Deep Learning-Based Liver Cancer Segmentation Using U-net

Kunlun Li(kli90), Boshen Pan(bpan8), Shiyang Lin(slin110)

***Abstract*—The liver, the largest organ in the human body, plays a vital role in metabolism and is consequently vulnerable to various diseases. Over 350 million people worldwide are affected by the Hepatitis B virus (HBV), as reported by the World Health Organization. Early detection and diagnosis of liver diseases remain challenging due to the reliance on CT and MRI scans, which generate numerous images for analysis, consuming significant time and effort. This often leads to prolonged examination times, potentially decreasing the accuracy of diagnoses. The development of deep learning technologies has revolutionized medical image segmentation, offering new methods to enhance diagnostic precision and efficiency. This report presents the application of the U-net architecture, augmented with basic image preprocessing techniques such as Histogram Equalization and Windowing, to improve the segmentation and diagnosis of liver diseases. These advancements aim to overcome the limitations of traditional imaging techniques, facilitating more accurate and efficient diagnostics.**

***Index Terms*—Medical Image, Segmentation, Liver Cancer, Windowing, Histogram Equalization, U-Net, DALU-Net**

I. Introduction

The rapid evolution of medical imaging has provided clinicians with powerful tools for diagnosing a range of conditions, yet the manual analysis of these images remains a time-consuming and error-prone task. Among these, semantic segmentation of medical images stands out for its utility in delineating specific regions of interest, distinguishing it fundamentally from other forms of image analysis like classification and object detection. Classification tasks predict a single label for an entire image, object detection identifies and locates multiple objects within an image, and instance segmentation identifies and delineates each object instance. In contrast, semantic segmentation assigns a label to each pixel in an image, making it particularly suitable for medical applications where precision is paramount, such as in liver cancer segmentation.

Liver cancer remains one of the most challenging cancers to diagnose due to the liver's complexity and the subtlety of early tumor signs. The application of deep learning, especially Convolutional Neural Networks (CNNs), to this problem has the potential to transform diagnostic practices by enhancing the speed and accuracy of image analysis. This report focuses on leveraging the U-net architecture, renowned for its effectiveness in medical image segmentation, to improve liver cancer detection and segmentation in CT scans. The report will be organized as following. First, the background of the medical image segmentation will be reviewed and then the data that have been used in ths task will be introduced. Then, the data preprocessing and augmentation method will be illustrated. Next, the U-Net model that have been implemented in this task will be elaborated and the parameter will be visualized as well. In the last part, the result of the segmentation from the U-Net will be concluded and compared with the state-of-art architecture such as DELU-Net in terms of the dice coefficient. The harvest and reflection will also be concluded in the last part.

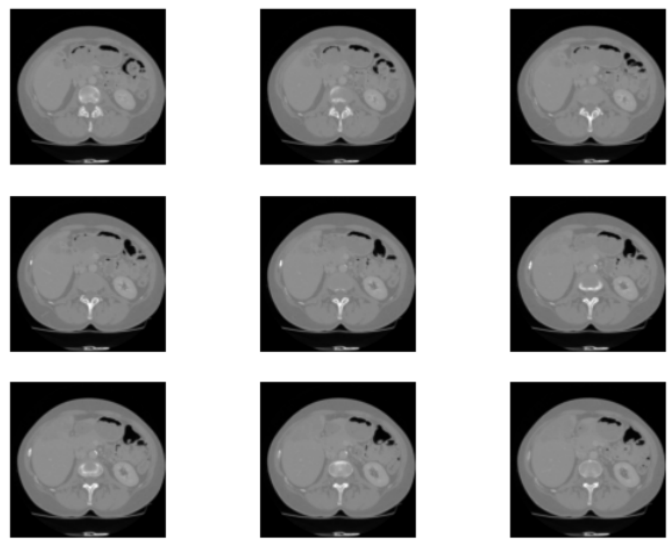

Fig. 1. The CT image for liver

II. Background

Medical image segmentation plays a critical role in computer-assisted diagnosis (CAD) systems, facilitating the extraction of anatomical structures from various imaging modalities such as CT, MRI, and ultrasound. Traditionally, segmentation was performed manually by skilled radiologists, a process that is not only labor-intensive but also prone to subjective interpretation and variability. Early automated segmentation approaches relied heavily on simple thresholding techniques and region-growing algorithms. However, these methods often struggled with image artifacts, noise, and the heterogeneity of medical images.

With advances in machine learning, more sophisticated techniques such as Support Vector Machines (SVM) and Random Forests were employed to improve segmentation accuracy. These methods utilized hand-crafted features to train the models, but they still required significant domain expertise and often failed to generalize across different types of medical images.

The advent of deep learning marked a significant turning point in medical image analysis. Convolutional Neural Networks (CNNs), in particular, have become the backbone of many modern medical image analysis tools due to their ability to learn hierarchical features directly from the data, eliminating the need for manual feature extraction. Among various architectures, Fully Convolutional Networks (FCN) first demonstrated the potential of CNNs for pixel-wise predictions, setting the stage for more specialized architectures.

Introduced by Olaf Ronneberger, Philipp Fischer, and Thomas Brox, the U-net architecture was specifically designed for medical image segmentation. The architecture features a symmetric encoder-decoder structure, which includes contracting and expansive path. Contracting Path: The encoder uses consecutive convolutions and pooling operations to capture context and reduce spatial dimensions. The Expansive Path: The decoder progressively reconstructs the spatial dimensions through up-sampling and convolutions, enhanced by skip connections from the encoder layers to retain and utilize fine-grained details.

The U-net's ability to efficiently use data and its robustness to the small size of medical imaging datasets quickly established it as a preferred choice for medical segmentation tasks.

III. Preliminary Methods and Results

The methodology we have used in this task include: Data Preprocessing, Data Augmentation, Building the U-net model, Design the loss function, Call Back function and Optimizer, and Test the Deep Attention LSTM U-Net (DALU-Net).

A. Data Preprocessing

The dataset utilized in this study consists of the 3D-IRCADb (3D Image Reconstruction for Comparison of Algorithm Database), which includes CT scans from 20 patients, comprising both male and female subjects, with liver tumors.

The preprocessing steps involve loading the CT images alongside their corresponding liver tumor masks, applying windowing techniques to enhance image contrast, performing histogram equalization to further improve visibility, and saving the processed images for model training. The algorithm will be further introduced in the following.

The first technique that have utilized to enhance the image contrast is windowing. Windowing, also known as intensity windowing or contrast stretching, is a technique used to improve the visibility of features in medical images by adjusting the image's intensity scale. This technique is particularly useful in CT imaging, where different tissues absorb varying amounts of radiation, resulting in a broad range of intensity values. By applying windowing, specific ranges of intensities can be highlighted while others are suppressed, making it easier to distinguish between normal and abnormal tissues.

In practice, windowing involves selecting an upper and lower intensity threshold, which defines the "window." Only the intensities within this window are displayed, with values below the threshold set to black and those above set to white. This process not only enhances the visibility of structures within the chosen range but also reduces the influence of outliers or noise outside this range. For liver tumor segmentation, the window is typically set to enhance soft tissue contrasts, which helps in clearly delineating the liver and any tumors from the surrounding structures. Another technique

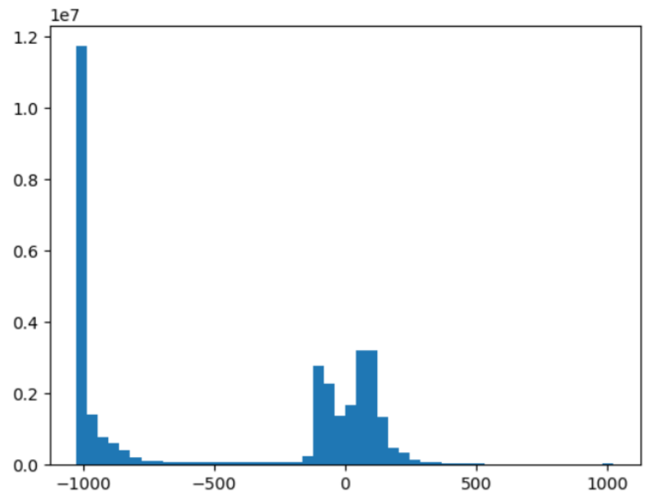

Fig. 2. The Histogram for the CT image after windowing with width=250 and center=0

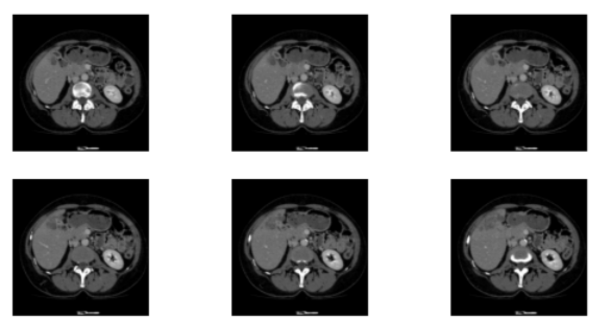

Fig. 3. The CT image after windowing

is called Histogram Equalization. Histogram Equalization is a technique used to improve the contrast of images. It works by effectively spreading out the most frequent intensity values, thus enhancing the global contrast of the image. This method is particularly beneficial in medical imaging, where contrast differences can be subtle yet critically important for accurate diagnosis.

HE adjusts the image intensities to achieve a more uniform distribution. This uniform distribution allows for areas hidden in darker regions to gain brightness and become more visible, which is crucial for detecting finer details necessary for accurate segmentation. In the context of liver cancer segmentation, HE ensures that subtle variations in tissue density are more discernible, which is vital for accurately identifying and segmenting tumors.

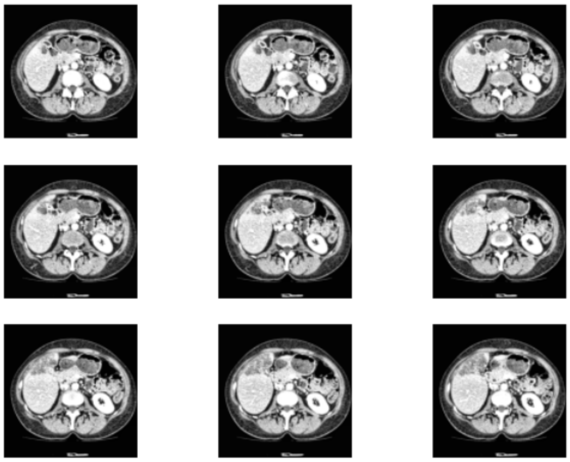

Fig. 4. The CT image after performing histogram equalization

B. Data Augmentation

Since there are only limited dataset can be accessed, the data augmentation should also be considered to further extend the dataset. Data augmentation is a critical technique in training deep learning models, especially in the field of medical image analysis, where the availability of large labeled datasets is often limited. The parameters shown in the provided code snippet indicate the application of various augmentation techniques such as rotation, width and height shifts, shear, and zoom. The reason of using data augmentation can be beneficial in terms of the following.

Increased Dataset Diversity: By applying transformations like small rotations (up to 10 percent), width and height shifts (up to 5 percent), and zoom (up to 5 percent), the model is exposed to a wider variety of possible image orientations and scales. This helps simulate different imaging conditions and patient positions that might not be adequately represented in the original dataset.

Robustness to Variability: Shear transformations introduce changes in the image geometry, mimicking variations that might occur due to patient movement or differences in imaging technique. This enhances the model's ability to generalize across images that have not been perfectly captured.

Prevention of Overfitting: Augmenting data effectively increases the amount of training data, helping to prevent the model from overfitting to the limited number of examples available. This is particularly important in medical imaging, where the nuances between normal tissues and pathological features can be subtle and highly variable.

Enhanced Model Performance: By training on augmented data, the model learns to recognize important features across a range of variations, leading to improved performance on new, unseen images. This is crucial for medical applications where accurate and reliable segmentation can directly impact clinical outcomes.

C. U-net training

Then we design the U-net architecture and train the model with our prepared data.

The U-net architecture is distinctively divided into two main components: the left side functioning as an encoder and the right side as a decoder. The encoder consists of four sub-modules, each comprising two convolutional layers followed by a max pooling layer for downsampling. The initial resolution of the input image is 572x572 pixels. The resolutions of the subsequent outputs from the first to the fifth sub-modules are 572x572, 284x284, 140x140, 68x68, and 32x32, respectively. Due to the use of valid padding in the convolution layers, the resolution of each subsequent sub-module is effectively reduced to half of its predecessor's size minus four pixels.

Conversely, the decoder includes four sub-modules that progressively increase the resolution of the feature maps through upsampling operations. The goal is to restore the resolution to match that of the original input image. However, due to valid padding in the convolutional process, the actual output resolution is slightly smaller than the input image.

Additionally, the network incorporates skip connections that link the output of each upsampling step in the decoder to the corresponding encoder output with the same resolution. These connections facilitate the propagation of context information and details across the network, improving the accuracy of the segmentation by leveraging features from multiple resolutions. After design the architecture of the model, we also include

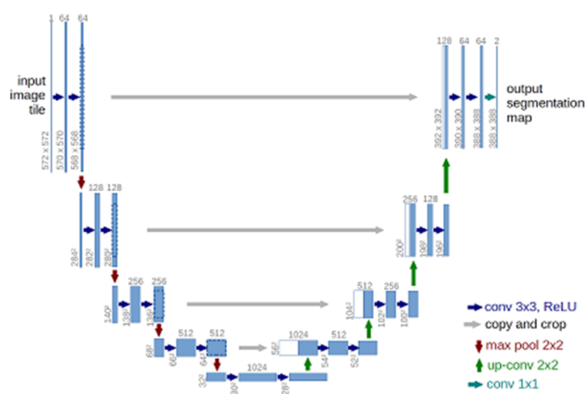

Fig. 5. The U-net architecture that have been implemented

callback function, Adagrad optimizer and cross-entropy loss function to enhance the performance of the model. The following are the brief introduction to the function used in this model.

Callbacks: Callbacks are essential in training deep learning models as they provide a way to automatically monitor and adjust the training process at specific stages. For instance, callbacks can be used to save model checkpoints at regular intervals, reduce the learning rate when a plateau in model performance is detected, or even stop training early if no improvement is observed over a certain number of epochs.

This automation enhances training efficiency and ensures that the model remains robust against overfitting.

Adagrad Optimizer: Adagrad is an optimizer with parameter-specific learning rates, which are adapted relative to how frequently a parameter is updated during training. This is particularly beneficial for deep learning applications in medical imaging, where the diversity in image features requires nuanced adjustments. Adagrad's approach helps in efficient training by allowing for larger updates for infrequent parameters and smaller updates for frequent ones, which can lead to faster convergence on complex tasks like image segmentation.

Cross-Entropy Loss: In the context of image segmentation, cross-entropy loss is a pivotal loss function for training classification models. It measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label, making it an ideal choice for segmentation tasks where the goal is to classify each pixel of an image as belonging to a particular class. This loss function is well-suited for tasks with multiple classes and unbalanced datasets, which are common in medical imaging scenarios.

After we pass the dataset into the model, it is obvious in the Fig.6 that with the increasing of the epoch, the accuracy for the prediction of the result becomes better.

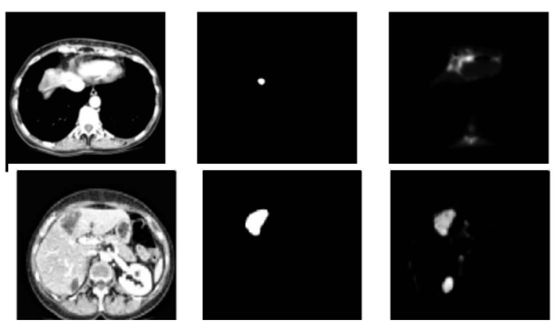

Fig. 6. From left to right are the Source Image, Ground Truth, Predicted segmentation. The first row is prediction result with epoch=5 and the second row is result with epoch=100

D. *Deep Attention LSTM U-Net*

Then, another consideration is to use Deep Attention LSTM U-Net (DALU-Net).

The DALU-Net is an advanced architectural variant of the traditional U-net, specifically engineered to address some of the limitations found in the standard U-net model by incorporating attention mechanisms, Deep Supervision, and ConvLSTM layers. This combination leverages the strengths of each component to improve the segmentation performance, particularly in complex medical imaging tasks. In the following, each components in the DALU-Net will be introduced.

Attention Mechanisms (AM): Attention mechanisms within DALU-Net help the model to focus selectively on regions of the image that are more pertinent to the task of segmentation. This is particularly useful in medical images where the areas of interest, such as tumors or lesions, occupy only a small part of the overall image but are critical for accurate diagnosis.

Deep Supervison(DS): Deep Supervison make the output of each layer of the decoder will be matched with the output and calculate the loss, which can help the network converge faster and relieve the gradient vanishing problem.

ConvLSTM (CLSTM): The inclusion of Convolutional LSTM (ConvLSTM) layers allows DALU-Net to maintain spatial relationships across image sequences, which is critical when dealing with volumetric data such as 3D medical images. ConvLSTM layers help in capturing the temporal dynamics and spatial features, thus preserving the integrity of spatial relationships across the layers of the network.

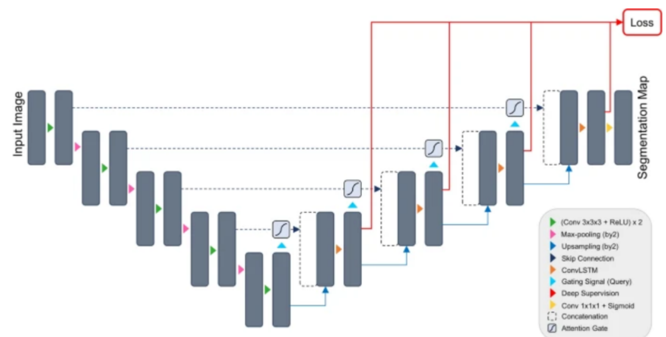

Fig. 7. The DALU-Net architecture that have been implemented

IV. Conclusion and Discussion

Basing on the Fig.8 and Table I, we can find compared with U-Net, the DALU-Net improves the Dice coefficient by about 10%. That is because compared to the standard U-net, DALU-Net offers enhanced feature extraction capabilities due to its dense connectivity and the ability to focus on important features through attention mechanisms. Additionally, the spatial coherence preserved by ConvLSTM layers makes DALU-Net particularly suitable for applications involving sequential image data, where understanding the spatial and temporal context can significantly enhance segmentation accuracy.

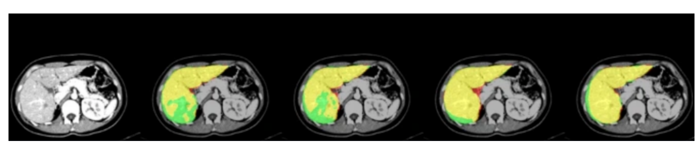

Fig. 8. The Segmented Result with different method. From left to right is Source Image,U-Net,AU-Net,AU-Net w/DS,DALU-Net

TABLE I
Evaluation for the segmentation by using different model

Model	U-Net	AU-Net	AU-Net w/DS	DALU-Net
Dice coefficient	0.822	0.852	0.887	0.899

Through this project, we have gained a comprehensive understanding of the workflow involved in a deep learning task. This included preparing the dataset, conducting data preprocessing and augmentation, developing the neural network architecture, selecting appropriate loss functions and optimizers, and observing and analyzing the results. Furthermore, we compared our model with other state-of-the-art models to identify areas for improvement and understand better how our model can be enhanced. This holistic experience has not only bolstered our theoretical knowledge but also enhanced our practical skills in implementing and refining deep learning models for complex tasks such as medical image segmentation.

REFERENCES

[1] Jeong, Jin Gyo, et al. "DALU-Net: Automated Liver Segmentation and Volumetry for Liver Transplantation in Abdominal Computed Tomography Volumes." (2021).

[2] Li, Huawei, and Changying Wang. "MIS-Net: A deep learning-based multi-class segmentation model for CT images." Plos one 19.3 (2024): e0299970.

[3] Nguyen, Hai Thanh, et al. "Denoising with Median and Bilateral on CT images for Liver segmentation." 2022 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE, 2022.