

Automatic Postoperative Brain Tumor Segmentation with Limited Data using Transfer Learning and Triplet Attention

Jingpeng Li^{*1,2} and Atle Bjørnerud^{1,2}

¹Computational Radiology and Artificial Intelligence (CRAI), Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway

²Department of Physics, University of Oslo, Oslo, Norway

Abstract

Accurate brain tumor segmentation is clinically important for diagnosis and treatment planning. Convolutional neural networks (CNNs) have achieved promising performance in various visual recognition tasks. Training such networks usually requires large amount of labeled data, which is often challenging for medical applications. In this work, we address the segmentation problem by applying transfer learning to downstream segmentation tasks. Specifically, we explore how knowledge acquired from a large preoperative dataset can be transferred to postoperative tumor segmentation on a smaller dataset. To this end, we have developed a 3D CNN for brain tumor segmentation, and fine-tuned the pretrained models on the target domain data. To better exploit the inter-channel and spatial information, triplet attention has been incorporated and extended into existing segmentation network. Extensive experiments on our dataset demonstrate the effectiveness of transfer learning and attention modules for improved postoperative tumor segmentation performance when only limited amount of annotated data is available.

1 Introduction

Glioblastoma is the most aggressive brain tumor and is commonly treated with surgery and chemoradiotherapy [5, 7, 19]. Accurate diagnosis and segmentation of glioblastoma is essential for treatment planning and postoperative analysis. Magnetic resonance imaging (MRI) provides high

soft tissue contrast and is the modality of choice for structural brain analysis [1, 4]. Manual tumor segmentation is challenging and time-consuming due to complex tumor structure and high inter-rater variability [6, 7, 19, 20], making automatic segmentation methods increasingly popular. Deep neural networks such as Convolutional Neural networks (CNNs) have achieved state-of-the-art performance in a range of vision recognition tasks, showing great potential for improved brain tumor segmentation performance. However, training such neural networks usually requires large amount of labeled data, which limits its application in medical imaging field.

Transfer learning leverages knowledge gained in a source domain to improve learning in a target domain without training the network from scratch. Zoetmulder *et al.* [21] assessed the transfer learning performance on multiple medical segmentation tasks by investigating various combinations of domains and tasks. Wacker *et al.* [19] applied fully-convolutional networks with pretrained encoders on ImageNet dataset for the task of brain tumor segmentation and achieved improved and more robust segmentation results. Ghaffari *et al.* [6] developed a 3D densely-connected U-net and transfer the knowledge acquired on preoperative brain tumor dataset to a target postoperative dataset.

Recent advances in deep learning, notably attention mechanism, have been shown to achieve performance gain in a variety of tasks, due to its capability of learning more discriminative representations by leveraging the semantic correlations among image regions. One of the most promising method is squeeze-and-excite networks (SENet) [9]. Triplet attention [14], a successor of SENet, learns

*Corresponding Author: jingpeng.li@fys.uio.no

more robust representation by exploiting the inter-dependencies among channels and spatial locations and provides computationally affordable and effective performance gains.

In this paper, we develop a popular 3D U-net variant architecture for brain tumor segmentation, and perform extensive experiments on various fine-tuning strategies on the target domain. Concretely, the network is first trained a large public preoperative dataset, the Brain Tumor Segmentation Challenge (BraTS) [13, 2, 3, 12], and then fine-tuned the model using transfer learning on an in-house target dataset of postoperative gliomas. Inspired by the success of attention-based methods, we propose to add the triplet attention module to the baseline network, introducing an additional branch of attention and adapting the attention module in the ResNet [8] backbone to fit into our network.

2 Method

2.1 Network Architecture

We use a 3D U-Net variant as our baseline network, which follows an encoder-decoder architecture with asymmetrically larger encoding pathways [16, 15]. The encoder structure comprises four stages of ResNet blocks. Each block consists of two convolutions with Instance Normalization [17] and Rectified Linear Unit (ReLU) activation, followed by additive identity skip connection. The input patches are progressively downsampled by convolutions with stride of 2. Each decoder level consists of a single ResNet block, and transpose convolutions with stride of 2 are used to double the spatial dimension and reduce the number of features. The endpoint of the decoder has the same spatial dimension as the input image with its channel dimension reduced to three after a $1 \times 1 \times 1$ convolution, followed by a sigmoid function. The network architecture is shown in Figure 1.

2.1.1 Loss

The Dice Similarity Coefficient (DSC) is used to measure the similarity between the predicted segmentation P and ground truth G :

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \quad (1)$$

The associated soft Dice loss can be expressed as:

$$\mathbf{L}_{dice} = 1 - \frac{2 \sum_i^n p_i * g_i}{\sum_i^n p_i + \sum_i^n g_i}, \quad (2)$$

where p_i and g_i are the predicted probability and ground truth label of the i -th voxel, respectively.

2.2 Transfer learning

Training a deep neural network from scratch with a small labeled dataset is challenging. The model is often not able to learn meaningful information when the training dataset is small, a problem called overfitting. One popular strategy to cope with limited data size and reduce overfitting is to apply transfer learning. The model is first trained on a large dataset, which allows the model to learn more general-purposed features. When directly apply the trained model on the local data, it might not work very well as the model weights are optimized based on the large dataset. The trained model is then tuned on the local data to adjust its weights to better fit the new training data. The fine-tuning step can stabilize the training process and improve its predictive performance where the model is less prone to overfit to the local data.

During transfer learning, we usually prevent certain part of the network from being trained so that the learned weights can be reused, where the rest of the network is trained as usual. The frozen part typically corresponds to the first shallow layers of the neural network, which tend to capture low-level features that are shared across domains, while deeper layers can learn high-level features that are more task-specific [18]. We only retrain those deep layers, thus avoiding losing those shared low-level information learned in the shallow layers. In other words, transfer learning adapts the pretrained model to the target domain, retaining shared information and fine-tuning domain-specific knowledge. More specifically, we partially transfer the pretrained model weights to the target domain by freezing the weights of initial shallow layers and fine-tuning the remaining layers.

2.3 Triplet attention

The triplet attention module was initially applied to ResNet backbone networks for classification and

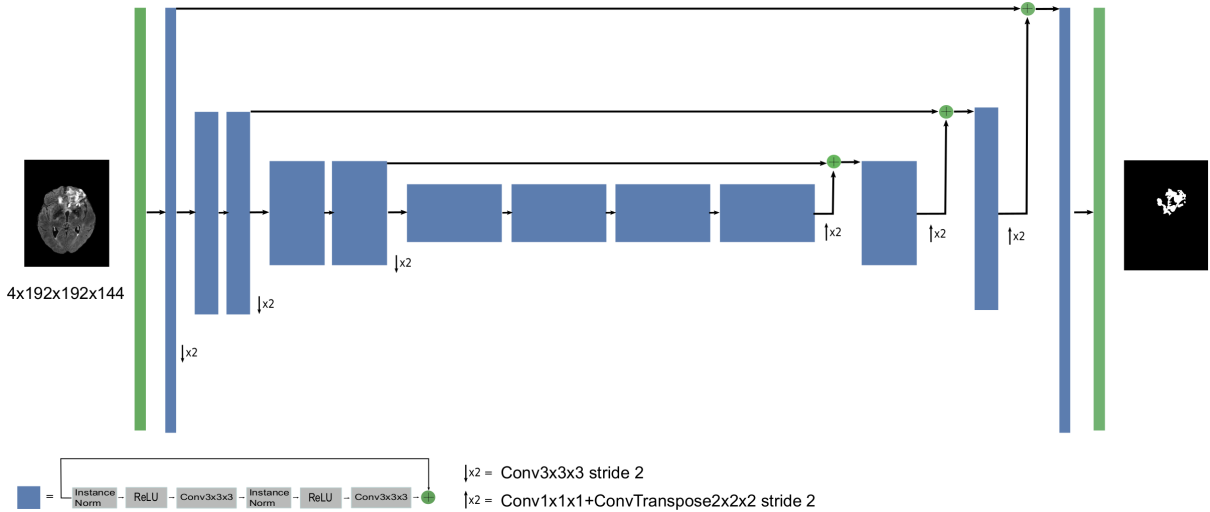


Figure 1: Schematic visualization of the baseline network architecture.

object detection [14]. We adapt the attention module to our 3D segmentation network by adding one additional branch capturing cross-dimensional interaction between channel and depth dimensions, as illustrated in Figure 2. We also replace the batch normalization with instance normalization in accordance with the baseline network.

The input feature x is passed to four branches, where three of them are responsible for computing attention weights across channel C and spatial dimensions H , W and D , and the last one captures spatial dependencies (H , W and D). In each channel-spatial branch, the feature is passed through a rotation operation followed by a residual transformation block, which consists of a Z-pool layer shrinking the depth dimension and a convolution layer, and the feature is rotated back afterwards to retain the same shape as input x . The last branch performs similar residual transformation, where rotation is not involved. The resulting outputs of all branches are averaged to generate the refined output of the attention module \tilde{x} . The process can be represented as follows:

$$\tilde{x} = \frac{1}{4}(\overline{\hat{x}_1\omega_1} + \overline{\hat{x}_2\omega_2} + \overline{\hat{x}_3\omega_3} + x\omega_4), \quad (3)$$

where ω_1 , ω_2 , ω_3 and ω_4 are the cross-dimensional attention weights, \hat{x}_1 , \hat{x}_2 and \hat{x}_3 are the rotated and Z-pooled features.

Following the original paper, the triplet attention module was appended to the bottleneck of the

encoder-decoder network.

3 Experiments

3.1 Datasets

3.1.1 Source Dataset

The publicly available BraTS dataset provides 3D multimodal MRI data with ground truth segmentation annotated by domain experts. The BraTS 2020 dataset comprises multi-parametric MRI (mp-MRI) scans with 369 cases for training and 125 cases for validation. The MRI scans were collected with different clinical protocols and from multiple institutions. Each MRI scan contains four modalities: native T1-weighted (T1), post-contrast T1-weighted (T1ce), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). Each 3D volume is skull-stripped, rigidly co-registered, and resampled to 1 mm^3 isotropic voxel resolution. Three tumor sub-regions were manually annotated by one to four raters: the Gd-enhancing tumor, peritumoral edema, and necrotic and non-enhancing tumor core. The annotations were combined into overlapping sub-regions for evaluation: enhancing tumor (ET), tumor core (TC), and whole tumor (WT).

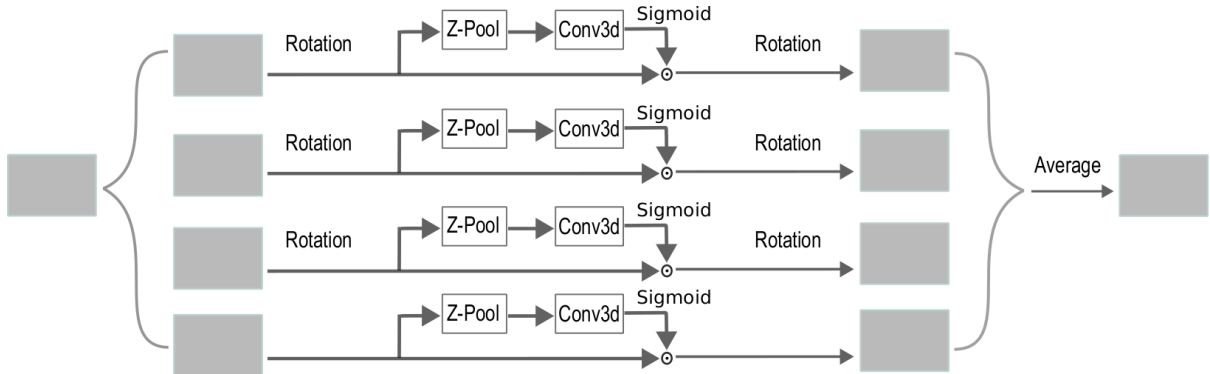


Figure 2: Illustration of the triplet attention module. The first three branches compute the attention weights between the channel dimension C and each of the three spatial dimensions (H , W , and D). The last branch captures the spatial dependencies (H , W and D). The output is obtained by averaging the computed weights from each branch. The attention module is added to the bottleneck of the encoder-encoder network.

3.1.2 Target Dataset

The target dataset is a subset of a glioblastoma study ongoing at our institution and consists of postoperative longitudinal brain MRI from 13 glioblastoma patients, with one to three time points per patient (see Table 1). The dataset was pre-processed similar to [12], including converting DICOM images to NIfTI format, skull-stripping using [10], and resampling to $1\text{mm} \times 1\text{mm} \times 1\text{mm}$ voxel resolution, rendering an image size of $256 \times 256 \times 190$. The same segmentation labels as in BraTS were generated. Here, an initial automated segmentation algorithm was applied to a larger dataset whereby a subset of this ($n = 13$ patients) was selected based on the review and segmentation quality of enhancing tumor and edema rating by a radiologist. Necrotic tumor was not specifically qualified by radiologist.

Table 1: Statistics of the target postoperative dataset.

Time-points per patient	Number of patients
One time-point scan	5
Two time-point scans	4
Three time-point scans	4
Total number of patients	13

3.2 Experimental Settings

All experiments were implemented in PyTorch, and the network was trained on NVIDIA A100 GPUs. We used five-fold cross-validation (CV) on 80% of the whole data and the remaining 20% as test set. All partitions were performed randomly at patient-level. In each experiment, there were roughly 8 patients in the training set, 2 in the validation set, and 3 in the test set.

3.2.1 Preprocessing

For both pretraining and fine-tuning, we used the same preprocessing steps. We randomly cropped the MRI images to a fixed size of $192 \times 192 \times 144$ and concatenated the four MRI modalities of each patient into a four-channel input. We independently normalized each channel by subtracting the mean intensity and dividing by the standard deviation of intensities within the brain region. Data augmentation techniques have shown to effectively reduce overfitting. We randomly flipped each spatial axis with a probability of 0.5 and applied a random intensity shift within $[-0.1, 0.1]$ of the standard deviation of each input channel, followed by a random intensity scaling in the range of $[0.9, 1.1]$.

3.2.2 Training

In both training cases, the Adam optimizer and a polynomial learning rate decay schedule, L2 regu-

larization with a decay rate of $1e-5$ were used. A batch size of 1 was used to compromise the large crop size, and the maximal number of epochs was set to 300. The learning rate were individually set to $1e-4$ and $1e-5$ for pretraining and fine-tuning. The model with best performance on the validation set was chosen. In our experiments, we found that freezing the first layer (level) achieved the best performance, which was discussed in the Appendix.

3.2.3 Evaluation and Comparisons

All experiments were evaluated as the average Dice score from the 5-fold cross validation results. To evaluate the segmentation performance using transfer learning, we trained the baseline network on the BraTS 2020 dataset, and tested its fine-tuned model on the target postoperative dataset. We then compared the performance of the baseline model trained on the target set from scratch with the fine-tuning approach on the pretrained model.

4 Results and Discussion

Table 2 presents the results of the baseline network trained and evaluated on BraTS 2020 (validation) dataset, showing performance close to the state-of-the-art in the BraTS challenge. The competitive results for the related preoperative brain tumor segmentation task convinced us that our baseline network is a good choice for the evaluation of the postoperative segmentation task. Table 3 compares the average Dice scores between the baseline and transfer learning, where fine-tuning achieves the highest value of 0.8018, with an absolute improvement of 0.04 from the baseline. It is observed that the Dices scores of tumor core increase most, compared to whole tumor and enhancing tumor, which It also shows that the Dices scores of the whole tumor increase most, which can be explained by the fact that the task of segmenting whole tumor is fairly easy in both domains and therefore the adaptation of the pretrained model weights to the target domain is more straightforward.

By incorporating the triplet attention into the baseline network, further performance gain is obtained. An example case of the segmentation result is provided in Figure 3.

5 Conclusion

Accurate segmentation of the different pathological components in postoperative glioblastomas from MRI is clinically important yet technically challenging. Starting with a network that achieves close to state-of-the-art performance in the preoperative brain tumor segmentation challenge, we here show that with limited labeled data transfer learning from pretrained model can be applied to improve segmentation performance in postoperative MRIs.

We add an attention module, triplet attention, to a 3D segmentation network. Triplet attention captures more discriminative features across channel and spatial dimensions, thus can enhance the model accuracy.

To note that transfer learning and triplet attention were independently experimented. For future work, we plan to integrate the two techniques and evaluate how attention mechanisms and transfer learning could potentially benefit each other.

Acknowledgements

The work was supported by Norwegian South-Eastern Health Authority under grant HSØ 2020029. The authors thank the Machine learning infrastructure (ML Nodes) of the University Centre for Information Technology at the University of Oslo for providing GPU resources.

6 Appendix

We experimented on the number of shallow layers to be frozen. As illustrated in Figure 4, fine-tuning with freezing the first layer achieved the highest Dice score on the target dataset. Also to note that the performance were very close when freezing 0, 1, and 2 layers.

Table 2: The baseline network evaluated on BraTS 2020 validation set, in comparison with the state-of-the-art results, as well as the ensemble of best performing variant models. All results were provided by the BraTS evaluation platform.

Method	WT	TC	ET	Mean
nnU-Net [11]	0.9124	0.8506	0.7989	0.854
Single baseline model (ours)	0.9051	0.8255	0.7821	0.8376
Ensemble of best baseline variants (ours)	0.9068	0.8425	0.7944	0.8479

Table 3: Quantitative results of the different methods on the target dataset. Pretrained model is trained on BraTS data. Baseline is trained from scratch on the target dataset. Transfer Learning (TL) fine-tunes the pretrained network with its first layer being frozen. Triplet attention (TA) is trained with added triplet attention to existing network. Note that TL and TA are experimented independently.

Model	WT	TC	ET	Mean
Pretrained	0.7136 ± 0.1146	0.7737 ± 0.107	0.6232 ± 0.1461	0.744 ± 0.0992
Baseline	0.8162 ± 0.079	0.7877 ± 0.085	0.7874 ± 0.080	0.7971 ± 0.079
TL	0.8301 ± 0.081	0.8321 ± 0.059	0.8244 ± 0.055	0.8289 ± 0.063
TA	0.8282 ± 0.0812	0.8060 ± 0.0693	0.8041 ± 0.0654	0.8042 ± 0.0695

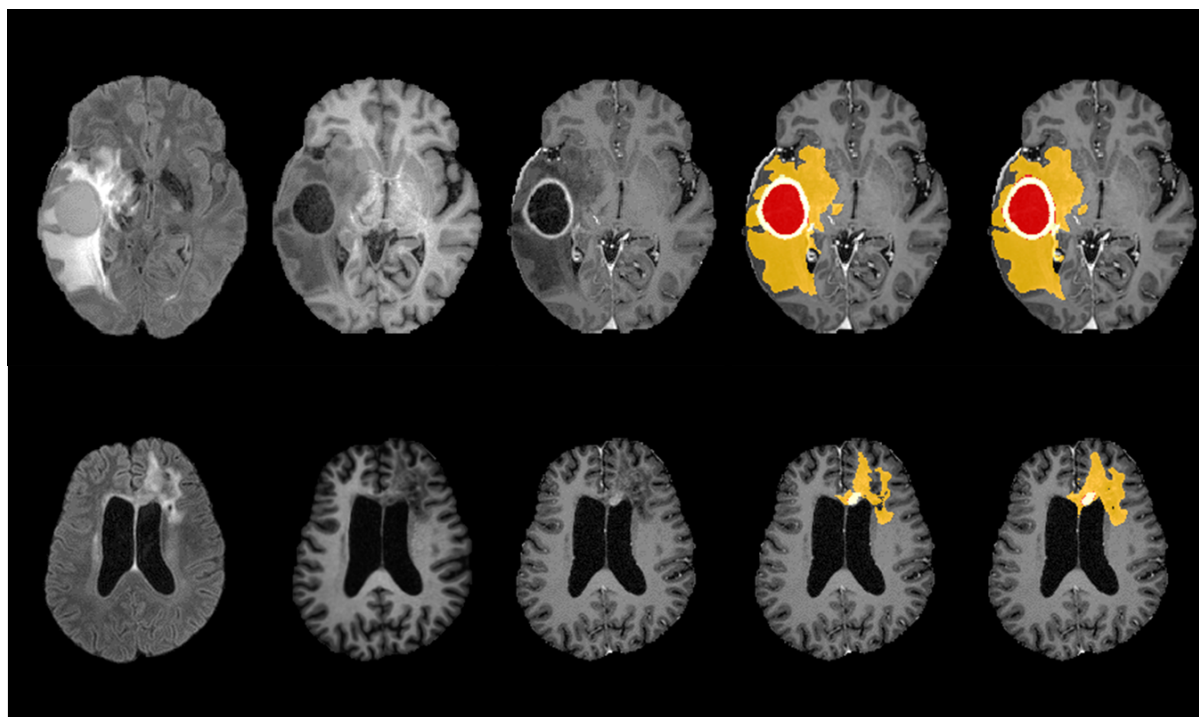


Figure 3: Visualization of the segmentation results of two sample cases. From left to right: FLAIR, T1, T1ce, ground truth overlaid on T1ce, predicted segmentation overlaid on T1ce.

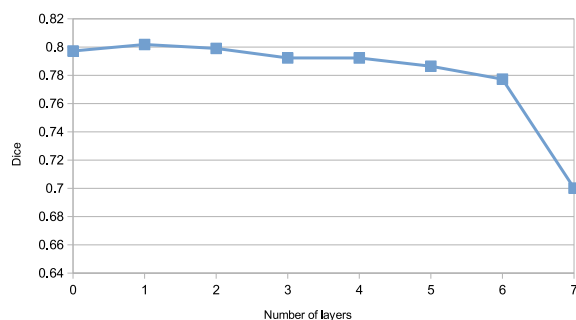


Figure 4: Average Dice scores corresponding to freezing different number of layers. In total, there are seven levels of convolutional layers in addition to the output endpoint.

References

- [1] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017. doi: 10.1007/s10278-017-9983-4.
- [2] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017. doi: 10.1038/sdata.2017.117.
- [3] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.
- [4] S. Deepak and P. Ameer. Brain tumor classification using deep cnn features via transfer learning. *Computers in biology and medicine*, 111:103345, 2019. ISSN 0010-4825. doi: 10.1016/j.compbimed.2019.103345.
- [5] A. K. Dhara, K. R. Ayyalasomayajula, E. Arvids, M. Fahlström, J. Wikström, E.-M. Larsson, and R. Strand. Segmentation of post-operative glioblastoma in mri by u-net with patient-specific interactive refinement. In *International MICCAI Brainlesion Workshop*, pages 115–122. Springer, 2018. doi: 10.1007/978-3-030-11723-8_11.
- [6] M. Ghaffari, G. Samarasinghe, M. Jameson, F. Aly, L. Holloway, P. Chlap, E.-S. Koh, A. Sowmya, and R. Oliver. Automated post-operative brain tumour segmentation: A deep learning model based on transfer learning from pre-operative images. *Magnetic resonance imaging*, 86:28–36, 2022. doi: 10.1016/j.mri.2021.10.012.
- [7] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017. doi: 10.1016/j.media.2016.05.004.
- [8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.
- [9] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. doi: 10.1109/CVPR.2018.00745.
- [10] F. Isensee, M. Schell, I. Pfueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H.-P. Schlemmer, S. Heiland, W. Wick, et al. Automated brain extraction of multisequence mri using artificial neural networks. *Human brain mapping*, 40(17):4952–4964, 2019. doi: 10.1002/hbm.24750.
- [11] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein. nnu-net for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 118–132. Springer, 2020.
- [12] R. Mehta, A. Filos, U. Baid, C. Sako, R. McKinley, M. Rebsamen, K. Dätwyler, R. Meier, P. Radojewski, G. K. Murugesan,

- et al. Qu-brats: Miccai brats 2020 challenge on quantifying uncertainty in brain tumor segmentation—analysis of ranking metrics and benchmarking results. *arXiv preprint arXiv:2112.10074*, 2021.
- [13] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. doi: 10.1109/TMI.2014.2377694.
- [14] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou. Rotate to attend: Convolutional triplet attention module. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3138–3147, 2021. doi: 10.1109/WACV48630.2021.00318.
- [15] A. Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*, pages 311–320. Springer, 2018. doi: 10.1007/978-3-030-11726-9_28.
- [16] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.
- [17] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.
- [18] J. M. Valverde, V. Imani, A. Abdollahzadeh, R. De Feo, M. Prakash, R. Ciszek, and J. Tohka. Transfer learning in magnetic resonance brain imaging: a systematic review. *Journal of imaging*, 7(4):66, 2021. doi: 10.3390/jimaging7040066.
- [19] J. Wacker, M. Ladeira, and J. E. V. Nascimento. Transfer learning for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 241–251. Springer, 2020. doi: 10.1007/978-3-030-72084-1_22.
- [20] G. Wang, W. Li, S. Ourselin, and T. Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI brainlesion workshop*, pages 178–190. Springer, 2017.
- [21] R. Zoetmulder, E. Gavves, M. Caan, and H. Marquering. Domain-and task-specific transfer learning for medical segmentation tasks. *Computer Methods and Programs in Biomedicine*, 214:106539, 2022. doi: 10.1016/j.cmpb.2021.106539.