Review Article

Akash Halder, Arup Sau, Surya Majumder, Dmitrii Kaplun*, and Ram Sarkar

An experimental study of U-net variants on liver segmentation from CT scans

https://doi.org/10.1515/jisys-2024-0185
received March 20, 2024; accepted September 5, 2024

Abstract: The liver, a complex and important organ in the human body, is crucial to many physiological processes. For the diagnosis and ongoing monitoring of a wide spectrum of liver diseases, an accurate segmentation of the liver from medical imaging is essential. The importance of liver segmentation in clinical practice is examined in this research, along with the difficulties in attaining accurate segmentation masks, particularly when working with small structures and precise details. This study investigates the performance of ten well-known U-Net models, including Vanilla U-Net, Attention U-Net, V-Net, U-Net 3+, R2U-Net, U^2-Net, U-Net++, Res U-Net, Swin-U-Net, and Trans-U-Net. These variations have become optimal approaches to liver segmentation, each providing certain benefits and addressing particular difficulties. We have conducted this research on computed tomography scan images from three standard datasets, namely, 3DIRCADb, CHAOS, and LiTS datasets. The U-Net architecture has become a mainstay in contemporary research on medical picture segmentation due to its success in preserving contextual information and capturing fine features. The structural and functional characteristics that help it perform well on liver segmentation tasks even with scant annotated data are well highlighted in this study. The code and additional results can be found in the Github https://github.com/akalder/ComparativeStudyLiverSegmentation.

Keywords: liver cancer, liver segmentation, medical imaging, CT scan, U-Net, deep learning

Abbreviations

AHCNet	Attention hybrid connection network
CHAOS	Combined healthy abdominal organ segmentation
CT	Computed tomography
CNN	Convolutional neural network
DICOM	Digital imaging and communications in medicine
DSC	Dice similarity coefficient
HCC	Hepatocellular carcinoma
HU	Hounsfield units
IoU	Intersection over union

*** Corresponding author: Dmitrii Kaplun,** Department of Automation and Control Processes, Saint Petersburg Electrotechnical University “LETI”, St Petersburg 197022, Russia, e-mail: dikaplun@etu.ru
Akash Halder: Department of Computer Science and Engineering, Jadavpur University, Jadavpur, Kolkata, West Bengal 700032, India, e-mail: akash.halder@tatasteel.com
Arup Sau: Department of Computer Science and Engineering, Institute of Engineering and Management, Kolkata 700032, India, e-mail: arup.sau.24@kgpian.iitkgp.ac.in
Surya Majumder: Department of Computer Science and Engineering, Heritage Institute of Technology, Kolkata, West Bengal 700107, India, e-mail: surya.majumder.cse24@heritageit.edu.in
Ram Sarkar: Department of Computer Science and Engineering, Jadavpur University, Jadavpur, Kolkata, West Bengal 700032, India, e-mail: ram.sarkar@jadavpuruniversity.in

MRI	Magnetic resonance imaging
PReLu	Parametric rectified linear unit
RCNN	Recurrent-CNN
ReLU	Rectified linear unit
RSU	ReSidual U
ROI	Region of interest
RVD	Relative volume difference
VOE	Volume overlap error

1 Introduction

The liver, as one of the largest and most intricate organs in the human body, performs numerous vital functions, including detoxification, metabolism, and the production of essential proteins. The anatomy of the liver is shown in Figure 1. It primarily comprises the left lobe and the right lobe. Both are further subdivided into many small lobules. These are attached to small ducts that connect to larger ducts, thereby forming the hepatic duct. The hepatic duct serves the purpose of transporting bile from the liver to the gall bladder.

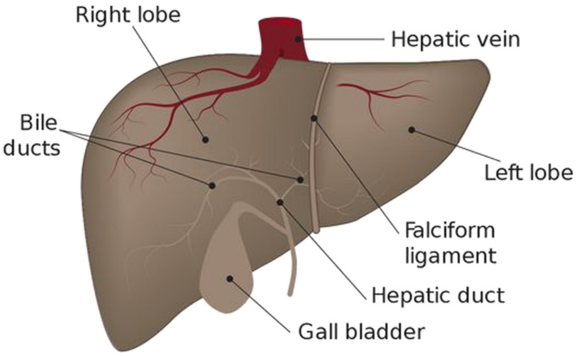


Figure 1: Anatomy of a typical human liver [1].

Accurate liver segmentation is a prerequisite for diagnosing a wide range of liver diseases and conditions. Medical professionals rely on segmented liver images to identify abnormalities, such as tumors, cysts, abscesses, and vascular malformations [2–5]. Hepatocellular carcinoma (HCC) comprises four phases [6]. Most of them start as single tumor cells, while others grow as small cells throughout the liver. Cholangiocarcinoma [7] can occur in bile duct cells, the hilum, where the bile duct exits, the distal bile duct, etc.

Segmented images enable clinicians to precisely measure and analyze the size, volume, and location of liver lesions, aiding in the early detection, staging, and prognosis of liver diseases, i.e., the practical utilization of liver segmentation task [8–14]. Because of this, liver segmentation in medical imaging is not only an essential technical task, but also a critical step with far-reaching clinical implications. Accurate segmentation empowers medical professionals with precise information that influences diagnosis, treatment planning, and patient outcomes.

In the field of medical image segmentation convolutional neural networks (CNNs) have revolutionized their ability to learn important features from input images. They have been widely used in various domains, including the medical field. CNNs have been used for various medical imaging tasks such as classification, such as lung cancer classification [15]. For example, Alzheimer detection using T1-weighted MRI images [16], for identifying dengue from peripheral blood microscopy [17], and others for early disease detection and diagnosis. The method EU^2-Net, as proposed by Roy et al. [18], introduces an efficient ensemble model with attention-aided triple feature fusion for tumor segmentation from breast ultrasound images. They have also been used for several segmentation tasks too like segmentation of brain MRI using moth-flame optimization with a modified cross entropy-based fitness function proposed by Bhattacharyya et al. [19] or an ensemble of U-Net framework for lung nodule segmentation proposed by Gautam et al. [20]. However, it may struggle with segmenting small structures or capturing intricate details due to pooling layers that reduce spatial resolution [21–26]. It can cause a loss of contextual information, impacting the accuracy of boundary delineation. To overcome these challenges, U-Net architecture is chosen over traditional CNNs for image segmentation due to its design that specifically provide solutions for accurate boundary delineation, capturing finer details, and preserving spatial information, making it a cornerstone in modern medical image segmentation research.

For this task, we used computed tomography (CT) images to conduct the comparative study. CT images offer high spatial resolution, which allows accurate segmentation of the liver and delineation of its highly varying boundaries. CT images provide information about tissue density through hounsfield units (HU). HU clipping can be performed on CT images, allowing us to highlight the liver portion from the image and then use this for further processing. HU clipping is not applicable to magnetic resonance imaging (MRI) images because it does not use HU for tissue density measurement. The pipeline of our proposed methodology is shown in Figure 2.

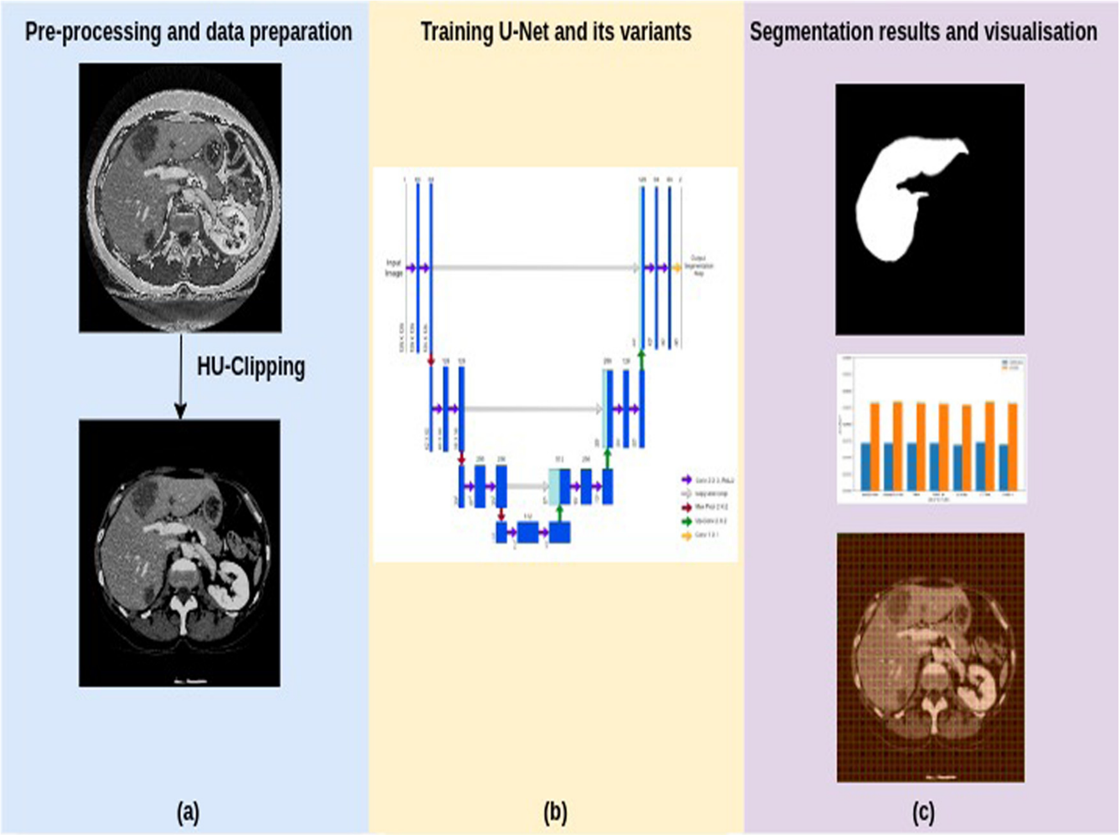

Pre-processing and data preparation
Training U-Net and its variants
Segmentation results and visualisation
HU-Clipping
(a)
(b)
(c)

Figure 2: Diagram depicting the workflow used in the present study: (a) we use 3DIRCADb, CHAOS, and LiTS datasets and perform HU clipping on the images to highlight the liver from the CT scans; (b) we train the U-Net and its variants for generating the segmentation task; and (c) results obtained include the liver mask, values of different evaluation metrics, and GradCAM visualizations, etc.

The U-Net architecture is a U-shaped neural network that works very well in different image segmentation tasks. It was first presented by Ronneberger et al. [27] in 2015 and has since grown to be a popular framework in applications for segmenting images. The name of the U-Net architecture comes from its distinctive U-shaped topology. The expanding path and the contracting path, also referred to as the encoder and decoder, respectively, make up its two fundamental components. U-Net can successfully collect small details in images and preserve contextual information because of this special structure. The effectiveness of U-Net lies in its capacity to produce incredibly precise segmentation maps even when dealing with limited annotated input. This is mostly accomplished through the architectural layout and features like hierarchical feature extraction [28–34] and skip connections [35–38]. The symmetric structure of U-Net allows for the preservation of tiny features while also allowing for the integration of data from various scales, enabling it to provide precise segmentation even in the absence of many training samples.

This comparative study can be extended to the task of liver tumor segmentation. The 3D Image Reconstruction for Comparison of Algorithm Database (3DIRCADb) dataset was originally created for the purpose of tumor segmentation. It consists of liver masks and tumor masks that are couninaud-segmented [39]. They are segregated into different regions, where each region can be tumorous or non-tumorous by nature.

Over time, researchers have explored various modifications and extensions to the original U-Net design to address specific challenges and optimize segmentation performance. In this study, we explore some important variants of the U-Net architecture such as Vanilla U-Net [27], Attention U-Net [40], V-Net [41], U-Net 3+ [42], R2U-Net [43], U^2-Net [44], U-Net++ [45], ResU-Net [46], Swin-UNet [47], and Trans-UNet [48]. These variants of U-Net have emerged as innovative solutions in the field of medical image segmentation.

The remainder of the study is organized as follows. In Section 1.1, we describe the motivation behind this study and the key contributions made here. In Section 1.2, we review some deep learning-based medical image segmentation methods and conclude with the performance of different variants of U-Net architecture. The datasets we use in our experimental study are discussed in Section 2.1. In Section 2.2, we mention the preprocessing steps used that make the images suitable for model training. Variants of U-Net architecture are discussed in Section 2.3. In Section 3, we describe the experimental results and the analysis of the results. In Section 3.9, we focus on the comprehensive comparative analysis of different liver segmentation approaches. The article is briefly summarized in Section 4 along with some insightful suggestions for further research.

1.1 Research gaps, motivation, and contributions

1.1.1 Research gaps

Despite significant advancements in liver segmentation using U-Net and its variants, several research gaps remain. First, the performance of existing methods can degrade in the presence of noise and artifacts in medical images, which is a common issue in real-world clinical settings. Second, the ability of these models to generalize across different imaging modalities and diverse patient populations is often limited, necessitating domain adaptation and transfer learning approaches. Third, while U-Net-based architectures have shown promise, there is a need for more robust methods to handle highly imbalanced datasets where liver lesions are significantly smaller compared to the liver. Finally, there is a lack of comprehensive studies comparing the effectiveness of different U-Net variants on standardized datasets, which hinders the identification of the most effective approaches for specific segmentation tasks.

1.1.2 Motivation

Liver segmentation is considered to be a very important task in the field of medical image analysis. Accurate liver segmentation is needed for the diagnosis of liver diseases like tumors, cirrhosis, and hepatitis. Obtaining the precise location of the liver is highly essential for identifying the characteristics of the liver and planning proper treatment. Segmentation is also required for image-guided procedures like biopsies. This would help doctors navigate instruments accurately within the desired regions. Liver segmentation is also essential for medical practitioners and professionals to understand its anatomy and explain treatment techniques. From the literature survey and above-mentioned research gaps, it can be said that there are various novel U-Net variants that have been proposed and extensively used in different medical image segmentation tasks, including liver segmentation. However, to the best of our knowledge, there is no research article that considers the new variants of the U-Net models and evaluates their performance on liver segmentation from CT images.

In this study, our main **contributions** include the following:

Number of models tested: The research explores the performance of various U-Net architectures, including vanilla U-Net, attention U-Net, V-Net, U-Net 3+, R2U-Net, U^2-Net, U-Net++, ResU-Net, swin-UNet, and trans-UNet. These models are evaluated for their effectiveness in segmenting and localizing the liver region in abdominal CT scans.

Number of datasets considered: Three publicly available datasets, namely 3DIRCADb, combined (CT-MR) healthy abdominal organ segmentation (CHAOS), and liver and tumor segmentation (LiTS) datasets, are used in the research. The choice of datasets aims to measure the performance of the U-Net architectures in liver segmentation across different data sources.

Evaluation metrics considered: The effectiveness of the U-Net models is measured using standard evaluation metrics, including dice similarity coefficient (DSC), recall, precision, $f1$ score, accuracy, intersection over union (IoU), volume overlap error (VOE), and relative volume difference (RVD). These metrics provide a comprehensive assessment of the models' segmentation accuracy.

Useful recommendations provided: Based on the experimental outcomes, the research provides recommendations in several areas: **Model selection:** Considerations include the number of parameters, training time, and other relevant factors. **Datasets:** Insights are provided regarding the suitability of datasets for liver segmentation tasks. **Data modalities:** Recommendations are made concerning the use of different data modalities in the context of liver segmentation. These recommendations aim to guide practitioners and researchers in making informed choices in their future work.

1.2 Literature review

Several deep learning-based segmentation strategies are proposed to improve the liver segmentation task. Jiang et al. [49] proposed an attention hybrid connection network (AHCNet) architecture that aims to leverage the benefits of soft attention's continuous weighting and hard attention's explicit selection, while also utilizing long and short skip connections. The proposed model obtained a 0.945 DSC on the 3DIRCADb dataset. Jiang et al. [50] proposed a residual multi-scale-attention U-Net (RMS-U-Net) that incorporates residual connections to counter gradient vanishing and captures inter-channel relationships among features. Tested on the LiTS [51] and the 3DIRCADb datasets, it produced DSCs of 0.9552 and 0.9697 for liver segmentation, respectively.

Hille et al. [52] introduced a hybrid network named SWTR-U-Net, combining a pre-trained residual network, a few transformer blocks, and a U-Net-like decoder path. It obtained a DSC of 0.98 on the LiTS dataset for the liver. In all these works, the utilization of the attention mechanism increases the network's capacity to record complex anatomical details and variations in the liver structure, producing segmentation results that are more accurate and resilient.

Li et al. [53] proposed a hybrid-variable structure, namely, the RDCTrans U-Net. The backbone comprises of ResNeXt50, which is basically a ResNeXt [54] with 50 layers. Dilated convolutions increase network depth, and transformers are used in down-sampling to increase accuracy. A DSC 0.9338 is obtained for the LiTS dataset. Jiang et al. [55] proposed a multi-dimensional hybrid network, namely the MDCF Net, which had the encoding part comprising CNN and CNNFormer [56]. A slice-by-slice novel feature stacking method is introduced to combine feature maps from encoder and decoder. A DSC of 0.968 is obtained for the segmentation of liver for the LiTS dataset.

Zhang et al. [57] proposed a star-shaped window transformer based on U-Net, namely the SWTRU, where the transformer is incorporated within the decoder section. A filtering feature integration mechanism for dimension reduction is utilized. A DSC of 0.972 is obtained for liver segmentation for the CHAOS + LiTS (CHLISC) dataset. Thereby, the usage of transformers helped to capture wide dependencies and pattern sequences within the image resulting in higher DSC.

Mulay et al. [58] presented a HED-Mask R-CNN approach in order to better identify edge maps from multimodal images. The holistically nested edge detection (HED) network was utilized to obtain an improved edge map from enhanced CT and MRI images, while the mask region-convolutional neural network (RCNN) is used to segment the liver from the improved edge maps. This architecture yields a DSC of 0.94 on CT and 0.91 on MRI images based on the CHAOS challenge dataset. Roy et al. [59] introduced a novel approach where a Mask-RCNN is used to segment the liver region, and then a maximally stable extremal regions network is utilized to identify the tumors in liver. The architecture correctly classified three categories of tumors with an accuracy of 0.878. These approaches based on RCNN are designed to concentrate on specific regions of interest (in our case, the liver) within the CT scans, thereby leading to more accurate segmentation results.

Chi et al. [60] proposed a novel architecture called X-Net for 2D intra-slice feature extraction of liver and tumors by incorporating an up-sampling branch (for identification of liver region) and a pyramid-like convolution network (for extracting inner-liver features) into the backbone dense U-Net. The method uses a joint objective function by combining Dice Loss and a contour-detection-based loss. The architecture achieves a DSC of 0.971 on liver and 0.843 on tumor based on the MICCAI 2017 LiTS Challenge dataset.

Manjunath et al. [61] presented a modified U-Net methodology comprising a 58-layer architecture for LiTS. The proposed model obtains a DSC of 0.9615 on liver and 0.8938 on tumor for the LiTS dataset and 0.9194 on liver and 0.6980 on tumor for the 3DIRCADb dataset. Khan et al. [62] presented a novel architecture called RMS-U-Net incorporating residual blocks and a multi-scale strategy to deal with inter-slice features with multi-channel input images. A combination of DSC and absolute volumetric difference is used as the loss function. The network shows a DSC of 0.9731 on 3DIRCADb, 0.9738 on LiTS, 0.9739 on SLiver07, and 0.9549 on CHAOS datasets for liver segmentation. In all these research works, the use of a U-Net-based approach involving a U-shaped architecture with skip connections enables a more efficient localization of the area of interest, leading to a better segmentation performance.

2 Materials and methods

In this section, a detailed description of datasets used for experimentation, data pre-processing, and the ten U-Net variants applied on the liver CT scan images have been discussed thoroughly.

2.1 Datasets

For this experimental study related to liver segmentation, we have taken into consideration three publicly available datasets, namely, 3DIRCADb, CHAOS, and LiTS. We split each of the datasets into training, validation, and test sets (80% as the training data, 10% for validation data, and 20% as test data). Only the liver masks have been taken into consideration for the purpose of liver segmentation. A mask is essentially used to highlight a desired region of interest (RoI) within a particular image. The number of images for training and testing is shown in Table 1. In this table, we have merged the training and validation as one set.

Table 1: Distribution of train and test data used for experimentation

Dataset	Number of images		Downloadable link
	Train	Test	
3DIRCADb	2,258	565	https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/
CHAOS	2,298	576	https://chaos.grand-challenge.org/
LiTS	15,324	3,832	https://competitions.codalab.org/competitions/17094

2.1.1 3DIRCADb

The 3DIRCADb dataset comprises CT scans of the abdominal area of ten women and ten men with hepatic tumors as well as non-tumorous scans. Images are segregated according to patients. Twenty folders are present for 20 patients, each containing CT images of that patient and their corresponding liver masks. A total of 2,823 CT images and liver masks are present in the dataset. All images are present in DICOM format.

2.1.2 CHAOS

The CHAOS dataset [63] comprises abdominal CT scans of people without any hepatic abnormalities. It contains the CT scans and their corresponding liver masks for 20 people. Images from CT scans are available in DICOM format, while liver masks are present in .png format. The data were collected from the Dokuz Eylul University Hospital, Turkey. CT scans were taken during the portal venous phase after the injection of a contrast agent. During this particular phase, liver tissue experiences its highest level of enhancement due to the increased blood supply from the portal vein. Although the portal veins exhibit strong enhancements, there are also noticeable enhancements observed in the hepatic veins. This phase is widely used for the purpose of segmenting the liver and its associated vessels, which is typically done before surgical interventions. The dataset has 2,874 CT images and liver masks.

2.1.3 LiTS

The LiTS [51] dataset comprises CT scan slices of the liver region. It comprises 131 training and 70 test scans. All the CT volumes and tumor mask slices are present in neuroimaging informatics technology format. The training images are labeled by medical professionals, whereas the test set is unlabeled. However, for research, we have considered only the 131 labeled CT volumes. Each CT volume consists of a series of 2D slices. The number of images per volume ranges from 42 to 1,026. The dataset has 19,156 CT images and liver masks.

2.2 Pre-processing

The CHAOS and 3DIRCADb datasets have images present in Digital Imaging and Communications in Medicine (DICOM) format. This causes a problem for visualizing the images. Thereby, we have converted all images into .jpg format. Furthermore, for both datasets, we have accumulated all CT images into one folder, and the liver masks into another folder. For the LiTS dataset, the images and masks are present in .nii format in the form of slices. They are extracted and also converted to .jpg format which were then splitted just like the other two datasets. Then, we remove all the images and masks where the mask has no RoI, that is, the masks are completely black. This significantly resolves the class imbalance problem.

Before feeding these CT scan images to the pipeline, we implement a windowing technique called HU Clipping [64–66] upon them. The HU is a measurement scale, used especially in CT scans, to express the radio density of various tissues in the body based on their attenuation of X-rays [67–69]. The intensity of X-rays varies as it traverses through matter. Different tissues have characteristic HU ranges, which help radiologists distinguish among various structures and detect abnormalities in a CT scan. The scale considers water as the reference point, assigning it a value of 0 HU. Air [70], which attenuates X-rays very weakly, is assigned a value of $-1,000$ HU, whereas dense materials like bone tissues [71], which strongly attenuate X-rays, can have HU values even above +3,000.

For all the datasets, the window width is opened in the range of [−200, +200], spanning a total of 400 HU $(200 - (-200) = 400)$. By using this clipping range, tissues with HU values between −200 and +200 are effectively highlighted on the CT image, while all other tissues are suppressed as they fall outside the selected window. Therefore, all irrelevant structures are removed, and our area of interest, i.e., the liver stands out from the surrounding viscera in the abdomen. A glimpse of the original CT scans in .jpg format versus the HU clipped images along with their corresponding liver masks for the said datasets is shown in Figure 3.

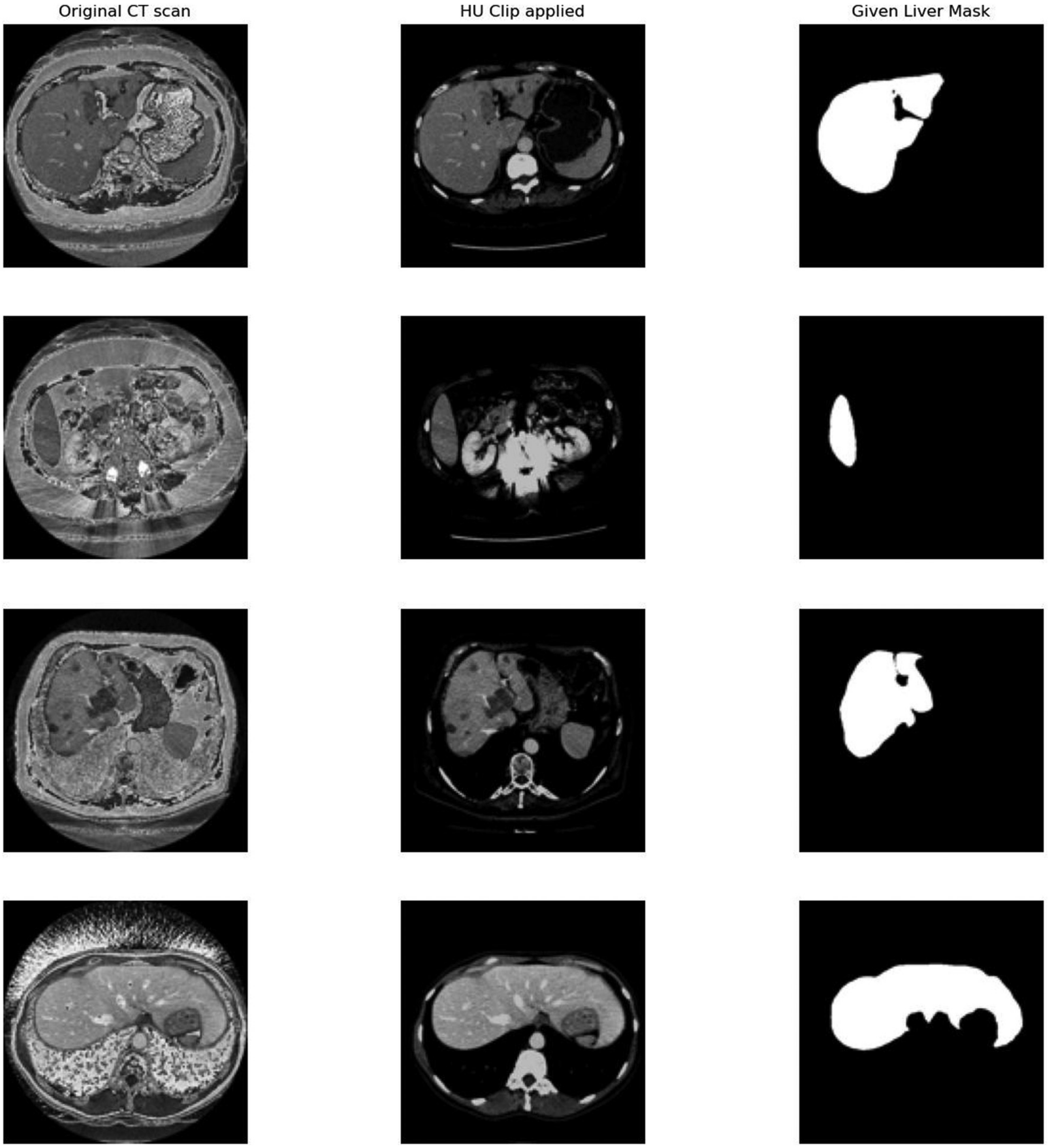


Figure 3: Visualization of the datasets – Column 1: Original CT scans, Column 2: HU-clipped CT images, and Column 3: Given liver masks. The images are examples from publicly available datasets (3DIRCADb, CHAOS, and LiTS) used in this study, where the first and second image is from 3DIRCADb, the third image is from CHAOS and the fourth image from LiTS.

2.3 Model architectures

In this section, we describe the architecture of Vanilla U-Net and its nine other variants that are used here for liver segmentation from CT scans. The pre-processed images are fed into these models for training.

2.3.1 Vanilla U-Net

The U-Net architecture introduced by Ronneberger et al. [27] is a CNN, designed particularly for semantic segmentation problems, wherein a given image needs to be segmented into different regions or classes. The architecture, as shown in Figure 4, consists of a contraction path called encoder and an expansive path called decoder, resembling a “U,” hence the name “U-Net.”

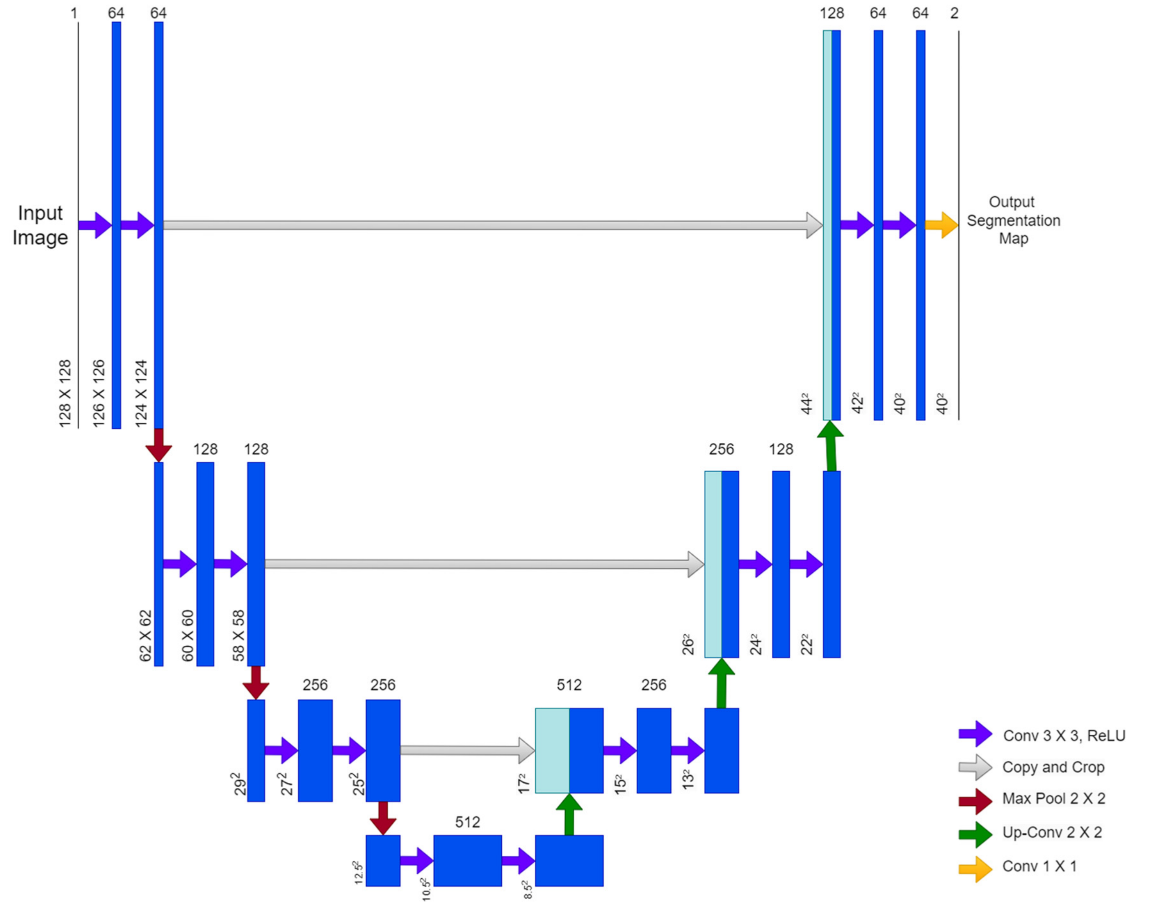


Figure 4: Diagrammatic representation of the U-Net model. Source: Created by the authors.

- **Encoder:** It is the left part of the U-Net model that typically consists of a series of down-sampling blocks, each of which consists of two successive 3 × 3 convolutions along with a rectified linear unit (ReLU) activation unit and then a 2 × 2 max-pooling layer.
  The convolution layers allow the network to learn more complex and abstract features, thereby increasing the number of feature channels, while the pooling layers help to reduce the spatial dimensions of the feature maps while retaining the most important information.
- **Decoder:** The right side of the U-Net contains the decoder, which uses transposed convolutions (or deconvolutions) to gradually up-sample the feature maps, thereby increasing the spatial resolution while reducing the number of feature channels. Each decoder layer consists of an upsampling operation followed by a series of convolutional layers.
- **Skip connections:** The novelty of U-Net lies in the usage of skip connections that link corresponding layers in the encoder and decoder paths. In the case of basic U-Net, skip connections are implemented by concatenating the feature maps from the encoder with the up-sampled feature maps in the decoder.

The final layer of the U-Net is a 1 × 1 convolutional layer followed by an activation function (sigmoid activation in our case for binary segmentation).

2.3.2 Attention U-Net

In order to capture increasingly complex picture properties, CNNs analyze localized information at each layer as they are built up. Rudimentary feature maps collect contextual information, focusing on the segmentation of conspicuous foreground objects. The combination of broad and detailed forecasts for an all-encompassing result is thus made possible by the use of skip connections to combine feature maps obtained at various scales. However, it might be difficult to successfully reduce false positives for small objects with large form fluctuations. Attention gates are used here. They are incorporated within the skip connections. Attention gates do not require an ROI to be cropped between network stages because they gradually reduce feature activation in unrelated background regions. Activation coefficients help to identify the important features and prune the irrelevant ones. The features transmitted over the skip connections are judiciously refined by them. The architecture of attention U-Net [40] is shown in Figure 5.

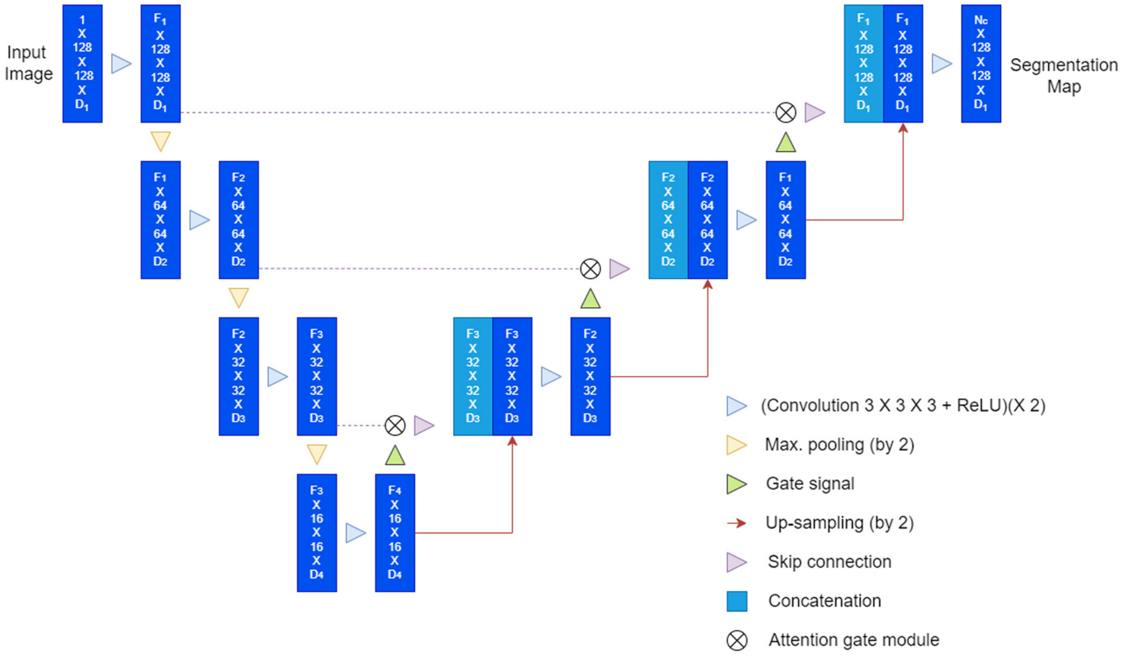


Figure 5: Diagrammatic representation of the Attention U-Net model created by authors based on [40].

2.3.3 V-Net

With a compression pathway on the left and a restoration path on the right, the V-Net [41] architecture uses parametric rectified linear unit (PReLu) non-linear activations [72–74]. The left path comprises 5 × 5 × 5 voxel kernel. However, instead of using max pooling for compression, it uses 2 × 2 × 2 voxel kernel, which aims to extract features by taking into consideration non-overlapping 2 × 2 × 2 patches with stride 2. The right path upsamples the image to its initial configuration. Features extracted from the left path are forwarded to the right path, which ensures that no data are lost during compression. The two feature maps of the last layer of the network having 1 × 1 × 1 kernel size and producing output with the same size as that of the input are converted to probabilities by applying the Softmax function [75,76].

2.3.4 U-Net 3+

The U-Net 3+ [42] architecture's usage of skip connections and deep supervision result in the segmentation map's position-awareness, and the border identification gets increased while efficiency with fewer parameters is maintained. The architecture of U-Net 3+ is shown in Figure 6. Both intra-connections among decoder sub-networks and the encoder–decoder connections are changed by full-scale skip connections in U-Net 3+. A fusion of feature maps from various scales, including smaller and same-scale maps from the encoder and larger-scale maps from the decoder, is made and incorporated into each decoder layer. Through this integration, the model can collect minute features and an extensive semantic context at all scales, resulting in a thorough knowledge of the data. For the implementation of deep supervision, the final layer of each decoder stage undergoes a sequence of operations: first passing through a standard 3 × 3 convolutional layer, then undergoing bilinear up-sampling, and finally being subjected to a sigmoid function [77,78].

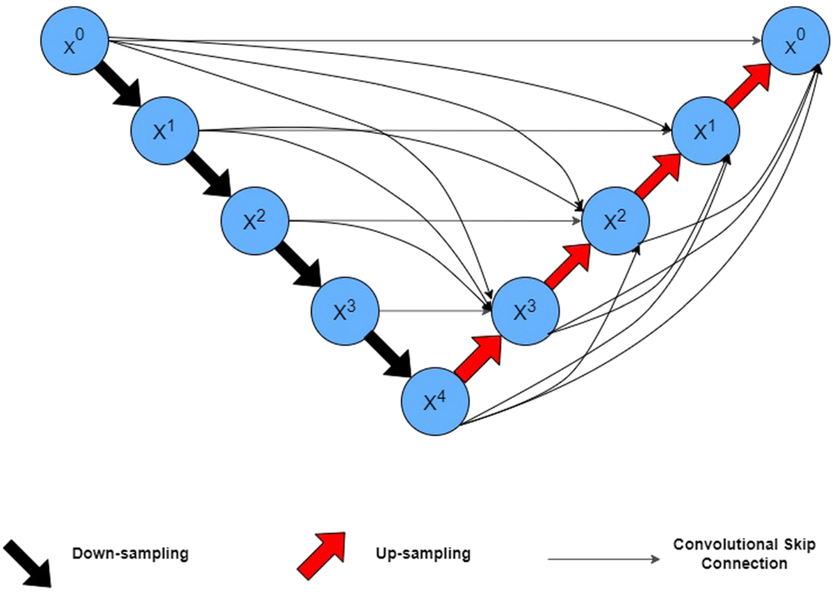


Figure 6: Diagrammatic representation of the U-Net 3+ model based on [42].

2.3.5 R2U-Net

Recurrent residual convolutional neural network based on U-Net (R2U-Net) [43] combines the potential of the vanilla U-Net, recurrent-CNN (RCNN) and residual units. This architecture consists of typical encoder and decoder pathways as Vanilla U-Net but replaces the forward convolutional blocks with recurrent convolutional layers with residual units in both the encoder and the decoder. The outputs from the recurrent convolutional blocks are passed through a residual block for a more efficient and deeper feature accumulation. This architecture design, in spite of having lesser parameters in comparison to the traditional U-Net, ensures better feature extraction using the recurrent residual convolutional layers. The model's efficiency in terms of the number of network parameters can be attributed to the fact that the recurrent operations do not consume additional parameters. The architecture of R2U-Net is shown in Figure 7.

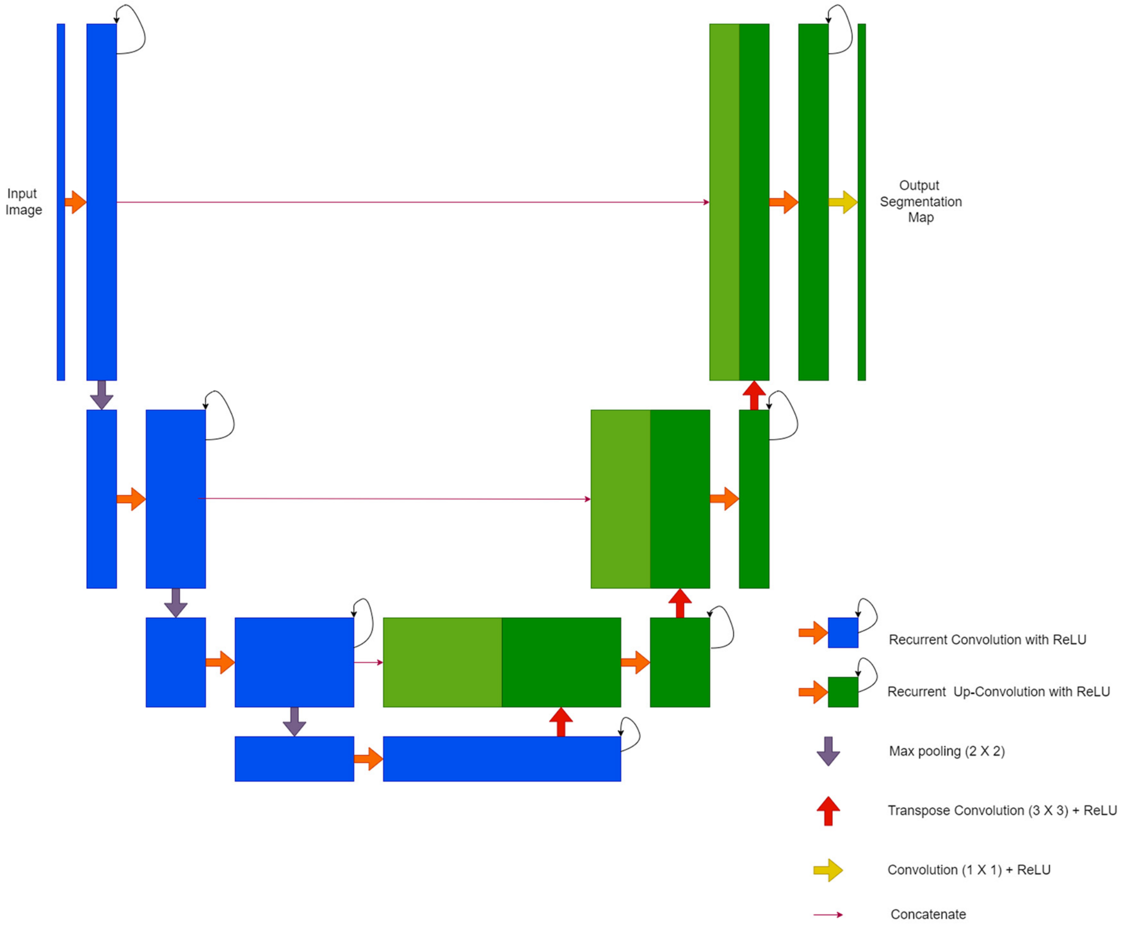

Input Image
Output Segmentation Map
Recurrent Convolution with ReLU
Recurrent Up-Convolution with ReLU
Max pooling (2 X 2)
Transpose Convolution (3 X 3) + ReLU
Convolution (1 X 1) + ReLU
Concatenate

Figure 7: Diagrammatic representation of the R2U-Net model created by authors based on [43].

2.3.6 U^2-Net

The U^2-Net architecture [44] is an extension of the U-Net architecture. The architecture of the U^2-Net is shown in Figure 8. It introduces the concept of a two-level nested U-structure, namely the top and bottom levels. The bottom-level basic building block comprises a novel ReSidual U (RSU) block. The RSU block contains a symmetric encoder and decoder architecture that helps in capturing multi-scale feature information alongside local feature information. These basic RSU building blocks have been incorporated into the encoder blocks colored green on the diagram. The decoder blocks, colored pink, have a symmetric architecture corresponding to its encoder block and take as input the upsampled feature maps from the decoder path and the feature maps obtained from its corresponding encoder block and concatenates them. These decoder blocks would generate gray-colored saliency maps, which are concatenated and passed through a 1 × 1 convolution and sigmoid activation to generate the final output, S_{fuse}.

2.3.7 U-Net++

U-Net++ [45] aims to enhance the segmentation performance of U-Net by incorporating nested skip pathways and multi-scale contextual information. U-Net++ retains the U-shaped architecture of the original U-Net, but the skip connections are extended to create nested skip pathways. Here, the encoder blocks and decoder blocks are connected through a series of nested densely connected convolutional blocks, which helps to bridge the

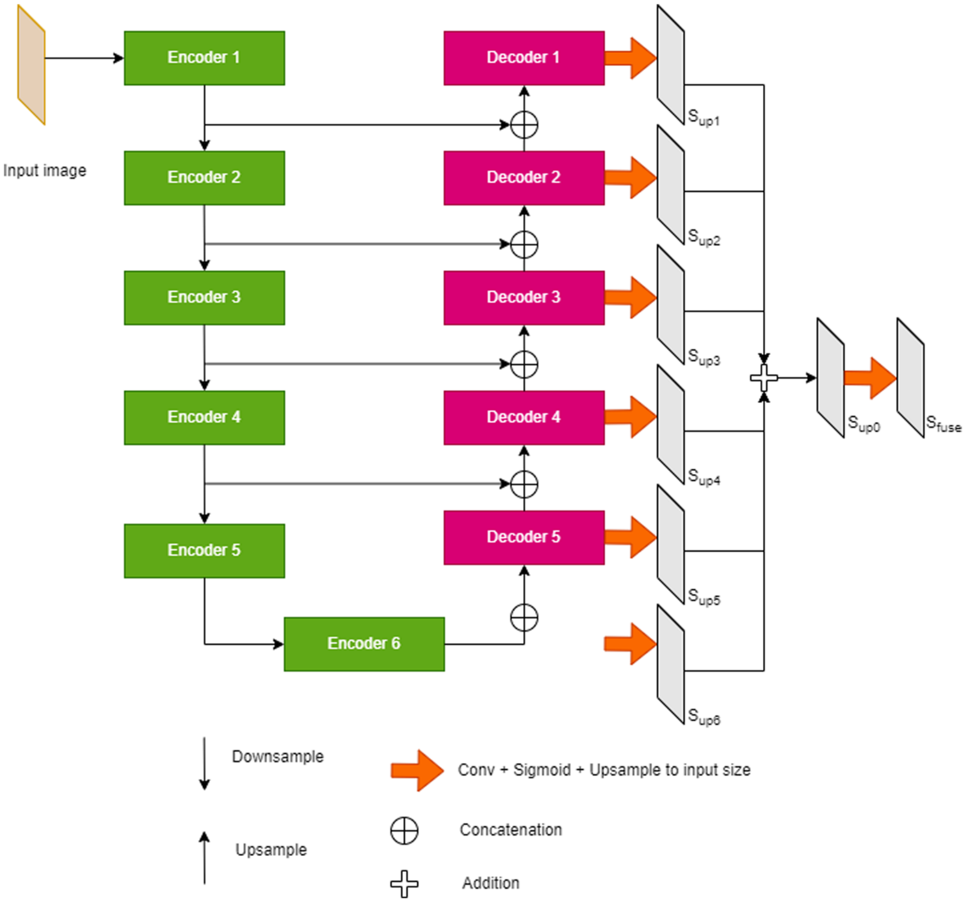


Figure 8: Diagrammatic representation of U^2-Net model created by authors based on [44].

semantic gap between the encoder and the decoder. This hierarchical decoder with nested skip connections enables the integration of features from various scales, enhancing detail preservation and capturing important contextual information. U-Net++ addresses scale variability and improves segmentation accuracy by hierarchically aggregating multi-scale information, making it suitable for diverse object sizes and scene complexities. The architecture of U-Net++ is shown in Figure 9.

2.3.8 Res U-Net

Res U-Net [46], an extension of the U-Net architecture, enhances semantic segmentation tasks by introducing residual connections within encoder and decoder blocks. Inspired by ResNets, these connections facilitate better gradient flow during training, improving feature propagation throughout the network. This integration of residual connections allows Res U-Net to efficiently capture multi-scale contextual information, enabling accurate segmentation of objects with varying sizes and complexities. By addressing challenges such as gradient vanishing and feature representation, Res U-Net proves to be a robust model for semantic segmentation tasks across diverse datasets and scenarios. The architecture of the Res U-Net is shown in Figure 10.

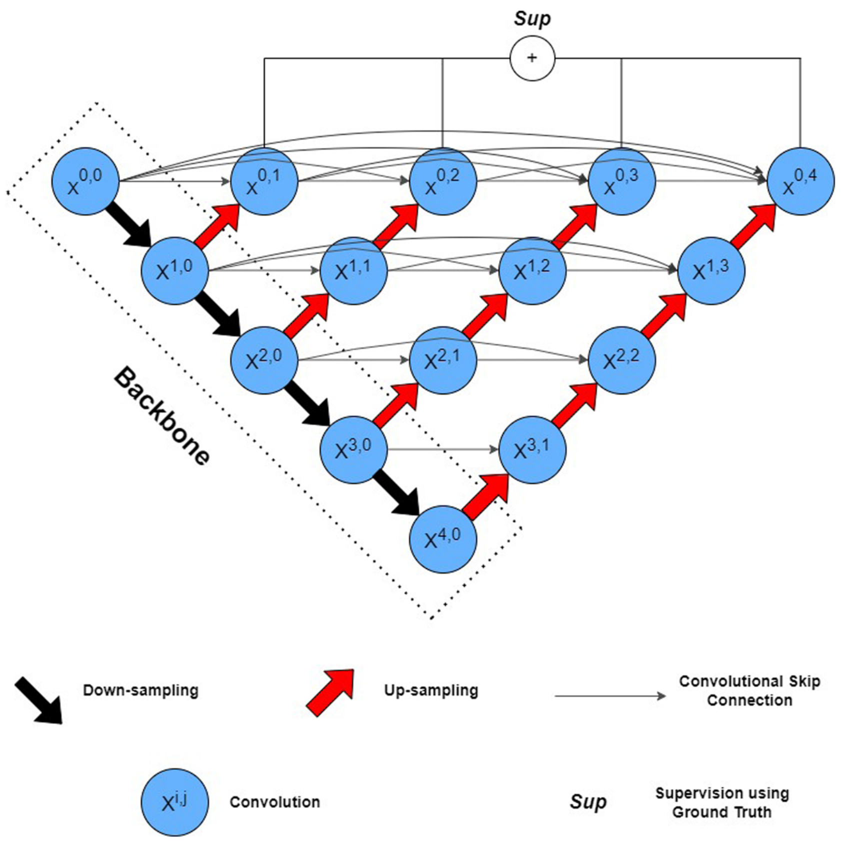


Figure 9: Diagrammatic representation of U-Net++ model created by authors based on [45].

2.3.9 Swin-UNet

The swin-UNet architecture [47] is a combination of sliding window-based transformer [79], a transformer-based architecture, and U-Net used for segmentation tasks. The architecture's central component is the swin transformer, as shown in Figure 11. This hierarchical transformer model successfully captures contextual information, both local and global. With the hierarchical design, an image is divided into non-overlapping patches at various sizes, each of which is processed separately before self-attention mechanisms enable interactions between the patches. The self-attention mechanism of the Swin transformer is utilized to capture long-range dependencies and enhance feature representations. The architecture is made to take advantage of U-Net's skip connections to preserve high-resolution details and Swin transformer's attention techniques to capture global context. The mask is computed across different layers and compared with the concatenated mask for computing the final loss as depicted in the figure.

2.3.10 Trans-UNet

Like other hybrid designs, Trans-UNet [48] aims to maximize the advantages of UNet and transformer-based models to achieve better results in medical image segmentation tasks. The encoder part of the model consists of transformer block that includes self-attention mechanisms to efficiently model relationships between different parts of the input image. The decoder part integrates higher-level encoder characteristics with lower-resolution feature maps while maintaining spatial information by using skip connections and upsampling layers. Trans-UNet uses a cascaded upsampler in addition to a hybrid CNN-transformer architecture as the encoder to provide accurate localization.

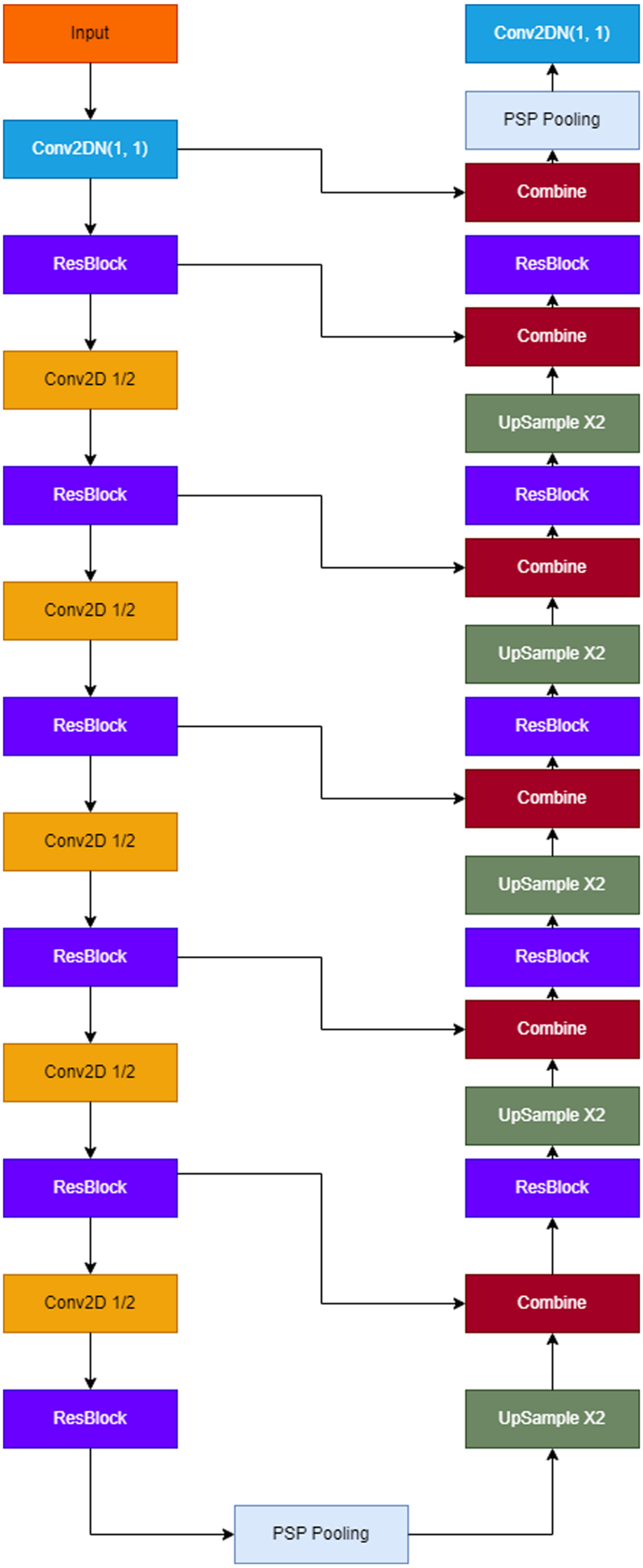


Figure 10: Diagrammatic representation of the ResU-Net model created by authors based on [46].

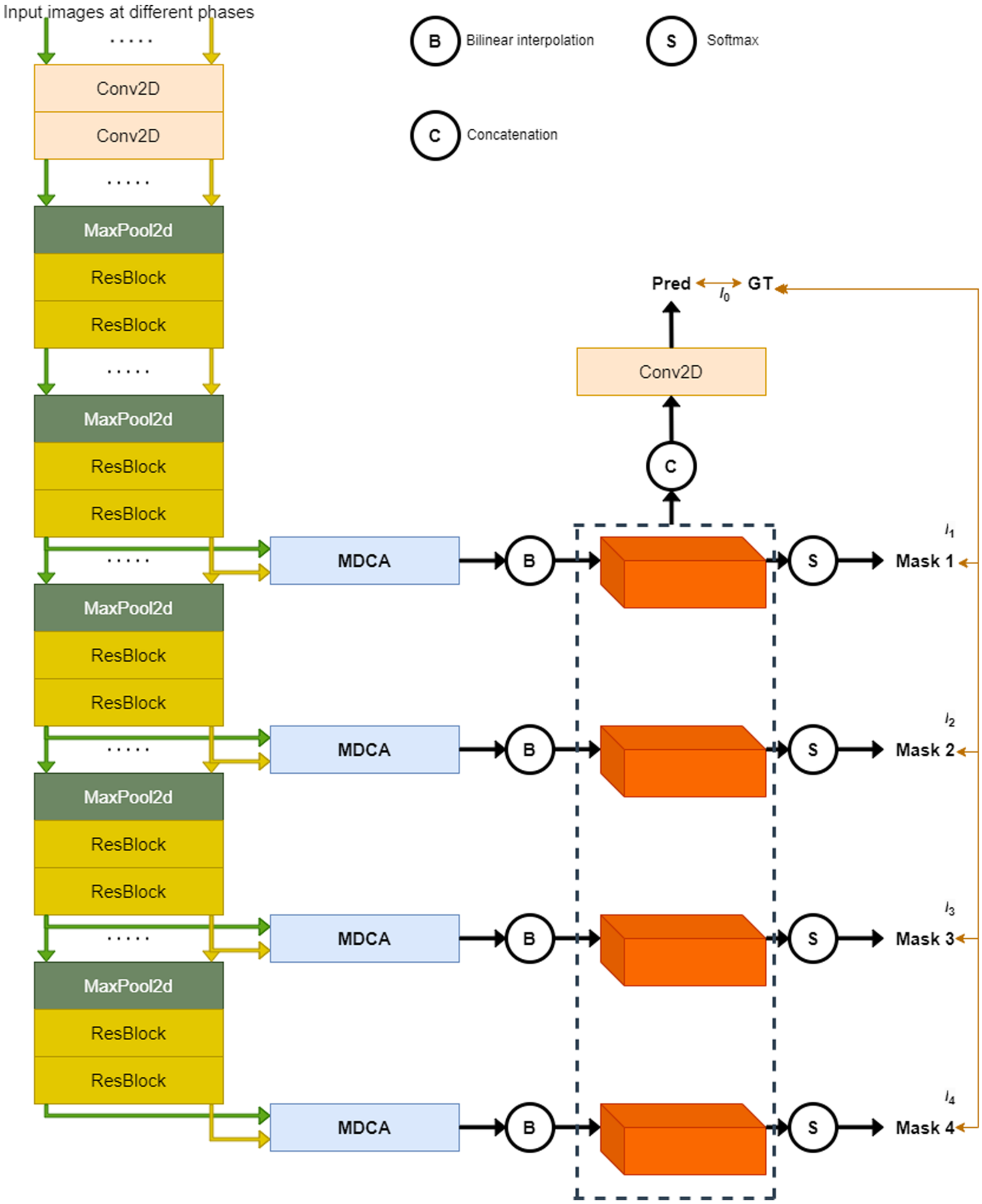


Figure 11: Diagrammatic representation of the Swin-UNet model created by authors based on [47].

3 Experimentation

A wide range of evaluation metrics have been considered for this comparative study. In this section, we discuss the results obtained on the three publicly available liver CT image datasets while applying the ten different U-Net variants. We also visualize the outputs obtained using GradCAM.

3.1 System configuration

A Jupyter Notebook with a 12 GB NVIDIA Tesla T4 GPU that is made available through Google's collaborative environment is used to run the entire set of experiments. Tensorflow, Keras, Matplotlib, Scikit, Numpy, and Pandas are the main open-source modules that are used in the tests to implement the U-Net variants in Python 3.

3.2 Evaluation metrics

Metrics for evaluation provide us with the ability to determine the predictive or learning potential. It tells us how well the system performs in predicting the proper class label. In regard to the segmentation task, we consider this as a pixel-wise two-class classification problem. For two labels, namely, "Positive" and "Negative," we have the following:

- **True positive** (T_P) refers to the number of accurately categorized positive pixels.
- **False positive** (F_P) refers to the number of negative pixels that are mistakenly labeled as positive.
- **False negative** (F_N) refers to those pixels that are mistakenly labeled as negative despite being positive.
- **True negative** (T_N) refers to the number of accurately categorized negative pixels.

In this comparative study, we have used the following metrics:

Dice score coefficient (DSC): DSC [50], also known as the Dice coefficient or Sørensen-Dice coefficient, is the similarity between the discovered region and the actual ground truth region. It entails calculating the ratio of the two regions' intersections to the sum of each region's individual areas. Higher values of the DSC suggest better overlap and accuracy and shed light on the degree of agreement between the anticipated and true regions.

Recall: Recall [80], often called the true positive rate or sensitivity, is a performance metric that measures a model's or algorithm's capacity to properly identify every pertinent occurrence present in a dataset. It is determined by dividing the total number of true positives (relevant cases that the model properly detected) by the total number of true positives and false negatives (relevant examples that the model missed). Recall measures the thoroughness of the model's output, and a high number is desired to reduce the possibility of missing important occurrences.

Precision: Precision [80] measures the accuracy of a model to correctly identify instances as relevant. It is determined by dividing the total number of relevant instances by the sum of relevant examples and false positives. The model's ability to prevent false positives and generate precise positive predictions is highlighted by precision.

***F*1 score:** The *F*1 score [80] is a metric that balances both precision and recall. It is calculated as two times the product of precision and recall divided by their sum. It is the harmonic mean of accuracy and recall. When the trade-off between precision and recall needs to be taken into account, the *F*1 score offers a fair assessment of a model's performance.

Accuracy: Accuracy [80] is a fundamental performance metric that calculates the ratio of accurately predicted cases (including true positives and true negatives) to all cases in the dataset. Although it provides a broad overview of a model's accuracy, it might not be appropriate for datasets with imbalances, where one class considerably outnumbers the other. Accuracy can be deceiving in certain situations, so other metrics such as precision, recall, and *F*1 score are frequently favored.

IoU: IoU [81], commonly referred to as the Jaccard index, is a metric widely used in the field of computer vision and image segmentation. By dividing the area of their intersection by the area of their union, it calculates the amount of overlap between two sets, often bounding boxes or regions. In essence, by measuring how much two items or regions coincide in comparison to how much they diverge, the IoU quantifies the degree of agreement or resemblance between them. It is frequently used to analyze how well object detection and segmentation algorithms perform, determining how closely the results they anticipate match the actual data.

VOE: VOE [50] is another measure that is frequently used in the field of medical image analysis. Its objective is to assess the accuracy of segmentation results. By calculating the percentage of the total volume of these regions that are not shared by their intersection, VOE analyzes the discrepancy between a segmented area and its matching ground truth. In essence, it measures how much the segmented region and the real region are out of alignment or in error. VOE's value equals (1 − IoU) when calculated as the complement of the IoU. A lower VOE value indicates more segmentation accuracy, whereas a value of 0 indicates perfect matching.

RVD: RVD [50] is a measure that gauges how much a segmented region's volume differs from the related ground truth. Finding the absolute difference between the segmented area's volume and the real region's volume, then dividing this difference by the real region's volume, are the steps in this calculation. RVD throws light on how much the segmented region differs from the actual region in terms of volume. Higher values imply more significant volume discrepancies between the segmentation and the ground truth, whereas values closer to 0 suggest a small volume distinction.

3.3 Loss function

We have considered Dice Loss [82] as the loss function for this comparative study. It is a typical loss function in image segmentation problems. It gauges how differently the expected segmentation mask compares with the actual segmentation mask. When working with unbalanced datasets or attempting to prioritize precise object segmentation while penalizing false positives and false negatives, dice loss is especially helpful. The Dice Loss is defined as the complement of the Dice coefficient, aiming to be minimized during the training of a model:

$$\text{Dice Loss} = 1 - \text{DSC}. \tag{1}$$

In the context of neural network training, the goal is to minimize the Dice Loss, encouraging the model to produce segmentations that closely match the ground truth.

3.4 Hyperparameters

Table 2 shows the values of different hyperparameters that we used for this comparative study. It should be noted that we have used the backbone as VGG16 [83] for the Vanilla U-Net and Attention U-Net models. This is mainly because for these two models, the DSC values that we obtained were less without the backbone. A backbone implies the base architecture used to extract the features of an image. The VGG16 architecture has already been trained on the large-scale ImageNet dataset. Therefore, the learned features from ImageNet help to capture the inherent feature information from the images in the dataset. The NAdam optimizer [84] used essentially combines the properties of the Adam optimizer [85] and the Nesterov momentum [84,86,87]. This helps to achieve faster convergence, adapt to the learning rates for individual parameters, and handle sparse gradients.

Table 2: Hyperparameters of the models

Hyperparameter	Value/name
Optimizer	NAdam [84]
Backbone	VGG16 [83]
Loss function	Dice Loss
Learning rate	0.0001
No. of epochs	70

3.5 Results

In this section, we present and discuss the findings, as well as thoroughly examine the assessment standards that are used in this study. Table 3 provides approximate parameter counts for the variants of U-Net architecture that are used in our experimental study. The trade-off between model complexity and computational resources must be taken into account when selecting U-Net variants because larger models require more memory and processing capacity for training and inference.

Table 3: Total number of parameters for each model

Model	Parameters
Vanilla U-Net [27]	31,172,547
Attention U-Net [40]	28,384,871
V-Net [41]	58,903,363
U-Net3+ [42]	7,875,651
R2U-Net [43]	23,054,531
U^2-Net [44]	55,375,439
U-Net++ [45]	9,042,243
ResU-Net [46]	**1,947,999**
Swin U-Net [47]	8,303,435
Trans U-Net [48]	208,206,211

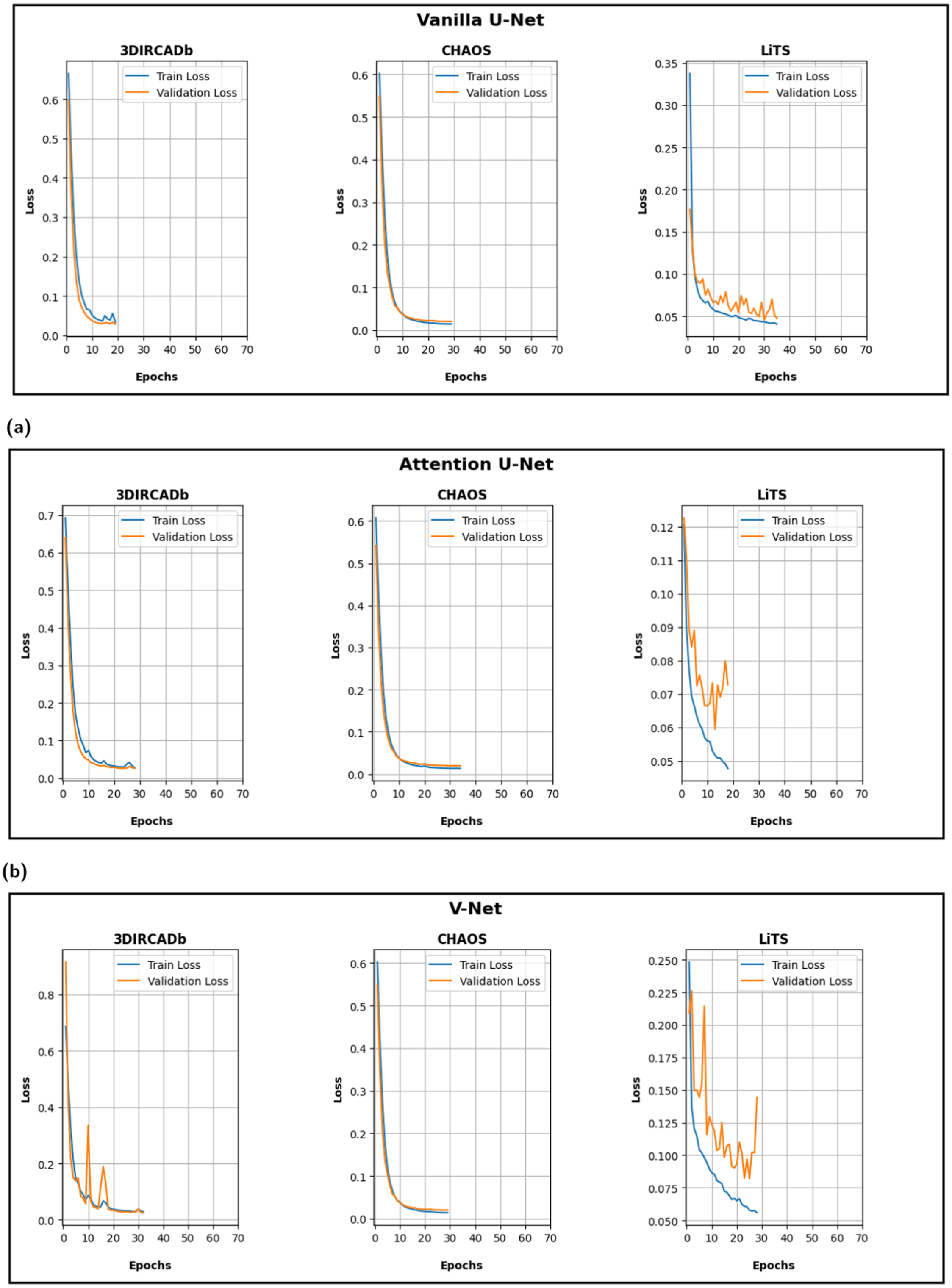
Bold value indicates the lowest number of parameters.

The loss curves for liver segmentation on 3DIRCADb, CHAOS, and LiTS datasets corresponding to all the U-Net variants are shown in Figure 12. Since early stopping is employed during the training process with a patience of 0.001 over the validation loss, we find that for most of the architectures, the loss curves do not span for the entire 70 epochs. We can observe that there is rarely any overfitting in the models from looking at their loss curves. Additionally, the fact that the training loss curve is decreasing with an acceptable steep slope, we can say that the models are easily able to learn the proper features reducing the loss in the process. Most of the loss curves are quite stable apart from the R2-UNet model. In the case of the R2U-Net model, we can see that there are several spikes in the validation loss curve for 3DIRCADb and LiTS datasets. However, it is performing better on the CHAOS dataset.

The observed differences in performance between the three datasets prompt further investigation into the underlying factors contributing to these variations. While the 3DIRCADb and LiTS datasets may pose greater challenges, possibly due to data irregularities or class imbalances, the CHAOS dataset appears to offer a more favorable environment for this model, potentially due to its data quality or distribution characteristics.

The experimental results for liver segmentation on 3DIRCADb, CHAOS, and LiTS datasets are shown in Tables 4, 5, and 6, respectively. Here, we analyze and interpret the experimental results obtained from our study on various U-Net variants for liver segmentation. On the basis of these results, we find that all variants of U-Net show reasonably good performance on liver segmentation in terms of DSCs. The obtained DSCs lie between 0.6304 (ResU-Net) and 0.9736 (Attention U-Net and V-Net) in the case of the 3DIRCADb dataset, 0.9056 (ResU-Net) and 0.9809 (Attention U-Net) in the case of the CHAOS dataset, and 0.5516 (ResU-Net) and 0.9545 (Vanilla U-Net in case of the LiTS dataset. Thus, it is evident that Attention U-Net is the best performing model, suggesting the fact that attention mechanisms could greatly benefit liver segmentation tasks that require precise organ delineation, as it effectively highlights relevant structures even in cases of low contrast with the surrounding viscera on abdominal CT scans. Table 7 throws light on a few analytic points and fallacies observed from the results so obtained.

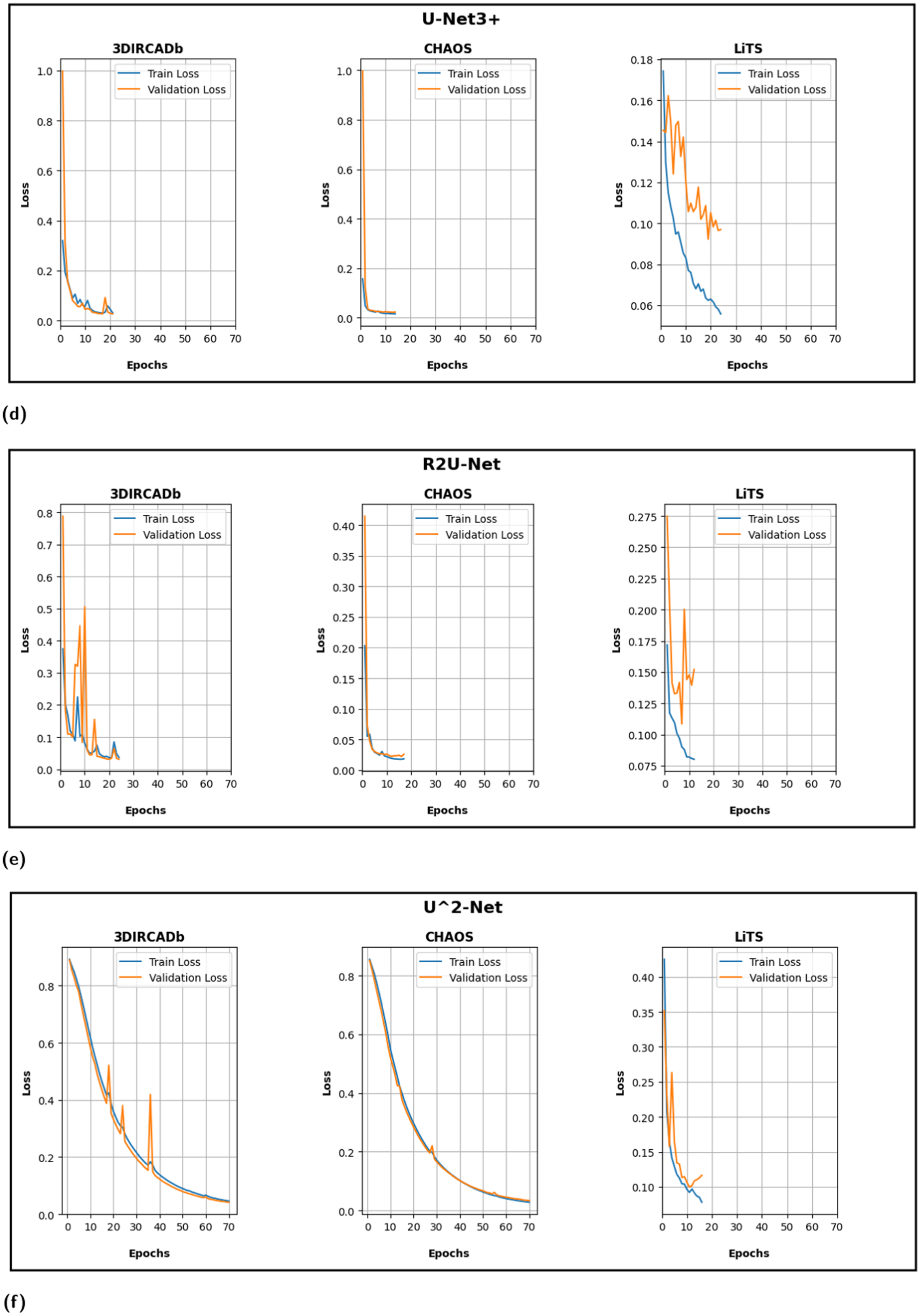
However, when it comes to parameter size, Attention U-Net is quite a heavy architecture with about 28 million parameters. Notably, U-Net 3+ is the lightest of all architectures in our experiment with only about 7.8 million parameters but still shows a comparative performance of 0.9725 DSC in the case of the 3DIRCADb


(a)

(b)

(c)

Figure 12: Loss curves corresponding to (a) Vanilla U-Net, (b) Attention U-Net, (c) V-Net, (d) U-Net 3+, (e) R2U-Net and (f) U^2-Net, (g) U-Net++, (h) ResU-Net, (i) Swin-UNet, and (j) Trans-UNet for 3DIRCADb, CHAOS and LiTS datasets. The curves depict training and validation loss, highlighting model convergence, stability, and comparative learning dynamics for liver segmentation. Source: Created by the authors.

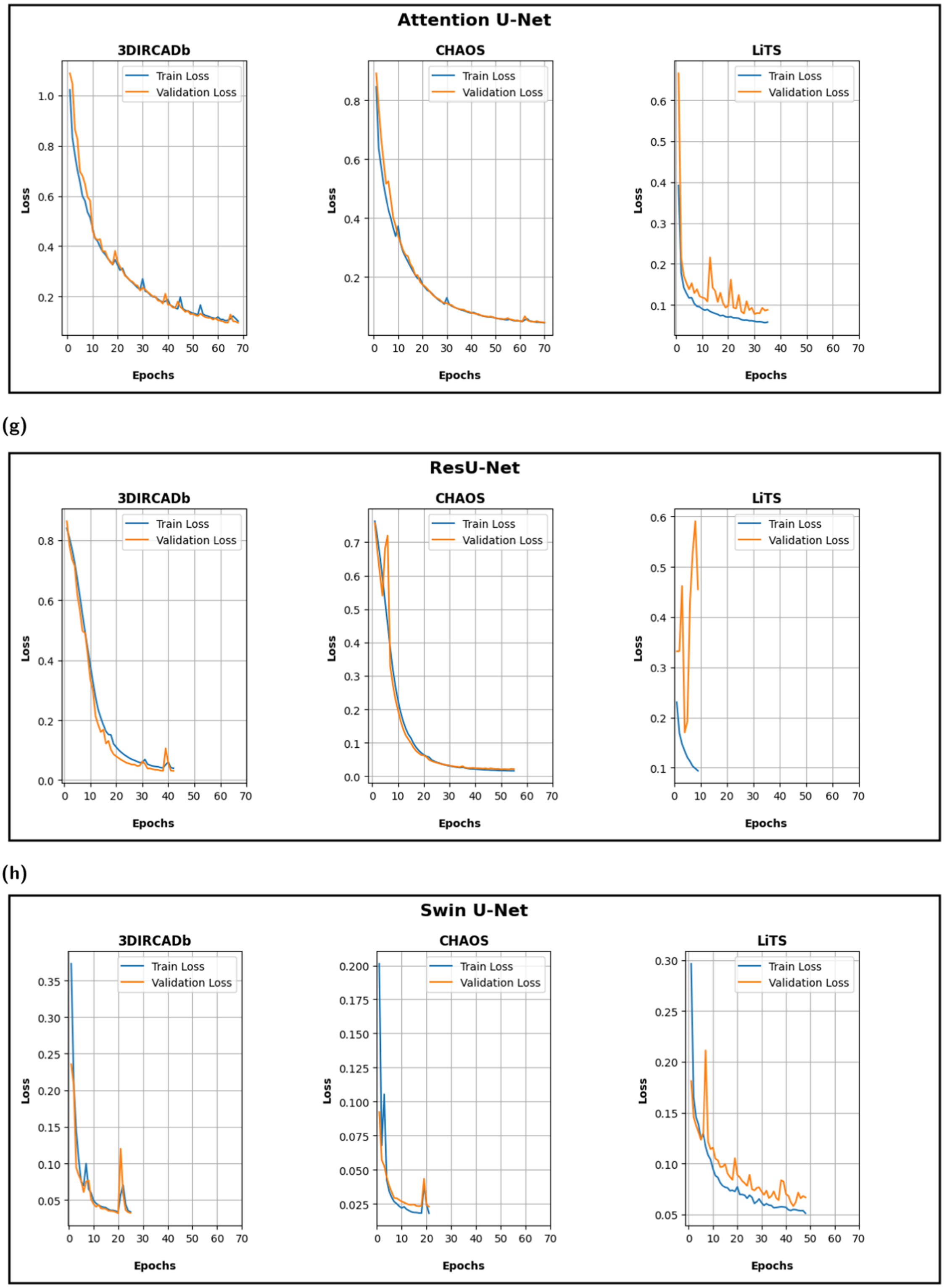


(d)

(e)

(f)

Figure 12: (*Continued*)

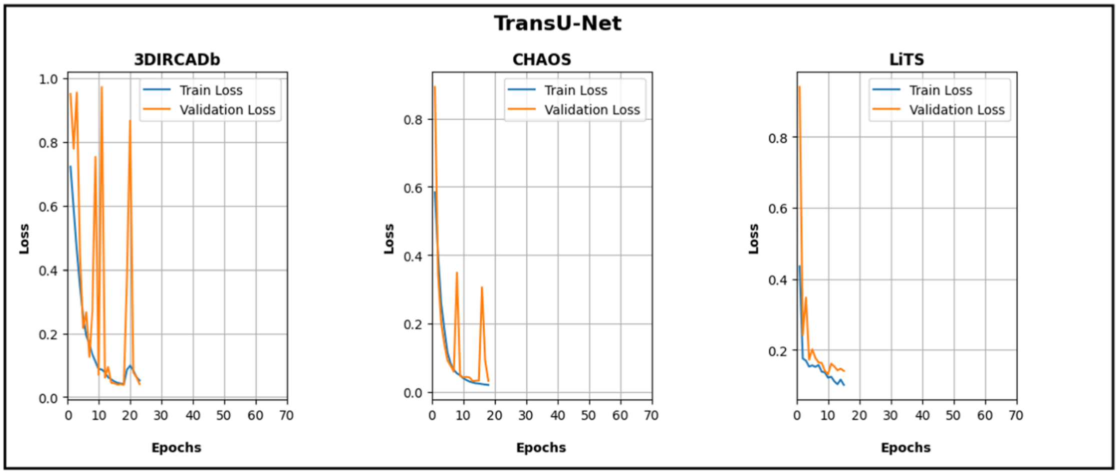


(g)

(h)

(i)

Figure 12: (*Continued*)



(j)

Figure 12: (*Continued*)

dataset and 0.9768 DSC in the case of the CHAOS dataset. The use of full-scale skip connections and full-scale deep supervision in U-Net 3+ results in faster feature accumulation in spite of fewer parameters, which makes it computationally economical.

We have shown the resulting metric scores for the test dataset in the form of bar graphs in Figure 13. We visualize the performance of the models in the form of bar charts specific to the datasets and evaluation metrics of interest. Here, we show accuracy, precision, recall, and $F1$ score for the three datasets. A few good segmented output masks are shown in Figure 14.

3.6 Qualitative analysis

In this section, we analyze the model performances in terms of some specific aspects. In Table 7, we consider four key points that are performance comparison, boundary accuracy, anatomical variations, and efficiency and convergence. From the experimental results, it is observed that the Attention U-Net model achieves the highest DSC as compared to the other U-Net variants showing higher segmentation accuracy. On the 3DIRCADb dataset, the V-Net model also achieves the same DSC score as the Attention U-Net. However, for the LiTS

Table 4: Segmentation performance of different U-Net variants on the 3DIRCADb dataset

Model	DSC	IoU	VOE	RVD
Vanilla U-Net [27]	0.9715	0.9447	0.0553	0.0079
Attention U-Net [40]	**0.9736**	**0.9487**	**0.0513**	**0.0113**
V-Net [41]	**0.9736**	0.9486	0.0514	0.0169
U-Net 3+ [42]	0.9725	0.9466	0.0534	0.0082
R2U-Net [43]	0.9666	0.9355	0.0645	0.0191
U^2-Net [44]	0.9572	0.9182	0.0818	0.0042
U-Net++ [45]	0.9690	0.9400	0.0600	0.0066
ResU-Net [46]	0.6304	0.5589	0.4411	0.0056
Swin U-Net [47]	0.9669	0.9361	0.0639	0.0062
Trans-UNet [48]	0.9575	0.9191	0.0809	0.0126

Bold values indicate best scores.

Table 5: Segmentation performance of different U-Net variants on the CHAOS dataset

Model	DSC	IoU	VOE	RVD
Vanilla U-Net [27]	0.9797	0.9603	0.0397	0.0059
Attention U-Net [40]	**0.9809**	**0.9625**	**0.0375**	**0.0014**
V-Net [41]	0.9800	0.9609	0.0391	0.0015
U-Net 3+ [42]	0.9768	0.9547	0.0453	0.0110
R2U-Net [43]	0.9744	0.9502	0.0498	0.003
U^2-Net [44]	0.9662	0.9348	0.0652	0.0023
U-Net++ [45]	0.9766	0.9544	0.0456	0.0029
ResU-Net [46]	0.9056	0.8707	0.1293	0.0014
Swin-UNet [47]	0.9765	0.9541	0.0459	0.0107
Trans-UNet [48]	0.9660	0.9345	0.0655	-0.0312

Bold values indicate best scores.

Table 6: Segmentation performance of different U-Net variants on the LiTS dataset

Model	DSC	IoU	VOE	RVD
Vanilla U-Net [27]	**0.9545**	**0.9153**	**0.0847**	**0.0103**
Attention U-Net [40]	0.9322	0.8814	0.1186	0.03574
V-Net [41]	0.8575	0.7748	0.2252	0.0498
U-Net 3+ [42]	0.6249	0.5996	0.4125	0.1053
R2U-Net [43]	0.8517	0.7723	0.2277	0.0158
U^2-Net [44]	0.8858	0.8125	0.1875	0.0587
U-Net++ [45]	0.9103	0.8485	0.1515	0.0327
ResU-Net [46]	0.5516	0.4379	0.5621	0.5167
Swin-UNet [47]	0.9352	0.8821	0.1179	0.0182
Trans-UNet [48]	0.8632	0.7768	0.2232	0.1104

Bold values indicate best scores.

dataset, the Vanilla U-Net achieves the highest DSC, while the Attention U-Net yields nearly comparable results. Capturing intricate boundaries and producing smooth and coherent segmentation results on the datasets, the U^2 -Net model produces fairly high-quality segmentation results.

It is crucial to develop segmentation algorithms that are robust and adaptable to various anatomical variants. This involves not only accurate boundary delineation but also the ability to handle irregular shapes, tissue textures, and vascular variations. Among the other U-Net variants, the Attention U-Net model is

Table 7: Analysis of the liver segmentation results over the three datasets on certain aspects

Discussion on	Dataset	Model	Remarks
Performance comparison	3DIRCADb	Attention U-Net and V-Net	Achieves the highest DSC,
	CHAOS	Attention U-Net	indicating superiority in
	LiTS	Vanilla U-Net	overall segmentation accuracy.
Boundary accuracy	3DIRCADb	U^2-Net	Excels in capturing
	CHAOS	Attention U-Net	intricate boundaries, resulting in
	LiTS	U^2-Net	smooth and coherent segmentation.
Anatomical variations	3DIRCADb	Attention U-Net	Showcases adaptability to
	CHAOS	Attention U-Net	diverse anatomies, accurately segmenting
	LiTS	Attention U-Net	liver images with varying shapes and sizes.
Convergence	3DIRCADb	Vanilla U-Net	Displays rapid convergence and requires
	CHAOS	U-Net3+	fewer training epochs, making it favorable
	LiTS	ResU-Net	for resource-constrained setups.

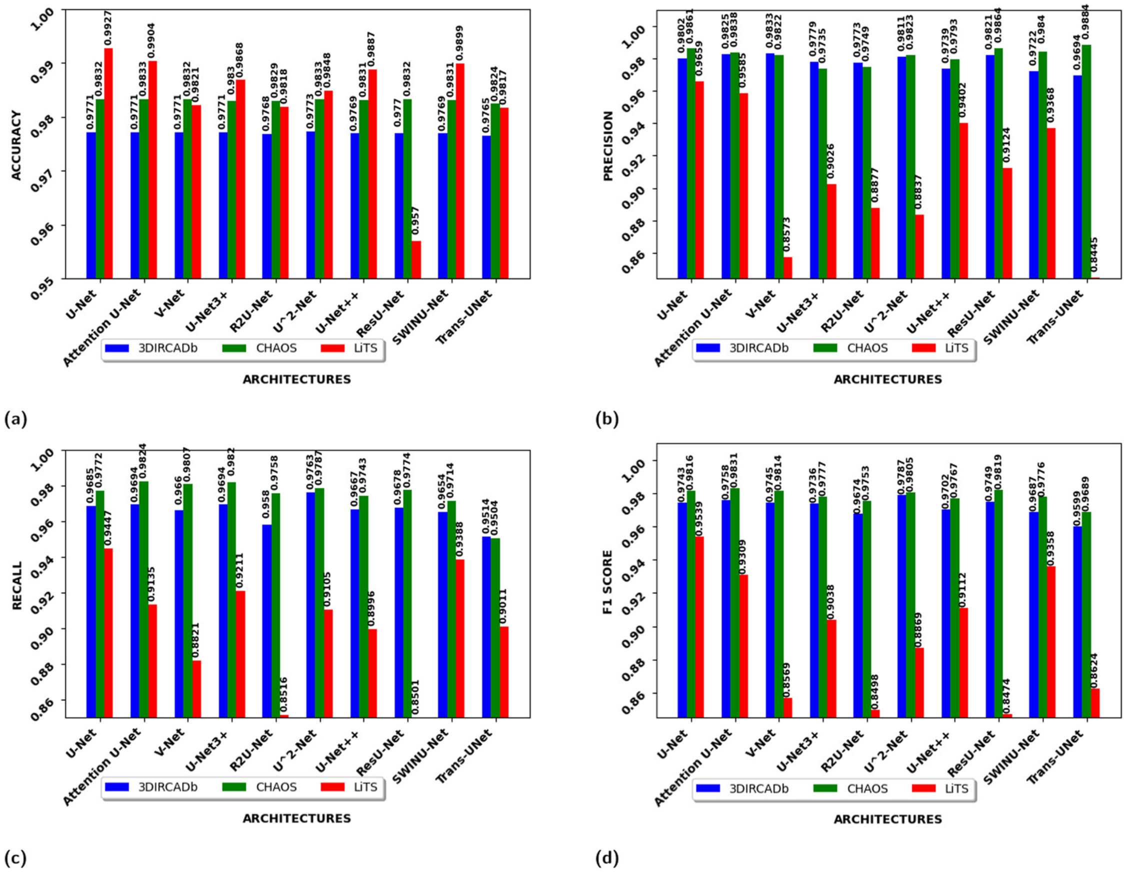


Figure 13: Performance comparison of the different U-Net models for liver segmentation in terms of: (a) Accuracy, (b) Precision, (c) Recall, and (d) *F*1 score. The metrics highlight the effectiveness of each model in accurately segmenting liver regions across the evaluated datasets. Source: Created by the authors.

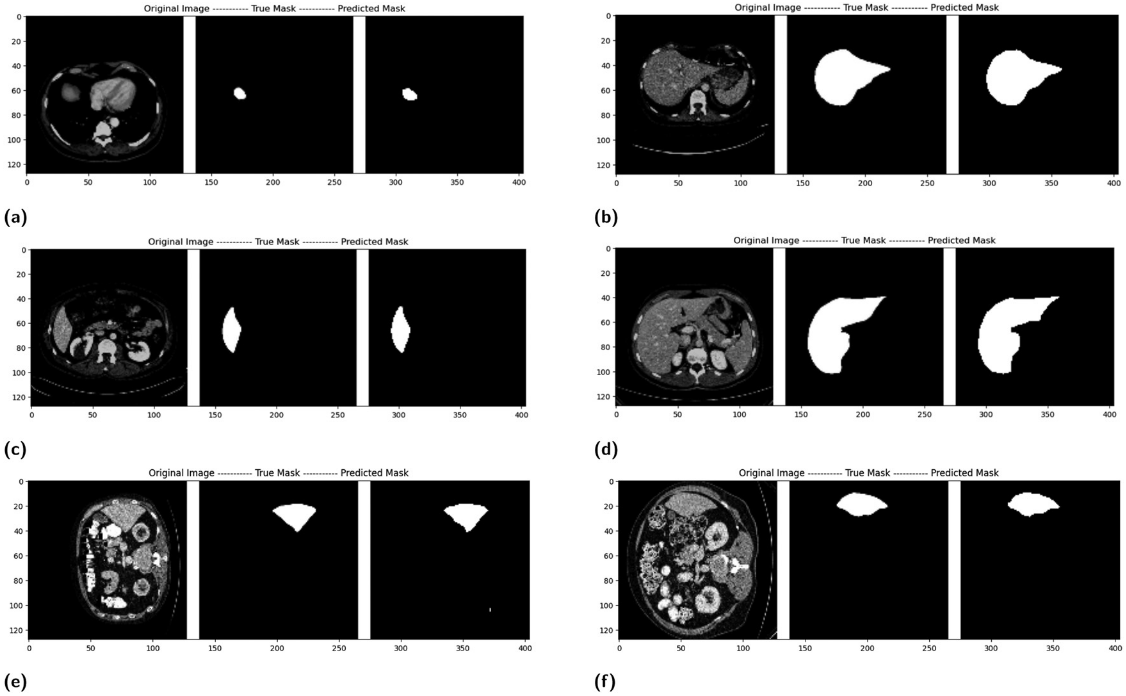

Original Image ---------- True Mask ---------- Predicted Mask
0
20
40
60
80
100
120
0
50
100
150
200
250
300
350
400
(a)
(b)
(c)
(d)
(e)
(f)

Figure 14: Examples of successful predictions for liver segmentation: (a) Vanilla U-Net and (b) Attention U-Net on the 3DIRCADb dataset, (c) U-Net3+ and (d) U-Net++ on the CHAOS dataset, and (e) U^2-Net and (f) ResU-Net on the LiTS dataset. These results demonstrate the models' ability to accurately segment liver regions across different datasets.

advantageous in anatomical variations on both datasets because of its ability to focus on and capture relevant features within an image. The Vanilla U-Net and U-Net3+ model easily converge with fewer number of training epochs on the datasets that make them efficient to conserve computational resources and enable quicker deployment and adaptation to various practical scenarios.

3.7 Error case analysis

The consistently lower DSC observed for the Res U-Net model across all datasets in Tables 4 and 5 suggests its limited ability to capture important features, likely stemming from its relatively low number of parameters. The model's diminished capacity to learn intricate patterns and structures within the data due to its sparse parameterization hampers its segmentation performance, indicating the need for increasing model complexity, to bolster feature extraction and improve segmentation accuracy. However, excessive model complexity can result in challenges such as increased computational demands, longer training times, and a higher risk of overfitting, necessitating a delicate balance between architecture weight and performance.

Despite having a heavy architecture, V-Net produces better DSCs in comparison to the Res U-Net. The V-Net is primarily designed for the segmentation of medical images. It captures contextual information more widely over the input image, which results in precise segmentation. Furthermore, the usage of 2 × 2 × 2 kernels with stride 2 instead of pooling ensures the learning of relevant features and capturing meaningful patterns without losing meaningful information. It also allows the incorporation of different activation functions, which unlike max-pooling does not have any learn-able parameters.

We have generated the segmented output masks for the test images, out of which only a few fail to generate the desired results. A few erroneous outputs generated have been shown in Figure 15. The weakly determined boundary points, a few extraneous-determined pixels, as well as unsegmented pixels, have been highlighted in this figure. The V-Net model uses 5 × 5 × 5 convolutions, which may be helpful in capturing local information, but it is not so efficient in capturing global information. For the models mentioned in Figure 15, they fuse different scale features through concatenation and upsampling which might lead to the deterioration of high-resolution features. Hence, at times, these models might fail to capture the precise location of the liver, especially the irregularities of the liver boundaries.

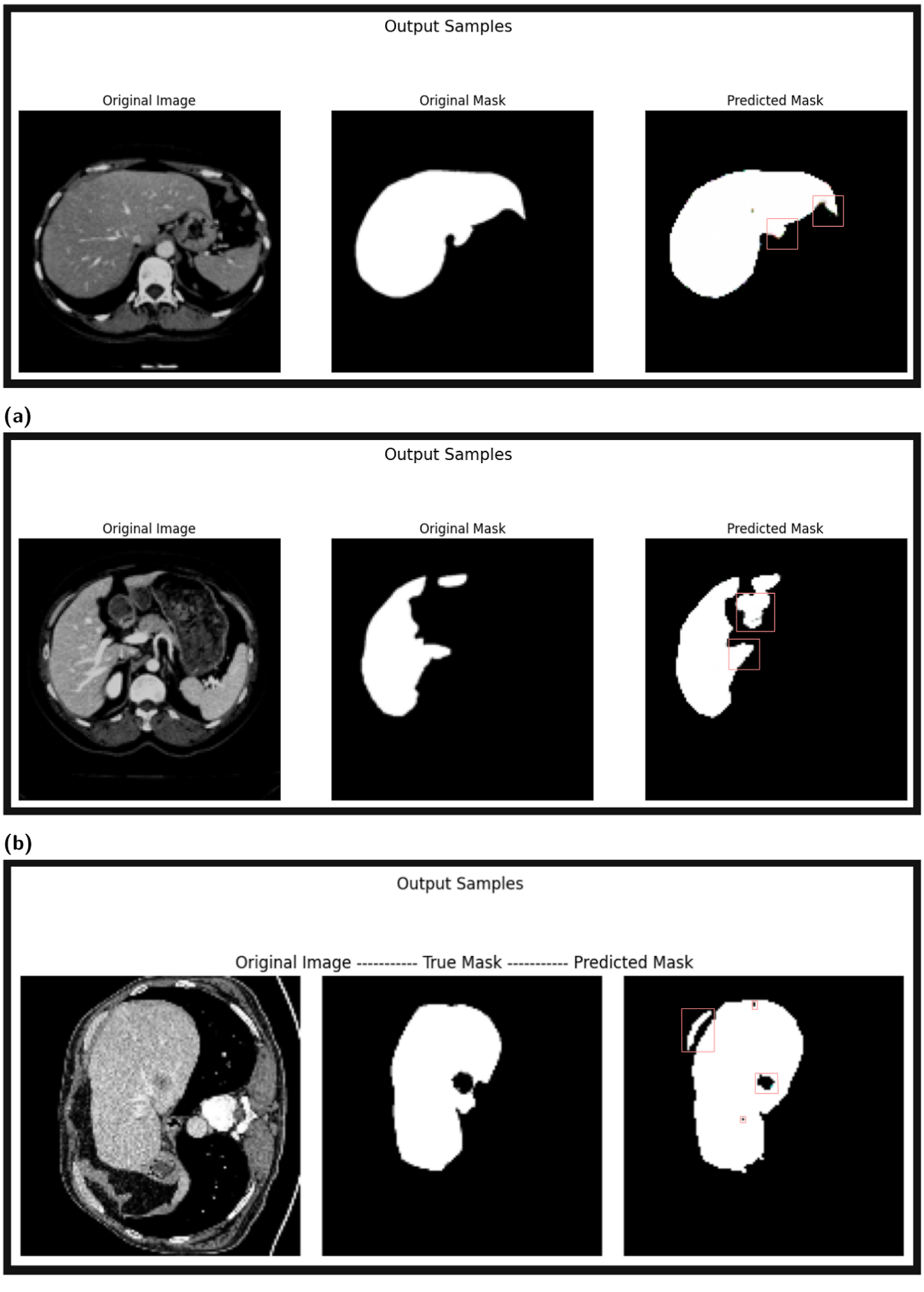
3.8 Data visualization

In this section, we use Gradient-weighted Class Activation Mapping (GradCAM) and IoU Heatmaps as data visualization tools to graphically display some results obtained by various U-Net models.

GradCAM analysis: To provide visual explanations of the model predictions in this investigation, we have used GradCAM [88], which aims to generate a gradient-weighted class activation map. These graphic representations help to explain how neural networks make decisions. We have used GradCAM to generate visualization for the models Vanilla U-Net, Attention U-Net, V-Net, U-Net 3+, R2U-Net, U^2-Net, Residual U-Net, and U-Net++ models. A CT scan image from the CHAOS dataset has been considered and the corresponding results are shown in Figure 16. Here, the models appear to focus on different segments of the input image. It is clear that the different models focus on distinct areas of the CT scans, suggesting that different models capture different but necessary information in their own unique way.

From Figure 16, the following things are observed:

(1) **Vanilla U-Net:** GradCAM for Vanilla U-Net reveals that the model emphasizes specific liver structures during segmentation, demonstrating a strong alignment between its predictions and important anatomical features.



(a)

(b)

(c)

Figure 15: Examples of erroneous outputs for liver segmentation: (a) R2U-Net on the 3DIRCADb dataset, (b) U-Net++ on the CHAOS dataset, and (c) U^2-Net on the LiTS dataset. These results highlight the limitations and challenges faced by the models in accurately segmenting certain liver regions in CT images.

(2) **Attention U-Net:** The GradCAM for the Attention U-Net indicates that this model places attention on critical liver regions, showing how it leverages its attention mechanism for accurate segmentation.

(3) **V-Net:** V-Net's GradCAM illustrates the regions of interest within the liver, showcasing the model's ability to capture fine details and contours during the segmentation process.

(4) **U-Net 3+:** The GradCAM for U-Net 3+ highlights areas within the liver, emphasizing the model's capacity to incorporate multi-scale information for precise segmentation.

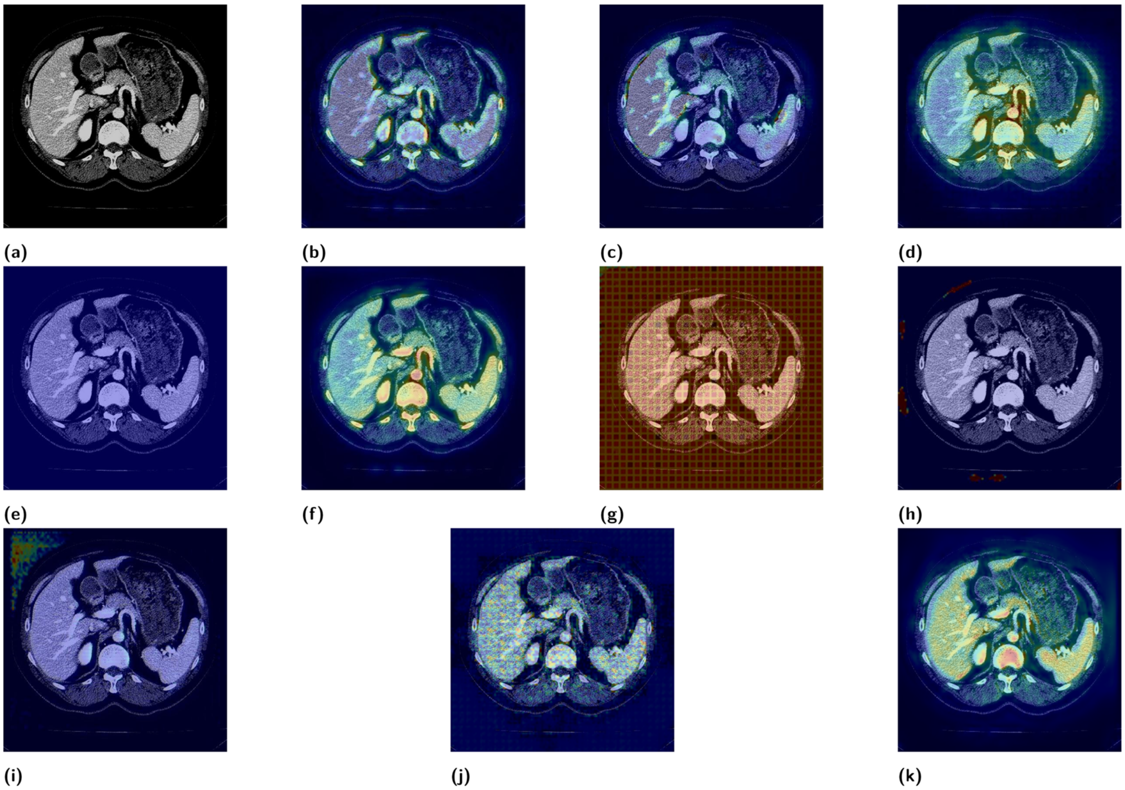


Figure 16: GradCAM visualizations for the 10 models used in this comparative study: (a) Original image, (b) Vanilla U-Net, (c) Attention U-Net, (d) V-Net, (e) U-Net 3+, (f) R2U-Net, (g) U^2-Net, (h) U-Net++, (i) ResU-Net, (j) Swin-UNet, and (k) Trans-UNet. These visualizations highlight the regions of the input images that contributed most to the models' predictions, providing insights into their decision-making processes based on the 3DIRCADb dataset.

(5) **R2U-Net:** R2U-Net's GradCAM shows a broader focus compared to other models, revealing its segmentation strategy that considers a wider context within the liver region.

(6) U^2-**Net:** GradCAM for U^2-Net displays the model's attention to liver boundaries and internal structures, underlining its segmentation capabilities, especially for challenging cases.

(7) **U-Net**++**:** U-Net++'s GradCAM highlights specific liver structures, indicating its proficiency in capturing complex anatomical features for accurate segmentation.

(8) **ResU-Net:** ResU-Net's GradCAM reveals the model's focus on intricate liver textures and patterns, showcasing its ability to exploit residual connections for robust segmentation.

(9) **Swin-UNet:** Swin-UNet's GradCAM demonstrates the model's attention to both global and local liver structures, highlighting its effectiveness in leveraging hierarchical representations for precise segmentation.

(10) **Trans-UNet:** GradCAM for Trans-UNet illustrates the model's adeptness in capturing spatial dependencies and long-range contextual information within the liver, indicating its capability to utilize transformer-based architectures for accurate segmentation.

IoU Heatmap analysis: When it comes to computer vision tasks like object identification, a model's prediction accuracy is visually represented as an IoU heatmap. Brighter colors indicate higher overlap between expected and ground truth bounding boxes, emphasizing accurate predictions. Colors are assigned based on IoU values. As shown in Figure 17, the heatmap is an effective tool for evaluating and improving model performance.

(1) **Vanilla U-Net:** The IoU heatmap for Vanilla U-Net shows high values corresponding to specific liver structures, indicating strong alignment between its predictions and important anatomical features during segmentation.

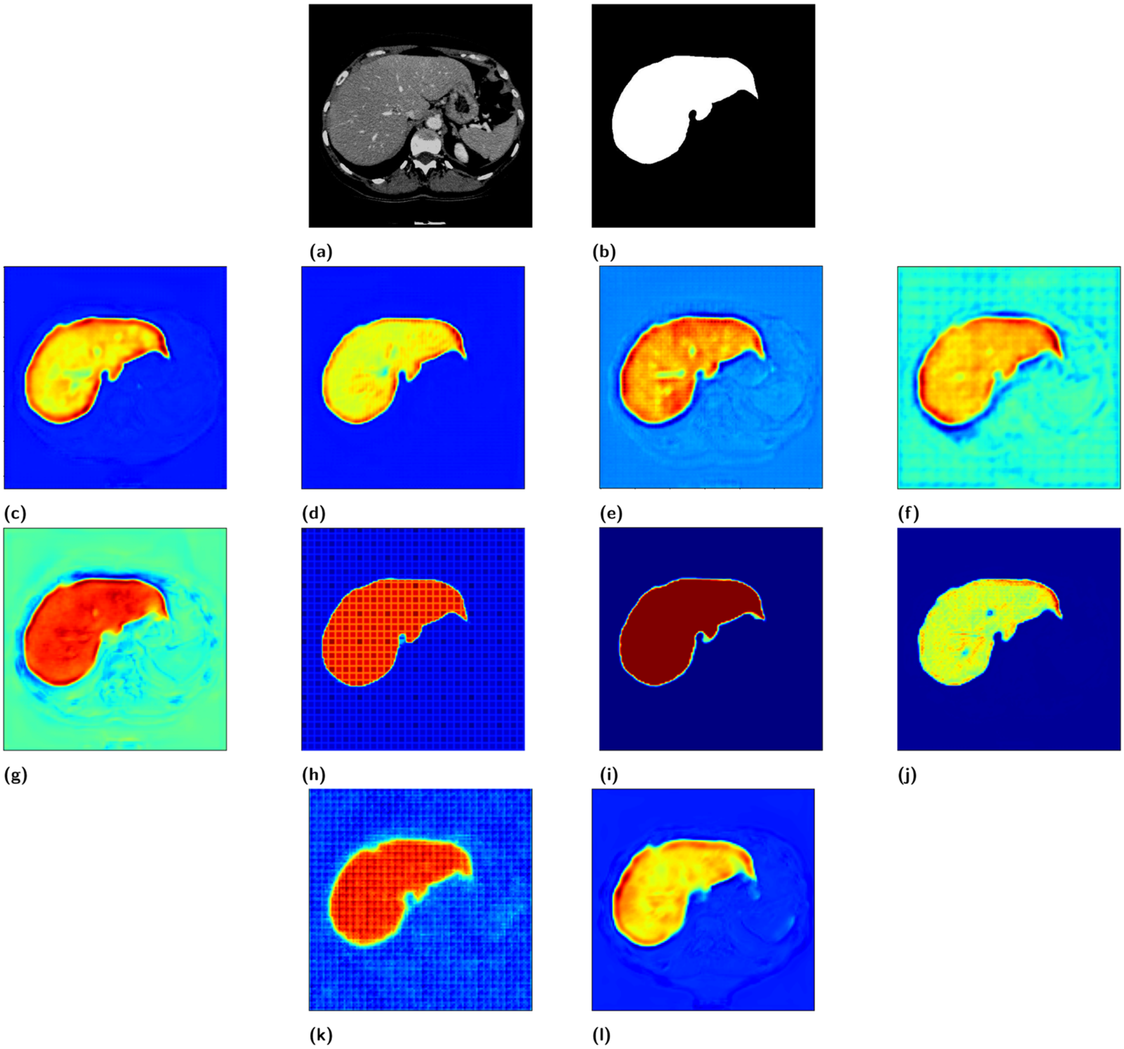


Figure 17: IoU heatmaps for the 10 models used in this comparative study: (a) Original image, (b) Original mask, (c) Vanilla U-Net, (d) Attention U-Net, (e) V-Net, (f) U-Net 3+, (g) R2U-Net, (h) U^2-Net, (i) U-Net++, (j) ResU-Net, (k) Swin-UNet, and (l) Trans-UNet. These heatmaps illustrate the Intersection over Union (IoU) performance of each model, highlighting their effectiveness in segmenting regions of interest compared to the ground truth based on the 3DIRCADb dataset.

(2) **Attention U-Net:** The IoU heatmap for Attention U-Net reveals elevated values in critical liver regions, demonstrating effective utilization of its attention mechanism for accurate segmentation.

(3) **V-Net:** The IoU heatmap generated for V-Net illustrates heightened values in regions of interest within the liver, reflecting the model's capability to capture fine details and contours during segmentation.

(4) **U-Net 3+:** The IoU heatmap for U-Net 3+ emphasizes elevated values within the liver, showcasing the model's proficiency in incorporating multi-scale information for precise segmentation.

(5) **R2U-Net:** The IoU heatmap for R2U-Net indicates a broader focus compared to other models, revealing its segmentation strategy that considers a wider context within the liver region.

(6) U^2**-Net:** The IoU heatmap for U^2-Net displays elevated values along the liver boundaries and internal structures, highlighting its segmentation capabilities, especially for challenging cases.

(7) **U-Net++:** The IoU heatmap for U-Net++ highlights elevated values corresponding to specific liver structures, indicating its proficiency in capturing complex anatomical features for accurate segmentation.

(8) **ResU-Net:** The IoU heatmap for ResU-Net reveals heightened values in intricate liver textures and patterns, showcasing its ability to exploit residual connections for robust segmentation.

(9) **Swin-UNet:** The IoU heatmap for Swin-UNet demonstrates elevated values in both global and local liver structures, indicating its effectiveness in leveraging hierarchical representations for precise segmentation.

(10) **Trans-UNet:** The IoU heatmap for Trans-UNet illustrates elevated values, highlighting the model's adeptness in capturing spatial dependencies and long-range contextual information within the liver, indicating its capability to utilize transformer-based architectures for accurate segmentation.

3.9 Recommendations

Accurate liver segmentation from medical images is of paramount importance in the diagnosis and treatment of various liver diseases. In this article, we have presented a comprehensive comparative analysis of several liver segmentation approaches, with a specific focus on different U-Net variants. Based on the findings and insights derived from this study, we offer the following recommendations:

(1) **Adoption of U-Net variants:** Our comparative analysis reveals that U-Net variants such as Vanilla U-Net, Attention U-Net, V-Net, U-Net 3+, R2U-Net, U^2-Net, U-net++, ResU-Net, Swin-U-Net, and Trans-U-Net have demonstrated remarkable performance in liver segmentation. We recommend the adoption of these latest U-Net architectures in the field of medical image segmentation, especially for tasks involving liver segmentation.

(2) **Dataset consideration:** The selection of appropriate datasets is crucial in evaluating the performance of liver segmentation algorithms. We recommend researchers and practitioners to consider the datasets like 3DIRCADb, CHAOS, and LiTS for benchmarking purposes, as these datasets provide diverse and challenging medical images that reflect real-world scenarios. Apart from these two, there is a wide variety of datasets that can be explored. A few among them are: PAIP2019 [89], and CholecSeg8k, etc. Comparative studies can also be conducted on them to check the performance of various models on these datasets.

(3) **Evaluation metrics:** This study emphasizes the importance of using established evaluation measures to compare the results of the segmentation. We recommend the use of well-established metrics, such as DSC, IoU, VoE, RVD, etc., to ensure a rigorous and meaningful comparison among different segmentation models.

(4) **Visual interpretation:** To enhance the interpretability of segmentation results, we recommend the use of visual tools like GradCAM, Activation Maps, IoU Heatmaps, etc. Such a tool can provide valuable insights into the performance of segmentation algorithms and facilitate a deeper understanding of the results. Other data visualization tools like t-distributed stochastic neighbor embedding (t-SNE) plots can be used in the future to know pixel-wise details segmentation results.

(5) **Further research:** While U-Net variants have shown significant promise in liver segmentation, there is always room for improvement. We encourage further research into the development of novel architectures and techniques that can address specific challenges in liver segmentation, such as the segmentation of small structures and intricate details.

(6) **Clinical integration:** Ultimately, the success of liver segmentation algorithms lies in their practical utility in the clinical setting. We recommend a close collaboration among researchers, medical professionals and healthcare institutions to ensure the seamless integration of advanced segmentation techniques into clinical practice. This collaboration can lead to better patient outcomes and more effective liver disease management.

Our comparative study provides valuable insights into liver segmentation techniques, with a focus on U-Net variants. By implementing the recommendations outlined above, researchers and practitioners can advance the field of liver segmentation, ultimately contributing to the improved diagnosis and treatment of liver diseases.

3.10 Challenges and future research directions

Despite significant developments, research gaps in liver and liver tumor segmentation still require thorough exploration. Accurate segmentation is crucial for diagnosis and treatment planning, but these computer-based methods are meant to assist, not directly plan, treatment. Consistently high segmentation accuracy is challenging due to anatomical variances, diverse tumor features, and changes in imaging modalities. To address these issues, specific algorithms are needed to improve patient care and medical decision-making. Key research gaps include the following:

(1) **Dataset limitations:** The datasets used, namely 3DIRCADb, CHAOS, and LiTS, are limited in size and diversity, potentially, thereby impacting result generalizability.
(2) **Hyperparameter uniformity:** The use of uniform hyperparameters across diverse models may not account for each model's sensitivities to specific settings.
(3) **Threat to validation:** Even with developments in CAD and deep learning-based models for medical imaging, medical experts' approval and supervision are still crucial. Even with their great efficiency, automated systems can sometimes result in false positives or negatives that need to be validated by a specialist in order to guarantee correct diagnosis and patient safety. Furthermore, there are serious privacy and security issues when using patient data to train and validate deep learning models. Maintaining public confidence in these technologies and protecting sensitive patient data depend on compliance with regulations like GDPR and HIPAA.
(4) **Model-specific constraints:** Specific model limitations, such as training time and robustness, should be acknowledged.
(5) **Generalizability:** The study's findings may not generalize well to other datasets or clinical scenarios, necessitating further exploration in future work.

Our research presents several interesting avenues for exploration in the domain of liver segmentation. However, there are future scopes in this domain that can give more insights to researchers.

(1) **Evaluation of latest U-Net variants:** Assessing the performance of the latest U-Net variants could provide valuable insights for improving segmentation accuracy.
(2) **Exploration of different pre-trained encoder backbones:** Experimenting with encoder backbones like DenseNet, ResNet, and MobileNet in U-Net architectures may enhance model robustness and efficiency.
(3) **Integration of recent models like ViT or Mamba:** Exploring recent models such as ViT (Vision Transformer) or Mamba within U-Net frameworks could lead to novel advancements in liver segmentation.
(4) **Investigation of alternative imaging modalities:** Studying other imaging modalities such as MRI or positron emission tomography scans may provide additional diagnostic insights.
(5) **Tumor detection and segmentation:** Probing the task of tumor detection and segmentation, especially for HCC, could significantly impact liver cancer diagnosis and treatment planning.

4 Conclusion

Liver cancer poses a significant global health challenge, ranking among the leading causes of cancer-related mortality worldwide with over 800,000 lives lost in 2020 alone. The prognosis and treatment outcomes for liver cancer vary greatly depending on the cancer's stage, underscoring the critical importance of early detection and intervention. Automated liver segmentation plays a crucial role in facilitating disease detection and localization, thereby aiding in timely medical interventions. Our research has validated the effectiveness of various advanced techniques such as attention mechanisms, dense convolutions, recurrent and residual blocks, as well as the utilization of pre-trained backbones within different U-Net architectures. These enhancements have shown notable improvements in segmentation performance, as measured by standard evaluation metrics like DSC and IoU.

These findings present promising avenues for both researchers and practitioners in the field. By demonstrating the efficacy of incorporating advanced architectural components and leveraging pre-trained models, our experiments provide insights into optimizing liver segmentation accuracy and efficiency. This research empowers stakeholders to select and adapt U-Net variants tailored to specific clinical or research requirements, thereby advancing the development of robust AI-driven tools for liver cancer diagnosis and treatment planning. Continued exploration and refinement in these areas hold potential to further enhance the capabilities of automated segmentation systems, ultimately improving patient outcomes in the fight against liver cancer.

Acknowledgements: We have had the honor of conducting research in the CMATER Research Lab, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India.

Funding information: This work was supported by the Russian Science Foundation (Project Number 22-41-04409).

Author contributions: Akash Halder – Conceptualization, Investigation, Writing original draft; Arup Sau – Conceptualization, Investigation, Validation, Writing review and editing; Surya Majumder – Investigation, Validation, Writing review and editing; Dmitri Kaplun – Funding acquisition, investigation, review and editing, project administration; Ram Sarkar – Conceptualization, Investigation, Writing review and editing, supervision.

Conflict of interest: All authors declare that there is no conflict of interest.

Data availability statement: We have used all publicly available datasets, link of which are as given below: 3DIRCADb: https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/ [90]. LiTS: https://competitions.codalab.org/competitions/17094. CHAOS: https://chaos.grand-challenge.org/ [91,92].

References

[1] Pharmacy Images. Labelled diagram of liver | liver images | human liver diagram. https://commons.wikimedia.org/wiki/File:Liver_Diagram.svg.

[2] Nayantara PV, Kamath S, Manjunath K, Rajagopal K. Computer-aided diagnosis of liver lesions using CT images: A systematic review. Comput Biol Med. 2020;127:104035. doi: 10.1016/j.compbiomed.2020.104035.

[3] Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. J Hepatol. 2019;70(1):151–71. doi: 10.1016/j.jhep.2018.09.014.

[4] Campadelli P, Casiraghi E, Esposito A. Liver segmentation from computed tomography scans: a survey and a new algorithm. Artif Intel Med. 2009;45(2–3):185–96. doi: 10.1016/j.artmed.2008.07.020.

[5] Arakeri MP. Recent advances and future potential of computer aided diagnosis of liver cancer on computed tomography images. In: International Conference on Information Processing. Springer; 2011. p. 246–51. doi: 10.1007/978-3-642-22786-8.

[6] Cleveland Clinic. Liver cancer. Accessed: December 12, 2023. [Online]. Available: https://my.clevelandclinic.org/health/diseases/9418-liver-cancer.

[7] Herbert Irving Comprehensive Cancer Center, Columbia University. Liver cancer. Accessed: December 20, 2023. [Online]. Available: https://www.cancer.columbia.edu/cancer-types-care/types/liver-cancer/about-liver-cancer.

[8] Ansari MY, Abdalla A, Ansari MY, Ansari MI, Malluhi B, Mohanty S, et al. Practical utility of liver segmentation methods in clinical surgeries and interventions. BMC Medical Imaging. 2022;22(1):1–17. doi: 10.1186/s12880-022-00825-2.

[9] Jayadevappa D, Srinivas Kumar S, Murty D. Medical image segmentation algorithms using deformable models: a review. IETE Tech Rev. 2011;28(3):248–55. doi: 10.4103/0256-4602.81244.

[10] Guo X, Schwartz LH, Zhao B. Automatic liver segmentation by integrating fully convolutional networks into active contour models. Med Phys. 2019;46(10):4455–69. doi: 10.1002/mp.13735.

[11] Wu W, Zhou Z, Wu S, Zhang Y. Automatic liver segmentation on volumetric CT images using supervoxel-based graph cuts. Comput Math Methods Med. 2016;2016:9093721. doi: 10.1155/2016/9093721.

[12] Thakur P, Madaan N. A survey of image segmentation techniques. Int J Res Comput Appl Robotics. 2014;2(4):158–65. https://api.semanticscholar.org/CorpusID:212446238.

[13] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2021;18(2):203–11. doi: 10.1038/s41592-020-01008-z.

[14] Zhang F, Wang Y, Yang H. Efficient context-aware network for abdominal multi-organ segmentation. 2021. arXiv: http://arXiv.org/abs/arXiv:210910601.

[15] Majumder S, Gautam N, Basu A, Sau A, Geem ZW, Sarkar R. MENet: AMitscherlich function based ensemble of CNN models to classify lung cancer using CT scans. Plos One. 2024;19(3):e0298527. doi: 10.1371/journal.pone.0298527.

[16] Goenka N, Sharma AK, Tiwari S, Singh N, Yadav V, Prabhu S, et al. A regularized volumetric ConvNet based Alzheimer detection using T1-weighted MRI images. Cogent Eng. 2024;11(1):2314872. doi: 10.1080/23311916.2024.2314872.

[17] Dsilva LR, Tantri SH, Sampathila N, Mayrose H, Muralidhar Bairy G, Belurkar S, et al. Wavelet scattering-and object detection-based computer vision for identifying dengue from peripheral blood microscopy. Int J Imaging Syst Tech. 2024;34(1):e23020. doi: 10.1002/ima.23020.

[18] Roy A, Pramanik P, Sarkar R. EU2-Net: a parameter efficient ensemble model with attention-aided triple feature fusion for tumor segmentation in breast ultrasound images. IEEE Trans Instrument Measurement. 2024;73:1–7. doi: 10.1109/TIM.2024.3421436.

[19] Bhattacharyya T, Chatterjee B, Sarkar R, Kundu M. Segmentation of brain MRI using moth-flame optimization with modified cross entropy based fitness function. Multimedia Tools Appl. 2024;83:1–22. doi: 10.1007/s11042-024-18461-z.

[20] Gautam N, Basu A, Kaplun D, Sarkar R. An ensemble of UNet frameworks for lung nodule segmentation. In: International Conference on Actual Problems of Applied Mathematics and Computer Science. Springer; 2022. p. 450–61. doi: 10.1007/978-3-031-34127-4.

[21] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8:1–74. doi: 10.1186/s40537-021-00444-8.

[22] Rozenwald MB, Galitsyna AA, Sapunov GV, Khrameeva EE, Gelfand MS. A machine learning framework for the prediction of chromatin folding in Drosophila using epigenetic features. PeerJ Comput Sci. 2020;6:e307. doi: 10.7717/peerj-cs.307.

[23] Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. Neurocomputing. 2017;234:11–26. doi: 10.1016/j.neucom.2016.12.038.

[24] Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, et al. A survey on deep learning: Algorithms, techniques, and applications. ACM Comput Surveys (CSUR). 2018;51(5):1–36. doi: 10.1145/3234150.

[25] Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. A state-of-the-art survey on deep learning theory and architectures. Electronics. 2019;8(3):292. doi: 10.3390/electronics8030292.

[26] Koppe G, Meyer-Lindenberg A, Durstewitz D. Deep learning for small and big data in psychiatry. Neuropsychopharmacology. 2021;46(1):176–90. doi: 10.1038/s41386-020-0767-z.

[27] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. Springer; 2015. p. 234–41. doi: 10.1007/978-3-319-24574-4.

[28] Liu X, Zhao H. Hierarchical feature extraction based on discriminant analysis. Appl Intell. 2019;49:2780–92. doi: 10.1007/s10489-019-01418-3.

[29] Yue L, Gong X, Li J, Ji H, Li M, Nandi AK. Hierarchical feature extraction for early Alzheimeras disease diagnosis. IEEE Access. 2019;7:93752–60. doi: 10.1109/ACCESS.2019.2926288.

[30] Wang Z, Zhang L, Fang T, Mathiopoulos PT, Tong X, Qu H, et al. A multiscale and hierarchical feature extraction method for terrestrial laser scanning point cloud classification. IEEE Trans Geosci Remote Sensing. 2014;53(5):2409–25. doi: 10.1109/TGRS.2014.2359951.

[31] Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: Artificial and Machine Learning-ICANN 2011: 21st International Conference on Artificial, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer; 2011. p. 52–59. doi: 10.1007/978-3-642-21735-7.

[32] Li H, Wei Y, Li L, Chen CP. Hierarchical feature extraction with local neural response for image recognition. IEEE Trans Cybernetics. 2013;43(2):412–24. doi: 10.1109/TSMCB.2012.2208743.

[33] Song HA, Kim BK, Xuan TL, Lee SY. Hierarchical feature extraction by multi-layer non-negative matrix factorization network for classification task. Neurocomputing. 2015;165:63–74. doi: 10.1016/j.neucom.2014.08.095.

[34] Chandra TB, Verma K, Singh BK, Jain D, Netam SS. Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. Expert Syst. Appl. 2020;158:113514. doi: 10.1016/j.eswa.2020.113514.

[35] Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. In: International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Springer; 2016. p. 179–87. doi: 10.1007/978-3-319-46976-8.

[36] Orhan AE, Pitkow X. Skip connections eliminate singularities. 2017. arXiv: http://arXiv.org/abs/arXiv:170109175.

[37] Tong T, Li G, Liu X, Gao Q. Image super-resolution using dense skip connections. In: Proceedings of the IEEE International Conference on Computer Vision; 2017. p. 4799–807.

[38] Mao X, Shen C, Yang YB. Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections. Adv Neural Inform Process Syst. 2016;29. https://proceedings.neurips.cc/paper_files/paper/2016/file/0ed9422357395a0d4879191c66f4faa2-Paper.pdf.

[39] Lebre MA, Vacavant A, Grand-Brochier M, Rositi H, Abergel A, Chabrot P, et al. Automatic segmentation methods for liver and hepatic vessels from CT and MRI volumes, applied to the Couinaud scheme. Comput Biol Med. 2019;110:42–51. doi: 10.1016/j.compbiomed.2019.04.014.

[40] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. 2018. doi: 10.48550/arXiv.1804.03999.

[41] Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE; 2016. p. 565–71. doi: 10.1109/3DV.2016.79.

[42] Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, et al. Unet 3.: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2020. p. 1055–9. doi: 10.1109/ICASSP40776.2020.9053405.

[43] Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. 2018. doi: 10.48550/arXiv.1802.06955.

[44] Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognit. 2020;106:107404. doi: 10.48550/arXiv.2005.09007.

[45] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer; 2018. p. 3–11. doi: 10.1007/978-3-030-00889-5.

[46] Diakogiannis FI, Waldner F, Caccetta P, Wu C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens. 2020;162:94–114. doi: 10.1016/j.isprsjprs.2020.01.013.

[47] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision. Springer; 2022. p. 205–18. doi: 10.1007/978-3-031-25066-8.

[48] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. 2021. doi: 10.48550/arXiv.2102.04306.

[49] Jiang H, Shi T, Bai Z, Huang L. Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes. IEEE Access. 2019;7:24898–909. doi: 10.1109/ACCESS.2019.2899608.

[50] Jiang L, Ou J, Liu R, Zou Y, Xie T, Xiao H, et al. RMAU-Net: residual multi-scale attention U-Net for liver and tumor segmentation in CT images. Comput Biol Med. 2023;158:106838. doi: 10.1016/j.compbiomed.2023.106838.

[51] Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, et al. The liver tumor segmentation benchmark (lits). Med Image Anal. 2023;84:102680. doi: 10.1016/j.media.2022.102680.

[52] Hille G, Agrawal S, Tummala P, Wybranski C, Pech M, Surov A, et al. Joint liver and hepatic lesion segmentation in MRI using a hybrid CNN with transformer layers. Comput Methods Programs Biomed. 2023;240:107647. doi: 10.1016/j.cmpb.2023.107647.

[53] Li L, Ma H. Rdctrans U-Net: A hybrid variable architecture for liver CT image segmentation. Sensors. 2022;22(7):2452. doi: 10.3390/s22072452.

[54] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 1492–500. doi: 10.48550/arXiv.1611.05431.

[55] Jiang J, Peng Y, Hou Q, Wang J. MDCF_Net: A Multi-dimensional hybrid network for liver and tumor segmentation from CT. Biocybernetics Biomed Eng. 2023;43:494–506. doi: 10.1016/j.bbe.2023.04.004.

[56] Han JJ, Zhang X, Cai X, Pan R, Xiao L, Nan Y, et al. Deep learning model based on CNN-former in the diagnosis and detection of liver fibrosis. Research Square. 2023. doi: 10.21203/rs.3.rs-2869666/v1.

[57] Zhang J, Liu Y, Wu Q, Wang Y, Liu Y, Xu X, et al. SWTRU: star-shaped window transformer reinforced U-net for medical image segmentation. Comput Biol Med. 2022;150:105954. doi: 10.1016/j.compbiomed.2022.105954.

[58] Mulay S, Deepika G, Jeevakala S, Ram K, Sivaprakasam M. Liver segmentation from multimodal images using HED-mask R-CNN. In: Multiscale Multimodal Medical Imaging: First International Workshop, MMMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 1. Springer; 2020. p. 68–75. doi: 10.1007/978-3-030-37969-8.

[59] Roy SS, Roy S, Mukherjee P, Roy AH. An automated liver tumour segmentation and classification model by deep learning based approaches. Comput Methods Biomech Biomed Eng Imag Vis. 2023;11(3):638–50. doi: 10.1080/21681163.2022.2099300.

[60] Chi J, Han X, Wu C, Wang H, Ji P. X-Net: Multi-branch UNet-like network for liver and tumor segmentation from 3D abdominal CT scans. Neurocomputing. 2021;459:81–96. doi: 10.1016/j.neucom.2021.06.021..

[61] Manjunath RV, Kwadiki K. Modified U-NET on CT images for automatic segmentation of liver and its tumor. Biomed Eng Adv. 2022;4:100043. https://www.sciencedirect.com/science/article/pii/S2667099222000196.

[62] Khan RA, Luo Y, Wu FX. RMS-UNet: Residual multi-scale UNet for liver and lesion segmentation. Artif Intel Med. 2022;124:102231. doi: 10.1016/j.artmed.2021.102231.

[63] Kavur AE, Gezer NS, Baríiş M, Aslan S, Conze PH, Groza V, et al. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. Med Image Anal. 2021;69:101950. doi: 10.1016/j.media.2020.101950.

[64] Won YD, Na MK, Kim CH, Kim JM, Cheong JH, Ryu JI, et al. The frontal skull Hounsfield unit value can predict ventricular enlargement in patients with subarachnoid haemorrhage. Sci Reports. 2018;8(1):10178. doi: 10.1038/s41598-018-28471-1.

[65] Ishihara H, Oka F, Kawano R, Shinoyama M, Nishimoto T, Kudomi S, et al. Hounsfield unit value of interpeduncular cistern hematomas can predict symptomatic vasospasm. Stroke. 2020;51(1):143–8. doi: 10.1161/STROKEAHA.119.026962.

[66] DenOtter TD, Schubert J. Hounsfield unit. StatPearls; 2019.

[67] Thomsen V, Schatzlein D, Mercuro D. Tutorial: attenuation of X-rays by matter. Spectroscopy. 2005;20(9).
[68] DeMarco JJ, Suortti P. Effect of scattering on the attenuation of X rays. Phys Rev B. 1971;4(4):1028. doi: 10.1103/PhysRevB.4.1028.
[69] Artem'ev V. Attenuation of X rays by ultradisperse media. Tech Phys Lett. 1997;23:212–3. doi: 10.1134/1.1261601.
[70] Molteni R. Prospects and challenges of rendering tissue density in Hounsfield units for cone beam computed tomography. Oral Surg Oral Med Oral Pathol Oral Radiol. 2013;116(1):105–19. doi: 10.1016/j.oooo.2013.04.013.
[71] Schreiber JJ, Anderson PA, Rosas HG, Buchholz AL, Au AG. Hounsfield units for assessing bone mineral density and strength: a tool for osteoporosis management. JBJS. 2011;93(11):1057–63. doi: 10.2106/JBJS.J.00160.
[72] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1026–34.
[73] Jiang T, Cheng J. Target recognition based on CNN with LeakyReLU and PReLU activation functions. In: 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). IEEE; 2019. p. 718–22.
[74] Crnjanski J, Krstić M, Totović A, Pleros N, Gvozdić D. Adaptive sigmoid-like and PReLU activation functions for all-optical perceptron. Optics Letters. 2021;46(9):2003–6. doi: 10.1364/OL.422930.
[75] Jang E, Gu S, Poole B. Categorical reparameterization with gumbel-softmax. 2016. arXiv: http://arXiv.org/abs/arXiv:161101144.
[76] Liu W, Wen Y, Yu Z, Yang M. Large-margin softmax loss for convolutional neural networks. 2016. arXiv: http://arXiv.org/abs/arXiv:161202295.
[77] Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: International Workshop on Artificial Neural Networks. Springer; 1995. p. 195–201. doi: 10.1007/3-540-59497-3.
[78] Yin X, Goudriaan J, Lantinga EA, Vos J, Spiertz HJ. A flexible sigmoid function of determinate growth. Ann Bot. 2003;91(3):361–71. doi: 10.1093/aob/mcg091.
[79] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 10012–22. doi: 10.48550/arXiv.2103.14030.
[80] Sun C, Xu A, Liu D, Xiong Z, Zhao F, Ding W. Deep learning-based classification of liver cancer histopathology images using only global labels. IEEE J Biomed Health Inform. 2019;24(6):1643–51. doi: 10.1109/JBHI.2019.2949837.
[81] Wang X, Zhang X, Wang G, Zhang Y, Shi X, Dai H, et al. TransFusionNet: Semantic and spatial features fusion framework for liver tumor and vessel segmentation under JetsonTX2. IEEE J Biomed Health Informatics. 2022;27(3):1173–84. doi: 10.1109/JBHI.2022.3207233.
[82] Zhao R, Qian B, Zhang X, Li Y, Wei R, Liu Y, et al. Rethinking dice loss for medical image segmentation. In: 2020 IEEE International Conference on Data Mining (ICDM). IEEE; 2020. p. 851–60. doi: 10.1109/ICDM50108.2020.00094.
[83] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. arXiv: http://arXiv.org/abs/arXiv:14091556.
[84] Dozat T. Incorporating Nesterov momentum into Adam. Caribe Hilton, San Juan, Puerto Rico: ICLR2016; 2016. https://openreview.net/forum?id=OM0jvwB8jIp57ZJjtNEZ&noteId=nx924kDvKc7lP3z2iomv.
[85] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014. arXiv: http://arXiv.org/abs/arXiv:14126980.
[86] Xie X, Zhou P, Li H, Lin Z, Yan S. Adan: Adaptive Nesterov momentum algorithm for faster optimizing deep models. 2022. arXiv: http://arXiv.org/abs/arXiv:220806677.
[87] Tang S, Shen C, Wang D, Li S, Huang W, Zhu Z. Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis. Neurocomputing. 2018;305:1–14. doi: 10.1016/j.neucom.2018.04.048.
[88] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision; 2017. p. 618–26. doi: 10.1109/ICCV.2017.74.
[89] Kim YJ, Jang H, Lee K, Park S, Min SG, Hong C, et al. PAIP 2019: Liver cancer segmentation challenge. Med Image Anal. 2021;67: 101854. doi: 10.1016/j.media.2020.101854.
[90] Soler L, Hostettler A, Agnus V, Charnoz A, Fasquel JB, Moreau J, et al. 3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. IRCAD, Strasbourg, France, Tech. Rep. 2010.
[91] Kavur AE, Selver MA, Dicle O, Barış M, Gezer NS. CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data (Version v1.03) [Data set]. Zenodo. 2019. doi: 10.5281/zenodo.3362844.
[92] Kavur AE, Gezer NS, Barış M, Şahin Y, Özkan S, Baydar B, et al. Comparison of semi-automatic and deep learning-based automatic methods for liver segmentation in living liver transplant donors. Diagn Inter Radiol. 2020;26:11–21. doi: 10.5152/dir.2019.19025.