J. Mantas et al. (Eds.)

doi:10.3233/SHTI230432

Brain Tumor Classification and Segmentation Using Dual-Outputs for U-Net Architecture: O2U-Net

Seyed Aman ZARGARI[a], Zahra Sadat KIA[b], Ali Mohammad NICKFARJAM[b,c,1], Daniel HIEBER[d] and Felix HOLL[d]
[a] *Electrical Engineering Department, Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran.*
[b] *Department of Health Information Management and Technology, School of Allied Medical Sciences, Kashan University of Medical Science, Kashan, Iran.*
[c] *Health Information Management Research Center, Kashan University of Medical Science, Kashan, Iran.*
[d] *Institute DigiHealth, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany.*
ORCiD ID: A. M. Nickfarjam https://orcid.org/0000-0003-3782-3038

Abstract. We propose a modified version of the U-Net architecture for segmenting and classifying brain tumors, introducing another output between down- and up-sampling. Our proposed architecture utilizes two outputs, adding a classification output beside the segmentation output. The central idea is to use fully connected layers to classify each image before applying U-Net's up-sampling operations. This is achieved by utilizing the features extracted during the down-sampling procedure and combining them with fully connected layers for classification. Afterward, the segmented image is generated by U-Net's up-sampling process. Initial tests show competitive results against comparable models with 80.83%, 99.34%, and 77.39% for the dice coefficient, accuracy, and sensitivity, respectively. The tests were conducted on the well-established dataset from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjin Medical University, China, from 2005 to 2010 containing MRI images of 3064 brain tumors.

Keywords. Deep Learning, U-Net, Multi-output Model, Brain Tumor Segmentation, Brain Tumor Classification.

1. Introduction

Image segmentation is a fundamental process for many medical research areas. Many recent clinical applications used medical image segmentation, including diagnostic interventions, treatment planning, and delivery [1]. With advances in automated segmentation, this time-consuming and tedious task can be shifted from physicians to AI. Modern approaches for tumor segmentation are trained to detect the tumor area and provide an infrastructure for automatically extracting valuable features. Besides image segmentation Machine-Learning (ML) models can further for classification tasks, e.g., the classification of brain tumors. A prominent category of ML architectures for image classification and segmentation using layered feature extraction is "deep learning" [2]. The focus of these architectures is to automatically discover abstractions from the lowest-level features to the highest-level concepts. Deep learning-based methods have achieved

[1] Corresponding Author: A. M. Nickfarjam, nickfarjam-a@kaums.ac.ir.

superior performance in segmenting region-of-interest (ROI) compared to conventional methods when applied to brain, lung, pancreas, and retinal pathologies [3].

One of the best models based on the dataset used in this article is presented by Diaz-Pernas et al. [4]. The model provides an automatic brain tumor segmentation and classification based on a Convolutional Neural Network (CNN) architecture designed for multiscale processing. Their approach uses three processing pathways that can successfully segment and classify the three kinds of brain tumors in the dataset: meningioma, glioma, and pituitary tumor. U-Net is one of the most efficient and popular deep-learning methods for medical image segmentation. Ronneberger et al. “presented a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture comprises a contracting path to capture context and a symmetric expanding path that enables precise localization” [5]. Another deep learning-based approach for the classification of brain tumors is presented by Rehman Khan et al [6]. Their method comprises three main steps: preprocessing, brain tumor segmentation -using k-means clustering-, and the classification of tumors into their two categories (benign/malignant). However, their method cannot determine the type of brain tumor. This method utilized features do not extract automatically and there is no reason for selecting these type of features. Lin et al. introduced the Dual Swin Transformer U-Net (DS-TransUNet), a U-shaped encoder-decoder-based framework for medical image segmentation. The model is based on the hierarchical Swin Transformer and does not perform classification [7].

2. Methods

This paper proposes a new variation of the U-Net architecture named O2U-Net. This approach combines the powerful segmentation of the U-Net with a classification output. The latter is realized by utilizing U-Nets down-sampling feature extraction process before up-sampling. The classical U-Net architecture was created for the segmentation of medical images. In contrast to other architectures for this use case, it uses no fully connected layers. This approach enables a very accurate segmentation of images with fewer data than comparable architectures. U-Net-based networks consists of two parts: 1) Encoder (down-sampling): In the encoding procedure, sampling layers are used instead of pooling layers. 2) Decoder (up-sampling): After passing through the encoder and extracting its features, the image enters the decoder. The encoding procedure of the network is the same as in usual convolutional networks. It applies convolutional filters to the images, and after each layer of convolution, a layer of RELU activating function is applied, followed by max-pooling. These three layers form a down-sampling block. The number of feature channels doubles in each down-sampling layer. In the up-sampling section, a convolution layer reduces and finds suitable features. It further receives the data which should be added to the image from its peer in the down-sampling section (gray arrow cf. Figure 1.). This is followed by a convolution layer and then a RELU activator. The bottom layer is a convolution with the number of filters equal to the number of classes and a SOFTMAX activator that maps all attribute vectors to one of the classes.

The U-Net takes the original image as an input and outputs its segmentation mask of the same size. In our scenario, the mask contains two areas, one for the tumor and one for healthy tissue/empty parts of the MRI. There is a large consensus that successful training of deep networks requires lots of training samples. U-Net establishes a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. “The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization” [5]. In the proposed O2U-Net architecture, input images of a 256×265 pixel resolution are processed to their tumor mask. Further, the type of tumor is generated as a second output. The model’s left side acts as an encoder, and the right side acts as a decoder, as shown in Figure 1.

As shown in the proposed O2U-Net architecture (cf. Figure 1.), the model branches at the end of the down-sampling process. Besides the normal up-sampling process, a convolution with 256 filters is applied. This output is then given to a global average pooling (GAP) layer, and its output is connected to 3 fully connected layers with 512, 256, and 3 units, respectively. Dropout (DP) layers are placed between the fully connected layers to prevent overfitting. In the last layer, the SOFTMAX activation function is used. Each block consists of two convolution layers on the encoder side with a max pooling layer. Each block on the decoder side starts with the Conv2DTranspose layer, which is applied to the output of the previous block. The last decoder block contains another 1×1 convolution layer. All O2U-Net convolution layers are associated with the RELU activation function except for the last convolution layer, which uses a sigmoid activation function. The learning rate is set to 0.0001, and the batch size to 16. Loss is the function of the dice coefficient segmentation part and is used for the categorical_crossentropy part.

3. Results and Discussion

In this section, the results of tests on the O2U-Net are compared with competing models. The proposed method is based on a 2D deep learning network that works on MRI images of 233 patients obtained from Nanfang Hospital, Guangzhou, China, and the General Hospital, Tianjing Medical University, China, from 2005 to 2010. This dataset contains 3064 slices: 1) meningiomas (708 sections), 2) gliomas (1426 sections) and 3) and pituitary tumors (930 sections) in common views (sagittal, coronal, and axial).

For the train and test, an 80/20 split is used (2452 train images, 612 test images). To enable a precise comparison with Díaz-Pernas et al. [4], we employed the same dataset split and omitted a validation data split. All images have an in-plane resolution of 512×512 pixels with a ratio of 0.49×0.49 mm^2 per pixel. The slice thickness is 6 mm, and the slice gap is 1 mm. The tumor area was labeled by three experienced radiologists manually [4]. Each slice in the dataset has an information structure that contains and the tumor label (1, 2, or 3 for meningioma, glioma, and pituitary tumor, respectively). Also, there is a binary image in which the position of the tumor is 1 and in healthy cases 0. To compare the segmentation results generated by our O2U-Net with other models, the dice coefficient (DC) and sensitivity (SE) were chosen. For classification, only accuracy (AC) was analyzed as it was available for all reverence models. The results are depicted in Table 1 and Table 2. It can be seen that the proposed O2U-Net architecture has superiority over other known methods regarding tumor type classification. With our approach, the network has to learn in a broader space, slightly disrupting the segmentation performance. For the classification, only accuracy was provided as a general metric and derived from [4]. Therefore, the emphasis of the proposed method is on correct classification rather than segmentation.

$$DC = \frac{2TP}{2TP+FP+FN} \quad (1)$$

$$SE = \frac{TP}{TP+FN} \quad (2)$$

$$AC = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

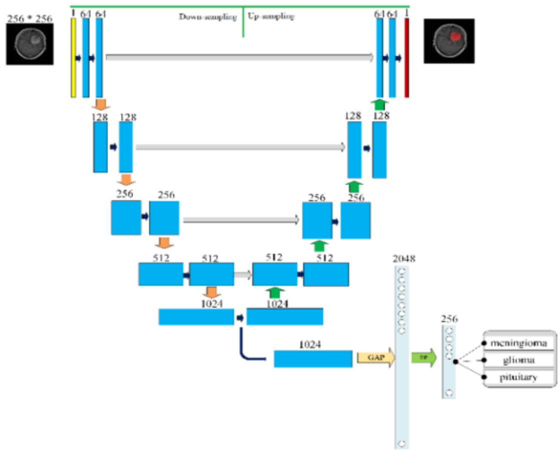

Figure 1. The proposed O2U-Net architecture.

Table 1: Accuracy comparison.

Method (classification)	Accuracy
Pashaei et al. [4]	93.6
Anaraki et al. [4]	94.2
Cheng et al. [4]	94.7
Sultan et al. [4]	96.1
Díaz-Pernas et al. [4]	97.3
Proposed O2U-Net	**98.04**

Table 2: Dice and sensitivity comparison.

Method (segmentation)	Dice
Díaz-Pernas et al. [4]	**82.8**
Proposed O2U-Net	80.83

Method (segmentation)	Sensitivity
Díaz-Pernas et al. [4]	**94.0**
Proposed O2U-Net	77.39

4. Conclusion

This paper proposes a modified version of U-Net called O2U-Net, which can segment and classify tumor images simultaneously. In O2U-Net, a specified 1D output is implemented between the down and up-sampling phases in U-Net architecture. This output relies on a fully connected layer to classify the tumor.

References

[1] Yan X., Tang H., Sun S., Ma H., Kong D., Xie X., After-unet: Axial fusion transformer unet for medical image segmentation, *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022.

[2] Dara S., Tumma P., Feature extraction by using deep learning: A survey. *In Second IEEE International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018.

[3] Khan M. K., Gajendran Y. Lee, Khan M. A., Deep neural architectures for medical image semantic segmentation, *IEEE Access*, 2021.

[4] Díaz-Pernas F. J., Martínez-Zarzuela M., Antón-Rodríguez M., González-Ortega D., A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. *In Healthcare Multidisciplinary Digital Publishing Institute*, 2021.

[5] Ronneberger O., Fischer P., Brox T., U-NET: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*, 2015.

[6] Khan A. R., Khan S., Harouni M., Abbasi R., Iqbal S., Mehmood Z., Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification, *Microscopy Research and Technique*, 2021.

[7] Lin A., Chen B., Xu J., Zhang Z., Lu G., DS-TransUNet: Dual swin Transformer U-Net for medical image segmentation. arXiv preprint arXiv:2106.06716, 2021.