Automatic Brain Tumor Segmentation using Convolutional Neural Networks with Test-Time Augmentation

Guotai Wang[1,2], Wenqi Li[1,2], Sébastien Ourselin[1], and Tom Vercauteren[1,2]

[1]School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[2]Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK
guotai.wang@kcl.ac.uk

Abstract. Automatic brain tumor segmentation plays an important role for diagnosis, surgical planning and treatment assessment of brain tumors. Deep convolutional neural networks (CNNs) have been widely used for this task. Due to the relatively small data set for training, data augmentation at training time has been commonly used for better performance of CNNs. Recent works also demonstrated the usefulness of data augmentation at test time, in addition to training time, for achieving more robust predictions. We investigate how test-time augmentation can improve CNNs' performance for brain tumor segmentation. We used different underpinning network structures and augmented the image by 3D rotation, flipping, scaling and adding random noise at both training and test time. Experiments with BraTS 2018 training and validation set show that test-time augmentation can achieve higher segmentation accuracy and obtain uncertainty estimation of the segmentation results.

Keywords: Brain tumor, convolutional neural network, segmentation, data augmentation

1 Introduction

Gliomas are the most common primary brain tumors that start in the glial cells of the brain in adults. They can be categorized according to their grade: Low-Grade Gliomas (LGG) exhibit benign tendencies and portend a better prognosis for the patient, while High-Grade Gliomas (HGG) are malignant and lead to a worse prognosis [22]. Medical imaging of brain tumors plays an important role for evaluating the progression of the disease before and after treament. Currently the most widely used imaging modality for brain tumors is Magnetic Resonance Imaging (MRI) with different sequences, such as T1-weighted, contrast enhanced T1-weighted (T1ce), T2-weighted and Fluid Attenuation Inversion Recovery (FLAIR) images. These sequences provide complementary information for different subregions of brain tumors [24]. For example, the tumor region and peritumoral edema can be highlighted in FLAIR and T2 images, and the tumor core region without peritumoral edema is more visible in T1 and T1ce images.

Automatic segmentation of brain tumors and substructures from medical images has a potential for accurate and reproducible delineation of the tumors, which can help more efficient and better diagnosis, surgical planning and treatment assessment of brain tumors [24,5]. However, accurate automatic segmentation of the brain tumors is a challenging task for several reasons. First, the boundary between brain tumor and normal tissues is often ambiguous due to the smooth intensity gradients, partial volume effects, and bias field artifacts. Second, the brain tumors vary largely across patients in terms of size, shape, and localization. This prohibits the use of strong priors on shape and localization that are commonly used for robust segmentation of many other anatomical structures, such as the heart [12] and the liver [30].

In recent years, deep Convolutional Neural Networks (CNNs) have achieved the state-of-the-art performance for multi-modal brain tumor segmentation [28,16]. As a type of machine learning approach, they require a set of annotated training images for learning. Compared with traditional machine learning approaches they do not rely on hand-crafted features and can learn features automatically. In [13], a CNN was proposed to exploit both local and global features for robust brain tumor segmentation. It replaces the final fully connected layer used in traditional CNNs with a convolutional implementation that obtains 40 fold speed up. This approach employs a two-phase training procedure and a cascade architecture to tackle difficulties related to the imbalance of tumor labels. Despite the better performance than traditional methods, this approach works on individual 2D slices without considering 3D contextual information. DeepMedic [17] uses a dual pathway 3D CNN with 11 layers to make use of multi-scale features for brain tumor segmentation. For post-processing, it uses a 3D fully connected Conditional Random Field (CRF) [20] that helps to remove false positives. DeepMedic achieved better performance than using 2D CNNs. However, it works on local image patches and therefore has a relatively low inference efficiency. In [28], a triple cascaded framework was proposed for brain tumor segmentation. The framework uses three networks to hierarchically segment whole tumor, tumor core and enhancing tumor core sequentially. It uses a network structure with anisotropic convolution to deal with 3D images, taking advantage of dilated convolution [31], residual connection [7] and multi-scale fusion [29]. It demonstrated an advantageous trade-off between receptive field, model complexity and memory consumption. This method also fuses the output of CNNs in three orthogonal views for more robust segmentation of brain tumors. In [16], an ensemble of multiple models and architectures including DeepMedic [17], 3D Fully Convolutional Networks (FCN) [21] and U-Net [26,2] was used for robust brain tumor segmentation. The ensemble method reduces the influence of the meta-parameters of individual CNN models and the risk of overfitting the configuration to a specific training dataset. However, it requires much more computational resources to train and run a set of models.

Training with a large dataset plays an important role for the good performance of deep CNNs. For medical images, collecting a very large training set is usually time-consuming and challenging. Therefore, many works have used data augmentation to partially compensate this problem. Data augmentation applies transformations to the samples in a training set to create new ones, so that a relatively small training set can be enlarged to a larger one. Previous works have used different types of transformations such as flipping, cropping, rotation and scaling training images [2]. In [32], a simple and data-agnostic data augmentation routine termed *mixup* was proposed for training neural networks. Recently, several studies have empirically found that the performance of deep learning-based image recognition methods can be improved by combining predictions of multiple transformed versions of a test image, such as in pulmonary nodule detection [15] and skin lesion classification [23]. In [14], test images were augmented by mirroring for brain tumor segmentation. In [27], a mathematical formulation was proposed for test-time augmentation, where a distribution of the prediction was estimated by Monte Carlo simulation with prior distributions of parameters in an image acquisition model. That work also proposed a test-time augmentation-based *aleatoric* uncertainty estimation method that can help to reduce overconfident predictions. The framework in [27] has been validated with binary segmentation tasks, while its application to multi-class segmentation has yet to be demonstrated.

In this paper, we extend the work of [28] and [27], and apply test-time augmentation to automatic multi-class brain tumor segmentation. For a given input image, instead of obtaining a single inference, we augment the input image with different transformation parameters to obtain multiple predictions from the input, with the same network and associated trained weights. The multiple predictions help to obtain more robust inference of a given image. We explore the use of different CNNs as the underpinning network structures. Experiments with BraTS 2018 training and validation set showed that an improvement of segmentation accuracy was achieved by test-time augmentation, and our method can provide uncertainty estimation for the segmentation output.

2 Methods

2.1 Network Structures

We explore three network configurations as underpinning CNNs for the brain tumor segmentation task: 1) 3D UNet [2], 2) the cascaded networks in [28] where a WNet, TNet and ENet was used to segment whole tumor, tumor core and enhancing tumor core respectively, and 3) adapting WNet [28] for one-pass multi-class prediction without using cascaded prediction, which is referred to as multi-class WNet.

The 3D U-Net has a downsampling and an upsampling path each with four resolution steps. In the downsampling path, each layer has two $3 \times 3 \times 3$ convolutions each followed by a Rectified Linear Unit (ReLU) activation function, and then a $2 \times 2 \times 2$ max pooling layer was used for downsampling. In the upsamping path, each layer uses a deconvolution with kernel size $2 \times 2 \times 2$, followed by two $3 \times 3 \times 3$ convolutions with ReLU. The network has shortcut connections between corresponding layers with the same resolution in the downsampling path and the upsampling path. In the last layer, a $1 \times 1 \times 1$ convolution is used to reduce the number of output channels to the number of segmentation labels, i.e., 4 for the brain tumor segmentation task in the BraTS challenge.

The WNet proposed in [28] is an anisotropic network that considers a trade-off between receptive field, model complexity and memory consumption. It employs dilated convolution [31], residual connection [7] and multi-scale prediction [29] to improve segmentation performance. The network uses 20 intra-slice convolution layers and four inter-slice convolution layers with two 2D down-sampling layers. Since the anisotropic convolution has a small receptive field in the through-plane direction, multi-view fusion was used to take advantage of the 3D contextual information, where the network was applied in axial, sagittal and coronal views respectively. For the multi-view fusion, the softmax outputs in these three views were averaged. In [28], WNet is used to segment the whole tumor. TNet for tumor core segmentation uses the same structure as WNet, and ENet for enhancing core segmentation is a variant of WNet that uses only one down-sampling layer. Compared with multi-label prediction, the cascaded networks require longer time for training and testing. To improve the training efficiency, we compare the cascaded networks [28] with the use of multi-class WNet, where a single WNet for multi-label prediction is employed without using TNet and ENet. Therefore, for this variant we change the output channel number from 2 to 4. Multi-view fusion is also used for this multi-class WNet.

2.2 Data Augmentation for Training and Testing

From the point view of image acquisition, an observed image is only one of many possible observations of the underlying anatomy that can be observed with different spatial transformations and noise. Direct inference with the observed image may lead to a biased result affected by the specific transformation and noise associated with that image. To obtain a more robust prediction, we consider different transformations and noise during the test time. Let $\boldsymbol{\beta}$ and $\boldsymbol{e}$ represent the parameters for spatial transformation and intensity noise respectively. We assume that $\boldsymbol{\beta}$ is a combination of f_l, r and s, where f_l is a random variable for flipping along each 3D axis, r is the rotation angle along each 3D axis, s is a scaling factor. We consider these parameters following some prior distributions: $f_l \sim Bern(0.5)$, $r \sim U(0, 2\pi)$, $s \sim U(0.8, 1.2)$. For the intensity noise, we assume $\boldsymbol{e} \sim N(0, 0.05)$ according to the reduced standard deviation of a median-filtered version of a normalized image [27].

For data augmentation, we randomly sample $\boldsymbol{\beta}$ and $\boldsymbol{e}$ from the above prior distributions and use them to transform the image. We use the same distributions of augmentation parameters at both training and test time for a given CNN. For test-time augmentation, we obtain N samples from the distributions of $\boldsymbol{\beta}$ and $\boldsymbol{e}$ by Monte Carlo simulation, and the resulting transformed version of the input was fed into the CNN. The N prediction results were combined to obtain the final prediction based on majority voting.

2.3 Uncertainty Estimation

Both model-based (*epistemic*) uncertainty and image-based (*aleatoric*) uncertainty have been investigated for deep CNNs in recent years [18]. The *epistemic* uncertainty is often obtained by Bayesian approximation-based methods such as test-time dropout [10]. In [27], test-time augmentation was used to estimate the *aleatoric* uncertainty of segmentation results in a consistent mathematical framework. In this paper, we use test-time augmentation to obtain segmentation results as well as the associated *aleatoric* uncertainty according to [27].

The uncertainty estimation is obtained by measuring the diversity of the predictions for a given image. Both the variance and entropy of the distribution can be used to estimate uncertainty. Since variance is not sufficiently representative in the context of multi-modal distributions, we use entropy for the pixel-wise uncertainty estimation desired for segmentation tasks. Let X denote the input image and Y denote the output segmentation. We use Y^i to denote the predicted label for the i-th pixel. With the Monte Carlo simulation described in Section 2.2, a set of values for Y^i are obtained $\mathcal{Y}^i = \{y_1^i, y_2^i, ..., y_N^i\}$. The entropy of the distribution of Y^i is therefore approximated as:

$$H(Y^i|X) \approx -\sum_{m=1}^{M} \hat{p}_m^i \ln(\hat{p}_m^i) \tag{1}$$

where $\hat{p}_m^i$ is the frequency of the m-th unique value in $\mathcal{Y}^i$.

3 Experiments and Results

Data and Implementation Details. We used the BraTS 2018[1] [24,3,4,5,6] dataset for experiments. The training set contains images from 285 patients, including 210 cases of HGG and 75 cases of LGG. The BraTS 2018 validation and testing set contain images from 66 and 191 patients with brain tumors of unknown grade, respectively. Each patient was scanned with four sequences: T1, T1ce, T2 and FLAIR. As a pre-processing performed by the organizers, all the images were skull-striped and re-sampled to an isotropic 1mm^3 resolution, and the four modalities of the same patient had been co-registered. The ground truth were provided by the BraTS organizers. We uploaded the segmentation results obtained by our method to the BraTS 2018 server, and the server provided quantitative evaluations including Dice score and Hausdorff distance compared with the ground truth.

[1] `http://www.med.upenn.edu/sbia/brats2018.html`

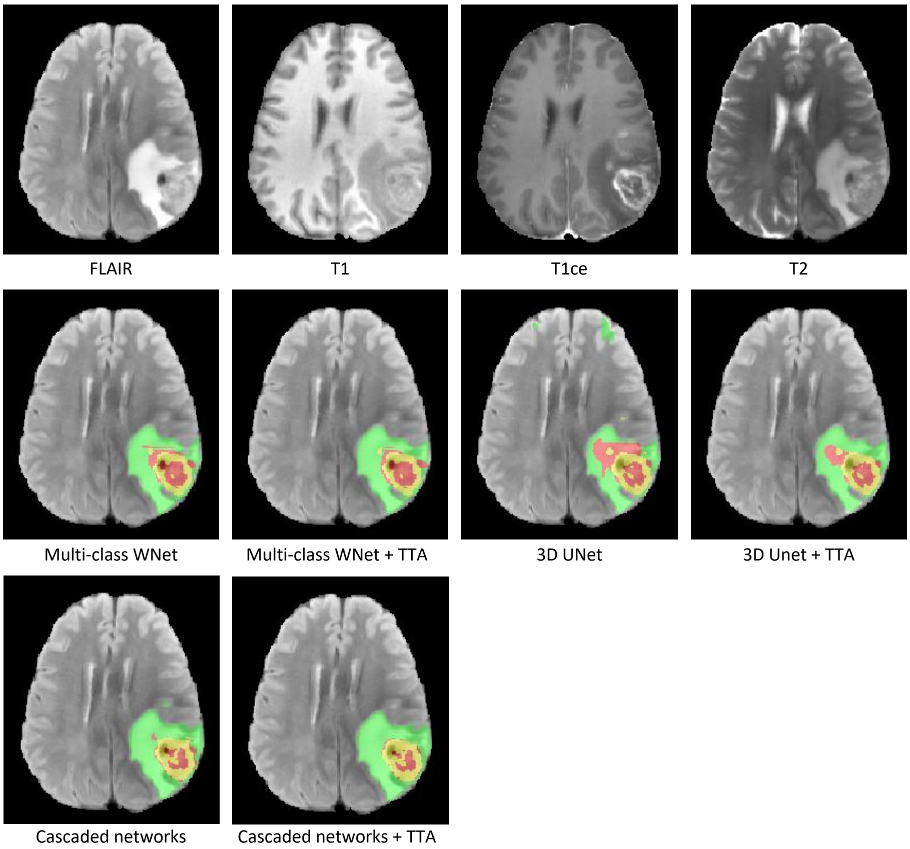

Fig. 1. An example of brain tumor segmentation results obtained by different networks and test-time augmentation (TTA). The first row shows the four modalities of the same patient. The second and third rows show segmentation results. Green: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core.

We implemented the 3D UNet [2], multi-class WNet and cascaded networks [28] in Tensorflow[2] [1] using NiftyNet[34] [11]. The Adaptive Moment Estimation (Adam) [19] strategy was used for training, with initial learning rate 10^{-3}, weight decay 10^{-7}, and maximal iteration 20k. The training patch size was 96×96×96 for 3D UNet and 96×96×19 for multi-class WNet. The batch size was 2 and 4 for these two networks respectively. For the cascaded networks, we followed the configurations in [28]. The training process was implemented on an NVIDIA TITAN X GPU. As a pre-processing, each image was normalized by the mean value and standard deviation. The Dice loss function [25,9] was used for training.

[2] https://www.tensorflow.org

[3] http://niftynet.io

[4] https://github.com/taigw/brats18

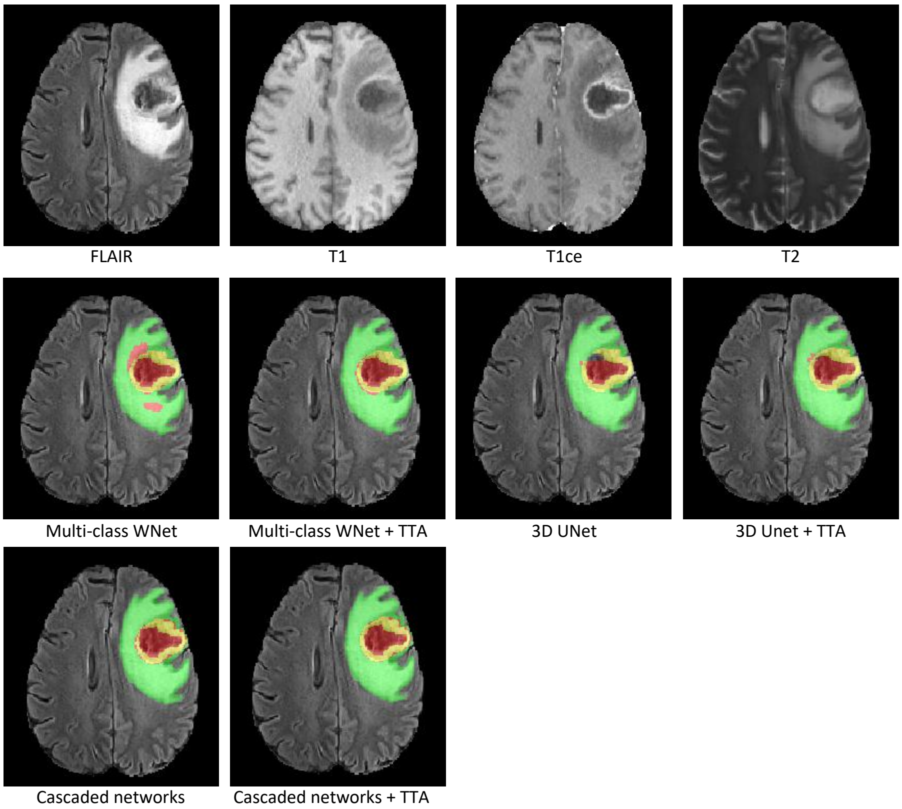

Fig. 2. Another example of brain tumor segmentation results obtained by different networks and test-time augmentation (TTA). The first row shows the four modalities of the same patient. The second and third rows show segmentation results. Green: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core.

At test time, the augmented prediction number was set to $N = 20$ for all the network structures. The multi-class WNet and cascaded networks were trained in axial, sagittal and coronal views respectively, and the predictions in these three views were fused by averaging at test time.

Segmentation Results. Fig. 1 shows an example from the BraTS 2018 validation set. The first row shows the input images of four modalities: FLAIR, T1, T1ce and T2. The second and third rows present the segmentation results of 3D UNet, multi-class WNet, cascaded networks and their corresponding results with test-time augmentation. It can be observed that the initial output of the 3D UNet seems to be noisy with some false positives of edema and non-enhancing tumor core. After using test-time augmentation, the result becomes more spatially consistent. The output of multi-class WNet also seems to be noisy for

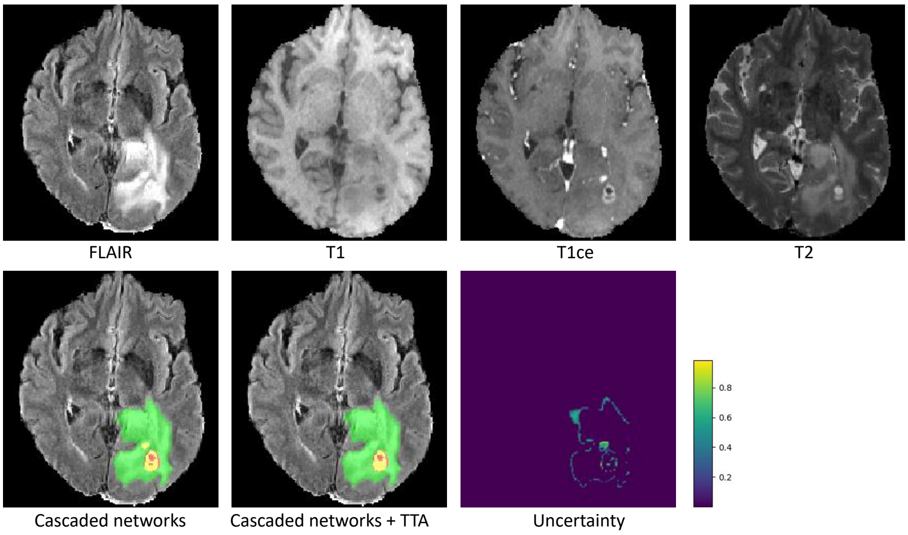

Fig. 3. An example of segmentation result and uncertainty estimation obtained by cascaded networks [28] with test-time augmentation.

the non-enhancing tumor core. A smoother segmentation is obtained by multi-class WNet with test-time augmentation. For the cascaded networks, test-time augmentation also leads to visually better resutls of the tumor core.

Fig. 2 shows another example from the BraTS 2018 validation set. It can be observed that the 3D UNet obtains a hole in the tumor core, which seems to be an under-segmentation. The hole is filled after using test-time augmentation and the result looks more consistent with the input images. The initial prediction by multi-class WNet seems to have an over segmentation of the non-enhancing tumor core. After using test-time augmentation, the over-segmented regions become smaller, leading to higher accuracy. Test-time augmentation also helps to improve the result of cascaded networks. Fig. 3 shows a case from the BraTS 2018 testing set, where test-time augmentation obtains a better spatial consistency for the tumor core. In addition, it leads to an uncertainty estimation of the segmentation output. It can be observed that most uncertain results focus on the border of the tumor and some potentially mis-segmented regions.

A quantitative evaluation of our different methods on the BraTS 2018 validation set is shown in Table 1. The initial output of 3D UNet achieved Dice scores of 73.44%, 86.38% and 76.58% for enhancing tumor core, whole tumor and tumor core respectively. 3D UNet with test-time augmentation achieved a better performance than the baseline of 3D UNet, leading to Dice scores of 75.43%, 87.31% and 78.32% respectively. For the initial output of multi-class WNet, the Dice score was 75.70%, 88.98% and 72.53% for these three structures respectively. After using test-time augmentation, an improvement was achieved, and the Dice score was 77.70%, 89.56% and 73.04% for these three structures respectively. For the cascaded networks, test-time augmentation leads to higher accuracy for the enhancing tumor core and tumor core. Table 2 presents the performance of our cascaded networks with test-time augmentation on BraTS 2018 testing set. The average Dice scores for enhancing tumor core, whole tumor and tumor core are 74.66%, 87.78% and 79.64%, respectively. The corresponding values of Hausdorff distance are 4.16mm, 5.97mm and 6.71mm, respectively.

Table 1. Mean values of Dice and Hausdorff measurements of different methods on BraTS 2018 validation set. ET, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively. TTA: test-time augmentation.

	Dice (%)			Hausdorff (mm)		
	ET	WT	TC	ET	WT	TC
3D UNet	73.44	86.38	76.58	9.37	12.00	10.37
3D UNet + TTA	75.43	87.31	78.32	4.53	5.90	8.03
Multi-class WNet	75.70	88.98	72.53	4.24	4.99	12.13
Multi-class WNet + TTA	77.07	89.56	73.04	4.44	4.92	11.13
Cascaded networks	79.19	90.31	85.40	3.34	5.38	6.61
Cascaded networks + TTA	79.72	90.21	85.83	3.13	6.18	6.37

Table 2. Dice and Hausdorff measurements of our cascaded networks with test-time augmentation on BraTS 2018 testing set. ET, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively.

	Dice (%)			Hausdorff (mm)		
	ET	WT	TC	ET	WT	TC
Mean	74.66	87.78	79.64	4.16	5.97	6.71
Standard deviation	25.85	11.92	24.97	7.07	8.56	10.27
Median	83.38	91.33	89.68	2.00	3.32	3.16
25 Quantile	72.87	86.69	78.24	1.41	2.24	2.00
75 Quantile	88.64	94.09	93.58	3.00	5.48	6.40

4 Discussion and Conclusion

For test-time augmentation, we only used flipping, rotation and scaling for spatial transformations. It is also possible to employ more complex transformations such as elastic deformations used in [2]. However, such deformations take longer time for testing and have a lower efficiency. The results show that test-time augmentation leads to an improvement of segmentation accuracy for different CNNs including 3D UNet [2], multi-class WNet and cascaded networks [28]. Test-time augmentation can be applied to other CNN models as well. The uncertainty estimation obtained by our method can be used for downstream analysis such as uncertainty-aware volume measurement [8] and guiding user interactions [29]. It would be of interest to assess the impact of test-time augmentation on CNNs trained with state-of-the-art policies such as in [14]. By using test-time augmentation, we investigated the test image-based (*aleatoic*) uncertainty for brain tumor segmentation. It is of interest to investigate how ensemble of CNNs [16] can produce *epistemic* uncertainty for this task. For a comprehensive study of uncertainty, it is promising to combine ensemble of models or test-time dropout with test-time augmentation. This will be left for future work.

In conclusion, we explored the effect of test-time augmentation on CNN-based brain tumor segmentation. We used 3D U-Net, 2.5D multi-class WNet and cascaded networks as the underpinning network structures. For training and testing, we augmented the image by 3D rotation, flipping, scaling and adding random noise. Experiments with BraTS 2018 training and validation set show that test-time augmentation helps to improve the brain tumor segmentation accuracy for different CNN structures and obtain uncertainty estimation of the segmentation results.

Acknowledgements. We would like to thank the NiftyNet team. This work was supported through an Innovative Engineering for Health award by the Wellcome Trust [WT101957, WT97914, 203145Z/16/Z, 203148/Z/16/Z], Engineering and Physical Sciences Research Council (EPSRC) [NS/A000027/1, NS/A000049/1, NS/A000050/1], the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative), hardware donated by NVIDIA, and the Health Innovation Challenge Fund [HICF-T4-275].

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., Brain, G.: TensorFlow: A system for large-scale machine learning. In: USENIX Symposium on Operating Systems Design and Implementation. pp. 265–284 (2016)
2. Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432 (2016)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)

5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data p. 170117 (2017)
6. Bakas, S., Reyes, M., et Int, Menze, B.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge (2018), `https://arxiv.org/abs/1811.02629`
7. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage 170, 446 – 455 (2018)
8. Eaton-Rosen, Z., Bragman, F., Bisdas, S., Ourselin, S., Cardoso, M.J.: Towards safe deep learning: Accurately quantifying biomarker uncertainty in neural network predictions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 691–699 (2018)
9. Fidon, L., Li, W., Garcia-peraza herrera, L.C.: Generalised Wasserstein Dice score for imbalanced multi-class segmentation using holistic convolutional networks. arXiv preprint arXiv:1707.00478 (2017)
10. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning. pp. 1050–1059 (2016)
11. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T.: NiftyNet: A deep-learning platform for medical imaging. Computer Methods and Programs in Biomedicine 158, 113–122 (2018)
12. Grosgeorge, D., Petitjean, C., Dacher, J.N., Ruan, S.: Graph cut segmentation with a statistical shape model in cardiac MRI. Computer Vision and Image Understanding 117(9), 1027–1035 (2013)
13. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Medical Image Analysis 35, 18–31 (2016)
14. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No newnet. arXiv preprint arXiv:1809.10483 (2018)
15. Jin, H., Li, Z., Tong, R., Lin, L.: A deep 3D residual CNN for false positive reduction in pulmonary nodule detection. Medical Physics 45(5), 2097–2107 (2018)
16. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 450–462. Springer International Publishing (2018)
17. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, 61–78 (2017)
18. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems. pp. 5580–5590 (2017)
19. Kingma, D.P., Ba, J.L.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015)

20. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. pp. 109–117. NIPS'11, Curran Associates Inc., USA (2011)
21. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
22. Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W.: The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathologica 131(6), 803–820 (Jun 2016)
23. Matsunaga, K., Hamada, A., Minagawa, A., Koga, H.: Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. arXiv preprint arXiv:1703.03108 (2017)
24. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993–2024 (2015)
25. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision. pp. 565–571 (2016)
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241 (2015)
27. Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T.: Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. arXiv preprint arXiv:1807.07356 (2018)
28. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 178–190. Springer International Publishing (2018)
29. Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: Interactive medical image segmentation using deep learning with image-specific fine-tuning. IEEE Transactions on Medical Imaging 37(7), 1562–1573 (2018)
30. Wang, G., Zhang, S., Xie, H., Metaxas, D.N., Gu, L.: A homotopy-based sparse representation for fast and accurate shape prior modeling in liver surgical planning. Medical Image Analysis 19(1), 176–186 (2015)
31. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. CoRR abs/1511.07122 (2015)
32. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: Mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 pp. 1–11 (2017)