Sub-Region Segmentation of Brain Tumors from Multimodal MRI Images Using 3D U-Net

Ammar Alhaj Ali [iD], Rasin Katta [iD], Roman Jasek [iD], Bronislav Chramco [iD], and Said Krayem [iD]

Faculty of Applied Informatics, Tomas Bata University in Zlin, Zlin, Czech Republic
aali@tech-ai.net
{katta,jasek,chramcov,krayem}@utb.cz
http://www.fai.utb.cz/

Abstract. Accurate segmentation of brain tumors from the magnetic resonance image (MRI) is an essential step for radionics analysis as well as finding the tumor extension is so necessary to plan the best treatment to improve the survival rate. Manually extracting sub-regions of the brain tumor from MRI is a tedious process and time-consuming, as the complex brain tumor images require extensive human expertise. In recent years, deep learning models have proved effective in medical image segmentation tasks. In brain tumor segmentation, the 3D multimodal MRI poses some challenges such as computation and memory limitations. This study aims to develop a deep learning model using 3D U-Net for brain tumor segmentation. The segmentation results on BraTS 2020 dataset show that the proposed model achieves promising performance.

Keywords: 3D image segmentation, 3D U-net, brain tumor segmentation, BraTS, deep learning.

1 Introduction

According to the American national brain tumor society about 700,000 people in the United States are living with a brain tumor, and approximately 85,000 more will be diagnosed in 2021. Approximately 70% of all brain tumors are benign [1]. Brain tumors can be fatal, significantly affect the quality of life, and change everything for a patient and their loved ones. The early detection and accurate diagnosis of tumor regions are vital and help in the treatment plan and increase the patient's chances are for successful treatment [2].

Glial cells are a type of cell that supports cells of the brain and the spinal cord, gliomas are tumors that start in these cells [3]. Gliomas are the most common and aggressive malignant tumors mostly occur in the brain and, rarely, in the spinal cord [4], and represent 40% of tumors of the central nervous system, and 80% of all malignant brain tumors [5]. Gliomas tumors contain different heterogeneous histological sub-regions, including peritumoral edema (ED), necrotic core (NCR), enhancing (ET), and non-enhancing tumor core (NET). These sub-regions reflect important and clinically relevant information [6]. Gliomas can

be further classified into low-grade (LGG) and high-grade (HGG). High-grade gliomas are an aggressive class of malignant brain tumors, usually requiring surgery and radiotherapy, and the prognosis for survival is poor [7].

Magnetic Resonance Imaging (MRI) has been extensively using for brain tumor diagnosis over the last few decades; it can effectively capture the intrinsic heterogeneity of Gliomas using multimodal scans with varying intensity profiles [5,6,7]. Usually, several 3D MRI scans are required to emphasize different tissue properties and areas of tumor spread [8,9]. Multimodal MRI scans consist of: 1- Native T1-weighted scan (T1). 2- Post-contrast T1-weighted scan (T1Gd). 3- Native T2-weighted scan (T2). 4- T2 Fluid Attenuated Inversion Recovery scan (T2-FLAIR).

Accurate segmentation of various sub-regions of gliomas from multimodal MRI scans has important clinical relevance in the diagnosis and treatment of brain tumors, and it is one of the most difficult tasks in analyzing medical images due to their heterogeneous appearance and shape [9]. Segmentation of the brain tumor manually to diagnose cancer from a big number of clinically produced MRI images is a challenging and time-consuming job, due to poor contrast, unknown noise, inhomogeneity, and unclear borders that are in clinical images, so sometimes manual segmentation doesn't give correct results [10], as well, brain tumors can appear anywhere in the brain, and their shape and size vary greatly. Furthermore, the intensity profiles of tumor regions are largely overlapped with healthy parts. Due to the challenge of brain tumor segmentation, many researchers have proposed various methods to solve the problem of brain tumor segmentation [11].

In recent years, deep learning models have emerged as a promising solution for segmentation in medical imaging. Generally, Convolutional neural networks (CNNs) have been outperformed traditional algorithms in computer vision tasks [12,13]. These days, biomedical 2D images can be segmented with accuracy close to human performance by CNNs. Due to this success; a lot of researches have been made to apply 3D CNNs to biomedical volumetric data [14].

U-Net has been offering impressive potential in segmenting images[15], to an extent that it has become the standard in image segmentation like the medical images, even with a small amount of labeled training data. U-Net - called in this way due to its U shape - is a type of convolutional neural network CNN that performs the task of semantic segmentation [16]; it consists of the encoder-decoder architecture. The encoder part extracts spatial features from the images, and the corresponding decoder and a decoder that forms the segmentation map from the encoded features. 3D U-Net is considered an extension of the previous U-Net (2D) and consists of the encoder-decoder architecture. Due to its out-performance and simple architecture, 3D U-Net is gradually being studied and applied for medical image segmentation tasks [17,18].

The contributions of this study can be summarized as follows: 1. Presenting a deep learning model for 3D segmentation of brain tumors. 2. Proposing a 3D U-Net architecture for segmentation of brain tumor and its sub-regions from

multimodal MRI scans. 3. Exploring the potential of the 3D U-Net architecture to achieve good segmentation with a small amount of data.

The rest of the paper is organized as follows: Section 2 presents the proposed methodology like network architecture, loss function and evaluation metrics. Section 3 presents training process. Section 4 provides the experimental results. Finally, Section 5 summarizes the study and draws a conclusion.

2 Methodology

Our methodology starts with exploration of dataset to discover and understand the data. Then we move on model architecture and explain deeply proposed 3DU-Net architecture. Then we proceed to explain how model is evaluated and which loss function is used for the optimization of the model. After that we move on with strategies to improve the model, such as data augmentations.

2.1 Database

The dataset in this study comes from the Multimodal Brain Tumor Segmentation Challenge (BraTS 2020) [19]. This dataset consists of 292 HGG and 77 LGG cases, for a total of 369 cases. Multimodal MRI scans are provided in 3D volumes; MRI volume is being composed of 155 slices of 240x240 resolutions. Multimodal MRI scans consisting of native T1weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and fluid-attenuated inversion recovery (FLAIR) [19,20]. See Figure 1.
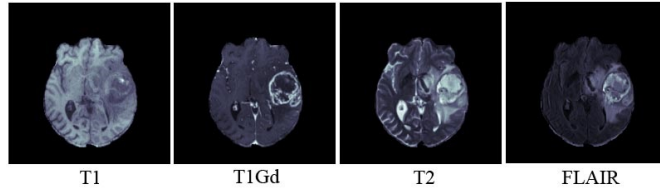

Fig. 1. A visualization of Brain Tumor Segmentation Challenge (BraTS) dataset

All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. the labels given in the dataset are: • '1' for Necrotic and Non-Enhancing Tumor (NCR/NET). • '2' for Peritumoral Edema (ED), • '4' for GD-Enhancing Tumor (ET), • 0 for everything else.

In this study, three glioma sub-regions are used to evaluate the segmentation processes that are WT, TC, and ET. • Whole tumor (WT): Is segmented by using the combination of labels 1, 2, and 4. • Tumor Core (CT): Is segmented by using the combination of labels 1 and 4. • Enhancing tumor (ET): Is segmented from label 4. See Figure 2.
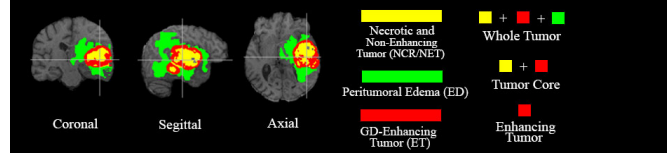
Fig. 2. Glioma sub-regions

2.2   Deep Learning Model Architecture

The data that is fed to deep learning model consists of images with slices. Each three dimensional regions are called voxels. A typical input is 240x240x155 voxels, where 155 is the number of slices and 240x240 is resolution of each image. The complete U-Net architecture is illustrated in Figure 3. There are two parts of network.
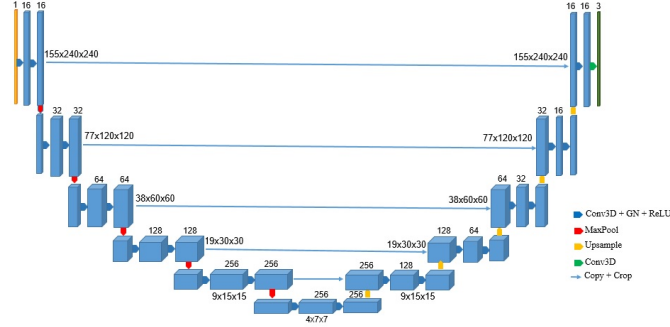

Fig. 3. Proposed U-Net architecture

The first part (left part) works as an encoder to extract the features. This part consists of convolution block and max pooling, a convolution block involves 3D convolution (conv3D), Group Normalization(GN), and rectified linear(ReLu) layer, this block is repeated two times. GN is used as alternative to Batch Normalization (BN) due to BN's error usually increases when use smaller batch size, this limit of usage BN comes for training large models and transfer features in detection and segmentation that require small batches constrained by memory consumption [21]. The max pooling layer (MaxPool3d) used for downsampling, it has the size 2x2x2 voxels with a stride 2 in each dimension [22]. After each downsampling the number of feature channels is double. The second part (right part) of the network acts as a decoder to aggregate the features and the segmentation mask. Each step in decoder consists of upsample, channel concatenation with the respective cropped feature map from encoder, and the two convolution blocks. Upsample layer increases the size of the feature map by inserting zero-

padding around each pixel followed by convolution on the padded image. The final convolution layer is designed to get the three classes from 16 features.

2.3   Evaluation Metrics

The evaluation metrics measure the success of a model. In semantic segmentation, a successful prediction is the one that maximizes the overlap between the predicted segmented and the true object (ground-truth). The Dice and Jaccard indices are two metrics used for evaluating the overlap and quality of model. The Jaccard index is also known as Intersection over Union (IoU). Both metrics are described in the formulas below:

$$(DiceScore(X,Y)) = \frac{2X \mid X \cap Y \mid}{\mid X \mid + \mid Y \mid} \tag{1}$$

$$(Jaccard(X,Y)) = IoU(X,Y) = \frac{\mid X \cap Y \mid}{\mid X \cup Y \mid} \tag{2}$$

Where X and Y are the predicted segmented and the ground-truth. Where $\cap$, $\cup$ denote the intersection and union operators, respectively. Both the Dice and IoU indices are ranged from 0 (there is no overlap) to 1 (X and Y match perfectly). The performance of the model is also evaluated by different metrics to analyse how well the model performs on test data. Sensitivity and specificity are calculated for determining over- or under-segmentations of tumor sub-regions, these metrics are based on a confusion matrix. Sensitivity (True Positive Rate) refers to the proportion of the positives that are correctly identified; this can be calculated by the following equation.

$$(Sensitivity) = \frac{TP}{TP + FN} \tag{3}$$

Where TP is the true positives that are the number of instances that are positive and are correctly identified, and FN refers to the false negatives that are the number of positive cases that are incorrectly identified as negative. Specificity (True Negative Rate) refers to the proportion of the negative label being true, this can be calculated using the following equation.

$$(Specificity) = \frac{TN}{TN + FP} \tag{4}$$

Where TN is the number of true negatives that are classified as negative, and FP is the number of false positives that refers to the negative instances that are incorrectly classified as positive cases.

2.4   Loss Function

Usually, segmentation models are trained with a cross-entropy loss function [23], however, the findings may be not ideal in terms Dice score. As an alternative, we

can employ Dice loss function to train our network that has given good results in segmentation challenges and it appears more frequently in many papers and projects. In order to formulate a Dice loss function that can be minimized, we simply use one Dice. The Dice loss function can be calculated for each class by the following equation.

$$L(Dice) = 1 - \frac{2 \sum PtruePpred}{\sum Ptrue + \sum Ppred + \varepsilon} \qquad (5)$$

Where Ptrue and Ppred are the ground truth and the predicted labels respectively. While cross-entropy addresses every pixel as independent prediction, dice-score looks at the resulting mask in a more holistic way and uses the predicted probabilities instead of threshold and converting them into a binary mask. Therefore, the combination of the two loss functions is often considered as a good practice and leads to more solid and segmentation models. The final loss function can be calculated as follows.

$$L = L(entropy) + L(dice) \qquad (6)$$

2.5   Data augmentation

Usually, deep learning models require a lot of data for training, in general, the more the data, the better the performance of the model. But sometimes acquiring huge amounts of data is reflected as a major challenge in deep learning. Basically, data augmentation is a technique that is very commonly used technique to increase the dataset size by adding slightly modified copies of already existing data without adding new data. This technique is one of the most practical approaches to enhance the model performance without spending days collecting data, especially when the collection, labeling, and segmentation of data are so hard and time-consuming. In addition, data augmentation helps to avoid overfitting. In this study, some data augmentation techniques were used to improve model performance, such as horizontal flip, contrast, gamma, and brightness.

3   Training Process

the experiments are implemented using the PyTorch framework, and the 3D U-Net model training was done in 400 epochs with a batch size of 1. The models was optimized via Adam optimizer [24] with a learning rate of 0.0005. The size of the training input was 240 x 240 x 155, and the dataset was split into three sets: a training set, validation set and test set with a ratio of 70:15:15. The training was executed on a Graphical Processing Unit GPU NVIDIA TESLA P100 having the Specifications: 16 GB memory, 3584 CUDA cores, 1190 MHz core clock, and with a memory bandwidth of 732.2 GB/sec [25].

4   Experimental Results and Discussion

We trained the model using the BraTS 2020 training set, then applied the trained model for prediction on the validation and test set. Table1 presents the quantitative analysis of the proposed model. We have obtained mean Dice scores of WT, TC, and ET as 0.9011, 0.8057, and 0.7645, respectively, on the validation dataset, while 0.8965, 0.8316, and 0.7503 on the test dataset. Also in Table 1, IoU, sensitivity, and specificity are presented on training, validation, and the test datasets.

Table 1. BraTS 2020 training, validation and testing results. Mean average scores on different metrics.

Metrics	Team	Training	Validation	Test
Dice	WT	0.9264	0.9011	0.8965
	TC	0.8804	0.8057	0.8316
	ET	0.7435	0.7645	0.7503
IoU	WT	0.8655	0.8268	0.8183
	TC	0.8015	0.7166	0.7387
	ET	0.6521	0.6616	0.6447
Sensitivity	WT	0.9219	0.9192	0.8962
	TC	0.8707	0.8280	0.8454
	ET	0.7696	0.8127	0.8220
Specificity	WT	0.9994	0.9991	0.9992
	TC	0.9996	0.9994	0.9994
	ET	0.9998	0.9997	0.9995

Table 2 shows the challenge testing set results. The proposed model achieves a good score for Dice and IoU. The high median values of the testing set show that the testing set may contain some low cases resulting in lower average scores.

Table 2. Dice and IoU for testing set.

	Dice			IoU		
	WT	TC	ET	WT	TC	ET
Mean	0.8965	0.8316	0.7503	0.8183	0.7387	0.6447
Median	0.9083	0.8985	0.8308	0.8319	0.8158	0.7106
StdDev	0.0643	0.1636	0.2405	0.1009	0.1984	0.2391
25quantile	0.8509	0.7971	0.7099	0.7405	0.6627	0.5502
75quantile	0.9464	0.9295	0.8972	0.8983	0.8683	0.8135

To observe the distribution of all results of the validation and test datasets clearly, the indicators of Dice, IoU, Sensitivity, and specificity are gathered into box plots, See Figure 4 and Figure 5.
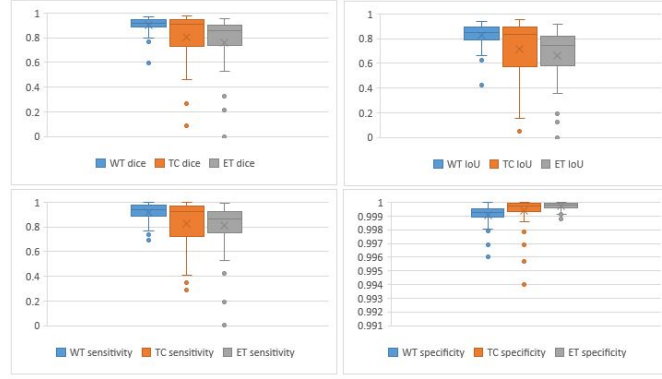

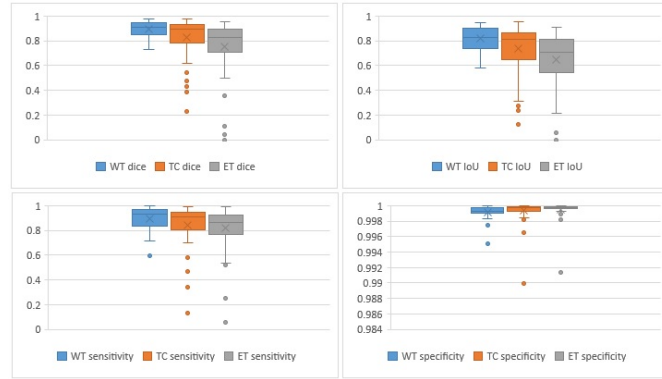**Fig. 4.** Quantitative metrics of validation dataset

Fig. 5. Quantitative metrics of testing dataset

Figure 6 shows the loss, Dice, and IoU curves for proposed model.

To provide qualitative results of our model, five-segmented images from test dataset are shown in Figure 7.

5   Conclusion

Manual segmentation of brain tumors from MRI is a time-consuming and hard work, often lead to inaccurate results due to human tiredness, size of MRI slices
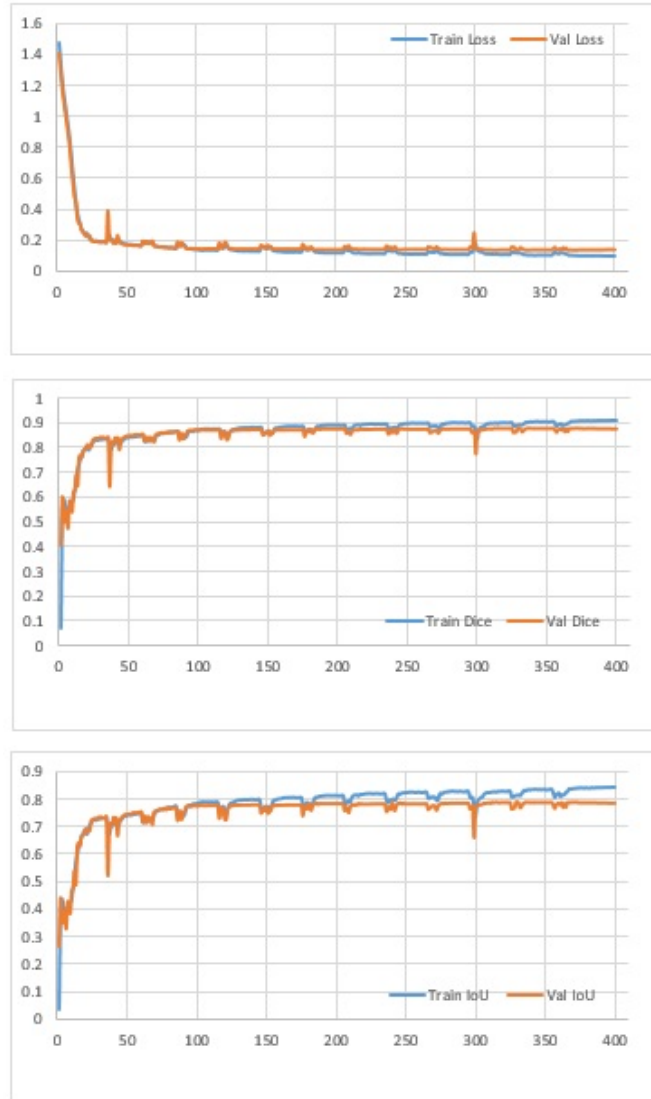
Fig. 6. Curves of proposed model. (a) Loss curves. (b) Dice curve. (c) IoU curve.
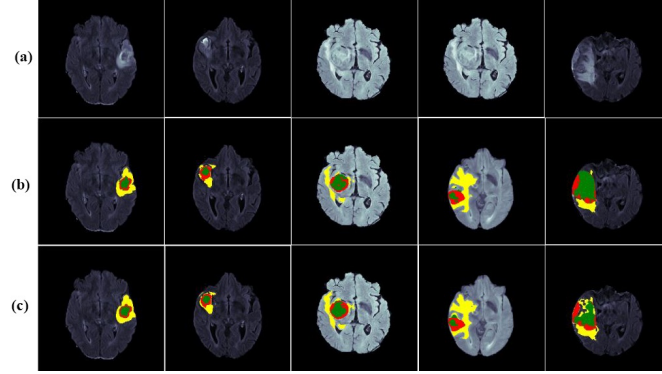
Fig. 7. Curves of proposed model. (a) Loss curves. (b) Dice curve. (c) IoU curve.

per patient. Automated methods for brain tumor segmentation is likely to have major clinical effect, since automated decision-making reduces human bias, errors and is faster. In this paper, we present 3D-UNet architecture for an automatic brain tumor segmentation. This model has been trained and evaluated on the BraTS 2020 Challenge dataset and achieved an average Dice scores of 0.8965, 0.8316, 0.7503 and IoU scores of 0.8183, 0.7387, 0.6447 for WT, TC, and ET respectively. In the future work, we will continue to improve the structure of the network and use GPU with higher specifications to improve the segmentation performance.

References

1. Quick Brain Tumor Facts. https://braintumor.org/brain-tumor-information/brain-tumor-facts/ (accessed Mar. 23, 2021).
2. Khalil, H. A., Darwish, S., Ibrahim, Y. M., & Hassan, O. F. (2020). 3D-MRI Brain Tumor Detection Model Using Modified Version of Level Set Segmentation Based on Dragonfly Algorithm. Symmetry, 12(8), 1256.
3. Glioma . https://www.cancerresearchuk.org/about-cancer/brain-tumours/types/glioma-adults (accessed Mar. 23, 2021).
4. Glioma . https://rarediseases.org/rare-diseases/glioma/. (accessed Mar. 23, 2021).
5. Banerjee, S., & Mitra, S. (2020). Novel volumetric sub-region segmentation in brain tumors. Frontiers in computational neuroscience, 14, 3.
6. Sun, L., Zhang, S., Chen, H., & Luo, L. (2019). Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning. Frontiers in neuroscience, 13, 810.
7. Qamar, S., Ahmad, P., & Shen, L. (2020). HI-Net: Hyperdense Inception 3D UNet for Brain Tumor Segmentation. arXiv preprint arXiv:2012.06760.
8. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., ... & Eaton-Rosen, Z. (2018). 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.

9. Feng, X., Tustison, N. J., Patel, S. H., & Meyer, C. H. (2020). Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. Frontiers in computational neuroscience, 14, 25.
10. Srinivas, B., & Sasibhushana Rao, G. (2020). Segmentation of Multi-Modal MRI Brain Tumor Sub-Regions Using Deep Learning. Journal of Electrical Engineering & Technology, 15, 1899-1909.
11. Chen, W., Liu, B., Peng, S., Sun, J., & Qiao, X. (2018, September). S3D-UNet: separable 3D U-Net for brain tumor segmentation. In International MICCAI Brainlesion Workshop (pp. 358-368). Springer, Cham.
12. Zhao, W., Jiang, D., Queralta, J. P., & Westerlund, T. (2020). MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net. Informatics in Medicine Unlocked, 19, 100357.
13. Ali, A.A., Chramcov, B., Jasek, R., Katta, R., Krayem, S. (2021). Classification of Plant Diseases Using Convolutional Neural Networks. In: Silhavy, R. (eds) Artificial Intelligence in Intelligent Systems. CSOC 2021. Lecture Notes in Networks and Systems, vol 229. Springer, Cham. https://doi.org/10.1007/978-3-030-77445-5-24
14. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016, October). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham.
15. Ali, A.A., Chramcov, B., Jasek, R., Katta, R., Krayem, S., Kadi, M. (2021). Detection of Steel Surface Defects Using U-Net with Pre-trained Encoder. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) Software Engineering Application in Informatics. CoMeSySo 2021. Lecture Notes in Networks and Systems, vol 232. Springer, Cham. https://doi.org/10.1007/978-3-030-90318-3-18
16. Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks, 121, 74-87.
17. Li, C., Chen, W., & Tan, Y. (2020). Point-Sampling Method Based on 3D U-Net Architecture to Reduce the Influence of False Positive and Solve Boundary Blur Problem in 3D CT Image Segmentation. Applied Sciences, 10(19), 6838.
18. Ngo, D. K., Tran, M. T., Kim, S. H., Yang, H. J., & Lee, G. S. (2020). Multi-Task Learning for Small Brain Tumor Segmentation from MRI. Applied Sciences, 10(21), 7790.
19. BraTS2020 Dataset. https://www.kaggle.com/awsaf49/brats20-dataset-training-validation (accessed Mar. 23, 2021).
20. Multimodal Brain Tumor Segmentation Challenge 2020: Data. https://www.med.upenn.edu/cbica/brats2020/data.html (accessed Mar. 23, 2021).
21. Wu, Y., & He, K. (2018). Group normalization. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).
22. https://pytorch.org/docs/stable/generated/torch.nn.MaxPool3d.html
23. Y. Ho and S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling," IEEE Access, vol. 8, pp. 4806–4813, 2020, doi: 10.1109/ACCESS.2019.2962617.
24. D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," Dec. 2015, Accessed: Mar. 13, 2021. [Online]. Available:https://arxiv.org/abs/1412.6980v9.
25. NVIDIA® TESLA® P100 GPU Accelerator. https://images.nvidia.com/content/tesla/pdf/nvidia-tesla-p100-PCIe-datasheet.pdf. Accessed 2021-01-16.