# DeepMedic for Brain Tumor Segmentation

Konstantinos Kamnitsas[1,2⋆], Enzo Ferrante[1], Sarah Parisot[1], Christian Ledig[1], Aditya Nori[2], Antonio Criminisi[2], Daniel Rueckert[1], and Ben Glocker[1]

[1] Biomedical Image Analysis Group, Imperial College London, UK
[2] Microsoft Research, Cambridge, UK

**Abstract.** Accurate automatic algorithms for the segmentation of brain tumours have the potential of improving disease diagnosis, treatment planning, as well as enabling large-scale studies of the pathology. In this work we employ DeepMedic [1], a 3D CNN architecture previously presented for lesion segmentation, which we further improve by adding residual connections. We also present a series of experiments on the BRATS 2015 training database for evaluating the robustness of the network when less training data are available or less filters are used, aiming to shed some light on requirements for employing such a system. Our method was further benchmarked on the BRATS 2016 Challenge, where it achieved very good performance despite the simplicity of the pipeline.

## 1 Introduction

Accurate estimation of the relative volume of the subcomponents of a brain tumour is critical for monitoring progression, radiotherapy planning, outcome assessment and follow-up studies. For this, accurate delineation of the tumour is required. Manual segmentation poses significant challenges for human experts, both because of the variability of tumour appearance but also because of the need to consult multiple images from different MRI sequences in order to classify tissue type correctly. This laborious effort is not only time consuming but prone to human error and results in significant intra- and inter-rater variability [2].

Automatic segmentation systems aim to provide a cheap and scalable solution. Over the years, automatic methods for brain tumour segmentation have attracted significant attention. Representative early work is the atlas-based outlier detection method [3]. Segmentation was later solved jointly with the registration of a healthy atlas to a pathological brain [4], making use of a tumour growth model and the Expectation Maximization algorithm. In [5] the problem was tackled as the joint optimization of two Markov Random Fields (MRFs). The state of the art was raised further by supervised learning methods, initially represented mainly by Random Forests, coupled with models such as Gaussian Mixtures for the extraction of tissue type priors [6], MRFs for spatial regularisation and a variety of engineered features [7].

⋆ Part of this work was carried on when KK was an intern at Microsoft Research. Email correspondence to: konstantinos.kamnitsas12@imperial.ac.uk

Recent years saw the success of deep learning, with the methods in [8] and [9] being the top performing automatic approaches in BRATS 2014 and 2015 [10], using 3D and 2D Convolutional Neural Networks (CNNs) respectively. The latter approached the accuracy of the winning semi-automatic method [11]. The fact that the employed models are rather simple in design reveals the high potential of CNNs. The method presented in [12] also exhibited good performance, based on a 3-layers deep 2D network that separately processes each axial slice. The authors empirically showed that the class bias introduced to a network when training with patches extracted equiprobably from the task's classes can be partially alleviated with a second training stage using patches uniformly extracted from the image. In [13] an ensemble of 2D networks is used to process three orthogonal slices of a brain MR image. Finally, in [1] we showed that multi-scale 3D CNNs of larger size can accomplish high performance while remaining computationally efficient. In that work we also analysed how the size of the input segments relates to the captured class distribution by the training samples. It was shown that this meta-parameter can be exploited for capturing a partially adaptive distribution of training samples that in practice leads to good performance in a variety of segmentation tasks. Our segmentation system exhibited excellent performance on stroke lesion segmentation, winning the first position in the SISS-ISLES 2015 challenge [14,15], brain tumours and traumatic brain injuries [1]. It's generic architecture and processing of 3D content also allow its use on diverse problems, such as the segmentation of the placenta from motion corrupted MR images [16], where it achieved very promising results.

In this work we further extend our network, the DeepMedic [1], with residual connections [20] and evaluate their effect. We then investigate the behaviour of our system when trained with less data or when its capacity is reduced to explore requirements for employing such a segmentation method. Finally, we discuss the performance of our method on the recent BRATS 2016 challenge where it was further benchmarked.

## 2 Method

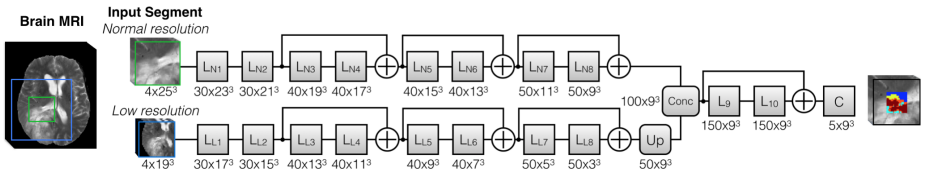


Fig. 1: The DeepMedic [1] extended with residual connections. The operations within each layer block are applied in the order: Batch-Normalization [17], non-linearity and convolution. [18] empirically showed this format leads to better performance. *Up* and *C* represent an upsampling and classification layer respectively. Number of filters and their size depicted as (*Number* × *Size*). Other hyper-parameters as in [1].

DeepMedic is the 11-layers deep, multi-scale 3D CNN we presented in [1] for brain lesion segmentation. The architecture consists of two parallel convolutional pathways that process the input at multiple scales to achieve a large receptive field for the final classification while keeping the computational cost low. Inspired by VGG [19], the use of small convolutional kernels is adopted. This design choice was shown [19] to be very effective in building deeper CNNs without severely increasing the number of trainable parameters, and we showed it allows building high performing yet efficient 3D CNNs thanks to the much smaller computation required for the convolution with small $3^3$ kernels [1]. The CNN is employed in a fully convolutional fashion on image segments in both training and testing stage[4].

We extend the DeepMedic with residual connections in order to examine their effect on segmentation. Residual connections were recently shown to facilitate preservation of the flowing signal and as such have enabled training of very deep neural networks [20,18]. In [20] the authors did not observe a performance improvement when a 18-layers deep network was employed, but only in experiments with architectures deeper than 34-layers. The networks employed on biomedical applications tend to consist of less layers than modern architectures in computer vision. However, the problem of preserving the forward and backwards propagated signal as well as the difficulty of optimization can be substantial in 3D CNNs due to the large number of trainable parameters in 3D kernels, as previously discussed in [1]. For this reason we set off to investigate such an architecture.

We extended the network by adding residual connections between the outputs of every two layers, except for the first two of each pathway to enforce abstracting away from raw intensity values. The architecture is depicted in Fig. 1. In our work, a residual connection after layer $l$ performs the element-wise addition $\oplus$ between corresponding channels of the output $out_l$ of layer $l$ and the input $in_{l-1}$ to the previous layer. This choice follows the investigation in [18] that found identity mappings on the residual path to perform best. More formally:

$$in_{l+1}^m = \begin{cases} out_l^m \oplus \hat{in}_{l-1}^m, & \text{if } m \leq M^{l-1} \\ out_l^m, & \text{otherwise} \end{cases} \tag{1}$$

where $l$ any layer after which a residual connection is added, the superscript m denotes the $m$-th channel and $M^l$ is the number of feature maps in the $l$-th layer. $\hat{in}_l$ is the input of the previous layer after padding in the (x,y,z) dimensions with reflection in order to match the dimensions of $out_l$.

[4] Code publicly available at: `https://github.com/Kamnitsask/deepmedic`

# 3 Evaluation

## 3.1 Data

The training database of BRATS 2015[5] (common with BRATS 2016) includes 220 multi-modal scans of patients with high (HGG) and 54 with low grade glioma (LGG). Scans include pre- and post-operative scans. T1-weighted, contrast enhanced T1c, T2-weighted and FLAIR sequences are available. The images were registered to a common space, resampled to isotropic 1 mm $\times$ 1 mm $\times$ 1 mm resolution with image dimensions 240 $\times$ 240 $\times$ 155 and were skull stripped by the organisers. Annotations are provided that include four labels: 1) necrotic core (NC), 2) oedema (OE), 3) non-enhancing (NE) and 4) enhancing core (EC). The annotations for the training database were obtained semi-automatically, fusing the predictions of multiple automatic algorithms, followed by expert review. The official evaluation is performed by merging the predicted labels into three sets: whole tumor (all 4 labels), core (1,3,4) and enhancing tumor (4).

The testing database of BRATS 2016 consists of 191 datasets. They are scans of 94 subjects, with 1-3 time points, including both pre- and post-operative scans. The scans were acquired in multiple clinical centers, some of which are distinct from those centers that provided the data for the training database. MRI modalities are the same as the training database. Ground truth annotations have been made manually by experts but were kept private for the evaluation. The MRI images have been preprocessed similarly to the training data and then provided to the participating teams. Interesting to note is that skull stripping had significant flaws in many cases, leaving behind portions of skull and extra-cerebral tissue at a significantly greater extent that what is observed in the training database. Such heterogeneity between testing and training data poses a significant challenge for automated machine learning methods that try to model the distribution of the training data. Yet they reflect a realistic scenario and an interesting benchmark for fully automated systems, which ideally should perform adequately even after inaccuracies introduced by individual blocks in the processing pipeline.

## 3.2 Evaluation on the BRATS 2015 Training Database

**Preprocessing and Augmentation:** Each scan was further individually normalized by subtracting the mean and dividing by the standard deviation of the intensities within the brain. Training data were augmented via reflection with respect to the mid-sagittal plane.

**Effect of Residual Connections:** To evaluate the effect of the residual connections we performed 5-fold cross validation on the mixed HGG and LGG data, while ensuring that all pre- and post-operative scans from a subject would only appear in training or validation of a fold. First we reproduced results similar to what was reported in [1] for the original version of DeepMedic. The extension

[5] links: http://braintumorsegmentation.org/

with residual connections gave a modest but consistent improvement over all classes of the task, as shown in Table 1. Important is that performance increases even on small challenging substructures like the necrosis and non-enhancing tumor, which may not be individually evaluated for the challenge but is interesting from an optimization perspective. The improvement seems mainly due to an increase in sensitivity, however at the cost of a lower precision. This is a positive side-effect as in practice it can prove easier to clear false positives in a post-processing step, for instance with a Conditional Random Field as performed in [1,14], rather than capturing areas previously missed by the CNN.

Table 1: Performance of the original *DeepMedic* (DM) and its extension with residual connections *DMRes*, evaluated with a 5-fold validation over the whole BRATS 2015 training database. For consistency with the online evaluation platform, cases that do not present enhancing tumor in the provided annotations are considered zeros for the calculation of the average, thus lowering the upper bound of accuracy for the class.

| | DICE | | | Precision | | | Sensitivity | | | DICE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Whole | Core | Enh. | Whole | Core | Enh. | Whole | Core | Enh. | NC | OE | NE | EC |
| DeepMedic | 89.6 | 75.4 | 71.8 | 89.7 | 84.5 | 74.3 | 90.3 | 73.0 | 73.0 | 38.7 | 78.0 | 36.7 | 71.8 |
| DMRes | 89.6 | 76.3 | 72.4 | 87.6 | 82.4 | 72.5 | 92.2 | 75.4 | 76.3 | 39.6 | 78.1 | 38.1 | 72.4 |

**Behaviour of the network with less training data and filters:** CNNs have shown promising accuracy when trained either on the extensive database of BRATS 2015 or on the rather limited of BRATS 2013. However, qualitative differences of the two databases do not allow estimating the influence of the database's size. Additionally, although various architectures were previously suggested, no work has investigated the required capacity of a network for the task. This factor is significant in practice, as it defines the computational resources and inference time required.

We evaluate the behaviour of our network on the tumour segmentation task with respect to the two above factors. To avoid bias towards subjects scanned at more than one time-point, only the earliest dataset was used from each subject. Out of the 198 remaining datasets, we randomly chose 40 (29 HGG, 11 LGG) as a validation fold. We then trained the original version of DeepMedic on the remaining 158 datasets, as well as on a reduced number of training data. Finally, versions of the network where all layers had their filters reduced to 50% and 33% percent were trained on the whole training fold.

It can be observed on Table 2 that although accuracy is negatively affected, the network still retains most of its performance for the three merged classes of the challenge even when trained with little data or its capacity is significantly reduced. A more thorough look in the accuracy achieved for the 4 non-merged classes of the task shows that the greatest decline is observed for the challenging necrotic and non-enhancing classes, which however does not influence the seg-

mentation of the overall core as severely. These experiments indicate that both training data and a large number of network filters to learn fine and detailed patterns are important for the segmentation of small and challenging structures[6].

Table 2: Exploring the performance of *DeepMedic* with reduced training data or number of filters at each layer. Note that the difference of the entry *DeepMedic* in comparison to Table 1 is due to the use of a subset of data (see text). Red color indicates reduction greater than 1% DICE.

| | DICE | | | Precision | | | Sensitivity | | | DICE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Whole | Core | Enh. | Whole | Core | Enh. | Whole | Core | Enh. | NC | OE | NE | EC |
| DeepMedic | 91.4 | 83.1 | 79.4 | 89.2 | 87.7 | 82.8 | 94.1 | 80.8 | 79.5 | 50.0 | 79.6 | 35.1 | 79.4 |
| DM(75% data) | 91.2 | 82.5 | 79.6 | 89.0 | 84.4 | 82.4 | 93.9 | 80.4 | 79.7 | 45.9 | 79.0 | 35.1 | 79.6 |
| DM(50% data) | 91.4 | 82.6 | 78.8 | 91.0 | 85.3 | 81.7 | 92.3 | 82.3 | 78.5 | 44.7 | 79.2 | 36.8 | 78.8 |
| DM(33% data) | 90.5 | 79.7 | 77.7 | 90.6 | 86.5 | 82.8 | 91.0 | 77.1 | 77.1 | 45.8 | 77.9 | 31.8 | 77.7 |
| DM(20% data) | 89.8 | 80.5 | 77.6 | 91.1 | 83.9 | 81.8 | 89.7 | 80.5 | 76.5 | 41.3 | 76.9 | 34.1 | 77.6 |
| DM(50% filters) | 91.4 | 80.8 | 79.8 | 92.2 | 89.0 | 82.5 | 91.3 | 76.3 | 80.2 | 49.0 | 79.2 | 29.4 | 79.9 |
| DM(33% filters) | 90.8 | 81.7 | 79.5 | 90.0 | 91.9 | 78.2 | 92.1 | 76.6 | 83.0 | 44.4 | 79.3 | 27.9 | 79.4 |

### 3.3 Evaluation on the BRATS 2016 Testing Database

Our team participated in the BRATS 2016 Challenge in order to further benchmark our system. In the testing stage of the challenge, each team is given 48 hours after they are provided the data, to apply their systems and submit predicted segmentations.

**Preprocessing:** Similarly to the preprocessing of the training data, we normalized each image individually by subtracting the mean and dividing by the standard deviation of the intensities of the brain. Additionally, for the subjects that had multiple scans acquired at different time-points, since the brains have been co-registered, we tried to mitigate the problem of the failed brain extraction by fusing with majority voting the brain masks from different time-points, hoping that the errors are not consistent at all time-points. Unfortunately this step did not resolve the issue. We decided not to apply ourselves an additional brain extraction step, since this is not guaranteed to work on the already (well or partially) stripped data and would require case-by-case visual inspection and intervention, which is against our interest in a fully automatic pipeline.

**Network Configuration and Training:** Prior to the testing stage of the challenge we had trained three models with identical architecture as shown in Figure 1. They were trained on the whole BRATS 2015 Training database, without dinstinction of HGG from LGG cases, or pre- from post-operative scans. Training

[6] Although these experiments were performed with the original version of the network, we expect the trends to continue after the extension with residual connections.

of a single model required approximately a day when using an NVIDIA GTX Titan X GPU. Our network can then segment a multi-modal dataset in less than 35 seconds when using CuDNN v5.0. The probability maps of the three models were then fused by averaging before the post-processing.

**Post-processing with a 3D Fully Connected CRF:** In previous work [1,14] we had implemented and evaluated a 3D fully connected Conditional Random Field (CRF) [21] for the segmentation of multi-sequence volumetric scans. Our evaluation had shown that the model consistently offers beneficial regularisation in a variety of lesion segmentation tasks. The CRF was found to be particularly beneficial in cases where the main segmenter underperforms. We employ the CRF in the same fashion. We provide as the CRF's unary term the whole-tumor probabity map, constructed by merging the multi-class predictions of the CNN. This way the CRF regularizes the overall size of the tumor and clears small spurious false positives. The whole-tumor segmentation mask produced by the CRF is used to mask the segmentation produced by the CNNs, leaving the internal structure of the tumor mostly intact [7]. Finally, any left out connected-components smaller than 1000 voxels were removed.

**Results:** In the testing stage of the challenge 19 teams participated. To evaluate the quality of their segmentations, the teams were ranked according to the statistically significant differences in the achieved DICE scores and Haussdorf distances with respect to each other. Additionally, they were assessed and ranked for their capability in following shrinkage or growth of the tumor between different time-points. The exact results on the testing database have not been made public by the organizers yet.

Our system achieved a place among the top ranking methods, performing well on the DICE metric. The high performing methods were rather close in terms of the DICE metric for the segmentation of the Whole Tumor. Greater were the differences for the Tumor Core and Enhancing Tumor. Our system achieved the top rank for the segmentation of the tumor core. The position was shared with another CNN-based approach [22], the overall most accurate model in the challenge. Note that in order to generalize, this model was trained on external data from a private brain tumor database with manual annotations, in addition to the BRATS 2015 data, and thus direct comparison may not be completely fair. Notice that out of the three main tumor classes, the core was found to be the one most influenced by the amount of training data in our experiments (Table 2), which explains the high DICE for this class achieved by the competing method. Moreover, since the testing data appear to have a significantly different distribution from the training data, the influence of additional external data should be significantly stronger than what was found in our experiments, where we only explored the effect of the amount of data when training and testing distributions are the same. Our system also achieved the top rank for the segmentation of enhancing tumor in terms of the DICE measure. This position was shared with

[7] Code of the 3D CRF available at: `https://github.com/Kamnitsask/dense3dCrf/`

the systems presented in [23] and [24]. The former employed a well engineered cascade of Random Forests tailored for segmentation of brain tumors, trained on a selected high quality subset of the training data to avoid learning from errors within the semi-automatically generated annotations (Fig. 2). The latter applied a CNN and a CRF similar to ours, but with the two jointly trained. Interestingly, we achieved higher performance in terms of DICE for the core and enhancing tumor and comparable accuracy for the whole tumor in comparison to [25], who employed an extended version of GLISTRboost [11], the semi-automatic method that won the first place in the BRATS 2015 challenge.

Less satisfying was the Haussdorf distance achieved by our method, with respect to which it achieved average ranking. We speculate a reason for this is false segmentation of skull and extra-cerebral tissue, portions of which were observed in many of the testing cases where the provided brain extraction was inaccurate (Fig. 2). Such tissues were not present in the training database to this extent, so our system never learnt to classify them. The resulting false positives decrease the DICE metric, but influence even more the Haussdorf distance since they lie well outside the brain. With further careful brain extraction we could alleviate the problem but this would require case-by-case inspection as it is prone to fail on the already (successfully or not) stripped testing cases, thus render the pipeline not fully automatic. At the time of writing it has not been reported yet how other teams dealt with this issue. Finally, semi-automatic systems that rely on manual initialization, such as the competing method in [25], should be less prone to this issue.

Finally, our system performed very well in predicting the temporal change of the volumes of whole and enhancing tumor, with its average performance for these classes achieving the third position. Interestingly, the two overall winners of this category [24,26] employed a CRF similar to ours, which indicates the effectiveness of this model for this task.

## 4 Conclusion

This paper has investigated the effect of residual connections on a recently presented 3D CNN model on the task of brain tumour segmentation, where their incorporation led to a small but consistent improvement. Our work reveals that an 11-layers 3D CNN gains from such an extension, mostly thanks to increased sensitivity, unlike the observation in [20] where benefits were found only for significantly deeper 2D networks.

In an attempt to explore the generalization and efficiency of CNNs for a task such as brain tumor segmentation, we also investigated the behaviour of DeepMedic when trained with smaller number of data or when less filters are used. Our experiments show that segmentation accuracy for the whole, core and enhancing tumour, even though affected, it is not severely hindered by the two factors to an extent that would render the system impractical. However, they are very important for segmenting challenging, fine substructures such as necrosis and non-enhancing tumour. On the other hand, in applications where

segmentation of such substructures is not required, small networks can be a suitable option, thanks to lower computational requirements and shorter inference times (35s versus 8s per multi-modal scan for the original and smallest model in Table 2 respectively). Note that in our experiments we only explored the effect of the amount of data when training and testing distributions are the same. Networks, similarly to most machine learning algorithms, face generalization problems when the two differ significantly, such as in cases shown in Figure 2. In this case, additional training data from new distributions should amplify generalization to an heterogeneous testing database more effectively. This is supported by the high performance of the rather small model of [22], overall winner of BRATS 2016 challenge, which was trained on an external private database along with the BRATS 2015 training set, as well as reports by the respective team that their incorporation amplified performance. It would be interesting to explore how much data is needed from a new source in order for a network to also generalize satisfyingly to the new distribution. Additionally, it would be worth investigating a relation between ideal network capacity versus the amount of available training data, as well as explore these factors on other segmentation tasks.

Finally, the version of DeepMedic with residual connections was further benchmarked on the BRATS 2016 challenge among 19 teams. It exhibited very good performance, achieving top ranking for the Core and Enhancing tumour classes in terms of DICE. Our system also performed very well in assessing the longitudinal change of the whole and enhancing tumor volume, ranking third for its average performance on these classes. Less satisfying was the achieved Hausdorf distance, which we mostly attribute to false segmentation of parts of the skull and extra-cerebral tissue that were left behind from incomplete skull stripping of the testing data. This performance is particularly satisfying considering the minimal preprocessing and the generic architecture of our system. Its performance is likely to benefit from a more extensive pre-processing pipeline, such as from careful skull stripping, bias field correction, as well as from careful selection of high quality training data. Finally, an interesting trend is that histogram matching to a common template has been a part of top performing pipelines in BRATS 2015 and 2016 challenges [9,11,22], even though the technique is often criticized as not well suited for the problem of tumour segmentation.

## Acknowledgements

This work is supported by the EPSRC First Grant scheme (grant ref no. EP/N023668/1) and partially funded under the 7th Framework Programme by the European Commission (TBIcare: http://www.tbicare.eu/; CENTER-TBI: https://www.center-tbi.eu/). Part of this work was carried on when KK was an intern at Microsoft Research Cambridge. KK is also supported by the President's PhD Scholarship of Imperial College London. We gratefully acknowledge the support of NVIDIA Corporation with the donation of two Titan X GPUs for our research.

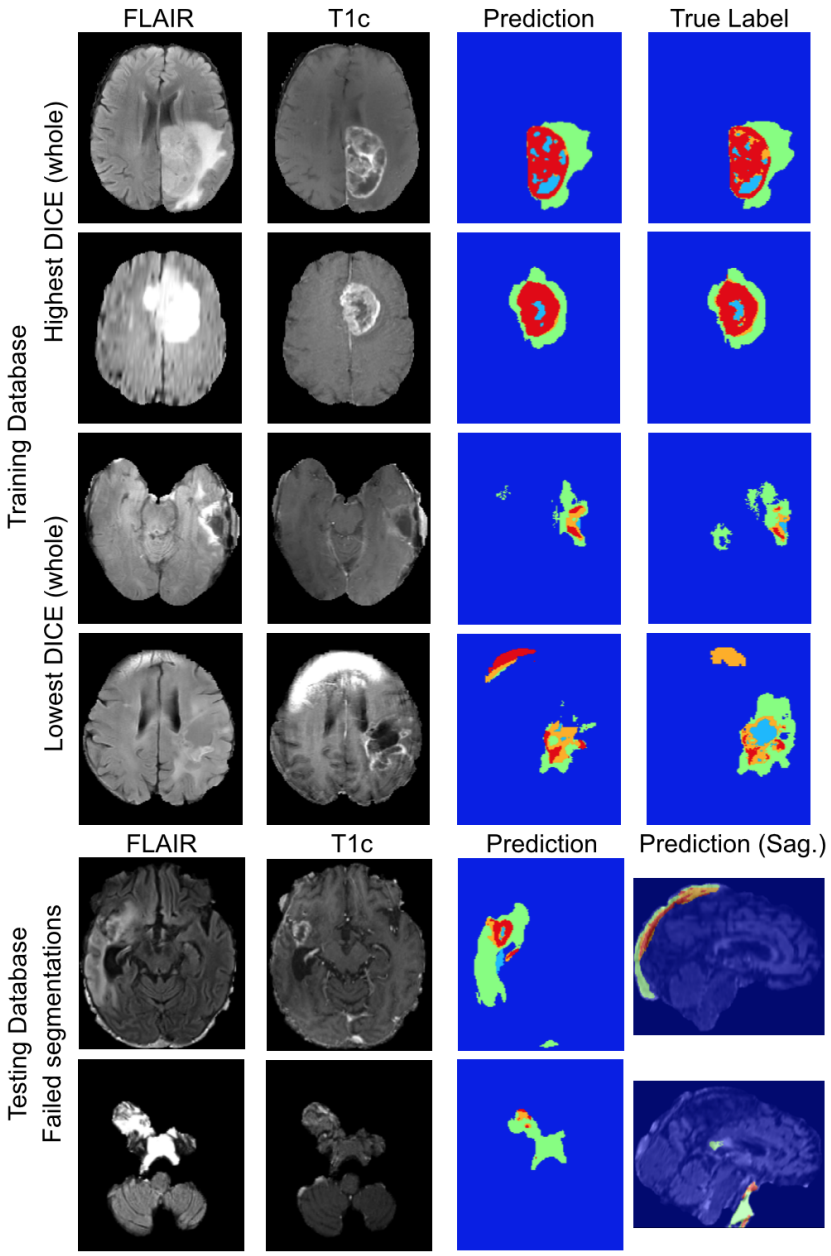


Fig. 2: Cases with the highest (two top rows) and lowest (third and forth rows) DICE for the whole tumor segmentation from the 5-fold validation on the training database. Note that the semi-automatically generated training labels also contain mistakes. (Two bottom rows) Two examples of the most common type of failed segmentation observed in the predictions on the testing data. Subset of the data provided in the testing stage of the challenge show significantly more remnants of skull and extra-cerebral tissue left behind from the brain extraction than what observed in the training database. The network fails to handle tissues that it has rarely seen during training. Colors represent: cyan: necrotic core, green: oedema, orange: non enhancing core, red: enhancing core. Cases from top to bottom row: tcia_242_01, tcia_479_01, tcia_164_01, tcia_222_304, cbica_ABH_341, cbica_AAM_285

## References


1. Kamnitsas, K., Ledig, C., Newcombe, V.F., 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** (2016) 61–78
2. Mazzara, G.P., Velthuizen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated mri segmentation. International Journal of Radiation Oncology* Biology* Physics **59**(1) (2004) 300–312
3. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical image analysis **8**(3) (2004) 275–283
4. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: MICCAI. Springer (2011) 532–540
5. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain mr images. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012. Springer (2012) 651–658
6. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O., Das, T., Jena, R., Price, S.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In: MICCAI. Springer (2012) 369–376
7. Tustison, N., Wintermark, M., Durst, C., Brian, A.: ANTs and Arboles. in proc of BRATS-MICCAI (2013)
8. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. in proc of BRATS-MICCAI (2014)
9. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging **35**(5) (2016) 1240–1251
10. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). Medical Imaging, IEEE Transactions on **34**(10) (2015) 1993–2024
11. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: Glistrboost: Combining multimodal mri segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer (2015) 144–155
12. 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. arXiv preprint arXiv:1505.03540 (2015)
13. Lyksborg, M., Puonti, O., Agn, M., Larsen, R.: An ensemble of 2d convolutional neural networks for tumor segmentation. In: Image Analysis. Springer (2015) 201–211
14. Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B.: Multi-scale 3d convolutional neural networks for lesion segmentation in brain mri. Ischemic Stroke Lesion Segmentation (2015) 13
15. Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al.: Isles 2015-a public eval-

uation benchmark for ischemic stroke lesion segmentation from multispectral mri. Medical Image Analysis **35** (2017) 250–269
16. Alansary, A., Kamnitsas, K., Davidson, A., Khlebnikov, R., Rajchl, M., Malamateniou, C., Rutherford, M., Hajnal, J.V., Glocker, B., Rueckert, D., et al.: Fast fully automatic segmentation of the human placenta from motion corrupted mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2016) 589–597
17. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
18. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. arXiv preprint arXiv:1603.05027 (2016)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
21. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. Advances in Neural Information Processing Systems 24 (2011)
22. Chang, P.D.: Fully convolutional neural networks with hyperlocal features for brain tumor segmentation. in proc of BRATS-MICCAI (2016)
23. Le Folgoc, L., Nori, A.V., Alvarez-Valle, J., Lowe, R., Criminisi, A.: Segmentation of brain tumors via cascades of lifted decision forests. in proc of BRATS-MICCAI (2016)
24. Zhao, X., Wu, Y., Song, G., Li, Z., Fan, Y., Zhang, Y.: Brain tumor segmentation using a fully convolutional neural network with conditional random field. in proc of BRATS-MICCAI (2016)
25. Zeng, K., Bakas, S., Sotiras, A., Akbari, H., Rozycki, M., Rathore, S., Pati, S., Davatzikos, C.: Segmentation of gliomas in pre-operative and post-operative multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. in proc of BRATS-MICCAI (2016)
26. Meier, R., Knecht, U., Wiest, R., Reyes, M.: Crf-based brain tumor segmentation: Alleviating the shrinking bias. in proc of BRATS-MICCAI (2016)