

Rowan University

Rowan Digital Works

Henry M. Rowan College of Engineering Faculty
Scholarship

Henry M. Rowan College of Engineering

8-15-2021

Dilated Inception U-Net (DIU-Net) for Brain Tumor Segmentation

Daniel E. Cahall

Ghulam Rasool

Rowan University, rasool@rowan.edu

Nidhal Carla Bouaynaya

Rowan University, bouaynaya@rowan.edu

Hassan M. Fathallah-Shaykh

Follow this and additional works at: https://rdw.rowan.edu/engineering_facpub



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Daniel E. Cahall, Ghulam Rasool, Nidhal C. Bouaynaya, Hassan M. Fathallah-Shaykh. 2021. Dilated Inception U-Net (DIU-Net) for Brain Tumor Segmentation. arXiv:2108.06772 [eess.IV]

This Article is brought to you for free and open access by the Henry M. Rowan College of Engineering at Rowan Digital Works. It has been accepted for inclusion in Henry M. Rowan College of Engineering Faculty Scholarship by an authorized administrator of Rowan Digital Works.

Dilated Inception U-Net (DIU-Net) for Brain Tumor Segmentation

D. Cahall¹, G. Rasool¹, N. C. Bouaynaya¹, and H. M. Fathallah-Shaykh²

1. Department of Electrical and Computer Engineering, Rowan University, New Jersey, USA

2. Departments of Neurology and Mathematics, University of Alabama at Birmingham, USA
danielenicocahall@gmail.com, rasool@rowan.edu, bouaynaya@rowan.edu, hfshaykh@uabmc.edu

Abstract— Magnetic resonance imaging (MRI) is routinely used for brain tumor diagnosis, treatment planning, and post-treatment surveillance. Recently, various models based on deep neural networks have been proposed for the pixel-level segmentation of tumors in brain MRIs. However, the structural variations, spatial dissimilarities, and intensity inhomogeneity in MRIs make segmentation a challenging task. We propose a new end-to-end brain tumor segmentation architecture based on U-Net that integrates Inception modules and dilated convolutions into its contracting and expanding paths. This allows us to extract local structural as well as global contextual information. We performed segmentation of glioma sub-regions, including tumor core, enhancing tumor, and whole tumor using Brain Tumor Segmentation (BraTS) 2018 dataset. Our proposed model performed significantly better than state-of-the-art U-Net-based model ($p < 0.05$) for tumor core and whole tumor segmentation.

I. INTRODUCTION

Machine learning techniques based on deep neural networks have become increasingly common in the medical imaging field in recent years [1]. One of the challenging problems in medical imaging is the pixel level segmentation of various biological structures in a given image, e.g., segmentation of brain tumors in MRIs [2, 3]. Accurate and timely segmentation of brain tumors can help physicians with the diagnosis, treatment planning, and post-treatment surveillance [2].

The accurate segmentation of various structures in an image is dependent upon the extraction of local structural and global contextual information. Several multi-path architectures have been proposed in the medical image segmentation literature which extract information from given data at multiple scales [4–6]. U-Net, proposed by Ronneberger et al., is commonly used for the segmentation of various structures in medical images [7]. U-Net is built using (1) a contracting path, which captures high-resolution, contextual features while downsampling at each layer, and (2) an expanding path, which increases the resolution of the output through upsampling at each layer [7]. The features from the contracting path are fused with features from the expanding path through long skip connections, ensuring localization of the extracted contextual features [8]. U-Net was originally developed and applied to cell tracking; however, more recently, the model has been applied to other medical segmentation tasks, such as

brain vessel segmentation, brain tumor segmentation, and retinal segmentation [9–11]. Variations of U-Net, such as 3D U-Net, GRA U-Net, RIC-UNet, PsLSNet, and SDResU-Net, have been proposed to tackle different segmentation problems in medical imaging [6, 12–14].

The concept of extracting and aggregating features at multiple scales has also been accomplished by Inception modules [15]. However, the mechanism of multi-scale feature extraction is different compared to multi-path architectures [4–6]. Each Inception module applies filters of various sizes at each layer and concatenates resulting feature maps [15]. Inception modules within U-Net have also been recently proposed for brain tumor segmentation, left atrial segmentation, and liver segmentation [3, 16–18].

Several extensions and modifications to the Inception module have been proposed, such as dilated (also known as atrous) convolutions [19]. Dilated convolutions enable the learned filters in a convolutional neural network (CNN) to have larger receptive fields with fewer parameters, thereby reducing the computational cost. Inception modules using dilated convolutions have also been utilized to improve image resolution, visual saliency prediction, change detection in multi-sensor images, and learning optical flow [20–23]. Recently introduced dilated residual Inception block accomplish multi-scale feature extraction in an end-to-end, fully convolutional retinal depth estimation model [24].

We introduce an end-to-end brain tumor segmentation framework based on U-Net architecture with dilated Inception modules, referred to as Dilated Inception U-Net (DIU-Net), to accomplish multi-scale feature extraction. We demonstrate that integrating dilated convolutions within Inception modules results in significant improvement ($p < 0.05$) in the segmentation of two of the three glioma sub-regions, i.e., tumor core and whole tumor.

II. METHODS

II-A. Dilated Inception U-Net (DIU-Net) Architecture

We propose to integrate dilated convolutions and Inception modules in the U-Net architecture [20]. In our settings, each dilated Inception module consists of three 1×1 convolution operations, each followed by one l -dilated convolutional filter with $l = 1, 2, \text{ and } 3$. The 1×1

convolution filters perform dimensionality reduction, while three l -dilated convolutional filters each of size 3×3 implement atrous convolutions. The schematic layout of a dilated inception module is provided in Fig. 1A and a detailed description of dilated convolution filters is provided in Section II-B. Finally, we use the rectified linear unit (ReLU) as the activation function and performed batch normalization in each dilated Inception module [25].

In Fig. 1, we present the detailed architecture of our proposed DIU-Net. We used a contracting-expanding architecture, resembling U-Net, with a bottleneck in the middle. The number of filters double at each layer on the contracting side and halve on the expanding side. On the other hand, the size of the output feature map (height and width) halves on the contracting side and doubles on the expanding side. We perform downsampling using max-pooling on the contracting path and upsampling on the expanding path. We also perform feature concatenation on the expanding path, i.e., features from the corresponding layer of the contracting path are concatenated with those on the expanding path. At the last layer on the expanding path, the output height and width are equal to the height and width of the original input images. At the output, we perform 1×1 convolutions to reduce the depth of the last feature map equal to the number of segmentation classes (i.e., tumor regions). Finally, a pixel-wise activation is performed to convert feature maps into binary segmentation outputs.

II-B. Dilated Convolutions

We consider an image I of size $m \times n$ and a discrete convolutional filter w of size $k \times k$. The linear convolution operation between the image I and the filter w is given by:

$$(I * w)(p) = \sum_s I[p+s]w[s]. \quad (1)$$

The simple convolution operation can be generalized to l -dilated convolution ($*_l$) as [19]:

$$(I *_l w)(p) = \sum_s I[p+ls]w[s]. \quad (2)$$

It is evident that for $l = 1$, we get the the simple convolutional operation given in 1. However, for $l > 1$, $l - 1$ zeroes are inserted between each filter element, creating a $k_s \times k_s$ scaled and sparse filter, where k_s is defined as:

$$k_s = k + (k - 1)(l - 1), \quad (3)$$

$$= l(k - 1) + 1. \quad (4)$$

The scaling s increases the receptive field of the filter by a factor $\left(\frac{k_s}{k}\right)^2$.

$$\frac{k_s}{k} = \frac{k + (k - 1)(l - 1)}{k}, \quad (5)$$

$$= l + \left(\frac{-l + 1}{k}\right). \quad (6)$$

The receptive field of the filter increases linearly with l , while the number of elements ($k \times k$) remains fixed. In Fig. 2, we present l -dilated convolution filters of size 3×3 for $l = 1, 2$, and 3.

II-C. Dataset and Pre-Processing

We used BRATS 2018 dataset for our experiments [26]. The dataset includes MRIs of 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) patients. Each patient's data consists of four MRI sequences: T2-weighted (T2), T1, T1 with gadolinium enhancing contrast (T1C), and Fluid-Attenuated Inversion Recovery (FLAIR) images. BRATS also provides pixel-level manual segmentation markings for three *intra-tumoral* structures: necrotic and non-enhancing tumor core (label = 1), peritumoral edema (label = 2), and enhancing tumor (label = 4). From the intra-tumoral structures, the following *glioma sub-regions* [27] were defined: whole tumor (WT) which encompasses all three intra-tumoral structures (i.e., label = $1 \cup 2 \cup 4$), tumor core (TC) that contains all but the peritumoral edema (i.e., label = $1 \cup 4$), and enhancing tumor (ET) (label = 4), where \cup represents union operation.

The BRATS dataset is provided in a preprocessed format, i.e., all the images are skull-stripped, resampled to an isotropic 1mm^3 resolution, and all four modalities of each patient are co-registered. We applied additional pre-processing that included (in order): 1) computing the bounding box of the brain in each image, and extracting the selected portion of the image, effectively zooming in on the brain and discounting excess background pixels, 2) re-sizing the cropped image to 128×128 pixels, 3) discarding images which contained no tumor regions in the ground truth segmentation, 4) applying an intensity windowing function to each image such that the lowest 1% and highest 99% of pixel values were mapped to 0 and 255, respectively, and 5) applying z-score normalization to each image i.e., subtracting the mean and dividing by the standard deviation of the dataset.

The input to DIU-Net is an $N \times M \times D$ pixel image, where $N = M = 128$ pixels and $D = 4$ which represents four MRI modalities. The output of the model is an $N \times M \times K$ tensor, where $K = 3$ and represents total number of segmentation classes, i.e., three intra-tumoral structures. Each slice of K is a binary image and represents the predicted segmentation for the i^{th} class where $0 \leq i \leq K - 1$.

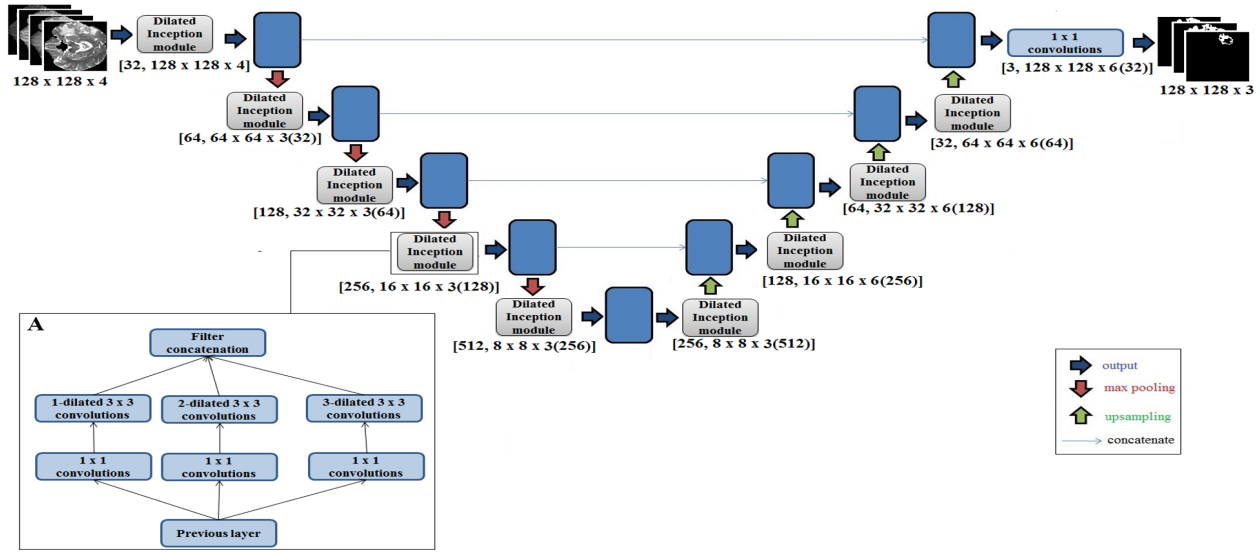


Figure 1. DIU-Net architecture with contracting and expanding path and a bottleneck in the middle is presented. The set of numbers shown below each Inception module indicates the total number of filters used, height, width, and depth of the input feature map. On the contracting path, the multiplication by 3 indicates three l -dilated convolutional filters. On expanding path, the concatenation of the feature maps from the contracting path doubles the depth of the output feature map, hence the multiplication by 6. (A) Dilated Inception module with three l -dilated convolutional filters and 1×1 dimensional reduction convolution filters is presented.

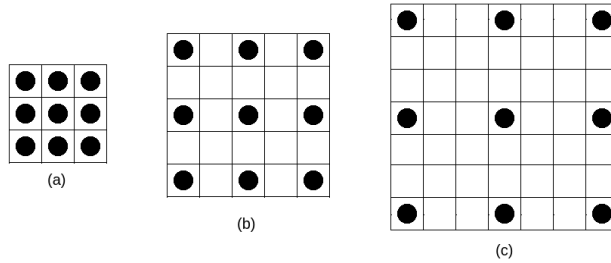


Figure 2. Three cases of a 3×3 dilated filter with $l = 1, 2$, and 3 , are presented in sub-figures (a), (b), and (c) respectively. We note that the number of filter elements, indicated by the black dots, stays constant while the receptive field increases proportionally to l .

II-D. Evaluation Metric and the Loss Function

Dice Similarity Coefficient or simply the Dice score is extensively used for the evaluation of segmentation algorithms in medical imaging applications [28]. The Dice scores between a predicted binary image P and a ground truth binary image G , both of size $N \times M$ is given by:

$$\text{Dice}(P, G) = \frac{2 \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P_{ij} G_{ij}}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P_{ij} + \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} G_{ij}}, \quad (7)$$

where i and j represent pixel indices for the height N

and width M . The value of Dice score ranges between 0 and 1 and a higher score corresponds to a better match between the predicted image P and the ground truth image G .

The loss function for DIU-Net is given by [3]:

$$\mathcal{L}_{\text{Dice}}(P, G) = -\log \left[\frac{1}{K} \sum_{i=0}^{K-1} \text{Dice}(P_i, G_i) \right]. \quad (8)$$

II-E. Training, Testing, and Evaluation of DIU-Net

We compared the performance of the proposed DIU-Net with Inception U-Net that did not incorporate dilated modules [3]. We trained both models under same conditions to ensure a fair comparison. Both models were trained using k -fold cross-validation scheme with $k = 10$. The dataset was randomly split into 10 mutually exclusive subsets of equal or near equal size. Each algorithm was run 10 times subsequently, each time taking one of the ten splits as the validation set and the rest as the training set. In our experiments, each model was trained 10 times using a different set of 90% of the data and validated on the remaining 10% data. This resulted in a total of 20 models, i.e., 10 models for U-Net with Inception modules, 10 models for DIU-Net. The Dice scores presented in the Results section are median values of the ten trained models.

We used stochastic gradient descent with an adaptive

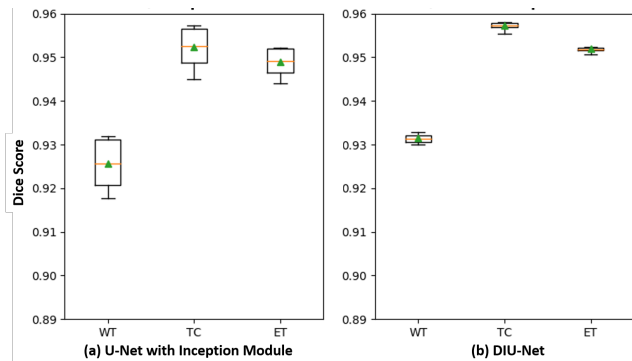


Figure 3. Dice scores are presented for both U-Net with Inception module and DIU-Net using box plot. On the x-axis, we present glioma sub-regions including whole tumor (WT), tumor core (TC), and enhancing tumor (ET). The median values are denoted by the horizontal orange line and the mean values are denoted by the green triangle. The increased dice score is statistically significant for WT and TC ($p < 0.05$). We also note a significant reduction in the variability for the DIU-Net.

moment estimator (Adam) for training all models [29]. The initial learning rate was set to 10^{-4} which was exponentially decayed every 10 epochs. The batch size was set to 64 and each model was trained for 100 epochs. All learnable parameters, i.e., weights and biases of the models were initialized based on He initialization method. We used Keras application programming interface (API) with TensorFlow backend for the implementation of all models. All models were trained on a Google Cloud Compute instance with 4 NVIDIA TESLA P100 graphical processing units (GPUs).

After training, each model was tested on the entire BRATS 2018 dataset. For each image, the intra-tumoral structures were combined to produce glioma sub-regions, and Dice scores were computed. The process was repeated for each image, and after evaluating all images, a median Dice score was calculated for each glioma sub-region. Overall, this process generates 2 sets of 10 Dice scores for each glioma sub-region. Each set was then evaluated for normality using the Shapiro-Wilk test, with the probability of Type-I error set to $\alpha = 0.05$. Based on the results of the Shapiro-Wilk test, we found that the set of Dice scores were not normally distributed. Therefore, we used non-parametric test, i.e., two-sided Wilcoxon signed rank test to compare Dice scores of two models.

III. RESULTS

We present cross-validation Dice scores for all three glioma sub-regions using the box plot for both models in Fig. 3. We note that DIU-Net showed significant improvement in the whole tumor sub-region, i.e., Dice score increased from 0.925 to 0.931 with $p < 0.05$. Similarly, for the tumor core sub-region, the Dice score

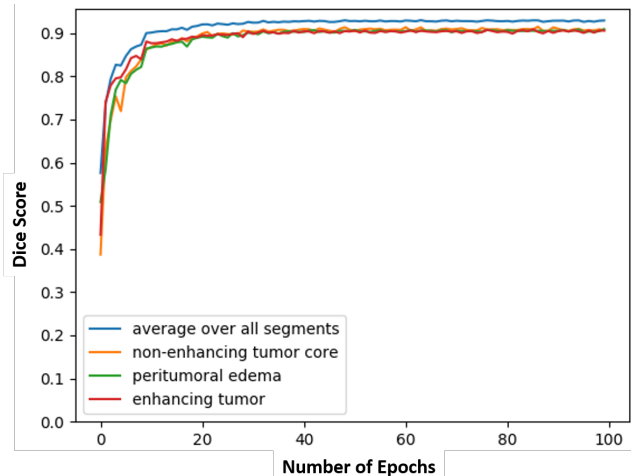


Figure 4. Dice scores are presented for an increasing number of epochs separately for each intra-tumoral structure during validation.

improved from 0.952 to 0.957 with $p < 0.05$. However, for the enhancing tumor, the change was not statistically significant, $p = 0.114$.

The validation Dice score curves plotted against the number of epochs for all intra-tumoral structures for the DIU-Net are presented in Fig. 4. The improvement in the segmentation over the number of epochs is evident. The segmentation results from one representative high-grade glioma and one low-grade glioma case are presented in Fig. 5. We note that the predicted segments (shown in the red block) of the glioma sub-regions are visually similar to the ground truth segments (shown in the black block).

IV. DISCUSSION AND CONCLUSIONS

We aimed to tackle the challenging problem of pixel-level segmentation in brain MRIs for tumor delineation, which, in turn, is essential for tumor diagnosis, identification, and surveillance. We introduced dilated convolutions in Inception modules and incorporated these modules into the U-Net architecture (DIU-Net). We extended our previously proposed framework and significantly improved its accuracy (measured using the Dice score) [3]. We used k -fold cross-validation and found that DIU-Net significantly improved ($p < 0.05$) the tumor segmentation performance in two of the three glioma sub-regions, i.e., whole tumor and tumor core.

We hypothesize that there is more contextual information for the whole tumor and tumor core, which DIU-Net was able to capture in the learning process. The results of enhancing tumor suggest that larger contextual information does not benefit model performance for this sub-region. This may be potentially linked to the small number of pixels in this sub-region relative to the

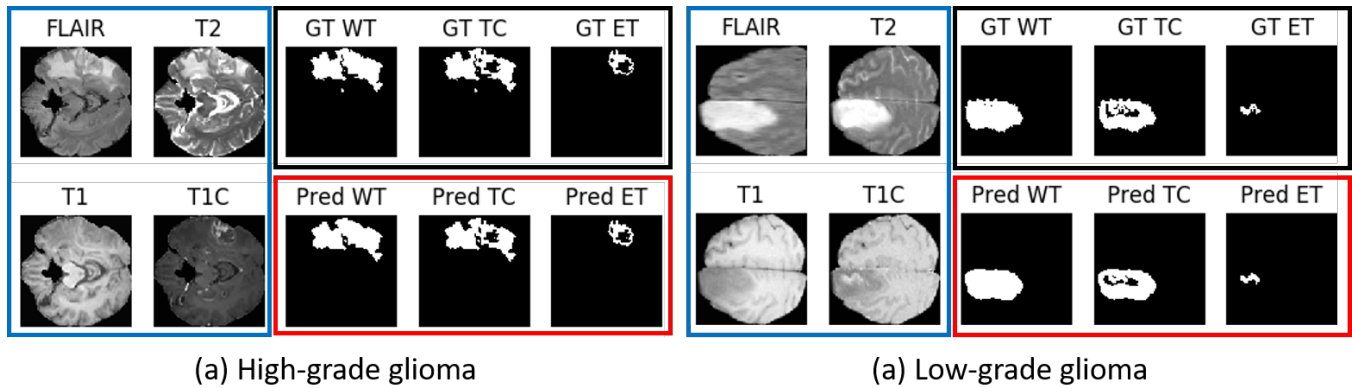


Figure 5. The segmentation results for a representative high-grade, and a low-grade glioma patient are presented. Images in blue blocks are the four MRI modalities. Images in the black blocks (top row) are the ground truth segments and are denoted by “GT” for all three glioma sub-regions (whole tumor - WT, tumor core - TC, and enhancing tumor - ET). Images in the red blocks (bottom row) are segmentation results and are denoted by “Pred”.

other glioma sub-regions. It is essential to mention that DIU-Net is computationally more efficient, i.e., DIU-Net has 2.5 million fewer parameters than the U-Net with Inception modules. DIU-Net achieves significantly better results at a lesser computational cost (15% fewer parameters). The dice scores for each glioma sub-region are comparable or exceed the results of other recently published architectures, including No New-Net, which achieved second place in the BRATS 2018 competition [30], SDResU-Net [13], and the ensemble approach proposed in [31].

REFERENCES

- [1] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, p. 60–88, 2017.
- [2] H. M. Fathallah-Shaykh, A. DeAtkine, E. Coffee, E. Khayat, A. K. Bag, X. Han, P. P. Warren, M. Bredel, J. Fiveash, J. Markert, N. Bouaynaya, and L. B. Nabors, “Diagnosing growth in low-grade gliomas with and without longitudinal volume measurements: A retrospective observational study,” *PLOS Medicine*, vol. 16, no. 5, pp. 1–16, 05 2019. (available at: <https://doi.org/10.1371/journal.pmed.1002810>).
- [3] D. E. Cahall, G. Rasool, N. C. Bouaynaya, and H. M. Fathallah-Shaykh, “Inception Modules Enhance Brain Tumor Segmentation,” *Frontiers in computational neuroscience*, vol. 13, p. 44, 2019.
- [4] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, “Auto-Context Convolutional Neural Network (Auto-Net) for Brain Extraction in Magnetic Resonance Imaging,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, p. 2319–2330, 2017.
- [5] M. Havaei *et al.*, “Brain tumor segmentation with deep neural networks,” *Medical image analysis*, vol. 35, pp. 18–31, 2017.
- [6] K. Kamnitsas *et al.*, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, p. 61–78, 2017.
- [7] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, p. 234–241, 2015, (Munich, Germany).
- [8] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, “The Importance of Skip Connections in Biomedical Image Segmentation,” *Deep Learning and Data Labeling for Medical Applications Lecture Notes in Computer Science*, p. 179–187, 2016, (Athens, Greece).
- [9] M. Livne *et al.*, “A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease,” *Frontiers in Neuroscience*, vol. 13, 2019.
- [10] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks,” *Communications in Computer and Information Science Medical Image Understanding and Analysis*, p. 506–517, 2017, (Edinburgh, United Kingdom).
- [11] F. Girard, C. Kavalec, and F. Chriet, “Joint segmentation and classification of retinal arteries/veins from fundus images,” *Artificial intelligence in medicine*, vol. 94, pp. 96–109, 2019.
- [12] M. Dash, N. D. Londhe, S. Ghosh, A. Semwal, and R. S. Sonawane, “PsLSNet: Automated psoriasis skin lesion segmentation using modified U-Net-based fully convolutional network,” *Biomedical Signal Processing and Control*, vol. 52, pp. 226–237, 2019.
- [13] J. Zhang, X. Lv, Q. Sun, Q. Zhang, X. Wei, and B. Liu, “Sdresunet: Separable and dilated residual u-net for mri brain tumor segmentation,” *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 15, 08 2019.
- [14] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, “RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images,” *IEEE Access*, vol. 7, p. 21420–21428, 2019.
- [15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, (Boston, MA, USA).
- [16] H. Li, A. Li, and M. Wang, “A novel end-to-end brain tumor segmentation method using improved fully convolutional networks,” *Computers in Biology and Medicine*, 2019.
- [17] C. Wang, M. Rajchl, A. Chan, and E. Ukwatta, “An ensemble of U-Net architecture variants for left atrial segmentation,” *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950. International Society for Optics and Photonics, 2019, p. 109500M, (San Diego, CA, USA).
- [18] L. Song, K. Geoffrey, and H. Kaijian, “Bottleneck feature supervised u-net for pixel-wise liver and tumor segmentation,” *Expert Systems with Applications*, vol. 145, p. 113131, 2020.
- [19] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2015.
- [20] W. Shi, F. Jiang, and D. Zhao, “Single image super-resolution with dilated convolution based multi-scale information learning inception module,” *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 977–981.
- [21] S. Yang and W. Lin, “Predicting visual saliency via a dilated

- inception module-based model,” *International Workshop on Advanced Image Technology (IWAIT) 2019*, vol. 11049. International Society for Optics and Photonics, 2019, p. 110491D.
- [22] M. Wang, K. Tan, X. Jia, X. Wang, and Y. Chen, “A deep siamese network with hybrid convolutional feature extraction module for change detection based on multi-sensor remote sensing images,” *Remote Sensing*, vol. 12, no. 2, p. 205, 2020.
- [23] M. Zhai, X. Xiang, R. Zhang, N. Lv, and A. El Saddik, “Learning Optical Flow Using Deep Dilated Residual Networks,” *IEEE Access*, vol. 7, pp. 22 566–22 578, 2019.
- [24] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, “Fully Convolutional Networks for Monocular Retinal Depth Estimation and Optic Disc-Cup Segmentation,” *IEEE Journal of Biomedical and Health Informatics*, 2019.
- [25] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv [Preprint]*, 2015.
- [26] S. Bakas *et al.*, “Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge,” *arXiv [Preprint]*, 2018.
- [27] B. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2014. (available at: <https://hal.inria.fr/hal-00935640>).
- [28] S. Bakas *et al.*, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Scientific Data*, vol. 4, p. 170117, 2017.
- [29] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv [Preprint]*, 2014.
- [30] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, “No New-Net,” *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 234–244, (Granada, Spain).
- [31] P.-Y. Kao, T. Ngo, A. Zhang, J. W. Chen, and B. Manjunath, “Brain Tumor Segmentation and Tractographic Feature Extraction from Structural MR Images for Overall Survival Prediction,” *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 128–141, (Granada, Spain).