Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
Vol. 27 (2024), No. 3, pp. 547–561
DOI : 10.47974/JIM-1767

Improved Global U-Net applied for multi-modal brain tumor fuzzy segmentation

Annu Mishra *
Department of Computer Science and Engineering
Birla Institute of Technology - Mesra
Noida
Uttar Pradesh
India

Pankaj Gupta †
Department of Computer Science and Engineering
Birla Institute of Technology - Mesra
Ranchi
Jharkhand
India

Peeyush Tewari §
Depatment of Mathematics
Birla Institute of Technology - Mesra
Jaipur
Rajasthan
India

Abstract

In this paper, we extended our work from Global U-Net combined with fuzzy amalgamation of Inception Model and Improved Kernel Variation for MRI Brain Image Segmentation [1] which was meant for single modality MRI images only to a brain tumor fuzzy segmentation. Many CNNs gives state of art results for a particular type of images. However, they cannot achieve the same result for the images captured from different imaging techniques. We experimented the Global U-Net model for MRI images earlier and this time we intended to make it applicable for other type of images too using the concept of fuzzy segmentation. The major concern was to overcome the limitations of single modality system

* *E-mail:* **phdcs10059.19@bitmesra.ac.in (Corresponding Author)**
† *E-mail:* **pgupta@bitmesra.ac.in**
§ *E-mail:* **peeyush@bitmesra.ac.in**

that is not all the kernels of U-Net are capable of generating clear feature vectors for different image modalities. The result generated was satisfactory and we would further extend it for colored images.

Subject Classification: *68U10.*

Keywords: *U-Net, Multi-modality, Fuzzy image segmentation, Pooling layer, Aggregation block.*

1. Introduction

Biomedical imaging currently employs a number of different techniques, including CT, MRI, X-rays, and others. Because of its pure imaging technology, MRI and CT scan are usually superior than other tactics, particularly for brain imaging. Even while there are associated good artifacts, we also have methods to lessen these artifacts. MRI collections can be T1, T2, or FLAIR series. The CSF appears moderate in the T2 sequence but dark in the T1 MRI collection. The FLAIR MRI sequence bears a striking resemblance to T2. On the other hand, the echo and repeat times are longer. In order to identify tumor cells, scientists and technocrats have focused their efforts on creating environmentally friendly methods for automatically segmenting CT and MRI images.. Among all the previous techniques, deep learning is the most popular and cutting-edge. The automatic segmentation problem is partially resolved by neural networks. As a result, CNNs like ResNet, V-Nets, and U-Nets are garnering a lot of interest in the industry. Segmentation is one of the most important processes in image processing, which is the methodical manipulation of virtual images. It splits the target image into several parts, and each of those parts has good records for texture, pixel density, and other characteristics. As a result, each piece wants to be isolated from the others, and in order to achieve this isolation, we require a reliable way to identify the individual parts. Therefore, the hardest task before starting the segmentation system is selecting an appropriate and satisfactory segmentation process. Furthermore, a model need not produce the same outcome for all kinds of photos. For instance, a model might function well with MRI scans but not with CT scan images, and so on. When we compared our model's output to that of traditional U-Net and our previously altered Global U-Net[1], we found that when multimodality was involved, our model outperformed the others in the given dataset and constraints. The unique component of image processing and clinical image segmentation is where we applied our version. The image is divided into multiple parts, each of which offers useful details on the image's texture, pixel depth, and other aspects. Every portion therefore wants to be isolated from other segments, and in order to achieve this isolation, we need a corporate method for identifying the component pieces. Thus, the most difficult task before beginning the segmentation process is selecting an excellent, ideal segmentation approach. Biomedical imaging can be achieved by a variety of methods nowadays, like CT, MRI, X-rays, and more. Because of its simple imaging methodology, MRI typically outperforms the other methods when it comes to mind imaging. Even though there are some related artifacts, we also have techniques to reduce those artifacts. An MRI sequence could be FLAIR, T1, or T2. In an effort to find tumor cells, researchers have expanded methods for automatic MRI and other model picture segmentation. Deep learning-enabled neural networks eliminate the headache of automated segmentation to produce superior outcomes. Nonetheless, multimodality remains a problem in the majority of image processing domains.Empirical techniques are utilized to understand the efficiency of the final image obtained, whereas analysis techniques are used just for image segmentation. While empirical approaches are highly efficacious, they are not without restrictions. Specifically, the need for outcome, input, and reference to be in one place adds to the framework's complexity. Through a detailed comparison of the input images and output images, these methodologies perform performance analysis to determine how effective the segmentation methodology was.

2. Related Work

The term image segmentation has caught the eyes of researchers in the recent years. Around 18000 articles are published since 2019 till date on this topic in various areas of biomedical imaging. In early years of image segmentation, the main focus of researchers was to find the ROI within the image and then analyse it on the basis of limited number of features based on texture, region and edges etc. [2][3]. The use of neural network has allowed the researchers to transcend in the area of image segmentation and classification [4]. Various recent approaches have been combined with the neural network to give rise to RCNN, VNet, UNet, Fast RCNN etc. [5][6][7]. The Fast RCNN for example firstly draws out the ROI from the whole image and then the output is fed to the Convolution Neural Network to extract the features of the image under interest and finally these features are exposed to SVM classifiers. The pooling layer present in the neural network pulls the maximum features out of input and thus makes the network work faster. In UNet, the approach was different where the researcher used the concept of extraction and compaction of the input image. Also, the skip connections helped to match the features faster [8]. Another modification was done by the authors of [9] wherein they implemented the UNet with inception and variant filters in each layer. They also incorporated accretion block to reduce the gradient descent problem that existed in traditional UNet due to deep hidden layers. The literature survey proves that all results vary drastically when we switch a neural network from one modality to another [9]. In [10] the author has worked on multi-modality CNN for image segmentation. The author has mentioned some key features of the the developed CNN as identifying the features and its use to classify the image into different classes for images belonging to multi modal systems. They also claimed that the model worked for visible images, Infrared images and X-ray images. The comparision was done based on Dice score(0.96 for heart image and 0.78 for Esophagus) and the results shown clearly depicted that the new method gave better results under the given condition of multi- modality. Figure 1 exhibit the collation of Mast RCNN and the modified Mask RCNN for CT images by the authors of [11][28].

Another work under the researchers of [12] mentioned an approach for 3D multi modal images. The experimented their idea on MRI and CT images to segment brain tumor. They divided the network in levels of input and decision-making levels. For this, the setup required was NVIDIA QuADro P2000 GPU and Intel Xeno Silver 4108 CPU. The results were based on Global accuracy, IoU or Jacquard score and BF score for both the types of images. In The MRI dataset was added with synthetic CT images. These synthesized CT images were generated by using UNet architecture. They achieved the state of art result like other efficient methods and they also mentioned multi class segmentation as the future scope of the model. Table 1 shows the comparison of multi-modal image segmentation including different organs with Thoracic Auto Segmentation Challenge. In

Table 1

Comparison of Multi-modal image segmentation model with Thoracic Auto segmentation Challenge model. Source:[11]

	Left lung	Right Lung	Heart	Esophagus	Spinal cord
Yang et al (2018)	0.97	0.97	0.93	0.72	0.88
Pemasiri et al(20121)	0.97	0.97	**0.96**	**0.78**	**0.91**

[13][14], the authors also dealt with 3D biomedical images and their segmentation. The key updates performed in their model which is considered as the modified version of UNet were:

A. The used Swin Transformer in both the sides of UNet that is encoder and decoder. In other modified versions of UNet the attention layers were used.
B. This Swin Transformer is combined with CNN which forms the network for segmentation process.

The architecture of the underlined model is shown below in Figure 1.The architecture clearly shows that the CNN is added to the transformer model. Also, the use of skip connection helps to map the features of the image from input to the output. Another thoughtful implementation could be seen in [15][16]. Here, the primary concern is to store the local and global information in the encoder side. The loss of local information results in poor image segmentation. This can be avoided by using an aggregation block. This block is basically a size preservative block and it is used at the bottom of the UNet architecture before decoding or upsampling. Figure 2 shows the aggregation block used with combination of UNet.The use of these aggregation block along with the inception model has proven to give results equivalent to state of art results given by other CNNs [17]. The elementary research in this implicates the involvement of the inception model for segmenting the MRI brain image[18][19].

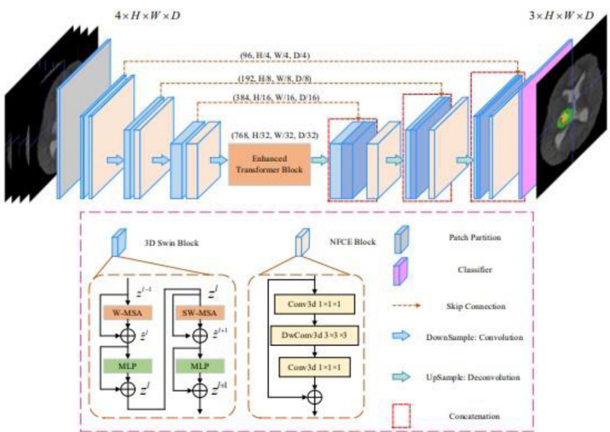

Figure 1

UNETR architecture. Source[12]

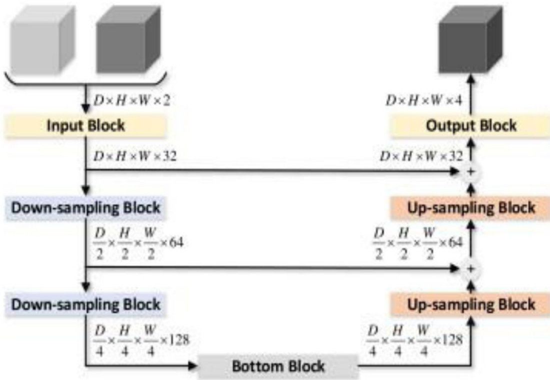

Figure 2

Aggregation Block for the implementation of Non local UNet

The authors also used the concept of variant or multiple sized filters [20]. These filters of different sizes helped the UNet model to become strenuous apropos the alteration in the scale.They firstly drew the activation featuresin detail.These features were applied to the whole image and also upon the small patches of variable size from the image. The output gained from each filter is then forwarded for batch normalisation (BN) [21][22]. Thereafter, the output of Batch Normalization is added to Maxout unit. In the end, activations of each patch of the image is pooled together by the help of Maxpool operation. Thus, the result is obtained by the combination of Maxpool operation and inception model parameters. Figure 3 represents the overall architecture of the Global UNet Model"

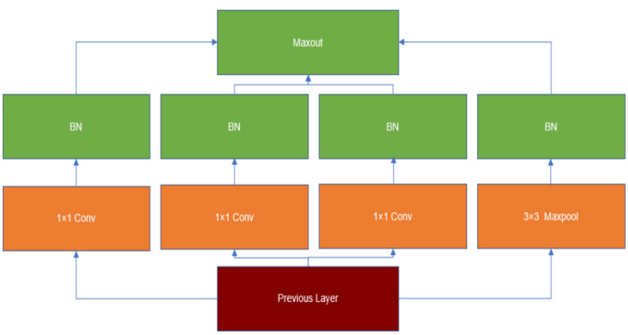

Figure 3

Architecture of Global U-Net with inception model

In another work [22] propoundedSegNet based non-linear multi modal image segmentation focussing on gastral diseases. The model involves bothdownsampling (consisting 13 conv layers) and upsamplingsimilar to U-Net for multi modal image [23][24]. The advantage of using this model was that even the images with low resolution could be mapped.

3. Proposed Model

We have extended our work and tried to incorporate the segmentation of multi modal images by changing some of the paradigms in the former model. The modification made earlier in [1] were variable size filters, use of Maxout function [25] as an activation function and incorporation of Batch Normalization [26]. The convolution layers involved has similar approach as in Global U-Net Inception Model. Apart from this, all these three paradigms mentioned above are also maintained in this proposed model too. The changes made further are as follows:

a. We have made a provision to link together the features gained based on the variable size filters/kernels. This gives us the new feature which can be used for different modalities, CT scan and MRI images in this case. This could be achieved by focusing on the number of kernels in each convolution layer.
b. One more change in the Global U-Net was done here which is related to the input fed prior to Maxpool. After gaining the individual features from each convolution layer, the features are resized to 128×128 during Batch normalization. This gives the infused output when the features are linked together.
c. Instead of using the self- attention as in [27], we have here used the cross-attention block. The cross-attention block like self-attention helps to preserve the features without degrading the image at the exchange point of the encoder and decoder. This not only solves the image resolution issue but also saves the details from each layer which could be utilized by the decoder side.

These changes when applied to the Global U-net gives results for the multi modal images too. We have also compared the accuracy for our new proposed model and Global U-Net along with other models too for image segmentation. The comparison table and the results are shown in the Result section of the article.

4. Machine configuration and training detail

The proposed U-Net consists of four GI U-Net (Global Inception U-Net), where two vestigial modules. Figure 4 and Figure 5 shows the two modules.

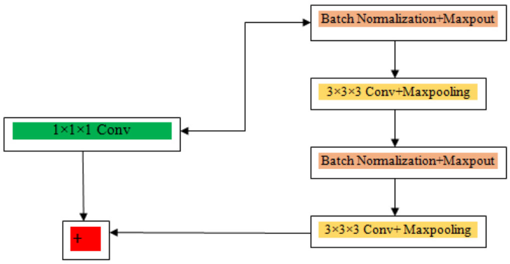

Figure 4

Vestigial Module of the proposed model

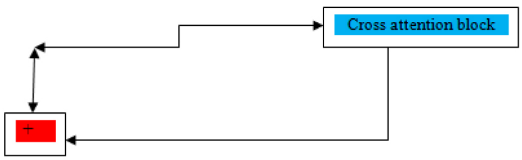

Figure 5

Cross attention block at the bottom of the proposed model

The overall architecture of the proposed model in shown in Figure 6.

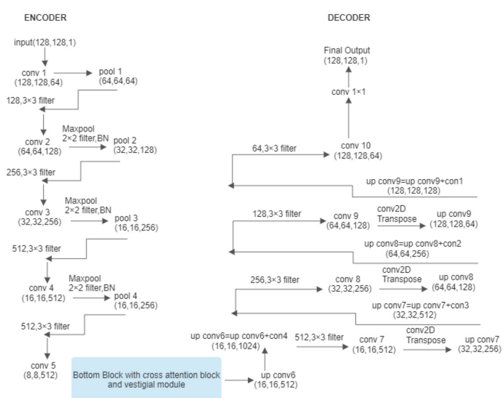

Figure 6

Architecture of our proposed model applicable for multi-modal image segmentation

The first and foremost task is to find out the number of filters required for specific image. This gives the opportunity to add features to the output image of each layer. After gaining the featured images the batch normalization is applied to resize the images. This result then goes for the Maxpool operation to gain the maximum out of it. We express the skip channel as follows: let represent the output of node Z, where *i* represents the down-sampling layer throughout the encoder and *j* represents the convolution layer of the dense block across the skip channel. The collection of feature maps denoted by is calculated as

$$z^{i,j} = \begin{cases} Z(x^{i-1,j}) & j = 0 \\ Z([z^{i,k}]^{j-1}, Z'([z^{i,k}]^{j-1}) & j > 0 \end{cases} \tag{1}$$

5. Datasets

Two sets of datasets are used here. One for the MRI images and the other for the CT Scan images. The MRI images are extracted from BRATs2020 and the CT Scan images were extractedfrom4500 2D CT slices etPreoperative MRI scans from many institutions are used in BraTS 2020, which principally focuses on the segmentation of gliomas, which are intrinsically heterogeneous brain tumors.It also focuses on the prediction of patient overall survival and the differentiation between pseudoprogression and true tumor recurrence via integrative analyses of radiomic features and machine learning algorithm,in order to pinpoint the clinical relevance of this segmentation task.

6. Result

The Dice score was calculated and compared with the previously developed model with single modality as well as multi-modality too. Apart from training our model, we trained the U-Net, Global U-Net and F-S-Net. The models were compared upon the Dice score, Ablation analysis. Apart from these, we used certain parameters like Accuracy Rate (ACC), Loss Function and Predictive Value (PV). Further the optimization was done using Adam Algorithm to gain the better results.The table clearly shows different parameters on the basis of which the models like U-Net, Global U-Net, F-S Net and Improved Global U-Net are compared. Figure 7 represents the precise location of the tumor using Improved Global U-Net.

Table 2

Comparison between various versions of U-Net along with F-S Net

Model	Dice Score	ACC	PV
U-Net	0.8532	0.9823	0.8012
Global U-Net	0.9856	0.9871	0.9976
F-S Net	0.9927	0.9916	0.99781
Improved Global U-Net	0.9935	0.9955	0.99813

It is percitible that the Improved Global U-Net has better conduct than the rest of the models taken under consideration. The Dice score is close to 1 which clearly pictures the pixel match.

The parameters of the Table 2 are calculated as follows:

$$\text{Dice Score} = \frac{2 \times True\_Positive}{2 \times True\_Positive + False\_Negative + False\_Positive} \tag{2}$$

$$\text{ACC} = \frac{True\_Positive + True\_Negative}{True\_Positive + False\_Positive + False\_Negative + True\_Negative} \tag{3}$$

$$PV = \frac{True\_Positive}{True\_Positive + False\_Positive} \tag{4}$$

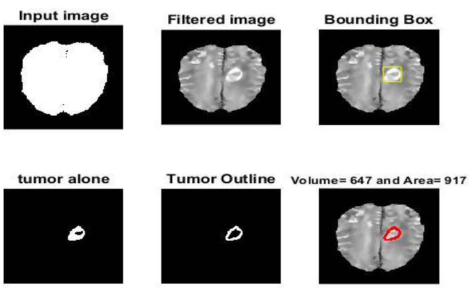

Figure 7

Locating the ROI with respect to brain tumor for MRI image using Improved Global U-Net

Figure 8 shows different division of Class 1 and Class 2 with respect to Training, Validation and Testing ROC for two parameters (True positive and False positive). The figure clearly depicts the clear classification of the classes along the regression line.Another result which is the histogram of several instances versus errors.The errors are considered as the Target (the

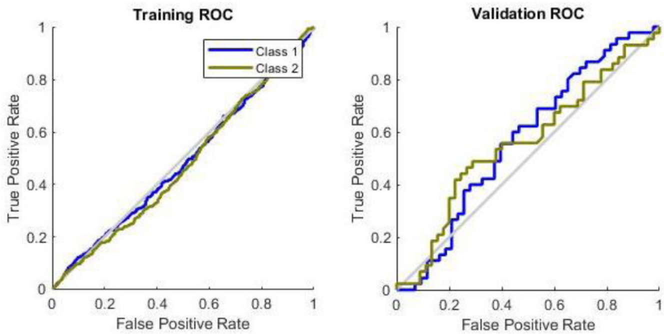

Figure 8

Training, Validation and Testing ROC for two parameters (True positive and False positive)

accurate result) minus the actual output gained for the CT Scan images of brain tumor.

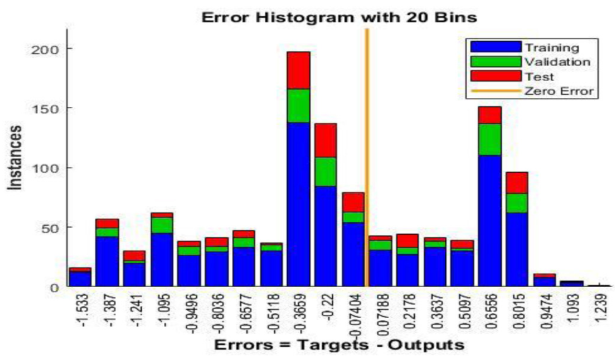

Figure 9

Training, Validation and Testing ROC for two parameters (True positive and False positive)

Figure 9 represents the graphical representation of Error identified. We have taken 20 bins into consideration here. This can be again observed that the number of errors was significantly minimal even if the number of instances were increased using the Improved Global U-Net.

7. Conclusion with Future Scope

The whole procedure of image segmentation for biomedical images is extremely pivotal particularly in the early time of the illness or during the

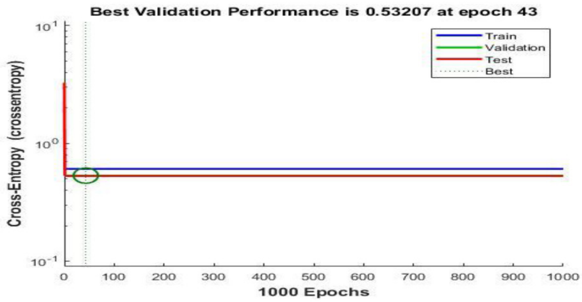

Figure 10

Training, Validation and Testing ROC for two parameters (True positive and False positive)

early signs of the disease. Figure 10 represents the Cross-Entropy achieved for the model for finding the validation performance. Numerous illnesses can be restored on the off chance that they are analysed appropriately and precisely. Hence, Image segmentation especially in case of biomedical images is one of the quickest developing spaces of exploration today. Different models are created and proposed under this area. Among them, U-Net is exceptionally well known as a result of its extensive variety of proficiency and properties. We advanced the U-Net and proposed Improved U-Net. We have applied the model to the MRI as well as the CT Scan images of the brain to segment the tumor area in the image and got the results that were satisfactory in terms of the errors and target output. The model could stand for both types of the images. The result section shows the training, validation and testing parameters and we could see that the difference between the all the three parameters were close to each other and the results were satisfactory. Among many, the significant constraint of U-Net is the gradient descent issue which emerges by adding additional pooling layers. The expansion brings about angle issue as well as makes the restriction issues. In due course, we utilized the Cross attention block to determine the issue and furthermore integrated the beginning model into our proposed model. Together with the Global U-Net model, these approaches resolve the problems of upsampling complexity and spatial data deficit.

References

[1] Mishra, Annu & Gupta, Pankaj & Tewari, Peeyush. Global U-net with amalgamation of inception model and improved kernel variation for

MRI brain image segmentation. *Multimedia Tools and Applications.* 81. 1-16. 10.1007/s11042-022-12094-w (2022).

[2] Pemasiri, A., Nguyen, K., Sridharan, S., & Fookes, C. Multi-modal semantic image segmentation. *Computer Vision and Image Understanding*, *202*, 103085 (2021).

[3] Fowlkes, C., Martin, D., & Malik, J. Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches. In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. (Vol. 2, pp. II-54). IEEE (2003, June).

[4] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks (2013). arXiv preprint arXiv:1312.6229.

[5] Alam, Khursheed & Kumar, Santosh & Kumar, Nitendra& Pandey, Shri & Pal, Dr-Surya. A Nonlinear Hybrid Diffusion Model for Image Denoising. Macromolecular Symposia (2023). 407. 10.1002/masy.202100511.

[6] Guan, X., Yang, G., Ye, J., Yang, W., Xu, X., Jiang, W., & Lai, X. 3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework. *BMC Medical Imaging*, 22(1), 1-18 (2022).

[7] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III (pp. 205-218). Cham: Springer Nature Switzerland (2023, February).

[8] Chaturvedi, R. P., & Ghose, U. A review of small object and movement detection based loss function and optimized technique. *Journal of Intelligent Systems*, 32(1), 20220324 (2023)

[9] Mishra, A., Gupta, P., & Tewari, P. An Amalgamated Deep Learning Approach for Lung Segmentation using X-Ray Images. Tuijin Jishu/ *Journal of Propulsion Technology*, 44(3), 1633-1639 (2023).

[10] Ramachandram, D., & Taylor, G. W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6), 96-108 (2017).

[11] Yang, J., Veeraraghavan, H., Armato III, S. G., Farahani, K., Kirby, J. S., Kalpathy-Kramer, J., ... & Sharp, G. C. Autosegmentation for tho-

racic radiation treatment planning: a grand challenge at AAPM 2017. *Medical Physics*, 45(10), 4568-4581 (2018).

[12] Jiang, Y., Zhang, Y., Lin, X., Dong, J., Cheng, T., & Liang, J. SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer. *Brain Sciences*, 12(6), 797 (2022)

[13] Guo, Z., Li, X., Huang, H., Guo, N., & Li, Q. Deep learning-based image segmentation on multimodal medical imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, *3*(2), 162-169 (2019).

[14] Wang, Z., Zou, N., Shen, D., & Ji, S. Non-local u-nets for biomedical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 6315-6322) (2020, April).

[15] Kumar, Santosh & Kumar, Nitendra& Alam, Khursheed. PDE-based time-dependent model for image restoration with forward-backward diffusivity (2021). 10.1201/9781003167488-56.

[16] Agnes, S. A., & Anitha, J. Efficient multiscale fully convolutional UNet model for segmentation of 3D lung nodule from CT image. *Journal of Medical Imaging*, 9(5), 052402-052402 (2022).

[17] Bjorck, N., Gomes, C. P., Selman, B., & Weinberger, K. Q. Understanding batch normalization. *Advances in neural information processing systems*, 31 (2018).

[18] Alalwan, N., Abozeid, A., ElHabshy, A. A., &Alzahrani, A. Efficient 3d deep learning model for medical image semantic segmentation. *Alexandria Engineering Journal*, 60(1), 1231-1239 (2021).

[19] Manoharan, S. Performance analysis of clustering based image segmentation techniques. *Journal of Innovative Image Processing* (JIIP), 2(01),14-24 (2020).

[20] Sujitha, S., Jayakumar, T., Maheskumar, D., & Kaviyan, E. V. Mathematical Model of Brain Tumor With Radiotherapy Treatment. Communications in Mathematics and Applications, 14(2), 1039 (2023).

[21] Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., &Terzopoulos, D. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and MachineIntelligence (2021).

[22] Pradhan, Pranita, et al. "Semantic segmentation of non-linear multimodal images for disease grading of inflammatory bowel disease: A segnet-based application." 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM), February 19-21,

2019, Prague, Czech Republic. [Sétubal]: SCITEPRESS-Science and Technology Publications Lda. (2019).

[23] Su, R., Liu, J., Zhang, D., Cheng, C., & Ye, M. Multimodal glioma image segmentation using dual encoder structure and channel spatial attention block. Frontiers in Neuroscience, 14, 586197 (2020).

[24] Chaturvedi, Ravi Prakash, and Udayan Ghose. "Small object detection using retinanet with hybrid anchor box hyper tuning using interface of Bayesian mathematics." *Journal of Information and Optimization Sciences* 43.8 : 2099-2110 (2022).

[25] Chaturvedi, Ravi Prakash & Ghose, Udayan. An effective framework for detecting the object from the video sequences by utilizing deep learning with hybrid technology, *Journal of Information and Optimization Sciences*, 44:1, 113-126 (2023).

[26] Mishra, A., Gupta, P., & Tewari, P. Biomedical Image Segmentation Using Integrated FCM Clustering Modified with Regularized Level Set Method. In 2023 International Conference on Disruptive Technologies (ICDT) (pp. 344-348). IEEE (2023, May).

[27] Chaturvedi, R. P., &Ghose, U. A novel model of texture pattern based object identification using convoluted multi-angular (CMA) pattern extraction method (2022).

[28] Ansari, M. A., Mehrotra, R., & Agrawal, R. Detection and classification of brain tumor in MRI images using wavelet transform and support vector machine. *Journal of Interdisciplinary Mathematics*, 23(5), 955-966 (2020).

Received, November 2023