Vol. 4 Issue No. 3, July-September 2022 e-ISSN 2456-7701

Journal of Science and Technological Researches

A Peer Reviewed Journal

Origin of Innovation

Domain: www.jstr.org.in, Email: editor@jstr.org.in

ADD-UNet A CONVOLUTIONAL NEURAL NETWORK FOR SEMANTIC SEGMENTATION OF MEDICAL IMAGES WITH ATROUS SPATIAL PYRAMID POOLING, ATTENTION-GUIDED DENSE UP-SAMPLING, AND DENSE CONTEXT LOCAL CONVOLUTION

M. Prasad[1*] and Vijaya J[2]

[1]Dept. of Biosciences and Bioengineering, Indian Institute of Technology Bombay

[2]Dept. of Data Science and Artificial Intelligence, International Institute of Information Technology, Naya Raipur

Email: mahavirmoond2001@gmail.com, vijaya@iiitnr.edu.in

Date of Received	**02 September, 2022**
Date of Revised	**19 September, 2022**
Date of Acceptance	**25 September, 2022**
Date of Publication	**30 September, 2022**

DOI : https://doi.org/10.51514/JSTR.4.3.2022.8-16

"together we can and we will make a difference"

I-3 Vikas Nagar, Housing Board Colony, Berasia Road, Karond Bhopal-462038

Domain: www.jstr.org.in, Email: editor@jstr.org.in, Contact: 09713990647

e-ISSN 2456-7701
Vol. 4 Issue No. 3, July-September 2022

ADD-UNet: A CONVOLUTIONAL NEURAL NETWORK FOR SEMANTIC SEGMENTATION OF MEDICAL IMAGES WITH ATROUS SPATIAL PYRAMID POOLING, ATTENTION-GUIDED DENSE UP-SAMPLING, AND DENSE CONTEXT LOCAL CONVOLUTION

M. Prasad[1]* and Vijaya J[2]

[1]Dept. of Biosciences and Bioengineering, Indian Institute of Technology Bombay
[2]Dept. of Data Science and Artificial Intelligence, International Institute of Information Technology, Naya Raipur

Email: mahavirmoond2001@gmail.com

ABSTRACT

Medical Imaging has widely revolutionized the medical practices for the diagnosis states such as cancer diagnosis and planning for the methodology for the treatment and monitoring. With the progress in medical imaging technologies, the size of high-quality medical data is increasing. Machine learning, particularly Deep Learning has set a new scope for understanding and utilizing medical imaging data smartly and extracting clinical information. Convolutional Neural Networks (CNNs) provides several methods for medical image segmentation to understand the desirable features and decision functions. U-Net is one of the top-performing convolution neural networks for exceptional segmentation of medical images. This paper proposes an ADD-UNet model based on a novel atrous spatial pyramid pooling, attention-guided dense up-sampling, and dense context local convolution for the semantic segmentation of medical images. We re-design the existing U-Net architecture to conserve the network's spatial information and apply the dense local contextual information methods for the fine recovery of localization information. An atrous spatial pyramid pooling is utilized to replace ordinary convolution filters in U-Net for extracting multi-scale information at the encoder-decoder portion of the process because it conserves the receptive field and details of the image in the network by inserting spaces in the convolution filters. It is possible to build intricate feature representations for dense prediction by precisely restoring the spatial resolution using an attention-guided dense up-sampling technique. The low details of the feature representations can be recovered to the input resolution for pixel-wise classification, which is useful for precise border localization, by learning the up-sampling procedure. A dense local context convolution is used on the output feature representation to extract multi-level features of the segmented image. We utilized the TCIA-TGCA Brain MRI Dataset to train and validate the proposed network and achieved 81.0%, 89.4%, 80.5%, and 89.1% in Training IoU, Train Dice Score, Testing IoU, and Testing Dice Score respectively.

***Keywords:** ADD-UNet, Atrous Spatial Pyramid Pooling, Attention-guided Dense Up-sampling, Dense C ontext Local Convolution, Medical Image Segmentation*

INTRODUCTION

Medical Imaging refers to the techniques that generate images of the external and internal biological tissues or a part of the human body to detect, diagnose, and treat diseases. The most commonly used medical imaging technologies include MRI Scans, Ultra-Sound, X-Ray, CT scans, etc. and the information density has increased. Imaging modalities such as MRI are used to examine multiple human organs while Retinal Photography is an example of organ-specific medical imaging modalities [1-5]. With the progress in medical imaging, the size of high-quality medical data is increasing that can be utilized for computations and predictive analysis. Many health advancements, such as novel medical techniques, the management of patient data, and the management of chronic diseases, heavily rely on machine learning. Machine Learning brings a new promise to medical practices with medical imaging data [5,6,7]. Particularly, Deep Learning has set a new scope for understanding utilizing the medical imaging data smartly as well as extracting clinical information. The key benefit of deep learning is that it automatically discovers complex relationships in data without the need for manual feature identification. MRI scans are clinically significant for the diagnosis and treatment of brain tumors. The extremely diverse shape and appearance of these tumors make it difficult to segment the sub-regions. There is a staggering growth in research on deep learning in computer vision for medical images that are improving clinical impact [5]. Deep Learning architectures provide various approaches for medical image segmentation.

*Author for correspondence

Segmentation provides the pixel-wise classification of the objects and scene semantics in an image [5, 8-10]. Recently Deep Convolution Neural Networks has produced promising results for semantic image segmentation with the goal of pixel-wise classification by assigning object labels at the pixel level. [11-15]. Modern state-of-the-art techniques for segmenting images semantically use an encoder-decoder structure. A fully convolutional network is employed in the encoder section to extract features. The decoder part is utilized to up-sample the feature representations acquired by the encoder part into a segmented output.

Literature Review

Applications for deep learning in healthcare address a diverse range of concerns, from tailored therapy recommendations to infection surveillance and cancer detection. Physicians now have access to a vast amount of data from many sources, including genetic sequencing, pathological imaging, and radiological imaging. However, to transform all of this information into useful knowledge. Deep learning uses neural networks made up of several convolutional nodes made of artificial neurons to discover patterns in data structures. This straightforward approach uses a nonlinear activation function before a linear source regular expression form. There are several frequently used nonlinear activation functions in a network, including the sigmoid conversion, ReLU, their variations, and hyperbolic tangent. Deep learning's development roots may be found in the work of Walter Pitts and Warren McCulloch (1943).

In U-Net proposed by Olaf Ronneberger et al. in 2015 described a network that focuses on data augmentation. An asymmetric expanding path and a contracting path that collect context and enable exact localization, respectively, constitute the layout [16].

A newer, more potent DCNN architecture for medical picture segmentation was presented by Zongwei Zhou et al. in 2018 [17]. The semantic imbalance between the encoder and decoder's feature representations is filled using skip pathways.

Dense concatenation was utilized by Sijing Cai et al. in 2020 [18]. In their proposed Dense-UNet, a novel CNN-based architecture for cellular picture segmentation to enable feature reuse and deepen the network design. This model comprises four expansion modules with four down-sampling layers each to extract features.

Three contributions from the work of Liang-Chieh Chen et al. in 2017 show significant practical usefulness [19]. They began by emphasizing the Atrous Convolution for the dense prediction problems. They also suggested the atrous spatial pyramid pooling to partition the objects at various sampling rates and useful fields of vision. Third, they used techniques from DCNNs and probabilistic graph models to enhance the localization of object boundaries

For the segmentation task, Liang-Chieh Chen et al. in 2017 took into account two different neural networks: one is a spatial pyramid pooling module that encodes contextual information by pooling the incoming features with pooling operations and filters at a different resolution, and the other is an encoder-decoder structure that can obtain the sharp object boundaries [20]. The DeepLab3+ model they suggest, which would refine the results of image segmentation, combines the benefits of both neural networks.

In a neural network, to conserve the spatial details of the image, K. M. Sediqi and Hyo Jong Lee in 2021 proposed a dense up-sampling convolution [21]. They also proposed a brand-new local context convolution method that precisely covers boundary delineation.

Proposed Work

In this proposed work, we address the issue of the spatial information loss brought on by convolution and pooling operations with a stride in U-Net. We propose ADD-UNet that conserves spatial features of the image, and also provides precise boundary delineation for better segmentation. We use high-resolution brain tumor images that contain rich details to be maintained in the output segmented image. Hence, we propose this network based on the Atrous Spatial Pyramid Pooling (ASPP), Attention-guided Dense Up-sampling Convolution (ADUC), and Dense Local Context Convolution (DLC)

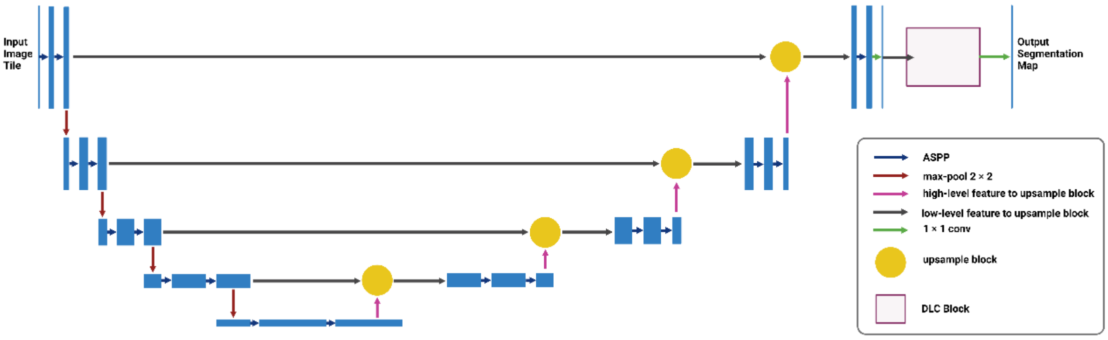


Fig. 1. The network architecture of the proposed ADD-UNet.

Therefore, to extract the image feature representations we propose a novel ADD-UNet model that uses three precise image instance segmentation strategies described as follows:

1. Atrous Spatial Pyramid Pooling (ASPP): We input the image through Spatial Pyramid Pooling the final input image is the scale. We replaced the ordinary 3 × 3 convolutions in U-Net with Atrous Convolution which expands the receptive field of the image in the network and obtains high-level semantic image features.
2. Attention-guided Dense Up-Sampling Block: We have introduced a Dense Up-sampling Block that utilizes the bilinear up-sampling that replaces the deconvolution up-sampling method utilized in U-Net which is not effective compared to bilinear up convolution. In this block, high-level feature representations are up-sampled and concatenated with the low-level feature representation from the skip connection after passing a convolution layer. Also, a Squeeze-and-Excitation Net [22] is utilized in the process, which explicitly models channel interdependencies and adaptively recalibrates channel-wise feature responses. This Up-sampling block enables the network to extract all important features from the low-level and high-level feature representation. It increases the target region's segmentation's accuracy and effectiveness.
3. Dense Local Context Convolution (DLC) Block: We also employ a dense local context Convolution block on the output feature representation at the end of the decoder block using several atrous convolutions combined in concurrent and concatenated form, we extract multi-level features of the segmented image.

Detailed Framework

Integration of ASPP

We utilized the ASPP module introduced in DeepLabv3 to have a larger receptive field for the feature extraction without image resolution loss because the resolution loss causes the loss of detailed image boundary information. Atrous convolutions can be employed as a method to control the feature resolution and to obtain a bigger perceptron without compromising resolution. Multiple concurrent atrous convolutions with varying rates constitute ASPP. It combines atrous convolution with spatial pyramid pooling that captures contextual information at different rates to extract multiscale features for precise image segmentation [19-20, 25].

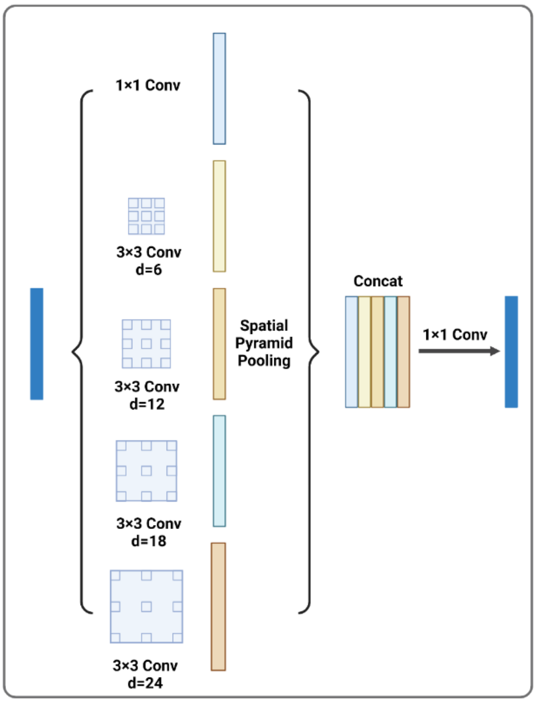


Fig. 2. : An overview of an atrous spatial pyramid pooling utilized in ADD-UNet to replace the ordinary convolution.

Proposed ASPP modules consist of four 3 × 3 concurrent atrous convolution layers with various rates of 6, 12, 18, and 24 respectively, and one 1 × 1 convolution. By average pooling of these five concurrent atrous convolutions, an image-level feature is produced.

Attention-guided Dense Up-Sampling Block

To recompense for the information loss caused by bilinear up-sampling, we utilized the attention-guided dense-up-sampling block that successfully merges the low-level and high-level features and simultaneously emphasizes the rich detail channels. The original U-Net utilized deconvolution to up-sample the high-level feature representation which is not effective compared to bilinear up convolution. This Up-sampling Block utilizes the bilinear up-sampling.

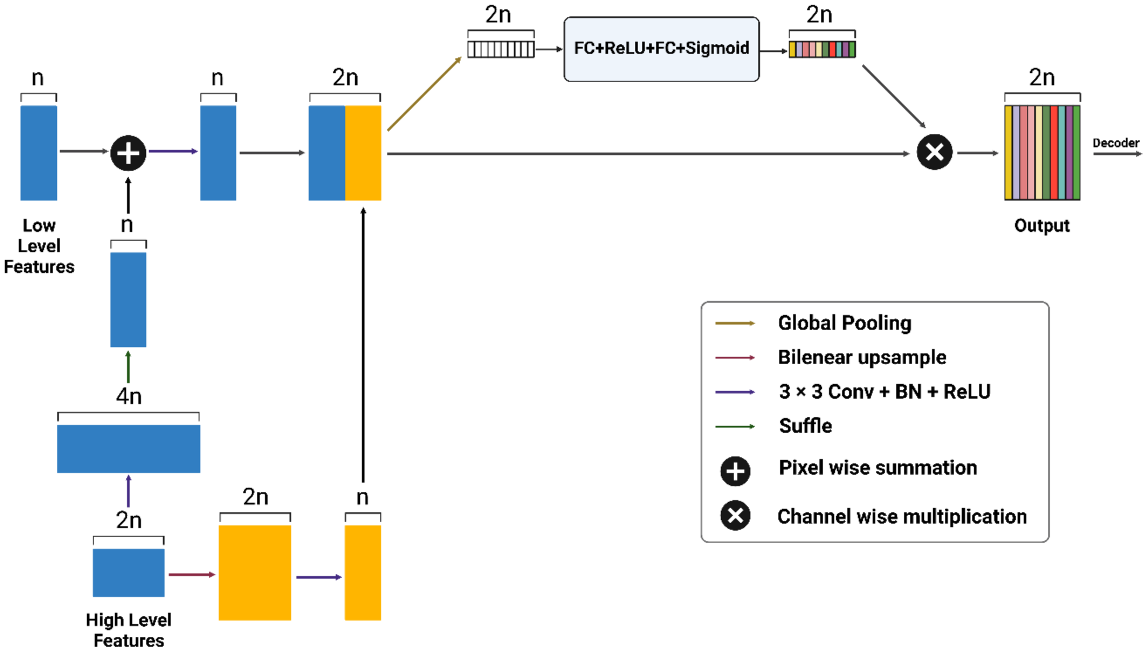


Fig. 3 : An illustration of an attention-guided dense up-sampling.

The first step involves up-sampling the low-resolution high-level feature representations (F_{high}) using two distinct techniques. One method is called dense up-sampling convolution (F_{duc}), while the other is called bilinear up-sampling with a convolution layer (F_{buc}). Batch normalization and ReLU activation are applied following the convolution layer, unless otherwise specified. The low-level characteristics (F_{low}) are then integrated with F_{duc} through summing (F_{sum}). Before F_{sum} and F_{buc} are concatenated (F_{concat}), a convolution layer is used to smooth the concatenation. Thus, F_{concat} includes the data from both F_{high} as well as, F_{low}. The next step is to select the required data from F_{concat}. We employ channel-wise attention, which is inspired by the Squeeze-and-Excitation Network.

Dense Local Context Convolution Block

We utilized a module to extract the dense multi-scale features to adapt to large-scale fluctuations. The localized unit enhances the multi-scale representation of objects in an image by combining the benefits of concurrent and concatenated dilated convolution layers for a deeper and richer perceptron [24, 27].

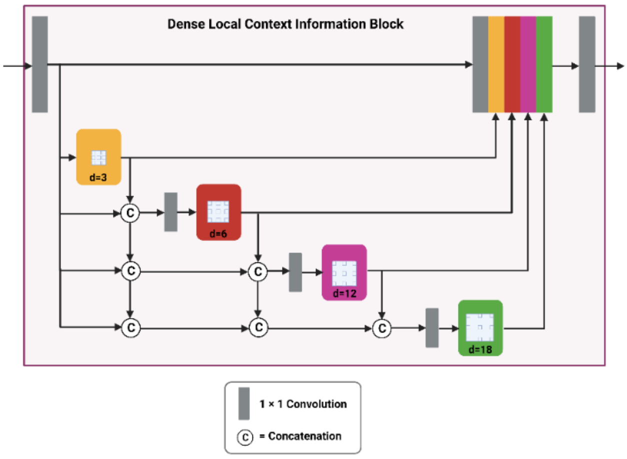


Fig. 4 : An illustration of a dense local context convolution block with numerous atrous convolutions combined in concurrent and concatenated form is utilized on the decoder output.

The final output of the decoder block is then used to sequentially convolve the input feature representation with dilated convolution at increasing dilation rates of 3, 6, 12, and 18. Concatenating the input feature representation with the output of earlier convolutions produces the input for each dilated convolution layer, which is ultimately convolved with a 1 × 1 convolution. Finally, the input feature representation is concatenated with the outputs from the four dilated convolutions [24-25].

Loss Function

We used the Jaccard Loss to the weight of the false predictions during the training of the model. This loss function evaluates each pixel measuring the distance between its ground truth $y \in \{0, 1\}$ and the current result of the model y. We define the Jaccard Loss Function (J_d) as given:

$$J_d(y, \hat{y}) = 1 - \frac{(y.\hat{y}) + \varepsilon}{(y + \hat{y} - y.\hat{y}) + \varepsilon}$$

Where, ε prevents zero division

or $J_d = 1 - IoU$

Where IoU is an evaluation index called Intersection-Over-Union (IoU) or Jaccard Index.

Evaluation Index

The evaluation matrices are intended to evaluate how well the proposed architecture is working. The performance is assessed in this research using the most used F-measure-based evaluation indices for medical image segmentation. We have used Intersection-Over-Union and the Dice Score to describe how the approaches perform on the test and test datasets to thoroughly compare our final proposed network to the current networks.

IoU stands for the intersection over union ratio between the expected segmentation and the actual segmentation. The ratio can be expressed as the intersection or number of pixels that overlap true positives divided by the sum of true positives, false positives, and false negatives (union).

The Dice Score is used to assess how well anticipated segmentation and actual segmentation match up pixel-for-pixel. The ratio can be expressed as the intersection or number of pixels that overlap twice as often true positives as false positives and negatives.

The brain tumor mask pixels are positive and the background pixels are negative because the brain tumor segmentation is thought to be a semantic segmentation problem. As a result, each prediction can be categorized into one of four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (FN). TP stands for the quantity of successfully identified brain tumor mask pixels. TN stands for the quantity of accurately identified background pixels. The number of background pixels designated as brain tumor pixels is given by the abbreviation FP. The number of brain tumor pixels designated as the background is known as FN. We define IoU and Dice Score as:

$$IoU = \frac{TP}{TP + FP + FN}$$

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Data Set

In this paper, the dataset is obtained from the TCIA and TCGA. In this collection, brain MR images and manual FLAIR abnormality segmentation masks are included. They relate to 110 individuals who are part of the TCGA lower-grade glioma collection and have one FLAIR sequence and one genomic cluster accessible. Each image is delivered in the ".tif" format and has three channels. Pre-contrast, FLAIR, and post-contrast sequences are each provided for 101 instances (in this order of channels). Both the pre- and post-contrast sequences are absent in 6 instances and 9 cases, respectively. All pictures are made 3-channel by replacing missing sequences with FLAIR sequences. One-channel, binary pictures are used as masks. They separate the present FLAIR anomaly from the FLAIR sequence (available for all cases).

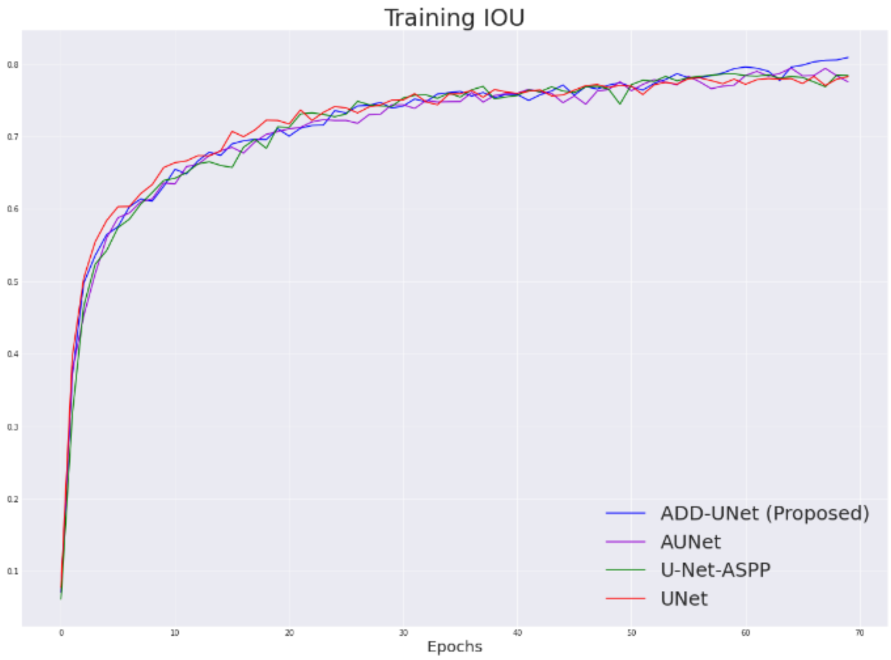
The image data sets with a resolution of 256 × 256. 20% of the images were used for the training of the networks with a training set and 5% of the images were used for the testing of the networks with a testing set.

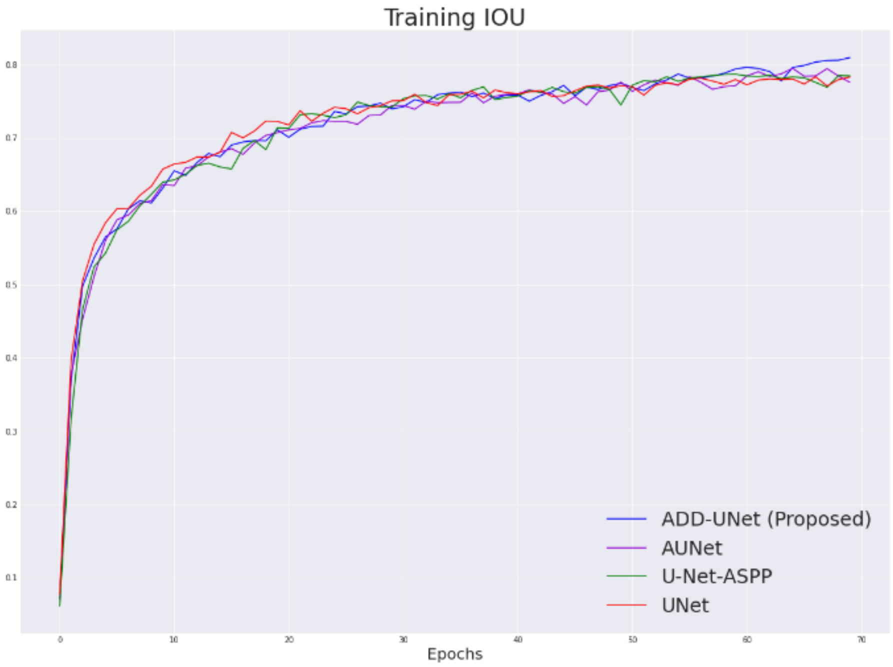
Experiment and Analysis

We evaluated the proposed ADD-UNet, U-Net [16], U-Net ASPP [25], and AUNet [26] on TCIA-TGCA Brain MR Images. From the comparison between different networks, the ADD-UNet achieves 81.0%, 89.4%, 80.5%, and 89.1% in Training IoU, Train Dice Score, Testing IoU, and Testing Dice Score respectively is better than U-Net, U-Net-ASPP, AUNet. The performance comparisons are summarized in Table 1. These results show that the proposed ADD-UNet facilitates the best image segmentation on a given dataset. We achieve the highest evaluation indices for our proposed ADD-UNet.

Model / Network	Training IoU	Training Dice Score	Testing IoU	Testing Dice Score
U-Net	78.5%	87.4%	73.2%	84.2%
U-Net-ASPP	78.6%	87.7%	78.0%	86.9%
AUNet	79.2%	88.2%	79.1%	87.8%
ADD-UNet (Proposed)	**81.0%**	**89.4%**	**80.5%**	**89.1%**

Table 1: The performance comparison between U-Net, U-Net-ASPP, AUNet, and proposed ADD-UNet..





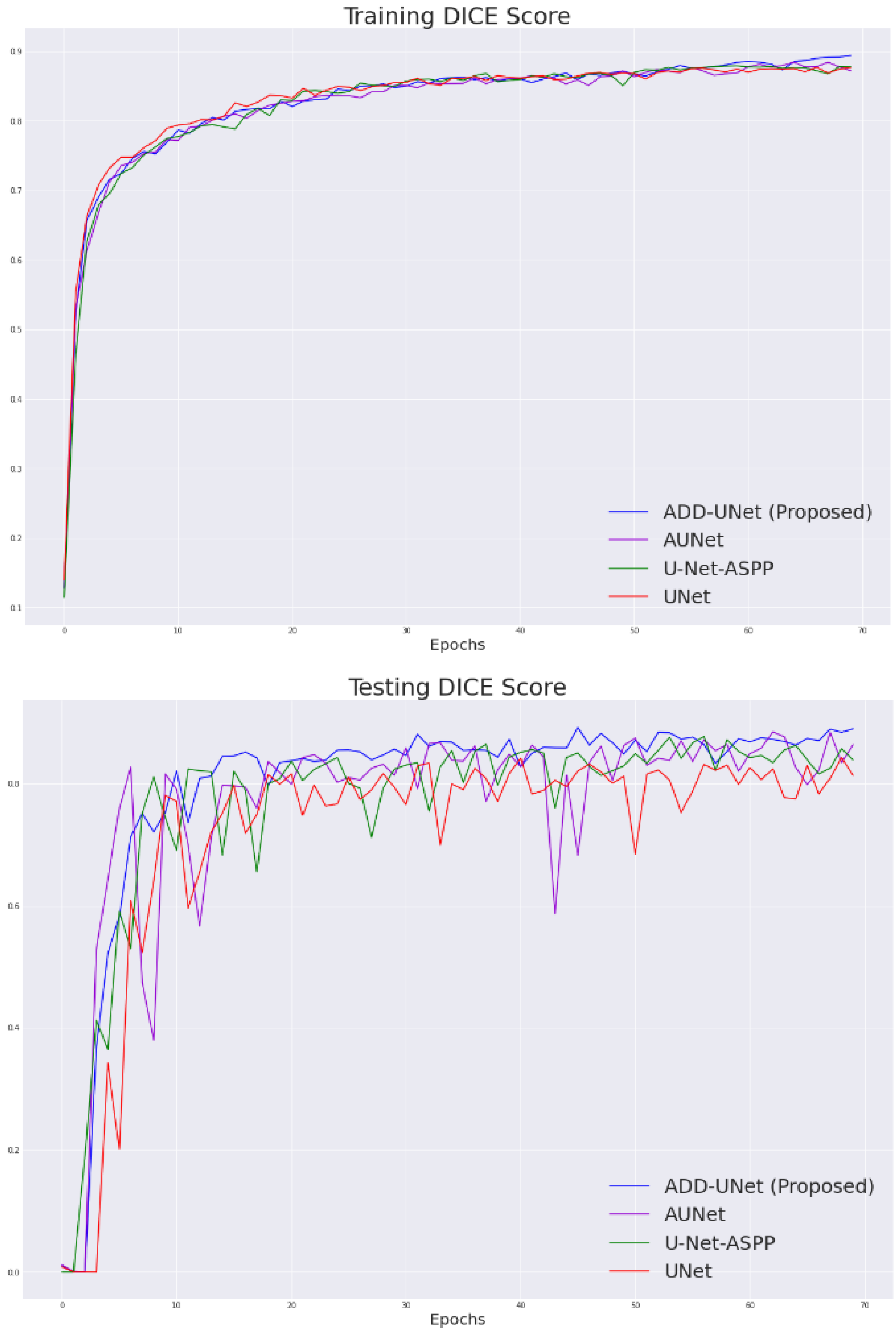


Fig. 5 : Performance Comparison of UNet, U-Net-ASPP, AUNet, and our proposed ADD-UNet based on evaluation indices Training IoU, Training Dice Score, Testing IoU, and Testing Dice Score (For interpretation of the different networks to the color of the legend in the figure).

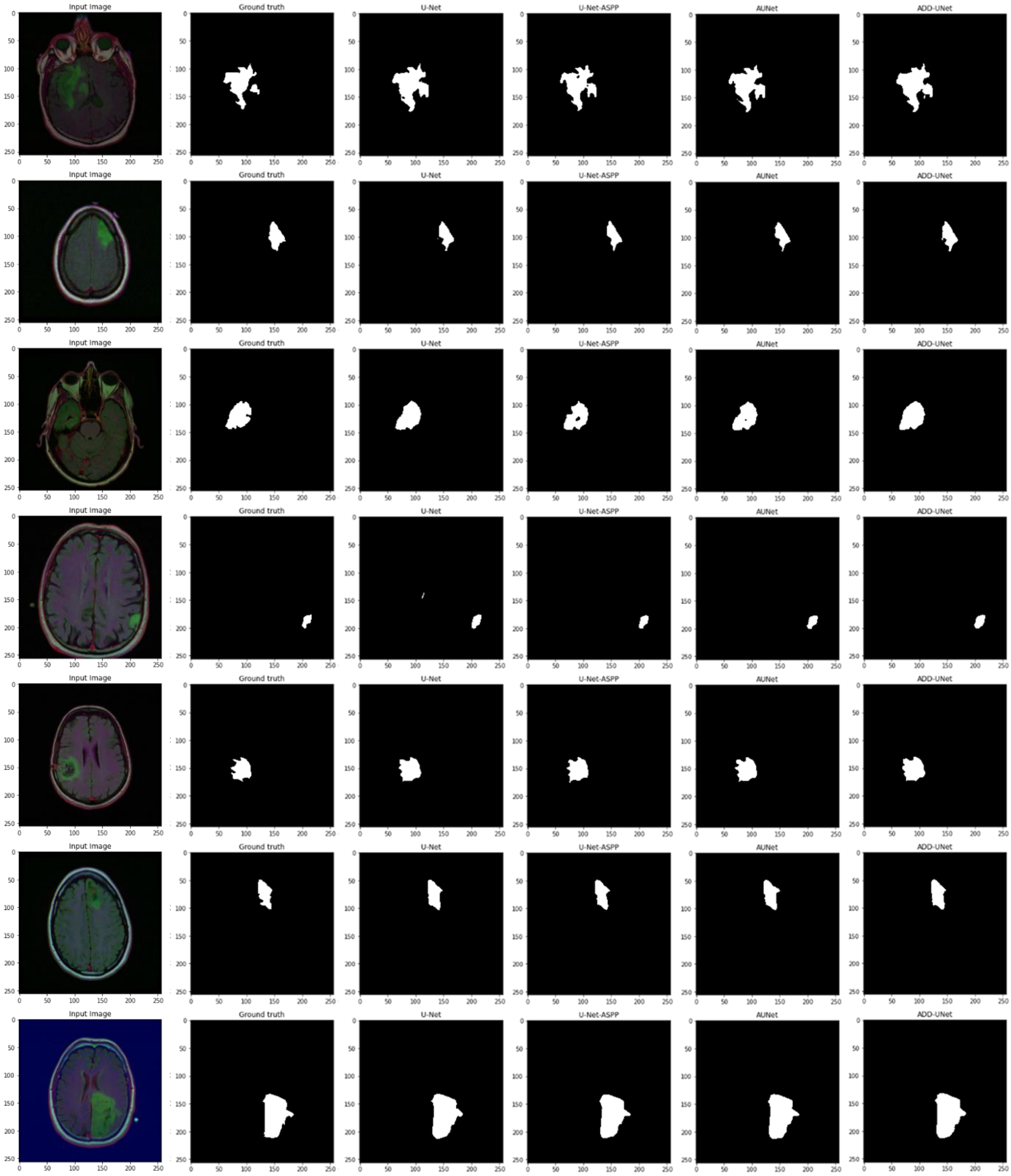


Fig. 6: Visualisation of segmentation outcomes. The columns represent, in order from left to right, the input images, the ground truth, and the segmentation outcomes of UNet, U-Net-ASPP, AUNet, and proposed ADD-UNet.

CONCLUSION

In the proposed ADD-UNet for medical image segmentation. We have utilized Atrous Spatial Pyramid Pooling which replaced the ordinary 3 × 3 convolutions which expands the image receptive field without resolution loss and recovers high-level semantic image features. The original U-Net utilized deconvolution to up-sample the high-level feature representation which is not effective compared to bilinear up convolution. We have introduced a Dense Up-sampling Block that utilizes the bilinear up-sampling which improves the precision and efficiency of segmenting the target region. We further use a DLC block on the output feature representation at the end of the decoder block using several atrous convolutions combined in concurrent and cascade form to extract multi-level features of the segmented image. Our network best performs on TCIA-TGCA Brain MR Images with comparison to different existing networks and achieved 81.0%, 89.4%, 80.5%, and 89.1% in Training IoU, Train Dice Score, Testing IoU, and Testing Dice Score respectively.

REFERENCES

[1]. Beutel, J., Kundel, H. L., Kim, Y., Van Metter, R. L., & Horii, S. C. (2000). *Handbook of medical imaging* (Vol. 3). Spie Press.

[2]. Suetens, P. (2017). *Fundamentals of medical imaging*. Cambridge university press.

[3]. Acharya, R., Wasserman, R., Stevens, J., & Hinojosa, C. (1995). Biomedical imaging modalities: a tutorial. *Computerized Medical Imaging and Graphics*, *19*(1), 3-25.

[4]. Sonka, M., & Fitzpatrick, J. M. (2000). Handbook of medical imaging. Volume 2, Medical image processing and analysis. SPIE.

[5]. Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, *21*(1), 4-21.

[6]. Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology*, *15*(3), 512-520.

[7]. Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, *37*(2), 505.

[8]. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*(1), 1-74.

[9]. Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), 102-127.

[10]. Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., ... & Giger, M. L. (2019). Deep learning in medical imaging and radiation therapy. *Medical physics*, *46*(1), e1-e36.

[11]. Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, *32*(4), 582-596.

[12]. Lai, M. (2015). Deep learning for medical image segmentation. *arXiv preprint arXiv:1505.02000*.

[13]. Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, *13*(3), 1224.

[14]. Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, *19*, 221.

[15]. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

[16]. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[17]. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3-11). Springer, Cham.

[18]. Cai, Sijing, et al. "Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network." *Quantitative imaging in medicine and surgery* 10.6 (2020): 1275.

[19]. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834-848.

[20]. Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

[21]. Sediqi, K. M., & Lee, H. J. (2021). A novel upsampling and context convolution for image semantic segmentation. *Sensors*, *21*(6), 2170.

[22]. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

[23]. Lian, X., Pang, Y., Han, J., & Pan, J. (2021). Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. *Pattern Recognition*, *110*, 107622.

[24]. Lin, C. Y., Chiu, Y. C., Ng, H. F., Shih, T. K., & Lin, K. H. (2020). Global-and-local context network for semantic segmentation of street view images. *Sensors*, *20*(10), 2907.

[25]. Qiu, X. (2022). U-Net-ASPP: U-Net based on atrous spatial pyramid pooling model for medical image segmentation in COVID-19. *Journal of Applied Science and Engineering*, *25*(6), 1015-1024.

[26]. Sun, H., Li, C., Liu, B., Liu, Z., Wang, M., Zheng, H., ... & Wang, S. (2020). AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. *Physics in Medicine & Biology*, *65*(5), 055005