An Overview of Image Segmentation Algorithms Used in the Medical Field

Omar Zayed, Nile Egyptian Schools - 6th of October Branch
Mohammed Ragab, STEM High School for Boys – 6th of October

Mentor: Youssef Amir, STEM High School for Boys – 6th of October

Abstract

There has been a rise in the usage of image segmentation algorithms in the medical field. Medical image segmentation includes many types, such as MRI and CR image segmentation. Studying those images help improve the ability to detect diseases earlier, in addition to increasing the knowledge about organ functions. Some of those applications include recognizing brain development, tumor segmentation, and surgical planning. This study aims to provide information about some of the most famously used techniques in the field of medical image segmentation, which can help improve the quality of applications constructed in the medical segmentation field. Four models were focused on in this review, which are: U-Net, V-Net, FCN, and DenseNet. The paper differentiates between the models through architecture, accuracy, and training environment.

I. Introduction

Having become one of the most famous research areas, deep learning is a type of machine learning model that can benefit from large amounts of data. It has become an essential part of many AI applications, of which is image segmentation, an image processing technique. It is a field used to improve machines' ability to recognize image contents and sectors. Image segmentation is used widely in the medical field with various purposes and goals. In addition, as several models are used in image segmentation for the medical field, this review paper is searching for the most effective image segmentation model in the medical field.

Searching for prior literature about medical image segmentation models enables the study to compare different ones. First, many discuss the U-Net architecture in medical image segmentation [1]–[3]. U-Net is used widely in that field, as this algorithm provides high accuracy [3]. Other research papers are concerned with the usage fully convolutional network (FCN) [4]. Finally, some other studies used the V-Net [5] and DenseNet [6] architectures to implement segmentation models. On the other hand, there was less medical research using architectures such as threshold-based and watershed segmentation. Using this literature can help compare the weaknesses and strengths of each model.

In this paper, we'll introduce the field of image segmentation and its role in the medical field, then compare how current techniques perform on medical datasets and give an overview of each model.

This paper will improve decision-making while creating medical image segmentation applications, as well as define the next steps for future research in the field. This paper, as well, will save compute resources, which in most cases are finite in the hands of developers and computer vision specialists.

As medical image segmentation is a commonly-used technique, this paper provides readers with the best-performing model on medical datasets for their specific application.

II. Background

Image segmentation is a complex computer vision task in which the algorithm or model has to assign a class to each pixel in the image. The image after running through an image segmentation task will look like the image below:

Figure 1: illustration example for a segmented image.

As you can see, each different class is assigned a different color.

i. Types of Segmentation

There are two main types of image segmentation, instance segmentation, and semantic segmentation. From its name, instance segmentation tries to label each entity separately, figuring out which pixel belongs to which entity. Semantic segmentation aims only to figure out to which class a certain pixel belongs.

ii. Machine learning In the Medical Field

The medical field is finding a use for machine learning. It can help in diagnosing diseases from medical images or data, drug discovery, and many more applications. As mentioned previously, we will give an overview of a particular subset of deep learning, which is a subset of machine learning, all under the domain of AI.

iii. Uses of Medical Image Segmentation in the Medical Field

Performing medical image segmentation is a tedious task, reducing the efficiency of getting results. Giving this task to machines can help reduce labor costs and get results on demand. Medical image segmentation is considered a big field that helps radiologists to visualize body parts such as the brain, liver, and neurons. In addition, medical segmentation helps in detecting different diseases and tracking their progress [7], [8]. Under each following algorithm, you'll find an application or data set on which this model was trained, perhaps highlighting a future application for each.

III. U-Net

The U-Net model is one of the most famous models used in the medical field. This model is used for the semantic segmentation of image data. As the name suggests, this model has a u-shaped architecture, as can be seen in figure 1.

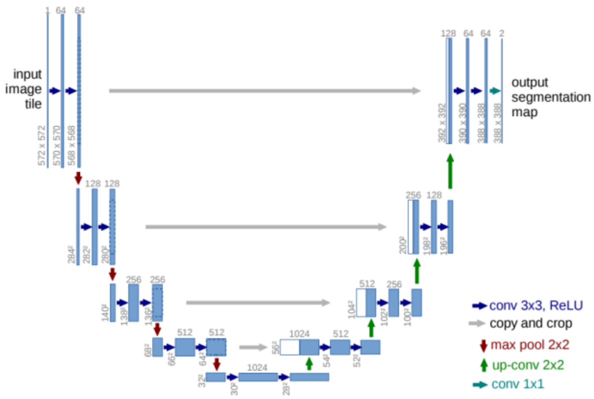

Figure 2: U-Net Architecture [1]

The model is comprised of 2 parts, the left side, and the right side. The left side is the contracting path where the image shrinks, while the right side is the expansive path, where the input to this section of the model gets larger along the rest of the model. The left side has the same architecture as that of a regular CNN. The activation function that is used throughout the model is the ReLU activation function. The last layer has a 1x1 convolution that is used to map the input getting into it to the desired number of classes [1].

i. Training

In the paper, it is mentioned that the model took 10 hours to train on a Nvidia Titan 6 GB GPU. The optimization algorithm used to train the model is SGD (Caffe implementation.) [1].

ii. Advantages

There are many convenient features the training strategy and network both have. Firstly, the training strategy relies on data augmentation, meaning that there is no need for a large abundance of data. Secondly, the architecture of the model allows it to have precise localization [1]. Moreover, the model is fast, as it takes under a second to do a segmentation of a 512x512 image [1]. Lastly, this network has various biomedical segmentation applications.

iii. Experiments

The following tables illustrate how well this type of model performed in different challenges:

1. EM Challenge

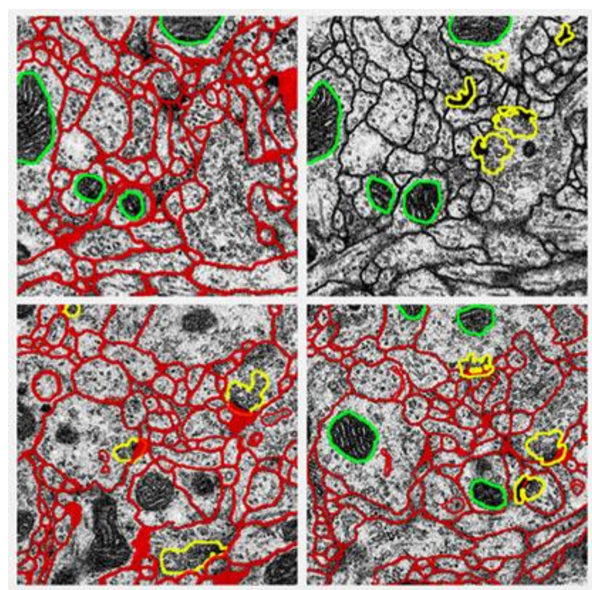

Figure 2: Segmented EM

The above image, figure 1.5 shows the output of an image segmentation model on one of the images of the EM segmentation challenge images. In this challenge, image segmentation models should be able to segment neuronal structures. This is the challenge that was introduced at the ISBI 2012 conference.

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

Table 1: EM Segmentation Challenge [1]

In Table 1, it can be seen that U-Net has the highest ranking among all its other competitors. U-Net can be seen to outperform all the models when the warping error is used as a metric, but it comes second when comparing using RE as a metric, as well as the pixel error.

2. ISBI Cell Tracking Challenge

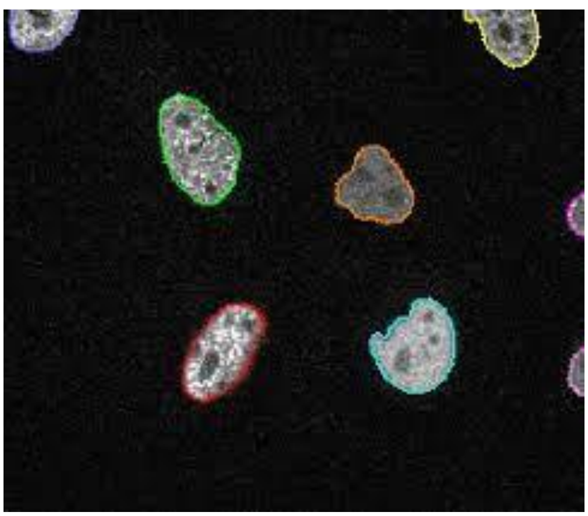

Figure 3: Cell Segmentation

This challenge was first introduced in 2013. Later versions were hosted in ISBI 2014, Beijing, and ISBI 2015, New York. The main application of this segmentation task is to track moving cells in real-time. This task can aid in observing cell behavior and cell lineages over a given time. This technique is required for many scientific and industrial applications.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

Table 2: ISBI Cell Tracking Challenge [1]

In Table 2, U-Net performs the best on both PhC-U373 and DIC-HeLa, having a score of 92.03% and 77.56% respectively.

IV. V-Net

Unlike all of the techniques discussed in this paper, this model can do segmentation on 3D images. This is very convenient as most medical data used in clinical practice consists of 3D volumes [5]. Figure 2 is from the original paper, showing the architecture of the V-Net model.

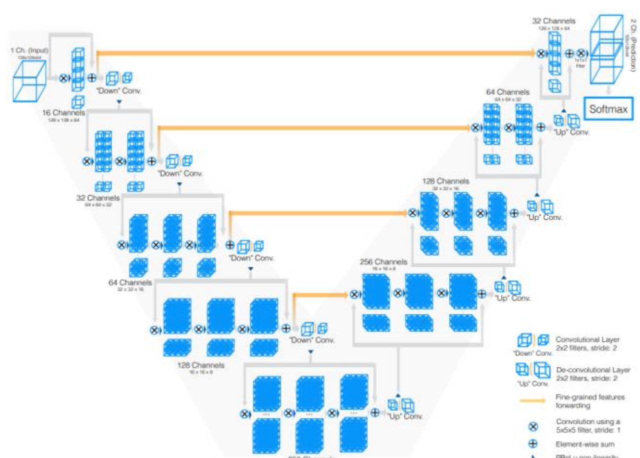

Figure 4: V-Net Architecture [5, p.]

i. Architecture

Essentially, the left part of the model is the compression path, while the right part decompresses the image to produce an output of the same size as the input. At each stage of the left part, there is residual learning. The input of each stage goes through 1-3 convolutional layers, then processed through the non-linearities coming up. The volumetric kernels used in the convolutions are of the size: 5x5x5. Through other stages though, 2x2x2 voxels applied with a stride of 2 are used. The activation functions used throughout the model are PReLU activation functions, which are a new version of the regular ReLU.

On the right side, after each stage, a deconvolution operation increases the size of the input, which is key for retaining the original size of the image. These deconvolution operations are followed by the same number of convolutional layers in each stage on the left side, which involve half the number of 5x5x5 kernels in the previous layer. The right side also utilizes residual functions in the convolutional stages [5].

ii. Training

The model was trained on a prostate MRI scans dataset, with the application of data augmentation. The training set consisted of 50 MRI volumes, using a mini-batch of size 2. A value of 0.99 for the momentum was used, starting with a learning rate of 10^3, which decreases by one order of magnitude every 25,000 iterations. The hardware used was as follows: memory of 64 GB, a processor of i7-5820K, and a Nvidia GTX 1080 8 GB. Training time was 48 hours, which was around 30,000 iterations [5].

iii. Advantages

One of the features that this model stands out with is that "this architecture ensures convergence in a fraction of the time required by a similar network that does not learn residual functions." [5]

iv. Results

Table 3 shows the scores of V-Net models in comparison with other models [5].

Algorithm	Avg. Dice	Avg. Hausdorff distance	Score on challenge task
V-Net + Dice-based loss	0.869 ± 0.033	5.71 ± 1.20 mm	82.39
V-Net + mult. logistic loss	0.739 ± 0.088	10.55 ± 5.38 mm	63.30
Imorphics [18]	0.879 ± 0.044	5.935 ± 2.14 mm	84.36
ScrAutoProstate	0.874 ± 0.036	5.58 ± 1.49 mm	83.49
SBIA	0.835 ± 0.055	7.73 ± 2.68 mm	78.33
Grislies	0.834 ± 0.082	7.90 ± 3.82 mm	77.55

Table 3: scores of V-Net models in comparison with other models

It can be seen that one of the V-Net models is in third place, with a score of 82.39. Using a different loss had a significant effect on the performance of the model, making the score differ by 19.09.

V. FCN

Fully convolutional networks (FCN) are a type of deep neural network used mainly in semantic segmentation. FCN is used in many tasks in medical image segmentation including liver, brain MRI, and lung segmentation.

First, liver image segmentation is the task of using MRI and CT images to identify different regions of the liver. Moreover, it helps in the diagnosis of liver tumors for early treatment of diseases. Tasks of liver segmentation take a long time if it is made manually. However, the difficulty in liver segmentation also appears to be difficult for automatic image segmentation. In addition, segmenting tumors from the liver adds more difficulty to this task [9].

Another task that FCN is used for is brain segmentation. Brain segmentation involves tasks of detecting brain development and classifying brain diseases like tumors. FCN achieves high accuracy in segmenting many brain regions.

Finally, lung image segmentation help in identifying lung diseases and tumors. Lung image segmentation has been used in many chest diseases. For example, lung segmentation is used mainly in COVID-19 diagnosis and cancer detection. Figure 5 shows a sample of images of chest x-rays that were segmented by FCN.

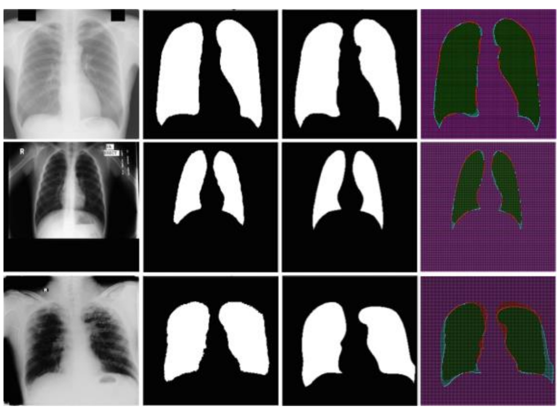

Figure 5: sample of lung segmentation results

i. Architecture

Each layer of the FCN is constructed from a 3-dimensional array with dimensions of height, width, and color channels. The structure of FCNs consists of convolution, pooling, and activation functions. Each of these components receives local input from the previous layer by equation (1)

$$y_{ij} = f_{ks}(\{X_{si+\delta i;\ sj+\delta j}\}\ 0 \le \delta i; \delta j \le k)\ (1)$$

where Xi j is the data of a layer and yi j is the output of the next one [4]. This method, with the use of convolutional layers, makes the network obtain a coarse map, which is later decoded to get the FCN output, which is of the same size as the input one [4], [10].

ii. Experiments

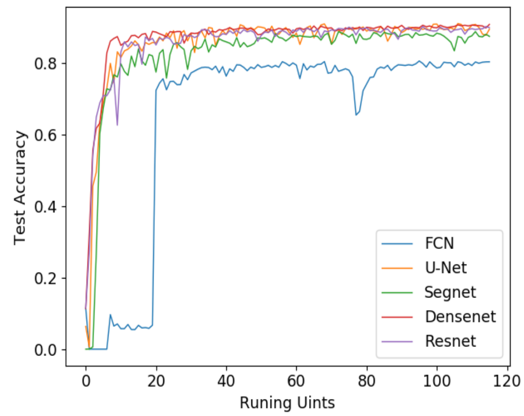

Figure 6: Comparison Between FCN and other models

According to figure 6 [11], after comparing FCN and four other models including U-Net, SegNet, ResNet, and DenseNet in liver segmentation, FCN has been found to have the least accuracy with a significant difference compared to the others. In addition, several other five "evaluation measures" were used in the comparison of the research: "Dice's similarity coefficient (DSC), Volume Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASD) and RootMean-Square Deviation (RMSD)." These results showed in table 4.

Model	DSC (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)
FCN-8s	81.23±1.20	24.30±4.79	24.25±6.01	16.04±1.40	21.01±1.62
U-Net	90.82±1.32	14.28±2.36	21.22±2.56	6.79±2.79	13.62±5.07
Segnet	89.46±0.95	18.31±1.39	52.39±1.20	9.61±1.18	16.97±2.11
Resnet	90.45±1.27	15.11±2.22	**5.56±6.06**	**5.84±0.88**	**9.42±0.81**
Densenet	**91.44±0.87**	**13.40±1.47**	11.20±4.60	6.80±0.06	12.56±0.75

Table 4: results of comparison between five models in medical image segmentation.

Despite FCN's flexible architecture inspired many modern models, according to research [11], FCN achieved the worst results in the experiment for two reasons. Those have only one deconvolutional layer for up-sampling and do not consider the relationship between pixels. The experiment ran on a computational configuration of “Ubuntu 16.04 with CPU i7 6700K, NVIDIA GTx1080ti GPU, 32G memory” [11].

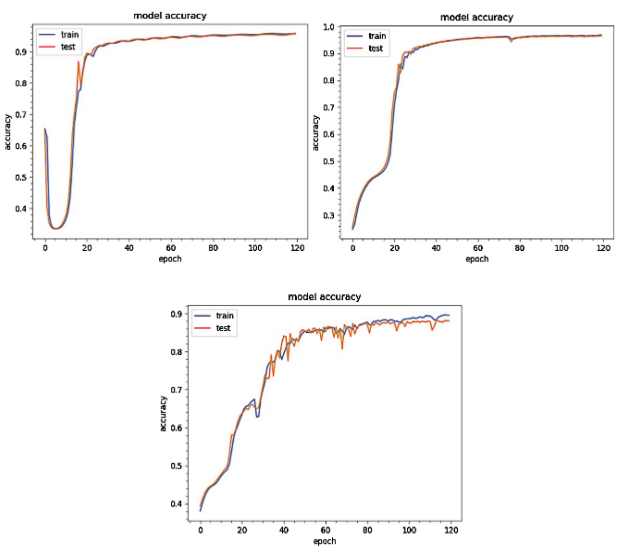

Figure 7: FCN for lung segmentation accuracy.

In another experiment made by Rabia Rashid et al. [12], they used FCN for “Lungs segmentation from chest x-rays.” The experiment was made on three datasets and achieved high accuracies of 97.1%, 97.7%, and 94.2% (figure 7). powered by two other methods for filtering images from unwanted parts, the model became able to segment the images with high accuracy. Lung segmentation, however, has few details to be segmented. In contrast, FCN model achieved low accuracy in liver image segmentation as it has more details, which makes it very sensitive. The computational configuration used in the training was “Ubuntu Kylin 14.04, GPU is Nvidia Quadro K6000, and CPU is Xeon E5-2697 v2.”

The sensitivity of FCN model can also be noticed in another experiment by Yabiao Wang et al. [7]. In this training, FCN is compared with traditional CNN in Magnetic Resonance Imaging (MRI) of the brain. As shown in figure 8, while FCN achieved the highest Dice coefficient in segmenting white and gray matter (WM and GM), it got lower in segmenting cerebrospinal fluid (CSF), which has more details than WM and GM.

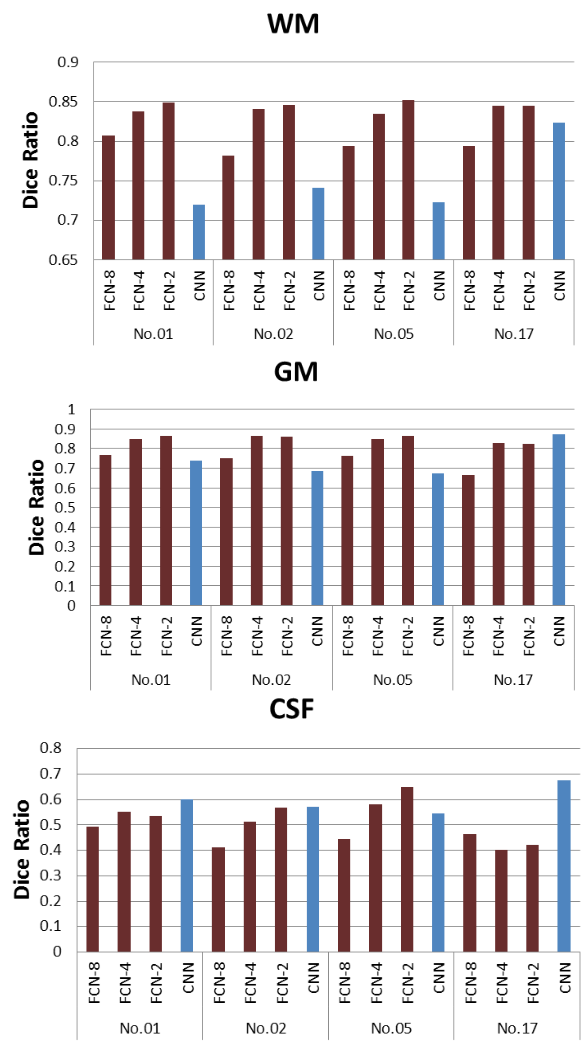

Figure 8: Comparison between FCN and CNN in brain MRI segmentation.

VI. DenseNet

The Dense Convolutional Network (DenseNet) is one of the models built on CNN architecture. DenseNet has many uses in medical image segmentation, as it handles the lack of big training datasets. Some of the uses of DenseNet are brain and liver image segmentation.

First, brain segmentation, nowadays, is used widely in surgical planning for detecting tumors from MRI brain images. While many models achieve low accuracy and are stuck in the task of brain segmentation due to the low number of training data, the complex structure of DenseNet makes it a good choice for detecting details without overfitting the data [13]. Figure 9 shows three samples of the training data with the truth segmentation and predicted output.

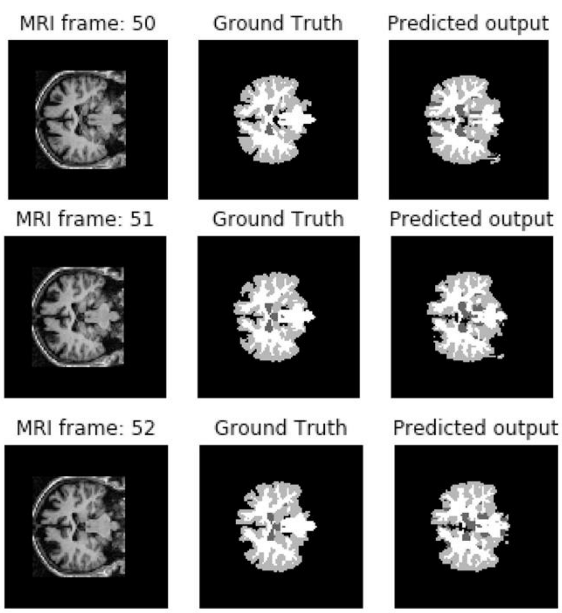

Figure 9: sample of segmented brain MRI images.

Another task that uses DenseNet architecture is liver segmentation. Liver segmentation is considered a hard task in medical image segmentation due to the specific details of the images. In addition, the details of detecting liver tumors are harder, so the complex architecture of DenseNet model achieved high accuracy in this task compared with other models. Figure 10 shows a sample of segmented liver images.

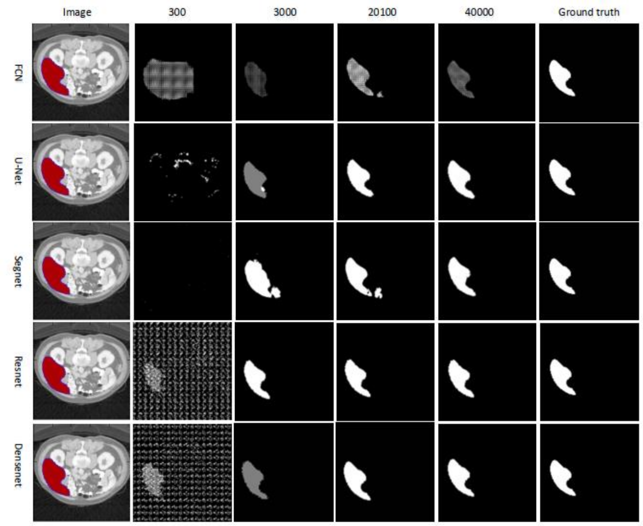

Figure 10: sample of segmented liver image segmentation.

iii. Architecture

Whilst traditional convolutional networks take the output of each layer as the input of the next one, DenseNet uses dense connectivity that takes the input of all the previous layers as the input of the current one [14]. As a result, using data from previous layers reduces the dependence of each layer on one layer. In addition, DenseNet benefits from features from all layers in its prediction. Consequently, it achieves better accuracy using less computational effort. On the other hand, some of the disadvantages of DenseNet are duplicating data while using it in each layer, causing linear growth in memory and computation by increasing the model size. Figure 11 shows an illustration of the DenseNet model with four dense blocks [15].

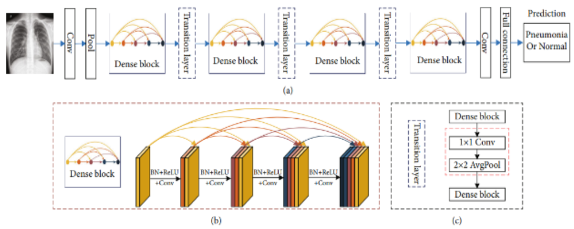

Figure 11: DenseNet Architecture [15]

iv. Experiments

As shown in figure 6 and table 4, comparing five models: DenseNet, FCN, U-Net, ResNet, and SegNet, DenseNet has shown the best results in many metrics [11]. One of the reasons that DenseNet achieved high accuracy is the high difficulty of the liver image segmentation task due to the high level of details included in the images. Dense blocks included in DenseNet model allow it to segment the image with consideration of specific details.

Layer	Loss	Accuracy
Architecture without compression	0.5972	0.9052
Architecture with compression	0.4577	0.9253

Table 5: Results of DenseNet training [6]

In addition, in another research paper about using DenseNet in brain segmentation, the results, shown in table 5, have shown an accuracy of about 92% and a dice coefficient of 0.673 [6]. The experiment was performed on a "NVIDIA Titan Xp GPU" [6].

VII. Conclusion

In the last several years, image segmentation has been used widely in the medical field. As image segmentation models give great benefits with classifying tasks that need trained eyes and a lot of time, there was a continuous development of the medical segmentation models. Although many models are used in medical image segmentation, different models achieve different accuracy and performance. Thus, it was important to do a review comparing some of the most used models. Through the four models this paper reviewed, U-Net and DenseNet have achieved high performance compared to other models. In addition, while V-Net and FCN have achieved acceptable performance, there were many better models.

Finally, as the research in medical image segmentation has various topics, it was important to mention some research gaps and model problems that were found. First, there were few research papers found about the use of the FCN model in medical image segmentation. In addition, FCN has a few deconvolution layers which can be improved: by doing more research about that model. Despite the high performance achieved by DenseNet, using it to make larger-size models will consume a significant amount of memory and increase computational time.

Another important problem while working on medical image segmentation is the scarcity of datasets. Providing larger datasets will help improve the performance of many models and achieve its goals.

VIII. References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, doi: 10.48550/ARXIV.1505.04597.

[2] H. Huang *et al.*, "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," 2020, doi: 10.48550/ARXIV.2004.08790.

[3] S. Subramaniam, K. B. Jayanthi, C. Rajasekaran, and R. Kuchelar, "Deep Learning Architectures for Medical Image Segmentation," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, MN, USA, Jul.

2020, pp. 579–584. doi: 10.1109/CBMS49503.2020.00115.

[4] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” 2014, doi: 10.48550/ARXIV.1411.4038.

[5] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” Jun. 2016, doi: 10.48550/arXiv.1606.04797.

[6] R. D. Gottapu and C. H. Dagli, “DenseNet for Anatomical Brain Segmentation,” *Procedia Comput. Sci.*, vol. 140, pp. 179–185, 2018, doi: 10.1016/j.procs.2018.10.327.

[7] A. Elnakib, G. Gimel’farb, J. S. Suri, and A. El-Baz, “Medical Image Segmentation: A Brief Survey,” in *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, A. S. El-Baz, R. Acharya U, A. F. Laine, and J. S. Suri, Eds. New York, NY: Springer New York, 2011, pp. 1–39. doi: 10.1007/978-1-4419-8204-9_1.

[8] A. Inglis, L. Cruz, D. L. Roe, H. E. Stanley, D. L. Rosene, and B. Urbanc, “Automated identification of neurons and their locations,” 2007, doi: 10.48550/ARXIV.0712.1845.

[9] H. Rahman, T. F. N. Bukht, A. Imran, J. Tariq, S. Tu, and A. Alzahrani, “A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet,” *Bioengineering*, vol. 9, no. 8, p. 368, Aug. 2022, doi: 10.3390/bioengineering9080368.

[10] Y. Wang, Z. Sun, C. Liu, W. Peng, and J. Zhang, “MRI image segmentation by fully convolutional networks,” in *2016 IEEE International Conference on Mechatronics and Automation*, Harbin, Heilongjiang, China, Aug. 2016, pp. 1697–1702. doi: 10.1109/ICMA.2016.7558819.

[11] Y. Shen *et al.*, “Empirical Comparisons of Deep Learning Networks on Liver Segmentation,” *Comput. Mater. Contin.*, vol. 62, no. 3, pp. 1233–1247, 2020, doi: 10.32604/cmc.2020.07450.

[12] R. Rashid, M. U. Akram, and T. Hassan, “Fully Convolutional Neural Network for Lungs Segmentation from Chest X-Rays,” in *Image Analysis and Recognition*, vol. 10882, A. Campilho, F. Karray, and B. ter Haar Romeny, Eds. Cham: Springer International Publishing, 2018, pp. 71–80. doi: 10.1007/978-3-319-93000-8_9.

[13] J. Stawiaski, “A Pretrained DenseNet Encoder for Brain Tumor Segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 11384, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds. Cham: Springer International Publishing, 2019, pp. 105–115. doi: 10.1007/978-3-030-11726-9_10.

[14] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” 2016, doi: 10.48550/ARXIV.1608.06993.

[15] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, “Dense Convolutional Network and Its Application in Medical Image Analysis,” *BioMed Res. Int.*, vol. 2022, pp. 1–22, Apr. 2022, doi: 10.1155/2022/2384830.