תמונה שמכילה גופן, לוגו, סמל, גרפיקה

התיאור נוצר באופן אוטומטי

Final Project Phase A

Outline of gastro-intestinal tumors based on image segmentation.

Supervisor:

Miri Weiss Cohen

Submitters:

Aviel Malayev – [aviel410aviel@gmail.com](mailto:aviel410aviel@gmail.com)

Omri Shalev – [shalevomri10@gmail.com](mailto:shalevomri10@gmail.com)

תוכן עניינים

[**ABSTRACT** 2](#_Toc137545322)

[**INTRODUCTION** 2](#_Toc137545323)

[**RELATED WORK** 5](#_Toc137545324)

[**BACKGROUND** 10](#_Toc137545325)

[**CT** 10](#_Toc137545326)

[**Image Segmentation** 11](#_Toc137545327)

[**U-net-Architectures** 12](#_Toc137545328)

[**3D-U-net** 15](#_Toc137545329)

[**NAS-U-net** 15](#_Toc137545330)

[**Dense-Inception U-net (DIUNet)** 16](#_Toc137545331)

[**TransUNet** 16](#_Toc137545332)

[**Dataset** 17](#_Toc137545333)

[**Process** 18](#_Toc137545334)

[**Challenges** 19](#_Toc137545335)

[**Product** 19](#_Toc137545336)

[**Use Case** 19](#_Toc137545337)

[**Sequence** 19](#_Toc137545338)

[**GUI** 20](#_Toc137545339)

[**Expected Achievements** 21](#_Toc137545340)

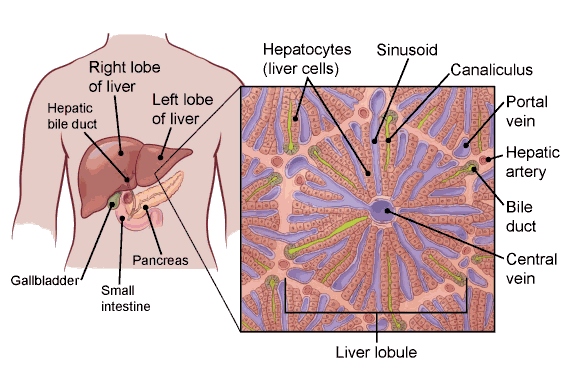
[**Evaluation / Verification Plan** 21](#_Toc137545341)

[**Bibliography** 22](#_Toc137545342)

# **ABSTRACT**

In Radiation therapy, radiation oncologists try to deliver high doses of radiation using X-ray beams pointed at liver tumors while avoiding surrounding tissues. Technology such as CT scans helps oncologists to visualize the daily position of the liver, which can vary from day to day due to organ motion. In these scans, radiation oncologists must manually outline the position of the liver and the tumors in order to adjust the direction of the X-ray beams to increase the dose delivery to the tumors and avoid the surrounding tissues. This is a time-consuming and labor-intensive process that can prolong treatments from 15 minutes a day to an hour a day, which can be difficult for patients to tolerate. Using Deep Learning to automate the segmentation process would make the treatment much faster and hopefully, a more accurate outline of the liver tumors, which will help patients to get better treatment.

# **INTRODUCTION**

The liver is the largest internal organ (the size of a football). [**[1]**](#_Bibliography) It lies under the right ribs just beneath the right lung and it has two lobes (sections). [**[2]**](#_References)

***Fig. 1. Liver anatomy***

The liver is a vital organ that performs various important functions in the body, including filtering toxins from the blood, producing bile to aid in digestion, and storing energy in the form of glycogen. The liver is made up mainly of cells called hepatocytes [**[3]**](#_References). It also has other types of cells, including cells that line its blood vessels and cells that line small tubes in the liver called bile ducts. The bile ducts carry bile from the liver to the gallbladder or directly to the intestines.

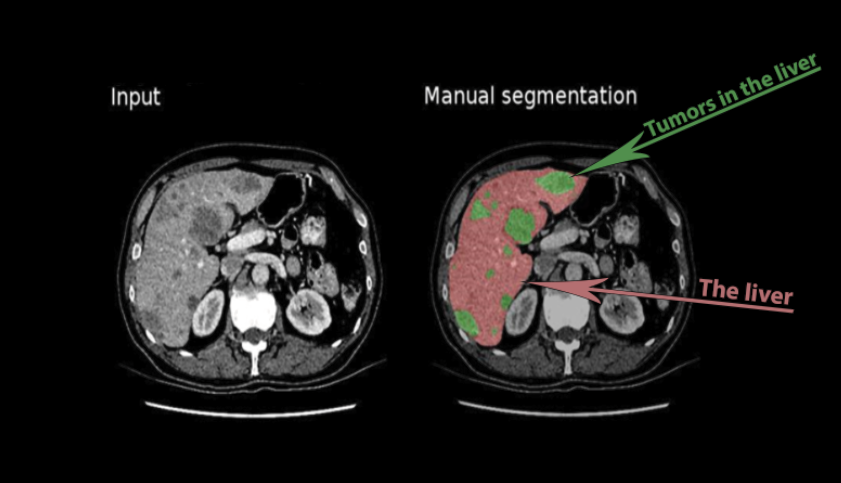
Liver cancer, also known as hepatic cancer, is a type of cancer that starts in the cells of the liver. Liver cancer can develop when healthy liver cells undergo mutations in their DNA, which causes the cells to grow and divide uncontrollably. Over time, these abnormal cells can form a tumor in the liver, which can interfere with its normal function and spread to other parts of the body.

There are several types of liver cancer, but the most common type is hepatocellular carcinoma, which accounts for about 75% of all liver cancer cases. Other less common types of liver cancer include cholangiocarcinoma, angiosarcoma, and hepatoblastoma.

Cancer that spreads to the liver is more common than cancer that begins in the liver cells. Cancer that begins in another area of the body such as the colon, lung, or breast, and then spreads to the liver is called metastatic cancer rather than liver cancer. This type of cancer is named after the organ in which it began, such as metastatic colon cancer to describe cancer that begins in the colon and spreads to the liver. The symptoms of liver cancer may include abdominal pain, jaundice, fatigue, weight loss, and loss of appetite.

Liver cancer is the seventh most commonly diagnosed cancer and the second leading cause of cancer death worldwide in 2020 [**[4]**](#_References). The main risk factors for hepatocellular carcinoma (HCC), which accounts for approximately 75%-85% of cases of liver cancer, are chronic infection with hepatitis B or hepatitis C viruses, heavy alcohol consumption, excess body weight, type 2 diabetes, and smoking.

Treatment options for liver cancer may include surgery, chemotherapy, radiation therapy, targeted therapy, and immunotherapy, depending on the stage and severity of the cancer. Radiation therapy is an important component of cancer treatment with approximately 50% of all cancer patients receiving radiation therapy during their course of illness. There are several radiation therapy techniques such as Image-guided radiotherapy (IGRT) [**[5]**](#_References). When treatment margins become tighter, the potential to miss tumors due to organ motion and patient setup variations becomes greater. When critical structures are close to the tumor, a slight positional error may also lead to inadvertent radiation of the normal organs. After radiation therapy has been ordered, a planning stage occurs. The patient will first undergo a simulation scan on a special CT scanner. The radiation oncologist then manually outlines the area to be treated, the tumors, and the areas to be avoided (such as normal organs).

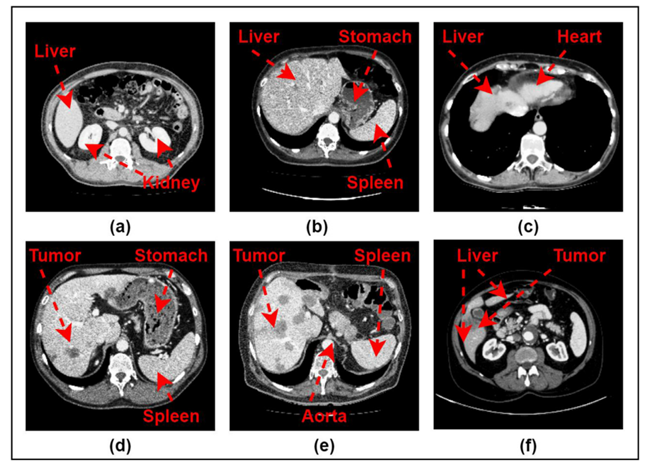


***Fig. 2. CT scans and manual segmentation***

Then the radiation plan is developed and checked by the medical team for quality and safety purposes. The procedure of outlining the relevant areas is tedious, time-consuming and may cause accuracy errors due to human limitations, opinions, knowledge, and experience.

By automating this procedure, we can prevent human errors such as missing small tumor areas, save precious time both for the physician and patients and hopefully get more accurate outlines to better target the tumor cells with radiation. Convolution neural networks (CNNs) have a promising future in this medical field. In recent years works have been conducted to automate CT and MRI scans outlining by segmenting the images with a variety of techniques and machine learning architectures, which we will further discuss later in this paper. Machine-learning implementation in medical fields is a relatively new field, while the machine-learning field itself is constantly improving with more advanced segmentation algorithms. Therefore, recent works that have been conducted in liver tumor segmentation may already be irrelevant.

Another reason is that the dice result in liver tumor segmentation, which is the most commonly used metric to evaluate the performance of image segmentation algorithms, is still a challenging problem since it suffers from a low score due to low contrast, vague boundaries between the tumors and surrounding tissues as well as the wide variations of the tumors in intensity, shape, and location across patients. as shown in fig. 3.



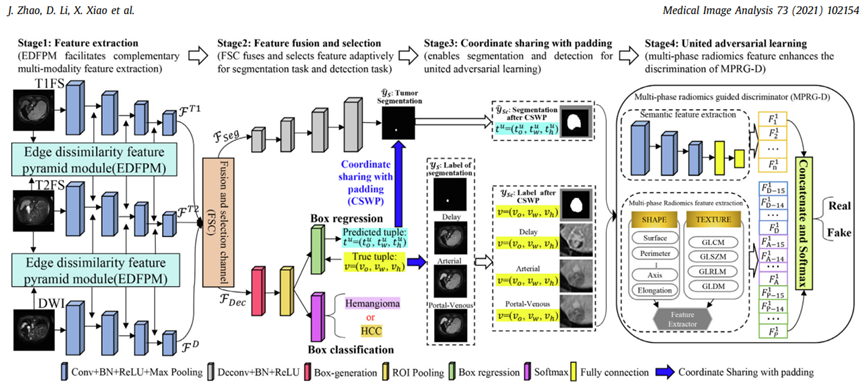
***Fig. 3. CT slices from (a) to (f) indicates different challenges in liver segmentation: (a) represents the abdominal organs surrounded by the liver, (b) intensity difference with other organs, (c) represents fuzzy boundaries and overlapping with other organs, (d), and (e) represents intensity difference between liver and tumor region and, complex liver shape, and (f) represents a discontinuity in the liver boundaries.***

# **RELATED WORK**

Jianfeng Zhao et al.[**[6]**](#_References), reported that there are 2 approaches for liver tumor segmentation and detection today. First, the most common approach is the use of CEMRI, where segmentation is done manually by physicians, which is heavy work associated with potential misjudgments. More importantly, patients are injected with gadolinium contrast agents during CEMRI, which is time-consuming, risky, expensive, and restricted for those with impaired kidney function.

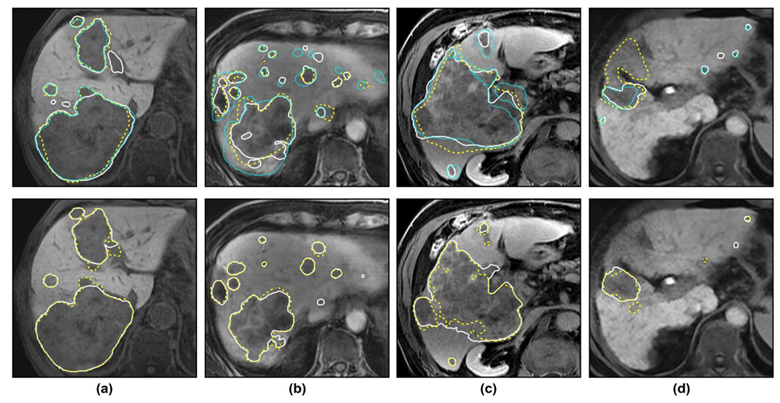
Second, to avoid the injection of gadolinium, work has been done attempting to segment or detect liver tumors using NCMRI with only one modality. However, the authors claim that these works have not taken into account that the information is insufficient when using a single modality. Therefore, these works are not very practical when the tumor is very small. In addition, these works are limited to separate segmentation or detection of liver tumors, which means that there is no work that does simultaneous segmentation and detection that clinics can use. Their solution is to use a United Adversarial Learning Framework (UAL) system for simultaneous liver tumor segmentation and detection with a multi-modal NCMRI consisting of T1, T2, and DWI modalities. Their main assumption is that integrating information from multimodal MRI can improve liver tumor segmentation and detection performance.

UAL first uses three parallel fusion channels for each modality in multi-modal NCMRI information extraction. Then, fusion and selection channels (FSC) are developed to make the final decision of feature fusion and selection. Then, there is the mechanism of Coordinate Sharing with Padding (CSWP) to integrate the segmentation task and the detection task, which enables multi-tasking to perform UAL. Finally, an innovative multiphase radiomics discriminator (MPRG-D) uses unique and specific tumor information to improve multi-tasking performance via the adversarial learning strategy.



***Fig. 4. Overview of the proposed UAL. It segments and detects liver tumors simultaneously and contains four stages. 1) Feature extraction via using three parallel convolution channels. 2) Feature fusion and selection via using an innovative FSC. 3) The operation of coordinate sharing via using the proposed CSWP. 4) UAL via using a newly designed MPRG-D.***

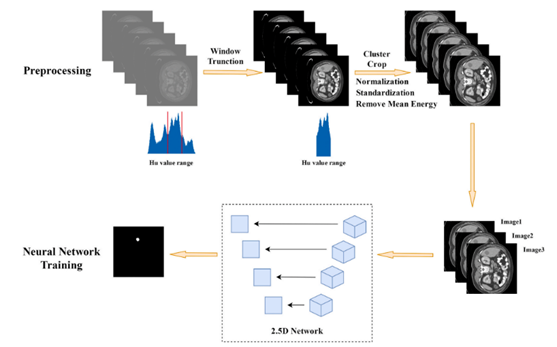
Annika Hänsch et al. [**[7]**](#_References)**,** Claim that liver segmentation based on deep learning methods combined with manual corrections can reduce the average interaction time required by 80%, compared to fully manual segmentation, with significantly lower inter-observer variability. According to the authors, most studies on the automatic segmentation of liver tumors have focused on CT. They are proposing a new approach for liver tumor segmentation in the late Hepatocellular Carcinoma (HCC) phase of DCE-MRI, using an anisotropic 3D U-Net architecture and a multi-model training strategy. From a clinical perspective, CE-MRI shows significantly higher sensitivity, in the diagnosis of HCC in patients with cirrhosis, compared to CT. The high accuracy of MRI in detecting and characterizing liver disease is mainly due to the higher soft tissue contrast compared to CT. MRI can use various contrast mechanisms to enhance the differentiation of various tissues, such as T1, T2, and DWI sequences. In addition, hepatocyte-specific contrast agents can be used to selectively stain healthy liver cells while leaving malignant tissue unstained and darkened. This increases the precision of volumetry of liver malignancies and allows more accurate dose planning for selective internal radiation therapy.

In the absence of public test data for MRI tumor segmentation, the automatic segmentation results are compared with the ground truth generated by three human raters, and uncertainty-based evaluation scores are used to account for inter-raters agreement. They propose to leverage the observation that the same model can produce different results when trained multiple times on the same data, by using a multi-model training strategy that starts with training sixteen models and narrowing them down to the best model using the Hyperband method (optimization algorithm - to find the most promising configurations of hyperparameters). The final model, which is selected based on validation data, significantly outperforms a single-trained model in an independent test set and achieves a higher average qualitative score.

***Fig. 5. Exemplary test cases with manual and automatic liver tumor segmentations. In the top row, white solid / yellow dashed / cyan finely dotted contours correspond to raters R1 / R2 / R3 (humans). In the bottom row, white solid / yellow dashed contours correspond to algorithms A1 / A2. Based on the average performance of A1 against all raters D (A1, Ri), these cases represent the (a) upper, (b, c) median, and (d) lower end of the Dice score range, not including outlier case 4****.*

Nowadays two main categories in U-nets exist: First, a 2D network based on U-Net is trained by slices, fully utilizing the two-dimensional spatial information in each slice, but ignoring the information between slices, resulting in a decrease in segmentation accuracy. Second, a 3D network based on V-Net utilizes intra-layer and inter-layer information to extract features through 3D convolution, learning 3D contextual information. However, the high memory and computational overhead in 3D networks is a major challenge for network training.

Chi Zhang et al[**[8]**](#_References), proposed a new 2.5D segmentation network, called UV-Net, which simultaneously integrates the 2D(U-Net) design and 3D(V-Net) design. A 3D encoder is used to capture the 3D spatial context while a 2D decoder is used to maintain the high in-plane resolution, this solution was proposed due to the limitations of 2D and 3D networks. The improved 2.5D network uses continuous multi-layer slices as inputs and only outputs one label prediction, using inter-layer information between adjacent slices for targeted prediction. A UV-Net-multi-scale based on UV-Net, which fuses the multi-scale features extraction into a network, is proposed. The result after realizing multi-scale convolutional is the extraction of features at multiple scales, which not only reduces the information redundancy but also strengthens the independence of features and realizes the sparseness of feature dimensions, which finally improves the adaptability of the network and accelerates the convergence. The working process of their method starts with sending CT images from the dataset to a pipeline for preprocessing, with the general methods: window truncation, cluster crop, standardization, and normalization, and finally being processed by the method of removing mean energy.

The inconsistency of data distribution has been effectively suppressed through the preprocessing pipeline, and then three continuous slices Image1, Image2, and Image3 with 400 × 400 spatial resolutions will be fed into the 2.5D network for forward propagation. Finally, the output of the 2.5D neural network is received as the prediction label of the middle slice.

***Fig. 6. The authors propose a method that combines the following: Window truncation, Normalization, Standardization, Removing mean energy, and Cluster crop functions. In the picture above there is an overview of the working process. The whole pipeline consists of two parts, the preprocessing module, and the neural network module.***

The authors propose a method that combines the following: Window Truncation, Normalization, Standardization, Removing Mean Energy, and Cluster Crop functions as we can see in the picture above.

Zhen Yang et al. [**[9]**](#_References)**,** proposed a methodology for semi-auto liver tumor segmentation, composed mainly via two commonly used image segmentation techniques. The first technique is Region of Interest (ROI) Using Adaptive Region Growing (ARG). This technique requires manually selecting the seed(base) points for each ROI (in this case - the tumors) and predefining the threshold that determines whether the neighboring pixels should be added or not to the segmented region. The second proposed technique is Graph Cuts. The purpose of graph cuts segmentation is to extract the tumor from the ROI accurately with enhanced information. Using this technique, each pixel in the image receives the probability to be a “source” or “background” (i.e., tumor or not) depending on the energy function. By minimizing the energy function, tumor regions can be extracted easily by the min-cut/max-flow algorithm.

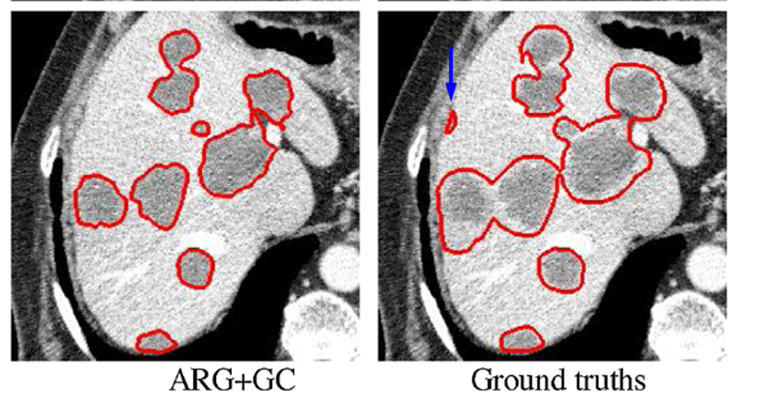
This method obtains great Dice results relative to the other solutions, though the proposed method is not fully automated which is again bound to misjudgment when selecting the seed points of interest.

תמונה שמכילה טקסט, גופן, מספר, קו

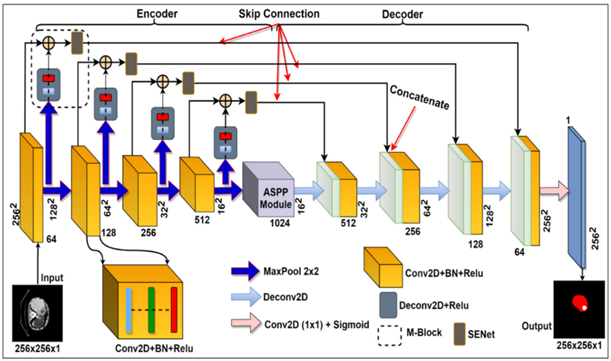
התיאור נוצר באופן אוטומטי

***Fig. 7. Comparison between multiple segmentation methods, which shows us that this method has better results than others.***

As we can see in the figure below, if the region which is marked in the blue arrow is not selected with a seed point of interest by the physician, then this region won't be segmented.



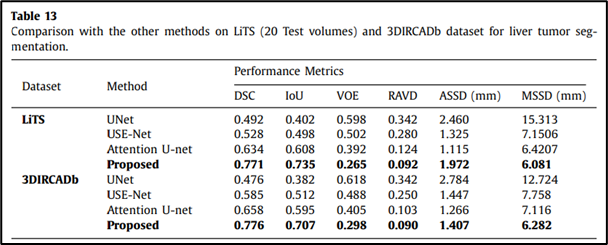
***Fig. 8. Segmentation results for ARG+GC and Ground truths.***

Devidas T. Kushnure et al.[**[10]**](#_References)**,** proposed to automate the segmentation procedure using a modified U-Net - HFRU-Net.

***Fig. 9. Proposed HFRU-Net architecture.***

There are two main modifications that the authors proposed: First, Modifying the skip pathways by using a Squeeze-and-Excitation Network (SENet), which has been proven as an improvement to the accuracy of the model at almost no computational cost. [**[11]**](#_References)SENet helps to adaptively recalibrate features by adding a content-aware mechanism to weigh each feature map (modeling the interdependencies between feature maps). Second, Modifying the bridge layer of the U-Net to atrous spatial pyramid pooling (ASPP) model to represent more contextual information on the up-sampling stages for better segmentation performance.

This method achieved better Dice results relative to the other solutions as shown in the table below:

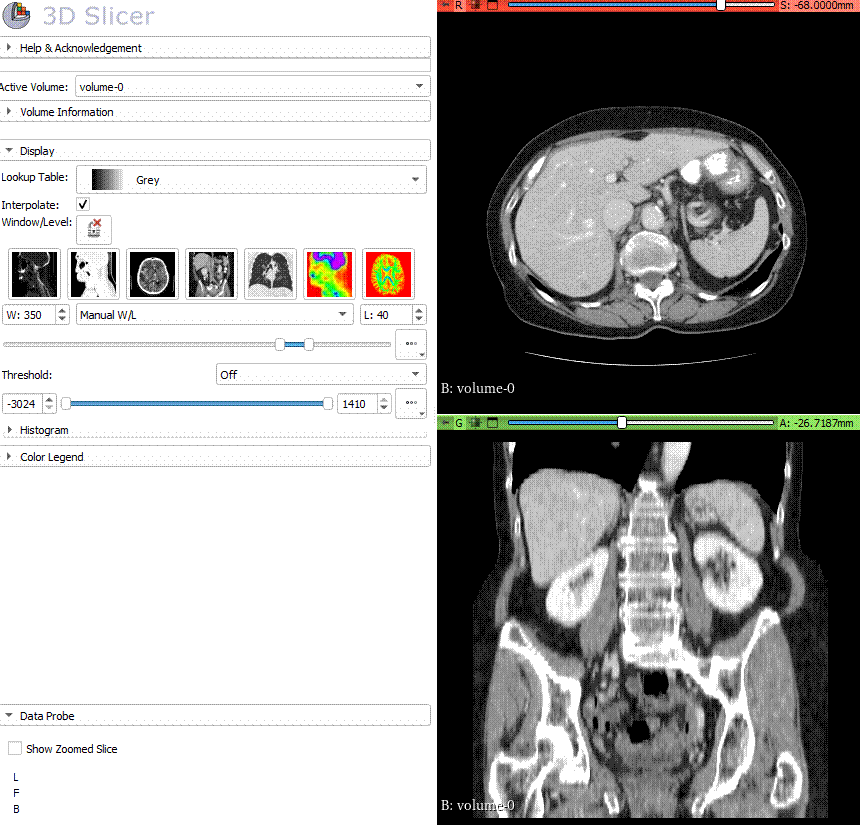


***Fig. 10. Comparison between multiple segmentation methods that work on two different data sets.***

# **BACKGROUND**

## **CT**

CT (Computed tomography) is a computerized X-ray imaging procedure that produces detailed cross-sectional images of the body, also known as "slices" [**[12]**](#_References). The procedure involves a narrow beam of X-rays being quickly rotated around the patient, while the signals it produces are processed by the machine's computer. This provides a clinician with more information than a conventional X-ray. Once a few slices have been taken, these can be used to create a 3D image of the region of interest, providing the clinician with even greater detail. CT scans can be used to diagnose a range of conditions, from broken bones to tumors.

CT scanners use a motorized X-ray source that rotates around a donut-shaped structure called a gantry. During a CT scan, the patient lies on a bed that slowly moves through the gantry while the X-ray tube rotates around the patient, shooting narrow beams of X-rays through the body. CT scanners use special digital X-ray detectors, which are located directly opposite the X-ray source. As the X-rays leave the patient, they are picked up by the detectors and transmitted to a computer. Each time the X-ray source completes one full rotation, the CT computer uses mathematical techniques to construct a 2D image slice of the patient. When a full slice is completed, the image is stored, and the motorized bed is moved forward incrementally into the gantry. The X-ray scanning process is then repeated to produce another image slice, and this process continues until the desired number of slices is collected. Image slices can either be displayed individually or stacked together by the computer to generate a 3D image of the patient that shows the skeleton, organs, and tissues as well as any abnormalities the physician is trying to identify, allowing for more accurate diagnosis and treatment plans.

***Fig. 11. Collection of 2D scans combined to represent a 3D image of the patient’s region of interest, where the volume bars in green and red are controlling which slice is presented***

CT scans can be used to measure the size and shape of organs, tumors, and other lesions. It has become a useful tool to diagnose many conditions, including cancer, heart disease, stroke, lung disease, liver tumors, and more. There are some soft tissues in the body that are very difficult to see through X-rays due to their ability to stop X-rays. Therefore, contrast agents have been developed which are highly visible in an X-ray or CT scan and are safe to use in patients. Contrast agents contain substances that can stop X-rays and are therefore more visible on an X-ray image. For example, Oral contrast agents, such as barium-based compounds, are used for imaging the digestive system, including the esophagus, stomach, and gastrointestinal (GI) tract.

## **Image Segmentation**

Image segmentation is the process of separating an image into different regions (segments). Objects in an image can have different shapes, sizes, colors, and textures, and they can be partially or fully occluded by other objects or by the image boundary itself. The Image can be noisy, ambiguous, and may contain a large number of objects or complex scenes with overlapping objects and complex backgrounds. As separating an object from the background in a given image is often a simple and intuitive task to perform by the human brain, it is a very challenging task for a computer.

In computer vision, there are several segmentation algorithms such as Thresholding, which is the most simple and basic segmentation technique where the image’s pixels are divided into two segments based on the threshold value.[**[13]**](#_References)

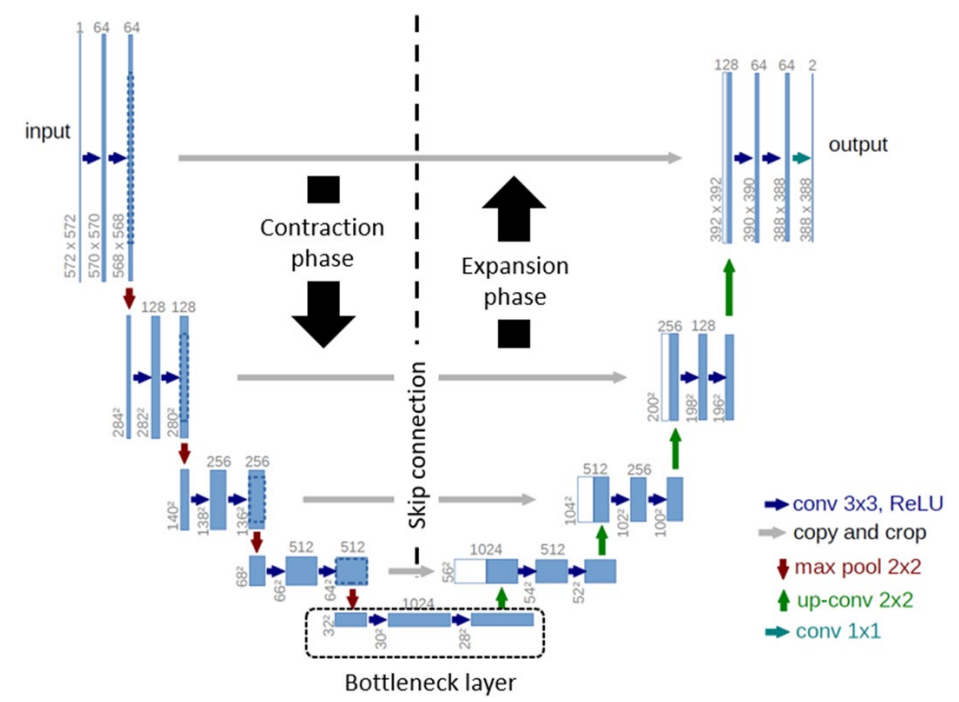
Another technique is Adaptive Region Growing (ARG), this algorithm starts with a seed point for each ROI and grows a region around each seed by adding adjacent pixels that meet certain criteria such as color, texture similarity, etc.[**[9]**](#_References)

In addition, there is the Graph cut, which is a very common method for image segmentation that works by dividing an image into regions or segments based on a graph representation.[**[9]**](#_References) In the graph, each pixel in the image is represented by a node, and the edges between the nodes, which are assigned weights that are calculated using many metrics such as color, distance, and more, represent the similarity between the pixels.

There is also a Convolutional neural network (CNN), though the typical use of a convolutional network is for classification tasks that output whether an object is in the image or not, regardless of its location in the image. [**[14]**](#_References) CNN-based segmentation models typically consist of encoder and decoder parts that extract features and produce a final segmentation map. This map separates the image into the desired object pixels and non-object pixels by classifying each pixel separately to the corresponding class. In that way, the desired output includes localization, which creates segmentation. We can see in Fig.12 the convolutional process.

# **U-net-Architectures**

U-Net is a CNN-based segmentation model originally developed for medical images.[**[15]**](#_References) It is a popular deep-learning architecture for semantic segmentation. As segmentation is basically a classification task where every pixel is classified as being part of the target region or background, the U-net model was proposed by Ronneberger et al. (2015) to distinguish every pixel, where input is encoded and decoded to produce output with the same resolution as input. The symmetrical arrangement of encoder-decoder blocks efficiently extracts and concatenates multi-scale feature maps, where encoded features are propagated to decoder blocks via skip connections and a bottleneck layer as we see in Fig.12.

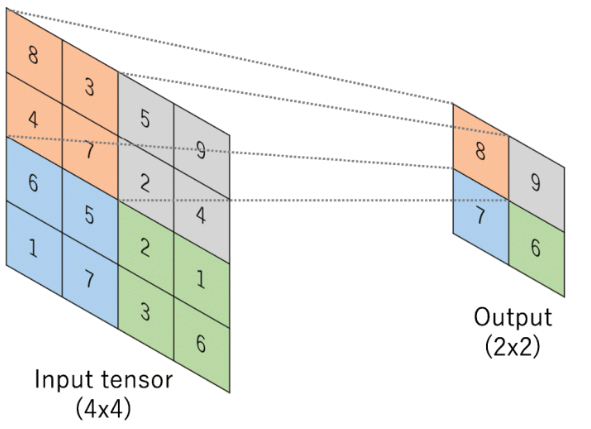


***Fig. 12. The left side is considered as the Encoder as the resolution of the image is compressing and the right side is the Decoder as the resolution is Expanding through each layer back to the original resolution. The extracted feature maps are propagated to the decoder block via a bottleneck layer that uses cascaded convolution layers***

Encoder for semantic segmentation is used to extract the features from an image, which represents the most important information in it.[**[16]**](#_References) This is done by using the convolutional layers and pooling layers. The convolution layers are composed of filters that go through each pixel in the image, those filters are learned through the process of training the model, then using a predefined activation function and the pooling layers to reduce the dimensionality of the output. By combining multiple convolutional and pooling layers, we can extract better information. The information is from small details like edges and colors to important features like ears, teeth, and eyes. The network will learn what features are important for classification and extract these to create a compact representation of an image. The encoder part is found in all the CNN architectures as Its job is to create a compact representation of the input image. One problem is this compact representation does not include the location of the features in an image. As mentioned before, this is fine for image classification. For example, to classify a dog, we only need to know if a tail, ear, or fur is in the image. It does not matter where in the image these features come from. In comparison, for segmentation location is important.

***Fig. 13. The convolution process. A filter is applied to the image and the result of calculating a dot product between the filter and the pixels is fed into an output result. Then, the filter moves to the next pixel area and repeats the process all over the image until we get the final input, which is a feature map.***תמונה שמכילה תרשים, טקסט, קו, מקביל

התיאור נוצר באופן אוטומטי

******

***Fig. 14. The pooling process. In U-net, the pooling function is typically max pooling, to reduce the resolution of the input without losing the most significant value from the region.***

The decoder, like the encoder, has convolutional layers. Except that instead of the pooling layers, we have deconvolution layers that increase the dimensionality of the image. Deconvolution is the same as convolution except that the input is padded with pixels to increase the resolution, thus the output size can be equal to or greater than the input size.

The goal of the decoder is to reconstruct the image back to the original resolution from the given compressed representation of the image as accurately as possible.

The problem is that the location of the features in the image is lost in the process of encoding the image. to address this problem enters the skip connection components which make the U-net successful in segmentation tasks.

Skip Connections in U-net are used to keep information from earlier convolutional layers and send them to the deconvolutional layers. Mainly, this information is the location of the important features which are extracted by the convolutional layers to tell the network where every feature belongs in the original image. This process is done by concatenating the last layer in the convolutional layers and the first layer of the opposite deconvolutional layers. Because the U-net is symmetrical, the dimensions of the layers are the same, which makes the process of concatenation easier. This process of concatenation is the most important phase of the U-net, which combines two important details. The first is Features extracted from the last layer by upsampling, and the second is Feature localization, which is the feature map from the opposite convolution layer. As this feature map is already in the needed size of the upsampled layer, it helps to localize the features correctly.

After U-net was introduced, U-net variants developed quickly to further enhance medical image segmentation, where each variant is designed to address different segmentation problems, in this paper we will discuss several variants that are most relevant for liver tumor segmentation via CT scans task.

## **3D-U-net**

In the biomedical image analysis field, images are often composed of 2D slices to represent a full 3D picture of the inspected organ or region of interest, thus, it is important to also extract contextual information from the images, which the traditional 2D-U-net lacks. Due to low efficiency and the loss of contextual information in 2D-U-net, Ozgun Cicek et al[**[17]**](#_References) proposed a 3D-U-net which is based on the original model, the 3D-U-net architecture is similar to the 2D-U-net architecture, except that 3D-U-net takes a 3D image as input and process it with corresponding 3D convolution, pooling, and up-sampling operations. The main problem with this suggested solution is that it requires a large number of parameters and therefore makes the network hard to train.

## **NAS-U-net**

Neural architecture search (NAS) is a research area that aims to automate the design process of neural network architectures.[**[18]**](#_References)It uses machine learning algorithms to search the optimal architecture for a specific given task by trying to find the most effective and efficient architecture without human intervention. NAS is composed of search space, search strategy, and performance estimation strategy which are the three major components. Search space is a set of predefined operations e.g. convolutional layers, pooling, fully connected, etc. Search strategy denotes the methodology used to search for the optimal architecture in the search space (i.e. Random search, Reinforcement learning, Evolutionary algorithms, Sequential model-based optimization, Gradient optimization). Performance estimation strategy denotes the accuracy of a model architecture trained over a referenced dataset.

The NAS-U-net sets a U-like architecture as a basic network and at the same time searches for an optimal two-cell-based architecture that is based on down-sampling cells and up-sampling cells.[**[19]**](#_References)

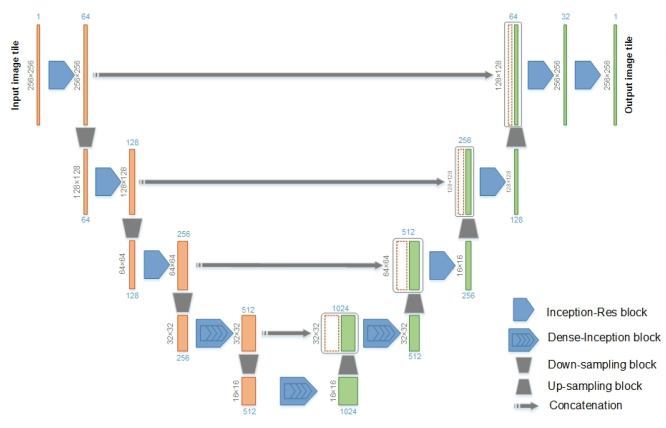
תמונה שמכילה תרשים, צילום מסך, תוכנית, טקסט

התיאור נוצר באופן אוטומטי

***Fig. 15. The U-like backbone of Nas-Unet architecture, the rectangle represents cell architectures that need to be searched. Notice that both DownSC and UpSC contain their down or up operation in NasUnet, and the green arrow merely represents the flow of the feature map (input image). The gray arrow that belongs to UpSC is a transform operation and is also automatically searched.***

## **Dense-Inception U-net (DIUNet)**

Dense-Inception U-net is a variant of a U-net that is based on the GoogLeNet Inception module and DenseNet model. [**[20]**](#_References) This architecture-based network combines the inception module and the densely connected convolutions model into a U-net architecture.



***Fig. 16. DIUNet architecture illustration.***

As with the U-net, the analysis and synthesis paths transmit feature maps directly from the down-sampling to the to-sampling process by the skip connections. In addition, based on GoogLeNet architecture (proposed the concept of an “Inception module” to build a sparse, high-performance computing network architecture), every convolutional layer, except from the three middle layers of the U-net, is replaced by an Inception-Res module with different sizes of convolutional kernels to make the network wider, where in the three middle layers of the U-net the Inception-Res blocks are replaced by Dense-inception blocks.

## **TransUNet**

TransUNet is a model which combines both Transformers and U-Net, as a strong alternative for medical image segmentation.[**[21]**](#_References)In the encoder Phase, a CNN-Transformer hybrid model is applied to leverage both the Transformers and the Unet approaches to build a network that can extract detailed high-resolution spatial information from CNN and also to better learn the global context information that a normal U-net lacks via a transformer, due to the self-attention mechanism that can distinguish between the importance and unimportance of the tokens in a sequence and thus, to capture and extract contextual relationship. This hybrid model was proposed simply due to trial and error where using only a transformer as the encoder gave poor results.

תמונה שמכילה טקסט, צילום מסך, תרשים, מפה

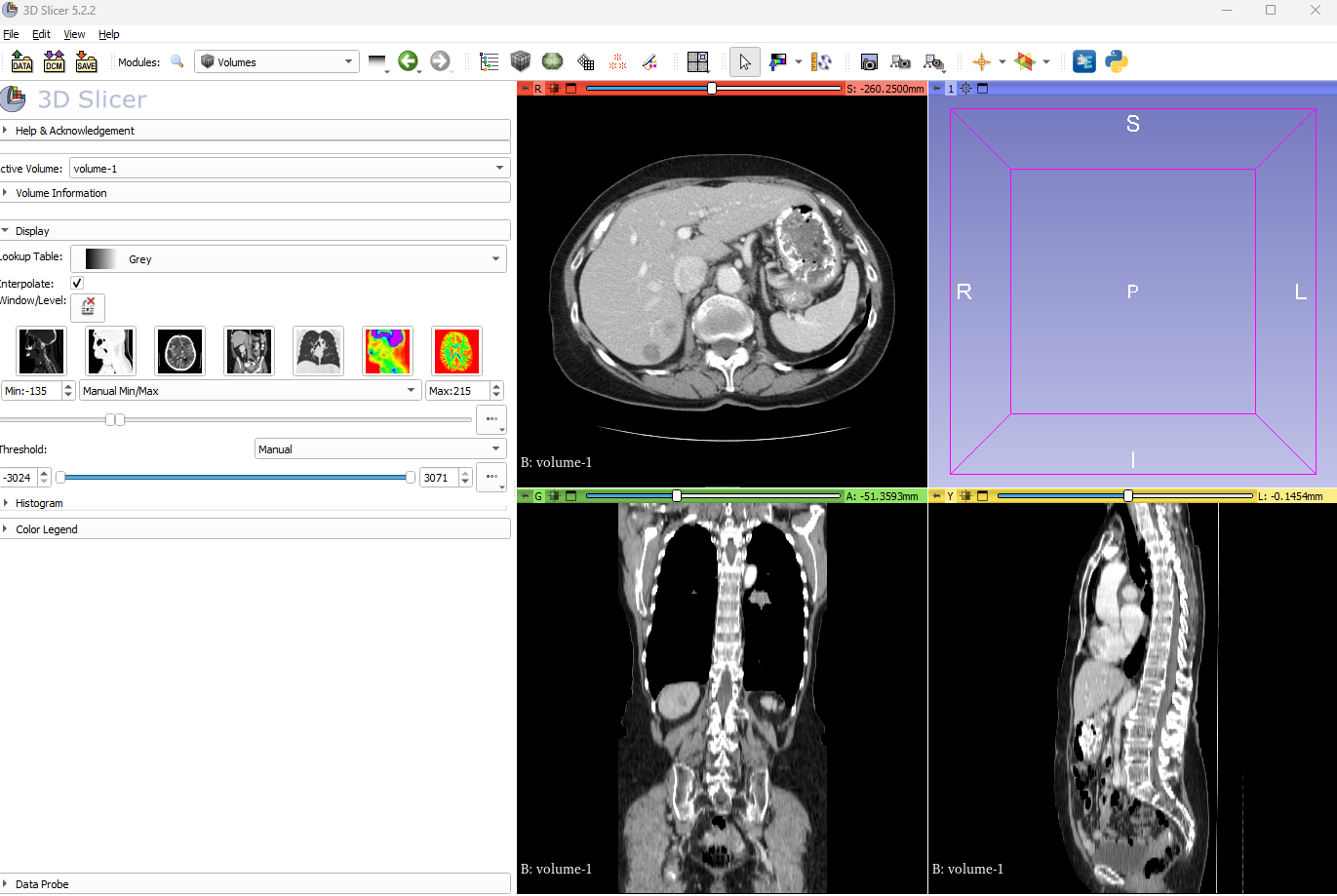
התיאור נוצר באופן אוטומטי

***Fig. 17. TransUNet architecture illustration, in the encoder, a CNN is first applied both for***

***the purpose of skip connection to the decoder phase and as input for the transformer layer.***

# **Dataset**

For our study, we will utilize the LiTS17 dataset, obtained from Kaggle[**[22]**](#_References), which consists of 130 3D CT scans of patients. These scans are stored as NIfTI-1 files (.nii). To view and analyze the images, we will use the "Slicer" program. This dataset provides us with a comprehensive collection of 3D CT scans, allowing us to effectively train and test our architectures. The "Slicer" program facilitates manual analysis and interpretation of the images. This combination of dataset and visualization tool enables us to conduct thorough evaluations and make informed decisions in our research.



***Figure 18: Visualization of a 3D CT scan from the LiTS17 dataset using the "Slicer" program.***

# **Process**

In our research, we will conduct a comparative study of three prominent architectures: TransUNet, Dense-Inception U-net, and Nas-U-net. Our objective is to identify the most effective architecture for our purposes. Subsequently, we will explore methods to enhance the chosen architecture by implementing structural improvements and modifying hyperparameters, specifically focusing on batch size, learning rate, and epoch size. Additionally, we will investigate the impact of transfer learning on the identified best architecture.

By undertaking this comprehensive analysis, we aim to not only identify the optimal architecture but also refine it further to achieve superior performance. The integration of structural improvements and fine-tuning of key hyperparameters will enable us to enhance the overall efficiency and effectiveness of the selected architecture. Additionally, exploring the potential benefits of transfer learning will provide valuable insights into leveraging pre-trained models to boost performance.

Through this research, we hope to contribute to the advancement of the field by providing valuable insights into architectural comparisons, optimizing hyperparameters, and leveraging the potential of transfer learning in medical image analysis.

# **Challenges**

During our project, we came across some challenges such as:

* Lack of knowledge in medical terms and processes such as understanding CT and MRI scans, hepatocellular carcinoma, gastro-internal anatomy, etc. We had to independently learn many terms to understand their connection to the articles.
* Lack of knowledge about liver and liver cancer. We wanted to understand the background of our work, so we searched for articles and medical sites on the internet to learn about the impact of the disease and the treatment method.
* The dataset format we used required a specific program to visualize the CT scans.

# **Product**

## **Use Case**

תמונה שמכילה טקסט, תרשים, צילום מסך, עיגול

התיאור נוצר באופן אוטומטי

## **Sequence**

תמונה שמכילה טקסט, צילום מסך, מקביל, מספר

התיאור נוצר באופן אוטומטי

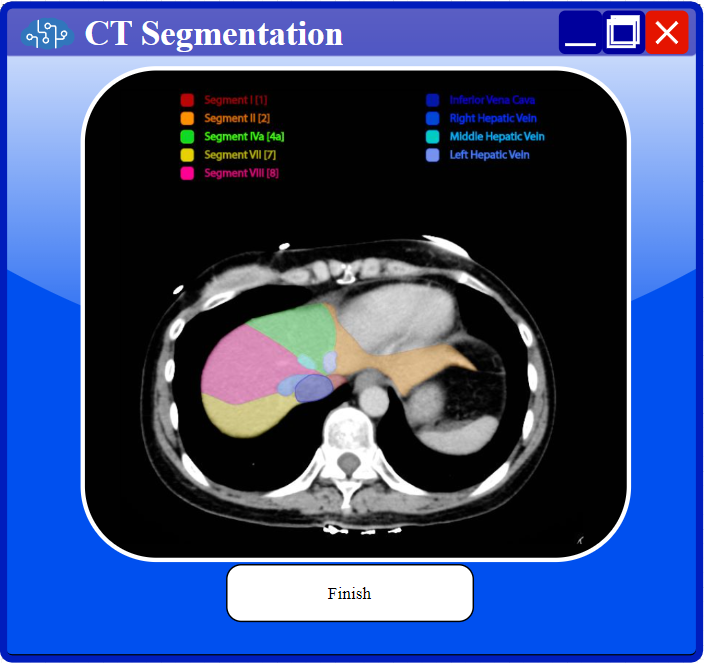
## **GUI**

תמונה שמכילה טקסט, צילום מסך, מערכת הפעלה, תוכנה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, מערכת הפעלה, תוכנה

התיאור נוצר באופן אוטומטי



# **Expected Achievements**

The objective of our project is to conduct a comprehensive analysis of three architectures: TransUNet, Dense-Inception U-net, and Nas-U-net, for the purpose of liver tumor segmentation in CT scans extracted from the LiTS17 dataset. We will compare these architectures using standardized hyperparameter configurations and evaluate their performance based on Dice scores. Through this comparative analysis, we aim to identify the architecture that demonstrates the highest performance.

Furthermore, upon identifying the best-performing architecture, we will implement a system that utilizes this architecture along with the optimal hyperparameter values determined during the comparison phase. The developed system will assist physicians in efficiently analyzing CT scans, leading to faster and potentially more accurate identification and segmentation of liver tumors. This, in turn, can significantly improve patient treatment outcomes.

The success of our project will be measured by achieving the following objectives: successfully comparing the different architectures mentioned, implementing a system that surpasses current accuracy benchmarks, and identifying the optimal hyperparameter values for our proposed network.

One of the significant challenges in liver tumor segmentation is accurately identifying and segmenting small tumors. The precise delineation of these tumors is crucial, as they can often go unnoticed due to their size, shape, and proximity to surrounding organs within the abdominal region.

# **Evaluation / Verification Plan**

|  |  |  |
| --- | --- | --- |
| **CASE #** | **Test explaination** | **Expected result** |
| **1** | drag&drop wrong file format | Nothing will happen  (will not drop the file) |
| **2** | Try to Press "Submit" before choosing file | Nothing will happen (“Submit” button is disabled) |
| **3** | Insert expected file (volume0.nii) (Can also insert more than one) | 1. “Submit” button will be enabled 2. A file logo will appear on the screen to show that a suitable file format was uploaded. |
| **4** | Click on Browse when there is a file uploaded already | Will open the find and open Explorer to add the chosen file to the batch |
| **5** | Click on the “x” button next to the file | will remove the file from the batch |
| **6** | Press “submit” when enabled | 1. Start the segmentation process. 2. When finished, save the result file in the same folder as the given input file and show the segmentation result with a Finish button to return to the main screen. |

# **Bibliography**

1. Mayo Clinic. Liver Cancer.  
<https://www.mayoclinic.org/diseases-conditions/liver-cancer/symptoms-causes/syc-20353659>

2. National Cancer Institute. Liver Cancer. <https://www.cancer.gov/types/liver/what-is-liver-cancer>

3. American Cancer Society. Liver Cancer. <https://www.cancer.org/cancer/liver-cancer.html>

4. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249.<https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660>

5. Baskar, R., Lee, K. A., Yeo, R., & Yeoh, K. W. (2012). Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, *9*(3), 193–199. https://doi.org/10.7150/ijms.3635<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3298009/>

6. Zhao, J., Li, D., Xiao, X., Accorsi, F., Marshall, H., Cossetto, T., ... & Li, S. (2021). United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI. *Medical image analysis*, *73*, 102154.‏

7. Hänsch, A., Chlebus, G., Meine, H., Thielke, F., Kock, F., Paulus, T., ... & Schenk, A. (2022). Improving automatic liver tumor segmentation in late-phase MRI using multi-model training and 3D convolutional neural networks. *Scientific Reports*, *12*(1), 1-10.‏

8. Zhang, C., Hua, Q., Chu, Y., & Wang, P. (2021). Liver tumor segmentation using 2.5 D UV-Net with multi-scale convolution. *Computers in Biology and Medicine*, *133*, 104424.‏

9. Yang, Z., Zhao, Y. Q., Liao, M., Di, S. H., & Zeng, Y. Z. (2021). Semi-automatic liver tumor segmentation with adaptive region growing and graph cuts. *Biomedical Signal Processing and Control*, *68*, 102670.‏

10. Kushnure, D. T., & Talbar, S. N. (2022). HFRU-Net: High-level feature fusion and recalibration unet for automatic liver and tumor segmentation in CT images. *Computer Methods and Programs in Biomedicine*, *213*, 106501.‏

11. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

12. National Institute of Biomedical Imaging and Bioengineering. <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>

13. Image Thresholding - OpenCV Documentation. <https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html>

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

15. Punn, N. S., & Agarwal, S. (2022). Modality specific U-Net variants for biomedical image segmentation: a survey. *Artificial Intelligence Review*, *55*(7), 5845-5889

16. U-Net Explained: Understanding its Image Segmentation Architecture  
https://towardsdatascience.com/u-net-explained-understanding-its-image-segmentation-architecture-56e4842e313a

17. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19* (pp. 424-432). Springer International Publishing.‏

18. Elsken, T., Metzen, J. H., & Hutter, F. (2019). Neural architecture search: A survey. *The Journal of Machine Learning Research*, *20*(1), 1997-2017.‏

19. Weng, Y., Zhou, T., Li, Y., & Qiu, X. (2019). Nas-unet: Neural architecture search for medical image segmentation. *IEEE access*, *7*, 44247-44257.

20. Zhang, Z., Wu, C., Coleman, S., & Kerr, D. (2020). DENSE-INception U-net for medical image segmentation. *Computer methods and programs in biomedicine*, *192*, 105395

21. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.‏

22. Dataset: https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation