Review on Deep Learning for Medical Image Segmentations

Vysakh Poolakkal Nair (21239041)
Data Analytics
National university of Ireland, Galway,
V.Nair1@nuigalway.ie

Anjaly Abraham (21251158)
Data Analytics
National University of Ireland, Galway,
a.abraham4@nuigalway.ie

Jinu Sivaraj (21238533)
Data Analytics
National University of Ireland, Galway,
j.sivaraj1@nuigalway.ie

***Abstract*— Image segmentation has been the cause of major advancements in the medical field, enabling better diagnosis and analysis of illnesses, which in turn allows doctors to validate and improve their treatment methods more effortlessly. Due to its capabilities, medical image segmentation has become a mainstream research subject in the areas of computer science and machine learning. With further experimentation and study in the deep learning field, convolutional neural networks have evolved as the algorithm that dominates most image segmentation models in recent years [2]. In this paper, we focus on the review of specific research papers that use different deep learning algorithms and techniques. Firstly, the ideas and terminology behind the topic is introduced, followed by related works, in which developments from machine learning model-specific approaches to deep learning data-based approaches are highlighted. Next the algorithms and processes used in the specific papers of focus are summarized, in which 2D and 3D imaging techniques, as well as semantic segmentation and instance segmentation techniques are demonstrated, along with the evaluation of the performance of the different techniques by analyzing their respective outcomes. Further limitations and suggestions of modifications, from our perspective, are also addressed, concluding with an overall comparison of performance of the various techniques discussed.**

Keywords—image segmentation, medical image processing, deep learning, convolutional neural networks

I. INTRODUCTION

Image segmentation is a process that splits an image into different regions of similar properties to differentiate an object from the background in an image. It is a crucial part in image processing. Currently image segmentation methods are emerging at a faster pace with more precision because of numerous theories and technologies [1].

Medical image segmentation is the extraction of pixels of an object i.e., organs, tissues, diseased bodies etc. from 2D or 3D images, manually or automatically, by partitioning the image into different regions based on resemblance between them. The developments in deep learning architectures for medical imaging, helps doctors perform various analysis of damaged organs or other regions of interest, to evaluate the differences quantitatively and qualitatively before and after treatments, enhancing efficiency, precision, and quality of diagnosis [4].

There are two types of image segmentation tasks, semantic segmentation, and instance segmentation. While instance segmentation differentiates instances based on different categories, semantic segmentation refers to classification involving pixels, where each pixel corresponds to a specific category.

With the recent improvements in treatments within the medical field, new medical imaging equipment that are widely used include Computer Topography (CT), Magnetic Resonance Imaging (MRI), X-ray and Ultrasound Imaging (UI). The information resulting from the use of these tools became an important aspect of medical diagnosis, which led to more research in this area within the branch of computer science. In comparison to the traditional machine learning and computational techniques, deep learning has managed to achieve the best results in image segmentation, both in terms of accuracy and speed [1].

Earlier methods on medical image segmentation involved techniques of edge detection, template matching, statistical shape modelling etc. According to the type of data involved, machine learning often categorizes these into supervised, unsupervised, and weakly supervised learning techniques. Since labelled data isn't common with medical images, it is required to use unsupervised learning, which has higher difficulty in learning compared to the other learning techniques.

Before deep learning became popular, a common method to use model driven implementations for medical segmentation. Although a lot of research have been done in these areas, there has always been challenges involving image segmentation due to the difficulty in feature representation. This is particularly true in terms in medical images due to the challenges associated with image noise, contrast, blur etc. Due to advancements in deep learning through convolutional neural networks (CNN), the hand-crafted feature implementations are no longer necessary, as CNN has managed to do these techniques successfully while also being indifferent to the issues of noise, contrast, blur etc [2].

In this paper, we review different deep learning techniques and algorithms by various models from specific papers, evaluate the performance of these methods, also include any limitations and recommendations for possible improvements in the future.

II. LITERATURE REVIEW

From the 1970s until now, there have been tremendous technological breakthroughs in the field of medical image analysis. The timeline can be viewed based on the introduction of new imaging modalities (2D image analysis, MRI, 3D images, UI) as well as advances in computer power and computational approaches [3]. The earlier work in medical image processing includes automatic tumor location using pattern recognition [5], display and enhancement techniques by Pizer and Todd-Pokropek [6], performing edge grouping using smoothing and contour search [7], image matching and subtraction [8], [9], among the initial works 3D edge detection for volumetric dataset proposed by Zucker and Hummel [10] in 1981 was quite remarkable.

Following the advent of MRI and computer-aided diagnosis, different clustering approaches such as hierarchical clustering [11], statistical clustering [12] have been studied by researchers in the field of region-based image segmentation for distinguishing grey and white matter. Using temporal boundary tracking and low-level feature extraction, a group of researchers developed a robust contour tracker from echocardiographic images in the late 1990s, using the texture of the image as a key feature [13]. Deformable models [14], active-contour-based segmentation [15], and intensity-based segmentation models were other significant advancements throughout these periods. However, traditional techniques, on the other hand, lacked the ability to segment many classes across a variety of datasets and needed human intervention to extract the key features, but they were quite popular and served as the foundation for many successful image analysis models.

Before the advent of deep learning (DL), there were many successful algorithms like Markov Random Fields (MRF) [16], Support Vector Machines (SVM) [17], K means clustering [18] which can be applied for certain image segmentation problems, many of these have been covered in the reference [19] along with the most advanced deep learning models. But most of the classical machine learning models faced difficulties in feature representation especially medical images with RGB pixels and were restricted to do specific tasks.

Semantic segmentation approaches based on artificial intelligence have become highly popular in the field of medical image segmentation, especially with the rapid growth of big data, computing power, and deep learning. DL models anymore do not require handcraft features and with a single model, it can be used for a variety of medical imaging modalities.

Convolutional neural networks (CNN) are the most successful image analysis models to date, as they are insensitive to visual distortions such as blur, noise, low contrast, etc and they can achieve hierarchical feature representation [20]. The basis of the CNN dates to 1959 were Hubel and Wiesel [21] studied the visual nervous system of cats and was inspired by this, Fukushima [22] created a neural network model for pattern recognition called neocognitron in 1980, which is considered the forerunner of CNN. Inspired by this, a group of researchers led by LeCun presented the LeNet5 [23] neural network with backpropagation for categorizing handwritten digits, which became the foundation for current CNNs. Over a decade later, in 2012, Krizhevsky [24] presented Alexenet for the imageNet challenge, which was a true breakthrough in classifying images using CNNs, which were previously limited only in academia.

Following AlexNet's success, subsequent research on CNN has focused on deepening the layers, lowering the memory footprint by stacking smaller kernels, employing a fixed kernel size throughout the network, and increasing the efficiency of the training parameters by layering complex building blocks on top of the existing neural network. VGGNet [25], GoogLeNet [26], and ResNet [27] were some of the more notable works in the field, with VGGNet having 19 layers and fixed kernel size, GoogleNet having 22 layers, and using inception blocks to replace the mapping function, and ResNet using residual blocks to improve the efficiency of the training parameters. However, the more complex models were susceptible to overfitting and difficult to optimize.

III. REVIEW OF SCIENTIFIC PAPERS

A. Semantic Segmentation of Pathological Lung Tissue with Dilated Fully Convolutional Network

This paper proposes a dilated Fully Convolutional N etwork (FCN) [28] which consists of only convolutional layers that uses dilated kernels to increase the receptive field, instead of down sampling the feature maps. A dilated kernel helps in expanding the area of the input image covered. It obtains more information from each convolution operation. The network consists of 13 convolutional layers. The first 10 layers has 32 kernels of size 3*3 and dilation rates in Fibonacci order. The output of these 10 convolutional layers along with the input to the network are then concatenated (creating 321 feature maps) which are then passed through a dropout layer with a rate of 0.5 and fed to the rest of the network. The last 3 layers contain convolutions of size 1*1 for dimensionality reduction which reduce the dimension from 321 to 6 which is the number of classes required in the problem [29].

A normalization approach proposed before called instance normalization has exhibited good performance in texture synthesis, image stylization and image to image translation. Therefore, a batch-normalization layer follows each convolution in the FCN architecture.

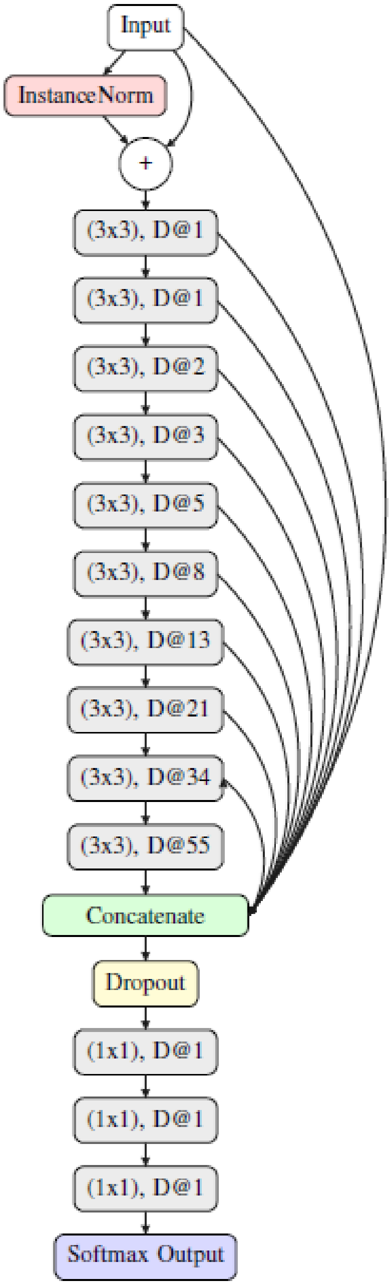

Fig. 1: The architecture of the proposed network. Each gray block corresponds to a block like the ones presented in Fig. 2 *[29]*.

The proposed network was reporting a balanced CV accuracy of 81.8% on the 172 HRCT scans compiled for the study.

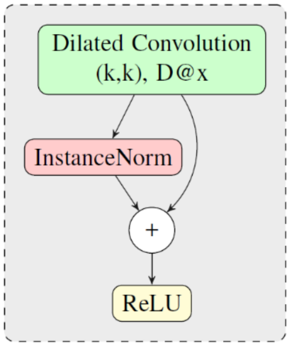

Fig 2: The block function of the proposed architecture *[29]*

B. *Deep learning for medical image segmentation using multi-modality fusion*

This paper focuses more on the multi-modal nature of the input data into the segmentation architecture. It introduces input-level, layer-level, and decision level fusion into the architecture. The input level architecture combines the images before feeding into the network, whereas the layer level fuses the learned feature representations of each individual image. The decision level fusion uses the output produced from the network for individual images. The U-Net architecture when used as the segmentation architecture was producing the best result with a dice score of 0.88 [30].

C. *U-Net: Convolutional Networks for Biomedical Image Segmentation*

The U-net architecture is a combination of an encoder (also known as a contracting path) and a decoder (also known as an expansive path). The encoder is like a normal convolution network, which consists of two 3X3 convolutions which are repeated several times, followed by a Rectified Linear Unit (ReLU) and max pooling.

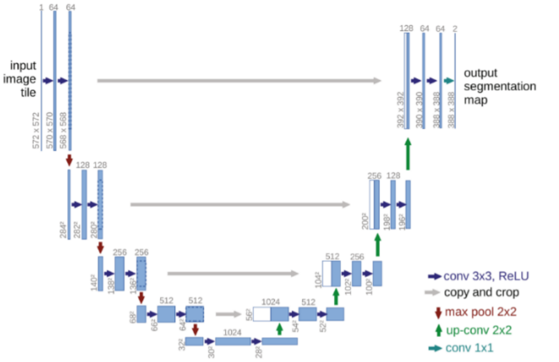

Fig. 3. U-net architecture. Each blue corresponds to a multi-channel feature map. The number of channels is denoted on top of the box *[31]*.

The decoder involves up-sampling of feature maps as well as a 2x2 convolution at each level, which is cropped and joined together with the corresponding feature map from the previous step(encoder) for better resolution in the output. This is followed by two 2x2 convolutions and a ReLU activation. In the last layer, a 1x1 convolution is applied to create a fully segmented image.

Without any pre-processing or post processing steps, the U-net architecture achieved a result of 0.0003529 for warping error and 0.0382 for rand error. This is a great improvement in comparison to the sliding window convolution network with warping error of 0.000420 and a rand error of 0.0504 [31].

D. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation

The V-Net is a variant of the U-Net architecture, but the paper is mainly proposed for medical images. Unlike 2D U-nets, the V-Net architecture is mainly designed for 3 dimensional images. It also follows the U-Net by superimposing the features from the compression path onto the expanded path to supplement the lost information. The main difference between both the architectures is the introduction of a short circuit connection of residual connections in V-Net. This is the biggest improvement of V-Net over U-Net. The V-Net architecture also replaces the up sampling and down sampling pooling layers with convolutional layers because it helps in having a smaller memory footprint during training.

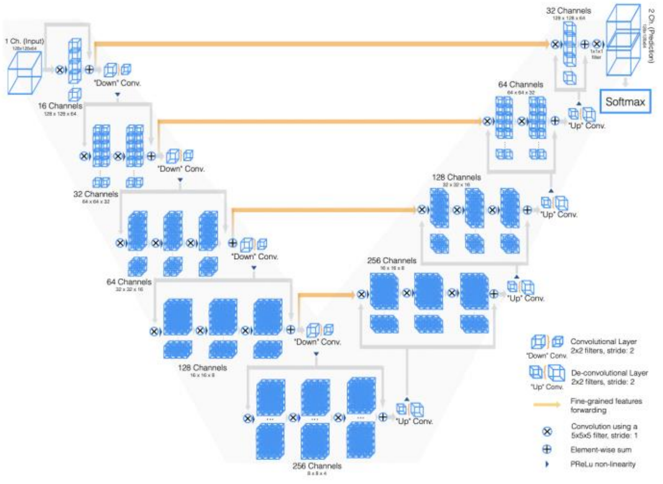

Fig. 4. Schematic Representation of the V-Net Architecture *[32]*

To train the network, they have used dice loss function which are commonly used in cases of medical images. According to the paper, the V-Net architecture along with the dice-based loss produced an average dice score of 0.869 [32].

E. CE-Net: Context Encoder Network for 2D Medical Image Segmentation

This paper introduces the CE-Net architecture which consists of three parts: A feature encoder, Context Extractor Module, and a Feature Decoder. The feature encoder is the stage where the images are fed into a pretrained network to obtain the image features or embeddings. The pre-trained network used for this study was ResNet-34. The paper reported that this helped in the faster convergence of the model. The Context Extractor Module (CEM) consists of a DAC and a RMP block. The DAC block is used to extract features from objects of different sizes and a RMP block is used to tackle the issue of variation of object sizes in medical images. The feature decoder is then used to bring back the semantic information lost in the previous stages of the architecture. This is done by upscaling and deconvolution. The information for this stage is obtained from the encoder part by skip connections [33].

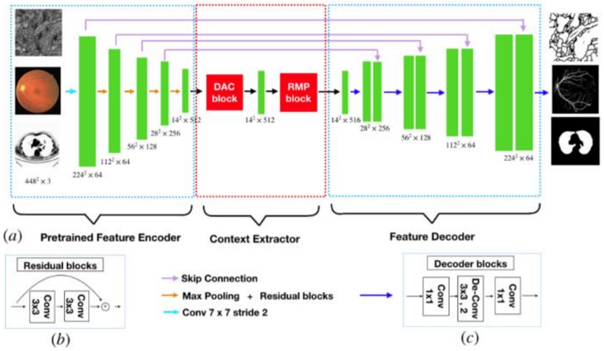

Fig. 5. Illustration of the CE-Net *[33]*.

F. FED-Net: Feature Fusion Encoder Decoder Network For Automatic Liver Lesion Segmentation

The proposed Feature fusion Encoder-Decoder network implements an encoder part that uses a pre-trained ResNet50 model trained on the ImageNet dataset. This encoder part is then divided into four blocks and a residual convolution block is added after 4 blocks separately. The output of each block is then fused together (feature fusion), and each unit is fed to the decoder.

Models like U-Net directly fuses the features from the encoder to the decoder, but they may not be able to efficiently fuse the features from different levels. Therefore, an attention based fusion network is proposed in this paper.

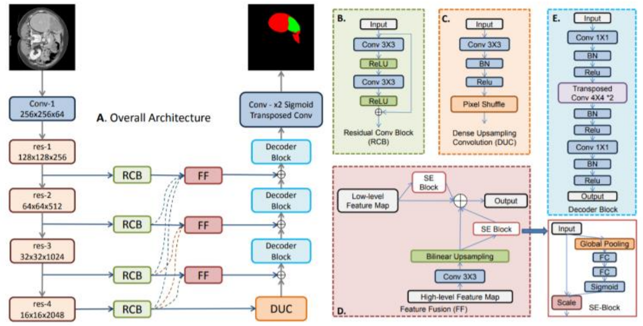

Fig. 6. The overall approach of the architecture and the individual components of the FED-Net *[34]*.

Dense upsampling convolutions are then used to compensate for the lost information during the upsampling process. Binary Cross-entropy combined with the Jaccard Index is used as the loss function. The model was producing a dice score of 0.766 on the dataset [34].

IV. LIMITATIONS AND MODIFICATIONS

The above section mentions a few scientific papers that produced some excellent results in the image segmentation task in the field of medical image processing. However not all papers are perfect and there are some shortcomings connected to each paper.

In the case of FCN, the upsampling technique has its own disadvantages since the results produced are fuzzy and insensitive to the details of the image. The authors have added high padding to the images which will result in a lot of unwanted noise. Multiple upsampling layers like what's used in a U-Net architecture with residual connections could be used to improve the fuzziness of the images. The residual connections might improve the architecture in the case of a deep architecture.

In case of the U-Net architecture, some semantic information might be lost between 2 paths while passing through the network. This loss of info could be minimised by using the U-Net++ architecture (influenced by DenseSet), which utilises skip connections between the encoder and decoder paths. The training time could be further reduced even with the less labelled data by using the 3D U-net architecture in which all 2D operations and replaced by 3D operations. (3D convolutions, max pooling etc. resulting in 3D segmented image). This is due to repetition in shapes and structures of 3D images, which aids in a faster training process. Pre-processing and post-processing of data wasn't performed on the data being analysed. The performance could have been further improved with these additional steps.

Data scarcity is a major problem in the medical field, and this normally leads to overfitting, this was a limitation mentioned in the multi-modality fusion implementation by Tongxue Zhou [30] as well as the V-Net study. High class imbalance was also reported in the multi-modality fusion paper which made the predictions challenging. The V-Net model was trained on 50 volumes of MRI images. Since the training data is so less, applying some transfer learning could improve the overall performance. Using a multi-modal dataset for each task combining with the ideas of paper 1(A) [30] could be used for extracting extra features thus improving the performance of the model.

V. CONCLUSION

With rapid developments in the medical image segmentation, many models and algorithms have become available in this field. In this paper, we discussed and compared some of those major algorithms.

A typical network consists roughly of convolutional layers, with variations depending on the model used. These modifications cater for information loss as the input passes through the convolution layers, different image dimensions (3D or 2D) and different priorities, i.e., cost, performance, speed etc.

Even with the numerous computational models presently available, the area of medical image segmentation still has some shortcomings associated with it. The lack of knowledge artificial intelligence researchers has of the medical conditions and requirements, and the lack of understanding medical practitioners have of AI techniques, makes it difficult to come up with the best solutions that match their sufficient needs.

The variance between normal images and medical images makes it harder to apply deep learning techniques due to difference in contrast, noise associated with data etc.
Although a lot of research have been done for 2D and 3D image segmentations, a fine balance is yet to be acquired as 3D methods, although exhibits better performance, tend to be very expensive in comparison to 2D processes.

The research of medical image segmentation in the deep learning field is consistently emerging and showing constant improvements with better analysis of diseases, in turn enabling the possibilities of better treatment of these diseases. With the continuous research in this field, more innovations can be expected in the future to contradict the flaws and limitations currently present and empower more efficient, accurate and economical models [1].

Author contributions: methodology of 6 papers-Vysakh Nair, Jinu Sivaraj, Anjaly Abraham; writing—Literature Survey-
Jinu Sivaraj; Summary of 6 papers, limitations, and modifications- Vysakh Nair; Introduction, Conclusion, Abstract- Anjaly Abraham; final editing and review- Jinu Sivaraj

VI. REFERENCES

[1] X. Liu, L. Song, S. Liu and Y. Zhang, "A Review of Deep-Learning-Based Medical Image," *sustainability,* 2021.

[2] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng and A. K. Nandi, "Medical Image Segmentaion Using Deep Learning: Asurvey," 2021.

[3] J. S. Duncan and N. Ayache, "Medical image analysis: Progress over two decades and the challenges ahead," *EEE transactions on pattern analysis and machine intelligence,* pp. 85-106, 2000.

[4] M. H. Hesamian, W. Jia, X. He and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," 2019.

[5] J. Sklansky and D. Ballard, "Tumor Detection in Radiographs, Computers and Biomedical Research," vol. 6, no. 4, pp. 299-321, Aug. 1973.

[6] S. Pizer and A. E. Todd Pokropek, "Improvement of Scinti- grams by Computer Processing," *Seminars in Nuclear Medicine,* vol. 8, no. 2, pp. 125-146, Apr. 1978..

[7] A. Martelli, "An Application of Heuristic Search Methods to Edge and Contour Detection," *Comm. ACM,* vol. 19, pp. 73-83, 1976.

[8] M. Yanagisawa and M. Levine, "Registration of Locally Distorted Images by Multiwindow Pattern Matching and Displacement Interpolation: The Proposal of an

Algorithm and Its Application to Digital Subtraction Angiography," *Proc. Seventh Int'l Conf. Pattern Recognition*, pp. 1,288-1,291, 1984..

[9] J. M. Fitzpatrick, J. J. Grefenstette, D. R. Pickens, M. Mazer and J. M. Perry, "A system for image registration in digital subtraction angiography. In Information processing in medical imaging," pp. 415-434, 1988.

[10] S. Zucker and R. Hummel, "A Three-Dimensional Edge Operator," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 3, no. 3, pp. 324-330, 1981.

[11] M. O'Donnell, J. Gore and W. Adams, "Towards an Automated Algorithm for NMR imaging," *Initial Segmentation Algorithm, Medical Physics,* vol. 13, pp. 293-297, 1986.

[12] D. Ortendahl and J. Carlson, "Segmentation of magnetic resonance images using fuzzy clustering. In Information processing in medical imaging," pp. 91- 106, 1987.

[13] G. Jacob, J. A. Noble, M. Mulet Parada and A. Blake, "Evaluating a Robust Contour Tracker on Echocardiographic Sequences, Medical Image Analysis," vol. 3, no. 1, pp. 63-75, 1999.

[14] T. McInerney and D. Terzopolous, "Deformable Models in Medical Image Analysis: A Survey, Medical Image Analysis," vol. 1, no. 2, pp. 91-108, 1996.

[15] C. Xu and J. L. Prince, "A Generalized Gradient Vector Flow for Active Contour Models," *Information Sciences and Systems,* pp. 885-890, 1997.

[16] S. Z. Li, "Markov random field models in computer vision," in *n European conference on computer vision*, Berlin, 1994.

[17] O. Chapelle, P. Haffner and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE transactions on Neural Networks,* vol. 10, no. 5, pp. 1055-1064, 1999.

[18] H. P. O. S. H. Ng, K. W. C. Foong, P. S. Goh and W. L. Nowinski, "Medical image segmentation using k-means clustering and improved watershed algorithm," *In 2006 IEEE southwest symposium on image analysis and interpretation,* pp. 61-65, 2006.

[19] H. Seo, M. Badiei Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia and L. Xing, "Machine learning techniques for biomedical image segmentation," *An overview of technical aspects and introduction to state-of-art applications*, vol. 47, no. 5, p. e148–e167, 2020.

[20] T. Lei, R. Wang, B. Zhang, H. Meng and A. K. Nandi, "Medical image segmentation using deep learning: a survey," 2020.

[21] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkeys triate cortex," *The Journal of physiology,* p. 215–243, 1968.

[22] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, Competition, and cooperation in neural nets," p. 267–285, 1982.

[23] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Handwritten digit recognition with a back-propagation network, Proceedings of the Advances in Neural Information Processing Systems (NIPS)," p. 396–404, 1989.

[24] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, 2012.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *The International Conference on Learning Representations (ICLR)*, 2015.

[26] C. Szegedy, Y. J. W. Liu, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[27] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

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

[29] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe and S. Mougiakakou, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *IEEE journal of biomedical and health informatics,* vol. 23, no. 2, pp. 714-722, 2018.

[30] T. Zhou, S. Ruan and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," 2019.

[31] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *In International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015.

[32] F. Milletari, N. Navab and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation.," in *In 2016 fourth IEEE international conference on 3D vision (3DV) (pp. 565-571)*, 2016.

[33] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. H., Y. Zhao and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging,* vol. 38, no. 10, pp. 2281-2292, 2019.

[34] X. Chen, R. Zhang and P. Yan, "Feature fusion encoder decoder network for automatic liver lesion segmentation," *In IEEE 16th international symposium on biomedical imaging (ISBI),* pp. 430-433, 2019.