This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Unsupervised Domain Adaptation for Pediatric Brain Tumor Segmentation

Jingru Fu[1[0000−0003−4175−395X]], Simone Bendazzoli[1,2[0000−0001−6673−1314]], Örjan Smedby[1[0000−0002−7750−1917]], and Rodrigo Moreno[1[0000−0001−5765−2964]]

[1] KTH Royal Institute of Technology, Stockholm, Sweden
{jingruf,simben,orsme,rodmore}@kth.se
[2] Karolinska Institutet, Stockholm, Sweden

Abstract. Significant advances have been made toward building accurate automatic segmentation models for adult gliomas. However, the performance of these models often degrades when applied to pediatric glioma due to their imaging and clinical differences (*domain shift*). Obtaining sufficient annotated data for pediatric glioma is typically difficult because of its rare nature. Also, manual annotations are scarce and expensive. In this work, we propose Domain-Adapted nnU-Net (DA-nnUNet) to perform unsupervised domain adaptation from adult glioma (*source domain*) to pediatric glioma (*target domain*). Specifically, we add a domain classifier connected with a gradient reversal layer (GRL) to a backbone nnU-Net. Once the classifier reaches a very high accuracy, the GRL is activated with the goal of transferring domain-invariant features from the classifier to the segmentation model while preserving segmentation accuracy on the source domain. The accuracy of the classifier slowly degrades to chance levels. No annotations are used in the target domain. The method is compared to 8 different supervised models using BraTS-Adult glioma (N=1251) and BraTS-PED glioma data (N=99). The proposed method shows notable performance enhancements in the tumor core (TC) region compared to the model that only uses adult data: ∼32% better Dice scores and ∼20 better 95th percentile Hausdorff distances. Moreover, our unsupervised approach shows no statistically significant difference compared to the practical upper bound model using manual annotations from both datasets in TC region. The code is shared at `https://github.com/Fjr9516/DA_nnUNet`.

Keywords: Unsupervised domain adaptation · Pediatric tumor segmentation · Gradient reversal layer.

1 Introduction

Pediatric brain tumors represent the most prevalent solid tumors and the leading cause of cancer-related mortality in children [5]. Accurate segmentation of these tumors from medical images is critical for surgical and treatment planning [14,5], facilitating the identification of tumor location, extent, and type.

Recent advances in automated brain tumor segmentation, driven by deep learning techniques and well-curated datasets, have shown promise in adult brain tumors [9,13,16]. In turn, an important issue for pediatric brain tumors is the scarcity of training data. For instance, while the incidence of adult glioblastomas is 3 in 100,000 people, pediatric diffuse midline gliomas are approximately three times rarer [11]. An additional issue is that obtaining accurate segmentation models involves expert annotation of data, which entails laborious processes.

An approach to deal with this data scarcity issue is to train deep-learning models with adult data and apply them to pediatric cases. This approach relies on the assumption that brain tumors are similar in both adults and children. Unfortunately, this approach often fails to generalize effectively to pediatric cases. For example, [16] reported a sharp decrease in the Dice similarity coefficient score (DSC) for the tumor core (TC) region, dropping from 0.8788 to 0.2639, when applying a model developed for adult gliomas to pediatric cases. Indeed, despite certain similarities between pediatric and adult tumors, there are also important differences in their appearance (see an example in Fig. 3 (right)). For instance, enhancing tumor regions on post-gadolinium T1-weighted MRI and radiologically-defined necrotic tissue regions are common in adult gliomas but are less common in pediatric cases [11].

International challenges, such as the brain tumor segmentation challenge (BraTS) [14], have been crucial for the development of successful approaches for adult tumor segmentation. BraTS provides participants with datasets of fully annotated, multi-institutional, multiparametric MR images (mpMRIs) of patients. Notably, winners of recent competitions have predominantly employed U-Net architectures, which demonstrate their superior performance, particularly for supervised learning with abundant training data. Among these, nnU-Net [9], a self-configuring framework that automatically adapts to specific datasets, has gained popularity for its robust performance [8,15,12]. Unsupervised and semisupervised approaches are promising for addressing problems with fewer training cases and/or with low to no annotation data, which is often the case of pediatric tumor segmentation.

Contributions: In this paper, we propose an unsupervised domain adaptation (UDA) method for pediatric glioma segmentation. In particular, we add a domain-adversarial task to a backbone nnU-Net. The goal of this task is to encourage the nnU-Net to learn domain-invariant features for segmentation. The inclusion of a domain classifier stems from principles of domain adaptation, which suggest that effective domain transfer requires predictions based on features that do not distinguish between the source and target domains [2]. This concept was initially introduced in neural architectures in [6] for natural image classification tasks, employing a gradient reversal layer (GRL) to learn domain-invariant intermediate features for the final task. Similar efforts have been made to extend this concept to natural image segmentation [3] and 3D medical binary segmentation problems [10]. In this paper, we adapt this approach to the 3D region-based brain tumor segmentation problem, which, to the best of our knowledge, has not been previously explored.

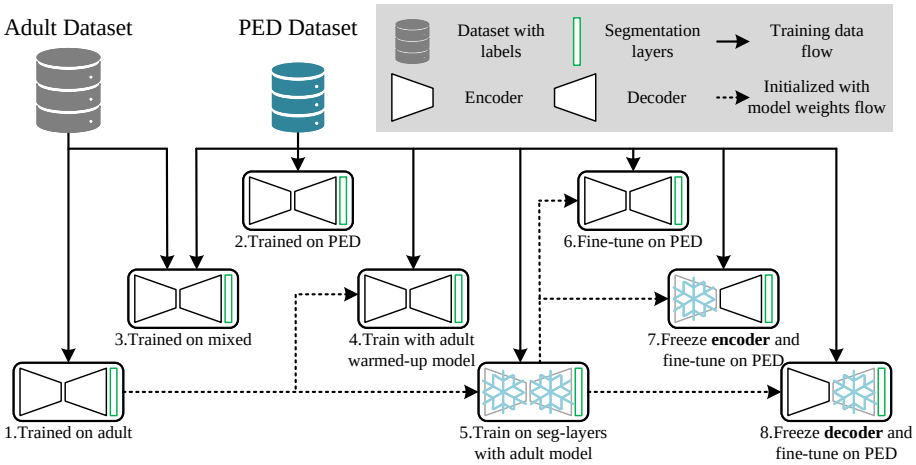

Fig. 1. Different strategies to obtain upper-bound and baseline models when assuming fully-annotated pediatric (PED) data are provided.

2 Method

In domain adaptation (DA), the objective is to learn a model from a source domain $\mathcal{D}_s$ and apply it to a related but different target domain $\mathcal{D}_t$ for the same task ($\mathcal{T}_s = \mathcal{T}_t$). In this study, we focus on UDA and use supervised domain adaptation (SDA) to establish upper-bound and baseline models.

2.1 Supervised Baseline Models

Despite their differences, adult and pediatric tumors share relevant characteristics. Thus, using abundant annotated adult data during training should be beneficial for pediatric tumor segmentation [4]. In case manual annotations of pediatric images are available, a standard procedure is to apply transfer learning from adult to pediatric gliomas by combining the two datasets during training [7,1]. This combination can be performed in different ways.

Fig. 1 shows the 8 different supervised models. Models 1-3 in Fig. 1 train 3D nn-UNets from scratch using either or both datasets. In turn, models 5-8 use standard transfer learning strategies. First, model 5 is trained only on task-related layers, based on the pre-trained nn-UNet for adult tumor segmentation as the feature extractor. Then different parts of the network are fine-tuned with pediatric data, with lower learning rates and fewer epochs in models 6-8. Model 4 is somewhat in the middle of the two approaches, where the whole network is retrained using pediatric data after pretraining with adult data.

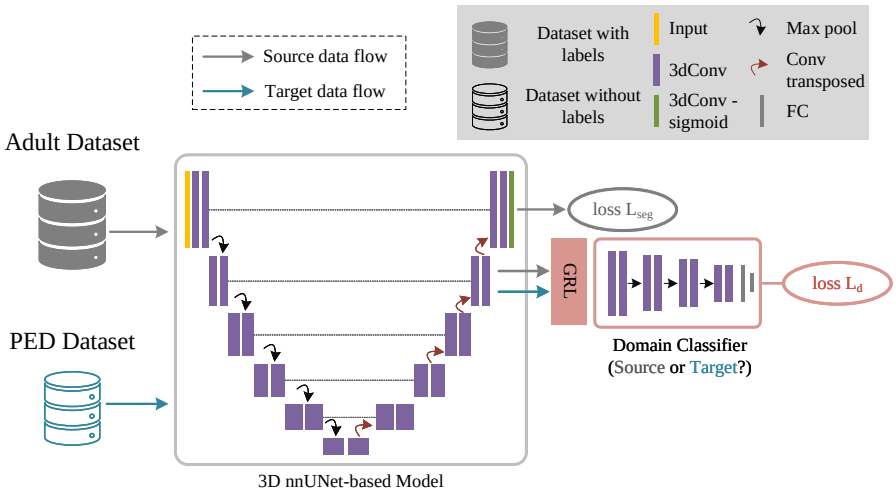

Fig. 2. The proposed DA-nnUNet comprises a 3D nnUNet-based backbone and a domain classifier. UDA is facilitated by integrating a gradient reversal layer (GRL) before the domain classifier.

2.2 Unsupervised Domain Adaptation for Pediatric Brain Tumor Segmentation

As illustrated in Fig. 2, the proposed architecture consists of a 3D nnUNet-based backbone and a domain classifier. During forward propagation, the GRL [6] acts as an identity function, while during backpropagation, it multiplies the gradient by a negative value α. This adversarial mechanism encourages the features passed to the domain classifier to become domain-invariant. At the final stage of the UNet decoder, which is responsible for the segmentation task, the segmentation loss L_{seg} is minimized only on adult samples, while the domain classifier minimizes a domain classification loss L_d across both domains.

The success of achieving domain adaptation with our model hinges on carefully balancing model learning between segmentation and domain classification tasks. Overfitting to the segmentation task may hinder domain adaptation, causing the model to excel only in source domain segmentation. Conversely, overfitting to the classification task may compel the model to prioritize learning domain-invariant features to deceive the classifier, potentially sacrificing segmentation performance. To address these concerns, we introduced several strategies to better balance the two tasks.

Inclusion of domain classifier weight in loss: The total loss comprises two primary components: a segmentation loss L_{seg} and a domain classifier loss L_d. We employed a combination of pseudo-Dice loss and binary cross-entropy loss as L_{seg}, akin to the default loss function in nnU-Net. For L_d, we utilized cross-entropy loss for the one-hot with two domain labels. The total loss is represented as follows:

Table 1. Datasets used in the experiments.

Dataset	Data Subset	# Subjects	Labels Used (Supervised)	Labels Used (Unsupervised)
BraTS Adult Glioma	$\mathcal{D}_{\text{Adult}}^{\text{source}}$	1251	✓	✓
BraTS-PEDs	$\mathcal{D}_{\text{PED}}^{\text{target}}$	99	✓	

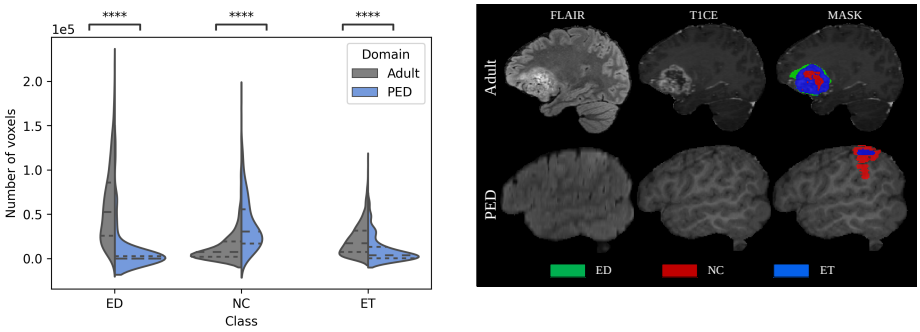

Fig. 3. Distribution of tumor sizes (left) and examples of the appearance of adult and pediatric tumors on FLAIR and T1w after contrast (T1CE) (right).

$$L_{total} = \sum_{x \in \mathcal{D}_s} L_{seg}(\hat{y}, y) + \lambda \sum_{x \in (\mathcal{D}_s \cup \mathcal{D}_t)} L_d(\hat{d}, d), \tag{1}$$

where λ denotes the domain classifier weight, controlling the relative importance of the segmentation and classification tasks; $\hat{d}$ and d are the predicted and actual domain labels; and $\hat{y}$ and y represent the predicted and ground truth segmentation masks for the adult dataset $\mathcal{D}_s$, respectively. We assume no annotations are available for the pediatric dataset $\mathcal{D}_t$, so they are not considered in L_{seg}.

Dynamic scheduler for α in GRL: A crucial component is the adjustment of α in the GRL, which dictates the strength of feature adaptation in response to the domain classifier. Following [10], we implemented a truncated ramp function in which the value of α grows linearly from 0 at epoch e_{min} to α_{max} at epoch e_{max}. e_{min} is selected such that the accuracy of the classifier is very high. Ideally, such an accuracy should decay to a chance level at e_{max}. Finally, we maintain α at its maximum value α_{max} to focus on refining further the segmentation.

3 Experimental Setting

3.1 Datasets

We utilized two databases provided by the BraTS challenge organizers: the BraTS 2021 adult glioma dataset [14], and the BraTS-PEDs 2023 dataset [11]. All data underwent identical preprocessing steps, including conversion to NifTI

format, skull stripping, defacing, co-registration to the SRI24 anatomical template, and resampling to isotropic resolution. Subsequently, the data were cropped to a size of [240, 240, 155]. Each dataset included four mpMRIs: pre- and post-gadolinium T1-weighted (T1w), T2-weighted (T2w), and T2 fluid attenuated inversion recovery (FLAIR) images. Additionally, expert annotated segmentation masks were provided, delineating three classes: enhancing tumor (ET), non-enhancing component (NC), and edema (ED).

We employed the BraTS training datasets to validate our method for the segmentation of three subregions: whole tumor (WT: ET+NC+ED), tumor core (TC: ET+NC), and enhancing tumor (ET). Details regarding each domain are summarized in Table 1. All models using PED data were trained using 5-fold cross validation (CV), while *adult* model was trained using all adult data. As shown in Fig. 3, there are significant disparities (i.e., *domain shift*) between adult and pediatric gliomas for tumor subregion sizes and appearance. Note that in the BraTS-PEDs set, 11 out of 99 cases lack the ET region (compared to 33 out of 1251 cases in the BraTS adult glioma set), while 42 out of 99 cases lack the ED region (compared to only 1 out of 1251 cases in the BraTS adult glioma set).

3.2 Implementation

Benefiting from the auto-configuring capability of nnU-Net, many preprocessing, postprocessing, and architecture hyperparameters can be robustly handled by nnU-Net's default strategies. However, to harness the DA capability of our proposed model, the introduction of a domain classifier necessitates careful selection of additional hyperparameters, along with extra effort to tailor current default segmentation configuration strategies for UDA.

Firstly, the architecture of the domain classifier must be designed and adjusted to accommodate the complexity of nnU-Net. We introduced the same block structure (i.e., double convolutional layers followed by max pooling) as in the backbone of nnU-Net. Two hyperparameters can be freely adjusted to modify the depth of the domain classifier: i) the number of blocks (set to 4 in our experiments); ii) the number of channels for convolutional layers and the final fully-connected (FC) layer (set to 100). Secondly, as discussed previously, the scheduler for α in the GRL involves three hyperparameters: $e_{min} = 100$, $e_{max} = 350$, and $\alpha_{max} = 3$. Additionally, we observed that the default deep supervision (DS) strategy in nnU-Net tends to prioritize the segmentation task, potentially bypassing the learning of domain-invariant features and leading to an imbalance between segmentation and classification tasks. Therefore, we disabled DS in our proposed model training. The hyperparameter λ is set to 0.01, the batch size is set to 4, the maximum number of epochs is set to 500, and a domain-balanced dataloader is utilized to prevent failure in domain classification. Finally, the learning rate is implemented following the same approach as described in [6], with an initial value of 0.01.

Table 2. Comparison of methods using 5-fold cross-validation (5CV) (excluding models 1 and 1*). The models are ranked separately based on whether they have access to adult source data for training, or if deep supervision (DS) was deactivated (marked with *). The best in each category is highlighted in bold.

Model	mean DSC			mean HD95			median DSC			median HD95		
	ET	TC	WT	ET	TC	WT	ET	TC	WT	ET	TC	WT
1	**0.583**	0.522	0.900	**58.77**	36.11	4.82	**0.742**	0.549	0.913	**4.12**	13.49	3.00
2	0.547	**0.872**	0.897	62.70	**5.53**	4.83	0.708	0.915	0.924	6.40	**3.61**	3.16
3	0.575	0.854	**0.908**	66.18	9.38	**4.22**	0.716	**0.918**	**0.927**	4.36	**3.61**	**2.83**
4	0.579	0.874	0.898	66.30	8.78	4.55	0.713	**0.914**	**0.922**	5.39	**3.32**	**3.00**
5	0.577	0.852	0.902	52.66	6.42	4.66	**0.732**	0.895	0.917	5.00	4.47	**3.00**
6	**0.584**	**0.879**	**0.906**	**51.79**	**5.23**	**4.26**	0.712	0.910	**0.922**	**4.24**	3.61	**3.00**
7	0.582	0.869	0.901	55.25	5.54	4.31	0.702	0.906	0.920	4.47	3.74	**3.00**
8	0.574	0.875	0.903	62.23	5.65	4.54	0.718	0.908	0.921	5.00	3.74	**3.00**
1*	0.571	0.526	0.890	59.61	22.25	7.07	**0.739**	0.555	0.909	**4.47**	13.49	3.61
2*	0.511	0.855	0.886	50.62	6.24	5.37	0.631	**0.909**	**0.921**	7.55	**3.74**	3.16
3*	**0.575**	**0.861**	**0.909**	**45.50**	**5.47**	**5.16**	0.721	0.905	0.920	5.10	**3.74**	**2.83**
UDA	0.580	0.846	0.892	50.87	16.55	8.90	0.713	0.916	0.923	4.24	3.32	3.00

4 Results and Discussion

We evaluated the performance of models on $\mathcal{D}_{\text{PED}}^{\text{target}}$ using average and median DSCs along with the 95th percentile Hausdorff distance (HD95). The results of all models are summarized in Table 2.

The results revealed that model 1, trained solely on adult data, mainly faces challenges in accurately segmenting the TC region, as evidenced by a decrease in mean DSC from 0.872 to 0.522 compared to model 2. In comparison to model 2, which was trained on a limited dataset of pediatric cases, model 3 benefits from combining both pediatric and adult datasets, resulting in improved median DSC and HD95 metrics due to its exposure to diverse domain characteristics. Since we deactivated the DS in the proposed UDA method, we also trained models 1-3 without DS (models 1*-3*) for fair comparison. As shown, these models perform very similarly to models 1 and 3 with DS with enough training data. Notably, models 3 and 3* are regarded as practical upper bounds for any DA model. Supervised DA models 4-8 achieve comparable results to the upper bound model trained with both datasets. Lastly, the proposed UDA method obtains segmentation accuracies which are close to the upper bound model 3* without requiring any annotations from pediatric domain. This underscores the necessity and efficacy of the adaptation.

From the results of Table 2, it is also evident that all tested methods have problems with segmenting ET: relatively low mean DSC and high mean HD95. Notably, there are significant disparities between median and mean metrics, with median values often considerably better than mean values (e.g., 4.12 median vs. 58.77 mean HD95 for model 1). As already mentioned, ET regions are usually very small or non-existent in pediatric cases. To assess the impact of small ET,

Table 3. Comparison methods after removing cases with no or small ET region. The best values are shown in bold, while the second-best values are underlined.

Model	mean DSC			mean HD95			median DSC			median HD95		
	ET	TC	WT	ET	TC	WT	ET	TC	WT	ET	TC	WT
1*	<u>0.661</u>	0.543	0.894	27.50	14.75	7.29	**0.768**	0.522	0.908	<u>3.61</u>	13.49	3.74
2*	0.616	0.857	0.894	14.05	6.36	5.27	0.729	0.900	0.916	5.10	4.24	3.46
3*	**0.685**	<u>0.863</u>	**0.913**	**12.98**	**5.74**	5.46	<u>0.760</u>	<u>0.902</u>	<u>0.918</u>	<u>3.61</u>	<u>4.12</u>	**3.00**
UDA	0.660	**0.864**	<u>0.908</u>	<u>13.22</u>	<u>5.90</u>	**5.15**	0.757	**0.913**	**0.919**	**3.16**	**3.74**	<u>3.16</u>

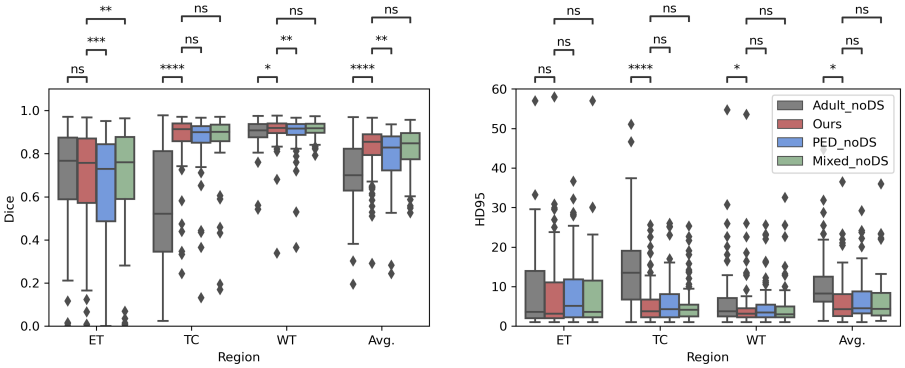

Fig. 4. Distributions of DSC (left) and HD95 (right) for Table 3. Pairwise t-tests were used to assess the statistical significance of differences between methods on $\mathcal{D}_{\text{PED}}^{\text{target}}$.

we reevaluated models 1*-3* and our UDA method without considering subjects with ET < 60 voxels (22 subjects were discarded, results shown in Table 3). We observe that DSC and especially HD95 are largely improved for all methods in the ET region. This suggests that the considerable variation in ET segmentation performance is largely influenced by small region size. Moreover, the proposed UDA method is not statistically different from the practical upper bound model 3* (see Fig. 4), except for the most challenging ET region in DSC metric.

5 Conclusion

Our experimental results show that the proposed unsupervised domain adaptation approach performs similarly to supervised approaches for pediatric tumor segmentation. The main advantage of our approach is that it has the potential to be used in cases where expensive manual annotations are not available without a significant drop in performance.

Acknowledgments. This study was partially funded by Barncancerfonden (grant number MT2022-0008) and the Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation. We thank the Knut and Alice Wallenberg Foundation for the use of Berzelius at NAISS in Sweden.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Amod, A.R., Smith, A., Joubert, P., Raymond, C., Zhang, D., Anazodo, U.C., Motchon, D., Mutsvangwa, T.E., Quetin, S.: Bridging the gap: Generalising state-of-the-art u-net models to sub-saharan african populations. arXiv preprint arXiv:2312.11770 (2023). https://doi.org/10.48550/arXiv.2312.11770
2. Ben-David, S., Blitzer, J., Crammer, K., Pereira, F.: Analysis of representations for domain adaptation. Advances in neural information processing systems **19** (2006). https://doi.org/10.7551/mitpress/7503.003.0022
3. Bolte, J.A., Kamp, M., Breuer, A., Homoceanu, S., Schlicht, P., Huger, F., Lipinski, D., Fingscheidt, T.: Unsupervised domain adaptation to improve image segmentation quality both in the source and target domain. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019). https://doi.org/10.1109/CVPRW.2019.00181
4. Boyd, A., Ye, Z., Prabhu, S., Tjong, M.C., Zha, Y., Zapaishchykova, A., Vajapeyam, S., Hayat, H., Chopra, R., Liu, K.X., et al.: Expert-level pediatric brain tumor segmentation in a limited data scenario with stepwise transfer learning. medRxiv (2023). https://doi.org/10.1101/2023.06.29.23292048
5. Fathi Kazerooni, A., Arif, S., Madhogarhia, R., Khalili, N., Haldar, D., Bagheri, S., Familiar, A.M., Anderson, H., Haldar, S., Tu, W., et al.: Automated tumor segmentation and brain tissue extraction from multiparametric mri of pediatric brain tumors: A multi-institutional study. Neuro-Oncology Advances **5**(1), vdad027 (2023). https://doi.org/10.1093/noajnl/vdad027
6. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016). https://doi.org/10.1007/978-3-319-58347-1_10
7. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)
8. Huang, Z., Wang, Z., Yang, Z., Gu, L.: Adwu-net: adaptive depth and width u-net for medical image segmentation by differentiable neural architecture search. In: International Conference on Medical Imaging with Deep Learning. pp. 576–589. PMLR (2022)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
10. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25. pp. 597–609. Springer (2017). https://doi.org/10.1007/978-3-319-59050-9_47
11. Kazerooni, A.F., Khalili, N., Liu, X., Haldar, D., Jiang, Z., Anwar, S.M., Albrecht, J., Adewole, M., Anazodo, U., Anderson, H., et al.: The brain tumor segmentation (BRATS) challenge 2023: Focus on pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). ArXiv (2023). https://doi.org/2305.17033v5
12. Kharaji, M., Abbasi, H., Orouskhani, Y., Shomalzadeh, M., Kazemi, F., Orouskhani, M.: Brain tumor segmentation with advanced nnu-net: Pediatrics and

adults tumors. Neuroscience Informatics p. 100156 (2024). https://doi.org/10.1016/j.neuri.2024.100156
13. Luu, H.M., Park, S.H.: Extending nn-UNet for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 173–186. Springer (2021). https://doi.org/10.1007/978-3-031-09002-8_16
14. 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). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2014). https://doi.org/10.1109/TMI.2014.2377694
15. Vossough, A., Khalili, N., Familiar, A.M., Gandhi, D., Viswanathan, K., Tu, W., Haldar, D., Bagheri, S., Anderson, H., Haldar, S., et al.: Training and comparison of nnu-net and deepmedic methods for autosegmentation of pediatric brain tumors. arXiv preprint arXiv:2401.08404 (2024). https://doi.org/10.48550/arXiv.2401.08404
16. Zeineldin, R.A., Karar, M.E., Burgert, O., Mathis-Ullrich, F.: Multimodal CNN networks for brain tumor segmentation in MRI: a BraTS 2022 challenge solution. arXiv preprint arXiv:2212.09310 (2022). https://doi.org/10.48550/arXiv.2212.09310