Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 732–740

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Multi-Modal Semantic Segmentation Model using Encoder Based Link-Net Architecture for BraTS 2020 Challenge

Gayathri Ramasamy[a], Tripty Singh[b], Xiaohui Yuan[c]

[a]*Department of Computer Sciecnce and Engineering, Amrita School of Engineering Bengaluru, Amrita Vishwa VidyaPeetham, India*
[b]*Department of Computer Sciecnce and Engineering, Amrita School of Engineering Bengaluru, Amrita Vishwa VidyaPeetham, India*
[c]*Department of Computer Science and Engineering, University of North Texas, Denton, Texas*

Abstract

Glioma, which is a malignant tumor, is present in the glial tissue region of the human brain. Segmentation of such tumor cells in the brain region is still challenging and needs experts. Because of the overlap between the intensity distributions of tissue with edema, non-edema, and enhancing features, the segmentation process is a significant challenge for neurosurgeons and radiologists. As per the current state of the art in medicine and surgery, artificial intelligence is gaining attention in effective detection and segmentation in the area of medical diagnosis. In MICCAI 2020, the authors prepare an algorithm for the semantic segmentation of brain tumors from multimodal MRI images for further treatments such as observing treatment, monitoring recovery, and evaluating the effects of the treatment on patients. This paper's objective is to develop an efficient deep learning model which performs semantic segmentation using a multi-modal modified Link-Net model which uses Squeeze and Excitation ResNet152 model is used as a backbone for the segmentation. A model developed by Manipal Hospital in Bangalore is compared with the traditional state-of-the-art models, and its accuracy is verified by neurosurgeons there. This model imbibes the multi-modal MRI dataset which includes T1 weighted images, Flair images, and T2-weighted MRI images of the human brain, model perform comparably well, which shows that our model is robust for tumor segmentation. The accuracy of this model is 99.2.

Keywords: BraTS 2020 Challenge; Link-Net Architecture; Multi-Modal Semantic Tumor Segmentation; Deep Learning

* Corresponding author.
E-mail address: iamrgayathri@gmail.com

1877-0509
10.1016/j.procs.2023.01.053

1. Introduction

It has been demonstrated that, when combined with Positron Emission Tomography (PET), computed tomography CT, and SPECT, MRI will be owned to study tumor regions in the human brain, but segmentation of the tumor remains challenging. Fusion of MRI and CT images was also adopted to extract detailed information about tumor regions as in [1], [2]. Contrast Enhanced MRI is the recently preferred imaging modality for the diagnosis of brain diseases because of the sensitivity and better contrast in soft tissue regions [3]. Glioma is the type of tumor that occurs in glial cells of the human brain. Glioma may be catalogued into Low Grade Glioma LGG and High Grade Glioma HGG, which are, dependent on the rate of the tumor development [4]. The rate of growth of the tumor is faster in the case of High Grade Glioma HCG when comparable with Low Grade Glioma LCG. A person having HGG will have a short life span comparable to LGG [5]. Automatic and Semiautomatic methods in segmentation of tumor portion from the imaging modality are therefore necessary since manual segmentation requires more time and effort and also that is subjective. The major challenge in the segmentation of brain tumors is that the tumor may occur in any part of the brain and of any size [6]. Recently FCN (Fully Convolutional Neural Networks) is found the most efficient for segmentation of biomedical images.

The demand for fully automatic brain tumour segmentation is increased for the diagnosis of Gliomas, which includes the usage of multi-modal MRI images [7]-[9]. Researchers found it difficult for the heterogeneous nature of the tumour such as appearance, size, shape, and location of tumours. In addition, obtaining the dataset is found to be difficult and limited in quantity. Biomedical images require expert reviews and it is time-consuming and labourintensive. Only a few publicly available datasets are available for bio-medical data until now.

Researchers are paying attention to Deep Learning models [10, 11] in medical image analysis in the past few years. Various Convolution Neural Network does bio-medical image segmentation operations. Deep CNNs comprehend various features of the image automatically which deals with complicated biomedical images. Deep learning is becoming increasingly popular for processing natural images in addition to biological images [12]-[16]. The evolution of biomedical imaging has been aided by the implementation of CNNs since then [17]-[19]. Using this approach, [20] has designed deep learning architectures tailored for different levels of glioma.

T-1, T-1c, T2c, and fluid-attenuated inversion recovery (FLAIR) are a few types of imaging techniques that reveal different factors about the tumour characteristics, which are shown in Fig 1. [21] – [23]. Researchers frequently ignore the significance of complex relationships among different image modalities, in spite of the high performance achieved by previous methods. For example, some modalities might be insensitive to changes in another, even though previous methods have achieved remarkable results.

To address the segmentation task, a novel deeper encoder cum decoder kind of network was developed, which adopts the co-related info of various multi-modal brain MRI images. This segmentation model is based on the MultiModal Link-Net segmentation algorithm, which comprises three main components: a multimodal encoding engine, a multimodal fusion engine, and the decoder that is shared. After that data acquired from each modality is sliced, and the depth values are compared. Following that layer, data are passed through four modal independent encoders. After that, the features are extracted cross-modally. A shared decoder is then used to up sample the shared multi-modality features. This study makes the following main suggestions.

- It is possible to exploit relationships between different modes when segmenting brain tumors in 3D MRI with an encoder cum decoder network that combines multi-modality encoders. In that each modality presents brain tumor separately, the multimodal encoder assists in disentangling info that elsewhere may have done fusion earlier, reducing the capacity of the network to capture relationships across modalities.
- The correlation between multiple modalities is modeled using a Link-Net format based on encoders and decoders. Combining encoders for different modalities, Link-Net integrates features in both a top-down and bottom-up manner. By exploiting the complex relationship between various modalities, Link-Net Model is able to improve segmentation accuracy.

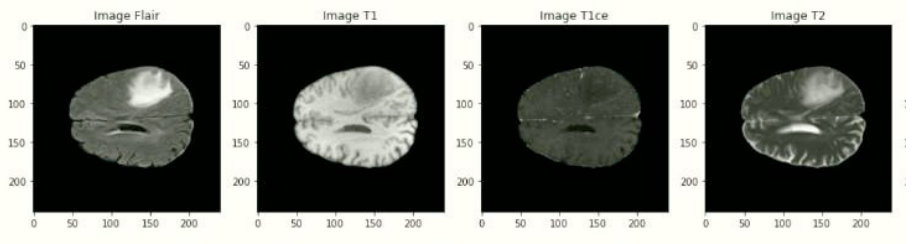

Fig. 1. MRI Image of various Modalities from BraTS2020 Dataset

2. Literature Survey

The segmentation of biological images, especially brain tumours, can be accomplished with deep learning algorithms or deep convolutional neural networks (DCNNs). In medical image segmentation, the vast majority of models work with 3D images [24]-[26]. Nevertheless, 3D networks require a lot of parameters, as well as computer processing power. A few models extract patches of 3D data rather than the entire volume of 3D data, in an effort to reduce computational costs during training. Patch-based methods [20], [27] involve creating a network of patches from an image; this re-computing is overlapping, resulting in slow training. The researchers neglect to think about the structure of information, which is global, such as the content of images along with co-relations among labels, instead of focusing on small sections of sliding windows. Additionally, segmenting a 3D volume in 2D can be achieved by splitting it up into multiple 2D slices [13], [28]. Adapting depth-slicing to fit this dataset would reduce this problem.

U-Net [29] and A-UNet [29] use encoder-decoder-like networks as segmentation tools. Both tools have been used not only for biomedical image segmentation. In order to perform the task of segmentation successfully, both semantic and spatial information is required, which is why U-Net [29] uses a decoder that gets its information via skip connections from the lower layers of the encoder. In recent years, researchers have focused on preserving important information during feature extraction from input images. A method was developed in [30] in which encoding and decoding phases were used simultaneously, leading to greater brain segmentation performance. The authors are the winners of the BraTS2018 Challenge with this method. A variation auto encoder (VAE) branch was recommended to regularize the model during training. In each stage of the hierarchy, these networks share the same parameters. Using this sharing technique, these networks become modules that can be used with any encoder-decoder unit to reduce the size of the unit.

The recent studies that have been conducted have typically used early fusion strategies, Assuming the relationship between the methodologies is linear, multiple methodologies are combined as one input. In many situations, different modalities are stacked with several different channels, which does not consider the co-relations among various modalities [31], [32]. The late fusion strategy thrives on the complex relationship between these modalities, unlike the early fusion strategy [33]. Recently, studies have effectively shown that a late fusion strategy outperforms early fusion with regard to medical segmentation problems because of the many representations and correlation structures present in each modality. Cross-modal convolution is proposed to combine information from multiple modalities using a technique known as late fusion [34][35]. This technique exploits the correlations between different modalities. The study demonstrate the limitations of modeling modalities with a single layer by incorporating each modality separately into this model and assessing the complex relationships among them by using the late fusion strategy.

The link-net algorithm is based on encoders and decoders, which are used to create semantic segmentation based on the responses between these complex multimodal correlations [36] The correlations among the modalities have yet to be fully investigated, despite the productivity of the late strategy method. In [37], [38] While multimodality is often viewed as separate channel; there is not much acknowledgment of their correlations. Nonetheless, [38] developed a convolution, which is cross-modal to reveal the mechanisms

underlying the response of various modalities. And hence the layers are down-sampled, and some information will be lost during the feature extraction process. Authors used link-net in a biomedical image segmentation process to optimize multi-scale feature fusion using a principled approach in their model. This model, therefore, incorporates a bi-directional Link-Net unit, as it was inspired by [35] and [39] to constitute multi modal fusion.

3. Proposed Methodology

MM-Link-Net is an encoder-decoder network that Encoder-decoder models composed of fully convolutional representations are completely convolutional. You can visualize the left hand portion in the image in which the model has 4 encoders each having the same mode of input (FLAIR, T1, T1c, and TM2). The multi-modal MRI features are fused together and are given to an encoder -decoder based LinkNet Architecture and pre-trained ResNET152 architecture is being used as backbone discriminator architecture for segmentation. The Proposed architecture for the segmentation model is shown in Fig 2.

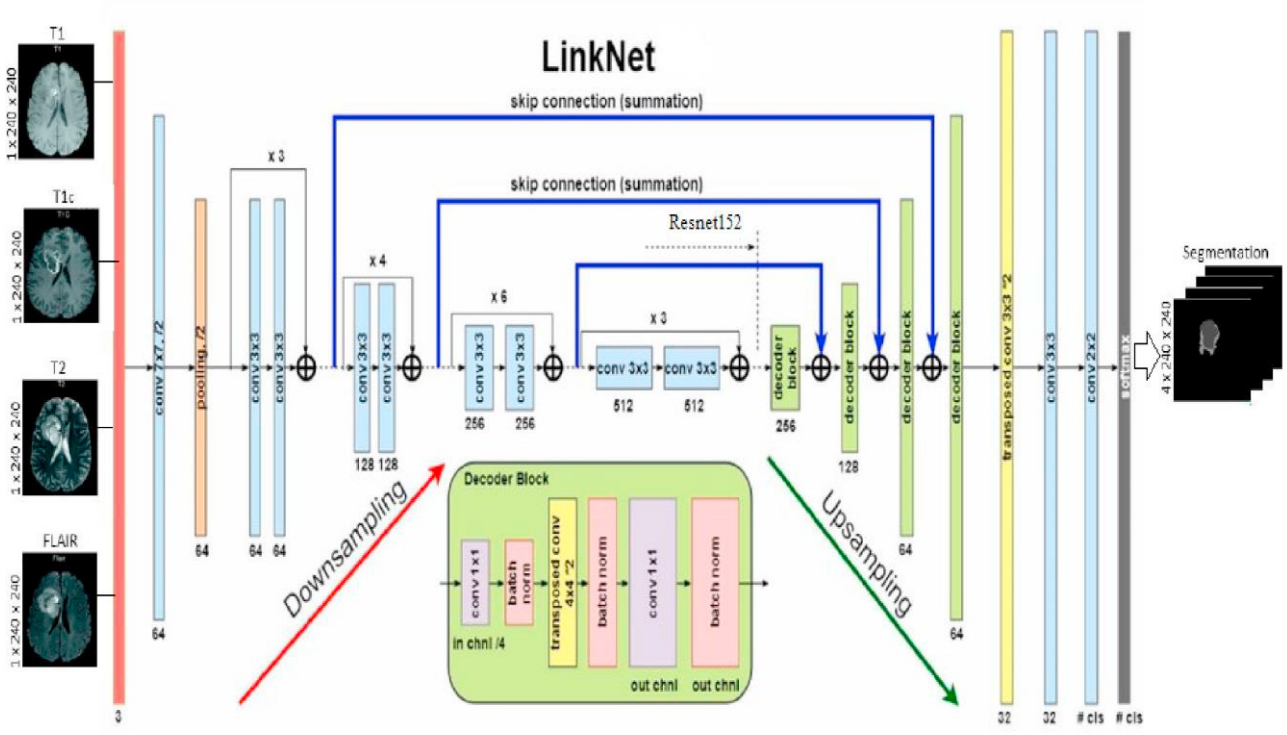

Fig. 2. Link-Net Architecture

3.1. Patch Generation

Since all the MRI images are three dimensional, segmentation of three-dimensional images may lead the training to be slow with the huge size of the dataset. To resolve this and to reduce the cost of computation, the architecture is generating the two dimensional volume slices or patches of MRI, which can have the size of 128×128 instead of having the original size240×240×155. After extracting the two-dimensional patches for the input images from different modalities, the MRI patches are given as input for the encoder region of the Link-Net architecture. .

3.2. Multi-Modal Fusion Encoder

Our MM-LinkNet architecture utilizes ResNet152 [13], which has two paths, a contracting path, and an expanding path. Contraction, also called an encoder, uses high-level feature representation for the input image to collapse it into compact representations, while expansion, or decoding, uses these feature values to build the mask for segmentation. Compared to a typical auto decoder, the Link Net encoder-decoder architecture [13] also includes skip connections to maintain missing data that can occur during compression. For Limited Dataset, ResNet152, as our baseline to minimize any risk of information loss is chosen, during the segmentation prediction. Link Net was constructed for assisting segmentation of biomedical datasets, and since our dataset is limited, we chose ResNet to assist with segmentation.

An important challenge for the process of multi-modal images was exploiting the compound interrelationship across the modality dimensions. As a result of the multi-modal encoding, each modal is encoded in the ResNet152 layer (Link-Net) to combine responses from all modality encoders. (inspire from [39]). Nodes on Link-Nets are not restricted to a single origin node, and can therefore combine information from a variety of sources, allowing for the fusion of low-level features with high-level features. Link-Net architecture blends feature both from the top-down direction and in the bottom-up directions to let the algorithm determine cross-scale features efficiently. The proposed architecture concatenate each convolutional block of every single modality to form the input for our Bi-Link-Net modeling model

3.3. Shared Encoder

In contrast with existing neural network architectures for segmentation, the study novel in that each encoder is linked with a decoder. Multiplying down sampling operations results in the loss of some spatial information. The encoder's output can be down sampled but this will not give you back all of the lost information. [41] Using indices that are not trainable, the encoder and decoder are linked together. To segment their audio, some methods use their encoder output directly as input to a decoder. For every encoder layer's input, we also bypass the output of the decoder layer. The decoding function and its up sampling functions will be able to make use of the recovered spatial information. Decoding may also require fewer parameters because the encoder shares knowledge learned at every layer. This results in a network that is more efficient and uses real-time operations, compared to the existing state-ofthe-art segmentation networks.

4. Details of Implementation

4.1. Dataset

Data is obtained from Brain Tumour Segmentation Challenge for 2020[43], which constitutes 369 MRI sets of Training phase and 125 MRI sets of validation phase. 5 times of cross validation is performed in the training phase for BraTS2020 and for validation phase respectively. Dataset comprises of scan in four different formats i.e .nii format including native T1-weighted MRI, ContrastEnhanced-T1 weighted MRI images, T2-weighted MRI, FluidAttenuation inversion recovery-FLAIR. For these MICCAI BraTS initiative, the layers are interpreting MRI brain images for the classification of Low Grade Glioma LGG and High Grade Glioma HCG. Each dataset contained 293 sets of High Grade Glioma with 79 sets of High Grade Glioma that were manually marked by expert neurologists and identified types of tumour labels I. Necrotic non-enhancing tumour II. Peritumoral edema III. Densely enhanced tumour ET, and label O for remaining parts. The area of intersection or union is obtained by calculating the ratio of intersections to unions, while the area of pixels incorrectly segmented as labels is determined by FPR.

4.2. Experimental Section

In our experiment, GeForce GTX TITAN Black of 6GB RAM is used . Computing system also used Python 3.6 and PyTorch 1.9.1. MM-LinkNet model were trained using Adam optimizer and used cross-category loss to train the network. This model is trained for 80 epoch for learning rate of 0.001.MM-LinkNet with backbone network ResNet152 is compared with U-Net , FCN (Fully Convolutional Network) and FPN (Feature Pyramid Network) and
SegNet with loss function - categorical cross entropy, Adam optimizer and learning rate of 0.001

4.3. Network Architecture

In Figure 2, the RESNET152 [13] structure is modified, as it is based on the RESNET152 structure. There are three main components of the network: 1) Each of the four encoders encodes information independently based on modality;2) the Link-Net layer, Correlation between multiple modalities is derived from this method, 3) Segmentation results are produced by means of one of two paths decoders: an expanding path decoder and a reclaiming path decoder.

4.4. Training Phase

A convolutional network needs to be trained to identify all non-tumor tissues in a BraTS dataset, which is very challenging. This model predicts non-tumor tissue in every pixel as a result, quickly reaching its minimum value. The model, therefore utilize median frequency balancing [47] to deal with the data imbalance problem, as proposed in [46]-[48], The weight is assigned for cross-entropy loss function to each class as follows:

$$W_c = \frac{MedianFrequnecy\ Frequency(c)}{} \quad (1)$$

Median-Frequency and Frequency (c) look at the class frequencies of all medians, respectively, along with the total count of pixels divides the total no. of pixels in classes in all image, having the entire count of pixels into account, for all image. The medial frequency balancing thus applied in our experiment should help maintain a balanced training process whereby the background of the image and tissues that are normal are assigns the smallest weight.

5. Experimental Results

5.1. Measurable Result

A performance evaluation for the network proposed was conducted by computing Dice coefficient, sensitivity, positive predictive value, and similarity index of the three regions. The values of Dice Similarity index, Sensitivity, Positive predicted value and Jaccard Index are evaluated for the model and compared with other state of art models which is shown in the Table 1.

Network	Dice Coefficient	Sensitivity	PPV	Jaccard Index
U-Net	0.7535	0.7400	0.8125	0.6312
FCN	0.7343	0.6616	0.8666	0.6071
FPN	0.8107	0.7108	0.8880	0.6972
SegNet	0.7320	0.716	0.827	0.6920
MM- LinkNet	0.7773	0.7662	0.8965	0.7169

- Dice Similarity Coefficient: In other words, Dice similarity coefficient is the measure of how closely segmented results and ground truth overlap.
- Sensitivity According to segmented results, sensitivity is a measure of the proportion of positive voxels in the ground truth
- Positive Predicted Value: Positive predicted value i.e PPV is the measure of the precision of the results.
- Jaccard Index: By dividing the ground truth result by the predicted result, this value measures the intersection between the results.

5.2. Subjective Result

In Fig 3, the visual segmentation of the trained model is illustrated using four input modalities for encoders on the BraTS2020 data set. Using the FLAIR modality, the methodology selected four BraTS2020 cases at random and presented the segmented results separated from their normal background. Red areas indicate the core tumor. Because of this visual comparison, the enhanced tumor segmentation structure is done by MM-LinkNet compared with other models.

6. Conclusion

This study presents, to make multi modal brain tumor images more useful, a bi-directional feature fusion method using multiple encoders and a shared decoder. In the testing of the proposed methods against the BraTS2020 datasets,

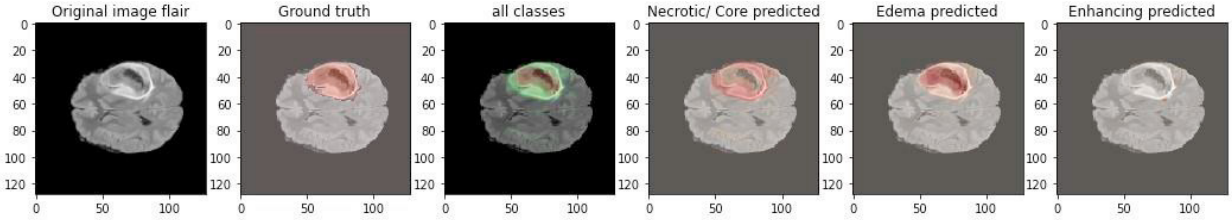

Fig. 3. Performance Results of Segmentation

it is found that they effectively integrated the tumor information in multimodality images and provided an improved segmentation. The RESNET152, which is used as a backbone of the LinkNet Architecture makes the architecture deep enough to segment the tumour portion with more accuracy. Experimental results on BraTS2020 validation dataset shows that our model outperformed in sematic segmentation and is relatively competitive with our compared models. Various other backbone Pre-trained models such as inceptionv3, vgg16 and resnet152 are also used for validation of the LinkNet Model in sematic segmentation.

7. Future Scope

Although, the proposed model performs well in segmentation, it is quite computationally expensive and time consuming because of the depth of layers in architecture and separation of encoders being used. Thus, It is important to extract information related to pixel-relationship and used to train the model to obtain high-accurate results.

References

[1] S. R. Muzammil, S. Maqsood, S. Haider, and R. Dama²evi£ius, "CSID: A novel multimodal image fusion algorithm for enhanced clinical diagnosis," Diagnostics, vol. 10, no. 11, p. 904, Nov. 2020.

[2] R. Zhu, X. Li, X. Zhang, and M. Ma, "MRI and CT medical image fusion based on synchronized-anisotropic diffusion model," IEEE Access, vol. 8, pp. 91336 91350, 2020.

[3] S. Cui, L. Mao, J. Jiang, C. Liu, and S. Xiong, "Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network," J. Healthcare Eng., vol. 2018, pp. 1 14, Mar. 2018.

[4] U. Baid et al., "Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model: A Novel Approach Using 3D UNET Based Deep Convolutional Neural Network for Predicting Survival in Gliomas," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, A. Crimi et al., Eds. Cham: Springer International Publishing, 2019, pp. 369–379.

[5] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, oct 2015.

[6] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, oct 2015.

[7] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3D dilated multi-fiber network for real-time brain tumor segmentation in MRI," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham Springer, 2019, pp. 184 192.

[8] K. Usman and K. Rajpoot, "Brain tumor classification from multimodality MRI using wavelets and machine learning," Pattern Anal. Appl.,vol. 20, no. 3, pp. 871 881, 2017.

[9] T. Zhou, S. Ruan, Y. Guo, and S. Canu, "A multi-modality fusion network based on attention mechanism for brain tumor segmentation," in Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI), Apr. 2020, pp. 377 380.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436 444, May 2015.

[11] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep Learning, vol. 1, no. 2. Cambridge, MA, USA: MIT Press, 2016.

[12] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in Proc. Adv. Neural Inf. Process. Syst., vol. 25, 2012, pp. 2843 ‒2851.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Pro-cess. Syst. (NIPS), vol. 25, Dec. 2012, pp. 1097 1105.

[14] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE Conf.´ Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117 2125.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2015,pp. 234 ‒241.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),Jun. 2016, pp. 770 778.

[17] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi,D. Rueckert, and B. Glocker, "Deepmedic for brain tumor segmentation,"in Proc. Int. Workshop Brainlesion, Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries. Cham, Switzerland: Springer, 2016,pp. 138 149.

[18] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi,M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sanchez,"A´ survey on deep learning in medical image analysis," Med. Image Anal.,vol. 42, pp. 60 ‒88, Dec. 2017.

[19] J.Ker, L.Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," IEEE Access, vol. 6, pp. 9375 9389, 2017.

[20] D. Polap, "Analysis of skin marks through the use of intelligent things," IEEE Access, vol. 7, pp. 149355 ‒149363, 2019.

[21] D. Po^{a}ap, "an adaptive genetic algorithm as a supporting mechanism for microscopy image analysis in a cascade of convolution neural networks," Appl. Soft Comput., vol. 97, Dec. 2020, Art. no. 106824.

[22] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1240 1251, May 2016.

[23] P. Y. Wen, D. R. Macdonald, D. A. Reardon, T. F. Cloughesy, A. G. Sorensen, E. Galanis, J. DeGroot, W. Wick, M. R. Gilbert, A. B. Lassman, C. Tsien, T. Mikkelsen, E. T. Wong, M. C. Chamberlain, R. Stupp, K. R. Lamborn, M. A. Vogelbaum, M. J. van den Bent, and S. M. Chang, "Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group," J. Clin. Oncol., vol. 28, no. 11, pp. 1963 1972, Apr. 2010.

[24] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," Phys. Med. Biol., vol. 58, no. 13, pp. R97 R129, Jul. 2013.

[25] H. Chen, Z. Qin, Y. Ding, L. Tian, and Z. Qin, "Brain tumor segmentation with deep convolutional symmetric neural network," Neurocomputing, vol. 392, pp. 305 313, Jun. 2020.

[26]] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in Proc. Int. MICCAI Brainlesion Workshop. Cham, Switzerland: Springer, 2017, pp. 287 297.

[27] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. MaierHein, “No new-net,” in Proc. Int. MICCAI Brainlesion Workshop. Cham, Switzerland: Springer, 2018, pp. 23 244.
[28] Y. Qin, K. Kamnitsas, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi, and A. Nori, “Autofocus layer for semantic segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2018, pp. 603 611.
[29] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks,” Med. Image Anal., vol. 35, pp. 18 31, Jan. 2017.
[30] H. Chen, X. Qi, L. Yu, and P.-A. Heng, “DCAN: Deep contour-aware networks for accurate gland segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2487 2496.
[31] A. Myronenko, “3D MRI brain tumor segmentation using autoencoder regularization,” in Proc. Int. MICCAI Brainlesion Workshop. Cham, Switzerland: Springer, 2018, pp. 311 320.
[32] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, “Joint sequence learning and cross-modality convolution for 3D biomedical segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 6393 6400.
[33] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 10781 10790.
[34] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431 3440.
[35] Chen Wang, Danfei Xu, Yuke Zhu, Roberto Mart´ın-Mart´ın, Cewu Lu, Li Fei-Fei, and Silvio Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3343–3352, 2019.
[36] Shuxin Wang, Shilei Cao, Dong Wei, Renzhen Wang, Kai Ma, Liansheng Wang, Deyu Meng, and Yefeng Zheng. Ltnet: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9162–9171, 2020.
[37] Weiyao Wang, Du Tran, and Matt Feiszli. What makes training multi-modal classification networks hard? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12695–12705, 2020.
[38] Xin Wang, Qiuyuan Huang, Asli Celikyilmaz, Jianfeng Gao, Dinghan Shen, Yuan-Fang Wang, William Yang Wang, and Lei Zhang. Reinforced cross-modal matching and selfsupervised imitation learning for vision-language navigation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6629–6638, 2019
[39] Yikai Wang, Wenbing Huang, Fuchun Sun, Tingyang Xu, Yu Rong, and Junzhou Huang. Deep multimodal fusion by channel exchanging. Advances in Neural Information Processing Systems, 33, 2020.
[40] Mike Wu and Noah Goodman. Multimodal generative models for scalable weakly-supervised learning. Advances in Neural Information Processing Systems, 2018.
[41] Yu Wu, Linchao Zhu, Yan Yan, and Yi Yang. Dual attention matching for audio-visual event localization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 6292–6300, 2019.
[42] Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
[43] Dingwen Zhang, Guohai Huang, Qiang Zhang, Jungong Han, Junwei Han, Yizhou Wang, and Yizhou Yu. Exploring task structure for brain tumor segmentation from multimodality mr images . IEEE Transactions on Image Processing, 29:9032–9043, 2020.
[44] Shunli Zhang, Xin Yu, Yao Sui, Sicong Zhao, and Li Zhang. Object tracking with multi-view support vector machines. IEEE Transactions on Multimedia, 17(3):265–278, 2015.
[45] Tan Zhi-Xuan, Harold Soh, and Desmond Ong. Factorized inference in deep markov models for incomplete multimodal time series. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 10334–10341, 2020.
[46] BABU, TINA; SINGH, TRIPTY; GUPTA, DEEPA; and HAMEED, SHAHIN (2022) ”Optimized cancer detection on various magnified histopathological colon imagesbased on DWT features and FCM clustering,” Turkish Journal of Electrical Engineering and Computer Sciences: Vol. 30: No. 1, Article 1. https://doi.org/10.3906/elk-2108-23 Available at: https://journals.tubitak.gov.tr/elektrik/vol30/iss1/1