Brain MRI 3D Segmentation using Patch Spatially Localized Network Tiles utilizing Pipelines on GPU Memory

1st Pandapotan Siagian
Department of Informatics
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
siagian.p@gmail.com

2nd Riyanarto Sarno
Department of Informatics
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

3rd Chastine Fatichah
Department of Informatics
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
chastine@if.its.ac.id

4th Aziz Fajar
Department of Informatics
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
azizfajar519@gmail.com

5th Muhammad Ibadurrahman Arrasyid Supriyanto
Department of Informatics
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ibad.arrasyid@gmail.com

RCID

2023 1st International Conference on Advanced Engineering and Technologies (ICONNIC) | 979-8-3503-0648-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICONNIC59854.2023.10467355

Abstract

Background and objective
Quantitatively detailed WBTS is a very important role in brain image analysis and non-invasive automatic identify of brain regions. Massive input training models represent a significant challenge to the development of CNNs to purpose a large MRI image and analysis in computational neuroscience. DCNN has been widely applied to WBTS and GPU memory limitations for studying 2D/3D CNN methods are based on down-sampling. However, 3D patch-based high-resolution 3D-CNN has achieved superior performance and high accuracy by performing model training and big data learning. However, GPU memory is limited, and training time is longers.
Methods
In study, propose a customized, sub-optimal "patch-based are able to improve performance and optimize input channels. The GPU memory collected to carrying out the training process and training model inference and a collective larger than the activation size on pipeline. A sub-optimal "patch network tiles" modeling approach is needed for small input data. The PSLNT method distributes data into several independent 3D-FCN, so that WBTS obtains high-resolution, each network learns contextual information about fixed spatial locations.
Results
The proposed method through extensive experiments on 316 MRI brats 2021. calculating the average score, standard deviation, mean-dice score range wt 0.785, tc 0.761, and et 0.754 for intact tumors, tumor cores, and adjunctive tumors, also, as we can see the specificity metrics are wt 0.841, tc 0.825, and et 0.873 which means two segmentation images are the same for this metric. the metric average sensitivity was similar to approximately 98% for the entire tumor, tumor nucleus, and tumor enhancers respective.
Conclusion
Multi data patch network tiles with PSLNT produce improved segmentation results using 3D CNN. 3D U-net has the advantage of sens, spec, and HD, namely 0.78-0.98 and the difference between DSC and HD measurements is 0.11%. Results had slight advantages across the data set compared to all other methods evaluated in this study.

Keywords— PSLNT, FCNN, pipeline, 3d-unet, brats challenge 2021, model optimization,WBTS.

I. INTRODUCTION

Deep learning (DL) has transformed many image processing applications, massive input image data, such as MRI analysis, facial, and ultrasound. Develop DL models with limited GPU memory. DL takes advantage of GPU computing, which offers a superior performance boost when compared to CPUs. However, GPUs have significantly less memory (usually 64 GiB or less). DL requires that the activations are calculated at each layer in the forward pass, and it calculates the gradient in the backward pass, because each layer is processed in reverse order [1],[18], and usually stores the forward activations in all layers until the reverse pass begins, so, the total set of activations must match the global memory of the GPU.

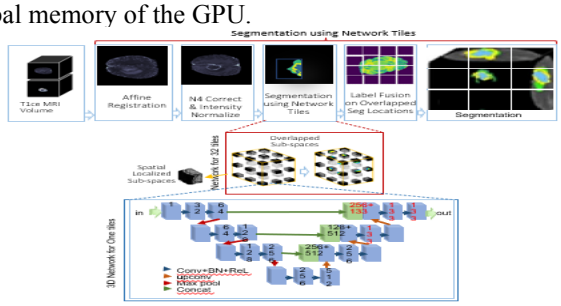

Fig. 1. The proposed 3D Patch-SLNT for whole brain segmentation method on T1ce, seg Dataset BraTS2021

Activation of host memory between the host and GPU drastically impacts performance. Limited memory, capable of processing with sub-optimal input data, either by smoothing the input data (e.g., brain MRI analysis [3]) or applying sub-optimal "patch"-based modeling to small slices of the entire image (e.g., digital MRI The Brain Tumor Segmentation Challenge 2021 (BraTS2021) [16]). To the best of our knowledge, propose a first approximation scheme for the limitations of available memory namely the large data patch spatially localized network tiles method (PSLNT). Distribute data across multiple independent 3D full convolutional networks (FCN), to obtain high-resolution whole-brain segmentation, using multiple spatially

distributed networks and each network performing contextual learning of a fixed spatial location. In addition, we propose effective utilization of GPU memory for model training and big data. The main contributions of our work are as follows:

- First, propose efficient utilization of GPU memory, performing DL processing, on three types of data based on access patterns (model parameters, input data, and results). Because of this, it allows efficient use of GPU memory in conducting training data. Conducted experiments and model training on T1ce BraTS2021 data [18] by optimizing memory and analyzing the experimental results.
- Second, implemented 3D Patch Spatially Localized Network Tiles (PSLNT) to process input DL training data in GPU memory and manage memory, because training large-scale models requires large memory. However, due to single GPU memory limitations and DL if the memory size is bigger than GPU memory. Thus, modified as well as verified memory efficiency in conducting large-scale model training. and achieve high accuracy.

II. RELATED WORK

In this section reviews the techniques for patch and segmentation models.

CNN-based medical image segmentation applies a contouring algorithm [1], then developed based on deep CNN U-Net, a simple model, superior performance to other Unets, such as DCNN [1], Res-UNet[2], [9]. Models use few parameters, such as 3D-Unet[1], V-Net[3], so that fused network learning, training is faster. CNN patch-wise based segmentation of,2D/3D patches [4],[5],[6]. 2D/3D patches extract various scales of information, local spatial context. Patch CNN carries out network training with patches of different sizes or multi-scale patches [7], the resulting patches are recombined, and connected to a soft-max layer. The network can also combine global spatial consistency with local detail and is based on pixel labels with the SegNet architecture [21], calculating a SegNet-based CNN, DSC score of 0.8 on the OASIS dataset. Also, DCNN computing [1],[5] has been widely applied to whole brain segmentation and performs training and learning of 2D/3D DCNN method [1],[8],[9] based on down-sampling with the brats 2020 dataset, DSC score of 0.8. 3D-CNN random patch as train by 2D-CNN are extracted features of long range, three orthogonal direction [9], additional input 3D-CNN[10]. Whole brain segmentation (WBTS) with 3D U-net [9] is superior by applying many GPUs and requires a long training time. Purposed an ensemble of 3D U-Nets with different hyper-parameters and patch data extracted non-uniformly[9]. The 3D tree structure of U-Net performs segmentation of edema faces, non-growing tumors, and tumor enhancement regions[11]. The main tool's GPU does the work of training the neural network inference and the GPU's collective memory must be larger than the activation size at each stage in the pipeline. for this, a sub-optimal “patch network tiles” modeling approach PSLNT distributed multiple independent 3D-FCNs, thereby producing whole-brain segmentation with high-resolution results.

Additionally, the 3D-WBTS model is annotated and produces dense segmentation. However, it is widely accepted that it is not responsive to user interactions [12],[4],[5] which means that brain segmentation analysis in 3D space is more difficult than in 2D space and interactive 2D segmentation is more suitable than direct 3D-WBTS due to distance and large inter-slice movements [13],[14],[15]. this study, applying manipulation of 3D individual patches is computationally much more complex than 2D patches. Moreover, compared with random patch selection that only focuses on a specific region, this study's proposed method using uniform patch selection can utilize information from the entire slice for better segmentation, non-expanding, and enhancement of the tumor region.

III. MATERIAL AND METHODS

In this section, to described dataset MRI data samples, pre-processing, 3D-Unet architecture, and 3D patch-based consists of sub-chapters including training, whole segmentation, and post-processing.

The conceptual framework and methods in the research are contained (in “Fig.1”). Currently, segmentation with CNNs is growing rapidly. Straightforward the whole brain segmentation strategy always adjusts from various aspects and capabilities for brain volume segmentation with 3D-CNNs, such as 3D-UNet [9] or V-Net[16], the results are impractical, unable to produce MRI high-resolution (eg ., 1mm or even higher isotropic voxel size) to 3D-FCN, due to common memory limitations GPUs and CNNs are not yet able to drill down into the details of a whole brain MRI scan with annotations (e.g. >100 labels). A separate challenge in terms of the device, specifically the memory capabilities of the GPU running limited for training data, many previous attempts have been made. CNN learns [17],[18] 2D, and 3D patches and layouts for whole brain segmentation[19][20] and the network has been extended to BrainSegNet [21], which uses 2.5D patches to train the CNN network.

This method performs whole-brain segmentation from a dataset PSLNT, Segmentation constraints a network tiles/voxel segmentation which are FCN has been entire whole brain segmentation [8] and method 2D FCN of a scale the size of the label on the data that was not originally labeled. training has entered the 2D method, recognizing images not only through random full brain segmentation, but using better labels that previously did not have labels [8], and 2D segmentation can display inferior 3D spatial consistency. So a lot of research is doing 3D-FCN (eg., 3D U-Net[9]).

Multi-task and complete learning [21],[22], realized in the aspect of visualization 3D whole brain segmentation. multi-task 3D structured learning network, capable of achieving decent segmentation performance. However, the maximum input_data size is $120\times120\times120$ mm^3, the total number of voxels is 2,065,189 on GPU memory limitations, that loses two-thirds of the spatial information compared to a 1mm isotropic brain volume in a $172\times220\times156$ mm^3 (5.903.040 voxel). MNI space (5,903 .040 voxels). The GPU is sufficient for data processing, multi-tasking can be done by avoiding over-fitting with the same size stage. We resized 240x240x150 mm^3 to $256\times256\times256$ mm^3 (in., **“**TABLE.1**”**), so the data would not overlap during training. Furthermore, applying 3D-UNet to whole-brain segmentation (for example, with 1mm isotropic resolution in MNI space or even higher resolution) is still limited by the memory current of the GPU. This obstacle is a new challenge, 3D-CNN learning-based slides are applied to a single network with a window base of patching at different spatial locations [23],[24],[25],[26],[27] and a single network design implicated learns two things : (i) “where is the patch located in the brain?”, (ii) “How do I set a label for

the patch". (i) decide on the patch candidate labels at the limit of 133 labels, while (ii) decide how to label each patch voxel. In terms of patches, we propose 3D-patches, 3D-UNet with a patch spatially localized network tiles selection strategy of 64 tiles. A method that alleviates the difficulty of patch-based learning is that 3D-UNet can work independently on each network, and is only responsible for certain spatial locations [5], simplification of network tasks focuses on the same patch on all brain regions with less spatial variation (e.g., networks associated with specific sub-sections of the complete brain, (in "Fig.1").

A. Dataset and Pre-processing

The brain MRI modality that can isolate edema from cerebrospinal fluid (CSF), is T1 [28], so we processed T1ce data from as many as 316 BraTS2021 datasets (in "TABLE.1").

Pre Processing: separating T1ce data, seg for 316 patient data and grouping T1ce data, seg from sequence number 0 to 316, Size 240x240x150 mm^3 and resizing the size 240x240x150 mm^3 to 256x256x256 mm^3, aiming to avoid overlap, gaps slices, etc [29]. The canonical voxel size is 1 mm^3 and the lesion size is 10 to 2.8×10^3 is set-manual and the demographic dataset is not available, adjust ITKSnap software. Then reduce some noise, N4 correction, and normal intensity to get high-quality and resolution images.

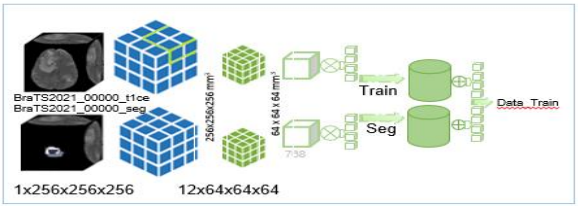

Fig. 2. Pre-processing 3D PSLNT of T1ce and seg BraTS2021.

Next, perform a 3D Patch-Based Spatially Localized Network Tiles (PSLNT) process for network tiles of64 voxels (ex., one T1ce MRI 256x256x256 mm^3, divided into 64 sub-voxels, 64x64x64 mm^3/voxels. So that the total sub-voxel are 12x64=768 voxels. Each network tile is stored in the order N_P_train.HD5, where, N is the initial sequence, P is the sub-voxels data sequence number for each patient, the train is labeled as training data and H5 is the data storage extension. Thus, each patient data sequentially countered for train, seg as validation consisting of (1_10_train.H5 to 64_10_train.H5) and (1_10_seg.H5 to 64_10_seg.H5), each file size is 2.1 MB/h5-tiles. Datapath accommodates 485,376 voxels (train = 242,688 voxels and seg 242,688 voxels) and 80% allocation is used as train, with 20% validation of 316 patient data.

TABLE I. HYPERPARAMETERS FOR MODEL OPTIMIZATION IN OUR EXPERIMENTS [28].

Optimization Stage	Stage
Input volume size mm^3	64x64x64 mm^3 (100% sub-volume)
Training Length	200epochs
Initial learning rate σ	1.00E−04
Optimizer	0.01
GPU	Adam optimizer and restart scheduler
Deep learning framework	NVIDIA AMD RTX, with 1 GB memory
	PyTorch

B. U-Net Network Architecture

Whole brain tumor segmentation (WBTS) used 3D-UNet [9]. 3D-UNet has an improved model of a FCN [30]. FCN performs sequential convolution to process sample data input and looks for max-pooling values. However, U-net performs a convolution operation with padding and class prediction results for each pixel taken from the activation feature map of the encoding and U-shape architecture on the decoded side (path of an encoder, decode) in charge of receiving, input process the activation feature prediction results of two classes (background) brain). As the decode processes the activation feature input, the pooling task is taken over by oversampling. Where the U-net will process the data received from the 3D Patch input in 64^3 dimensions and there are two segmentation map outputs with 64^3, dimensions, then the softmax function is in charge of equating the segmentation map values and class predictions in the U-Net network (in "Fig. 2").

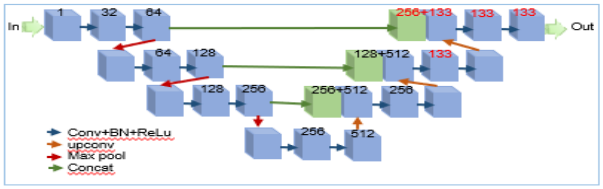

Fig. 3. U-Net Architecture Implemented.

C. Patch-Based

The success of image segmentation was carried out by a 3D-UNet architecture with a high level of performance [9]. If the size of the input data image is getting larger, the memory capacity of the GPU (Graphics Processing Unit) is often overloaded in training, because the memory will hold all the MRI images or MRI pieces, so the model that is run tends to lose detailed information from several image areas, this occurs due to the input of the predicted activation features from the two classes (background, brain) being unbalanced [9]. Over-coming unbalanced classes, we have a suggestion, namely the input data is from network tiles, which are obtained from 3D-PSLNT, contained in pre-processing (in Section 2. A), and if the class prediction features are not balanced with small voxels input, then proceed with the patching process. Where the network patches local information so that the GPU memory occupation is relatively small during training and inference. 3D-patch enriches the context of information stored in the patch itself during training, which is advantageous in terms of three-way context (axial, coronal, and sagittal)[9].

1) Training: 3D-PSLNT dataset, the result of pre-processing (in Section 2. A), will contribute to the regulatory effect of augmentation by extracting various sub-volumes from the original volume shape. A unit of sub-units and sub-optimal GPU devices perform small volume patches (PVS) and are trained at the same patch location. A PVS training results as a dataset as input_data to the 3D-UNet and the dataset consists of train and seg, and input_data (p) and seg or ground ruth ($\hat{p}$).

Overlapping patches (in Section 2. C), will help with local (eg.,64×64×64 mm^3 wasteful local neighbors) and global context convolution [20], however, more patches in the brain segmentation area, will provide accuracy of global information. Of note is the voxel prediction result during training, the decision received from the overlapping patches, (eq. (2)). Voxels in each stored patched, so that 3D-UNet is called having a brain because every parameter in it can be

adapted so that the binary cross-entropy loss function gets smaller [24].

$$L(U,\hat{u}) = \sum_i -(U_i \log(\hat{u}_i) + (1-U_i)\log(1-\hat{u}_i)) \quad (1)$$

where U_i is the score voxel x_i in the class real times (0 background value, 1 brain value) and $\hat{u}_i$ is voxel x_i from brain probability. If the patch values overlap, of belonging to the brain and if patches overlap, values $\hat{u}_i$ average for all related patches:

$$\hat{u}_i = \frac{\sum_{p \in P_i} Softmax_p(u_i)}{P_i} \quad (2)$$

Where P_i is a set of patches with value x_i ; and $Softmax_p(u_i)$ is voxel x_i the voxel patched by p values, as the prediction belongs to the brain:

$$Softmax_p(u_i) = \frac{exp^{u_i}}{\exp^{(1-u_i)} + exp^{u_i}} \quad (3)$$

where u_i is value voxel x_i where is the voxel value classified as the brain $(1-u_i)$ is the value as background. It should be remembered that the brain is in a very small MRI volume, and u_i is not met, the class representative is not fulfilled [20]), voxel x_i voxels are masked against the set-seg parameter for the whole brain (in "Fig.3") with the dataset train 80%, seg 20 % (in Section 2. A). The training was carried out by randomly selecting patches in each region with a combination of 256, 128, and 64, starting from the center voxel medical images (CVMI).

2) Inference: we evaluate adjacent patches, review the relevance of adjacent patches, and segmentation matters [9][31]. 3d-patching is done on adjacent values on the MRI images because some of the whole segmentation network tiles cover the whole image. u-net architecture will return the probability v_i vector belonging to each class (background or brain) of the mask. The value of each voxel x_i of the value of each patch p.

$$\hat{u}_i = argmax v_i \quad (4)$$

3) Post-processing: segmented patch aggregation after deep learning generates a mask on the T1ce-BRATS2021, still many false positive displays are found. Improvements were made in this regard, processing the training dataset [25][26][32]. Segmentation accuracy is a challenge, especially in the sub-region, of the smallest voxel, et, which is very difficult to obtain precise segmentation, the success rate is 40% [26][32], and precision can be maintained with a stable score of 1 dice when not using labels and scores die 0, single false positive voxel, then no tumor found. therefore, if it is conditioned that the dice value < 0.65, it will mask the result to a data set mask consisting of 4 image masks (in "Fig.3").

IV. Experiment and Result

In this section, we explain the segmentation, and evaluation metric (Section 4. A), the results of our experiments consist of sub-chapters (among other things 3D whole segmentation results, prediction performance), result of calculation average score (Section 4. B). and (Section 3. C), A results related to the use of 3D patches and Comparison existing method.

A. Segmentation and Evaluation Metric

This research was conducted with the Jupyter platform, PyTorch library, regulated spatial monai-library on Windows 11 64-bit with AMD Ryzen 7 6800H, GPU Nvidia Geforce RTX 3060 GDDR6 @6GB (192bits). The 3D-UNet models training on working with parameters (in "TABLE 1"). Considering the loss aspect with the cross-entropy function (in Section 2.C.1) (Equation (1)) The quality of whole brain segmentation with the conceptual framework has been described previously (in Section 2.C.1-4). Evaluating the performance of the 3D-UNet model, we perform calculations, DSC, mDSC, HD, ASSD, TPR, and precision for each MRI scan. DSC, HD, and ASSD, using surface distance computing libraries [29][33][34]. Similarity predicted $\hat{p}$ with Ground truth p [35]:

$$DSC(p,\hat{p}) = \frac{2|p \cap \hat{p}|}{|p|+|\hat{p}|} \quad (5)$$

where |.| represents set.

We previously evaluated the accuracy of train_data, Ground_ruth prediction for the Hausdorff distance (HD), between the set of maximum-distance points p to the nearest point $\hat{p}$ [29] as well the average symmetrical surfaces distance [36], which can be defined :

$$HD(p,\hat{p}) = \max\left(\max_{a \in p} \min_{a \in \hat{p}} |v-u|, \max_{b \in \hat{p}} \min_{a \in p} |u-v|\right) \quad (6)$$

$$ASSD(p,\hat{p}) = \frac{1}{|p|+|\hat{p}|} x\left(\sum_{a \epsilon p} D_{\hat{p}}(u) + \sum_{b \in \hat{p}} D_{\hat{p}}(v)\right) \quad (7)$$

The average of all symmetrical surface distances between p and the nearest $\hat{p}$, also the inverse of $\hat{p}, -\ p$ and $D_{\hat{p}}$(a). gives the voxel distance to the surface point $\hat{p}$. 3D-UNet training needs to be done because some of the information affects local information [31] and we do patch learning and we compare the results with 64 voxel-sized patches at different spatial locations [9][10][11][12][19][20].

B. Result of calculation average score

In this section, we present the results of 3D-PSLNT, 3D whole segmentation of brain tumors, and post-processing. Architectural evaluation applies the calculation of Dice Score (DS), Sensitivity (Sens), Specificity (Spes), and HD.

1) Whole Segmentation Result: in this section, we summarize and display the results of the whole brain segmentation (in section 2), the results of applying the 3D-UNet model (in "Fig.4") and evaluate the results by calculating the average score, standard deviation, median, for four metrics: SD, Sens, Spes, and HD and the mean-dice score for architecture. in the validation set were: WT 0.785, TC 0.761, and ET 0.754 for intact tumors, tumor cores, and adjunctive tumors (in "TABLE. 2"). also, as we can see the specificity metrics are WT 0.841, TC 0.825, and ET 0.873 which means two segmentation images are the same for this metric. The metric average sensitivity was similar to approximately 98% for the entire tumor, tumor nucleus, and tumor enhancers respectively. in addition, the first-time post-processing will partially remove components according to the global threshold limit of 110, to partially reduce the false positive region, the impact of the process produces SD, and Spec has been improved after applying this post-processing technique. In addition, after applying post-processing, we notice an increase in the HD, accumulating the average of all symmetrical surface distances between p and $\hat{p}$ nearest points, also the inverse of $\hat{p}, -p$ dan $D_{\hat{p}}$(a), and gives the voxel distance to the surface point $\hat{p}$.

Also, it presents the median value for the validation dataset, and the model, thus obtaining a high median value (in "TABLE 3"). However, the uncertainty between EN, and ET, so that the results of whole brain segmentation are not

acute > 87 %, even though the sensitivity is high. Needs further testing, and sensitivity is high. Needs further testing. The results have not obtained the expected results for the whole segmentation of the data validation set. A whole tumor (WT), core tumor (TC), and enhancing tumor (ET) respectively (in "Fig.4").

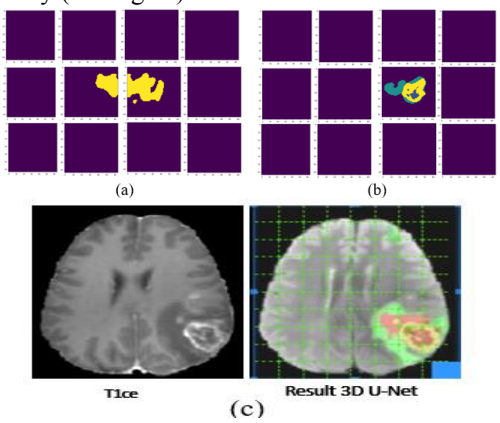

Fig. 4. Visual : (a) PSLNT for Eendema (b) PSLNT for Enhancing Tumor and Non-Enhancing Tumor Core the Necrotic (c) Result Of The Whole Tumor Structures

2) Prediction performance: evaluation metrics of the u-net model summarized (in "TABLE 2"). where the median DS of our best architecture are WT 0,785, TC 0,761, ET 0,753. We can detect and segment complete WT ± 71.7% - 99.5%, as observed in dice score metrics, sens, spec, and HD ±84.9% - 95.7% and ET segmentation ± 88.1% - 95.9%, Also, can increase the yield and observe the metrics DS, sens, spec and HD ± 83% - 96%. [12][13][24].

C. Result in Comparison with Existing Methods

The tests are similar in detail to segmentation, but they look at brain structure, while we focus on looking at 3D whole brains. To further increase the network size results in increased performance using a small margin, higher GPU power requirements, as well as, having an impact on the 3D-UNet model train, on PSLNT dataset training, obtaining shorter segmentation results and providing the best performance, and shorter inference time fast. In addition, it allocates datasets in classes and sub-groups, so that each memory works optimally during training[25][26][28].

TABLE 2..EVALUATION RESULT FOR TRAIN_DATA, SEG AS VALIDATION-SET, POST-PROCESSING OF T1CE-BRATS-2021 AND INPUT ARCHITECTURE IS TRAINED ON PSLNT-64 × 64 × 64 MM^3.

	DS			Sens			Spec			HD		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
Mean	0.785	0.761	0.754	0.841	0.825	0.853	**0.972**	**0.965**	**0.987**	21.205	15.163	11.142
Standard deviation	0.113	0.214	0.273	0.153	0.244	0.246	0.004	0.006	0.003	24.284	10.254	23.324
Median	0.872	0.822	0.831	0.834	0.841	0.847	0.996	0.994	0.997	9.213	10.208	2.665

TABLE 3. EVALUATION METRIC SUMMARY. MAXIMUM SCORE IS THE BEST FOR DSC,MDSC, TPR AND PRECISION. SMALLEST SCORE WILL BE BETTER FOR DISTANCE METRIC (HD, ASSD), BEST SCORES ARE MARKED IN BOLD. THE MODEL'S PERFORMANCE, BASED ON THE PRIMARY BRAIN LOCATIONS

Methods	mDSC	DSC	HD (mm)	ASSD (mm)	TPR	Precision
FCN		78.88 ± 1.07	21.50 ± 4.69			
U-Net		**79.42** ± 1.12	16.25 ± 4.00			
MA_FCN		81.19 ± 1.06	14.11 ± 3.22			
VoxHRNet		**81.19 ± 1.06**	14.11 ± 3.22			
3D U-Net	0.875	0.765	20.6	3.6	0.81	0.63

V. DISCUSSION AND CONCLUSION

Challenging WBTS, and minimum over-fitting we adopted augmentation, early termination, functioning for minimum-error estimation of DL learning, training on the T1ce-MRI dataset, and more data. At the beginning of the training we only used 50 patient data, our model test results had an accuracy of ± 0.42, and we had not tested other variables, including DSC, and HD. By adding 75% data we get a better result. Our 3D-UNet model experiment, with a set batch-size (BZ) of 1, due to the limited GPU memory to process volumetric input data the number of Patches is 40,448 tiles, produces a better computational gradient, in contrast to previous studies that used large data size data as input data to GPU memory, resulting in more uncertainty in gradient computing [28] and the overall learning process. However, we still consider adopting an optimization algorithm that functions with more advanced variance reduction [28]. Then we focus, on increasing the amount of training data the results of the accuracy increase are 0.721 and also apply post-processing, so we see an increase in HD due to the effect of this operator on the ET class, the median value percentage is relatively high, and the achievement of good segmentation for most patient images.

Comparison of the results [9][19], obtained regarding the overall segmentation of the data validation set, 3D U-net testing, they obtained DSC, HD (voxel) average DSC 0.81, HD 0.14. While our test results obtained a DSC of 0.763 and HD 35.3 and 20.6 and we tested the ASSD parameter, precision. Our comparison of DSC and HD has a lower difference of ± 0.034.

ACKNOWLEDGMENT

This research was funded by the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program; and Institut Teknologi Sepuluh Nopember (ITS) under the project scheme of the Publication Writing and IPR Incentive Program (PPHKI).

REFERENCES

[1] K. N. Qodri, I. Soesanti, and H. A. Nugroho, "Image Analysis for MRI-Based Brain Tumor Classification Using Deep Learning," vol. 5, no. 1, 2021.

[2] A. Arora, A. Jayal, M. Gupta, P. Mittal, and S. C. Satapathy, "Brain tumor segmentation of MRI images using processed image driven u-net architecture," *Computers*, vol. 10, no. 11, 2021, doi: 10.3390/computers10110139.

[3] X. Guan *et al.*, *3D AGSE - VNet : an automatic brain tumor MRI data segmentation framework*. BioMed Central, 2022. doi: 10.1186/s12880-021-00728-8.

[4] L. M. Ballestar and V. Vilaplana, “MRI Brain Tumor Segmentation and Uncertainty Estimation Using 3D-UNet Architectures,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 376–390. doi: 10.1007/978-3-030-72084-1_34.
[5] Y. Huo *et al.*, “3D whole brain segmentation using spatially localized atlas network tiles,” *Neuroimage*, vol. 194, pp. 105–119, 2019, doi: 10.1016/j.neuroimage.2019.03.041.
[6] S. Roy, D. Kügler, M. Reuter, and M. R. De, “Are 2.5D approaches superior to 3D deep networks in whole brain segmentation?,” *Proc. Mach. Learn. Res. Rev.*, pp. 1–17, 2022.
[7] J. Wu and X. Tang, “Brain segmentation based on multi-atlas guided 3D fully convolutional network ensembles 1 Introduction,” pp. 1–41.
[8] Y. Huo *et al.*, “NeuroImage 3D whole brain segmentation using spatially localized atlas network tiles,” *Neuroimage*, vol. 194, no. February, pp. 105–119, 2019, doi: 10.1016/j.neuroimage.2019.03.041.
[9] B. Lee, N. Yamanakkanavar, and J. Y. Choi, “Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture,” *PLoS One*, vol. 15, no. 8 August, pp. 1–20, 2020, doi: 10.1371/journal.pone.0236493.
[10] A. Kanezaki, Y. Matsushita, and Y. Nishida, “RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints,” 2016, doi: 10.1109/CVPR.2018.00526.
[11] C. Laiton-Bonadiez, G. Sanchez-Torres, and J. Branch-Bedoya, “Deep 3D Neural Network for Brain Structures Segmentation Using Self-Attention Modules in MRI Images,” *Sensors*, vol. 22, no. 7, 2022, doi: 10.3390/s22072559.
[12] Y. Huo, S. M. Resnick, and B. A. Landman, “4D multi-atlas label fusion using longitudinal images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2017, pp. 3–11. doi: 10.1007/978-3-319-67434-6_1.
[13] N. M. Rajkumar and A. Professor, “A MEDICAL IMAGE RETRIEVAL SYSTEM USING HYBRID FOCS FRAMEWORK,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 411, no. 11, pp. 256–266, 2015.
[14] H. Lu, H. Wang, Q. Zhang, S. W. Yoon, and D. Won, “A 3D convolutional neural network for volumetric image semantic segmentation,” in *Procedia Manufacturing*, Elsevier B.V., 2019, pp. 422–428. doi: 10.1016/j.promfg.2020.01.386.
[15] U. Baid *et al.*, “A Novel Approach for Fully Automatic Intra-Tumor Segmentation With 3D U-Net Architecture for Gliomas,” *Front. Comput. Neurosci.*, vol. 14, 2020, doi: 10.3389/fncom.2020.00010.
[16] X. Liu, R. Yin, and J. Yin, “Attention V-Net: A Modified V-Net Architecture for Left Atrial Segmentation,” *Appl. Sci.*, vol. 12, no. 8, 2022, doi: 10.3390/app12083764.
[17] L. Atikah, N. A. Hasanah, R. Sarno, A. Fajar, and D. Rahmawati, “Brain segmentation using adaptive thresholding, k-means clustering and mathematical morphology in MRI Data,” *Proc. - 2020 Int. Semin. Appl. Technol. Inf. Commun. IT Challenges Sustain. Scalability, Secur. Age Digit. Disruption, iSemantic 2020*, pp. 161–167, 2020, doi: 10.1109/iSemantic50169.2020.9234303.
[18] U. Salamah, R. Sarno, A. Z. Arifin, A. S. Nugroho, I. E. Rozi, and P. B. S. Asih, “Incorporating index of fuzziness and adaptive thresholding for image segmentation,” *Int. J. Electr. Comput. Eng.*, vol. 8, no. 4, pp. 2406–2418, 2018, doi: 10.11591/ijece.v8i4.pp.2406-2418.
[19] A. Fahmi *et al.*, “Vim line technique thalamotomy for Parkinson tremor: Case series,” *Int. J. Surg. Case Rep.*, vol. 77, pp. 573–575, 2020, doi: 10.1016/j.ijscr.2020.11.096.
[20] H. Mzoughi *et al.*, “Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI Gliomas Brain Tumor Classification,” *J. Digit. Imaging*, vol. 33, no. 4, pp. 903–915, 2020, doi: 10.1007/s10278-020-00347-9.
[21] M. Lee *et al.*, “Split-attention u-net: A fully convolutional network for robust multi-label segmentation from brain mri,” *Brain Sci.*, vol. 10, no. 12, pp. 1–22, 2020, doi: 10.3390/brainsci10120974.
[22] A. Bhardwaj, A. Tiwari, H. Bhardwaj, and A. Bhardwaj, “A genetically optimized neural network model for multi-class classification,” *Expert Syst. Appl.*, 2016, doi: 10.1016/j.eswa.2016.04.036.
[23] U. Baid *et al.*, “A Novel Approach for Fully Automatic Intra-Tumor Segmentation With 3D U-Net Architecture for Gliomas,” *Front. Comput. Neurosci.*, vol. 14, 2020, doi: 10.3389/fncom.2020.00010.
[24] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, “Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation,” *Front. Comput. Neurosci.*, vol. 13, no. August, pp. 1–13, 2019, doi: 10.3389/fncom.2019.00056.
[25] C. H. Yeh *et al.*, “White matter microstructural and morphometric alterations in autism: implications for intellectual capabilities,” *Mol. Autism*, vol. 13, no. 1, pp. 1–16, 2022, doi: 10.1186/s13229-022-00499-1.
[26] H. J. Lee, J. U. Kim, S. Lee, H. G. Kim, and Y. M. Ro, “Structure boundary preserving segmentation for medical image with ambiguous boundary,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4816–4825, 2020, doi: 10.1109/CVPR42600.2020.00487.
[27] X. Chen, S. Jiang, L. Guo, Z. Chen, and C. Zhang, “Whole brain segmentation method from 2 . 5D brain MRI slice image based on Triple U-Net,” *Vis. Comput.*, 2021, doi: 10.1007/s00371-021-02326-9.
[28] D. Sunaryono *et al.*, “Enhanced Gradient Boosting Machines Fusion based on the Pattern of Majority Voting for Automatic Epilepsy Detection,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, pp. 595–604, 2022, doi: 10.14569/IJACSA.2022.0130771.

BIBLIOGRAPHY OF AUTHORS

FIRST A. AUTHOR, Pandapotan Siagian received he Master degree in 2010. Currently a doctoral student of Institut Teknologi Sepuluh Nopember Surabaya, Indonesia. His research interests include image processing and biomedical informatic and medical analysis.

SECOND B. AUTHOR, Riyanarto Sarno (Senior Member, IEEE) received the Ph.D. degree, in 1992. He is currently a Professor with the Informatics Department, Institut Teknologi Sepuluh Nopember (ITS). He was the author of more than five books and over 300 scientific articles led him incorporated in the top 2% world ranking scientist by Standford University, in 2020. He has researched process mining for a period of five years. His research interests include machine learning, the Internet of Things, knowledge engineering, enterprise computing, and information management.

THIRD C. AUTHOR, Chastine Fatichah (Member, IEEE) received the Ph.D. degree from the Tokyo Institute of Technology, Japan, in 2012. She is currently an Professor with the Institut Teknologi Sepuluh Nopember, Surabaya. She has published more than 110 computer science-related journal articles and conference papers. Her research interests include artificial intelligence, image processing, and data mining.

FOURTH D. AUTHOR, Aziz Fajar. I am a Computer Scientist currently working on Biomedical Informatics. I am mostly working on image processing of DICOM images such as MRI, CT, and DWI images. I have Published some papers about the topics in recent years. Currently, a doctoral student of Institut Teknologi Sepuluh Nopember.

FIFTH E. AUTHOR, Muhammad Ibadurrahman Arrasyid Supriyanto. I have Published some papers about the topics in recent years. Currently, a doctoral student of Institut Teknologi Sepuluh Nopember.