ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

O-Net: a Brain Tumor Segmentation Architecture Based on U-Net Using Alternated Pooling

Omar EL BARRAJ[1], Aya HAGE CHEHADE[2], Jean Marie MARION[1]
Mohamad OUEIDAT[3], Pierre CHAUVET[1] and Nassib ABDALLAH[4] *

1- LARIS, University of Angers
Angers - France

2- LabISEN, Yncrea Ouest
Brest - France

3-Faculty of Technology, Lebanese University
Beirut - Lebanon

4-LaTIM, INSERM UMR 1101, University of Brest
Brest - France

Abstract. Deep Learning (DL) offers promising tools for improving diagnostic processes in healthcare. Automated brain tumor segmentation using multi-parametric multimodal Magnetic Resonance Imaging (mpMRI) plays a vital role in the clinical management of brain tumor patients, enabling precise delineation of tumor regions. In this paper, we present O-Net, a deep learning model inspired by the U-Net architecture. O-Net employs an ensemble of two mirrored U-Nets with alternating pooling strategies –Max and Average Pooling– to enhance feature extraction. Our approach demonstrates the potential to improve segmentation accuracy using the BraTS 2021 training dataset and highlights the advantages of combining complementary pooling strategies for this task.

1 Introduction

In brain tumor segmentation, the ultimate goal is to save lives through early diagnosis and precise treatment. Manual segmentation is time-consuming, taking 45–60 minutes per case [1], while automated methods greatly reduce this time and provide consistent, reliable results. DL techniques have made remarkable strides in computer vision tasks [2], including brain tumor segmentation, with models like U-Net. The RSNA-ASNR-MICCAI BraTS challenges, ongoing since 2012, aim to advance segmentation methods for gliomas and other critical brain tumors. The BraTS 2021 challenge utilizes mpMRI scans to evaluate cutting-edge methods for two primary tasks, with task 1 focuses on segmenting glioblastoma sub-regions, including Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). U-Net has become a cornerstone of medical image analysis due to its encoder-decoder structure and skip connections, which effectively capture spatial context [3]. Researchers have extended U-Net by integrating

*This work is supported by the research unit of "Together for CHEHIM" Association and was financially funded by the Averroes College Foundation.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

different techniques. Also, the choice of pooling layers is critical for reducing spatial dimensions and enhancing feature extraction. Alternating between Max and Average Pooling is a relatively unexplored yet promising strategy. Our proposed O-Net model builds upon U-Net by introducing alternating pooling layers, capturing both dominant features and broader contextual information. O-Net aims to refine tumor segmentation, ultimately contributing to more accurate treatment planning and improved patient outcomes using BraTS 2021 dataset.

2 Related Work

Several surveys have been conducted in the field of DL to tackle the problem of image segmentation [4, 5], and numerous networks based on U-Net have been proposed, including U-Net++ [6], Reciprocal Adversarial Learning [7], Hybrid Attention-Based Residual U-Net (HA-Runet) [8], improved Residual Network (ResNet) [9], and 3D Cascade Dynamic Attention U-Net (ICU-Net) [10], all of which demonstrate advancements in feature extraction and segmentation performance. Recently, researchers have focused on integrating transformers with U-Net, resulting in architectures such as (HRSTNet) [11] and Swin-UNETR [12]. On the other hand, pooling layers play a pivotal role in downsampling within Convolutional Neural Networks (CNNs) [13], as they reduce spatial dimensions, facilitate efficient feature extraction, and improve computational efficiency [14]. Max Pooling, as explained by [15], extracts the maximum value (the most prominent feature), whereas Average Pooling computes the mean (providing a more generalized view). Combining these methods takes advantage of the strengths of both techniques, balancing detail preservation with contextual information [16].

3 Method

Given the historical success of U-Net (encoder-decoder structure) in medical imaging, which is based solely on Max Pooling, and recognizing that Max and Average Pooling, each offering distinct advantages [17], and since neither of them outperforms the other, there is a possible advantage to perform different types of pooling operations [18]. This approach may prevent information loss and preserve pixel adjacency relationship effectively. Our method combines U-Net's strengths with the advantages of alternating pooling methods to optimize semantic representation and enable the network to learn more diverse and richer feature representations. Our proposed architecture is called O-Net, a simple and enhanced U-Net model based on [3]. It represents a reflection of two U-Nets as shown in figure 1, making an ensemble where each U-Net alternates the use of Max and Average Pooling in parallel by extracting different features at the same level/depth of the network (vertically), leading to a possible preservation of spatial information. Finally, the feature maps from both reflected U-Nets are concatenated at the end.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

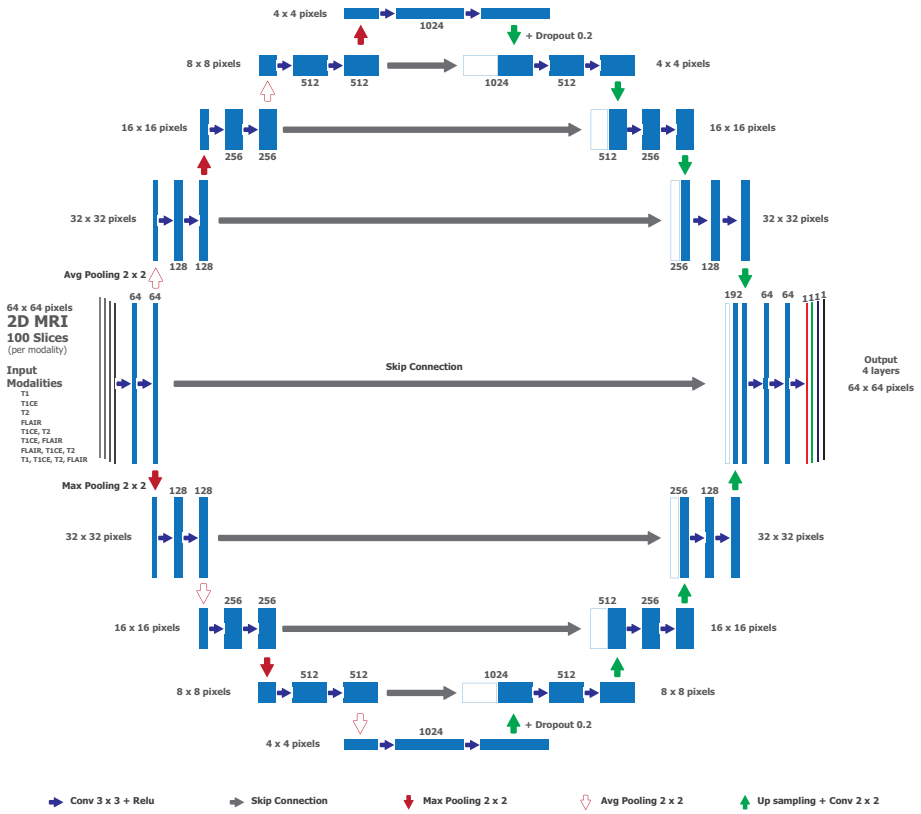

Fig. 1: O-Net architecture

Dataset: The BraTS 2021 training dataset [19], is used in this paper, which includes 3D MRI NIfTI images (T1, T1ce, T2, Flair and segmentation mask) for 1251 patients, each with 155 grayscale 2D slices at a resolution of 240x240 pixels. Limited processing resources led us to downsize and crop each image, using 100 middle 2D slices to remove irrelevant areas and focus on the axial plane. The images were normalized to 0–1. To evaluate model performance, we trained it separately, on each of the four MRI modalities, as each provides unique and complementary information about the tumor and surrounding tissues. We also trained it on various combinations of modalities (T1ce-T2, T1ce-Flair, T1ce-Flair-T2, Flair-T1ce-T2, T1-T1ce-T2-Flair, and T1ce-Flair-T2-T1), reflecting clinical practice where multiple modalities are used for diagnosis and treatment.
Metrics used: Our model is compiled with the categorical crossentropy loss function, suitable for multi-class classification problems like brain tumor sub-regions classification and evaluated using the standard accuracy metric, the Dice Similarity Coefficient (DSC), and the Mean Intersection Over Union (IOU).
Configurations: Our model was trained, validated, and tested with a batch size of 1 using the Adam optimizer (Lr: 0.001). It has $49,889,604$ trainable

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

parameters and was trained for 1,000 epochs on 64×64 pixel images. The dataset included 850 training, 251 validation, and 150 testing images, with early stopping applied (patience: 5).

4 Results and discussion

The results demonstrate that the best values across metrics are achieved by different modalities and their combinations, highlighting the specific strengths of each model (U-Net and O-Net). The results show that the order of combination, specially for (T1ce,Flair and T2) affected the result obtained. While O-Net leads in (T1ce-Flair-T2), U-Net leads in (Flair-T1ce-T2). O-Net outperforms U-Net in T1, while they compete in the T1ce and T2 and Flair. U-Net outperform O-Net in (T1ce-Flair). The high and highest (bold and underlined) DSC scores obtained are represented in table 1, showing an overall DSC (0.7363) belonging to O-Net (T1ce-Flair-T2-T1), underscoring its precision in segmenting different tumor regions. For individual sub-regions, U-Net (T1ce-Flair) combination achieves the highest DSC for necrotic tissue (0.7104) and TC (0.8677). In contrast, O-Net in (T1ce-Flair-T2-T1) excels in capturing ED (0.7808), ET (0.8024), and WT (0.9001), demonstrating its robustness in segmenting larger tumor structures. These results reveal that O-Net configurations generally outperform in overall tumor segmentation, particularly (T1ce-Flair-T2-T1), while U-Net performs effectively in targeted regions, with (T1ce-Flair) leading for necrotic and TC. O-Net with the (T1ce-Flair-T2-T1) combination, offers sev-

Model	DSC	NEC	ED	TC	ET	WT
U-Net(T1ce-Flair)	0.7347	**0.7104**	0.7696	**0.8677**	0.7977	0.8932
O-Net (T1ce-Flair-T2-T1)	**0.7363**	0.6934	**0.7808**	0.8620	**0.8024**	**0.9001**

Table 1: Best DSC result for U-Net and O-Net

eral advances over existing architectures with a TC DSC of 86.2%, comparable to HRSTNet [11] (86.9%) and Swin-UNETR [12] (87.6%). It outperforms other models like Reciprocal Adversarial Learning [7] (85.3%) and 3D Cascade Dynamic Attention U-Net [10] (82.8%), U-Net++ [6] (78.17%), HA-Runet [8] (81.3%). In WT segmentation, O-Net achieves 90.010%, which is competitive with models like Reciprocal Adversarial Learning [7] (90.46%) and surpasses others like HA-Runet [8] (86.7%), Improved ResNet [9] (86.4%) and U-Net++ [6] (87.12%). O-Net achieves a DSC of (80.24%) for ET, outperforming models like U-Net++ [6] (71.92%), HA-Runet [8] (78.7%) and [10] (78.6%) although it falls short of models like Improved ResNet [9] (94.5%). Transformer-based architectures are often complex, whereas O-Net utilizes the simpler U-Net structure to enhance feature extraction and preserve spatial information. Although O-Net may not consistently outperform transformer-based models across all metrics, its balanced and robust performance in WT and TC regions highlights its value in medical image segmentation. Further investigation of alternating pooling strategies in transformer U-Net based architectures could offer valuable insights.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

5 Conclusion

In this work, we presented O-Net, a simple architecture based on U-Net encode-decoder structure with an alternating pooling mechanism. This represents a meaningful extension to the field of U-Net-based models, which typically rely on either Max or Average Pooling independently. This promising area should be further explored by researchers as it offers a valuable avenue for advancing medical image segmentation. Future work should investigate the use of original input dimensions, additional modality combinations, and alternative parameter configurations. Eventually, conducting experiments with other types of pooling methods proposed in the literature is certainly our next step.

References

[1] N. S. Punn and S. Agarwal, "Modality specific u-net variants for biomedical image segmentation: a survey," *Artificial Intelligence Review*, vol. 55, pp. 5845–5889, 10 2022.

[2] I. El Hachimy, D. Kabelma, C. Echcharef, M. Hassani, N. Benamar, and N. Hajji, "A comprehensive survey on the use of deep learning techniques in glioblastoma," *Artificial Intelligence in Medicine*, vol. 154, p. 102902, 2024.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241.

[4] E. J. van Kempen, M. Post, M. Mannil, R. L. Witkam, M. ter Laan, A. Patel, F. J. A. Meijer, and D. Henssen, "Performance of machine learning algorithms for glioma segmentation of brain mri: a systematic literature review and meta-analysis," *European Radiology*, vol. 31, pp. 9638–9653, 12 2021.

[5] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to mri: A survey," *Journal of Imaging*, vol. 8, no. 8, 2022.

[6] N. Micallef, D. Seychell, and C. J. Bajada, "Exploring the u-net++ model for automatic brain tumor segmentation," *IEEE Access*, vol. 9, pp. 125 523–125 539, 2021.

[7] H. Peiris, Z. Chen, G. Egan, and M. Harandi, "Reciprocal adversarial learning for brain tumor segmentation: A solution to brats challenge 2021 segmentation task," *Lecture Notes in Computer Science*, vol. 12962 LNCS, pp. 171–181, 1 2022.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

[8] W. R. Khan, T. M. Madni, U. I. Janjua, U. Javed, M. A. Khan, M. Alhaisoni, U. Tariq, and J.-H. Cha, "A hybrid attention-based residual unet for semantic segmentation of brain tumor," *Computers, Materials Continua*, vol. 76, pp. 647–664, 2023.

[9] M. Aggarwal, A. K. Tiwari, M. P. Sarathi, and A. Bijalwan, "An early detection and segmentation of brain tumor using deep neural network," *BMC Medical Informatics and Decision Making*, vol. 23, p. 78, 4 2023.

[10] H. Byeon, M. Al-Kubaisi, A. K. Dutta, F. Alghayadh, M. Soni, M. Bhende, V. Chunduri, K. S. Babu, and R. Jeet, "Brain tumor segmentation using neuro-technology enabled intelligence-cascaded u-net model," *Frontiers in Computational Neuroscience*, vol. 18, 4 2024.

[11] C. Wei, S. Ren, K. Guo, H. Hu, and J. Liang, "High-resolution swin transformer for automatic medical image segmentation," *Sensors*, vol. 23, no. 7, 2023.

[12] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds. Springer International Publishing, 2022, pp. 272–284.

[13] M. Javan, M. Toews, and M. Pedersoli, "Balanced mixture of supernets for learning the cnn pooling architecture," 6 2023. [Online]. Available: https://arxiv.org/abs/2306.11982

[14] L. Zhao and Z. Zhang, "A improved pooling method for convolutional neural networks," *Scientific Reports*, vol. 14, p. 1589, 1 2024.

[15] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," *ArXiv*, vol. abs/2009.07485, 9 2020.

[16] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Rough Sets and Knowledge Technology*, D. Miao, W. Pedrycz, D. Ślezak, G. Peters, Q. Hu, and R. Wang, Eds. Springer International Publishing, 2014, pp. 364–375.

[17] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018.

[18] A. Deliège, M. Istasse, A. Kumar, C. D. Vleeschouwer, and M. V. Droogenbroeck, "Ordinal pooling," *CoRR*, vol. abs/2109.01561, 9 2021.

[19] D. Schettler, "Brats 2021 task 1 dataset." [Online]. Available: https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1/data