1) forget the old inaccurate student weights quickly and (2) benefit from a longer memory when the student improvement slows down after the ramp-up phase. The final teacher model is used to perform lesion segmentation for test samples.

2.3   Implementation Details

The proposed method is implemented using TensorFlow (https://www.tensorflow.org). In the training of student models, we followed the settings in [6] and minimized the loss with an RMSProp optimizer [12], where the learning rate is 0.0001 and the decay rate is 0.9. The batch size is 16, which consists of eight annotated and eight unannotated samples. Both the annotated and unannotated training patches were sampled from the lesion region and healthy tissue with equal probability, which mitigates class imbalance [6]. Note that since the lesion region is unknown for unannotated data, it is approximated by the DeepMedic prediction.

The noise injection for the proposed method was applied as follows. We applied Gaussian noises to the inputs of the student and teacher models. Noise η consists of two different types: the additive noise η_s and the multiplicative noise η_m. Both are sampled from Gaussian distributions. At each voxel of the input patch, noise was applied independently, and the noisy intensity I' is computed from the original intensity I as follows

$$I' = (I + \eta_s) \times \eta_m. \tag{5}$$

3   Experiments

3.1   Data Description

For demonstration, the proposed method was evaluated on a task of ischemic stroke lesion segmentation. A total number of 246 *diffusion weighted*

images (DWIs) of ischemic stroke patients were acquired on a 3T Siemens Verio scanner, where a *b*-value of 1000 s/mm^2 was applied and a *b*0 image (the image without diffusion weighting) was also acquired. The image resolution is 0.96 mm × 0.96 mm × 6.5 mm and the image dimension is 240 × 240 × 21. Manual delineations of stroke lesions were performed by an experienced radiologist on 50 DWIs, and the rest 196 DWIs are unannotated.

The image intensities were normalized for each scan. Specifically, a brain mask was extracted with the Dipy software [3]. Then, the mean and standard deviation of the intensity in the brain were computed. The mean was subtracted from the intensity at each voxel in the skull-stripped image and the resulting intensities were further divided by the standard deviation. The patch sizes for the normal resolution and downsampled pathways are 37 × 37 × 21 and 23 × 23 × 18, respectively, so that the multi-scale features can be concatenated. The additive noise was sampled from a Gaussian distribution which has a zero mean and a standard deviation of 0.05, whereas the multiplicative noise was sampled from a Gaussian distribution which has a mean of 1.0 and a standard deviation of 0.01.

3.2   Training Phase

We randomly selected 20 annotated subjects as training scans, and used the rest 30 annotated data as test scans. The 196 unannotated scans were included in the training process of the proposed approach as well. The training was performed on an NVIDIA GeForce GTX 1080Ti GPU, and it took about 12 h. The training process was evaluated and the Dice coefficients of the training data are shown for the student and teacher models in Fig. 2. We can see that both models better fit the training data as the training continues and become stable in the end. In addition, the teacher model consistently achieves higher Dice coefficients than the student model, until it is close to the end of the ramp-up phase (400 steps), where the two Dice coefficients are close. These observations are consistent with the assumption in self-ensembling and the settings of the EMA decay.

3.3   Evaluation of Lesion Segmentation

Then, we evaluated the segmentation results of the proposed method. The proposed method was compared with three methods. The DeepMedic approach [6] was included as the baseline method that does not use unannotated data for training. The strategy used by [14] was also integrated with DeepMedic for comparison, which performs SSL-based image segmentation. This strategy assumes that the segmentation of unannotated data should follow a distribution that is similar to that of annotated data. Such similarity is enforced by a separate evaluation network with adversarial learning. We replaced the segmentation network in [14] with DeepMedic for lesion segmentation. Due to the use of an evaluation network, this strategy is referred to as DeepMedic-EN. The network in [5] for unsupervised domain adaptation was also considered, because although it is originally developed for transfer learning, it can be directly used for SSL. This strategy applies an idea that is similar to [14], where adversarial learning is
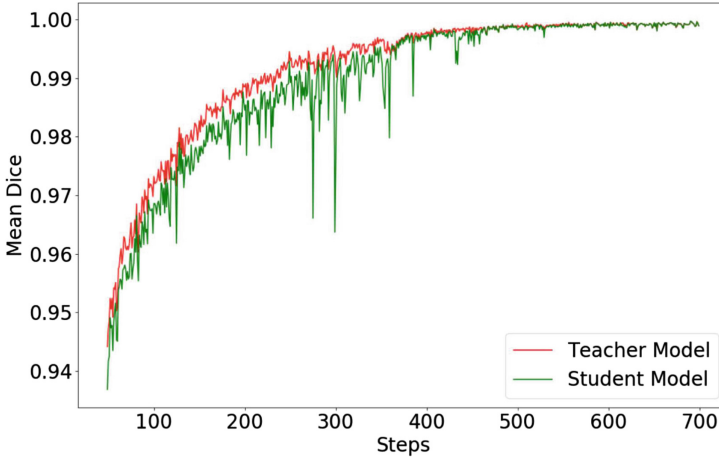
Fig. 2. The mean Dice coefficients of the training data for the teacher model and the student model after each training step.

applied so that the features extracted from the source (annotated) and target (unannotated) data follow similar distributions. Since the deep network in [5] is based on the DeepMedic architecture, we used the structure directly, and the method is referred to as DeepMedic-UDA, where UDA stands for unsupervised domain adaptation as described by [5].

We first qualitatively evaluated the proposed method. Cross-sectional views of the segmentation results overlaid on DWIs are shown in Fig. 3 for two representative test subjects with different sizes of lesions. The gold standard of the manual delineation and the results of the competing methods are also shown for comparison. It can be seen that the proposed method produced segmentation that better agrees with the gold standard.

Next, the proposed method was quantitatively evaluated. We computed the Dice coefficients on the test scans for the proposed and competing methods, and the results are summarized in Table 2. Here, the means and standard deviations of the Dice coefficients computed from the 30 test subjects are listed. The proposed method has the highest mean Dice coefficients, which indicates its better segmentation quality than the competing methods. In addition, the results were compared between the proposed method and each competing method with a paired Student's t-test. In all cases, the difference is significant ($p < 0.05$). Note that DeepMedic-EN and DeepMedic-UDA have smaller mean Dice coefficients than the baseline DeepMedic. This is possibly due to the limited number of training scans, which cannot adequately represent the distribution of annotated data. Thus, the adversarial learning in DeepMedic-EN and DeepMedic-UDA may incorrectly modify the segmentation result.
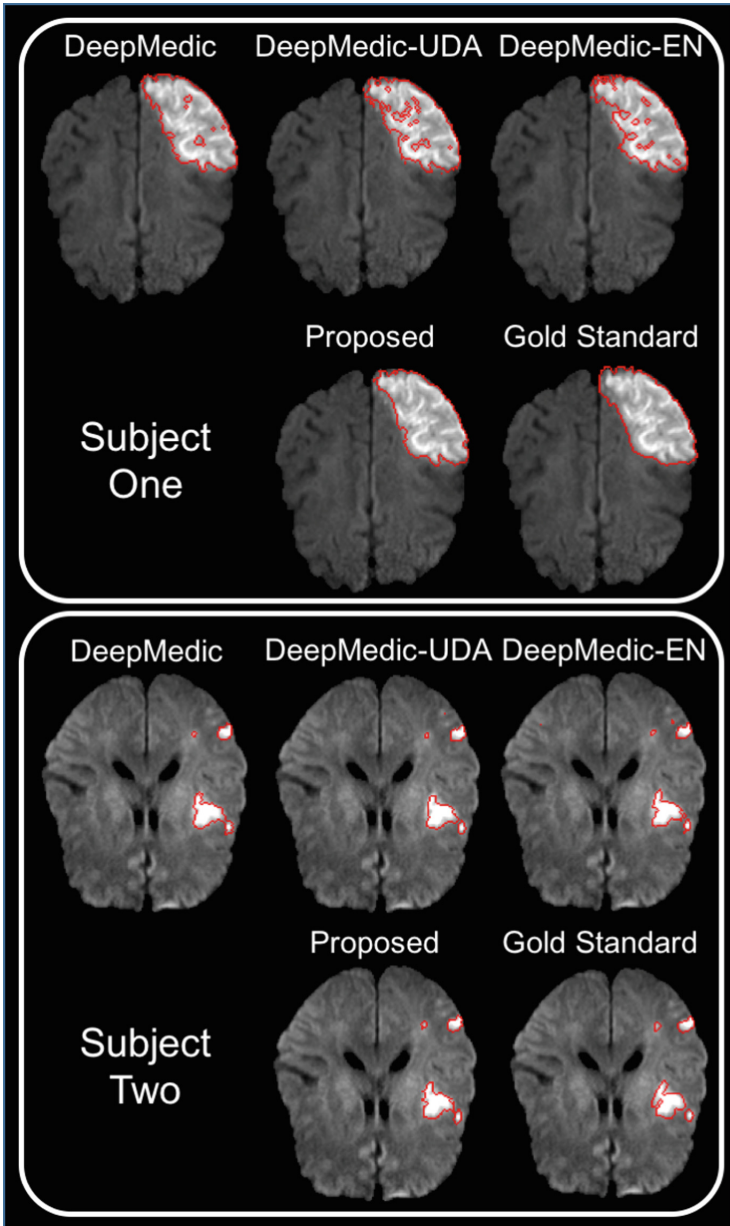
Fig. 3. Cross-sectional views of the segmentation results overlaid on DWIs for two representative test subjects with different sizes of lesions.

Table 2. Means and standard deviations of the Dice coefficients on test scans when 20 annotated scans were used for training. Best results are highlighted in bold font. Asterisks (*) indicate that the difference between the proposed method and the competing method is statistically significant ($p < 0.05$) using a paired Student's t-test.

DeepMedic	DeepMedic-UDA	DeepMedic-EN	Proposed
$0.6312 \pm 0.2617^*$	$0.6096 \pm 0.2819^*$	$0.6236 \pm 0.2577^*$	**0.6676 ± 0.2392**

Table 3. Means and standard deviations of the Dice coefficients on test scans when 10 and 30 annotated scans were used for training. Best results are highlighted in bold font. Asterisks (*) indicate that the difference between the proposed method and the competing method is statistically significant ($p < 0.05$) using a paired Student's t-test.

	10 annotated training scans	30 annotated training scans
DeepMedic	$0.5912 \pm 0.2243^*$	$0.6551 \pm 0.2514^*$
DeepMedic-UDA	$0.5562 \pm 0.2878^*$	$0.6627 \pm 0.2208^*$
DeepMedic-EN	$0.5684 \pm 0.2873^*$	$0.6503 \pm 0.2650^*$
Proposed	**0.6518 ± 0.2484**	**0.6879 ± 0.2334**

3.4   Impact of the Amount of Training Data

Lastly, we investigated the impact of the number of training scans. Specifically, we investigated two additional cases, where 10 and 30 randomly selected annotated scans were included in training and the rest annotated data were used for testing. For SSL-based methods, all the unannotated data were also included in training. The results are shown in Table 3, where the means and standard deviations of the Dice coefficients are listed. In all cases, the proposed approach has higher mean Dice coefficients than the competing methods, and the difference is significant using a paired Student's t-test. These results indicate that the proposed method outperforms the competitors.

4   Discussion

The original MT model is developed for semi-supervised image classification, and its consistency loss is simply the difference between class predictions. In our task, however, the consistency needs to be enforced for segmentation. Thus, we have adapted the MT model by defining the consistency loss based on the Dice coefficient. The results indicate that the adaption can be successfully applied to semi-supervised image segmentation.

We have performed experiments with different numbers of training scans. As expected, a greater number of training scans leads to more accurate segmentation for all the methods considered in the experiment. In addition, when the number of training scans is small, the SSL-based approaches DeepMedic-EN and DeepMedic-UDA perform even worse than the baseline DeepMedic model.

It is possibly due to the small number of training scans, which cannot adequately represent the distribution of desired features and segmentation. Thus, the adversarial learning strategy in DeepMedic-EN and DeepMedic-UDA cannot enforce proper regularization based on the unannotated data, where it is very likely that the segmentation or the extracted feature of unannotated data does not resemble that of the annotated data. As the number of training scans increases, the margin between the baseline DeepMedic and DeepMedic-EN or DeepMedic-UDA becomes smaller, possibly because the annotated data can better represent the distribution of expected features and segmentation. With 30 training scans, DeepMedic-UDA is able to outperform the baseline DeepMedic.

Unlike DeepMedic-EN and DeepMedic-UDA, the proposed approach relies on the assumption that similar inputs should produce consistent outputs, and the use of such consistency is further improved with a self-ensembling strategy. In contrast to the adversarial learning strategies in DeepMedic-EN and DeepMedic-UDA, the adapted MT model in the proposed work is less affected by the limited number of training scans, because it does not require the comparison between annotated data and unannotated data. This is confirmed by the results, where the propose approach is robust to the decrease in the number of training scans.

We have also observed that with only 10 annotated training scans, the proposed method outperforms the baseline DeepMedic model trained by 20 annotated scans and performs comparably to the baseline DeepMedic model trained by 30 annotated scans. This highlights the importance of the incorporation of unannotated data for training CNNs. It can potentially greatly reduce the annotation cost or increase the segmentation quality with existing annotated data.

We applied Gaussian noise to the input samples to generate a pair of similar inputs for the student and teacher models. Other strategies for generating pairs of similar inputs are possible. For example, dropout provides a convenient way for noise injection [13]. Future works may explore additional approaches to enforcing the consistency regularization to more efficiently use unannotated data.

5   Conclusion

We have proposed an SSL-based approach to brain lesion segmentation. A teacher model and a student model are constructed and updated alternately. By minimizing the segmentation loss computed from annotated data and segmentation consistency loss computed from unannotated data, the student model learns from the teacher model at each step. The teacher model is then updated with an EMA strategy, and the final teacher model performs lesion segmentation on test samples. The proposed method was applied to ischemic stroke lesion segmentation, and the results demonstrate the benefit of incorporating unannotated data using the proposed method.

References

1. Baur, C., Albarqouni, S., Navab, N.: Semi-supervised deep learning for fully convolutional networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 311–319. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_36

2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49

3. Garyfallidis, E., et al.: Dipy, a library for the analysis of diffusion MRI data. Front. Neuroinformatics **8**(8), 1–17 (2014)

4. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)

5. Kamnitsas, K., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 597–609. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_47

6. 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)

7. Kuang, H., Najm, M., Menon, B.K., Qiu, W.: Joint segmentation of intracerebral hemorrhage and infarct from non-contrast CT images of post-treatment acute ischemic stroke patients. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 681–688. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_78

8. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. In: International Conference on Learning Representations (2016)

9. Maier, O., et al.: ISLES 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. **35**, 250–269 (2017)

10. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging **35**(5), 1240–1251 (2016)

11. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems, pp. 1195–1204 (2017)

12. Tieleman, T., Hinton, G.: Lecture 6.5-RMSProp, coursera: neural networks for machine learning. University of Toronto, Technical Report (2012)

13. Wager, S., Wang, S., Liang, P.S.: Dropout training as adaptive regularization. In: Advances in Neural Information Processing Systems, pp. 351–359 (2013)

14. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 408–416. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_47

15. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., Fan, Y.: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med. Image Anal. **43**, 98–111 (2018)

16. Zhou, Z.H.: A brief introduction to weakly supervised learning. Nat. Sci. Rev. **5**(1), 44–53 (2017)