Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 06, June-2021

Anomaly Segmentation from 3D Biomedical Images: A Literature Survey

Vishwa Mohan Singh
School of Computer Engineering and Technology,
MIT World Peace University
Pune, India

Shivam Shrirao
School of Computer Engineering and Technology,
MIT World Peace University
Pune, India

Atharva Barve
School of Computer Engineering and Technology,
MIT World Peace University
Pune, India

Prithviraj Murthy
School of Computer Engineering and Technology,
MIT World Peace University
Pune, India

Mangesh Bedekar
School of Computer Engineering and Technology,
MIT World Peace University
Pune, India

***Abstract*— Three Dimensional Biomedical Images are a crucial means to diagnose anomalies and diseases like tumors, Alzheimer's, Parkinson's disease and many more. There are many methods to record such scans including Magnet-ic Resonance Imaging, Computer Tomography and Positron Emission To-mography. However, precisely analyzing such scans can be a difficult and time-consuming job. Therefore, computer vision methods, specifically image segmentation can be really helpful at assisting in such tasks. Segmentation refers to extracting specific segment (foreground) from the rest of the image (background). Over the years, there have been many methods to draw the segments from the biomedical image, aiming to determine the affected parts in the 3D render. Our goal with this work is to study and compare the ap-proaches for segmentation in biomedical images including MRI, CT and PET in order to determine a direction in which improvements can be made to such methods.**

Keywords*— Computer Vision; Image Segmentation; Magnetic Resonance Imaging (MRI); Computer Tomography (CT); Positron Emission Tomography (PET); Auto-encoders

I. INTRODUCTION

Operating on conditions like tumors is a critical and complicated task. Therefore, the diagnosis method is required to be timely and precise. The development of methods that can create a full three-dimensional rendition of the organs under observation can make the results a lot more accurate. Some of the most prominent methods are MRI, CT and PET scans which are used to diagnose a variety of diseases. Although this has increased the scope of biomedical imaging significantly, analyzing and segmenting these images is still a meticulous task. There-fore, appropriate computer vision algorithms can help a lot with assisting such diagnosis. Since we need a precise location and not just an estimated bounding box, the best approach to utilize computer vision will be to use image segmentation methods which aim at separating the foreground from the background. So far, in the biomedical domain, the U-Net architectures [1] and its variants [2] have been some of the most consistent networks for this job. However, their architectures leave a lot of space of improvements which significantly improves their performances in specific tasks. This is why, over the years, there have been many competitions in order to recognize the best methods to extract the anomalous segment form the scans. Many algorithms have been put forward and have achieved great accuracy with timeliness. Although, with the rapid new developments in deep learning methodologies, there is still some scope of improvements. The goal of this work is to study the existing methods using various neural net architectures and compare their results to identify some gaps or areas where improvements can lead to an improved accuracy of the segmentation model.

II. SCANS AND DATASET

A. Magnetic Resonance Imaging

These scans are done using radio waves, magnetic field gradients and strong magnetic fields [3]. This forms pictures of the anatomy and the physiological processes of the body. The main organs scanned using MRI are Brain, Chest and Abdomen. Diseases like Brain Tumors, Lung cancers, Strokes, Parkinson's Dis-ease, Blood vessel issues, Chirossis, abnormalities in bile duct and duct inflam-mation.

The most prominent competition for tumor segmentation in MRI scans is the MICCAI Brain Tumor Segmentation Challenge (BraTS). The dataset from BraTS contains for every scan, 4 – 3-dimensional channels (namely flair, t1, t1ce and t2) and the extracted segment, all stored in NIFTI (.nii.gz) format. The goal is to maximize the Dice similarity coefficient (dsc) [4] and minimize the Hausdorff distance [4] between the predicted and the actual tumorous segment.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 06, June-2021

B. Computer Tomography

Commonly referred to as CT scans, these uses X-Rays to take the image of organs from different angels. These images are then concatenated to create a 3D cross sectional image of the organs. These scans are used to detect abnormalities like intracranial bleeding, aberrations in structure, and interstitial diseases in organs like lungs and brain.

Liver Tumor Segmentation (LiTS) challenge is the resource which the work studied under CT focuses on. This challenge uses Dice score, Jaccard and Volume Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASSD) and Maximum Symmetric Surface Distance (MSSD) as mentioned by the evaluation section. All these metrics are explained in [5]. Here, there is only one channel in the scan unlike the BraTS dataset. The output contains 3 segments in total. The first one is of the background, the second one is for the liver and the third is for the lesion (tumors).

C. Positron Emission Tomography

It is a functional imaging technique that uses radioactive substances known as radiotracers to visualize and measure changes in metabolic processes, and in other physiological activities including blood flow, regional chemical composition and absorption. These are mainly used to analyze the conditions of skins and some activity disorders like Alzheimer's and Parkinson's disease.

ADNI PET dataset is used in the work studied in this paper. This dataset does not contain any output for the segments which therefore requires an unsupervised method to detect the targeted disease i.e. Alzheimer's.

III. METHODOLOGY

In this section, we will discuss the components and architectures used in creating the segmentation network and the results of these architecture on the datasets discussed in the previous section.

A. Basic Concepts

The main backbone of all the models will be a 3D Convolution blocks. This al-lows us to create feature maps from a 3-dimensional image (which in this case is our scans). A simple 3D CNN architecture is seen in Fig. 1.

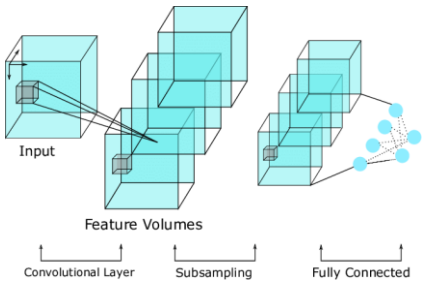

Fig. 1: 3D Convolutional Neural Network

All architectures bulit for image segmentation of 3D scans are based on 3D-CNN.

The recuring structure which appears frequently in the architectures as either a method of creating the segments or as a method for regualarization of the neural network is the encoder-decoder model. The most basic enc-dec model called the autoencoders was studied by Yuan F. et al. in [6] where the input and the output shape of the network is the same. An image recreation autoenocder can be directly used to generate the segment as demonstrated by Mallick et al. in [7]. However, this method is not as effective due to the information loss between the encoder and the decoder blocks. Therefore, this method is mostly used as a template or a performance enchancing method in the segmentation models under study.

B. Methods for Segmentation in MRI

The BraTS challenge for MRI Brain Tumor Segmentation is launched every year with some great models producing some great results. We will be analyzing all the best solutions from every year since 2018.

The BraTS 2018 dataset was won by the work of Andriy Myronenko presented in [8]. This model works by first creating a standardized green block with basic layers and using those blocks to create the whole architecture. After a series of green blocks creating a common encoder, this model is split into two parts. The first one is tasked with generating the tumor segment with the output from the last common green block which is to be used as the final result. The second part is used to recreate the input and is meant to regularize the shared encoder.

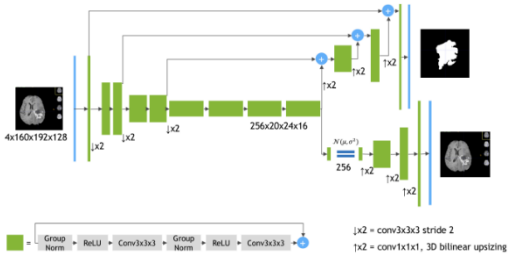

Fig. 2: Segmentation model with AE ragularization [8]

The next model that won the BraTS'19 challenge was the two-stage cascaded U-Net by Z. Jiang et. al. [9]. It builds on top of the U-Net architecture originally proposed by O. Ronnenberger et. al. [1] which uses an encoder-decoder architecture to generate the segment. The U-Net model features shortcuts form the encoder to the decoder part to feed it the information which might have been lost during the downscaling and upscaling.

In the two stage architecture, the first stage U-Net predicts the segmentation map roughly. This map is fed to the second stage along with the raw image. This can provide a more accurate segmentation map.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 06, June-2021

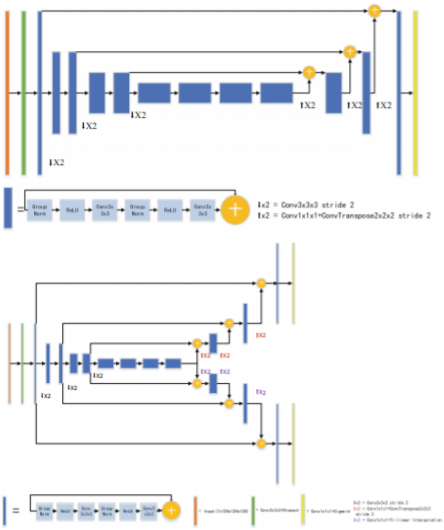

Fig. 3: The stage 1 (top) and stage 2 (bottom) of the cascade network [9]

The winning model of the BraTS'20 was created by F. Isensee et. al. in [10] by using a model called the nnU-net proposed by Isensee himself in [11] which tries to address the problem that the proposed networks for the previous challenges are higly specialized for brain tumor segmentation. The nnU-Net architecture automatically configures segmentation pipelines for arbitrary biomedical datasets.

They do this by getting some infered parameters such as the resampling size, batch size, normalzation parameters etc by using the data fingerprints and creating and testing pipeline fingerprints that are trained and cross validated. nnU-Net found the architecture that can perform best with the BraTS'20 dataset.

The main criterias used to compare and rank these models is the Dice Score Coefficient and Hausdorff distance (95%) computed on Enhanced Tumor (ET), Tumor Core (TC) and Whole Tumor (WT) [12]. The following Table 1 shows the comparison of the aforementioned models using the given metrics on the validation data.

TABLE I

COMPARISON OF ANNUAL WINNERS OF BRATS CHALLENGE

Models	DSC (ET)	DSC (WT)	DSC (TC)	H95 (ET)	H95 (WT)	H95 (TC)
AE Reg. [8]	**0.825**	**0.912**	**0.870**	3.997	4.537	6.761
Cascaded UNet [9]	0.802	0.909	0.864	**3.145**	4.263	5.439
nnU-Net [10]	0.798	**0.912**	0.850	23.49	**3.692**	**3.692**

C. *Methods for Segmentation in Computer Tomography*

There have been a lot of submission on the LiTS challenge, using multiple methods. The best submission was made by M. Bellver et. al. in [13] where they propose a 2-stage model. The first model is supposed to extract the segment of the liver which helps in narrowing down the look-up area and the second model is applied on the extracted liver segment to get the lesion area. Both the models are based on the Deep Retinal Image Understanding (DRIU) proposed by Kevis-Kokitsi Maninis et. al. in [14] which aims at segmenting both the retinal vessel and the optic disc by first passing the input image through a base network which is VGG16 [15] pretrained on the ImageNet set [16] and then using specialized layers to extract specific segments.

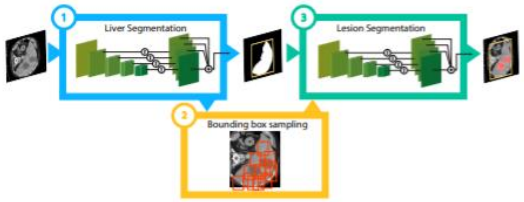

Fig. 4: Two stage model for liver and lesion segment detection [13]

Another recent U-Net based architecture that performs well on the LiTS dataset is the KiU-Net proposed by J. Valanarasu et. al in [17]. They try to address the problem that the conventional U-Net and it's variants fail to detect the tiny structures which are present in the segmentation maps specially when the bounadries are blurry. To tackle this, they use the concept of overcomplete network proposed in [18] to create Kite-Net which is an overcomplete version of the U-Net which uses the encoder to project the input impage onto a spatially higher dimesion. The KiU-Net contains the two networks, the U-Net and the Kite-Net which are connected with each other on each stage using Cross Residual Fusion Blocks proposed in [17] to learn the complimentary features from both the networks to improve the segmentation accuracy.

The following table includes the score obtained by both the networks on the LiTS challenge for liver and Lesion Segmentation.

TABLE II

COMPARISON OF ANNUAL WINNERS OF LITS CHALLENGE

Models	DSC	VOE	ASSD	MSD
2-Stage DRIU [13]	0.850	0.335	**0.873**	**5.294**
KiU-Net [17]	**0.942**	**0.105**	1.771	29.98
U-Net [1]	0.9346	0.117	1.945	33.88

D. *Methods for Segmentation in Positron Emission Tomography*

The Since the dataset in observation, i.e. the ADNI set of PET Scans, does not have labels or precalculated segments, we cannot train a supervised model for the same. Work by A. Meena et. al. presented in [19] and [20] shows how clustering algorithms can be used to find the anomalous segments. The work test k-means and fuzzy c-means clustering algorithms to segment different areas based on their locations and their intensity. The intensity/pixel value in an area represents the amount of amyloid protein present between the brain. A large quantity of the same can damage the brain cells and cause Alzheimer's disease. Therefore, using this method, we can find how much area of the brain is covered with amyloid and confirm the diagnosis. In the work presented in [19] and [20], the number of clusters considered in K-means algorithm is 5.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 06, June-2021

Another study by E. Pfaehler et. al. presented in [21] studies two for tumor segmentation in PET scans. The two methodologies under study consideration is the classical U-Net architecture and the Textural Feature Segmentation which was also proposed by E. Pfaehler et. al in [22].

The performances were evaluated using JC Median and test-retest coefficient. TF achieved a JC Median score of 0.7 and the U-Net/CNN achieved the score of 0.73. The TRT% scores for U-Net and TF are 13.9% and 13.0% respectively.

IV. Results Discussion and Possible Improvements

As seen in the work, most 3D imaging can be segmented and diagnosed using autoencoders and architectures similar to U-Net. Additions like cascading (as seen in [9] and [13]), regularization and a parallel network to enhance the results of the model. However, some improvements which are already helping many models in various computer vision tasks can be made here to improve different aspects of the model. Following are some examples of such improvements.

Recently use of Attention Mechanism originally proposed in [23] has been very useful in language models allowing it to pay attention to important parts of input. Similar mechanism has been proved to be useful in computer vision tasks in Attention Augmented Convolution [24] and Vision Transformers [25].

CNN Architectures like ResNeXt [26] have used Grouped Convolution and current state of the art like MobileNet, Efficient, etc. have made use of Depthwise Separable Convolutions presented in [27] to greatly decrease the number of pa-rameters and computation required while maintaining higher accuracy.

Segmentation architectures like Deeplab have used Spatial Pyramid Pooling presented in [28] to feed input image at different spatial resolutions directly to deeper layers in network and Deep Supervision [29] to get multiple segmentation maps at different resolution scales from deeper layers in network. This allows the network to process the objects in images at different scales.

Sparse Convolution has been used in segmenting Point Clouds to reduce com-putation by only operating on non-empty voxels in the 3D Point Cloud Space [30].

Pixel Shuffle proposed by Shi et. al in [31] is a technique which reshuffles the channels to increase the resolution during the upsampling path of the decoder.

V. CONCLUSION

In this work, we discussed the work done and models used in anomaly segmenta-tion in 3D biomedical imaging and possible ways to improve the existing meth-ods.

Using the Methods discussed in the result section, we can further improve the capabilities of such networks to improve upon aspects like dsc, training time and model weight.

REFERENCES

[1] Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4_28. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Siddique, Nahian & Sidike, Paheding & Elkin, Colin & Devabhaktuni, Vijay. (2020). U-Net and its variants for medical image segmentation: theory and applications.

[3] Grover VP, Tognarelli JM, Crossey MM, Cox IJ, Taylor-Robinson SD, McPhail MJ. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. J Clin Exp Hepatol. 2015;5(3):246-255. doi:10.1016/j.jceh.2015.08.001

[4] Taha, A.A., Hanbury, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 15, 29 (2015). https://doi.org/10.1186/s12880-015-0068-x

[5] Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. J Med Imaging (Bellingham). 2018;5(1):015006. doi:10.1117/1.JMI.5.1.015006

[6] Yuan, F.-N & Zhang, L. & Shi, J.-T & Xia, X. & Li, G.. (2019). Theories and Applications of Auto-Encoder Neural Networks: A Literature Survey. Jisuanji Xuebao/Chinese Journal of Computers. 42. 203-230. 10.11897/SP.J.1016.2019.00203.

[7] Mallick, Pradeep Kumar & Ryu, Seuc & Satapathy, Sandeep & Shruti, Mishra & Nhu, Nguyen & Tiwari, Prayag. (2019). Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2902252.

[8] Myronenko A. (2019) 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi A., Bakas S., Kuijf H., Keyvan F., Reyes M., van Walsum T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol 11384. Springer, Cham. https://doi.org/10.1007/978-3-030-11726-9_28

[9] Jiang, Zeyu & Ding, Changxing & Liu, Minfeng & Tao, Dacheng. (2020). Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task. 10.1007/978-3-030-46640-4_22.

[10] Isensee, F., Jaeger, P. F., Full, P. M., Vollmuth, P., and Maier-Hein, K. H., “nnU-Net for Brain Tumor Segmentation”, arXiv e-prints, 2020.

[11] Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z

[12] Tasks' Description and Evaluation Framework : https://www.med.upenn.edu/cbica/brats2020/tasks.html

[13] Bellver, Miriam & Maninis, Kevis-Kokitsi & Pont-Tuset, Jordi & Giró-i-Nieto, Xavier & Torres, Jordi & Van Gool, Luc. (2017). Detection-aided liver lesion segmentation using deep learning.

[14] Maninis, Kevis-Kokitsi & Pont-Tuset, Jordi & Arbelaez, Pablo & Van Gool, Luc. (2016). Deep Retinal Image Understanding. 10.1007/978-3-319-46723-8_17.

[15] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[16] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[17] Valanarasu, Jeya Maria Jose & Sindagi, Vishwanath & Hacihaliloglu, Ilker & Patel, Vishal. (2020). KiU-Net: Overcomplete Convolutional Architectures for Biomedical Image and Volumetric Segmentation.

[18] M. S. Lewicki and T. J. Sejnowski, “Learning overcomplete representations,” Neural computation, vol. 12, no. 2, pp. 337–365, 2000.

[19] Arumugam, Meena & Kothandaraman, Raja. (2013). Segmentation of Alzheimers Disease in PET scan datasets using MATLAB. International Journal on Information Sciences and Computing. 6. 10.18000/ijisac.50121.

[20] Arumugam, Meena & Kothandaraman, Raja. (2013). K Means Segmentation of Alzheimers Disease in PET scan datasets: An implementation. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST. 117. 10.1007/978-3-319-11629-7_24.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 06, June-2021

[21] Pfaehler, E., Mesotten, L., Kramer, G. et al. Repeatability of two semi-automatic artificial intelligence approaches for tumor segmentation in PET. EJNMMI Res 11, 4 (2021). https://doi.org/10.1186/s13550-020-00744-9

[22] Pfaehler, E., Mesotten, L., Kramer, G., Thomeer, M., Vanhove, K., de Jong, J., Adriaensens, P., Hoekstra, O. S., & Boellaard, R. (2020). Textural Feature Based Segmentation: A Repeatable and Accurate Segmentation Approach for Tumors in PET Images. In B. W. Papiez, A. I. L. Namburete, M. Yaqub, J. A. Noble, & M. Yaqub (Eds.), Medical Image Understanding and Analysis - 24th Annual Conference, MIUA 2020, Proceedings (pp. 3-14). (Communications in Computer and Information Science; Vol. 1248 CCIS). Springer. https://doi.org/10.1007/978-3-030-52791-4_1

[23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[24] I. Bello, B. Zoph, Q. Le, A. Vaswani and J. Shlens, "Attention Augmented Convolutional Networks," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 3285-3294, doi: 10.1109/ICCV.2019.00338.

[25] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[26] Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "Aggregated residual transformations for deep neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492-1500. 2017.

[27] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258. 2017.

[28] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." IEEE transactions on pattern analysis and machine intelligence 37, no. 9 (2015): 1904-1916.

[29] Wang, Liwei, Chen-Yu Lee, Zhuowen Tu, and Svetlana Lazebnik. "Training deeper convolutional networks with deep supervision." arXiv preprint arXiv:1505.02496 (2015). J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[30] Graham, Benjamin, Martin Engelcke, and Laurens Van Der Maaten. "3d semantic segmentation with submanifold sparse convolutional networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9224-9232. 2018.

[31] Shi, Wenzhe, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1874-1883. 2016. K. Elissa, “Title of paper if known,” unpublished.