תמונה שמכילה גופן, לוגו, סמל, גרפיקה

התיאור נוצר באופן אוטומטי

Capstone Project Phase B

Outline of liver tumors based on image segmentation.

R-2-23-2

Supervisor:

Miri Weiss Cohen

Submitters:

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

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

**Table of Content**

[**Abstract** 3](#_Toc165634805)

[**Introduction** 3](#_Toc165634806)

[**Related Work** 5](#_Toc165634807)

[**Background** 7](#_Toc165634808)

[**CT** 7](#_Toc165634809)

[**Image Segmentation** 8](#_Toc165634810)

[**U-net** 9](#_Toc165634811)

[**TransUNet** 11](#_Toc165634812)

[**ResUNet** 11](#_Toc165634813)

[**Research Process and Achievements** 12](#_Toc165634814)

[**Objective** 12](#_Toc165634815)

[**Dataset** 12](#_Toc165634816)

[**Hyperparameters** 13](#_Toc165634817)

[**Loss Functions** 13](#_Toc165634818)

[**Binary Cross-Entropy** 13](#_Toc165634819)

[**Weighted Binary Cross-Entropy** 14](#_Toc165634820)

[**Dice Loss** 15](#_Toc165634821)

[**Learning Rates** 15](#_Toc165634822)

[**Results** 16](#_Toc165634823)

[**U-Net Model results** 16](#_Toc165634824)

[**ResUNet Results** 18](#_Toc165634825)

[**TransUNet Results** 20](#_Toc165634826)

[**Comparison of the results** 22](#_Toc165634827)

[**Challenges** 23](#_Toc165634828)

[**User Documentation** 28](#_Toc165634829)

[**Introduction** 28](#_Toc165634830)

[**Prerequisites** 28](#_Toc165634831)

[**Projects Overview** 28](#_Toc165634832)

[**Configurable Parameters** 29](#_Toc165634833)

[**Operating Instructions** 29](#_Toc165634834)

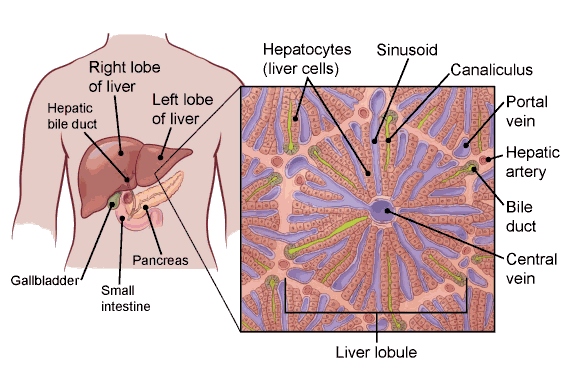
[**Notes** 30](#_Toc165634835)

[**Bibliography** 31](#_Toc165634836)

# **Abstract**

In radiation therapy, radiation oncologists attempt to administer high doses of radiation using X-rays directed at liver tumors while sparing the surrounding tissue. Technologies such as CT scans help oncologists visualize the daily position of the liver, which can change from day to day due to the organ's movement. These scans require radiation oncologists to manually outline the position of the liver and tumors to adjust the direction of the X-rays to increase the dose to the tumors and avoid surrounding tissue. This is a time-consuming and labor-intensive process that can extend the treatment time from 15 minutes per day to an hour per day, which is difficult for patients to tolerate. Using deep learning to automate the segmentation process would significantly speed up the treatment and hopefully lead to a more accurate visualization of the liver tumors, which would help patients receive 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]**](#_Bibliography)

***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]**](#_Bibliography). 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.

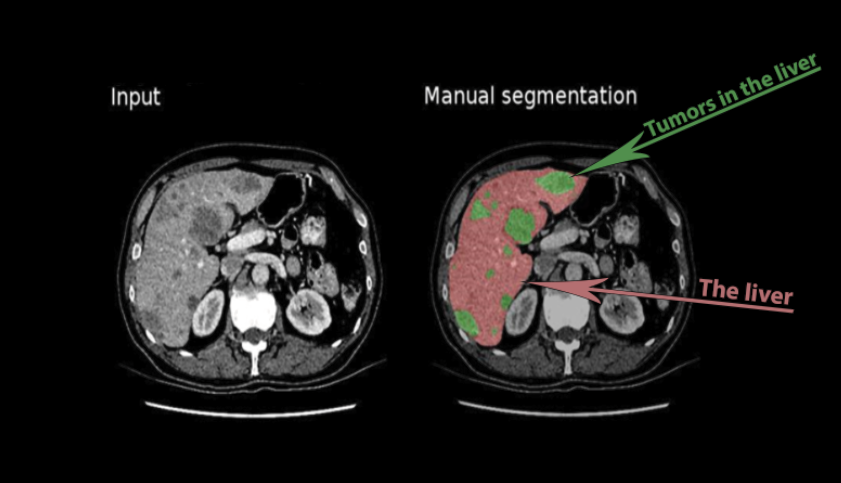
Liver cancer, also known as hepatic cancer, is a type of cancer that starts in the cells of the liver. Liver cancer can develop when healthy liver cells undergo mutations in their DNA, which causes the cells to grow and divide uncontrollably. Over time, these abnormal cells can form a tumor in the liver, which can interfere with its normal function and spread to other parts of the body. [**[3]**](#_Bibliography)

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]**](#_Bibliography). 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]**](#_Bibliography). 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 project. 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.

# 

# **Related Work**

Annika Hänsch et al. [**[6]**](#_Bibliography)**,** 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.

Chi Zhang et al[**[7]**](#_Bibliography), 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.

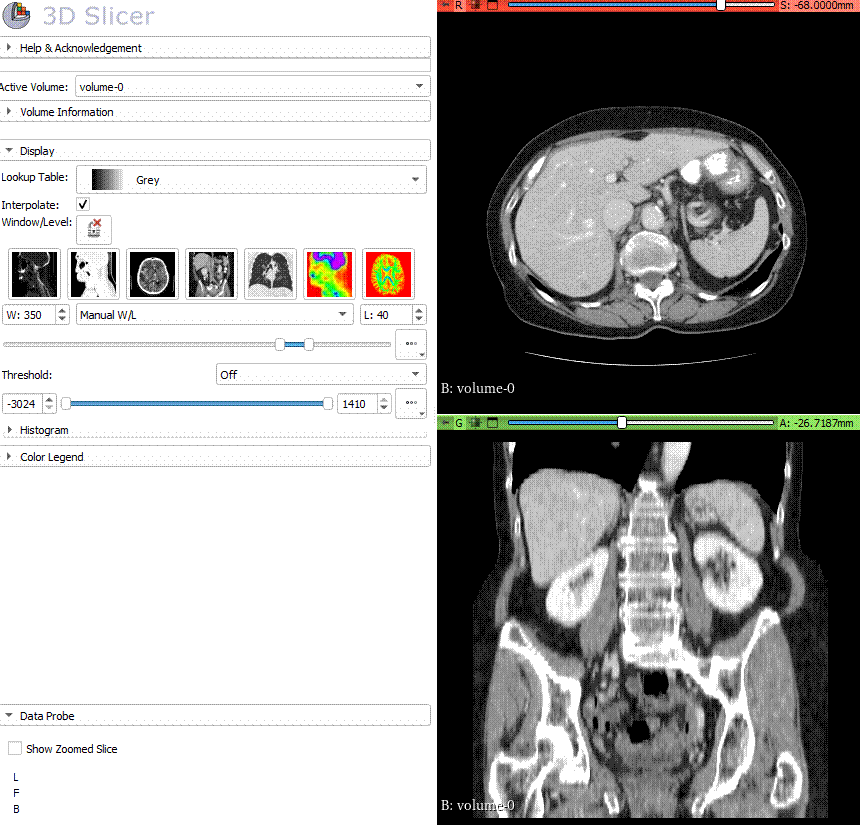
The authors propose a method that combines the following: Window Truncation, Normalization, Standardization, Removing Mean Energy, and Cluster Crop functions as we can see in the picture above.

Devidas T. Kushnure et al.[**[8]**](#_Bibliography)**,** proposed to automate the segmentation procedure using a modified U-Net - HFRU-Net. 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. [**[9]**](#_Bibliography)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.

# **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" [**[10]**](#_Bibliography). 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. 3. 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 many objects or complex scenes with overlapping objects and complex backgrounds. As separating an object from the background in each 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.[**[11]**](#_Bibliography)

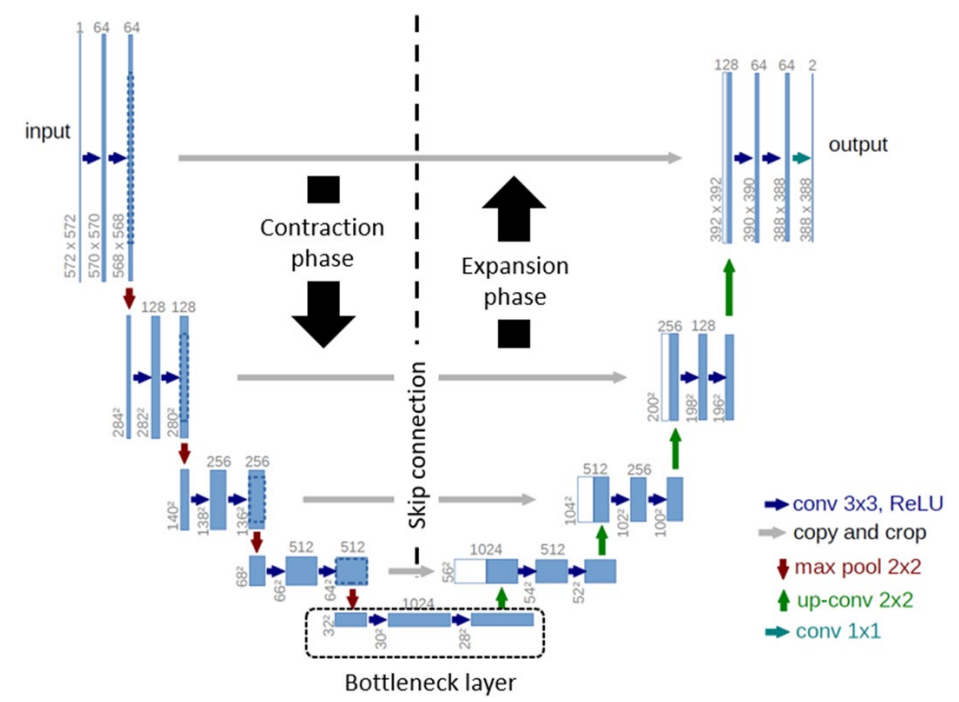
Another technique is Adaptive Region Growing (ARG), this algorithm starts with a seed point for each ROI and grows a region around each seed by adding adjacent pixels that meet certain criteria such as color, texture similarity, etc.[**[12]**](#_Bibliography)

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.[**[12]**](#_Bibliography) 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. [**[13]**](#_Bibliography) 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**

U-Net is a CNN-based segmentation model originally developed for medical images.[**[14]**](#_Bibliography) 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. [**[14]**](#_Bibliography) 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.4.

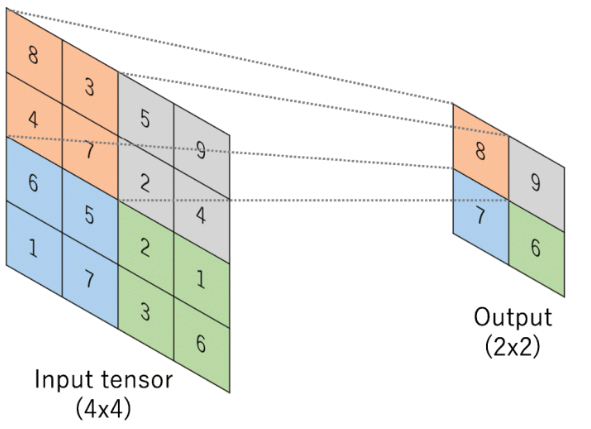


***Fig. 4. The left side is considered as the Encoder as the resolution of the image is compressing and the right side is the Decoder as the resolution is Expanding through each layer back to the original resolution. The extracted feature maps are propagated to the decoder block via a bottleneck layer 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.[**[15]**](#_Bibliography) 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.

תמונה שמכילה תרשים, טקסט, קו, מקביל

התיאור נוצר באופן אוטומטי***Fig. 5. 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.***

******

***Fig. 6. 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 up-sampling, 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 up-sampled layer, it helps to localize the features correctly.

## **TransUNet**

TransUNet is a model which combines both Transformers and U-Net, as a strong alternative for medical image segmentation.[**[16]**](#_Bibliography)In the encoder Phase, a CNN-Transformer hybrid model is applied to leverage both the Transformers and the U-net 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. 7. 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.***[***[16]***](#_Bibliography)

## **ResUNet**

The ResUNet **[[17]](#_Bibliography)** architecture is an innovative adaptation of the U-Net model, which has been widely recognized for its efficiency in medical image segmentation tasks. This advanced architecture incorporates residual blocks, commonly referred to as Res-blocks, which are designed to facilitate the training of deeper neural networks by mitigating the vanishing gradient problem—a common obstacle in traditional deep learning models. Residual blocks consist of a series of convolutional layers where the input to the block is added to the output, effectively allowing the gradients to bypass several layers through a direct pathway. This concept of "skip connections" or "shortcut connections" is a fundamental component of Res-blocks, ensuring that the signal necessary for learning is not diluted as it propagates through the network. By doing so, ResUNet can learn more complex features without the degradation of performance that typically accompanies increased depth.

# **Research Process and Achievements**

## **Objective**

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

## **Dataset**

For our project we used the LiTS17 dataset, obtained from Kaggle[**[18]**](#_Bibliography), 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 used at first the "Slicer" program. This dataset provides us with a comprehensive collection of 3D CT scans, allowing us to effectively train and test our architectures.

תמונה שמכילה טקסט, צילום מסך, הדמיה רפואית, רדיולוגיב

התיאור נוצר באופן אוטומטי

***Fig. 8. We used "Slicer"*** ***program to create a demonstration.***

To adapt the LiTS17 dataset for our project's specific needs, we undertook a series of preprocessing steps designed to optimize the data for our machine learning models.

1. We began by cropping each CT scan image to a uniform resolution of 255x255 pixels, centering the crop to ensure that the most relevant anatomical features, particularly the liver and any present tumors, remained in focus. We then applied a thresholding technique to exclude images with masks that contained insufficient content for liver and tumors, focusing our model training on scans that provided more significant and relevant data.
2. To increase the robustness of our model and to prevent overfitting, we implemented a basic augmentation technique, specifically rotation, which helped create a more diverse set of images from the existing scans.
3. In addition to automated preprocessing, we conducted a manual review of the images and removed any that had inaccurately defined masks. This manual curation was crucial to ensuring the quality of the training data, as poor masks could mislead the learning algorithm, resulting in inaccurate model training.
4. Finally, we normalized the intensity values of the CT images based on Hounsfield units (HU), adjusting them to a specific window ranging from 54 to 66 HU. This step is critical in medical imaging as it enhances the contrast and clarity of different tissues, making it easier for the model to learn from consistent and standardized data. These preprocessing steps were essential for preparing the LiTS17 dataset to effectively train and validate our deep learning architectures, ensuring optimal performance on relevant, high-quality data.

## **Hyperparameters**

In medical image segmentation, the performance of neural network architectures, specifically the U-Net and its variants, is substantially influenced by the choice of hyperparameters. Hyperparameters are the adjustable parameters that control the learning process of the model, and selecting the optimal set is crucial for achieving high accuracy and efficiency.

Our study focuses on two primary hyperparameters: loss functions and learning rates.

## **Loss Functions**

The cornerstone of training neural networks, loss functions quantify the difference between the predicted outputs and the actual data. In image segmentation tasks, the correct choice of a loss function is pivotal for the model to learn precise and reliable segmentations.

We choose these three types of loss functions:

### **Binary Cross-Entropy**

In the pursuit of an effective model for medical image segmentation, particularly for the complex task of liver tumor delineation, we grounded our approach in proven methodologies. Binary Cross-Entropy (BCE) serves as one of the most fundamental loss functions for binary classification problems. It’s widely recognized for its ability to measure the divergence between the predicted probabilities and the actual binary outcomes—making it an essential starting point for our experiments. We chose BCE because it's a time-tested method that provides a solid baseline from which to assess and refine the model's performance. It was important for us to benchmark our model against this standard to understand its capabilities and limitations before considering more complex or tailored loss functions. The simplicity and ubiquity of BCE in binary classification tasks make it a logical first choice. It allowed us to set a comparative standard for our model's ability to segment liver tumors. This step was crucial for establishing a baseline performance, ensuring that any subsequent improvements could be accurately attributed to our innovations and adaptations in the model's architecture and loss function strategy.

The BCE mathematical formula **[19]**:

BCE mathematical formula for a single pixel:

This function works by comparing the model's predicted probability that a given sample belongs to the positive class against the actual outcome. If the actual label is 1 and the model predicts 0, or vice versa, the loss becomes large. Conversely, if the prediction is close to the actual label, the loss becomes small.

### **Weighted Binary Cross-Entropy**

We started with Binary Cross-Entropy because it's well-established and reliable for telling if our model is on the right track. But we noticed it doesn't always give enough attention to the rarer, but critical, features like tumors in our liver scans. So, we introduced Weighted BCE to give those important parts the focus they need. This way, our model learns to spot tumors more effectively, balancing out the learning process since tumors occur less frequently than normal liver tissue in our images. Its addresses the issue of class imbalance by assigning different weights to the classes, thus emphasizing the minority class during training.

Where weight\_one and weight\_zero are coefficients that represent the weights for the positive and negative classes, respectively.

We chose to set weight\_one to 1.0 and weight\_zero to 0.625  
and therefore to increase the weight of true label pixels.

### **Dice Loss**

Dice Loss was chosen for its ability to enhance the alignment between predicted segmentation and ground truth, a vital factor for discerning fine details in medical images. Its design counteracts the common issue of imbalance in medical datasets, specifically by emphasizing smaller but critical areas such as liver tumors, ensuring they are not overshadowed by the larger background. Additionally, the direct relation of the Dice coefficient to segmentation quality means that improvements in Dice scores are likely to translate to more precise and clinically useful segmentations.

## **Learning Rates**

This hyperparameter determines the size of the model's steps during optimization. An adequately chosen learning rate ensures that the model converges to a solution that is both accurate and generalizable. We tested three distinct learning rates:

0.001: A standard learning rate that offers a balance between the speed of convergence and the accuracy of the solution.

0.0001: A lower rate, allowing the model to take smaller steps during training, potentially leading to better generalization at the expense of a longer training time.

0.00001: A conservative approach to training where the model incrementally adjusts its weights, minimizing the risk of overshooting optimal solutions but requiring more epochs to converge.

## **Results**

The efficacy of our segmentation models can be visually appreciated through the comparison between the ground truth masks and the predicted masks for each U-Net configuration. This section outlines how each model utilizes the original images and their corresponding ground truth masks:

* Original Image: These scans represent the inputs to our models, showcasing varied densities and structures within the liver. Darker regions within these images often indicate the presence of potential tumors, which are the primary focus of our segmentation efforts.
* Ground Truth Mask: These masks are manually segmented by expert radiologists and serve as the gold standard for both training and evaluating our models. They precisely delineate the boundaries of liver tumors, ensuring that our models learn to accurately identify and segment these critical areas.

The following segments will detail the performance and outcomes of each model, using these images to demonstrate the models' abilities to replicate expert-level tumor identification and segmentation.

### **U-Net Model results**

The U-Net model, across various configurations, exhibited reliable TPRs, underscoring its potential for clinical applications where high sensitivity is critical. The differences in performance metrics highlight the nuanced impact of hyperparameter selection on segmentation outcomes, informing future model optimizations and research directions in the field of medical imaging.

A screenshot of a graph

Description automatically generated

***Fig. 9. Comparison between the U-net executions with different Loss functions and Learning Rates.  
Marked the best Learning Rates for each Loss function.***

**Best Performing Loss Function and Learning Rate**

For the U-Net model, the **Weighted Binary Cross-Entropy** loss function has proven to be the most effective. It shows a strong ability to highlight important areas within medical images, which is critical given the challenge of class imbalance. The learning rate of **0.001** stands out as the best among the rates we tested, aligning with both high true positive rates and Dice coefficients, indicating a successful segmentation by numerical standards. Yet, it's crucial to remember that these metrics don't always fully represent the model’s ability to capture smaller details, such as tumors, as evidenced in the results. Therefore, while 0.001 is the optimal learning rate according to the data, the visual results also need to be considered to ensure the model's practical effectiveness in clinical applications.

**Visual Segmentation Results**

A close-up of a brain

Description automatically generated

A gray and white image of a person's face

Description automatically generated

A grayscale shot of a whale

Description automatically generated

***Fig. 10. Examples of U-net with Weighted Binary Cross-Entropy and Learning Rates of 0.001***

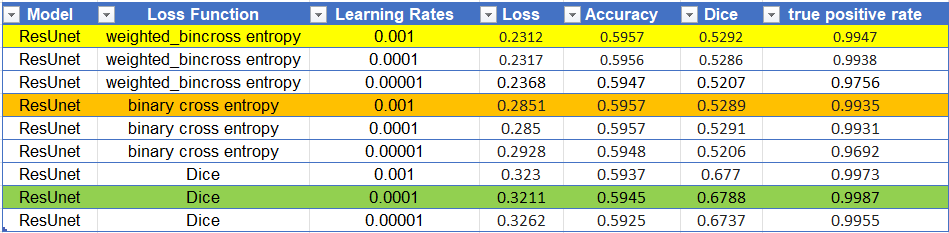
**First Example (Predicted Mask):** The predicted mask closely mirrors the ground truth but with noticeable smoothing of boundaries. This results in a more generalized liver shape, potentially obscuring finer details such as small protrusions or irregularities at the liver’s edge.

**Second Example (Predicted Mask):** Similar to the first, this mask shows a robust capability in identifying the main tumor regions but with a smoothing effect that slightly over-generalizes the smaller tumor regions. This effect can lead to less precision in capturing the exact morphological features of the liver and tumors.

**Third Example (Predicted Mask):** Here, the predicted mask aligns well with the larger structures of the ground truth, yet the smoothing at the edges can be observed. This may result in under-segmentation, particularly noticeable at the bottom part of the liver, where smaller details are smoothed over.

The smoothing observed in these predicted masks suggests a trend in the model’s segmentation approach, favoring a smoother, more uniform delineation of the liver and tumor boundaries. While this might be beneficial for reducing noise and avoiding overfitting to minute details that could be anomalies, it also highlights a potential area for improvement. Enhancing the model’s ability to preserve fine details without introducing noise could lead to more precise and clinically useful segmentation outputs.

### **ResUNet Results**

An advanced version of the U-Net with integrated residual learning blocks, has shown promise in addressing the vanishing gradient problem, allowing the training of deeper neural network architectures. Our study put this enhanced model to the test.

***Fig.11.Comparison between the ResUNet executions with different Loss functions and Learning Rates.  
 Marked the best Learning Rates for each Loss function.***

**Best Performing Loss Function and Learning Rate**

For ResUNet, choosing the right settings is key to success. When we paired the model with the **Weighted Binary Cross-Entropy** loss function, we found that a learning rate of **0.001** gave us the best results. This setup did a great job at picking up important details in the medical images, especially with tricky tasks like outlining tumors. This careful approach seems to match the precision needed for medical imaging, capturing the subtle yet crucial features that are vital for accurate diagnoses.

**תמונה שמכילה סרט רנטגן, הדמיה רפואית, צילום מסך, שחור ולבן

התיאור נוצר באופן אוטומטיVisual Segmentation Results**

A close-up of a person's face

Description automatically generated

תמונה שמכילה שחור ולבן, שחור, גרפיקה, עיצוב גרפי

התיאור נוצר באופן אוטומטי

***Fig. 12. Examples of ResUNet with Weighted Binary Cross-Entropy and Learning Rates of 0.001***

**First Example (Predicted Mask):** The ResUNet model successfully captures the primary anatomical structure of the liver as well as the main tumor region, demonstrating close alignment with the ground truth mask. The boundaries of the liver and tumor are well-defined, though a slight smoothing is observed, which reduces the detail that might be critical for detecting minute pathological features.

**Second Example (Predicted Mask):** This segmentation showcases the model's capability to delineate large tumor masses effectively, with a high degree of overlap with the ground truth. Nevertheless, the predicted mask here also exhibits a subtle smoothing effect around the smaller tumor masses, potentially leading to an underestimation of tumor extents or missing small, but clinically significant, lesions.

**Third Example (Predicted Mask):** In this example, the ResUNet model has achieved robust segmentation of the liver and tumors, mirroring the ground truth with high accuracy. The model demonstrates improved edge detection over standard U-Net models, yet it continues to display some degree of smoothing at the liver's boundary and around the tumor edges, which might obscure finer details crucial for detailed medical analysis.

These results highlight the ResUNet model's enhanced capability for liver and tumor segmentation compared to simpler models, attributed to its integrated residual learning blocks that help mitigate the vanishing gradient problem. This enhancement allows the model to train more deeply and effectively, capturing complex features necessary for precise medical imaging.

### **TransUNet Results**

A screenshot of a graph

Description automatically generatedTransUNet integrates the robust feature extraction capabilities of Convolutional Neural Networks (CNNs) with the global contextual awareness of transformers, creating a synergistic approach to the complex task of medical image segmentation.

***Fig. 13. Comparison between the TransUNet executions with different Loss functions and Learning Rates.  
 Marked the best Learning Rates for each Loss function.***

**TransUNet: Best Loss Function and Learning Rate**

In our trials with the TransUNet, which is designed to utilize both CNN and transformer strengths, the combination of the **Weighted Binary Cross-Entropy** loss function and a more cautious learning rate of **0.0001** stood out. This approach led to the lowest loss values and the highest accuracy and true positive rates in our numerical evaluations.

However, when we look at the segmented images the model produced, we get a more nuanced story. These images highlight the model’s challenges in accurately segmenting finer details like tumors, which are critical for successful medical applications. While the learning rate of 0.0001 with Weighted Binary Cross-Entropy has shown potential, it’s important to recognize that the numbers only tell part of the story. The segmentation accuracy, particularly for critical features, requires closer examination to ensure that the TransUNet model is not just theoretically competent but also practically reliable for clinical use.

A close-up of a skull

Description automatically generated**Visual Segmentation Results**

A close-up of a mask

Description automatically generated

A close-up of a grey whale

Description automatically generated

***Fig. 14. Examples of TransUNet with Weighted Binary Cross-Entropy and Learning Rates of 0.001***

**First Example (Predicted Mask):** The TransUNet demonstrates strong performance in capturing the overall shape and location of the liver and tumors, as seen in the alignment with the ground truth. However, there is a noticeable degree of smoothing, particularly around the smaller tumor features, which could potentially result in missing critical diagnostic details. The model’s blending of edges may reduce the precision needed for detailed clinical analysis.

**Second Example (Predicted Mask):** In this case, the TransUNet inaccurately segments an area as a tumor that does not correspond to any tumor presence in the ground truth mask. This indicates a false positive, which is a significant concern in clinical diagnostics. While the segmentation of the actual liver structure is relatively accurate, the erroneous identification of tumor tissue highlights potential issues in the model's specificity and reliability.

**Third Example (Predicted Mask):** Here, the TransUNet has notably missed segmenting two small tumors that are present in the ground truth mask, indicating a false negative result. This failure to detect smaller tumors could lead to significant issues in clinical diagnostics, where comprehensive tumor detection is essential. The model successfully segments the larger liver structure but its inability to capture these smaller tumors necessitates improvements in sensitivity and detection capabilities.

These revised observations highlight both the strengths and critical areas for improvement in the TransUNet model. The integration of CNNs with transformers allows for robust general feature extraction but also points to specific challenges such as avoiding false positives and enhancing tumor detection sensitivity.

These revised observations highlight both the strengths and critical areas for improvement in the TransUNet model. The integration of CNNs with transformers allows for robust general feature extraction but also points to specific challenges such as avoiding false positives and enhancing tumor detection sensitivity.

### **Comparison of the results**

|  |  |  |  |
| --- | --- | --- | --- |
| Ground Truth | U-Net | ResUNet | TransUNet |
|  |  |  |  |
|  | The U-Net model achieves a reasonable segmentation of the liver and identifies the primary tumor accurately, aligning well with the ground truth mask. However, it misses finer details in the liver's outline and does not segment a smaller tumor present in the ground truth. | The ResUNet model performs exceptionally well in this instance, providing highly accurate segmentation of both the liver and the tumors. It closely matches the ground truth mask, capturing both the large and smaller tumors with high accuracy. | The model's segmentation is broad but fails to accurately capture the full structure of the liver, missing significant parts of its boundary. Additionally, it exhibits false negatives by not identifying all tumors present in the ground truth mask, which is crucial for reliable medical diagnostics. |

This comparative evaluation of U-Net, ResUNet, and TransUNet illustrates their varied efficiency in liver tumor segmentation. U-Net offers good baseline accuracy but misses smaller tumors. ResUNet excels in detecting and segmenting tumors with high precision, making it ideal for detailed clinical analysis. TransUNet, while advanced, falls short in accuracy and completeness, suggesting a need for further optimization. Selecting the right model hinges on the specific demands of clinical imaging tasks.

## **Challenges**

Despite the generally high-performance metrics reported for the U-Net, ResUNet, and TransUNet models when using Dice loss, the qualitative assessment of the segmentation outputs revealed significant challenges. Specifically, while these models exhibited proficiency in segmenting larger anatomical structures such as the liver, they consistently failed to accurately delineate smaller, yet clinically significant features like tumors. This limitation was present across all learning rates and highlighted by the substantial black areas found within the CT scan NII files, which posed additional complications to the training process.

**U-Net**

The U-Net model, even with its capacity for detailed segmentation, displayed a substantial discrepancy between the high Dice coefficients and the practical outcomes. The segmentation images illustrate that the model, while adept at capturing the broader liver region, did not successfully segment the tumors. This highlights a potential limitation of the U-Net in differentiating between regions with subtle differences in pixel intensities or small-scale features, which is crucial in tumor detection.

**ResUNet**

Similarly, the ResUNet architecture, despite its enhanced ability to learn complex patterns through residual learning, did not effectively improve the segmentation of tumors. The residual connections that typically facilitate training deeper networks did not translate into better localization of these small and less defined features in the medical images.

**TransUNet**

The TransUNet, with its incorporation of transformer models to capture global dependencies, also struggled with the precise segmentation of tumors. While transformers are renowned for their ability to manage contextual information over large regions, this trait did not extend to the accurate identification of smaller, localized targets such as tumors within the images.

These results indicate a common challenge faced by U-Net architectures: while they can achieve high scores on metrics like the Dice coefficient, which suggests good performance on larger, more homogeneous regions, they may still lack in the precision required for more intricate tasks like tumor segmentation. The failure to segment smaller areas accurately could be due to several factors, including the imbalance in the representation of such features in the training data, the loss of functions prioritizing larger region segmentation, or a need for more sophisticated feature extraction mechanisms that can capture nuances at a smaller scale.

Further research could explore the integration of attention mechanisms, more advanced data augmentation techniques, or the incorporation of additional training data specifically focused on the features of interest to overcome these limitations. Additionally, incorporating multi-scale analysis and specialized post-processing techniques could improve the model's ability to delineate tumors more effectively.

In conclusion, while the tested U-Net variants demonstrated proficiency in segmenting large structures, their ability to segment small-scale features like tumors was limited. This emphasizes the need for more focused model tuning and possibly architectural adjustments to enhance the segmentation of clinically relevant features that are critical for diagnostic purposes.

**Noise in the Dataset**

Upon manual examination of the liver segmentation dataset, several instances of noise were identified. Notably, some liver masks were improperly segmented with a white color, which was supposed to indicate liver tumors, instead of the correct gray color for healthy liver tissue. This mislabeling represents a significant source of noise, as it directly impacts the model's learning process by providing incorrect information about the liver's anatomy.

תמונה שמכילה שרטוט, ציור, אומנות, שחור ולבן

התיאור נוצר באופן אוטומטיתמונה שמכילה שרטוט, ציור, אומנות, צללית

התיאור נוצר באופן אוטומטי

***Fig. 15. Example for image with inaccurate mask***

**Impact on U-Net Model Performance**

The presence of noise in the dataset has a direct and detrimental effect on the performance of U-Net models. These models are designed to learn from accurate, consistent representations of the data. However, when faced with inaccuracies and inconsistencies, the models struggle to generalize from the training data, leading to poor segmentation performance. This is particularly problematic in medical image segmentation, where precision is crucial for clinical applications.

Furthermore, the discovery of noise raises concerns about the overall accuracy of the segmentation masks provided as true labels. If some masks are incorrectly labeled, it is reasonable to assume that others might also contain errors, albeit less obvious ones. This undermines confidence in the dataset's reliability and, by extension, the model's predictive accuracy.

**תמונה שמכילה טקסט, שרטוט, שחור ולבן, להדפיס

התיאור נוצר באופן אוטומטי**תמונה שמכילה טקסט, צללית, איור

התיאור נוצר באופן אוטומטי

***Fig. 15. Another example of a mask which might be inaccurate.***

**Computational Constraints**

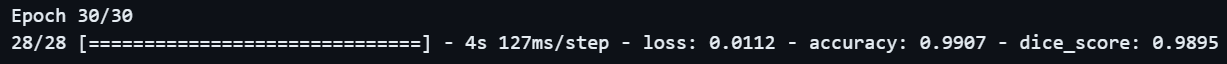
Computational constraints limited our ability to enhance the dataset and model architecture. Due to reliance on Google-Colab and personal computing resources, there were significant limitations on the extent of data augmentation that could be performed. Similarly, the computational constraints restricted our ability to explore more complex U-Net architectures, such as deeper networks that might be more resilient to noise.

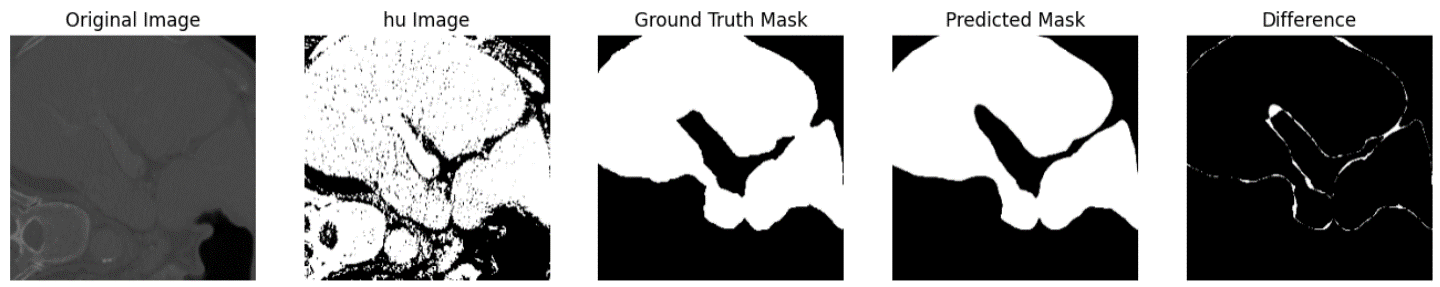
**New Approach**

During our research, and through a detailed examination of our models' performance, we identified a new strategy to address the challenges encountered in segmenting liver tumors. Our observations indicated that while models utilizing Dice loss were proficient at delineating the liver, they were less effective at identifying tumors. Conversely, models employing Binary Cross-Entropy based loss functions experienced higher-than-expected losses, which we attributed to the non-binary nature of the dataset values.

To refine our approach, we proposed a dual-model strategy. The first model employs Dice loss to segment the liver exclusively. Subsequently, the segmented output from the first model serves as an additional input for the second model, which utilizes Weighted Binary Cross-Entropy to segment tumors. This Dual-Model approach allows for more focused and specialized processing by each model according to the distinct characteristics of the liver and tumors, respectively.

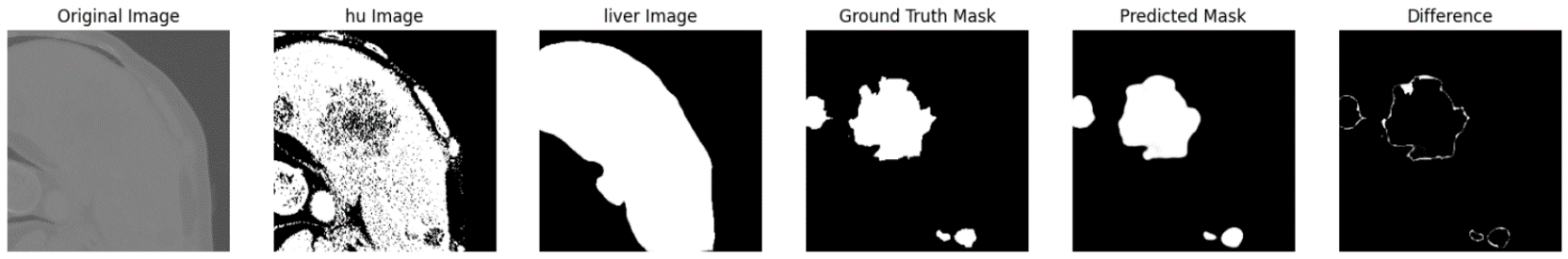
Upon integrating the outputs from both models—combining the liver segmentation from the first with the tumor segmentation from the second—we observed significant improvements. Both models independently achieved high accuracy and Dice scores. This innovative method offers a promising solution to enhance the precision and reliability of liver tumor segmentation, aligning with our project's objectives.

**First Model Results**

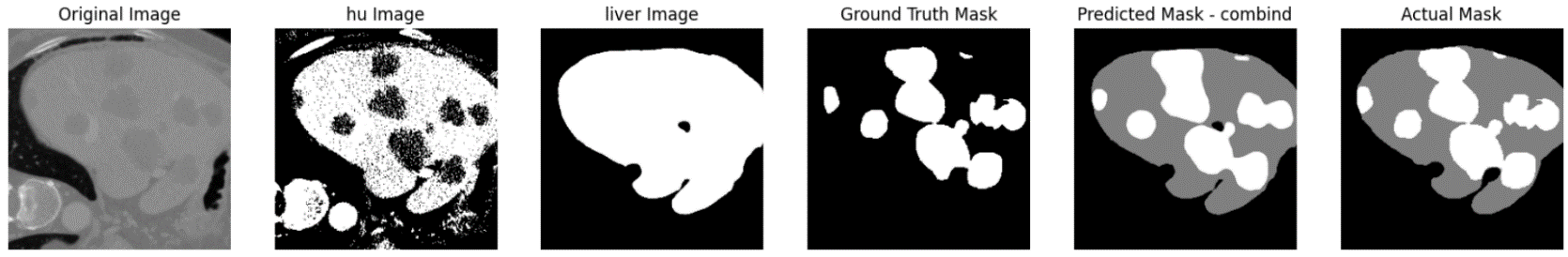
****

***Fig. 16. Example of an almost accurately predicted complex shape of liver.***

**Second Model Results**



***Fig. 17. Example of an almost accurately predicted complex shape of liver tumor.***

**Combined Results**

***Fig. 16. The combined result of the tumor in the liver***

The segmentation results from the first model reveal an incomplete liver delineation, with a small portion of the liver accidentally classified as background. Nonetheless, this outcome demonstrates the effectiveness of the second model, which reliably identifies tumors within the boundaries defined by the first model's liver segmentation, without attempting to re-segment the liver. Consequently, the second model exclusively focuses on detecting tumors within the liver regions previously identified, aligning perfectly with our project's objectives.

**Conclusion**

This project was dedicated to an exhaustive comparative analysis of three distinct architectural frameworks—U-net, ResUNet, and TransUNet—tailored for the segmentation of liver tumors in CT scans, leveraging the LiTS17 dataset. The study focused on implementing a variety of hyperparameters, specifically different loss functions and learning rates, to ascertain the most effective model in terms of Accuracy, True Positive Rates, and Dice scores.

Through our systematic examination, we identified notable differences in the performance capabilities of each model. Our findings underscored the significance of selecting appropriate loss functions and learning rates, which directly influence the precision and effectiveness of the segmentation process. Notably, our exploration led us to a novel dual-model approach where the segmentation task was bifurcated; the first model employing Dice loss focused solely on liver segmentation, while the second utilized Weighted Binary Cross-Entropy to target tumor delineation. This method significantly enhanced the accuracy and specificity of tumor detection, reflecting a promising direction for future research and application.

Learning lessons: Initially, our work suffered from organizational issues as we started by writing the project in an unreadable flow, which later required time-consuming adjustments to hyperparameters and models between executions. To resolve this, we developed three distinct notebooks based on a generic workflow that included hyperparameter and U-net model blocks, which could be adapted depending on the specific model notebook. This restructuring greatly improved our project management and execution efficiency.

Project benchmarks: We believe we have met our project benchmarks despite the tight schedule and challenging time that we are facing. Our proposed approach to the segmentation task demonstrated that we even surpassed our initial objectives.

The project also highlighted the crucial role of extensive and diverse datasets in the training process, where the quality and variance of data notably affect the learning outcomes. Through the deployment of our models, it became evident that adaptations in the preprocessing steps such as image normalization based on Hounsfield units and data augmentation were pivotal in refining the model's learning efficacy.

In conclusion, the project not only provided deep insights into the architectural strengths and weaknesses of each model but also pushed the boundaries of current methodologies by introducing a new, dual-model approach to segmentation tasks.

# **User Documentation**

## **User Guide**

### **Introduction**

This documentation assists users in operating our Google Colab projects designed for the segmentation of liver tumors using three different models: U-Net, ResUNet, and TransUNet. Each project utilizes the LiTS17 dataset and includes customizable hyperparameters for optimal performance evaluation.

### **Projects Overview**

There are four separate Google Colab notebooks, one for each model:

1. Unet Notebook
2. ResUnet Notebook
3. TransformerUnet Notebook
4. DualModelApproach Notebook – **Not** **Configurable.**

Each notebook is pre-configured with a specific set of operations to preprocess data, train the model, and evaluate its performance based on Accuracy, True Positive Rates, and Dice scores.

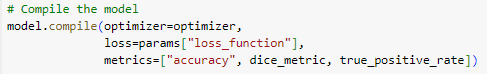
### **Configurable Parameters**

Users can configure the following parameters based on their requirements:

* **Loss Functions:** Binary Cross-Entropy, Weighted Binary Cross-Entropy, Dice Loss.
* **Learning Rates:** 0.001, 0.0001, 0.00001.
* Other parameters such as batch\_size, epochs, threshold, and image cropping dimensions (crop\_x, crop\_y) are set as defaults but can be adjusted as needed.

תמונה שמכילה טקסט, צילום מסך, גופן, מספר

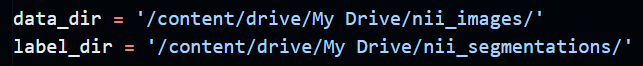
התיאור נוצר באופן אוטומטי



Refer to the uploaded images in the documentation for default parameter settings and configurations.

### **Operating Instructions**

1. Download files from GitHub: [https://github.com/AvielMalayev/FInalProject.git](https://github.com/AvielMalayev/FInalProject.git%20)
2. Download dataset from Kaggle [**[18]**](#_Bibliography).
3. Organize the data in Google Drive in 2 folders.



1. **Open the Google Colab Notebook:** Choose the notebook corresponding to the model you wish to use (U-Net, ResUNet, or TransUNet).
2. **Configure Parameters:**

* Navigate to the cell containing the params dictionary.
* Modify the loss function and "learning\_rate" according to your needs. Other parameters like "batch\_size" and epochs can also be adjusted here.

1. **Load and Preprocess Data:**

* Ensure the dataset path is correct.
* Run the preprocessing steps which include cropping, thresholding, augmentation, and normalization as per the default setup or your adjustments.

1. **Train the Model:**

* Execute the training cell. Ensure that the correct loss function and optimizer are selected.
* Monitor training progress through output logs which display loss and accuracy metrics.

1. **Evaluate the Model:**

* After training, evaluate the model using the test data split.
* Review the performance metrics such as Accuracy, Dice score, and True Positive Rate to assess the model's effectiveness.

**Notes**

* Ensure that each Colab notebook is connected to a Google Drive containing the necessary LiTS17 dataset in the correct format.
* Consider using GPU acceleration in Colab for faster computation by adjusting the notebook settings:

Runtime > Change runtime type > Hardware accelerator > GPU.

## **Maintenance Guide**

### **Prerequisites**

To ensure the effective maintenance and operation of the segmentation models within this project, users must have the following:

* Access to Google Colab.
* A basic understanding of Python programming.
* Familiarity with deep learning concepts, particularly convolutional neural networks.

### **Documentation and Code Structure**

Each segment of the code is accompanied by a detailed header that describes the purpose of the block. This structured documentation is designed to aid users in comprehending the overall flow of the project, ensuring clarity in navigation and modification of the code.

### **Libraries and Dependencies**

This project employs key libraries essential for processing and analyzing medical imaging data:

**TensorFlow:** Utilized for building and training deep learning models. Due to its rapid development, TensorFlow functions may evolve or become deprecated, necessitating regular updates to maintain compatibility and functionality.

**NumPy and Nibabel:** These libraries are stable, supporting numerical computations and handling of NIfTI data, respectively. While changes are less frequent, staying informed about updates is recommended to ensure ongoing stability and security.

# **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. 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.‏

7. 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.‏

8. 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.‏

9. 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).

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

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

12. 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.‏

13. 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).‏

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

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

16. 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*.‏

17. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).‏

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

19. Y. Ho and S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling," in IEEE Access, vol. 8, pp. 4806-4813, 2020, doi: 10.1109/ACCESS.2019.2962617.