Liver Tumor Segmentation from CT Images using ResUNet

Anwesha Samaddar, Sharan Rajamanoharan and Manpreet Singh
University of Minnesota, Twin cities

***Abstract*—Medical image segmentation plays an important role in field of medical diagnostics by enabling identification, interpretation and analysis of complex structures. Different type imaging techniques like magnetic resonance imaging (MRI),ultrasound and computer tomography (CT) are used to detect malignant or benign tumors present in various organs. We will specifically explore the image segmentation techniques to liver CT scans. The complex structure of the liver tissues, combined with the lack of color in the images, makes analyzing the CT output images difficult. So the goal of liver image segmentation is to detect abnormal patterns in CT images and segment them for in-depth analysis. This becomes an important area of research with significant implications for disease diagnosis and treatment planning. Our focus will be on integration of deep neural networks in automating the segmentation process using a large dataset of CT images. We will explore a popular architecture in image segmentation known as U-net and investigate an enhanced version of this architecture called ResUnet. We will showcase the resulting segmentation produced by this architectures. Additionally, we will analyze their accuracy using dice loss.**

***Index Terms*—U-net, liver segmentation, ResUnet, Dice loss**

I. Introduction

The liver is a vital organ in our abdomen which plays crucial roles in metabolism and digestion. Its connection with other digestive organs emphasizes the significance of understanding issues associated with its functioning. One primary complication that arises in the liver is cancer, which can lead to the uncontrolled growth of harmful cells. Oncologists employ different types of imaging techniques, such as MRI and CT scans to identify these abnormal growths and determine whether they are malignant or benign. However, the understanding of these images requires a high level of expertise in the field of oncology. Moreover, high-resolution imaging is necessary to study complex structures like the liver.

While these imaging approaches are useful, they present few difficulties. The manual interpretation of medical images makes the process subjective and susceptible to errors. Additionally, obtaining high-resolution images to study complex structures increases the overall cost. These challenges highlight the need for image processing techniques to improve the readability of images, reducing the subjectivity, error and requirement of expensive equipment. Techniques such as image segmentation are used to highlight crucial parts of an image, typically anomalies like tumours.

To address these challenges, research has focused on creating automatic segmentation of tumours using deep neural networks. The advantage of deep neural networks is that they develop over time to improve their performance by learning from various inputs and conditions. These are called convolution neural networks(CNN). This deep learning network learns to recognize complex features by classifying and integrating features derived from earlier layers. Convolution neural networks are versatile and can be used for complex image segmentation tasks, object detection/localization, image super-resolution, image generation, style transfer and medical image analysis.

CNNs and traditional image processing methods have been utilized to achieve state-of-the-art results for applications like MRI image segmentation, tumour detection in CT scans, and lung segmentation in X-ray images for different diagnosis purposes. Since CNN networks perform well, their performance depends on the quality and quantity of data provided and the computing power available. Hybrid approaches have emerged to leverage the strengths of classical methods and deep learning techniques and address issues like low data availability and resource constraints for computation.

Krizhevsky et al. [1] achieved high performance in CNN network by supervised training of AlexNet, which had eight layers and millions of parameters using the ImageNet dataset with 1 million training pictures. In some biomedical image processing applications, the intended output should contain localization, meaning a class label should be applied to each pixel. Ciresan et al. [2] used a sliding-window design to train a network to predict the class label of each pixel by supplying a local area (patch) around that pixel.

The method presented by Ciresan et al. [2] has two significant flaws. It is not computationally efficient since the network requires independent execution for each image patch, resulting in redundant feature accumulation from overlapping patches. Secondly, there is a trade-off between determining object location and including contextual information. Larger patches may require more max-pooling layers, reducing localization accuracy, but when patches are too small, it restricts the network's learning of contextual information. Some approaches have addressed these limitations (Seyedhosseini et al. [3], Hariharan et al. [4]) by introducing classifier outputs that leverage features from multiple layers, enabling accurate localization while also making effective use of contextual information, overcoming the challenges posed by the previous strategy.

Region-based Convolutional Neural Network (R-CNN) designed by Girshick et al. [5] takes an input image and proposes around 2000 regions/blobs to provide meaningful data for

the classification algorithm to detect objects. However, the drawback of this method is that it is relatively slow and requires thousands of forward convolutions, making it unreliable for practical applications. Even end-to-end training might be difficult because of the two-step training procedure (it requires pre-training on a CNN for feature extraction and then fine-tuning for object recognition).

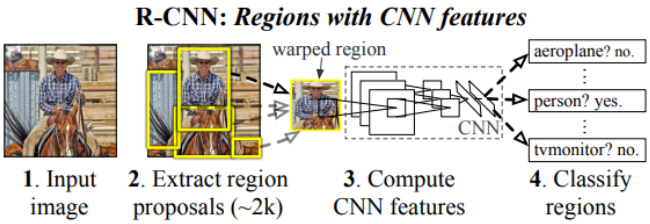

Fig. 1: R-CNN architecture. Source from [5].

The U-Net [6] is a convolutional neural network architecture that has become famous in medical imaging due to its unique U-shape network design and high performance in image segmentation with few training images. In Fig. 2, we can see this special design and how U-net works. This network captures fine details and spatial relationships in an image. In medical imaging, it became successful because it uses less data to generate good results and can capture complex structure of images. This capture of intricate details allows U-net to segment complex tissues of organs from tumours present.

Despite giving good results, U-net can fail to generalize the capturing of all the details for various types of organs. So, we might need to modify the U-net architecture or add some other type of network to improve its performance for our liver CT scans.

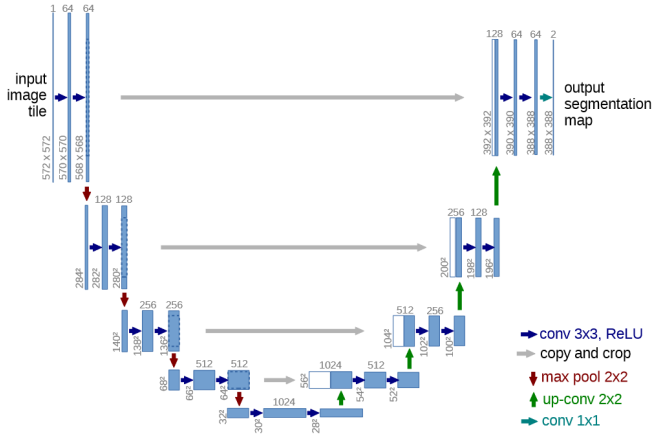

Fig. 2: U-net architecture. Source from [6]. U-net architecture. Each blue box represents a multi-channel feature map. The number of channels is indicated on the box's top. The x-y size is specified at the box's lower left side. White boxes indicate the duplicated feature maps. The arrows represent various operations as indicated in the diagram.

Wen Li et al. [7] used a Gaussian smoothing filter to reduce noise in CT scans. The resulting photos were then normalized and downsampled to speed up the training process. These pre-processed images ware fed into five CNNs each having different patch sizes 13×13, 15×15, 17×17, and 19×19. The CNN with the patch size of 17×17 outperformed others.

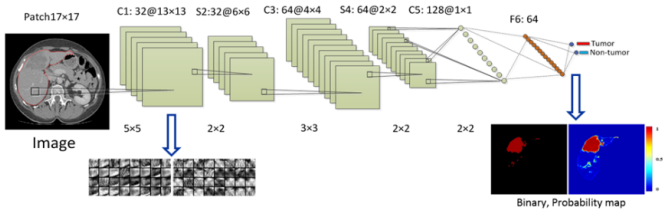

Fig. 3: The architecture of CNNs for liver tumor segmentation. Source from [7].

Avi Ben-Cohen et al. [8] used a Fully Convolutional Network (FCN) for liver segmentation and metastasis identification in CT scans in their study. They used FCN which displayed significant success in the segmentation process. The assessment was carried out on a very limited dataset of 20 patients, which included 68 lesions and 43 livers in a single slice, as well as 20 separate individuals for 3D liver segmentation. Following data augmentation, two networks were trained: one for liver segmentation, which segregated surrounding organs, and another for tumor and lesion segmentation, which used the first network's output. The findings used cross-validation that gave a true positive rate of 0.86 and a false positive rate of 0.6 for each patient.

Zhengxin Zhang et al. [9] proposed a unique solution to handle the difficult challenge of extracting roads from aerial photos via remote sensing. Road extraction is critical for applications such as autonomous vehicle navigation, map generation, and geographical data updates. U-Net model was unable to give satisfactory results, so the authors developed ResUNet, a revolutionary technique. ResUNet deviates from the standard convolutions used in classic U-Net models by employing residual blocks with skip-connections. This novel approach results in speedier training and better performance, especially in settings with limited data.

II. Methods

A. Dataset Description

We shall be using the 3D-IRCADb-01 [10] liver segmentation dataset for tumor detection. The dataset includes 3D CT scans from ten female and ten male patients, 75% of whom had hepatic tumors. Hepatic tumors, commonly referred to as liver tumors, are abnormal cell growths in the liver. These tumors are may be as benign (non-cancerous) or malignant (cancerous). The dataset is divided into 20 files, each of which corresponds to a separate patient. The dataset also contains information about the images, such as liver dimensions (width, depth, and height), tumor sites determined by Couninaud's

segmentation, and possible obstacles for liver segmentation software. Each CT scan is divided into 2,800 2D slices, each with a mask for various organs (Liver, Tumors, Bones, Arteries, Kidneys, and Lungs).

The folders within the dataset follow the naming convention of "3D-IRCADb-01-number" where the number ranges from 01 to 20. Each "3D-IRCADb-01-number" directory consists of four subfolders called "PATIENT_DICOM," "LABELLED_DICOM," "MASKS_DICOM," and "MESHES_VTK." Patient's 3D CT scans in DICOM format, labeled images representing segmented zones of interest in DICOM format, a collection of subdirectories named after the segmented zones containing DICOM images for each mask, and finally all files corresponding to surface meshes of the segmented zones of interest in VTK format can be found within these subdirectories.

An illustration of our proposed workflow is shown in Fig.4.

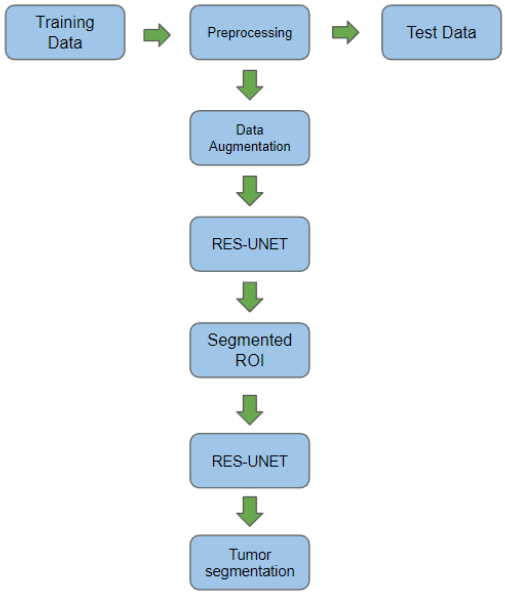

Fig. 4: Our proposed workflow for liver segmentation and tumor detection.

B. Data Preprocessing

1) **Hounsfield Windowing:**The Hounsfield Unit (HU) is a standardized metric in CT imaging that measures X-ray attenuation as it passes through tissues. This scale is supported by two reference points; distilled water, which has a value of zero HU, and air, which has a value of -1000 HU due to its low X-ray attenuation. It is encoded in a 12 bits scale, with a total of 4096 potential values ranging from -1024 HU for air to 3071 HU. Hounsfield Units give a quantifiable way to quantify tissue density, which aids in structure distinction in medical imaging. Hounsfield Units are calculated by comparing the attenuation of X-rays in a certain voxel to that of pure water under typical circumstances. The formula includes the linear attenuation coefficient (μ), which represents the X-ray absorption properties of the tissue. This mathematical connection gives the exact definition of tissue density, which aids diagnostic interpretation in CT images. HU [11] is formulated as:

$$HU = 1000 \times \frac{\mu_{\text{tissue}} - \mu_{\text{H2O}}}{\mu_{\text{H2O}}}$$

μ_{tissue} : linear attenuation coefficient of the tissue
μ_{H2O} : linear attenuation coefficient of distilled water

The grayscale intensity given to each pixel is determined by these HU values, with higher HU values resulting in brighter pixels. After reading DICOM slides, Hounsfield Windowing is used to limit the display range to $[-100, 400]$. This windowing improves the visibility of picture details, as seen in Figures 3.2 and 3.3.

HU values for body organs	
Bone	1000
Liver	40 to 60
White Matter	46
Grey Matter	43
Blood	40
Muscle	10 to 40
Kidney	30
Cerebrospinal	15
Water	0
Fat	-50 to -100
Air	-1000

Fig. 5: HU values for different body organs. Source from [12].

2) **Hisogram Equalization:** Histogram Equalization (HE) is a method used to improve visual contrast in images, especially when contrast is confined to a small range and is not uniformly distributed [13]. Applying histogram equalization after Hounsfield windowing to an image improves contrast notably between the liver and surrounding organs.
This method aids in bringing out features that may be concealed in areas with low contrast at first. Following that, we normalize our image to a range of [0, 1], maintaining pixel uniformity and allowing subsequent image processing or analysis. These steps enhance visual representation and processing of medical images especially while identifying the liver from the surrounding tissues.
3) **Data Augmentation:** In order to add variation to our dataset and prevent overfitting we use augmentation

techniques; this helps the model to generalize well on unseen samples or perform well on the validation set and solve class imbalance issues. In our dataset, the images have pixels that do not contain tumors much higher in number than the pixels that do contain tumors. To create a balanced dataset we applied a set of image transformation techniques like random rotation, cropping, flipping along y-axis, adding noise and translation.

This dataset contains separate masks for every tumor. Therefore, for images containing multiple tumors, we had to merge the different masks into a single mask before applying augmentation techniques. A few representative images of our augmented data after applying the preprocessing steps are shown in Fig. 6.

C. ResUnet

To improve upon the performance of U-net further Research was done and a Residual connection based model was developed in the literature Zhang et al [14] called ResUnet. ResUnet takes the advantage of both the UNET architecture and the Deep Residual Learning. ResUnet is designed to give high perfoemance metrics with less parameters . The ResUnet similar to ResNet solves the important problem of vanishing gradient and exploding gradients while optimising the loss. Skip connections within the ResUnet facilitate improved information flow between all layers, enhancing the smooth propagation of gradients during the training process.. Similar to Unet ResUnet is also has the encoder-decoder network as its backbone.

The **encoder** in this architecture processes the input image through three encoder blocks, each constructed with pre-activated residual blocks. These blocks aid the network in learning abstract representations. Importantly, the output of each encoder block serves as a skip connection for the corresponding decoder block. Next this is passed to the Middle layer called Bridge which consists of pre-activated Residual block of stride 2.

The **decoder** in this architecture receives the feature map from the bridge and incorporates skip connections from various encoder blocks. This process facilitates the learning of an enhanced semantic representation, ultimately utilized in generating a segmentation mask. Comprising three decoder blocks, the decoder undergoes a spatial dimension doubling and reduction of number of feature channels after each block.

Initiating with a 2×2 upsampling, each decoder block doubles the spatial dimensions of the feature maps. Subsequently, these expanded feature maps undergo concatenation with the corresponding skip connection from the encoder block. These skip connections play a crucial role in providing the decoder blocks with the features learned by the encoder network. Following the concatenation operation, the feature maps pass through a pre-activated residual block.

D. Dice Loss

The general Loss functions used with Segmentation are Cross-entropy loss and dice loss.First introduced in Milletari et al[15] as a loss function for highly unbalanced segmentations, ie where the positive and negative classes are not evenly distributed., it has become more popular for Medical Image segmentation purposes. It measures the dissimilarity between the predicted segmentation and the Ground truth of the input Image. The Dice score coefficient (DSC) is a measure of overlap widely used to assess segmentation performance [16].

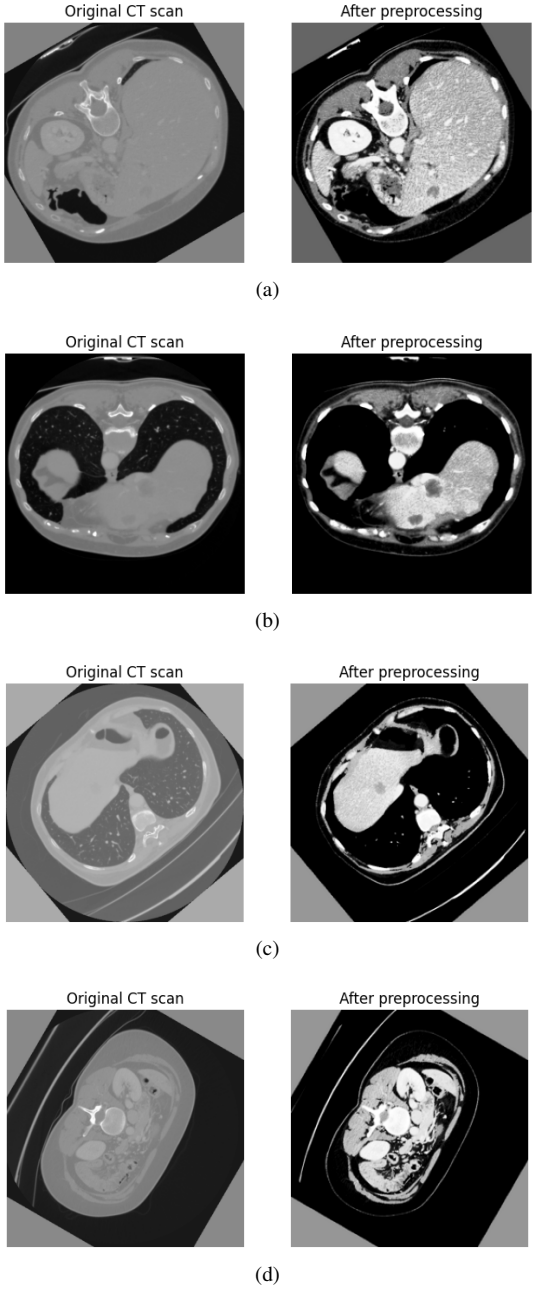

Fig. 6: Examples of original vs. pre-processed images of three different patients (after applying image augmentation functions).

$$DL_2 = 1 - \frac{\sum_{n=1}^{N} p_n r_n + \epsilon}{\sum_{n=1}^{N} (p_n + r_n) + \epsilon}$$

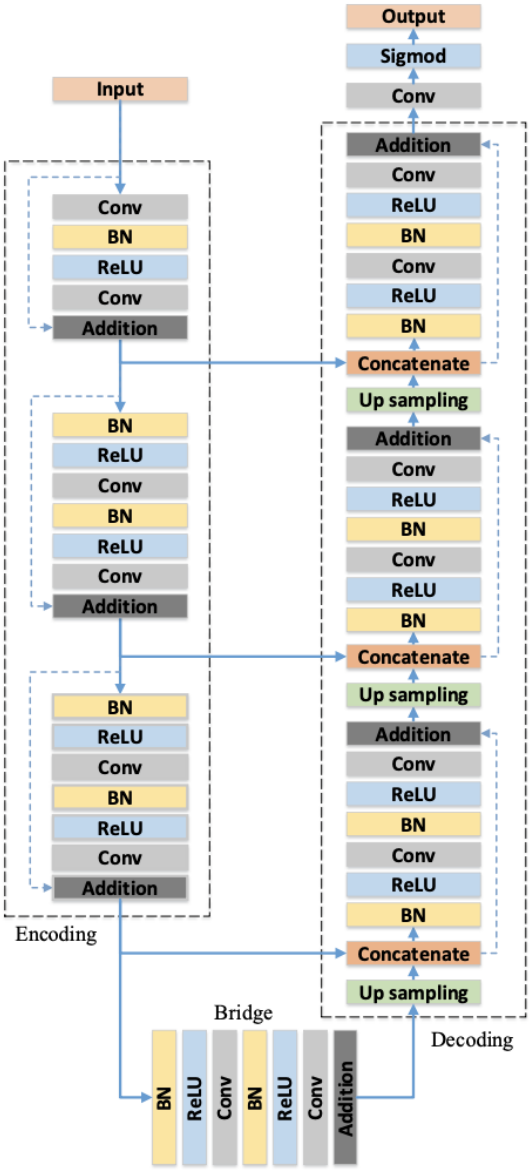

Fig. 7: Res U-net architecture. Source from [14]

where ϵ is a small perturbation which ensures the stability of the Loss function. Using this Loss can help to reduce the model's tendency to prioritize one class over the other. Take, for instance, an image segmentation task where the negative class (background) significantly outweighs the positive class (foreground). Even if a model achieves high accuracy, it might still exhibit subpar performance with respect to Dice Loss if it struggles to correctly segment the foreground [17].

E. Liver Segmentation with ResUnet

We perform liver segmentation with ResUnet to extract the mask corresponding to our region of interest i.e to first segment the liver and then perform tumor segmentation on our region of interest.

F. Tumor Segmentation

Finally we perform tumor segmentation by training a separate ResUnet model to segment liver tumors based on region of interest detected by previous network and report our results.

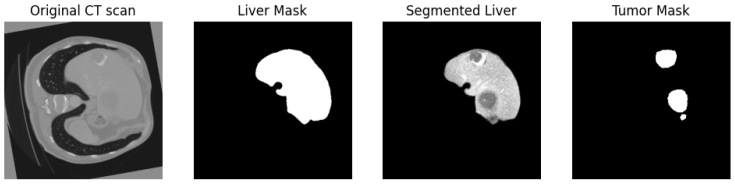

Fig. 8: Our target is to first find liver mask and then tumor mask by training ResUnet

III. Results

A. Evaluation Metrics

Before we discuss our results its imperative to know the evaluation metrics which we used to evaluate the performance of our Model.

The accuracy metric used here is the count of correctly classified pixels regardless of their class, making it a suboptimal metric due to the overwhelming dominance of true negatives, which can skew the evaluation.

One main metric used in this research in **Dice Loss**. it computes the overlap input X and the output Y [18]

$$DiceCoeff = \frac{2|X \cap Y|}{|X| + |Y|}$$

$$\text{Dice} = \frac{2\ X\ Area\ of\ overlap}{Total\ area} = \frac{2\ \text{x}}{}$$

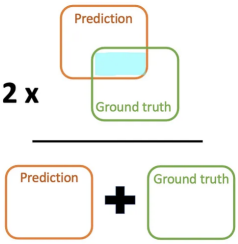

Fig. 9: The visual representation of Dice Coefficient metric. Source:[19]

The Higher dice Coefficient indicates high similarity between the Ground truth and segmented output. The Other metrics used are the common metrics such as **Confusion Matrix**. we also the Jaccard Index (**IoU**) metric. IoU stands for **Intersection over union**, it is calculated as the ratio of the intersection between the predicted segmentation and the ground truth to the total area encompassed by their union [19].

$$IoU = \frac{TP}{TP + FN + FP}$$

$$\text{IoU} = \frac{Area\ of\ overlap}{Area\ of\ union} = \frac{\text{Prediction} \cap \text{Ground truth}}{\text{Prediction} \cup \text{Ground truth}}$$

Fig. 10: The visual representation of IoU metric. Source: [19]

B. Libraries and tools used

We used the tensorflow framework, keras and Google CoLabPro for data extraction and preprocessing, data augmentation, model training, and experimenting with data and number of epochs to check our losses, validation accuracies, dice coefficient and plot confusion matrix.

C. Liver segmentation results

Training ResUnet for liver segmentation with 25% of our data gives the results as shown in Fig. 11. We display the **Input image**, **ground truth mask** and the **segmentation obtained from ResUnet**.

Metric	Training	Validation
Loss	0.1540	0.1560
Accuracy	0.9565	0.9557
Dice Coefficient	0.8460	0.8440

TABLE I: Metrics for ResUnet trained on 25% of data

Metric	Training	Validation
Loss	0.0907	0.1119
Accuracy	0.9555	0.8881
Dice Coefficient	0.9093	0.8881

TABLE II: Metrics for ResUnet trained on 50% of data

Metric	Training	Validation
Loss	0.0242	0.0224
Accuracy	0.9879	0.9878
Dice Coefficient	0.9758	0.9776

TABLE III: Metrics for ResUnet trained on full data

We selected a random sample Fig.12 to visualize the confusion matrix and calculate other comparison metrics like pixel accuracy, IOU score, true positive rate and dice coefficient.

Metric	Value
IoU	88%
Pixel Accuracy	98.86%
True Positive Accuracy	95.80%
Dice Coefficient	0.9365

TABLE IV: Evaluation Metrics on selected sample of liver segmentation - Fig.12.

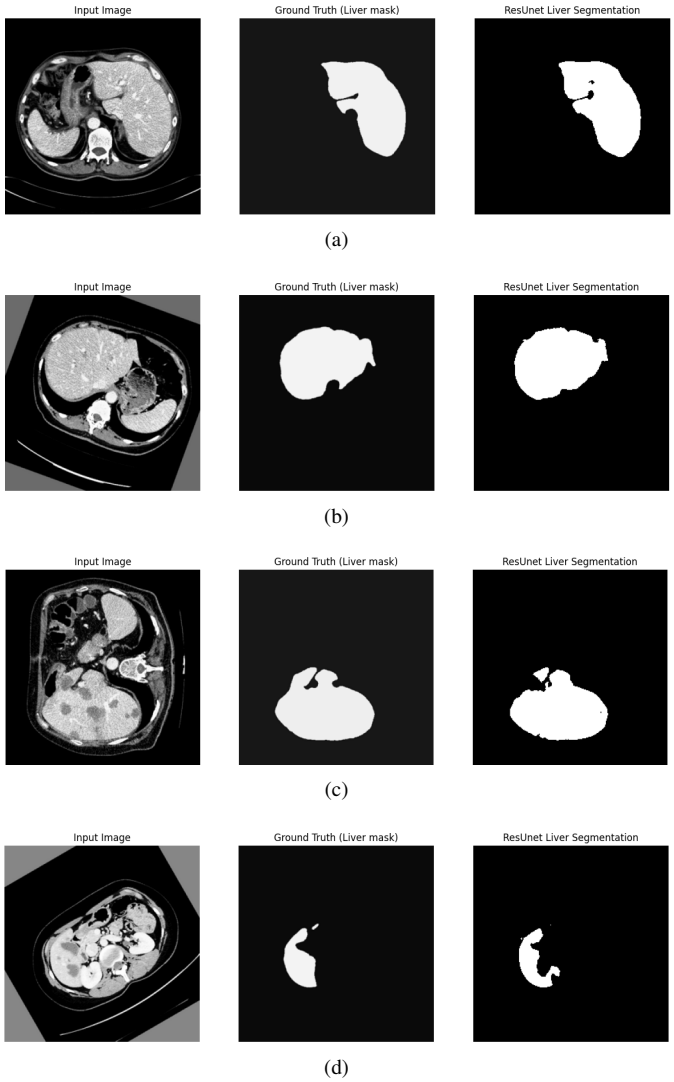

Fig. 11: Liver Segmentation results with ResUnet on the test set.

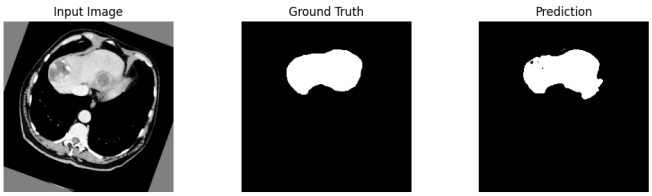

Fig. 12: We calculated our performance metrics for liver segmentation and plot a confusion matrix for this sample. Refer to table IV.

The confusion matrix plot for this liver segmentation example is given in Fig.13

D. Tumor segmentation results

We have trained ResUnet with tumor masks and liver segmented from the original diacom images and masks supplied by the dataset. We plot the confusion matrix for a randomly

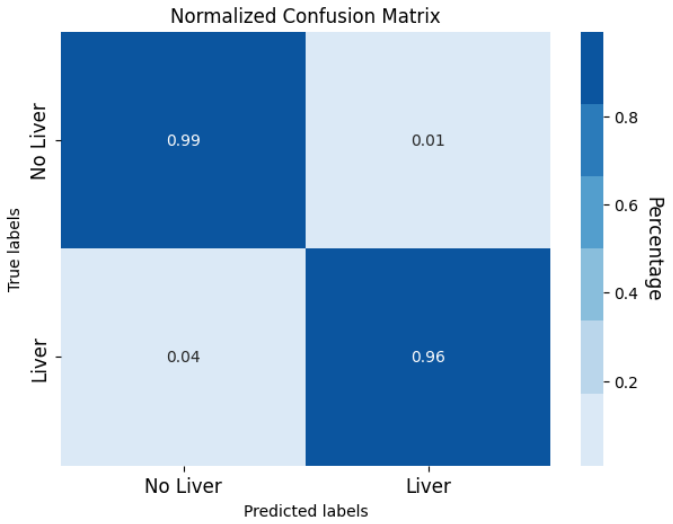

Fig. 13: Normalized Confusion matrix for the liver segmentation sample used in Fig.12.

selected samples. Fig. 16 shows our results obtained for tumor segmentation by training ResUnet.

We selected a random sample Fig.14 to perform tumor segmentation and evaluated the pixel accuracy, IOU score, true positive rate, and dice coefficient.

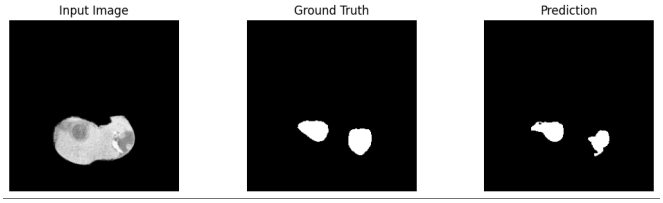

Fig. 14: We calculated our performance metrics of tumor segmentation and plotted a confusion matrix for this sample. Refer to Table V and Fig. 15.

Metric	Value
IoU	71%
Pixel Accuracy	98.98%
True Positive Accuracy	73.41%
Dice Coefficient	0.8274

TABLE V: Evaluation Metrics on selected sample of tumor segmentation- Fig.14

The confusion matrix plot for this example is given in Fig.15.

Metric	Training	Validation
Loss	0.4004	0.4399
Accuracy	0.9950	0.9945
Dice Coefficient	0.5996	0.5601

TABLE VI: Training and Validation Metrics of ResUnet (trained on 25% data) for Tumor Segmentation.

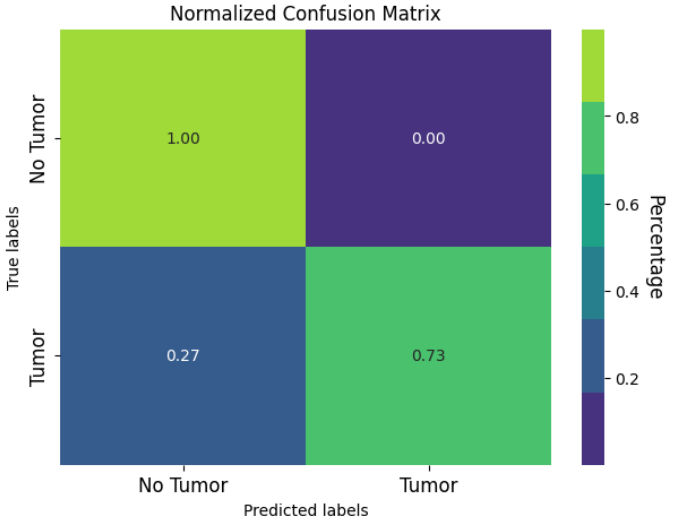

Fig. 15: Normalized Confusion matrix for the tumor segmentation sample used in Fig.14.

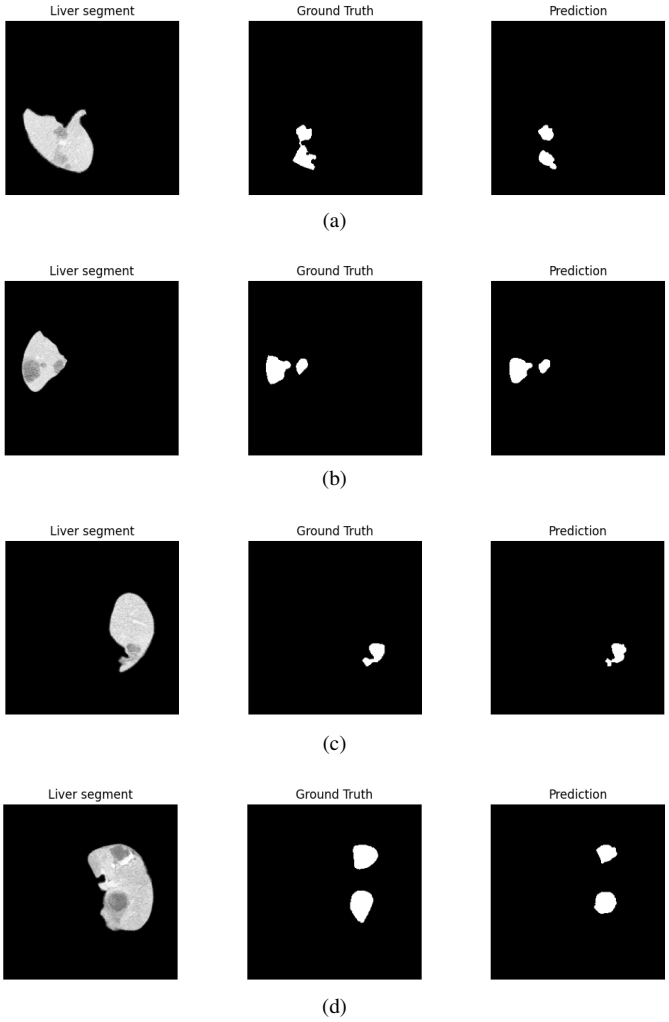

Fig. 16: Tumor Segmentation results with ResUnet on the test set. We display the segmented Liver, the ground truth mask for tumor segmentation and the segmentation using ResUnet.

IV. Discussion

Through the experiment of training with different sizes of dataset we wanted to find the effect which the size of data has on the accuracy metrics and how efficiently our model learns through varying size of the same dataset.

The clear discriminator for the Models trained with 25%, 50% or 75% is the increasing trend of accuracies , IoU metric and dice Coefficient. Better results for each of these models could have been obtained if we have increased the number of epochs, but due to GPU constraints we were unable to do that.

Although ResUnet provides very competitive results compared to other models which we came across in many literature's, But there are also few data's which were not segmented properly and can be seen in Fig-17, Fig-18. This can be explained by the limitations which comes with ResUnet

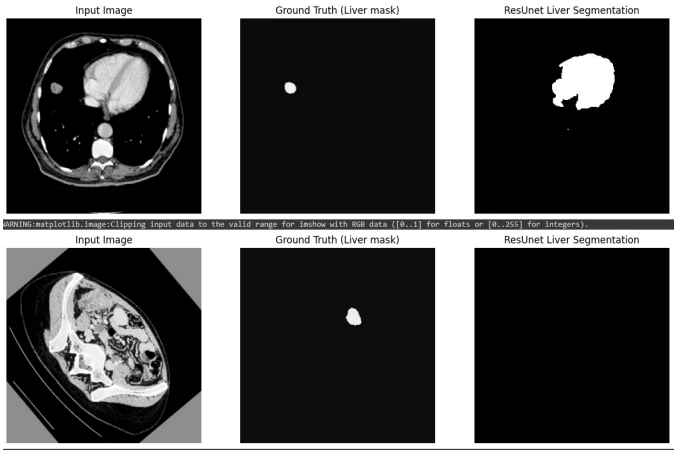

Fig. 17: Few liver segmentation data which were not segmented properly

Few of these limitations are, ResUnet which is a combination of U-Net and residual connections increases the number of parameters, leading to a higher computational cost during training and testing. Also the performance of ResUNet can be sensitive to hyperparameter choices, and finding the optimal set of hyperparameters may require few trial and errors. While ResUNet captures local features effectively, it may struggle with capturing global contextual information. These are just few of the Limitations of ResUnet.

we can overcome these limitations by introducing regularisation techniques such as **Dropout**. we can also systematically search for optimal hyperparameters using techniques like grid search. Now to capture the Global contextual information we can do the following

1) Implement attention mechanisms to allow the model to focus on relevant parts of the input, helping capture long-range dependencies and improving the integration of contextual information.
2) (or) Introduce dilated convolutions to increase the receptive field of the model, enabling it to capture larger contextual information.

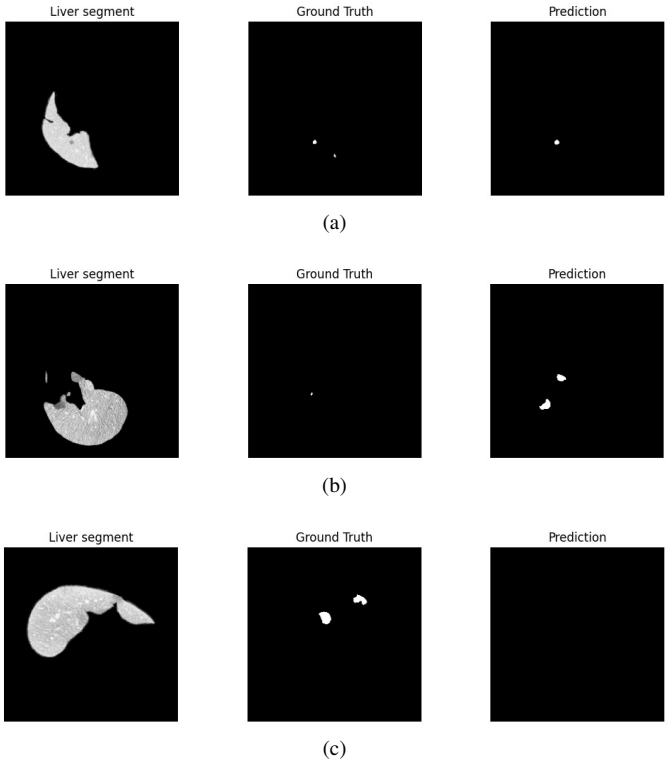

Fig. 18: This Images represents few wrongly segmented tumour on the segmented Liver Image by our ResUnet model.

V. Conclusion

In this study we have conducted in-depth research regarding different segmentation models and different types of Loss function, after reading many literature's we chose ResUnet with Dice Loss as the Loss function. We have taken different percentage of data from the main dataset, to test the Liver segmentation capabilities of our Model at 25%,50%,100% of our original data set. For the tumour segmentation after the Liver segmentation we only used 25% of the data because of computational constraints, ResUNet can be computationally expensive, especially for large and high-resolution images.

We also saw that ResUnet although gives us good segmentation results with limited and complex data, it has a few limitations such as Computational Complexity, Hyperparameter sensitivity and Limited contextual information. Further improvements can be made in the ResUnet architecture by using algorithm such as Grid search to find the optimal hyperparameter values, we can also introduce regularisation using the concept of dropout. We can also research on using Attention based ResUnet Architecture in-order to improve the performance metrics of the ResUnet Architecture.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1106–1114.

[2] D. C. Ciresan, L. M. Gambardella, A. Giusti, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 2852–2860.

[3] M. Seyedhosseini, M. Sajjadi, and T. Tasdizen, "Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks," in *2013 IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 2168–2175.

[4] B. Hariharan, P. Arbelez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," 2014.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2014.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[7] W. J. Li, F. Jia, and Q. Hu, "Automatic segmentation of liver tumor in ct images with deep convolutional neural networks," 2015.

[8] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Fully convolutional network for liver segmentation and lesions detection," in *Deep Learning and Data Labeling for Medical Applications*, G. Carneiro, D. Mateus, L. Peter, A. Bradley, M. R. S. Tavares, V. Belagiannis, P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and Cornebise, Eds.

[9] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *CoRR*, vol. abs/1711.10684, 2017.

[10] "Liver segmentation 3d ircadb-01," https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/.

[11] M. Lev and R. Gonzalez, "Ct angiography and ct perfusion imaging," in *Brain Mapping: The Methods (Second Edition)*, 2nd ed., A. W. Toga and J. C. Mazziotta, Eds. San Diego: Academic Press, 2002, pp. 427–484.

[12] I. J. De Backer, I. W. Vos, O. Vanderveken, D. A. Devolder, M. Braem, D. van Dyck, and W. Backer, "Combining mimics and computational fluid dynamics (cfd) to assess the efficiency of a mandibular advancement device (mad) to treat obstructive sleep apnea (osa)," 05 2019.

[13] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. t. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[14] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *CoRR*, vol. abs/1711.10684, 2017. [Online]. Available: http://arxiv.org/abs/1711.10684

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565–571.

[16] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, *Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations*. Springer International Publishing, 2017, p. 240–248. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-67558-9_28

[17] I. Schumacher, "Dice Loss: A Comprehensive Overview," https://serp.ai/dice-loss/, 2008, [Online; accessed 18-Dec-2023].

[18] B. A. T. C. K. M. H. S. W. W. r. J. F. Zou K.H., Warfield S.K. and K. R., "Statistical validation of image segmentation quality based on a spatial overlap index," 2004.

[19] N. Huynh, "Understanding Evaluation Metrics in Medical Image Segmentation," https://medium.com/mlearning-ai/understanding-evaluation-metrics-in-medical-image-segmentation-d289/#:~:text=In%20contrast%2C%20dice%20coefficient%20and,class%20imbalanced%20datasets%20like%20MIS./, 2023, [Online; accessed 18-Dec-2023].