A Study on the Effects of U-Net Skip-Connections on Brain Tumor BraTS Dataset

Jawaher Albanki and Ebrahim Mattar

College of Engineering, University of Bahrain, P. O. Box 32038, Kingdom of Bahrain
20143551@stu.uob.edu.bh; ebmattar@uob.edu.bh

***Abstract* - Skip-connections play an important role in the gradient flow and convergence of U-Net model. They bridge the semantic gap between the encoder and decoder, allowing the recovery of lost information. Previous research has explored various approaches to altering these connections to improve the overall performance of U-Net. In this study, we propose an efficient skip-connections configuration by analyzing their effect on the BraTS dataset. We varied the architecture of the skip-connection by omitting specific layers and testing different combinations of configurations to assess performance. Our findings suggest that eliminating the first skip-connection in U-Net results in the most efficient and optimal performance. We suggest that future research build upon this finding as a base model upon which further improvements can be made.**

Keywords - U-Net, Glioblastoma, Segmentation, Skip-connections.

I. INTRODUCTION

Brain tumors are abnormal growth of cells happening in the brain. There are many types of brain tumor. Cancerous brain tumors are often gliomas, as they are the most frequent. They arise from the glioma cells and have a varying level of aggressiveness upon which they are accordingly classified as either low-grade type or high-grade type. High grade gliomas are rapid and aggressive forming abnormal vessels and often a necrotic core, accompanied by surrounding oedema and swelling [1]. They have high mortality rate. Low grade gliomas grow slower and can be benign as well as cancerous but can evolve into high grade gliomas. Patients need to undergo surgery, chemotherapy, or radiotherapy for treatment. Magnetic resonance imaging MRI is often used to facilitate tumor analysis and localization which is done by segmenting the MRI image [2]. Segmentation of a brain tumor is a challenging procedure and requires high accuracy. Doing this procedure manually is time consuming and due to differences in protocol and practitioners the precise details of the boundaries of the tumor may differ from one clinic to the other. Also, the complexity of structure of a tumor may require parallel consultation to make an informed judgment. These things make it invadable to resort to automation of the procedure [3].

Automatic segmentation of brain tumor aims to help clinicians obtain an objective and a unified segmentation. However, it is difficult to create an algorithm to do the segmentation as the procedure tends to require a level of visual intuition that humans outperform computers at. With the advent of deep learning, convolutional neural networks have showed outstanding human level performance in visual tasks, and therefore have gain popularity. There is various implementation of CNNs in models and architectures that are used for imaging and vision tasks.

U-Net model which is based on CNNs has been extensively utilized in many biomedical image segmentation tasks. We refer the reader to survey by Siddique et al. [4]. It continues to show outstanding results when it serves as the base model. It has been modified in many ways to achieve better performance and continues to give outstanding results as a base model. In this study, we argue that researchers should shift to a slightly different variation from the original U-Net and consider it to be the starting point of their work. We focus on the skip-connection layers and their effects. Different datasets tend to have different signals and unique characteristics. Wang et al. [5] showed the effect of varying the skip-connection layers of U-Net on different datasets. We will repeat their study specifically on BraTS dataset from Decathlon library, as it is the most prominent dataset in the field of brain tumor segmentation.

II. RELATED WORK

In 2018, Myronenko et al. [6] implemented a U-Net which has an additional variational auto encoder branch that was only used during training to regularize the model and achieved the highest score to win the competition. BraTS 2019's winner proposed implementing a cascaded structure of U-Net with two encoders. In 2020 and 2021 versions of the BraTS competition, the models that scored the highest did not apply significant modification to the architecture of U-Net. Instead, they focused on optimizing other aspects such as training setting, augmentation, and loss functions which proves that performance can be enhanced without unnecessary complexities in architecture which tend to produce computational burden.

In particular, winner of 2020, Isensee et al. [7] proposed the model nnU-Net which stands for no new net. Their model is not very different from the original U-Net, except they proposed using region-based deeply supervised training, data augmentation, and added only minor modifications such as the use of ReLU activation with leakiness 0.01 and used instance normalization instead of batch normalization. Winner of 2021, Futrega et al. [8] run an extensive ablation study to test on deep supervision loss, focal loss, decoder attention, drop-out block, and residual connections on U-Net. They have found that the optimal model is U-Net with deep supervision which can be improved by adding an additional input channel with one-hot encoding for foreground, increasing depth and designing a post-processing strategy.

In line with the simple modifications to the U-Net architecture studied in the successful work of [8] and [7], we propose an additional small modification in U-Net that can have the potential to further improve the performance. We argue that U-Net itself with its default number of skip-connection layers is not necessarily the optimal configuration for the nature of BraTS dataset. We inspire this fact from Wang et al. [5] who showed the effect of varying the number of skip-connections built into U-Net on three different datases; Synapse, Glas, and MoNuSeg. They observed that different layers of skip-connections have different contributions to the overall performance and therefore dice score, as shown in figure 1. Their results implied that full connections of all U-Net layers are not necessarily the optimal configuration for all datasets. In fact, for GlaS, the optimal performance was achieved by omitting all skip-connections. This is because not each skip connection setting is effective due to the issue of incompatible feature sets of encoder and decoder stage, even some skip connection may result in negatively effecting the performance. Inspired by their study, we will repeat their experiment of U-Net skip-connection analysis on BraTS dataset. The optimal configuration that we find in this study should serve as a starting point base model to future researchers interested in implementing U-Net architecture.

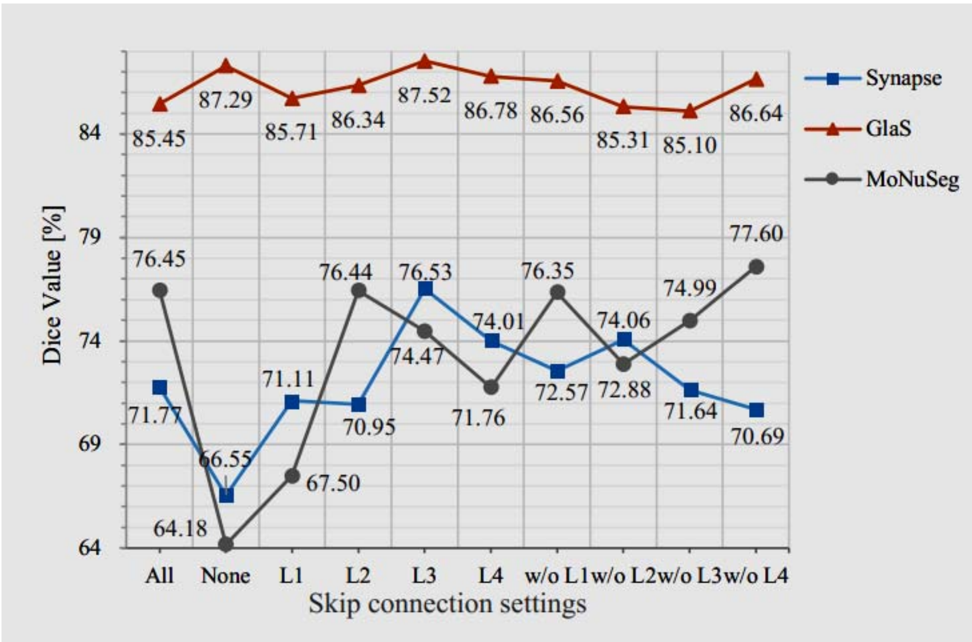

Figure 1. ablation study run by [5] to investigate skip-connection effect. Plot shows the dice score vs. model configuration (original source: [5]).

III. METHOD

The network architecture we used is as proposed on the original paper of U-Net [9]. The model is made of an extracting and a contracting path. Along with skip-connections in between, as shown in figure 2. The contracting path is called an encoder which consists of four blocks, each with two convolutional layers with kernel size 3 followed by ReLU activation and a max-pooling layer at the end with strides 2 and kernel size 2. After each block, the number of feature channels is doubled. Therefore, the assuming a feature of *n*, the input channel size of each block is *n*, *2n*, *4n*, and *8n*, respectively. The expensive path is a decoder which is consistent of also four blocks that are arranged in a symmetrical manner with the encoder blocks. Each block consists of an up-sampling with kernel size of 2 and two convolutional layers of kernel size 3 followed by ReLU activation. A cropping is applied which is necessary due to the loss of border pixels in every convolution. At the final layer, a convolution of size 1. The network was originally designed for 2D images; thus, we modified it into 3D network as done in [10].

The advantage of our proposed modifications is that they do not involve any major architecture alteration, nor any increase in training or inferencing time. There has been a lot of previous work which focused on the modification of U-Net without effecting its architecture majorly or inferencing time [7][8].

A. Skip-Connection Modifications

In addition to the original architecture, we designed another ten architectures all of which have only the skip connections being modified. In the first variation, we remove all skip connection layers which results in a network identical to FCN model as proposed in [11]. We designed four other networks that keep only a single layer of skip-connection the first, second, third, and fourth which we denote as L1, L2, L3, and L4, respectively. We also designed four other variants to include all skip-connection layers, except one layer. In each variant we removed a different skip-connection layer and denoted the models as no-L1, no-L2, no-L3, and no L4.

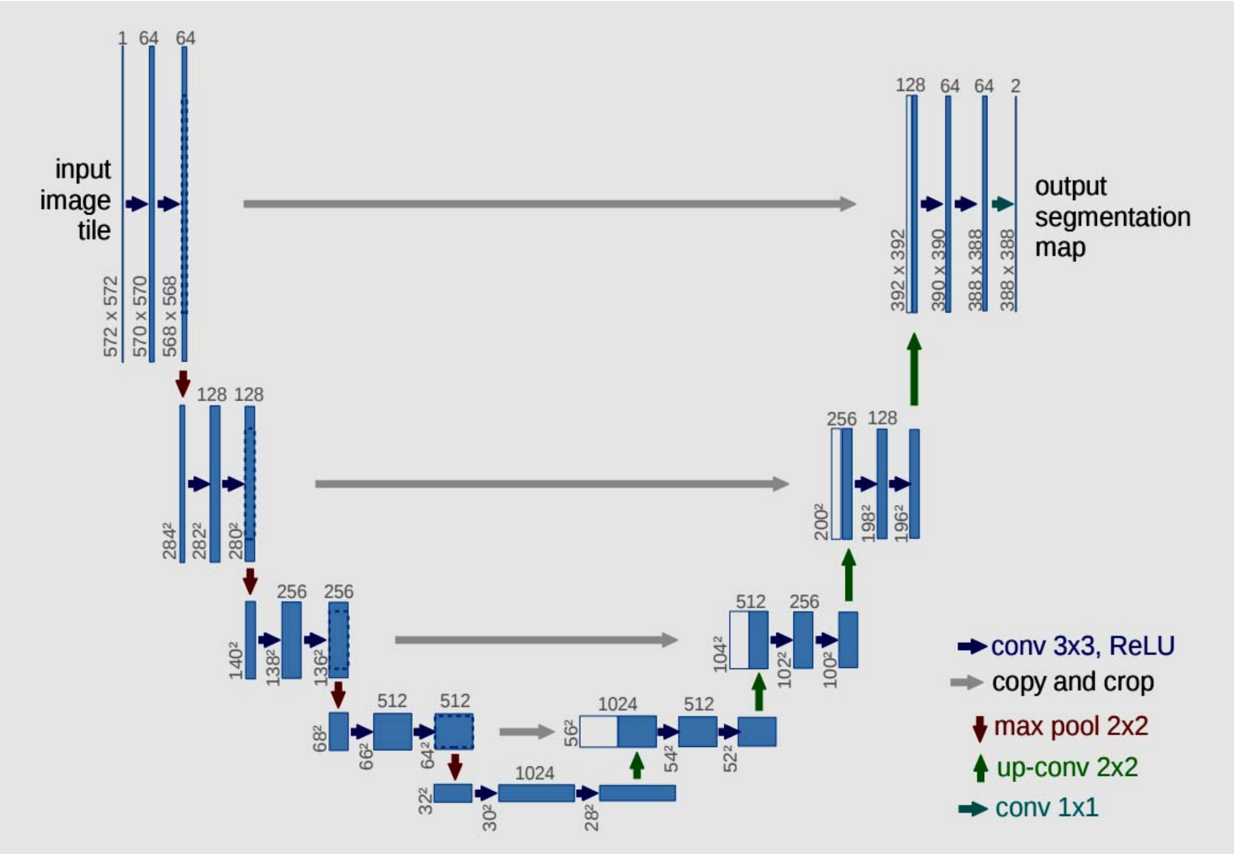

Figure 2. U-Net architecture with contracting path (left), expensive path (right), and skip-connections (middle). (Original source [9]).

B. Dataset

We obtained BraTS dataset from Decathlon library [12]. The BraTS dataset consists of inputs with four modalities of MRI scan. The modalities are T1-w, T2-w, T1gd, FLAIR. There are 750 images with 484 for training and 266 for testing.

The dataset is originally from 2016 and 2017 BraTS competitions. The challenge in those images is their complexity and heterogeneously located target. The output is a dense predication mask of three categories tumor core (TC), whole tumor (WT), and enhancing tumor (ET). This labeling is a protocol followed by BraTS and created by the dataset collectors. Since the data is collected from various institutions and clinics with various protocols and practices, creating a heterogenous dataset.

IV. LOSS FUNCTION

The loss function used is Dice loss function [13] which is computed voxel-wise. It is expressed as

$$\mathcal{L}(G,Y) = 1 - \frac{2}{J}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I} G_{i,j}Y}{\sum_{i=1}^{I} G_{i,j}^2 + \sum_{i}^{I}} \tag{1}$$

Where I denote voxels; J denotes classes; Y and G are probability of output and ground through output, respectively.

V. IMPLEMENTATION DETAILS

Our study is implemented using PyTorch and MONAI and trained using Google Colab Pro+ which uses a V100 NIVIDIA GPU limited the number of compute units which if a user exceeds may experience a drop in performance, we kept the number of epochs to 100 and frequency of validation to once per every 10 epochs.

The input has a 48-feature size with image re-sized to 128x128x128 pixels. The optimizer used is ADAM optimizer with learning weight of 0.0001 and weight decay of 0.00001. The training inputs to the model were normalized to speed up training. The images were resampled to a pixel-dimension (1,1,1) with interpolation mode being bilinear and nearest. For augmentation, random cropping, random flipping with probability 0.5, random scaling intensity of factor 0.1 and probability 1, and random intensity shifting of off-set 0.1 and probability 1 were applied to the training dataset. The dataset used is taken from Decathlon library [12]. A seed of 0 for deterministic output was set by MONAI built-in determinism function to make the results of the experiment reproducible and consistent.

VI. EXPERIMENTAL RESULTS

All models were trained on the same dataset with the same training settings. The aim of this experiment is to study the effect of skip-connection and each layer's contribution on BraTS dataset. The results are shown in Table 1. Table shows validation results. The first column indicates the model being trained. To illustrate what each model label indicates, "U-Net" is the original model with full connection, "None" is no skip-connections, "L1 only" is only first skip-connection is included while all others are not, and "No L1" is all skip-connection layers are included expect L1. The same terminology convention applies to L2, L3, and L4. In the table, TC indicates tumor core, WT whole tumor, ET enhancing tumor, and final column is the average dice score.

The results of average dice in table 1 are plotted as a graph in figure 3. An example of segmented outputs from the highest average dice scoring model "No L1" is shown in figure 4.

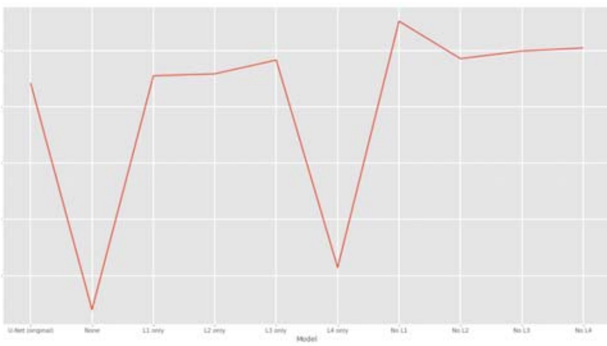

Figure 3. Average Dice Score vs U-Net model skip-connection modification.

TABLE I. DICE SCORE RESULTS OF VALIDATION

Modification	Dice score			
	TC	*WT*	*ET*	Average
U-Net (original)	0.7458	0.8836	0.4959	0.7084
None	0.7093	0.8409	0.3336	0.6279
L1 only	0.7518	0.8853	0.4960	0.7110
L2 only	0.7495	0.8867	0.4990	0.7117
L3 only	0.7681	0.8815	0.5001	0.7166
L4 only	0.6735	0.8717	0.3836	0.6429
No L1	0.7732	0.8985	0.5196	0.7304
No L2	0.7484	0.8915	0.5113	0.7171
No L3	0.7525	0.8888	0.5182	0.7198
No L4	0.7583	0.8823	0.5222	0.7209

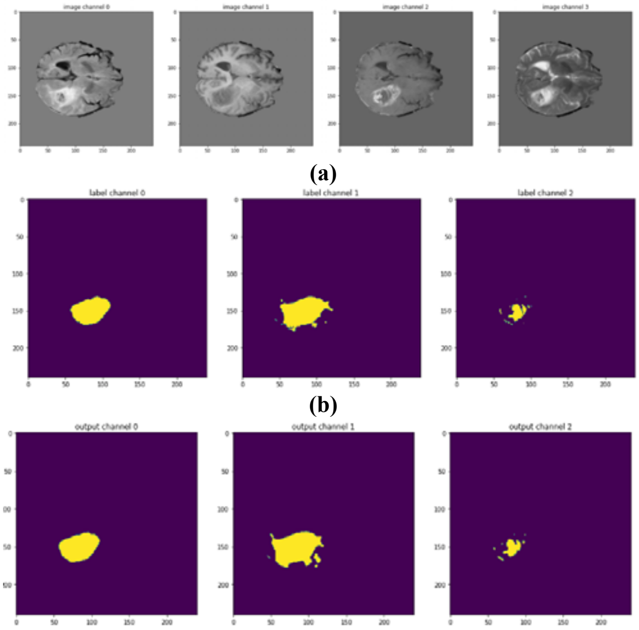

Figure 4. a sample of segmentations from "no L1" model. (a) The four MRI input modalities. (b) ground truth segmentation. (b) model output predicted segmentation.

VII. DISCUSSION

It is clear from Table 1 that the original U-Net is not an optimal design for BraTS dataset in terms of skip-connection. Generally, the tumor of least dice score is the enhancing tumor ET. And the tumor that is easiest to detect is the whole tumor WT which is normally the case in all models trained on BraTS due to the imbalance in the dataset which is expected in medical imaging dataset. Though it can be noticed that for model "no L4" the ET is at its highest score indicating that the fourth layer of skip-connection may have a negative impact on this mask type.

It is important to highlight that understanding what makes omitting the first skip-connection beneficial for U-Net in BraTS dataset requires deeper analysis of the dataset and U-Net architecture that is beyond the scope of this study. However, one possible explanation could be that eliminating the first-skip connection helped the model in focusing more on important features of the dataset, leading to an improved performance, especially on TC and WT.

It can be inferred that to achieve optimal performance on all masks (TC, WT, and ET), one may need to build rather a cascaded network than a single one to classify each different network. For example, an optimal configuration may be using two decoder branches one for TC and WT segmentation, and the other branch for ET segmentation, where the first branch feeds from all layers except the first, and the second branch feeds from all layers except the fourth. In fact, cascading U-Net structure has been proposed several times in previous works, such as [14] and [15]. However, the use of a cascaded U-Net may compromise computational efficiency and speed of training. As such, to balance between high performance and computational cost, U-Net with “No L1” itself offers the highest average score. In fact, it also reduces computational cost slightly, in that the concatenation between two largest feature maps in the first layer encoder and decoder is eliminated.

It is important to mention though that our experiment faced some limitations. We had very scarce computational resources upon conducting this study. Therefore, study was done on quite a small number of epochs which is 100. A larger number of epochs could give more accurate results of each model’s performance. After longer epochs, we may or may not observe performances reflecting table 1 as the results may tend to stabilize further and give different outcomes.

VIII FUTURE WORK

Due to the computational resource limitation of this experiment, we leave to future studies to investigate the implementation of a UCTransNet as proposed in [5] instead of the skip-connections and prove whether it is more effective than skip-connections. Though, it must be noted that transformers suffer expensive computational cost, even if they perform better than CNNs and are able to capture long-range dependencies.

Also, we propose for future work to implement this analysis on previous U-Net variants that were used on BraTS datasets to confirm the consistency of the experiment results on other modifications of U-Net. For example, omitting the first skip-connection in nn-Net or other U-Net based models such as [6], and [8]. Finally, we leave for future work the implementation of the same ablation studies on datasets of BraTS from years other than 2016 and 2017 which were used in this study to confirm whether the same thing applies to other versions of the dataset, as datasets tend to have different signals resulting in different responses from a model.

IX. CONCLUSION

We conducted a comprehensive analysis of U-Net skip-connection on BraTS dataset to investigate their effect on the dataset. Our experiment which involved the testing of different combinations of skip-connections revealed that certain layers were redundant and could slightly impair performance, confirming that the original U-Net architecture does not have the optimal skip-connection configuration for this particular dataset. We found that the most efficient and optimal performance is achieved by eliminating the first skip-connection from U-Net which we labeled as “no L1”. Therefore, we recommend using "no L1" as the base U-Net model for future studies involving U-Net on the BraTS dataset. Moreover, we highlight that existing state-of-the-art models can also benefit from this base model. Therefore, our findings have practical implications for researchers seeking to improve the performance of U-Net on the BraTS dataset.

REFERENCES

[1] Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of mri-based medical image analysis for brain tumor studies. Physics in med. and biol. 58(13) (2013) R97

[2] DeAngelis, L.M.: Brain tumors. New Engl. Journ. of Med. 344(2) (2001) 114{123)

[3] 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 TMI 34(10) (2015) 1993-2024

[4] Siddique, Nahian, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. "U-net and its variants for medical image segmentation: A review of theory and applications." Ieee Access 9 (2021): 82031-82057.

[5] Wang, Haonan, Peng Cao, Jiaqi Wang, and Osmar R. Zaiane. "Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 3, pp. 2441-2449. 2022.

[6] Myronenko, Andriy. "3D MRI brain tumor segmentation using autoencoder regularization." In International MICCAI Brainlesion Workshop, pp. 311-320. Springer, Cham, 2018.

[7] Isensee, Fabian, Paul F. Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H. Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature methods 18, no. 2 (2021): 203-211.

[8] Futrega, Michał, Alexandre Milesi, Michał Marcinkiewicz, and Pablo Ribalta. "Optimized U-Net for brain tumor segmentation." In International MICCAI Brainlesion Workshop, pp. 15-29. Springer, Cham, 2022.

[9] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham, 2015.

[10] Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation." In International conference on

[11] medical image computing and computer-assisted intervention, pp. 424-432. Springer, Cham, 2016.

[12] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440. 2015.

[13] Antonelli, Michela, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens et al. "The medical segmentation decathlon." Nature communications 13, no. 1 (2022): 1-13.

[14] Sudre, Carole H., Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations." In Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 240-248. Springer, Cham, 2017.

[15] Talha Bukhari, Syed, and Hassan Mohy-ud-Din. "E1D3 U-Net for Brain Tumor Segmentation: Submission to the RSNA-ASNR-MICCAI BraTS 2021 Challenge." arXiv e-prints (2021): arXiv-2110.

[16] Jiang, Zeyu, Changxing Ding, Minfeng Liu, and Dacheng Tao. "Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task." In International MICCAI brainlesion workshop, pp. 231-241. Springer, Cham, 2019.