Medical Image Segmentation using Novel U-Net Variants

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Engineering
in
Computer Science and Engineering

Submitted by

Akash Halder
Regn. No. - 163557 of 2022 - 2023
Exam Roll No. - M4CSE24003
Class Roll No. - 002210502005
Session - 2022 - 2024

Under the supervision of

Prof. Ram Sarkar

Department of Computer Science and Engineering
Jadavpur University
188, Raja S.C. Mallick Rd,
Kolkata - 700032, West Bengal, India

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF RECOMMENDATION

This is to certify that the work embodied in this thesis entitled **"Medical Image Segmentation using Novel U-Net Variants"** has been satisfactorily completed by **Akash Halder** (Registration Number 163557 of 2022 - 2023; Class Roll No. 002210502005; Examination Roll No. M4CSE24003). It is a bonafide piece of work carried out under my supervision and guidance at Jadavpur University, Kolkata, for partial fulfilment of the requirements for the awarding of the **Master of Engineering** in **Computer Science and Engineering** degree of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, during the academic year 2022 - 2024.

Prof. Ram Sarkar,
Department of Computer Science and Engineering,
Jadavpur University.
(Supervisor)

Forwarded By:

Prof. Nandini Mukhopadhyay,
HOD, Department of Computer Science and Engineering,
Jadavpur University.

Prof. Dipak Laha,
DEAN, Faculty of Engineering and Technology,
Jadavpur University.

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled **"Medical Image Segmentation using Novel U-Net Variants"** is a bonafide record of work carried out by **Akash Halder** (Registration Number 163557 of 2022 - 2023; Class Roll No. 002210502005; Examination Roll No. M4CSE24003) in partial fulfilment of the requirements for the award of the degree of **Master of Engineering** in **Computer Science and Engineering** in the Department of Computer Science and Engineering, during the period of September 2022 to May 2024. It is understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the thesis only for the purpose of which it has been submitted.

_____

Signature of Examiner 1
Date:

_____

Signature of Examiner 2
Date:

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that the thesis entitled **"Medical Image Segmentation using Novel U-Net Variants"** contains literature survey and original research work by the undersigned candidate, as a part of his degree of **Master of Engineering** in **Computer Science and Engineering, Jadavpur University**. All the information have been obtained and presented in accordance with academic rules and ethical conduct.
I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Akash Halder

Examination Roll No.: M4CSE24003

Registration No.: 163557 of 2022 - 2023

Thesis Title: Medical Image Segmentation using Novel U-Net Variants

Signature of the Candidate:

. . . . . .

Dedicated to my unwavering pillars of support,
my parents,
whose love and sacrifice have fueled every chapter of my journey.

. . . . . .

Acknowledgements

I tender my deepest gratitude and high sense of indebtedness to my guide, **Dr. Ram Sarkar**, Professor, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for his constant support and supervision in carrying out this thesis, which has been instrumental in the materialisation of this work. His expertise and guidance was crucial in shaping both the direction and the quality of this thesis, and I am profoundly thankful for his steadfast dedication.

I express my sincere gratitude to the Head of the Department, **Dr. Nandini Mukhopadhyay**, Professor, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for being kind enough to permit me to carry out my research in this prestigious institution.

I must express my sincere gratitude to **Prof. Mahantapas Kundu**, Coordinator of **Centre for Microprocessor Applications for Training, Education and Research (CMATER) Lab**, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for granting me permission to conduct this study in the renowned CMATER lab and rendering valuable suggestions.

My special thanks to my teachers, **Prof. Debotosh Bhattacharya**, **Prof. Subhadip Basu**, and **Dr. Mita Nasipuri (former Professor)**, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for imparting their wealth of knowledge to me.

I thank my seniors, Mr. Sujan Sarkar, Dr. Neelotpal Chakraborty, Ms. Payel Pramanik, Mr. Arup Sau, and Ms. Rakhi Tewari, labmates, Mr. Surya Majumder, Mr. Shivam Singh, Mr. Sutirtha Mukherjee, Mr. Debasmit Roy, Mr. Amartya Ray, Mr. Soumyajit Gayen, Mr. Diptarka Mandal, Ms. Debarati Paul, Mr. Abhijit Sarkar, Mr. Suman Ghosh, Ms. Priyanka Ghosh, Mr. Siddhanth Majumder, Mr. Soumyadip Sahoo, Ms. Sanchita Das, Ms. Godhuli Das, and Mr. Ankit Das, and friends, Ms. Tirtha Das, Mr. Shankha Shubhra Das, Mr. Saikat Mazumdar and Mr. Rajarshi Mitra for their unwavering support, validation and understanding throughout this journey. Their companionship and camaraderie have been invaluable to me.

I must also extend a special word of thanks to my parents and family members for the encouragement and support they have rendered in preparing this thesis.

Akash Halder

Examination Roll No.: M4CSE24003

Registration No.: 163557 of 2022 - 2023

Department of Computer Science and Engineering,

Jadavpur University.

Abstract

In the realm of medical imaging, precise segmentation holds the key to unlocking invaluable insights for diagnosis and treatment. Leveraging the power of deep learning, this thesis explores innovative methodologies to enhance the accuracy and efficiency of medical image segmentation. Recently, deep learning models, particularly U-Net-like architectures, have obtained notable success in medical image segmentation. Ensemble learning is a powerful approach that helps leverage the performance of an overall model by incorporating the decisions of multiple models. We propose an ensemble learning model, called BLUE-Net, using a combination of three models, namely the basic U-Net, Link-Net, and FPN. We combine the decisions of these models with a fuzzy-ranking scheme by using a modified Blumberg function to decide whether a pixel in an image is a part of a foreground pixel of the segmentation mask. We also propose a novel model, called SEAU-Net, which employs attention to the lower skip connections only, effectively integrating SE attention blocks within the skip connections of the encoder backbone. High-level semantic features are ensured by these attention-aided feature-maps from the lower layers to the higher layers in the decoder that are effectively utilized for the reconstruction of segmentation maps. Recognizing key features from input images is a potent strategy to enhance overall model performance. One way to achieve this is by integrating decisions from multiple outputs. Hence, we introduce the W^2CU-Net, which is a dual decoder network based on Cosine Similarity, Weibull function and Weierstrass' Elliptic Curve Equation. It is designed for liver and liver tumor segmentation from CT scan images. The model employs an ensemble scheme, which combines the decisions of two decoders using a modified Weierstrass' elliptic curve equation while selecting important features using Cosine Similarity and Weibull function, to determine whether a pixel in a CT scan image belongs to the foreground of the segmentation mask. Three models are evaluated across six datasets: 3DIRCADb, CHAOS, LiTS for liver and liver tumor segmentation, BUSI for breast cancer segmentation, OCDC for oral cancer segmentation, and ISIC 2018 for skin cancer segmentation.

Contents

Abstract i

Table of Contents iii

List of Figures vii

List of Tables ix

List of Abbreviations xi

1 Introduction 1
- 1.1 Objectives . . . 1
- 1.2 Scope of the Thesis . . . 2
- 1.3 Organization of the Thesis . . . 5

2 Literature Survey 9
- 2.1 Liver Cancer . . . 9
- 2.2 Breast Cancer . . . 11
- 2.3 Oral Cancer . . . 12
- 2.4 Skin Cancer . . . 13

3 Methods and Materials 17
- 3.1 Datasets . . . 17
  - 3.1.1 3DIRCADb . . . 18
  - 3.1.2 CHAOS . . . 20
  - 3.1.3 LiTS . . . 20
  - 3.1.4 BUSI . . . 21
  - 3.1.5 OCDC . . . 22
  - 3.1.6 ISIC 2018 . . . 22
- 3.2 Methods . . . 24
  - 3.2.1 BLUE-Net . . . 24
    - 3.2.1.1 U-Net . . . 25
    - 3.2.1.2 Link-Net . . . 25
    - 3.2.1.3 Feature Pyramid Network . . . 25
    - 3.2.1.4 Blumberg Function based Ensemble Method . . . 25
    - 3.2.1.5 Significance of Blumberg Function . . . 26
    - 3.2.1.6 Implementation of the Ensemble Model . . . 26
    - 3.2.1.7 Loss functions . . . 28

3.2.2 SEAU-Net . . . 30
3.2.2.1 Tri-Step Residual Block . . . 32
3.2.2.2 Squeeze-and-Excitation Block . . . 33
3.2.2.3 Loss Function . . . 34
3.2.3 W^{2}CU-Net . . . 35
3.2.3.1 Deep Neural Based Backbone for Encoder . . . 35
3.2.3.2 MobileNetV2 . . . 36
3.2.3.3 Proposed Model Architecture . . . 36
3.2.3.4 Implementation of the Entire Model . . . 36
3.2.3.5 Detailed Cosine Similarity based Feature Pruning Algorithm . . . 37

4 Results and Discussion **47**
4.1 System Configuration . . . 47
4.2 Evaluation Metrics . . . 47
4.3 BLUE-Net . . . 49
4.3.1 GradCAM Analysis . . . 52
4.3.2 Error Analysis . . . 53
4.4 SEAU-Net . . . 54
4.4.1 Ablation Study . . . 54
4.4.2 GradCAM Analysis . . . 55
4.4.3 Segmentation Qualitative Analysis . . . 55
4.5 W^{2}CU-Net . . . 57
4.5.1 Ablation Study . . . 61
4.5.2 Comparison with other Models . . . 61
4.5.3 t-SNE Plots . . . 64
4.5.4 IoU Heatmaps . . . 64

5 Conclusion **67**
5.1 Limitations . . . 68
5.2 Scope for Future Work . . . 69

List of Publications **71**

Bibliography **73**

List of Figures

1.1 A few complicated segmentation instances. The datasets considered are: (a) 3DIRCADb, (b) LiTS, (c) BUSI, (d) ISIC 2018 and (e) OCDC. 4

3.1 Visualization of the 3DIRCADb dataset - Column1: Original CT scans, Column2: HU-clipped CT images, Column3: Liver Masks, Column4: Tumor Masks 19
3.2 Examples of breast ultrasound images and associated masks in the BUSI dataset. 21
3.3 A sample image with multiple masks, and the single mask obtained after merging them. 22
3.4 A few sample histopathology images from the OCDC dataset. (a) WSIs and (b) their corresponding masks. 23
3.5 A few sample dermoscopy images from the ISIC 2018 dataset. Column1: the images and Column2: their corresponding masks. 23
3.6 Pipeline of the proposed Blumberg function based ensemble model used for liver and tumor segmentation 24
3.7 Graphical representation of the modified Blumberg function used in the present work. 27
3.8 Architecture of the proposed SEAU-Net 30
3.9 Decoder block architecture 32
3.10 Architecture of the proposed TSR block. 33
3.11 Architecture of Squeeze-and-Excitation block 34
3.12 Architecture of the MobileNetV2 model 36
3.13 Architecture of the proposed W^2CU-Net 37
3.14 Residual_Conv_Block in W^2CU-Net 39
3.15 Modified Weibull function used in the present work. 40
3.16 Modified Weierstrass' Elliptic Curve Equation. 43

4.1 Loss curves for BLUE-Net obtained by using the hybrid loss function. Here x-axis denotes epoch number and y-axis denotes loss value 51
4.2 GradCam of the Models on 3DIRCADb Dataset: (a) Original Image, (b) Heatmap of U-Net, (c) HeatMap of Link-Net, (d) HeatMap of FPN, (e) GradCam of U-Net, (f) GradCam of Link-Net, (g) GradCam of FPN, for Liver Segmentation. 52
4.3 Some undesirable outputs for liver segmentation (the first 2 images) and tumor segmentation (the next 2 images) 53

4.4 Class activation maps for different images using the trained models. The first two columns show the original image and its corresponding mask. The class activation maps are captured at different layers, namely, the *batch_ normalization_ 8* and *conv2d_ 14*. The first two rows show the results for the 3DIRCADb images and the last two rows show the results for the CHAOS images. 57
4.5 A few good segmentation results obtained using SEAU-Net on the BUSI dataset. 58
4.6 A few superimposed segmentation results obtained using the SEAU-Net on the ISIC 2018 dataset. 59
4.7 Loss and Dice Coefficient curves for the W^{2}CU-Net across the datasets 60
4.8 Qualitative analysis presented for different models on the 3DIRCADb dataset. 63
4.9 Pixel-wise t-SNE for liver and tumor segmentation on 3DIRCADb and LiTS dataset for both decoders. 64
4.10 IOU heatmaps for liver segmentation on 3DIRCADb and LiTS and tumorous cell segmentation on the OCDC dataset for both decoders and final ensemble 65

List of Tables

4.1 Hyperparameters of the base classifiers for the BLUE-Net model. . 49
4.2 Segmentation results obtained for BLUE-Net. 50
4.3 Ablation study for liver segmentation on the 3DIRCADb dataset for BLUE-Net. 50
4.4 Ablation study for tumor segmentation on the 3DIRCADb dataset for BLUE-Net. 50
4.5 Hyperparameters considered for training the SEAU-Net model . . . 54
4.6 Segmentation results obtained for SEAU-Net. 54
4.7 Segmentation results by ablation study on the 3DIRCADb dataset. 56
4.8 Segmentation results by ablation study on the CHAOS dataset. . . 56
4.9 Hyperparameters considered for training the W^2CU-Net model. . . 57
4.10 Segmentation results obtained for W^2CU-Net. 59
4.11 Ablation study conducted on the 3DIRCADb dataset for liver segmentation. 61
4.12 Ablation study conducted on the 3DIRCADb dataset for liver tumor segmentation. 61
4.13 Ablation study conducted on the LiTS dataset for liver segmentation. 62
4.14 Ablation study conducted on the LiTS dataset for liver tumor segmentation. 62
4.15 Quantitative analysis on the 3DIRCADb dataset for the segmentation of liver. 62
4.16 Quantitative analysis on the 3DIRCADb dataset for the segmentation of liver tumor. 62
4.17 Quantitative analysis on the LiTS dataset for the segmentation of liver. 63
4.18 Quantitative analysis on the LiTS dataset for the segmentation of liver tumor. 63

List of Abbreviations

Symbols

3DIRCADb	**3D I**mage **R**econstruction for **C**omparison of **A**lgorithm **D**atabase
A	
AL	**A**ctive **L**earning
ANU-Net	**A**ttention-based **N**ested **U**-**N**et
AS-Net	**A**ttention **S**ynergy **N**etwork
ASN	**A**utomatic **S**egmentation **N**etwork
ASPP	**A**trous **S**patial **S**yramid **P**ooling
B	
BCDU-Net	**B**i-**D**irectional **C**onvLSTM **U**-Net
BLUE-Net	**B**lumberg function-based **E**nsemble **N**etwork
BUSI	**B**reast **U**ltrasound **I**mages
C	
CHAOS	**C**ombined (CT and MRI) **H**ealthy **A**bdominal **O**rgan **S**egmentation
CIE	**C**ommission **I**nternationale de l'**E**clairage
CL	**C**osine Similarity Feature Pruning **L**ayer
CNN	**C**onvolutional **N**eural **N**etwork
CSAF	**C**hannel and **S**patial **A**ttention **F**usion
CT	**C**opmuted **T**omography
D	
DBU-Net	**D**ual **B**ranch **U**-Net
DICOM	**D**igital **I**maging and **C**ommunications in **M**edicine

DL	**D**eep **L**earning
DRAU-Net	**D**eep **R**esidual **A**ttention-based **U**-shaped **N**etwork
DSC	**D**ice **S**core **C**oefficient
E	
EPSA	**E**fficient **P**yramid **S**queeze **A**ttention
F	
FCN	**F**ully **C**onvolutional **N**etwork
FD-Net	**F**eature **D**istillation-**N**et
FPN	**F**eature **P**yramid **N**etwork
FRAU-Net	**F**ractal **R**esidual **U**-**N**et
G	
GAN	**G**enerative **A**dversarial **N**etworks
GPU	**G**raphics **P**rocessing **U**nit
GradCAM	**G**radient-weighted **C**lass **A**ctivation **M**ap
H	
HSV	**H**ue **S**aturation **V**alue
HU	**H**ounsefied **U**nit
I	
IoU	**I**ntersection over **U**nion
ISIC	**I**nternational **S**kin **I**maging **C**ollaboration
L	
LCAF	**L**ocal **C**ross-**C**ttention **F**eature fusion
LiTS	**L**iver and **T**umor **S**egmentation
LSTM	**L**ong **S**hort-**T**erm **M**emory
M	
MA-Net	**M**ulti-scale **A**ttention **N**et
MDSC-Net	**M**ulti-scale **D**epthwise **S**eparable **C**onvolutional neural **N**etwork
MFS-Net	**M**ulti-**F**ocus **S**egmentation **N**etwork

MICCAI	**M**edical **I**mage **C**omputing and **C**omputer **A**ssisted **I**ntervention
MLF	**M**ulti-**L**ayer **F**usion
MRI	**M**agnetic **R**esonance **I**maging
Ms RED	**M**ulti-scale **R**esidual **E**ncoding and **D**ecoding **N**etwork
N	
NII	**N**euro-**I**maging **I**nformatics
O	
OCDC	**O**ral **C**avity-**D**erived **C**ancer WSI dataset
OED	**O**ral **E**pithelial **D**ysplasia
P	
PGMF	**P**rior **G**uided **M**ulti-scale **F**usion
R	
R-CNN	**R**egion-based **C**onvolutional **N**eural **N**etwork
RASM	**R**esidual **A**ttention **S**kip **M**odules
RGB	**R**ed **G**reen **B**lue
RL	**R**andom **L**earning
RNN	**R**ecurrent **N**eural Network
RoI	**R**egion of **I**nterest
RVD	**R**elative **V**olume **D**ifference
S	
SBANet	**S**egmentation-**B**ased **A**ttention **N**etwork
SE	**S**queeze-and-**E**xcite
SEAU-Net	**S**queeze-and-**E**xcitation aided **A**ttention **U**-Net
SoTA	**S**tate-of-**T**he-**A**rt
SPIR	**S**pectral **P**resaturation with **I**nversion **R**ecovery
SVS	Aperio **S**canScope **V**irtual **S**lide
T	
t-SNE	**t**-Distributed **S**tochastic **N**eighbor **E**mbedding

W

W^{2}CU-Net	**W**eibull **W**eirstrass function with **C**osine Similarity-based **N**etwork
WL	**W**eibull **L**ayer
WSI	**W**hole **S**lide **I**mages

CHAPTER 1

Introduction

In recent years, the usage of DL techniques has revolutionized the landscape of medical image analysis, particularly in the domain of cancer segmentation. Accurate delineation of tumor regions from medical images is paramount for precise diagnosis, treatment planning, and monitoring of cancer progression. DL, with its remarkable ability to automatically learn hierarchical features from raw data, has emerged as a powerful tool for addressing the challenges inherent in cancer segmentation tasks. This work presents a pioneering exploration into the application of DL methodologies for segmentation of medical images of different organs, introducing three novel approaches aimed at enhancing segmentation accuracy and robustness. Drawing upon insights from previous research endeavors [1] , as well as leveraging advancements in DL architectures and methodologies, the study endeavors to push the boundaries of current capabilities in cancer segmentation.

1.1 Objectives

The significance of accurate cancer segmentation cannot be overstated. It facilitates the localization and characterization of tumor regions, enabling clinicians to make informed decisions regarding patient management and treatment strategies. However, the task of cancer segmentation is fraught with challenges, including the inherent complexity and heterogeneity of tumor tissues, variability in imaging modalities and protocols, as well as the presence of noise and artifacts in medical images. Traditionally, cancer segmentation rely on handcrafted features and heuristic algorithms, which often struggle to capture the nuanced characteristics of tumors and may yield sub-optimal results, particularly in the presence of variability and noise in medical images. In contrast, DL-based approaches offer a promising alternative by leveraging large-scale annotated datasets and powerful neural network architectures [2] to automatically learn discriminative features

directly from the data. Research motivations behind this thesis work include:

- Medical images often contain structures with ambiguous boundaries or variations in appearance, making it challenging for segmentation models to accurately delineate boundaries between different anatomical structures or pathological regions.

- Some medical imaging modalities, such as ultrasound or MRI, may have lower spatial resolution compared to others like CT, leading to difficulties in accurately segmenting small or intricate structures.

- Segmentation models are sensitive to imaging artifacts, such as noise, motion artifacts, or imaging artifacts caused by equipment malfunctions, which can distort image features and compromise segmentation accuracy.

- The human body exhibits complex anatomical structures and pathological conditions with varying shapes, sizes, and appearances, posing challenges for segmentation models to accurately differentiate between different tissues or identify subtle abnormalities.

- Annotated medical image datasets for segmentation tasks are often limited in size due to the difficulty and cost of manual annotation, hindering the training of DL models with sufficient data to learn complex patterns effectively.

- Medical image segmentation tasks frequently involve imbalanced classes, where certain anatomical structures or pathological regions are disproportionately represented in the dataset, leading to biased model training and suboptimal segmentation performance on minority classes.

- Segmentation models trained on data from one imaging modality or patient population may struggle to generalize to unseen data from different modalities or patient demographics due to domain shifts, potentially leading to poor segmentation accuracy and reliability in clinical practice.

1.2 Scope of the Thesis

In this study, three novel DL approaches tailored specifically for medical image cancer segmentation are introduced. These approaches incorporate innovative methodologies and architectures aimed at addressing key challenges in tumor delineation, including the robust handling of heterogeneous tumor tissues, the

mitigation of noise and artifacts in medical images, and the enhancement of segmentation accuracy and generalization capabilities.

A few complicated segmentation instances are shown in Figure 1.1. As observed from the figure, for liver tumor segmentation using CT scans, the challenge lies in accurately delineating inconspicuous tumorous regions. The task is further complicated by the need to distinguish tumorous areas from healthy tissue amidst complex anatomical structures. In breast cancer segmentation with ultrasound images, the absence of clear boundaries between tumorous cells and non-tumorous regions poses a significant challenge. Identifying tumorous regions accurately requires overcoming the ambiguity present in ultrasound images. In histopathology images for oral cancer segmentation, precise delineation is hindered by tissue structure variations and staining artifacts, requiring sophisticated analysis techniques. Accurate segmentation in these images is crucial for effective diagnosis and treatment planning in oral cancer cases. For skin cancer segmentation, distinguishing between malignant and benign lesions is crucial for accurate diagnosis and treatment. High-resolution dermoscopic images present challenges in precisely identifying tumorous regions amidst complex skin structures and various textures.

A thorough evaluation of the three novel DL models is conducted across six distinct datasets, each presenting unique challenges in cancer segmentation across various modalities. For liver and liver tumor segmentation, the 3DIRCADb, CHAOS, and LiTS datasets are employed, all derived from CT scans, where the complexity lies in accurately delineating liver anatomy and detecting tumors amidst irregular boundaries amid organs. Additionally, evaluations are extended to ultrasound imaging with the BUSI dataset, focusing on breast cancer segmentation, where challenges arise from imaging noise. Furthermore, histopathological images from the OCDC dataset for oral cancer segmentation is also included, addressing the intricacies of analyzing tissue microstructures and identifying malignant regions. Finally, the models are assessed on the dermatology domain with the ISIC 2018 dataset for skin cancer segmentation, navigating the complexities of lesion detection from dermoscopic images.

- Ensemble learning is used to address the limitations of individual models by combining their predictions, which can enhance segmentation accuracy and generalization capability. By aggregating predictions from basic U-Net[3], Link-Net[4] and FPN [5] models, the proposed BLUE-Net leverages the strengths of each architecture to improve segmentation accuracy and reliability. Ensemble learning can help mitigate the effects of ambiguous boundaries, low spatial resolution, imaging artifacts, and imbalanced classes by leveraging diverse model architectures and training strategies.

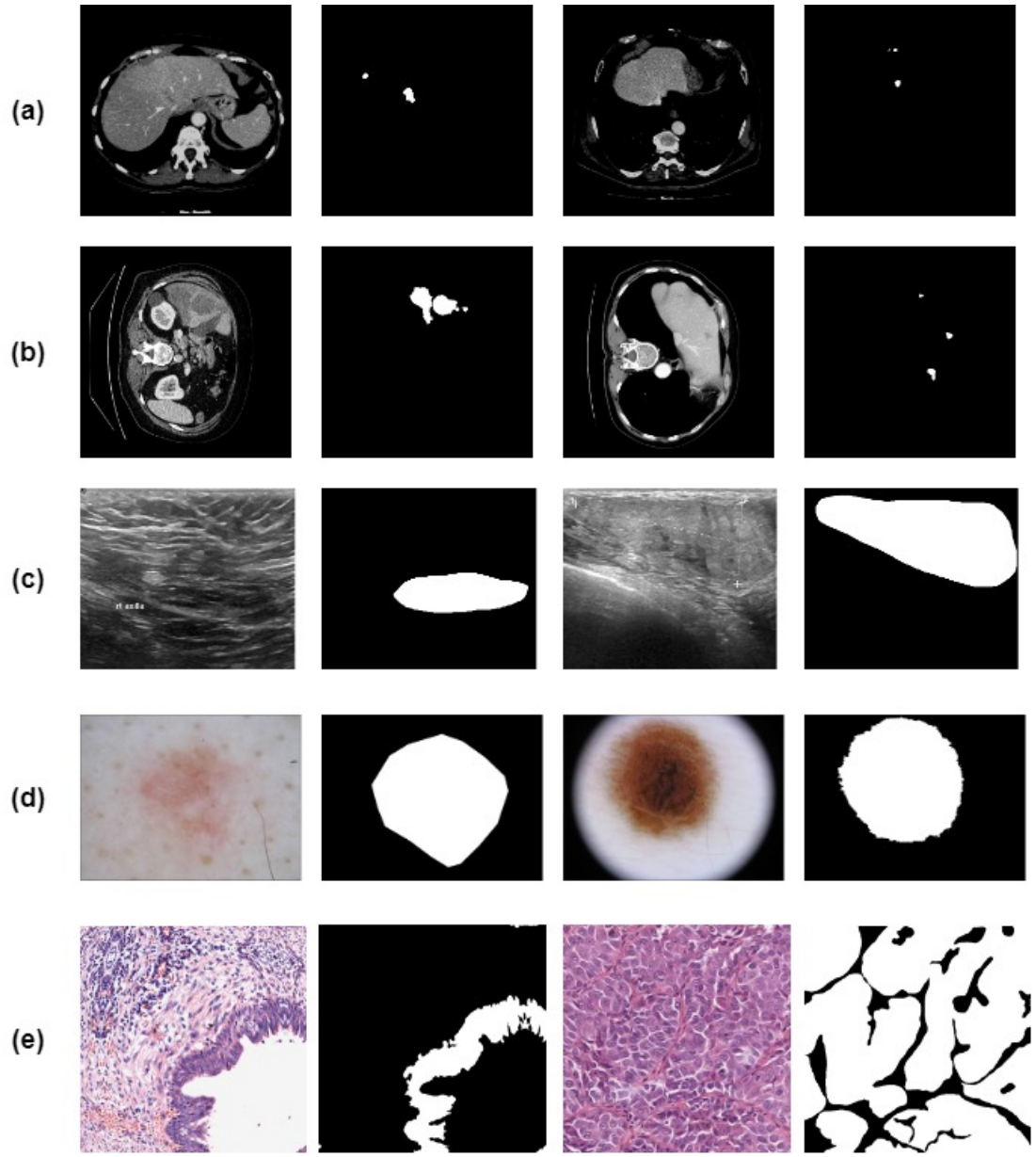


Figure 1.1: A few complicated segmentation instances. The datasets considered are: (a) 3DIRCADb, (b) LiTS, (c) BUSI, (d) ISIC 2018 and (e) OCDC.

- Introducing SE attention blocks within skip connections selectively applies attention to lower skip connections, enhancing feature utilization while avoiding redundancy. By selectively applying attention to lower skip connections, SEAU-Net aims to retain and effectively use high-level semantic features in segmentation, which can help mitigate challenges related to ambiguous boundaries, low resolution, and complex anatomical structures. The attention mechanism can help the model focus on relevant features and ignore noise or irrelevant information, thereby improving segmentation accuracy.
- W^2CU-Net employs two decoders with specific modifications, including a Weibull layer for fuzzy rank-based weighting and a cosine similarity-based feature pruning layer. By incorporating fuzzy rank-based weighting and feature pruning techniques, W^2CU-Net aims to generate more precise segmentation masks. These modifications can help mitigate challenges related to ambiguous boundaries, low resolution, imaging artifacts, and imbalanced classes by improving the model's ability to capture relevant features and ignore noise or irrelevant information. Additionally, the fusion technique using Weierstrass' elliptic curve equation may enhance segmentation accuracy by integrating complementary information from multiple decoders.

1.3 Organization of the Thesis

The remainder of this thesis is organized as follows:

- In chapter 2, a comprehensive literature survey is presented. This survey delves into existing research endeavors, exploring various methodologies, challenges, and breakthroughs in the field of medical image cancer segmentation.
- The chapter 3 sheds light on novel methodologies and materials introduced in this study. This chapter outlines the innovative approaches and techniques proposed to address the complexities inherent in medical image cancer segmentation.
- Subsequently, chapter 4 provides a comprehensive presentation of the results obtained from experiments conducted on four distinct datasets. This chapter not only showcases the performance metrics and quantitative evaluations but also engages in critical discussions surrounding the implications of the results.
- Finally, chapter 5 encapsulates essential conclusions drawn from the study and offers insights into future research directions. This concluding chapter

synthesizes key findings, highlights overarching themes, and underscores the significance of the study's contributions to the field of medical image cancer segmentation.

Through this research, the aim is to contribute to the advancement of medical image analysis techniques, ultimately facilitating more accurate and efficient diagnosis and treatment of cancer patients.

CHAPTER 2

Literature Survey

Over the past few years, significant advancements have been made in the field of medical image segmentation, particularly concerning the application of DL models to various cancer types such as liver cancer, breast cancer, oral cancer, and skin cancer. A literature survey focusing on recent research papers reveals a rich landscape of innovative methodologies and techniques employed to improve the accuracy and efficiency of cancer segmentation. Across the spectrum of cancer types, researchers have explored novel approaches utilizing CNNs, RNNs, and attention mechanisms to extract meaningful features and delineate tumor boundaries from medical images. Moreover, the integration of multi-modal imaging data, including MRI, CT scans, and histopathological images, has enabled more comprehensive and accurate segmentation results. However, challenges persist, including the need for large annotated datasets, robust model generalization, and interpretability of DL models in clinical settings. Despite these challenges, recent advancements underscore the potential of DL-based segmentation techniques to revolutionize cancer diagnosis, treatment planning, and prognosis assessment. A few recent publications related to liver cancer are highlighted in section 2.1, breast cancer in section 2.2, oral cancer in section 2.3 and skin cancer in section 2.4.

2.1 Liver Cancer

Entrenched in the multifaceted domain of liver cancer diagnostics, DL emerges as a pivotal asset, offering potential advancements in detection and prognostic precision. Within the expansive arena of medical imaging studies, the convergence of DL methodologies with liver cancer analysis presents a paradigm shift, poised to redefine early detection and treatment modalities, with select methods showcasing remarkable liver cancer segmentation capabilities. Liu et al. [6] propose a global context and hybrid attention network for the segmentation of the liver. Kavur

et al. [7] propose an ensemble model using four base models such as U-Net, Deepmedic[8], V-Net[9], and Dense V-Networks[10]. This ensemble method uses majority vote, average, product, and min/max as a basic non-trainable ensemble combiners, and results are evaluated with two publicly available datasets, namely CHAOS and 3DIRCADb. For automatic liver and tumour segmentation from CT images, the authors of [11] propose a modified U-Net architecture by altering the skip paths and utilising a feature fusion approach that represents the fine-grained contextual information in the high-level features and local feature reconstruction. Chen et al. [12] suggest a FRAU-Net network for two-step segmentation of the liver and tumours.

At times, it is difficult to segment by using one individual model; hence Di et al. [13] propose a hybrid model using a convolution network that integrates transformer and direction information so that liver tumours from CT images can be segmented automatically. A method to feature map stacking that concentrates on the correlation of neighbouring channels between two feature maps is proposed in [14]. For the segmentation of liver tumours, Li et al. [15] suggest an H-DenseUNet that combines 2D U-Net and 3D U-Net. While the 3D U-Net was created for inter-slice contextual feature representation, the 2D U-Net was utilised to extract tumour characteristics from inside slices of tissue.

Jiang et al. [16] use residual SE, where the residual module helps in resolving the gradient vanishing problem inside the SE block. Oktay et al. [17] propose a novel attention gate model for medical imaging, which can be integrated into the U-Net architecture that increases the model sensitivity and prediction accuracy. This enabled to eliminate the necessity of using explicit external tissue/organ localisation modules of cascaded CNNs. Hettihewa et al. [18] propose a multi-attention network, which aims to fuse different attention mechanisms. Zhu et al. [19] use a channel attention module within the decoder layers, with convolution operations replaced by residual structures.

A U-Net based multi-scale attention mechanism is used in [20], which uses RASM to better fuse low and high level feature maps. To capture rich contextual dependencies, Fan et al. [21] propose a MA-Net. Liu et al. [6] propose a hybrid attention mechanism for liver segmentation, which uses global attention module to capture the global characteristics and local attention module to capture spatial information. To enhance its performance in image segmentation tasks, Li et al. [22] suggest ANU-Net that leverages attention mechanisms, nested structures, deep supervision, and a redesigned dense skip connection. A combination of hard and soft attention and long and short skip connections is introduced here. Dual effect attention module and deep residual block are used in a biplane joint approach with a DRAUNet [23] for segmenting the liver.

A modified U-Net [24] addresses class imbalance by reducing filters and discarding slices without tumor information. A multi-modal co-learning method has been proposed for liver lesion segmentation in [25]. The decoder branch with a hierarchical feature co-learning module is used for combining features. A 2.5D perpendicular U-Net [26] model is proposed for liver tumor segmentation consists of three Res-UNets trained separately in the Coronal, Sagittal, and Axial planes, utilizing multi-scale convolutions in the Coronal and Sagittal planes. These individual Res-UNets are trained and their segmentation outputs are integrated for comprehensive segmentation results. In [27], a cascaded DL method for liver segmentation is presented, which employs an initial U-Net for segmentation, focusing on reducing false positive pixels and eliminating liver-connected organs. Subsequently, two cascading U-Nets are utilized, followed by an additional U-Net for refining the liver region reconstruction.

2.2 Breast Cancer

Nestled within the complex realm of breast cancer diagnostics, DL shines as a potent tool, holding the promise of improved detection and prognostic accuracy. Amidst the expansive landscape of medical imaging research, the integration of DL techniques with breast cancer analysis emerges as a transformative force, poised to redefine early detection and treatment paradigms, with a few notable methods offering insightful breast cancer segmentation capabilities. Podda te al. [28] propose a segmentation ensembling method built on five backbones, namely the ResNet50 [29], [30], VGG19 [31], DenseNet121 [32], a U-Net [3] comprising of Inception blocks combined with traditional convolutional layers and a U-Net built from scratch. This is followed by a soft voting mechanism and thresholding step. Hossain et al. [33] propose an automatic method for segmenting micro-calcification regions from mammogram images. The positive patches containing micro-calcifcation pixels are used to train a modifed U-net segmentation network.

Pérez-Benito et al. [34] discuss the development of a fully automated framework based on DL to estimate breast density from mammograms. The framework includes breast detection, pectoral muscle exclusion, and fibroglandular tissue segmentation. Wang et al. [35] have demonstrates the effectiveness of novel DL models with the addition of ASN for morphological analysis and determine the performance for diagnosis breast cancer in BUS images. To enhance the feature extraction capabilities of the encoder module of U-Net, Pramanik et al. [36] introduce a novel approach involving the utilization of two distinct encoding paths. In the first path, the original image is employed, while in the second path, an image created using the Roberts edge filter is used, in which edges are highlighted.

The DBU-Net's dual branch encoding strategy helps to extract the semantic rich information through a mutually informative learning process.

Luo et al. [37] propsose the SBANet, where a segmentation network is trained to produce images that enhance the segmentation of tumors. Then two parallel networks extract features for the original images and segmentation enhanced images and one channel attention based feature aggregation network is to automatically integrate the features extracted from two feature networks to improve the performance of recognizing malignant tumors in the breast ultrasound images. BCDUNet [38] applies a block of dense convolutions to the bottom encoding layer. Instead of a simple concatenation in the skip connection of U-Net, BConvLSTM to combine the feature maps extracted from the corresponding encoding path and the previous decoding up-convolutional layer in a non-linear way.

2.3 Oral Cancer

Amidst the intricate web of oral cancer's diagnostic challenges, DL emerges as a beacon of hope, promising enhanced detection and prognosis capabilities. Within the vast expanse of medical imaging literature, the fusion of DL methodologies with oral cancer analysis stands poised to revolutionize early detection and treatment strategies, offering a glimpse into a future where a few oral cancer segmentation methods provide invaluable insights. Albishri et al. [39] propose a U-Net with novel CSAF module for refining channel features and capturing long-range dependencies to exploit contextual information effectively. Another network, called Multiencoder U-Net [40], is a multi-encoder single decoder network that takes as input an image and splits it in tiles. For each tile, there is an encoder responsible for encoding it in the latent space, then a convolutional layer is responsible for merging the tiles into a single layer. Each layer of the decoder takes as input the previous up-sampled layer and concatenate it with the layer made by merging the corresponding layers of the multiple encoders.

Sunny et al. [41] use the encoder-decoder models based on the well-established U-Net architecture, modified U-Net and ResNet-34 [42] for multi-class segmentation. Zhang et al. [43] propose the FD-Net to segment the cancerous and non-cancerous nuclei in the learning stage. In the decision-making stage, the segmentation results are post-processed, and the lesions are effectively distinguished based on the prior. Shephard et al. [44] introduce a DL framework to simultaneously segment (and classify) nuclei and (intra-)epithelial layers in H & E stained slides from OED cases. The proposed architecture consists of an encoder branch and four decoder branches for simultaneous instance segmentation of nuclei and semantic segmentation of the epithelial layers.

Dos et al. [45] propose a method that uses color features in the HSV color model to identify tissue regions in a pre-processing step to remove background and non-relevant areas. The identified tissue regions are then transformed into the CIE $l \times \alpha \times \beta$ color model and split into image-patches. A FCN-based model to segment cancer and pre-cancer lesion regions in the oral cavity is proposed [46]. In addition to the RGB channels of the input image, they append features of Gabor filter and wavelet filter that show strong response at cancer and pre-cancer regions. They also propose a refine stage before the decision layer of FCN to preventing weight dominating problem when reducing high dimension features to small number of classes.

Matias et al. [47] compare three different DL approaches: segmentation, object detection, and image classification.The results show that the binary object detection with Faster R-CNN is the best approach for nuclei detection and localization. Folmsbee et al. [48] use AL, a process for iterative interactive training, to create a modified U-Net classifier on the RoI scale, which is then compared with RL, where images for addition to the dataset for retraining are randomly selected.

2.4 Skin Cancer

Within the intricate realm of skin cancer diagnosis, DL emerges as a transformative ally, promising heightened accuracy in detection and prognosis. Within the vast expanse of medical imaging research, the fusion of DL techniques with skin cancer analysis holds the potential to revolutionize early intervention strategies, with several notable methods showcasing robust skin cancer segmentation capabilities. Wu et al. use a transformer-based encoder approach for improving segmentation of skin cancer [49]. [50] present the AS-Net framework, which demonstrates improved discriminative capabilities for skin lesion segmentation applications, by combining spatial and channel attention approaches. Superior skin lesion segmentation performance is also achieved by MFSNet [51] and Ms RED [52] by integrating the attention mechanism with the multi-scale approach.

Jiang et al. [53] propose the MDSC-Net which uses a novel multi-scale depthwise separable residual convolution module in skip connection, conveying more detailed features to the decoder. To compensate for the loss of spatial location information in down-sampling, a novel spatial adaption module is proposed. Furthermore, a multi-scale decoding fusion module is also introduced in the decoder to capture contextual information. To capture the local and global contextual information, Yu et al. [54] employ inverted residual blocks and an EPSA block as the main encoders at different stages, while ASPP is utilized after the last encoder and a soft-pool method is introduced for downsampling. A novel method

named MLF module is also proposed to effectively fuse the feature distributions and capture significant boundary information of skin lesions in different encoders to improve the performance of the network. Furthermore, a reshaped decoders fusion module is used to obtain multi-scale information by fusing feature maps of different decoders to improve the final results of skin lesion segmentation.

Ma et al. [55] propose a novel LCAUnet to improve the ability of complementary representation with fusion of edge and body features. First, two separate branches are set for edge and body segmentation with CNNs and transformer-based architecture respectively. Then, LCAF module is utilized to fuse feature maps of edge and body of the same level by local cross-attention operation in encoder stage. Furthermore, PGMF module is embedded for feature integration with prior guided multi-scale adaption. A generic novel deep medical segmentation framework, denoted as Ψnet is introduced [56]. This model introduces a novel parallel encoder-decoder structure that draws up the power of triple U-Nets. In addition, a multi-stage squeezed-based encoder is employed to raise the network sensitivity to relevant features and suppress the unnecessary ones. Moreover, ASPP is employed in the bottleneck of the network which helps in gathering more effective features during the training process, hence better performance can be achieved in segmentation tasks.

CHAPTER 3

Methods and Materials

The framework utilized to evaluate the three novel DL methods across six diverse datasets for cancer segmentation is outlined. The architectural designs, training procedures, and optimization strategies tailored for each method to address specific challenges in medical image analysis are detailed. Leveraging datasets such as 3DIRCADb, CHAOS, and LiTS for liver and liver tumor segmentation from CT scans, as well as the BUSI dataset for ultrasound-based breast cancer segmentation, adaptations are explained to accommodate complexities in organ segmentation and lesion detection. Additionally, methodologies for histopathology-based oral cancer segmentation using the OCDC dataset and skin cancer segmentation from dermoscopic images using the ISIC 2018 dataset are discussed. Through concise elaboration of the methods, a comprehensive overview of the experimental setup and procedures for evaluating the performance and robustness of the novel approaches across diverse cancer types and imaging modalities is aimed to be provided. section 3.1 elaborates the different datasets that are considered and section 3.2 elaborates the three novel methods that are introduced.

3.1 Datasets

This section provides a comprehensive description of the datasets considered in the study. Each dataset represents a distinct imaging modality and cancer type, spanning CT scans for liver and liver tumor segmentation (3DIRCADb, CHAOS, and LiTS), ultrasound imaging for breast cancer segmentation (BUSI), histopathological images for oral cancer segmentation (OCDC), and dermoscopic images for skin cancer segmentation (ISIC 2018).

3.1.1 3DIRCADb

The 3DIRCADb dataset [57], [58] comprises 3D CT scans of 10 women and 10 men with hepatic tumours as well as non-tumorous CT scans. Images are present in DICOM format. A total of 2823 CT scans are present. Liver masks of all CT scans are present. However, all CT scans might not be tumorous. Thereby, for the non-tumorous CT images, there is no tumor mask present. However, for the tumorous cases, there can be more than one tumor mask present. Since the liver and tumor segmentation task was undertaken, for training the model, consideration was given to CT scans of only those patients diagnosed with liver cancer, while omitting scans from patients without this condition. Thereby, 2063 CT images, their corresponding liver masks and tumor masks are considered for the segmentation task here.

All images are present in DICOM format. This causes a problem of visualization. So, all the images are converted into .jpg configuration. The tumor masks are Couninaud-segmented, which implies that a CT scan of a patient having a tumor may have more than one tumor mask. Therefore, all these masks are superimposed, producing one tumor mask per CT scan.

CT images are highly prone to noise and not homogeneous by nature because of fallacies in the machines used, unwanted light penetrating within the scans, etc. To reduce such inconsistencies, a windowing technique, namely HU clipping, is performed on these DICOM images. This technique, referenced in [59]–[61], helps to standardize the Hounsfield Unit values across the dataset.

The radio density of different biological tissues is expressed on a measuring scale called the HU, which is primarily employed in CT scans. When X-rays pass through different materials, their intensity changes. Different tissues have unique HU ranges that aid radiologists in readily distinguishing among different structures and identifying anomalies in a CT scan. Water is designated as the reference point on the scale with a value of 0 HU. Whereas dense materials like bone tissues [62], which attenuate X-rays considerably, might have HU values even exceeding +3000, air [63], which attenuates X-rays extremely weakly, is assigned a value of -1000 HU.

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 [64]–[66]. The intensity of X-rays vary as it traverses through matter. Different tissues have characteristic HU ranges, which help radiologists distinguish clearly 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 [63], which attenuates X-rays very weakly, is assigned a value of -1000 HU, whereas dense

materials like bone tissues [62], which strongly attenuate X-rays, can have HU values even above +3000.

The window in the HU range of [-200,+200] is opened for the 3DIRCADb dataset so that irrelevant structures are removed and the area of interest i.e., the liver stands out from the surrounding viscera in the abdomen. Figure 3.1 shows a visualization of the original CT images vs the HU-clipped CT images and their corresponding liver and tumor masks for the 3DIRCADb dataset. On these final pre-processed HU-clipped images, a train-validation-test split is performed with a ratio of 8:1:1. Thereby, the training set has 1239 images, the validation set has 412 images and the test set has 412 images.

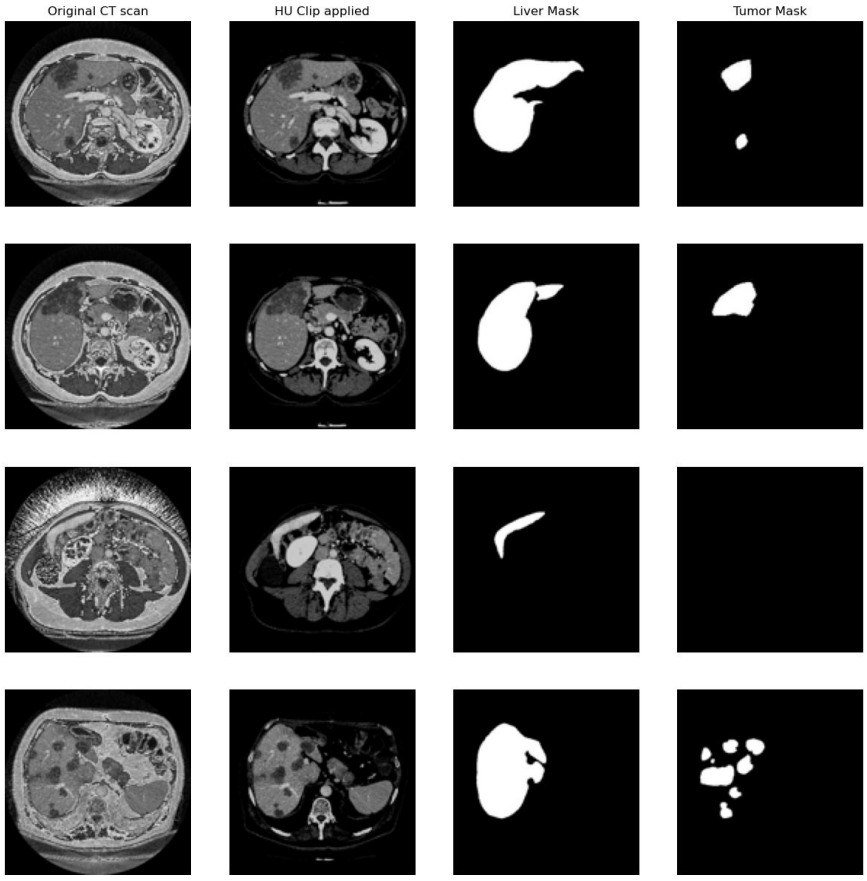


Figure 3.1: Visualization of the 3DIRCADb dataset - Column1: Original CT scans, Column2: HU-clipped CT images, Column3: Liver Masks, Column4: Tumor Masks

3.1.2 CHAOS

The CHAOS dataset [67], [68] is a collection of medical imaging data specifically focused on the segmentation of abdominal organs, such as the liver, from CT and MRI scans. It is used for training and testing algorithms and models in the field of medical image analysis. The CHAOS dataset consists of 40 cases in total, with 20 cases allocated for training and 20 cases for testing. For the CT scan dataset, only the liver is annotated, while the MRI dataset contains both T1-Dual and T2 SPIR sequences. The training set includes DICOM images along with their corresponding ground truth masks, which indicate the labeled RoIs such as the liver. The testing set, on the other hand, only contains the DICOM images without the ground truth masks. For this research, only the training set of the original dataset is considered for training purposes.

The window width is opened within the range of [-200, +200], covering a total of 400 units. Tissues with HU values between -200 and +200 can be efficiently emphasized on the CT picture by employing this clipping range, while all other tissues are muted since they lie beyond the chosen window. As a result, all extraneous structures are eliminated, allowing the liver RoI—to be distinguished from the other viscera in the abdomen. The dataset is split into training, validation and testing sets in the ratio of 7:1:2. Thereby, the training set has 2011 images, the validation set has 287 images and the test set has 576 images.

3.1.3 LiTS

The LiTS dataset [69], [70] comprises 131 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 NII technology format. The training images are labelled by medical professionals, whereas the test set is unlabelled. However, for research purposes, only the 131 labeled CT volumes are considered. Each CT volume consists of a series of 2D slices. The number of images per volume ranges from 42 to 1026.

In preprocessing the .NII medical imaging files, HU clipping within the range of [-100, 400] is crucial using PyDICOM in Python. The process involves loading the .NII files, converting pixel values to HU considering RescaleSlope and RescaleIntercept attributes, and clipping HU values to -100 and 400. The preprocessed samples is stored in .jpg format, ensuring focus on clinically relevant tissue densities and excluding artifacts. Furthermore, every image in the dataset is traversed, and all black masks for liver tumor segmentation are eliminated.

The datasets are then split into train-validation-test sets in an 8:1:1 ratio for all subsequent analysis, enhancing accuracy and interpretability of medical imaging

data. Thereby, the traning set has 5,752 images, the validation set has 719 images and the test set has 719 images for liver tumor segmentation.

3.1.4 BUSI

The BUSI dataset [71] comprises 780 ultrasound images belonging to three classes, namely normal, malignant, and benign. The instruments used in the scanning process are LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system, producing image resolution of 1280 × 1024. The transducers are 1e5 MHz on ML6-15-D Matrix linear probe. A glimpse of a few images is shown in Figure 3.2. For this research, all images and masks belonging to the normal class are removed. It is to be noted that each image can consist of one mask or two. Thereby, two masks are superimposed into one mask as shown in Figure 3.3. The images are then split into training and testing in the ratio of 7:3. Data augmentation is then applied on the training images in the form of horizontal flip, vertical flip, random rotation between -45 degrees and 45 degrees, zoom, shear, brightness adjustment and contrast adjustment. Thereby, the training set has 3416 training images and test set has 209 images.

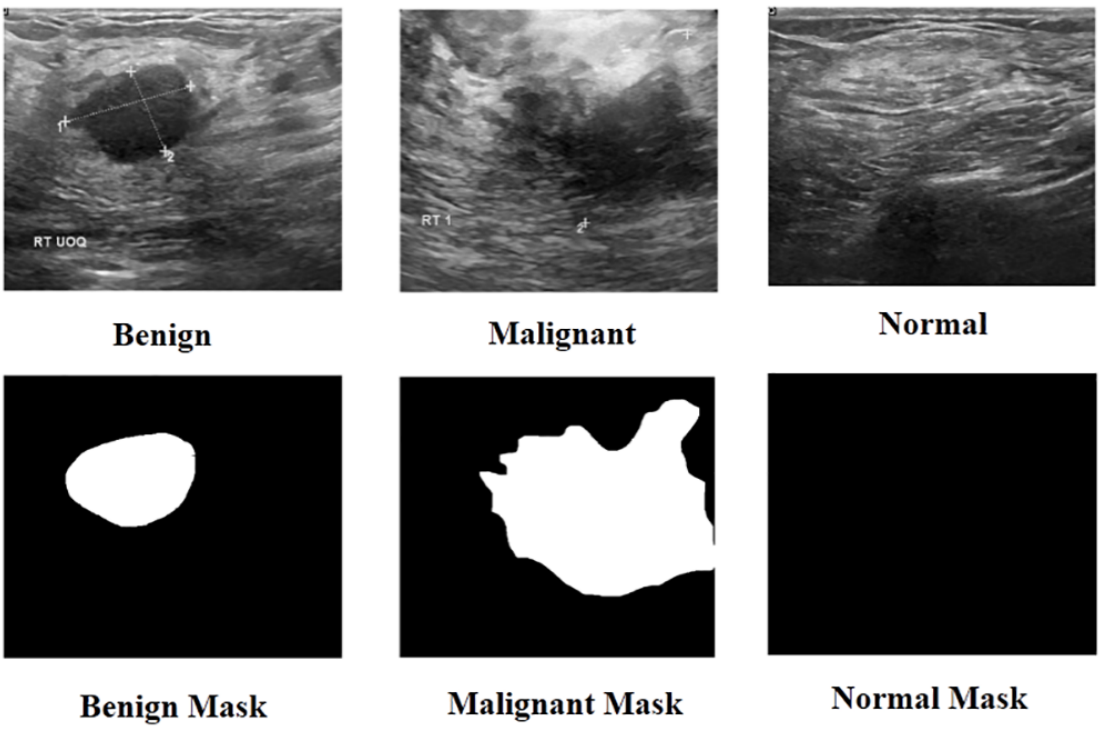


Figure 3.2: Examples of breast ultrasound images and associated masks in the BUSI dataset.

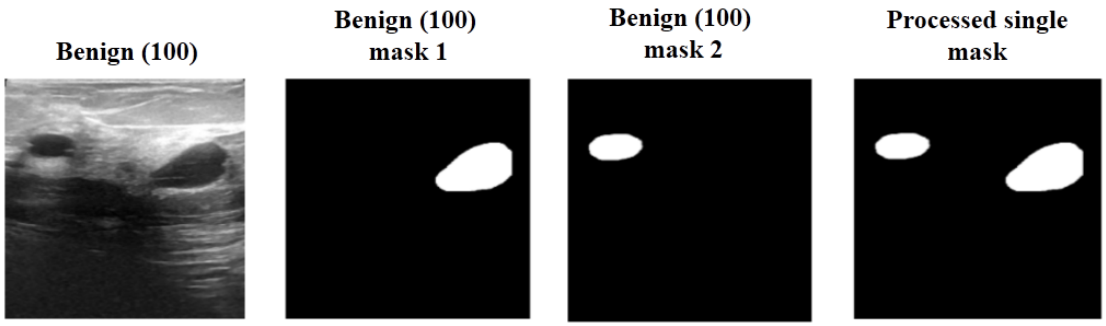


Figure 3.3: A sample image with multiple masks, and the single mask obtained after merging them.

3.1.5 OCDC

The OCDC cancer WSI dataset [45] comprises 15 WSIs were digitized using the Slide Scanner Aperio AT2 (Leica Biosystems Imaging, Inc., Nussloch, Germany) coupled to a computer (Dell Precision T3600) at 20× magnification and pixel-level resolution of 0.5025 μm × 0.5025 μm. The digitized images have different sizes – the larger one has almost three billions of pixels (63,743 × 45,472 pixels) – and were stored in SVS format using the RGB color model. A total of 1050 image-patches of size 640 × 640 pixels were randomly extracted from these 15 WSI and the content of each image-patch was hand-annotated by a well-trained pathologist. The image-patch dataset was split into two subsets: the training set with 840 image-patches and the test set with 210 image-patches. From these, the images with black masks are removed from the dataset. Then augmentation is carried out on the training set using horizontal flip, vertical flip, elastic transformation, grid distortion and optical distortion. Thereby, the training set has 2340 images and test set has 61 images. A glimpse of the OCDC dataset is shown in Figure 3.4.

3.1.6 ISIC 2018

The ISIC dataset is part of the world's largest public repository of dermoscopic images of skin. The challenge was hosted in 2018 at the MICCAI conference in Granada, Spain. For part 1, 2,594 dermoscopic images with ground truth segmentation masks are provided for training. For validation and test sets, 100 and 1,000 images are provided, respectively, without ground truth masks. For this research, only the training images with the masks are considered. This dataset is split into training and test sets in the ratio of 4:1. Augmentation is then applied on the training dataset in the form of vertical flip, horizontal flip, transpose and random rotation of the image by 90 degrees one or more times. Finally, the training set has 10,375 images and test set has 519 images. A glimpse of the OCDC dataset is shown in Figure 3.5.

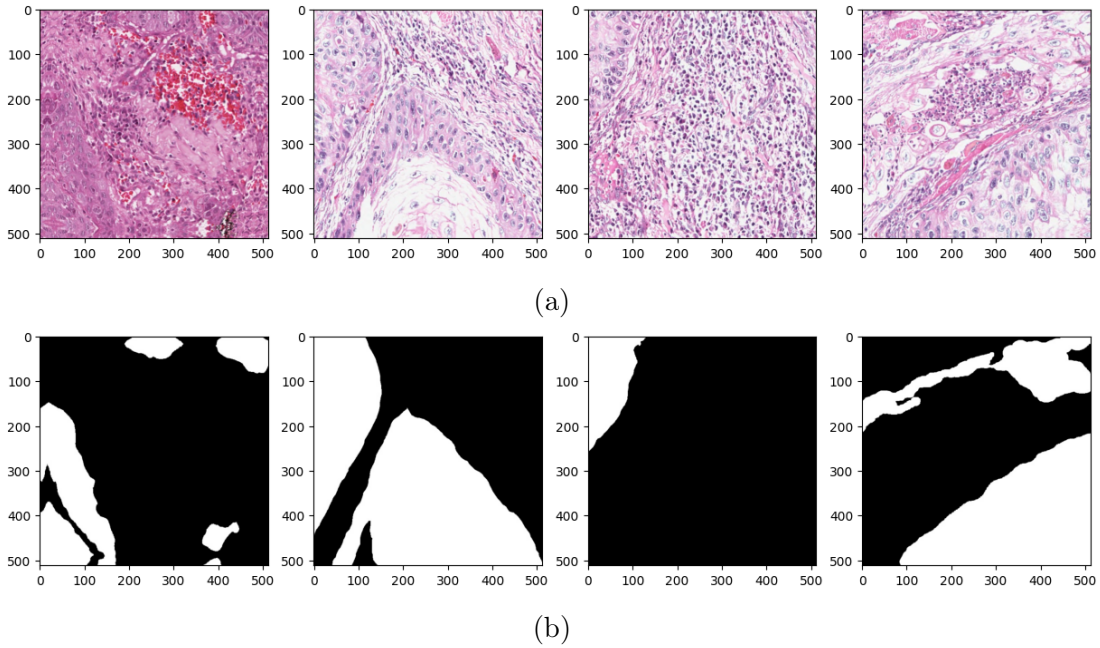


Figure 3.4: A few sample histopathology images from the OCDC dataset. (a) WSIs and (b) their corresponding masks.

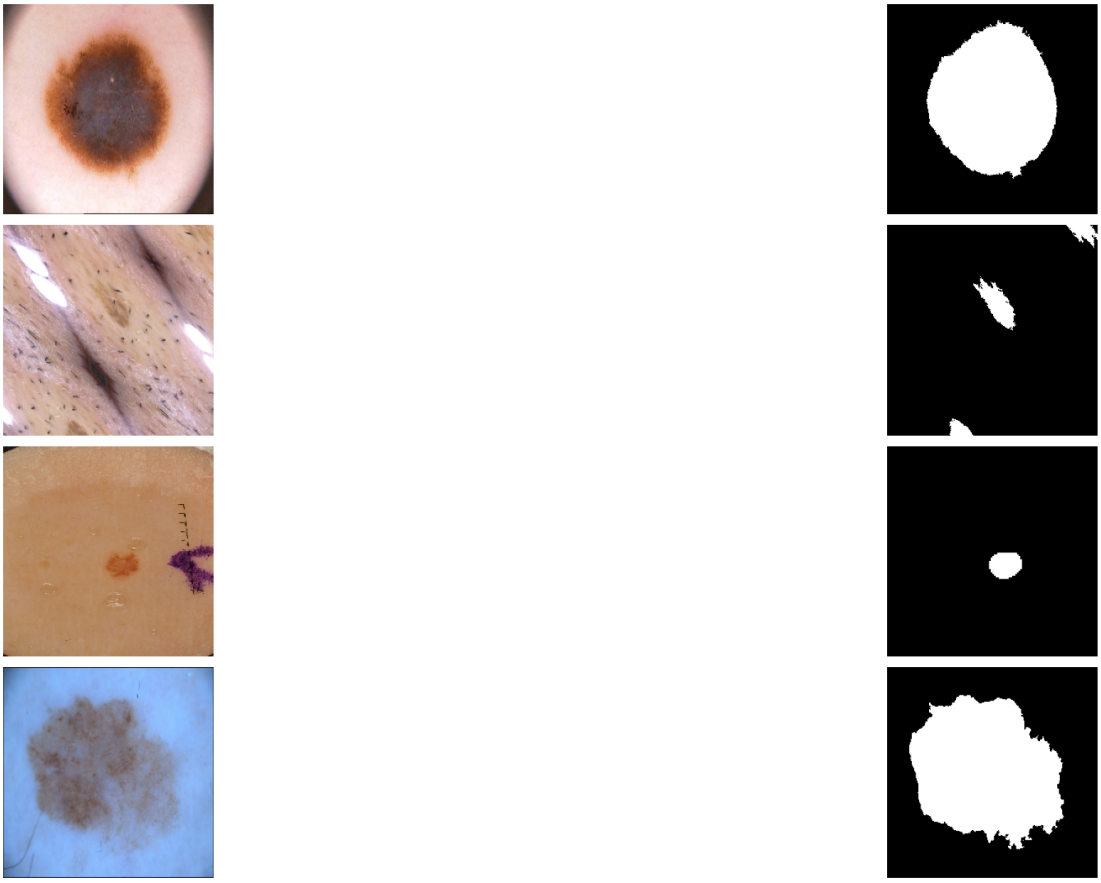


Figure 3.5: A few sample dermoscopy images from the ISIC 2018 dataset. Column1: the images and Column2: their corresponding masks.

3.2 Methods

This section elaborates three novel DL approaches tailored for enhancing cancer segmentation accuracy. Each method is designed to address specific challenges in medical image analysis, leveraging DL architectures for improved performance across diverse datasets. A detailed exposition of the architectural designs, training procedures, and optimization strategies employed for each method is provided, aiming to assess their efficacy and robustness in accurately delineating cancerous regions. Through this presentation, a contribution is made to advancing state-of-the-art techniques in cancer segmentation, facilitating improved diagnostic and treatment strategies in oncology.

3.2.1 BLUE-Net

In this work, an ensemble learning model, named BLUE-Net, is proposed. It utilizes three base models: Vanilla U-Net [3], Link-Net [4], and FPN [5], with MobileNetV2 serving as the backbone. These models are imported from the segmentation models' library [72]. The results from these models are then combined using the Blumberg function-based ensemble method, resulting in a dice score higher than those obtained by the individually trained models. The complete flowchart of the proposed methodology is shown in Figure 3.6.

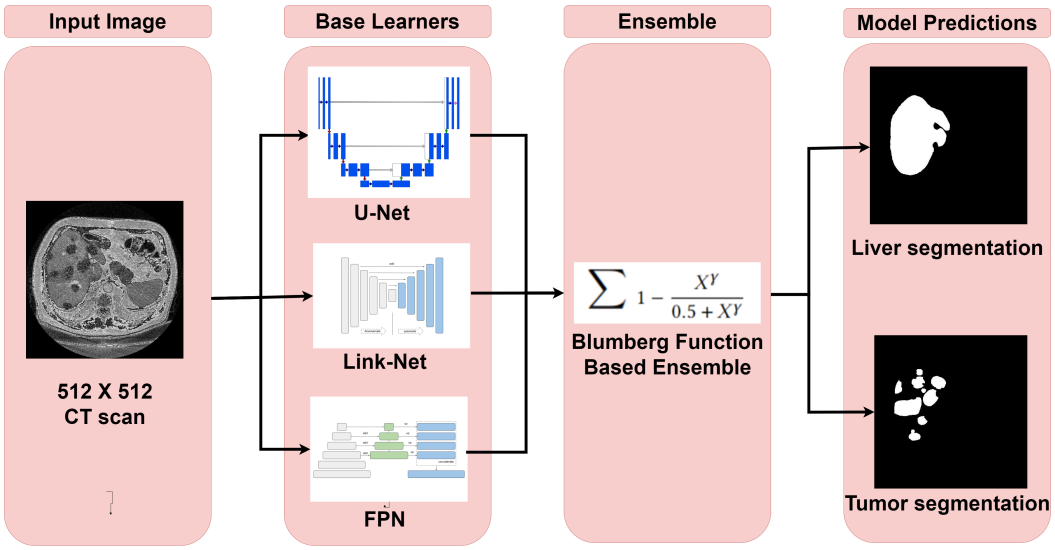


Figure 3.6: Pipeline of the proposed Blumberg function based ensemble model used for liver and tumor segmentation

3.2.1.1 U-Net

The U-Net architecture is a DL CNN architecture, primarily designed for semantic image segmentation tasks. It was proposed by Ronneberger et al. [3]. The U-Net architecture is characterized by its U-shaped structure, which consists of a contracting path and an expanding path. This unique architecture allows the network to capture both global context information and local details, making it particularly effective for tasks like biomedical image segmentation, where precise localization and segmentation of structures are essential. It comprises of a contracting path, a bottleneck layer, an expanding path and skip connections.

3.2.1.2 Link-Net

Introduced by Eduardo et al. [4], Link-Net is another CNN architecture that is specifically designed for semantic image segmentation tasks. Link-Net builds upon the encoder-decoder architecture, similar to U-Net, but introduces some modifications to improve segmentation performance and efficiency. The main focus of Link-Net is to strike a balance between accuracy and computational efficiency, making it suitable for real-time or resource-constrained applications. It comprises the encoder-decoder structure, skip connections, bottleneck with shortcut connections and residual blocks.

3.2.1.3 Feature Pyramid Network

The FPN architecture was introduced by Tsung-Yi et al. [5]. It is a DL architecture designed to address the challenges of object detection and semantic segmentation tasks, particularly when dealing with objects at various scales. The main goal of FPN is to enhance the ability of a CNN to detect and segment objects of different sizes in an image. Traditional CNN architectures, such as those used for image classification, tend to lose fine-grained details as they go deeper into the network. FPN addresses this issue by creating a multi-scale representation of features that captures both low-level and high-level information. It comprises a bottom-up path, a top-down path, lateral connections and the output, which is a feature pyramid map.

3.2.1.4 Blumberg Function based Ensemble Method

The fundamental objective of the suggested strategy is to offer greater versatility and independence for managing datasets with different levels of complexity. The unpredictability of each base learner's predictions is taken into account, and each one is assigned a varied level of weightage based on its performance on a

certain test case. This is achieved using a fuzzy ranking-based technique. In the suggested methodology, livers and tumors are segmented from CT scans based on their respective masks using the three CNN-based segmentation models (U-Net, Link-Net, and FPN with the MobileNetV2 as a backbone), and fuzzy ranks are generated for each pixel in the final masks produced by the base models using the re-parameterized Blumberg function. The Blumberg function, which has been applied previously to the field of machine learning [73], has been used to describe several demographic, economic, ecological and other processes [74]. Here it has been used for combining the outputs of different base models with various strengths and weaknesses for liver and tumor segmentation from CT scan images.

3.2.1.5 Significance of Blumberg Function

The performance of each of the base models is shown in the Blumberg function by a dose response curve between 0 and 1 on the x-axis that shows performance as a function of the input strength. The fit between the forecasts and the actual outcomes of each pixel is determined by the weighted sum of the predictions of every pixel in an image, which is measured by the Blumberg function that then returns a value. To do this, each model in the ensemble generates a fuzzy rank of a class for the pixels, and the decision of different base models are then combined to make predictions on a validation set for each rank inside the top K ranks. Each base model applies the Blumberg function to the confidence scores of a pixel class to produce fuzzy ranks for that class. By using this technique, an adaptive ensemble model that can change the ranking of individual models in response to the particulars of each input instance can be created. The Blumberg function's drastically declining nature in the domain ranging from 0 to 1 helps to generate an ensemble of the decision scores from the learning models.

3.2.1.6 Implementation of the Ensemble Model

There should be M different decision scores (also known as confidence scores of the base models) such as $CoF^{(1)}, CoF^{(2)}, ..., CoF^{(M)}$ for each input image pixels $Pi_c^{(i)}$, where c is the number of classes in the image from the ith model. Since three CNN-based models produce the confidence scores on the image pixels, M in this case equals 3. In Equation 3.1, the decision scores for images are normalized, where C is the total number of classes for the dataset under consideration.

$$\sum_{c=1}^{C} CoF_c^{(i)} = 1; \forall i,\ i = 1, 2, 3, ..., M \tag{3.1}$$

The confidence scores of all the pixels in each of the test images, belonging to

various classes, are used to produce fuzzy ranks. The fuzzy rank is produced by the Blumberg function for a class c using the i^{th} model's decision scores as shown in Equation 3.2.

$$R_c^{(i)} = 1 - \frac{(CoF_c^{(i)})^\gamma}{0.5 + (CoF_c^{(i)})^\gamma}; \tag{3.2}$$

$$\forall i, c,\ i = 1, 2, 3, ..., M; c = 1, 2, ..., C;\ here\ \gamma = 1$$

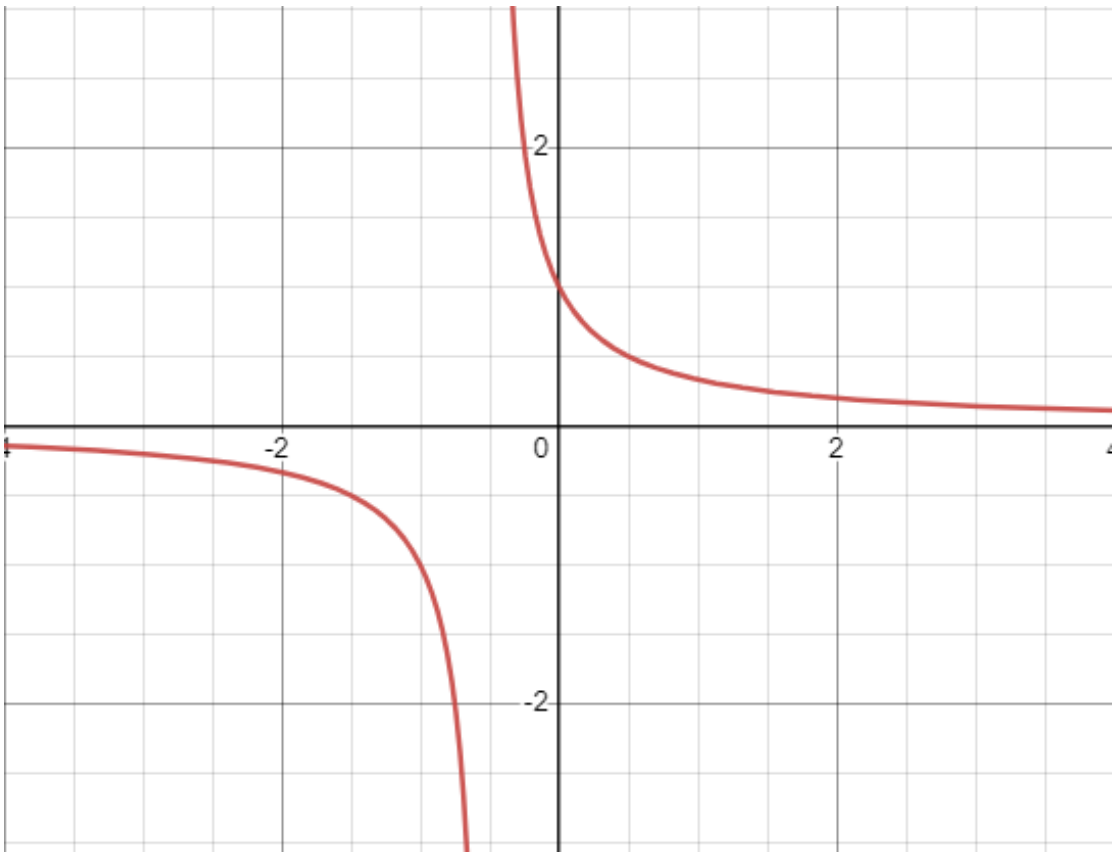


Figure 3.7: Graphical representation of the modified Blumberg function used in the present work.

Figure 3.7 shows the pictorial depiction of the re-parameterized Blumberg function graph, as mentioned in Equation 3.2.

The value of $R_c^{(i)}$ is between $0.3\overline{33}$ and 1, with the smallest value $0.3\overline{33}$ being equivalent to rank 1 (best rank), i.e., greater confidence results in a lower (better) value of rank. The fuzzy rank sum (FRS_c) and complement of confidence factor sum ($CCoFS_c$) are computed as follows if the top k ranks are represented by $K^{(i)}$, that is, ranks 1, 2,..., k, corresponding to class c are as follows [75]:

$$FRS_c = \sum_{i=1}^{M} \begin{cases} R_c^{(i)}, & \text{if } R_c^{(i)} \in K^{(i)} \\ P_c^R, & \text{otherwise} \end{cases} \tag{3.3}$$

$$CCoFS_c = \frac{1}{M}\sum_{i=1}^{M}\begin{cases} CoF_c^{(i)}, & \text{if } R_c^{(i)} \in K^{(i)} \\ P_c^{CoF}, & \text{otherwise} \end{cases} \tag{3.4}$$

If class c does not fall among the top k class ranks, penalties P_c^R and P_c^{CoF} are applied on it. Using the aforementioned Blumberg function, the value of P_c^R is set to 1, which is determined by setting $CoF_c^{(i)} = 0$; and the value of P_c^{CoF} is set to 0. The penalty values prevent class c from emerging as an improbable winner. The combination of FRS_c and $CCoFS_c$, which is used to produce the ensemble model's final predictions, yields the confidence score in question. Equation 3.5 determines the final decision score (FDS).

$$FDS_c = FRS_c * CCoFS_c \tag{3.5}$$

Finding the class with the lowest FDS value yields the final projected class for the particular pixel in the data instance $\mathbf{I}$, which is provided as shown in Equation 3.6.

$$class(\mathbf{I}) = \underset{c=1,2,...,C}{\arg\min}\, FDS_c \tag{3.6}$$

The overall process of the ensemble approach has been described in algorithm 1.

3.2.1.7 Loss functions

Dice Loss: The Dice loss is a metric used in the field of image segmentation, particularly in medical imaging and computer vision tasks. It's used to evaluate the similarity between the predicted segmentation masks and the ground truth masks. The Dice loss is derived from the Dice coefficient, which is a measure of the overlap between two sets.

$$DiceLoss = 1 - DSC \tag{3.7}$$

Focal Loss: The binary Focal loss is a modification of the standard binary cross-entropy loss that aims to address class imbalance, and focus on harder-to-classify examples. It was introduced by Lin et al. [76]. It uses a modulating factor that downweights the contribution of well-classified pixels and focuses more on the mis-classified or challenging ones. If p denotes the predicted probability of the positive class, γ denotes the ground truth label, α denotes the balance factor that adjusts the weight of the positive class, and γ denotes the focusing parameter

Algorithm 1 Blumberg function based Ensemble Method

$C \leftarrow Total\ no.\ of\ classes\ in\ an\ image$
$M \leftarrow No.\ of\ base\ learners$
$p1 \leftarrow Test\ results\ of\ U-Net$
$p2 \leftarrow Test\ results\ of\ Link-Net$
$p3 \leftarrow Test\ results\ of\ FPN$

function FUZZY_RANK(CF)
    $R_L \leftarrow 1 - \frac{CF}{0.5+CF}$ ▷ Applying Blumberg function to each class of every pixel
    $K_L \leftarrow Initialise\ array\ of\ R_L\ shape\ with\ penalty\ values$
    if $R_L_c^i \in top\ 1\ rank$ **then** ▷ R_L for a pixel results by ith base model for cth class
        $K_L_c^i \leftarrow R_L_c^i$
    end if
    return K_L
end function

function CFS_FUNC(CF, K_L)
    if $R_L_c^i \notin top\ 1\ rank$ **then**
        $CF_c^i \leftarrow 0$ ▷ Applying penalty values
    end if
    $CFR \leftarrow \sum_{c=1}^{C} \sum_{i=1}^{M} CF_c^i$
    $CFS \leftarrow 1 - \frac{CFR}{M}$
    return CFS
end function

function BLUMBERG($p1, p2, p3$)
    $CF \leftarrow array(array(p1, p2, p3), array(1-p1, 1-p2, 1-p3))$ ▷ Elementwise subtract from 1
    $pred \leftarrow Array\ initialized\ with\ shape\ of\ p1$
    $R_L \leftarrow$ **fuzzy_rank(CF)**
    $RS \leftarrow \sum_{c=1}^{C} \sum_{i=1}^{M} R_L_c^i$
    $CFS \leftarrow$ **CFS_func(CF, R_L)**
    $FS \leftarrow RS \times CFS$
    if $FS_c\ is\ minimum$ **then**
        $pred \leftarrow class\ c$ ▷ Predicting the class for each pixel
    end if
    return $pred$
end function

results ← **Blumberg(p1, p2, p3)**

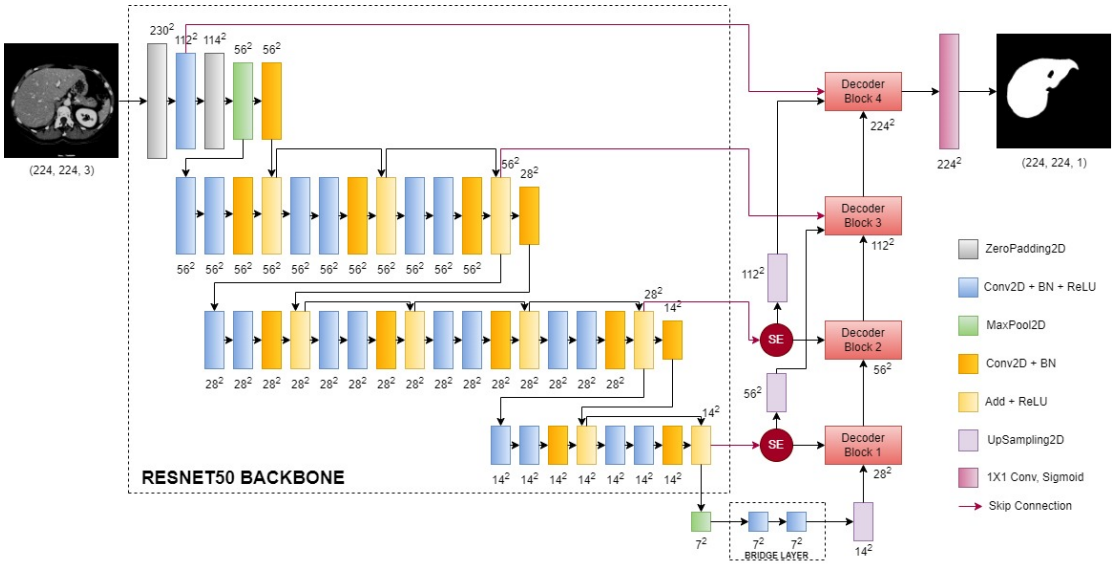


Figure 3.8: Architecture of the proposed SEAU-Net

that controls the rate at which the loss for well-classified examples is reduced, then the binary Focal loss can be defined as:

$$BinaryFocalLoss = -\alpha \times (1-p)^{\gamma} \times \log(p) - (1-\alpha) \times p^{\gamma} \times \log(1-p) \quad (3.8)$$

Hybrid Loss: The Hybrid loss combines two well-known loss components, the Dice Loss and the Focal Loss, in an attempt to improve the training and performance of a segmentation model.

$$HybridLoss = DiceLoss + (2 \times FocalLoss) \quad (3.9)$$

3.2.2 SEAU-Net

In this section, the detailed implementation of the SEAU-Net is elaborated, including the decoder block and TSR module. The process commences with the initial pre-processing of the dataset. Subsequently, the prepared training and validation images are supplied to the model for the training phase. Following the training process, the model's efficacy is evaluated by testing it on a separate dataset dedicated to assessing its performance. The testing phase on the independent test dataset aims to validate the generalization capabilities of the model beyond the training data.

A novel model for medical image segmentation is proposed, named Squeeze-and-Excitation aided Attention U-Net (SEAU-Net). It is based on the encoder-decoder architecture of U-Net, utilizing and enhancing the advantages of the U-shaped structure. Figure 3.8 shows the structure of the SEAU-Net.

The SEAU-Net architecture consists of three main parts: an encoder, a decoder and skip connections.

- **Encoder** : The conventional U-Net contracting path consisting of encoder blocks is replaced with a ResNet50 backbone. The residual modules integrated within the backbone help in reducing the gradient vanishing problem. The backbone can extract rich feature representations and is pre-trained having ImageNet weights. Skip connections are drawn from different layers in the encoder architecture, producing feature maps of dimensions (112^2,64), (56^2,256), (28^2,512) and (14^2,1024) respectively. The entire ResNet50 backbone is not considered as it doubles the number of parameters.

- **Skip connections**: In this approach, SE attention blocks are strategically employed in the two lower skip connections. This selective application of attention to lower-level connections helps reduce redundancy and computational load. The feature maps from these layers are upscaled and then fed into the upper two layers of the decoder. Specifically, the feature map from the first skip connection is directed to the third decoder layer, and similarly, the map from the second connection is sent to the fourth layer. Integrating attention-enhanced feature maps from the lower to the upper layers improves the spatial resolution in the final output. This means that in addition to the standard horizontal skip connections at each layer, two vertical, upscaled connections that bridge the lower and upper layers are also incorporated.

- **Decoder**: The network includes four decoder blocks, each designed to reconstruct the high-level semantic feature map from the encoder back to the original image's dimensions. The process starts with the feature map from the bridge layer being upscaled using a bilinear upsampling layer. This upscaled map is then merged with the feature map from the attention block through concatenation. Subsequently, the map undergoes further restoration involving another upsampling step, a 1x1 convolutional layer, and the TSR block(Figure 3.10). The same process holds for the other three decoder blocks. Each decoder block is composed of this sequence: a concatenation layer, a bilinear upsampling layer, a 1x1 Conv layer, and the Tri-Step Residual (TSR) block. To generate the final probability segmentation map, a 1x1 Conv layer followed by a sigmoid activation layer is used. Moreover, every layer in the decoder integrates features from the corresponding encoder layer, enhancing the overall feature fusion.

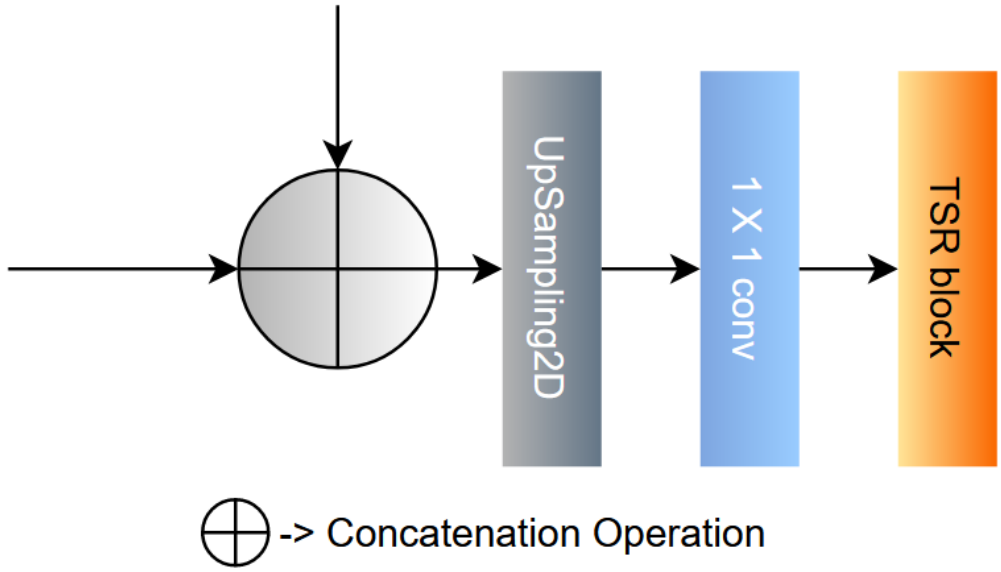


Figure 3.9: Decoder block architecture

3.2.2.1 Tri-Step Residual Block

The TSR block structure, as proposed, deviates from the conventional residual architecture. The key distinction lies in the concatenation of three feature maps. Typically, a standard residual structure incorporates two 3×3 convolution layers. However, this network employs a single sequence of Conv2D operations with 3×3 filters, followed by batch normalization and rectified linear unit (ReLU) activation to extract features. Initially, the input is passed through a 3×3 convolution layer, batch normalization, and then activated using ReLU. This produces an intermediate feature map. The process is repeated with another 3×3 convolution layer and batch normalization, without an immediate activation step.

The network diverges from the standard structure by introducing an element-wise addition of the initial feature map, the intermediate feature map and the final map. This step effectively creates a shortcut connection, allowing the network to bypass potential performance-degrading layers by directly transferring the original features to deeper layers. The combined output is then passed through a ReLU activation function.

This design choice aims to maintain feature integrity across layers and prevent degradation in deeper networks, a common issue in very deep architectures. By integrating the initial input (x-init) and the intermediate output (x-break) with the current layer output, the network ensures more effective feature propagation and potentially better learning outcomes.

To describe the residual network block in mathematical terms, the operations applied to the input x are considered in a sequential manner. Note that in the

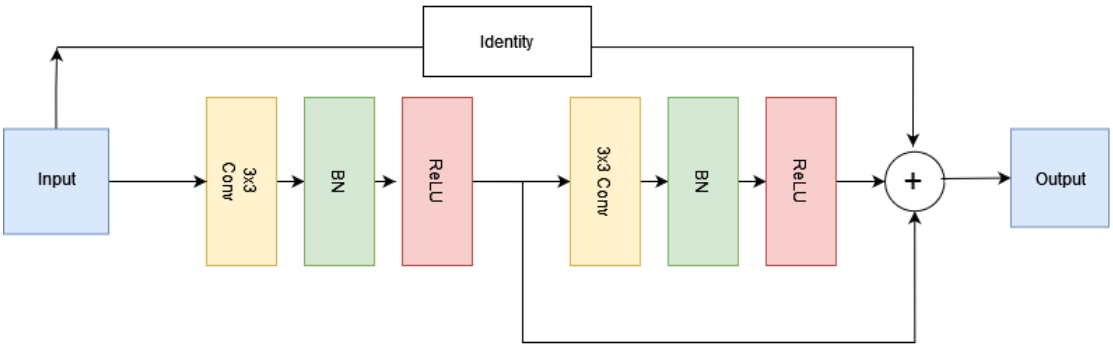


Figure 3.10: Architecture of the proposed TSR block.

context of CNNs, the operations are applied on a per-feature map basis, and the input x and the intermediate variables are tensors. The operations are as follows:

$$x_1 = Conv2D_{3\times3,num_filters}(x_{input}) \tag{3.10}$$

This represents a convolution operation with a 3×3 kernel and '*num-filters*' number of filters.

$$x_{intermediate} = ReLU(BatchNorm(x_1) \tag{3.11}$$

This step stores the intermediate output.

$$x_{final} = BatchNorm(Conv2D_{3\times3,num_filters}(x_{intermediate})) \tag{3.12}$$

They are concatenated as follows:

$$x_{final} = x_{final} + x_{intermediate} + x_{input} \tag{3.13}$$

This is a key step in a residual block where the output of the second batch normalization, the intermediate output, and the original input are added together. This step helps in addressing the vanishing gradient problem by allowing gradients to flow directly through the shortcut connections.

3.2.2.2 Squeeze-and-Excitation Block

The SE block describes a squeeze and excitation block, designed to enhance the representational capacity of the network (Figure 3.11). The feature map is first processed through a GlobalAveragePooling2D (GAP) layer. This operation computes the spatial average for each channel in the input feature map, effectively squeezing the spatial dimensions to 1x1 while retaining the depth (channel-wise information). The squeezed output is then fed into a dense (fully connected) layer. This layer reduces the channel dimensionality. This layer does not use a bias term

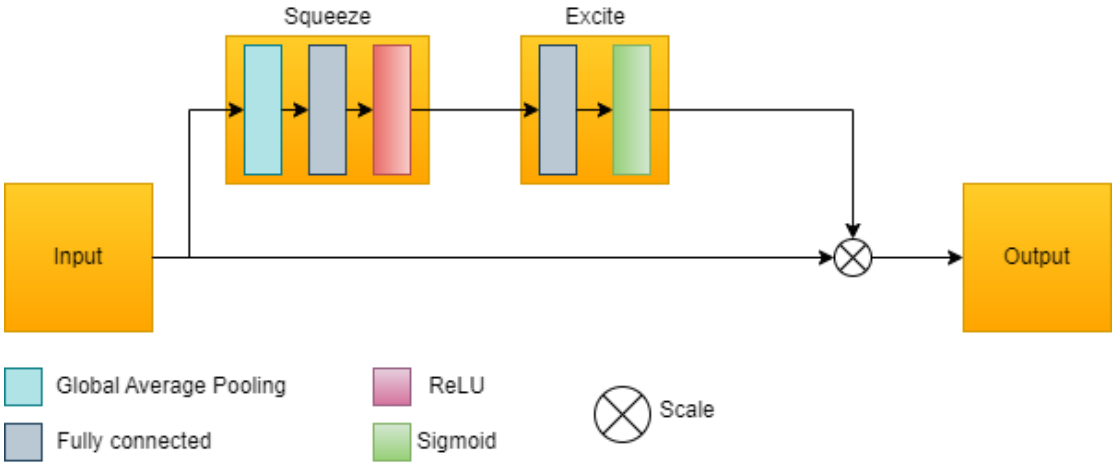


Figure 3.11: Architecture of Squeeze-and-Excitation block

and uses ReLU as the activation function. The output from the first dense layer is passed to another dense layer. Similar to the first dense layer, this layer does not use a bias term and employs a sigmoid activation function.

The output from the second dense layer, which is a channel-wise weighting factor, is multiplied (element-wise) with the original input. This scaling process allows the network to perform channel-wise feature recalibration dynamically.

The SE block, through this sequence of operations, learns to emphasize informative features and suppress less useful ones, thereby improving the representational power of the network. This mechanism is particularly beneficial in CNNs, where it helps to focus on the most relevant features for the task at hand.

$$x_{final} = \sigma(Dense(ReLU(Dense(GAP(x_{input}))))) \times x_{input} \tag{3.14}$$

3.2.2.3 Loss Function

A loss function is introduced based on the combination of dice loss and volumetric loss, which helps in reducing the class imbalance problem that is prevalent in medical imaging datasets and capture irregular boundaries that prevail between liver and its surrounding organs. The dice loss formula for binary segmentation can be rewritten as:

$$DiceLoss = 1 - \frac{2 \times Intersection + Smooth}{\sum_i^N y_{\text{true}} + \sum_i^N y_{\text{pred}} + Smooth} \tag{3.15}$$

where:

$$Intersection = \sum_i^N (y_{true} \cdot y_{pred}) \tag{3.16}$$

and 'smooth' is a smoothing factor kept as 0.001. In this context:

- 'y_{pred}' represents the predicted label of each pixel belonging to the foreground class.
- 'y_{true}' is the actual label of each pixel, with 1 indicating the foreground class and 0 indicating the background.
- N denotes the total number of pixels.

Volumetric loss measures the difference in volume between the segmented region obtained by the algorithm and the actual ground truth.

$$Loss_{Volumetric} = 1 - \frac{\sum_i^N (y_{true} \cdot y_{pred})}{\sum_i^N y_{true} + \sum_i^N y_{pred} - \sum_i^N (y_{true} \cdot y_{pred})} \tag{3.17}$$

This formula essentially measures the degree of overlap between the predicted and true volumes. A lower volumetric loss indicates a better overlap and, consequently, a more accurate segmentation. Volumetric loss is related to the Dice coefficient but is more specific to volumetric data, which is common in medical imaging (e.g., MRI, CT scans).

The final loss function is formulated as:

$$Loss_{Total} = \alpha \cdot Loss_{Dice} + \beta \cdot Loss_{Volumetric} \tag{3.18}$$

where 'α' and 'β' denote weights assigned to dice and volumetric loss respectively. In this study, they are both kept as 1.

3.2.3 W^{2}CU-Net

The suggested liver and tumor segmentation model (W^2CU-Net) is described in-depth in this section.

3.2.3.1 Deep Neural Based Backbone for Encoder

CNNs are favored for image segmentation, especially in medical imaging tasks like liver and tumor segmentation. Their design, incorporating convolutional and pooling layers, enables effective capture of local patterns and spatial relationships in complex images. CNNs automate feature learning, eliminating manual engineering. The MobileNetV2 backbone is selected for the encoder due to its satisfactory performance during the initial stages of the process.

3.2.3.2 MobileNetV2

MobileNetV2, introduced by [77], is designed for embedded and mobile vision applications. It enhances efficiency with linear bottlenecks, inverted residuals, and depthwise separable convolution. The architecture includes two types of blocks—shrinking blocks with a stride of two and residual blocks with a stride of one, each with three levels. Utilizing a width multiplier, it adjusts for hardware limitations. As a lightweight and efficient model, MobileNetV2 is suitable for deployment on mobile devices and in resource-constrained scenarios, delivering high accuracy in tasks like semantic segmentation, object detection, and image categorization. Figure 3.12 displays the MobileNetV2 architecture.

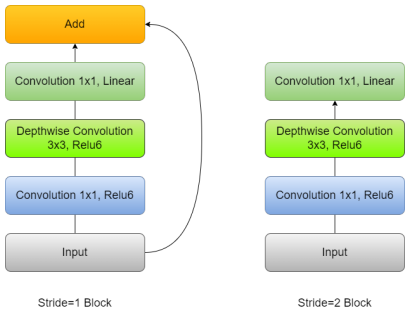


Figure 3.12: Architecture of the MobileNetV2 model

3.2.3.3 Proposed Model Architecture

The proposed model, W^2CU-Net, enhances flexibility in handling datasets of varying complexity. A Weibull function is used to implement a fuzzy ranking-based approach, considering uncertainty in feature maps' predictions and assigning different importance levels. Cosine similarity-based feature pruning dynamically removes less important feature maps, allowing the model to focus on crucial features. Results undergo range-based ensemble using an elliptic curve (Weierstrass equation), enhancing confidence score distribution for a better ensemble of outputs. The two-decoder approach allows the model to obtain outputs for both a decoder that can learn and one that does not learn, respectively. The Figure 3.13 details the base model, including the repetitive Residual_Conv_Block illustrated in Figure 3.14.

3.2.3.4 Implementation of the Entire Model

The model is created using a part of MobileNetV2 pre-trained model (till block_14_add layer) as the encoder. Two decoder types are employed: trainable and non-trainable. The non-trainable decoder utilizes Upsampling layers

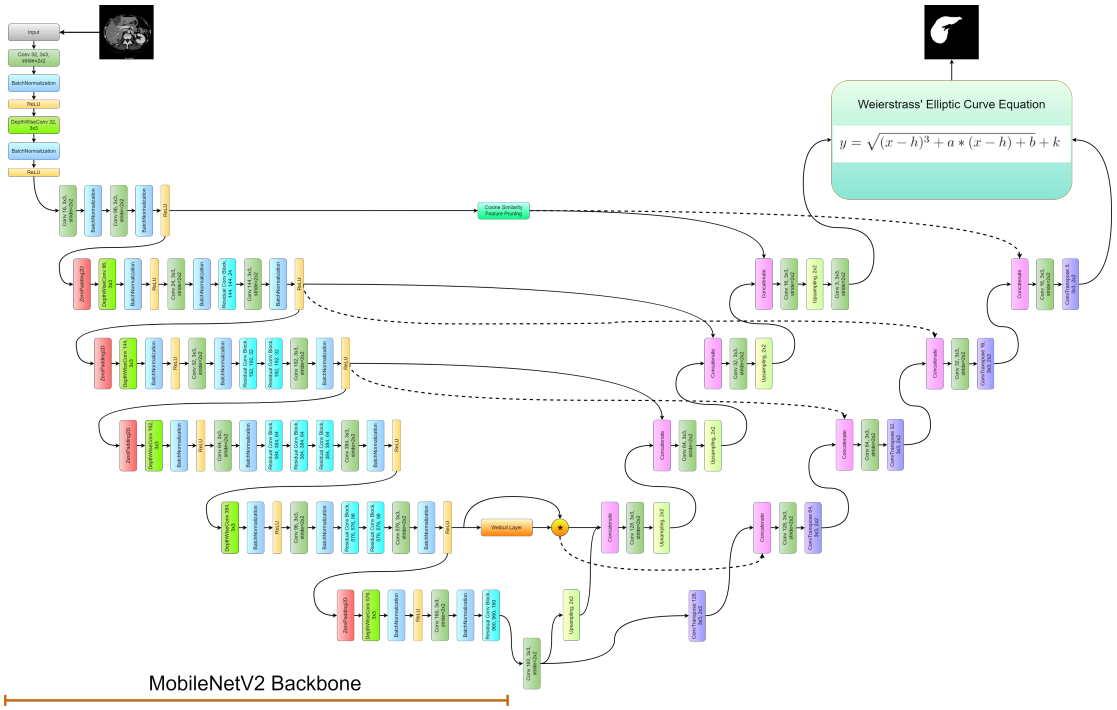


Figure 3.13: Architecture of the proposed W^2CU-Net

for upsampling encoder outputs and skip connections. In contrast, the trainable decoder employs ConvTranspose layers to upsample encoder outputs and skip connections, learning from them simultaneously. For a single CT image given as an input, the first skip connection (i.e., the outermost skip connection) will have some shape with 96 feature maps as the last dimension. The tensors that go as inputs to the cosine similarity feature pruning layer have a shape of $[batch, height, width, channels]$, where channels represent the total number of feature maps (i.e., 96).

3.2.3.5 Detailed Cosine Similarity based Feature Pruning Algorithm

In this section, a detailed description of the Cosine Similarity based feature pruning layer is provided. Algorithm 2 provides a detailed steps of the same.

Significance of Cosine Similarity based Feature Pruning: The proposed model incorporates a unique cosine similarity-based feature pruning layer to eliminate non-informative feature maps. Positioned in the outermost skip connection, known for less useful feature maps, it aggregates input feature maps through pixel-wise averaging. Cosine similarity scores between each input feature map and the aggregate feature map are computed, considering the top $\alpha\%$ based on these scores. This process retains feature maps closer to the mean while discarding those farther away, concentrating on more relevant information. Despite the computational complexity, the GPU's tensor operations ensure negligible overall

Algorithm 2 Cosine Similarity based Feature Pruning

$N \leftarrow Total\ no.\ of\ feature\ maps$
$top \leftarrow Percentage\ to.\ accept$

function COSINE_SIMILARITY(x, y)
    $x_flat \leftarrow Flatten(x)$ ▷ Flattened version of x
    $y_flat \leftarrow Flatten(y)$ ▷ Flattened version of y
    $score \leftarrow \frac{\sum_{i=1}^{N} x_flat_i * y_flat_i}{\|x_flat\| * \|y_flat\|}$ ▷ Calculating Cosine Similarity Score between flattened x and y.
    return *score*
end function

function SIMILARITY_COMPUTE($mean, argv$)
    $similarities \leftarrow array()$
    for $arg_i \in argv$ **do**
        $cosine_sim \leftarrow$ **cosine_similarity(mean, arg_i)** ▷ Calculating the Cosine Similarity Score for ith feature map with aggregate feature map
        $similarities \leftarrow Add\ cosine_sim\ to\ this\ array$
    end for
    $res \leftarrow Concatenate(arg_i\ if\ similarities[i] \in top$
    return *res*
end function

function CALL($inputs$)
    $L \leftarrow array(inputs[:, :, :, i])$ ▷ Extracting feature maps from input
    $agg_feature_map \leftarrow \frac{1}{N} * \sum_{i=1}^{N} inputs[:, :, :, i]$
    $results \leftarrow$ **similarity_compute(agg_feature_map, *L)** ▷ Passing mean and unpacked list as function input
    return *pred*
end function

results ← **Call(inputs)**

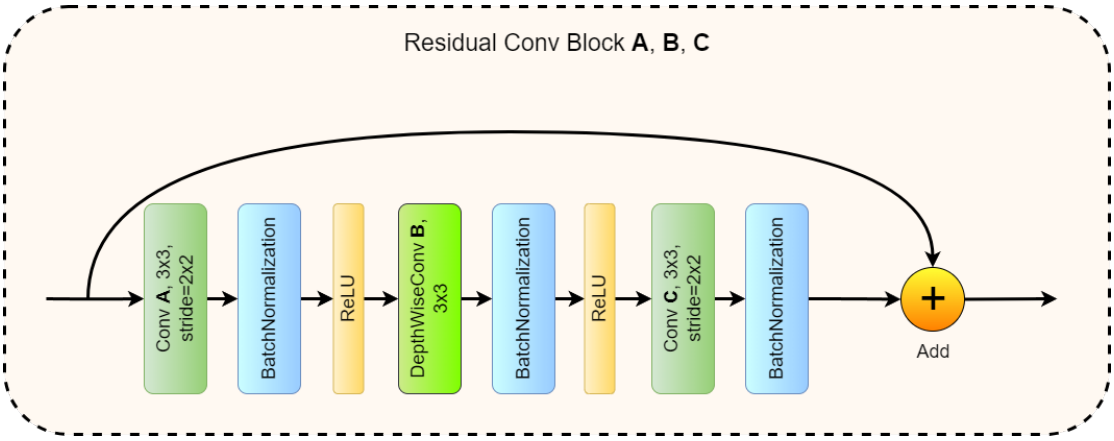


Figure 3.14: Residual_Conv_Block in W^2CU-Net

processing time.

$$Agg.FeatureMap = \frac{1}{C} * \sum_{i=1}^{C} inputs[:, :, :, i]; \tag{3.19}$$

where C = No. of channels (feature maps), $inputs$ = the input tensor with the feature maps.

Using Equation 3.19, the aggregate feature map is created by extracting each feature map based on their dimensions and then averaging them pixel-wise. Then, using Equation 3.20, the cosine similarity of each extracted feature map with the aggregate feature map obtained from Equation 3.19 is found.

$$Cosine\ Similarity\ Score\ (x, y) = \frac{\sum_{i=1}^{n} x_i * y_i}{\|x\| * \|y\|}; \tag{3.20}$$

where x and y are the flattened version of input feature maps, n = the total number of elements in the flattened version of x or y, $\|x\|$ and $\|y\|$ = Euclidean norms of the flattened tensors x and y, respectively.

The cosine similarity scores are computed as depicted in Equation 3.21. Based on these scores in the similarity_compute array, the feature maps that belong to the top ceiling of α% are chosen else pruned out.

$$\begin{aligned} &similarity_compute(mean, argv) = \\ &[cosine_similarity(mean, arg_1), \ldots, \\ &cosine_similarity(mean, arg_N)]; \end{aligned} \tag{3.21}$$

where arg = the input feature maps, $mean$ = the aggregate feature map, N = total no of feature maps.

So from the 96 input feature maps, the top α feature maps with the highest

cosine similarity scores are chosen and passed as outputs. This clips out the feature map with the worst features while keeping the more important features.

In the innermost skip connection, the Weibull layer processes tensors with a shape of $[batch, height, width, channels]$, where channels represent 576 feature maps.

Significance of Weibull Function: The Weibull function generates weights for feature maps based on a dose response curve in 0 to 1 domain on x axis. It measures the fit between feature maps and final decided weights by assigning fuzzy ranks to class predictions for each pixel. The adaptive weight generation process adjusts ranks based on input specifics, enhancing adaptability. Stacking identical results or weights and employing the Hadamard product [78] with the input yield final output. This method creates an adaptive weight-generating layer for nuanced adjustments in feature map ranking. Weights are influenced by the Weibull function's sharp decline, particularly in the innermost skip connection, dealing with enriched feature maps.

Now, each pixel $Pix_c^{(i)}$ in the input has M distinct confidence scores, denoted as $CoS^{(1)}, CoS^{(2)}, ..., CoS^{(M)}$, where M is 576. The decision scores are normalized in Equation 3.22, where C is the total number of classes for the dataset.

$$\sum_{c=1}^{C} CoS_c^{(i)} = 1; \forall i,\ i = 1, 2, 3, ..., M \tag{3.22}$$

The confidence scores of all the pixels in each of the test images, belonging to various classes, are used to produce fuzzy ranks. The fuzzy rank is produced by the Weibull function for a class c using the i^{th} feature map's decision scores as shown in Equation 3.23.

$$Rank_c^{(i)} = \frac{e^{-2*(CoS_c^{(i)})^2}}{2}; \tag{3.23}$$

$\forall i, c,\ i = 1, 2, 3, ..., M; c = 1, 2, ..., C$

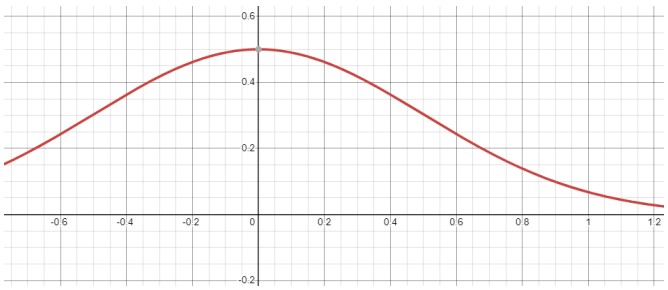


Figure 3.15: Modified Weibull function used in the present work.

Figure 3.15 depicts the graphical depiction of the re-parameterized Weibull function, as mentioned in Equation 3.23.

The value of $Rank_c^{(i)}$ is between 0.0677 to 0.5, with the smallest value 0.0677 being equivalent to rank 1 (best rank), i.e., greater confidence results in a lower (better) value of rank. The fuzzy rank sum (FS_c) and complement of confidence factor sum ($CCoS_c$) are computed as follows if the top k ranks are represented by $K^{(i)}$, that is, ranks 1, 2,..., k, corresponding to class c are as follows [75]:

$$FS_c = \sum_{i=1}^{M} \begin{cases} Rank_c^{(i)}, & \text{if } Rank_c^{(i)} \in K^{(i)} \\ P_c^{Rank}, & \text{otherwise} \end{cases} \tag{3.24}$$

$$CCoS_c = \frac{1}{M} \sum_{i=1}^{M} \begin{cases} CoS_c^{(i)}, & \text{if } Rank_c^{(i)} \in K^{(i)} \\ P_c^{CoS}, & \text{otherwise} \end{cases} \tag{3.25}$$

If class c does not fall among the top k class ranks, penalties P_c^{Rank} and P_c^{CoS} are applied on it. Using the aforementioned Weibull function, the value of P_c^{Rank} is set to 0.5, which is determined by setting $CoS_c^{(i)} = 0$; and the value of P_c^{CoS} is set to 0. The penalty values prevent class c from emerging as an improbable winner. The combination of FS_c and $CCoS_c$, which is used to produce the method's final predictions, yields the confidence score in question. Equation 3.23 determines the final decision score (FD).

$$FD_c = FS_c * CCoS_c \tag{3.26}$$

Finding the class with the lowest FD value yields the final projected class for the particular pixel in the data instance $\mathbf{I}$, which is provided as shown in Equation 3.24.

$$class(\mathbf{I}) = \underset{c=1,2,...,C}{\arg\min}\, FD_c \tag{3.27}$$

This output will have only one channel in its shape so to use them as weights this same output is stacked up to the total number of feature maps as the input. This stacked up tensor is then Hadamard product with the input tensor for generating final layer output. The overall process of the approach has been described in 3.

The model is trained with these layers included. The trained model now generates to outputs from its two decoders. These two results are then fused together using the Equation 3.28, called the Weierstrass' Elliptic Curve Equation.

Significance of Weierstrass' Elliptic Curve Equation: The final outcomes or the reults of final Sigmoid output from both of the model's decoders are

Algorithm 3 Weibull Function based Layer

$C \leftarrow Total\ no.\ of\ classes\ in\ an\ image$
$M \leftarrow No.\ of\ feature\ maps$
$p1, p2, .., pN \leftarrow Feature\ maps\ from\ previous\ layer$

function FUZZY_RANK(CF)
    $R_L \leftarrow \frac{e^{-2*CF^2}}{2}$ ▷ Applying Weibull function to each class of every pixel
    $K_L \leftarrow array(shape = R_L, 0.5)$ ▷ Initialize array of R_L shape with penalty values
    if $R_L_c^i \in top\ 1\ rank$ **then** ▷ R_L for a pixel results by ith feature map for cth class
        $K_L_c^i \leftarrow R_L_c^i$
    end if
    return K_L
end function

function CFS_FUNC(CF, K_L)
    if $R_L_c^i \notin top\ 1\ rank$ **then**
        $CF_c^i \leftarrow 0$ ▷ Applying penalty values
    end if
    $CFR \leftarrow \sum_{c=1}^{C} \sum_{i=1}^{M} CF_c^m, i$
    $CFS \leftarrow 1 - \frac{CFR}{M}$
    return CFS
end function

function WEIBULL($p1, p2, ..., pN$)
    $CF \leftarrow array(array(p1, p2, ..., pN), array(1 - p1, 1 - p2, ..., 1 - pN))$ ▷ Element-wise subtract from 1
    $pred \leftarrow Array\ initialized\ with\ shape\ of\ p1$
    $R_L \leftarrow$ **fuzzy_rank(CF)**
    $RS \leftarrow \sum_{c=1}^{C} \sum_{i=1}^{M} R_L_c^i$
    $CFS \leftarrow$ **CFS_func(CF, R_L)**
    $FS \leftarrow RS \times CFS$
    if $FS_c\ is\ minimum$ **then**
        $pred \leftarrow class\ c$ ▷ Predicting the class for each pixel
    end if
    $predictions \leftarrow tensor()$
    for $i \leftarrow 1\ to\ M$ **do**
        $predictions \leftarrow Concatenate(preds)$
    end for
    return $predictions$ ▷ Predictions have M (N) channels or feature maps all equal to pred
end function

results ← **Weibull(p1, p2, ..., pN)**

fused using a range-based ensemble method with a re-parameterized Weierstrass' elliptic curve equation. Each decoder sigmoid result undergoes this equation, expanding the distribution and then added elementwise with weights. The result is interpolated to a range of +1 to -1. The interpolated result determines the final class for each pixel, emphasizing smaller variations for better decision-making. If the value surpasses a threshold, it belongs to one class; otherwise, it belongs to the other.

$$Res_S_r = \sqrt{(x-h)^3 + a*(x-h) + b} + k; \tag{3.28}$$

where r = rth input result, $h = -0.89,\ a = -1.57,\ b = 0.705$ and $k = -1.115$.

The distribution of each of the two decoder sigmoid results is increased into a larger range than 0 to 1 using Equation 3.28. This has been applied to each pixels of the two outputs. The results (RS) get expanded in the range of -1.002 to 1.004. This gives even smaller distribution to have larger effect in final ensemble result. Figure 3.16 provides a pictographic representation of Weierstrass' Elliptic curve equation.

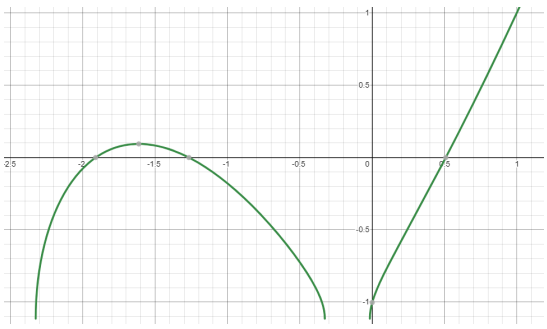


Figure 3.16: Modified Weierstrass' Elliptic Curve Equation.

Two results Res_S1 and Res_S2 are obtained after applying Equation 3.28 on each of the two inputs. They are now added pixel-wise with weights (dice coefficients) to generate final result i.e., $Res_S = \alpha*RS1+\beta*RS2, where\ \alpha\ and\ \beta\ are\ the\ weights\ ad$ Now this final result is then interpolated into the range of -1 to +1 using the linear interpolation formula in Equation 3.29.

$$MV = \frac{(GS - Xmin)*(Ymax - Ymin)}{Xmax - Xmin} + Ymin; \tag{3.29}$$

where GS = each pixel value, $Xmax$ and $Xmin$ are max and min score of Res_S respectively, $Ymax = 1.0,\ Ymin = -1.0$.

The final class of a pixel is predicted by its value. If value is above 0.0 then it belongs to class 1 else class 2. 4

Employing a unique feature pruning technique, along with the incorporation of fuzzy rank weight generation and consideration of two decoder types, alongside a

Algorithm 4 Weierstrass Elliptic Curve Equation based Ensemble

$C \leftarrow Total\ no.\ of\ classes\ in\ an\ image$
$q \leftarrow Total\ no.\ of\ pixels$
$p1 \leftarrow result\ of\ 1st\ decoder$
$p2 \leftarrow result\ of\ 2nd\ decoder$

function WEIERSTRASS_ELLIPTICAL_FUNC($CS1, CS2$)
$h \leftarrow -0.89;\ a \leftarrow -1.57;\ b \leftarrow 0.705;\ k \leftarrow -1.115$
$Res_S1 \leftarrow \sqrt{(CS1-h)^3 + a*(CS1-h) + b} + k$ ▷ Applying Weierstrass function to each pixel value of 1st input
$Res_S2 \leftarrow \sqrt{(CS2-h)^3 + a*(CS2-h) + b} + k$ ▷ Applying Weierstrass function to each pixel value of 2nd input
$Res_S \leftarrow Res_S1 * \alpha + Res_S2 * beta$ ▷ Pixel-wise weighted addition; (α and β) $\rightarrow$ weights (dice coefficient)
return Res_S
end function

function LINEAR_INTERPOLATION(GS)
$Xmax \leftarrow GS.max; Xmin \leftarrow GS.min; Ymax \leftarrow 1.0; Ymin \leftarrow -1.0$
$MV \leftarrow \frac{(GS-Xmin)*(Ymax-Ymin)}{Xmax-Xmin} + Ymin$
return MV
end function

function FUSION($p1, p2$)
$GS \leftarrow$ **Weierstrass_Elliptical_Func(p1, p2)**
$IS =$ **Linear_Interpolation(GS)**
if $IS_q > 0.0$ **then** ▷ Predicting the class for each pixel
$pred \leftarrow class\ c_1$
else
$pred \leftarrow class\ c_1$
end if
return $pred$
end function

results $\leftarrow$ **Fusion(p1, p2)**

novel ensemble method, enhances the superiority of the propoed model compared to others.

CHAPTER 4

Results and Discussion

The results presented in this section highlight the efficacy of the proposed models in medical image segmentation. section 4.1 highlights the system configuration that is considered for this study and section 4.2 elaborates the evalution metrics that are used. section 4.3, section 4.4 and section 4.5 shows the results and highlights the analysis that is drawn from the three methods.

4.1 System Configuration

A Jupyter Notebook with a 12 GB NVIDIA Tesla T4 GPU in Google's collaborative environment is used to run the whole set of experiments. Tensorflow, Keras, Matplotlib, Scikit, Numpy, and Pandas are the main open-source modules that have been used for implementing the suggested method in Python.

4.2 Evaluation Metrics

Metrics for evaluation are crucial in determining a predictive/learning model's potency and usefulness. These metrics are used to assess how well a system performs in predicting the correct class label for an input in classification challenges. Considering a situation with two classes, where one class is designated as "positive" and the other as "negative."

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

- **True Negative (T_N)** refers to the number of accurately categorized negative samples.

For the present work, the following performance metrics are used:

Dice Score Coefficient: The DSC, also known as the Sørensen-Dice coefficient, is a commonly used metric for measuring the similarity between two sets. In the context of image segmentation, it is used to evaluate the similarity between a predicted segmentation mask and a ground truth segmentation mask. If the total number of pixels predicted as foreground in the predicted mask is $|P|$ and the total number of pixels labeled as foreground in the ground truth is $|GT|$, then the Dice Score can be formulated as:

$$DSC = \frac{2 \cdot |P \cap \mathrm{GT}|}{|P| + |\mathrm{GT}|} \tag{4.1}$$

Here $|P \cap \mathrm{GT}|$ indicates the total number of common pixels present as foreground in both the mask and the predicted image.

Intersection over Union The IoU is the ratio of the area of overlap between the predicted bounding box and the ground truth bounding box to the area of union between the two bounding boxes. For pixel-wise binary classification, the IoU can be formulated as:

$$IoU = \frac{T^P}{T^P + F^N + F^P} \tag{4.2}$$

Relative Volume Difference: RVD [16] is especially significant in areas such as medical imaging and segmentation assessment. Its task is to estimate the volume difference between a segmented region and the associated ground truth. The stages in this method are to find the absolute difference in the volume between the segmented area and the real region, and then divide this difference by the volume of the real region. RVD sheds information on the volumetric differences between the divided region and the real region.

Accuracy: The percentage of events that are correctly predicted over all instances is known as accuracy. It is employed to assess how well a model performs in terms of producing accurate predictions. It is determined by dividing the total number of predictions made by the model by the number of predictions that were accurate. Mathematically, Equation 4.3 represents the accuracy.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{4.3}$$

Precision: The proportion of accurate positive predictions amongst all positive predictions made by a model is known as its precision. It is determined

by dividing T_P by the product of T_P and F_P. Mathematically, precision can be represented as in Equation 4.4.

$$Precision = \frac{T_P}{T_P + F_P} \tag{4.4}$$

Recall: Recall is the percentage of actual positive events that the model correctly identifies. It is the proportion of T_P to T_P plus F_N as shown in Equation 4.5. A higher recall demonstrates that the model is better at identifying positive examples, whereas a lower recall raises the possibility that some positive cases may have been missed by the model.

$$Recall = \frac{T_P}{T_P + F_N} \tag{4.5}$$

4.3 BLUE-Net

The ensemble of three U-Net based models is applied for liver and liver-tumor segmentation tasks using the Blumberg function. Table 4.1 shows the hyperparameter for the base models. Based on the hybrid loss function, the segmentation results obtained for the BLUE-Net is shown in Table 4.2. It shows the results of different base models and the final result after ensembling these base models.

Table 4.1: Hyperparameters of the base classifiers for the BLUE-Net model.

Hyperparameter	**Value/Name**
Optimizer	Adam [79]
Loss function	Hybrid Loss
Learning rate	0.001
No. of epochs	60

The loss curves so obtained are shown in Figure 4.1

Some ablation studies are conducted on the 3DIRCADb dataset to ensure the suggested network is adequate. Table 4.3 and Table 4.4 report the results of ablation studies for liver and tumor segmentation tasks based on the different loss functions. In both tables, the model with the hybrid loss function yields the best DSC score as compared to the other two loss functions.

Table 4.2: Segmentation results obtained for BLUE-Net.

Dataset	**DSC**$_{U-Net}$	**DSC**$_{Link-Net}$	**DSC**$_{FPN}$	**DSC**
3DIRCADb (liver)	0.9410	0.9439	0.9520	0.9774
3DIRCADb (tumor)	0.7102	0.7679	0.7309	0.8359
CHAOS	0.9563	0.9113	0.7886	0.9750
LiTS (liver)	0.9728	0.9527	0.9751	0.9775
LiTS (tumor)	0.8975	0.8957	0.8290	0.9144
BUSI	0.4681	0.4534	0.4606	0.4923
OCDC	0.7258	0.8549	0.7674	0.8076
ISIC 2018	0.6898	0.7500	0.7011	0.7847

Table 4.3: Ablation study for liver segmentation on the 3DIRCADb dataset for BLUE-Net.

Model	**Loss function**	**DSC**
BLUE-Net	Focal Loss	0.9470
	Dice Loss	0.9662
	Hybrid Loss	**0.9774**

Table 4.4: Ablation study for tumor segmentation on the 3DIRCADb dataset for BLUE-Net.

Model	**Loss function**	**DSC**
BLUE-Net	Focal Loss	0.5659
	Dice Loss	0.6324
	Hybrid Loss	**0.8359**

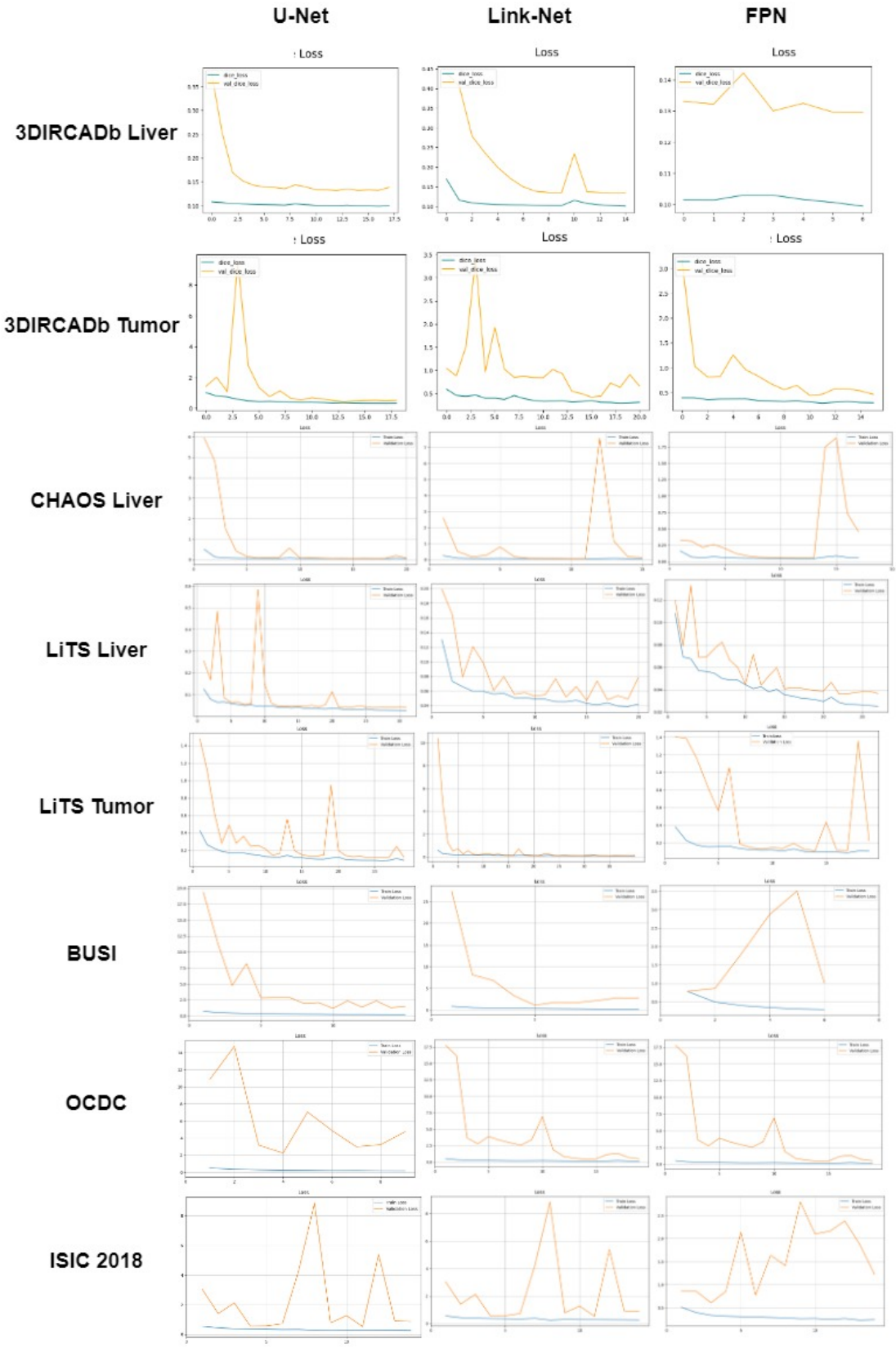


Figure 4.1: Loss curves for BLUE-Net obtained by using the hybrid loss function. Here x-axis denotes epoch number and y-axis denotes loss value

4.3.1 GradCAM Analysis

In this section, assistance is taken from a data visualization tool to visually display some results of the proposed method, namely GradCAM. GradCAM, a technique proposed by [80], is utilized to generate visual explanations of the model predictions by creating a gradient-weighted class activation map. These visualizations help to demonstrate how neural networks arrive at decisions. As shown in Figure 4.2e, Figure 4.2f, and Figure 4.2g, GradCAM is used to generate visualizations for the U-Net, Link-Net, and Feature Pyramid Network (FPN), respectively, from the 3DIRCADb dataset, using the three base models that have been used to construct the ensemble model.

It is evident that the various models concentrate on distinct areas of the Liver CT scans, thereby suggesting that individual learners capture unique and complementary information that is required to form an ensemble-learning-based model.

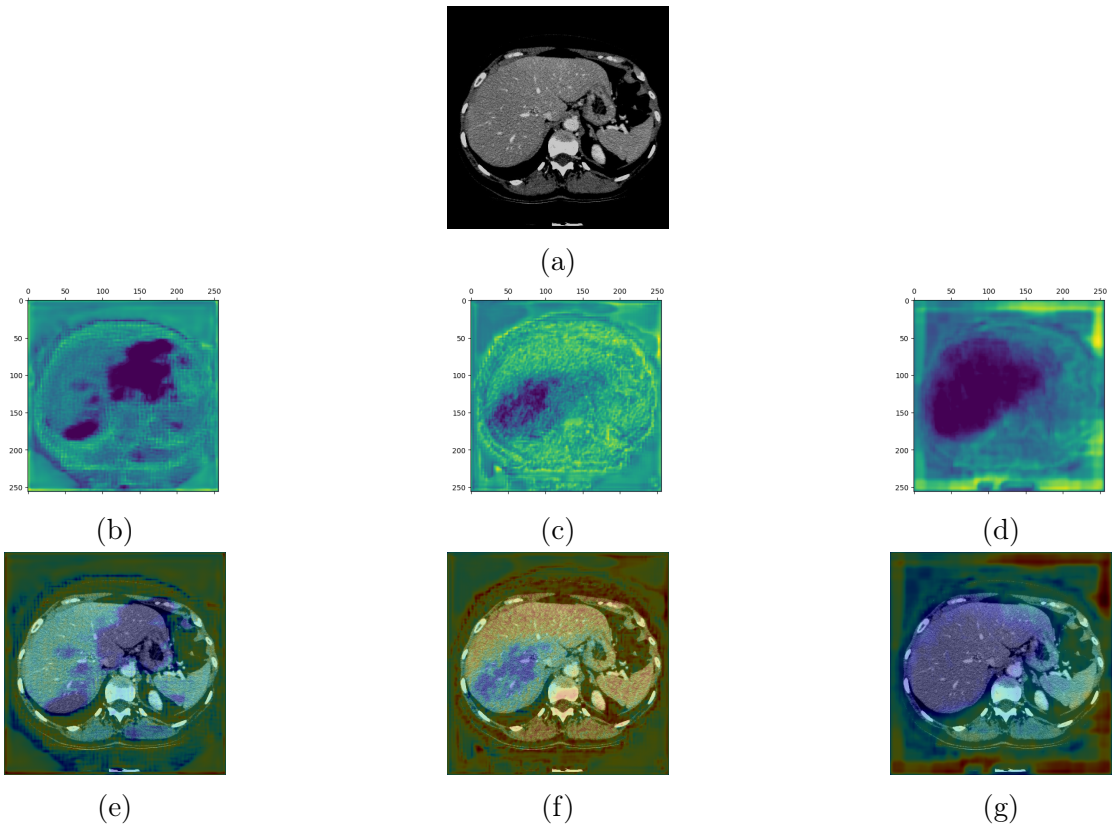


Figure 4.2: GradCam of the Models on 3DIRCADb Dataset: (a) Original Image, (b) Heatmap of U-Net, (c) HeatMap of Link-Net, (d) HeatMap of FPN, (e) GradCam of U-Net, (f) GradCam of Link-Net, (g) GradCam of FPN, for Liver Segmentation.

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

2. **Link-Net:** The GradCam for the Link-Net indicates that this model places attention on critical liver regions, showing how it leverages its mechanisms for accurate segmentation.

3. **FPN:** FPN'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.

4.3.2 Error Analysis

Despite outperforming the state-of-the-art results for liver and tumor segmentation tasks, the proposed model renders a few undesirable results, as shown in Figure 4.3.

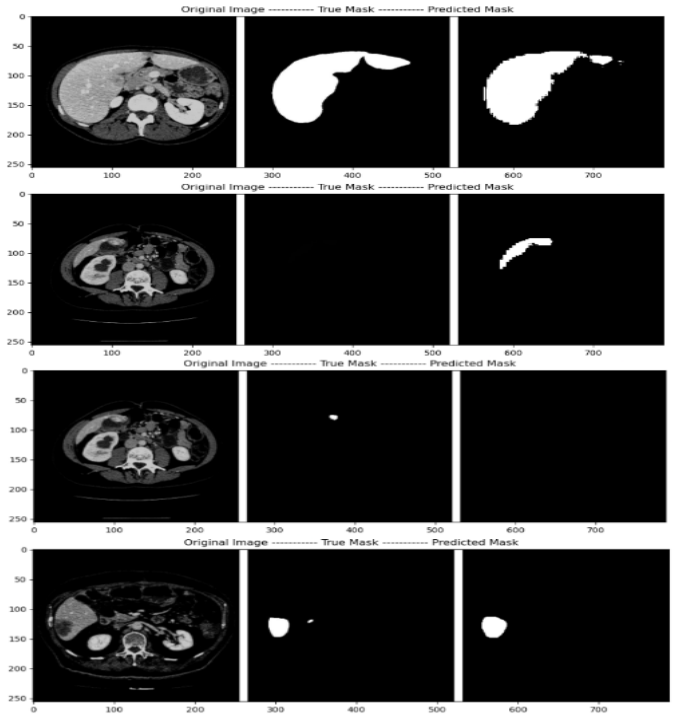


Figure 4.3: Some undesirable outputs for liver segmentation (the first 2 images) and tumor segmentation (the next 2 images)

.

Segmentation models can struggle with images containing uneven object boundaries, poor quality images, occlusions, or significant background clutter. Even well-trained models can produce noisy segmentation masks with fragmented or disconnected regions. While an ensemble method often leads to improved performance, it is not a guarantee that it will always provide better results than a single well-tuned model (despite providing a better DSC as observed here). Thereby,

base models that capture varied information about the input images might provide better results than solely depending on the ensemble.

4.4 SEAU-Net

This section outlines the empirical findings derived from the training and testing phases of the SEAU-Net architecture proposed in this study. The Table 4.5 shows the hyperparameters chosen for training the model for both datasets.

Table 4.5: Hyperparameters considered for training the SEAU-Net model

Hyperparameter	**Value/Name**
Optimizer	NAdam [81]
Loss function	Hybrid Loss
Learning rate	0.0001
No. of epochs	50 (With early stopping)

The segmentation results obtained for the SEAU-Net is shown in Table 4.6.

Table 4.6: Segmentation results obtained for SEAU-Net.

Dataset	**DSC**	**IoU**	**RVD**	**Recall**	**Precision**	**Accuracy**
3DIRCADb (liver)	0.9799	0.9606	-0.0096	0.9788	0.9867	0.9773
3DIRCADb (tumor)	0.7315	0.5864	0.0014	0.6405	0.7480	0.9977
CHAOS	0.9832	0.9670	0.0041	0.9834	0.9874	0.9835
LiTS (liver)	0.9794	0.9596	0.0021	0.9809	0.9791	0.3458
LiTS (tumor)	0.9120	0.8407	-0.0968	0.8693	0.9628	0.7854
BUSI	0.7611	0.6216	-0.0022	0.7601	0.7819	0.9530
OCDC	0.9324	0.8735	-0.0111	0.9275	0.9382	0.9356
ISIC 2018	0.8719	0.7780	-0.1115	0.8260	0.9340	0.9320

4.4.1 Ablation Study

Table 4.7 and Table 4.8 show the results obtained using the proposed SEAU-Net architecture.As part of ablation, different modules are removed, and the training process is conducted for each of the variants. It is observed that the model performs significantly better with all of the modules present on different parameters such as DSC score, recall, precision, IoU score, and others. The SEAU-Net with SE attention blocks gives better results than the standard Attention U-Net

which has a grid based attention mechanism. The Attention U-Net as described in [17] comprises of attention applied across all levels within skip connections. For implementing Attention U-Net, keras-unet-collection [82] has been imported, which comprises the TensorFlow implementation of the Attention U-Net with optional ImageNet trained backbone architecture. With roughly 24 million parameters, it is observed that the DSC score obtained from Attention U-Net for both datasets is significantly lower compared to the results obtained using other methods. Further, the TSR block in the decoder block helps to address the vanishing gradient problem especially in the expanding path.

4.4.2 GradCAM Analysis

Further, Figure 4.4 shows the class activation maps from different layers of the proposed model obtained via GradCAM. As observed, the activation maps depicted in the 3rd column are derived from a lower layer compared to the fourth layer. This indicates that the accuracy of the model increases as we move up the layers.

4.4.3 Segmentation Qualitative Analysis

A few good segmentations for the BUSI dataset is shown in Figure 4.5. This highlights the superiority of the model. Also, a few outputs rendered by the SEAU-Net on the ISIC 2018 is shown in Figure 4.6. The obtained segmentation masks are overlaid on the original test images with a transparency of 0.4 using the 'jet' colormap.

Table 4.7: Segmentation results by ablation study on the 3DIRCADb dataset.

Model	**DSC**	**Recall**	**Precision**	**F1-score**	**Accuracy**	**IoU**	**RVD**
Attention U-Net	0.9682	0.9674	0.9746	0.9708	0.9770	0.9387	-0.0084
Without TSR	0.9793	0.9788	0.9856	0.9822	0.9773	0.9594	-0.0084
Without SE	0.9797	0.9782	**0.9871**	0.9826	0.9773	0.9603	-0.0106
SEAU-Net (only Dice Loss)	0.9795	**0.9830**	0.9813	0.9821	0.9773	0.9598	**1.4035e-04**
SEAU-Net (Ours)	**0.9799**	0.9788	0.9867	**0.9827**	**0.9773**	**0.9606**	-0.0096

Table 4.8: Segmentation results by ablation study on the CHAOS dataset.

Model	**DSC**	**Recall**	**Precision**	**F1-score**	**Accuracy**	**IoU**	**RVD**
Attention U-Net	0.9537	0.9223	**0.9919**	0.9554	0.9809	0.9122	-0.0701
Without TSR	0.9816	0.9809	0.9868	0.9838	0.9835	0.9640	-0.0060
Without SE	0.9828	**0.9862**	0.9839	0.9850	0.9835	0.9662	0.0024
SEAU-Net (only Dice Loss)	0.9829	0.9855	0.9848	0.9851	0.9835	0.9665	**6.6643e-04**
SEAU-Net (Ours)	**0.9832**	0.9834	0.9874	**0.9854**	**0.9835**	0.9670	-0.0041

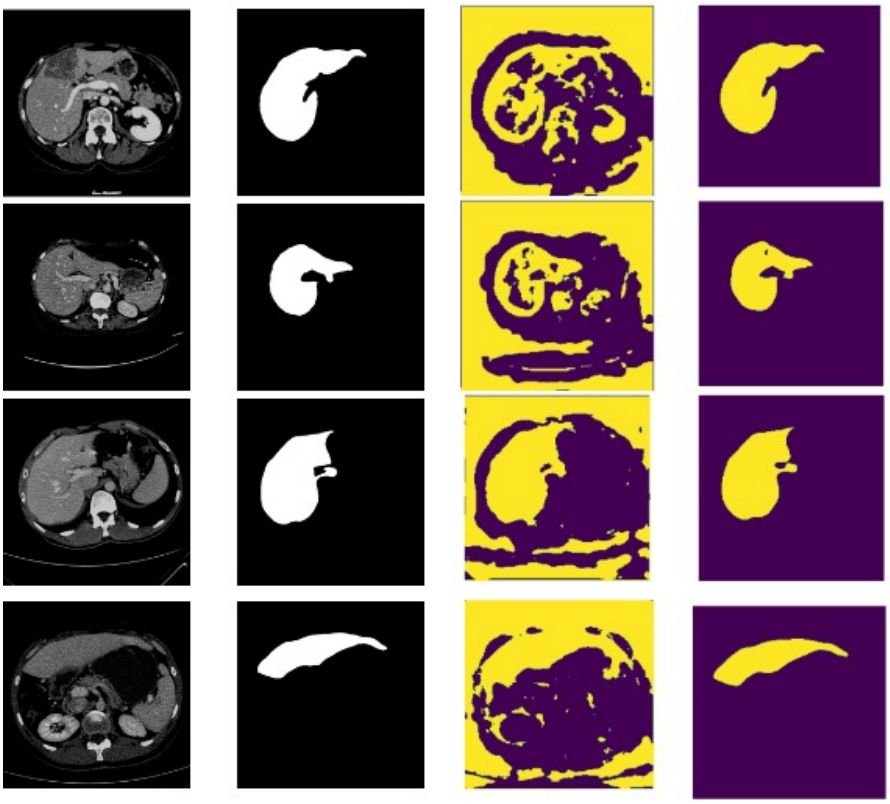


Figure 4.4: Class activation maps for different images using the trained models. The first two columns show the original image and its corresponding mask. The class activation maps are captured at different layers, namely, the *batch_normalization_8* and *conv2d_14*. The first two rows show the results for the 3DIRCADb images and the last two rows show the results for the CHAOS images.

4.5 W^{2}CU-Net

The same hyperparameters have been used throughout all datasets for training the models, as shown in Table 4.9.

Table 4.9: Hyperparameters considered for training the W^{2}CU-Net model.

Hyperparameter	**Value/Name**
Optimizer	Adam [79]
Loss function	Hybrid Loss {(Dice Loss) + 2 * (Focal Loss)}
Learning rate	0.001
No. of epochs	150 (With early stopping)

The segmentation results obtained for the W^{2}CU-Net is shown in Table 4.10.

The resultant loss and dice coefficient curves obtained by the model have been pictorially depicted in Figure 4.7.

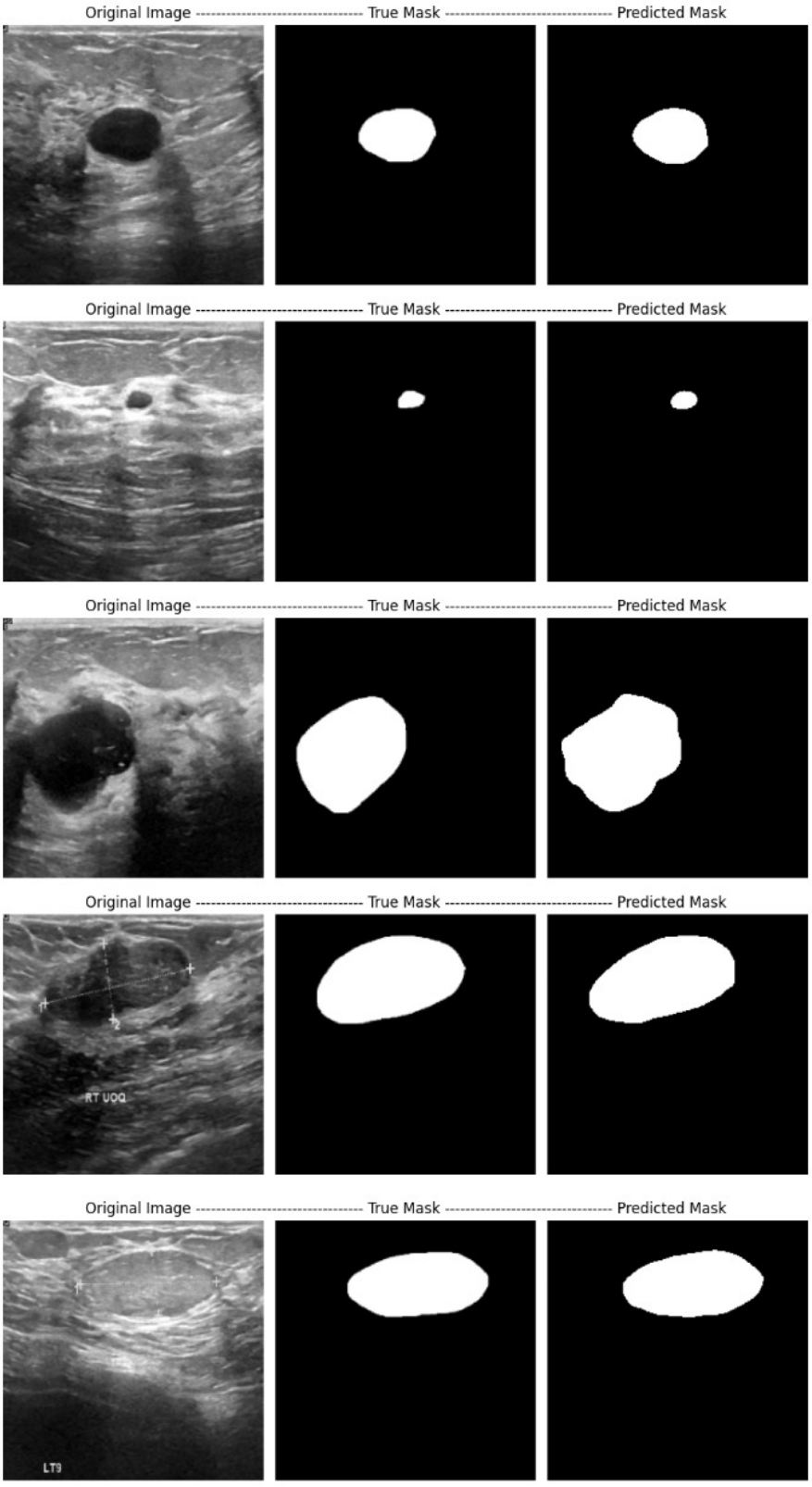


Figure 4.5: A few good segmentation results obtained using SEAU-Net on the BUSI dataset.

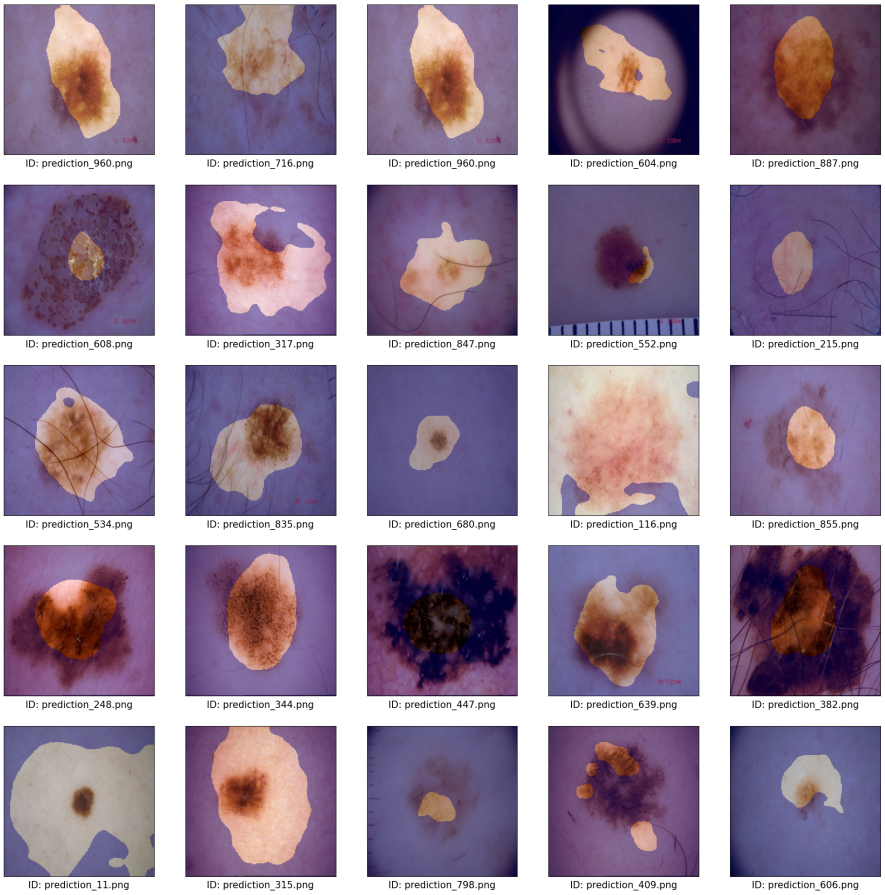


Figure 4.6: A few superimposed segmentation results obtained using the SEAU-Net on the ISIC 2018 dataset.

Table 4.10: Segmentation results obtained for W^{2}CU-Net.

Dataset	**DSC**	**IoU**	**Recall**	**Precision**	**Accuracy**
3DIRCADb (liver)	0.9720	0.9503	0.9930	0.8684	0.9773
3DIRCADb (tumor)	0.8491	0.7464	0.9405	0.5955	0.9973
CHAOS	0.9743	0.9538	0.9954	0.8626	0.9835
LiTS (liver)	0.9837	0.9592	0.9912	0.9561	0.9952
LiTS (tumor)	0.9164	0.8549	0.9783	0.6633	0.9890
BUSI	0.7642	0.6185	0.7447	0.7847	0.9542
OCDC	0.7478	0.5973	0.8260	0.6767	0.7511
ISIC 2018	0.8819	0.7887	0.8675	0.8965	0.9357

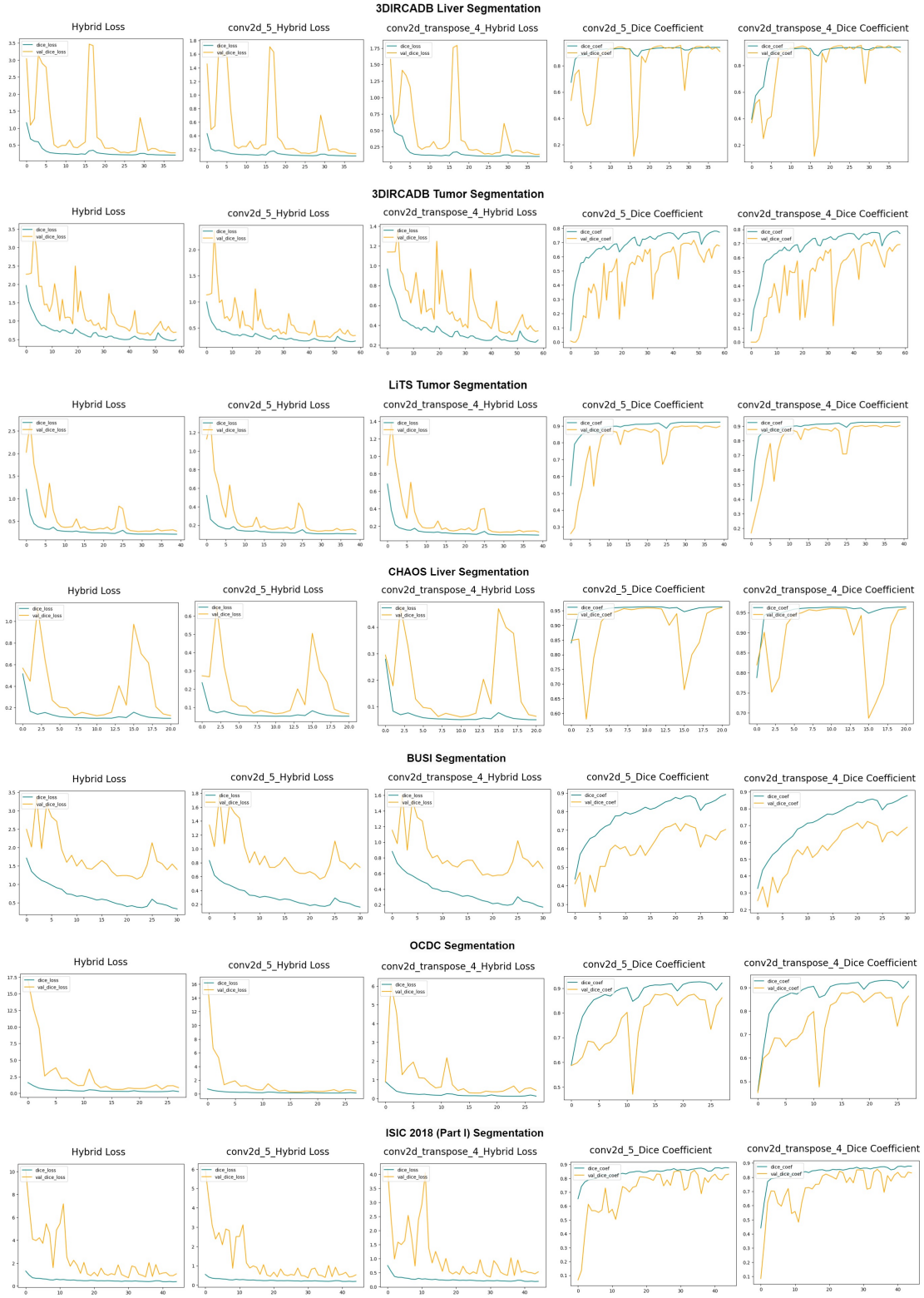


Figure 4.7: Loss and Dice Coefficient curves for the W^{2}CU-Net across the datasets

4.5.1 Ablation Study

This section highlights an ablation study of the W^{2}CU-Net model on 3DIRCADb and LiTS datasets. Table 4.11 and Table 4.12 depict the ablation study on 3DIRCADb liver and tumor segmentation, and the ablation study on LiTS tumor segmentation has been shown in details in Table 4.13 and Table 4.14. In tables the representations are: WL and CL. Here, the various permutations are explored by deactivating these two modules independently to gauge their individual and collective impacts on the system's performance. This systematic analysis aims to elucidate the functional significance and interplay between the modules under investigation.

Table 4.11: Ablation study conducted on the 3DIRCADb dataset for liver segmentation.

Method	DSC	IOU	Accuracy	Precision	Recall
No WL	0.9655	0.9376	0.9770	0.8607	0.9890
No CL	0.9707	0.9476	0.9771	0.8775	0.9885
No WL & CL	0.9320	0.8754	0.9749	0.8752	0.9452
W^{2}CU-Net	**0.9720**	0.9503	0.9773	0.8684	0.9930

Table 4.12: Ablation study conducted on the 3DIRCADb dataset for liver tumor segmentation.

Method	DSC	IOU	Accuracy	Precision	Recall
No WL	0.8366	0.7270	0.9973	0.5519	0.9498
No CL	0.8049	0.6805	0.9971	0.5393	0.9336
No WL & CL	0.7957	0.6573	0.9958	0.5993	0.9104
W^{2}CU-Net	**0.8491**	0.7464	0.9973	0.5955	0.9405

4.5.2 Comparison with other Models

Table 4.15 and Table 4.16 show the results of the comparative study conducted among various models taken from keras-unet-collections [82] respectively. Figure 4.8 show the qualitative results obtained by the proposed model.

Table 4.13: Ablation study conducted on the LiTS dataset for liver segmentation.

Method	DSC	IoU	Accuracy	Precision	Recall
No WL	0.9779	0.9598	0.9953	0.9629	0.9889
No CL	0.9788	0.9636	0.9954	0.9564	0.9929
No WL & CL	0.9785	0.9626	0.9954	0.9576	0.9921
W^{2}CU-Net	**0.9837**	0.9592	0.9952	0.9561	0.9912

Table 4.14: Ablation study conducted on the LiTS dataset for liver tumor segmentation.

Method	DSC	IoU	Accuracy	Precision	Recall
No WL	0.9105	0.8775	0.9979	0.8898	0.9607
No CL	0.9124	0.8481	0.9889	0.6536	0.9797
No WL & CL	0.9086	0.8540	0.9977	0.8906	0.9417
W^{2}CU-Net	**0.9227**	0.8639	0.9978	0.8779	0.9569

Table 4.15: Quantitative analysis on the 3DIRCADb dataset for the segmentation of liver.

Model	DSC	IOU	Accuracy	Precision	Recall
U-Net [3]	0.8610	0.7963	0.3630	0.9480	0.7915
U-Net++ [83]	0.9257	0.7975	0.4522	0.9880	0.7377
V-Net [84]	0.9396	0.8010	0.6336	0.9942	0.7449
R2U-Net [85]	0.8527	0.7975	0.4841	0.9707	0.8522
W^{2}CU-Net	**0.9720**	0.9503	0.9773	0.8684	0.9930

Table 4.16: Quantitative analysis on the 3DIRCADb dataset for the segmentation of liver tumor.

Model	DSC	IOU	Accuracy	Precision	Recall
U-Net [3]	0.5063	0.4463	0.0632	0.6930	0.3815
U-Net++ [83]	0.5732	0.5037	0.8148	0.6376	0.5002
V-Net [84]	0.5603	0.4937	0.0757	0.9572	0.3328
R2U-Net [85]	0.5363	0.4835	0.0662	0.7339	0.3659
W^{2}CU-Net	**0.8491**	0.7464	0.9973	0.5955	0.9405

Table 4.17: Quantitative analysis on the LiTS dataset for the segmentation of liver.

Model	DSC	IoU	Accuracy	Precision	Recall
U-Net [3]	0.9781	0.7359	0.3592	0.9903	0.9714
U-Net++ [83]	0.9597	0.6765	0.9402	0.9734	0.9519
V-Net [84]	0.9766	0.6930	0.1049	0.9888	0.9703
R2U-Net [85]	0.9798	0.6765	0.6359	0.9908	0.9742
W^{2}CU-Net	**0.9837**	0.9592	0.9952	0.9561	0.9912

Table 4.18: Quantitative analysis on the LiTS dataset for the segmentation of liver tumor.

Model	DSC	IoU	Accuracy	Precision	Recall
U-Net [3]	0.8244	0.9391	0.2405	0.9918	0.6780
U-Net++ [83]	0.8538	0.8657	0.1896	0.9481	0.7077
V-Net [84]	0.8691	0.9437	0.2811	0.9906	0.6957
R2U-Net [85]	0.8746	0.8477	0.0059	0.9896	0.6969
W^{2}CU-Net	**0.9164**	0.8549	0.9890	0.6633	0.9783

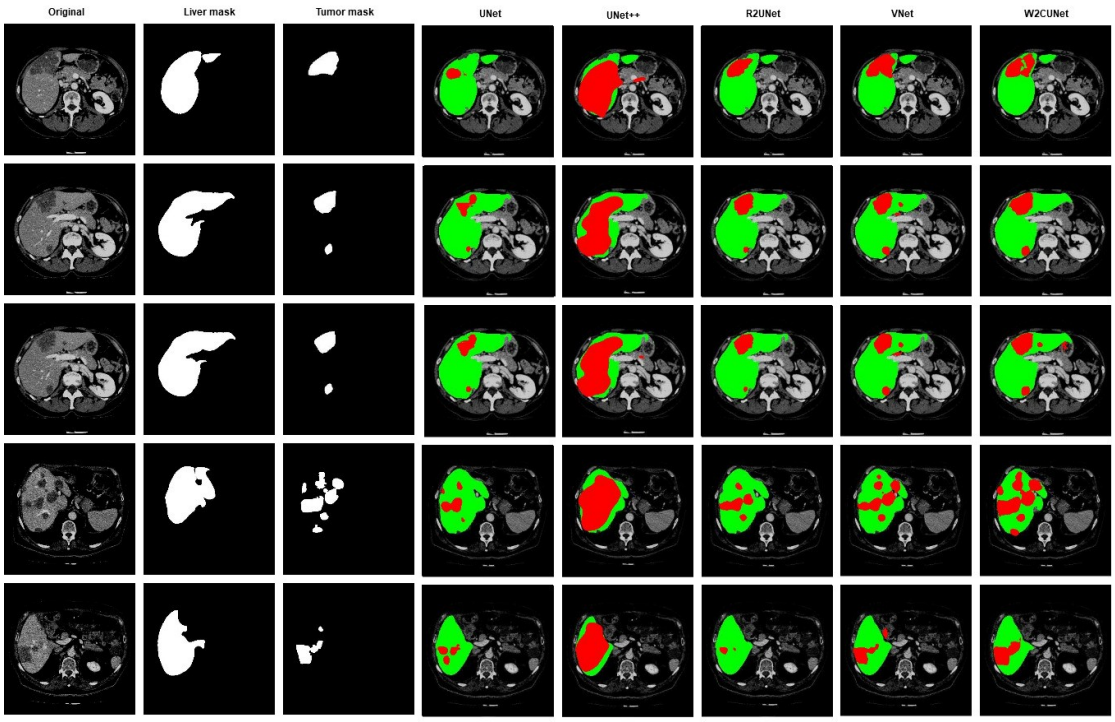


Figure 4.8: Qualitative analysis presented for different models on the 3DIRCADb dataset.

4.5.3 t-SNE Plots

As per [86], t-SNE is a notable dimensionality reduction method for visualizing high-dimensional data in a lower-dimensional space. It transforms Euclidean distances between data points into conditional probability scores using SNE. Figure 4.9 illustrates the pixel-wise t-SNE plot of two decoders on 3DIRCADb and LiTS datasets.

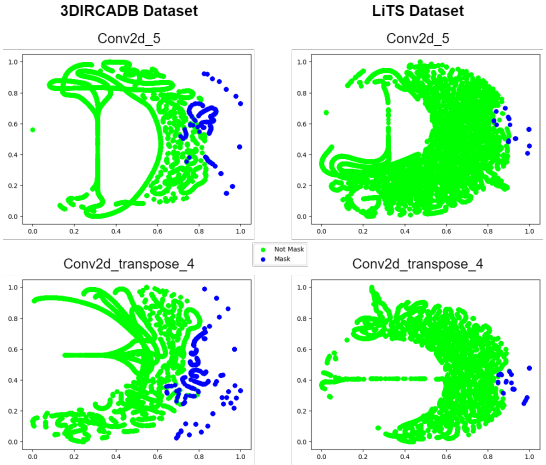


Figure 4.9: Pixel-wise t-SNE for liver and tumor segmentation on 3DIRCADb and LiTS dataset for both decoders.

4.5.4 IoU Heatmaps

An IoU heatmap visually represents the accuracy of a model's predictions in computer vision tasks like object detection. It assigns colors based on IoU values, where brighter colors indicate higher overlap between predicted and ground truth bounding boxes, highlighting accurate predictions. The heatmap, exemplified in Figure 4.10, serves as a valuable tool for assessing and enhancing model performance.

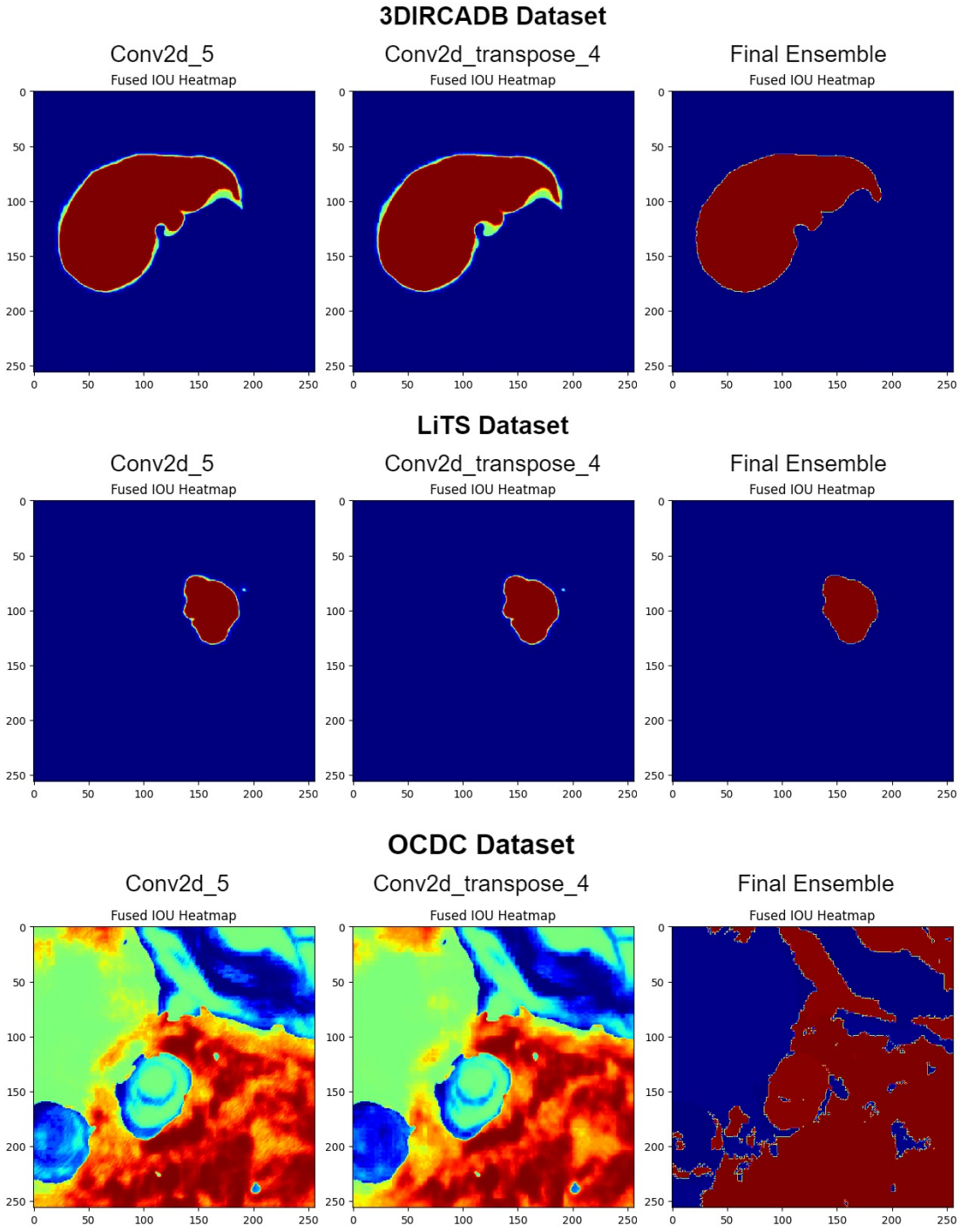


Figure 4.10: IOU heatmaps for liver segmentation on 3DIRCADb and LiTS and tumorous cell segmentation on the OCDC dataset for both decoders and final ensemble

CHAPTER 5

Conclusion

This thesis presents novel DL approaches tailored for medical image segmentation across diverse datasets encompassing liver, skin, oral, and breast cancers. Through rigorous experimentation and validation, these methodologies demonstrate remarkable efficacy in accurately delineating anatomical structures and pathological regions, thereby contributing significantly to the field of medical imaging analysis.

By addressing the challenges posed by varying image modalities and complex tissue interactions, these innovations pave the way for enhanced diagnostic precision, treatment planning, and patient care. Moving forward, the integration of these DL techniques promises to revolutionize medical image analysis, offering clinicians invaluable tools to combat cancer and improve healthcare outcomes on a global scale. Better results are obtained in comparison to the SoTA for most of the datasets.

The BLUE-Net combines the decisions of three variants of U-Net models, namely the basic U-Net, Link-Net, and FPN. A fuzzy-rank based method is designed by using the Blumberg function in order to combine the decisions of the said models.For the experimental evaluation, the 3DIRCADb liver segmentation dataset is considered, and the method outperforms many recently proposed methods. DSC scores of 0.9774, 0.8359, 0.9750, 0.9775, 0.9144 and 0.8076 are obtained for 3DIRCADb liver segmentation, 3DIRCADb tumor segmentation, CHAOS, LiTS liver segmentation, LiTS tumor segmentation, and OCDC dataset respectively.

The SEAU-Net efficiently integrates SE attention blocks within the skip connections of a U-Net architecture, by selectively employing attention to the lower skip connections. These attention aided feature-maps from the lower layers to the higher layers in the decoder ensure high-level semantic features. DSC scores of 0.9799, 0.9832, 0.9794, 0.9120 and 0.9324 for 3DIRCADb liver segmentation,

CHAOS, LiTS liver segmentation, LiTS tumor segmentation, and OCDC dataset respectively.

The W^2CU-Net model employs a cosine similarity and Weibull function-based layers. It combines decisions from two different decoders through an ensemble method utilizing Weierstrass' Elliptic curve equations. DSC scores of 0.9720, 0.8491, 0.9743, 0.9837 and 0.9164 3DIRCADb liver segmentation, 3DIRCADb tumor segmentation, CHAOS, LiTS liver segmentation, and LiTS tumor segmentation respectively.

5.1 Limitations

While the mentioned methods—BLUE-Net, SEAU-Net, and W^{2}CU-Net—offer innovative approaches to medical image segmentation, they still have some limitations:

- **BLUE-Net:** Despite leveraging ensemble learning, BLUE-Net's DSC score may still be affected by inconsistent or conflicting predictions from the base models, particularly in areas with complex anatomical structures or ambiguous boundaries. As seen for the BUSI dataset, the BLUE-Net struggles to handle problems related to ultrasound images and class imbalance effectively, as they rely on combining predictions from multiple models, which may inherit biases or errors from individual models.

- **SEAU-Net:** While SE attention blocks in SEAU-Net aim to enhance feature utilization and focus on relevant areas, they may inadvertently introduce noise or artifacts into the segmentation masks, particularly in regions with complex or irregular shapes, as observed for the 3DIRCADb dataset for tumor segmentation. The computational overhead introduced by SE attention blocks within skip connections and the heavy ResNet50 backbone in itself may impact real-time segmentation applications or require additional optimization to maintain efficiency without sacrificing DSC.

- **W^{2}CU-Net:** The modifications in W^{2}CU-Net, such as the Weibull layer and feature pruning based on cosine similarity, may not always generalize well, as observed for the OCDC dataset, leading to suboptimal segmentation performance in certain scenarios. The fusion technique using Weierstrass' elliptic curve equation in W^{2}CU-Net may introduce additional complexity to the fusion process, potentially leading to challenges in effectively integrating information from multiple decoders, especially in cases where complementary information is not well-aligned or consistent.

5.2 Scope for Future Work

However, in the medical field, to apply a model to real-life applications, it should be perfect, i.e., there is no scope for error. Hence, research can be done to continue improvement of the performance of the model. Exploration of other U-Net variants to capture more varied information and the utilization of different loss functions to train models efficiently can be pursued. To overcome the problem of limited annotated data in the medical field, the use of GAN for creating synthetic images is also considered.

As the number of parameters that the SEAU-Net bears is around 45M, making it bit heavy by nature. Hence, the aim can be to develop lighter-weight models, thereby making it deployable in resource-constrained environments. Acknowledging the imperfections inherent in medical applications, the researchers express the need for continuous improvement. Future research directions include exploring other U-Net variants, employing diverse loss functions for efficient training, etc.

List of Publications

Accepted for Publication

1. S. Majumder, A. Sau, **A. Halder**, P. Saha, R. Sarkar, "BLUE-Net: BLUmberg function based Ensemble Network for Liver and Tumor Segmentation from CT scans", 2nd International Conference on Data Electronics and Computing (ICDEC-2023), December 15-16, 2023, Mizoram University, Aizawl, Mizoram. **Best Paper**

Communicated (under review)

1. **A. Halder**, S. Singh, M. Kundu, R. Sarkar, "SEAU-Net: A Squeeze-and-Excitation aided Attention U-Net for Liver Segmentation", Evolving Systems.

2. S. Majumder, **A. Halder**, S. Singh, M. Kundu, R. Sarkar, "W^2CU-Net: A Weibull and Weierstrass functions with Cosine Similarity based Network for Liver and Tumor Segmentation from CT Scans", IEEE Transactions on Artificial Intelligence.

3. **A. Halder**, A. Sau, R.Sarkar, "Computer-based Liver and Tumor Segmentation: From Past Endeavors to Future Research Challenges", Computers In Biology and Medicine.

4. **A. Halder**, A. Sau, S. Majumder, D. Kaplun, R. Sarkar, "An Experimental Study of U-Net Variants on Liver Segmentation from CT Scans", Journal of Intelligent Systems.

Bibliography

[1] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[2] H.-C. Shin, H. R. Roth, M. Gao, *et al.*, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[3] O. Ronneberger, P. Fischer, and T. Brox, "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, pp. 234–241.

[4] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE visual communications and image processing (VCIP)*, IEEE, 2017, pp. 1–4.

[5] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[6] H. Liu, Y. Fu, S. Zhang, *et al.*, "Gcha-net: Global context and hybrid attention network for automatic liver segmentation," *Computers in Biology and Medicine*, vol. 152, p. 106 352, 2023.

[7] A. E. Kavur, L. I. Kuncheva, and M. A. Selver, "Basic ensembles of vanilla-style deep learning models improve liver segmentation from ct images," in *Convolutional Neural Networks for Medical Image Processing Applications*, CRC Press, 2022, pp. 52–74.

[8] K. Kamnitsas, E. Ferrante, S. Parisot, *et al.*, "Deepmedic for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction*

with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 2, Springer, 2016, pp. 138–149.

[9] A. Abdollahi, B. Pradhan, and A. Alamri, "Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179 424–179 436, 2020.

[10] E. Gibson, F. Giganti, Y. Hu, *et al.*, "Automatic multi-organ segmentation on abdominal ct with dense v-networks," *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.

[11] D. T. Kushnure and S. N. Talbar, "Hfru-net: High-level feature fusion and recalibration unet for automatic liver and tumor segmentation in ct images," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106 501, 2022.

[12] Y. Chen, C. Zheng, F. Hu, *et al.*, "Efficient two-step liver and tumour segmentation on abdominal ct via deep learning and a conditional random field," *Computers in Biology and Medicine*, vol. 150, p. 106 076, 2022.

[13] S. Di, Y.-Q. Zhao, M. Liao, F. Zhang, and X. Li, "Td-net: A hybrid end-to-end network for automatic liver tumor segmentation from ct images," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1163–1172, 2022.

[14] J. Jiang, Y. Peng, Q. Hou, and J. Wang, "Mdcf_net: A multi-dimensional hybrid network for liver and tumor segmentation from ct," *Biocybernetics and Biomedical Engineering*, 2023.

[15] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[16] L. Jiang, J. Ou, R. Liu, *et al.*, "Rmau-net: Residual multi-scale attention u-net for liver and tumor segmentation in ct images," *Computers in Biology and Medicine*, vol. 158, p. 106 838, 2023.

[17] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[18] K. Hettihewa, T. Kobchaisawat, N. Tanpowpong, and T. H. Chalidabhongse, "Manet: A multi-attention network for automatic liver tumor segmentation in computed tomography (ct) imaging," *Scientific Reports*, vol. 13, no. 1, p. 20 098, 2023.

[19] J. Zhu, Z. Liu, W. Gao, and Y. Fu, "Cotepres-net: An efficient u-net based deep learning method of liver segmentation from computed tomography images," *Biomedical Signal Processing and Control*, vol. 88, p. 105 660, 2024.

[20] J. Wu, S. Zhou, S. Zuo, *et al.*, "U-net combined with multi-scale attention mechanism for liver segmentation in ct images," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–12, 2021.

[21] T. Fan, G. Wang, Y. Li, and H. Wang, "Ma-net: A multi-scale attention network for liver and tumor segmentation," *IEEE Access*, vol. 8, pp. 179 656–179 665, 2020.

[22] C. Li, Y. Tan, W. Chen, *et al.*, "Anu-net: Attention-based nested u-net to exploit full resolution features for medical image segmentation," *Computers & Graphics*, vol. 90, pp. 11–20, 2020.

[23] Y. Chen, C. Zheng, T. Zhou, *et al.*, "A deep residual attention-based u-net with a biplane joint method for liver segmentation from ct scans," *Computers in Biology and Medicine*, vol. 152, p. 106 421, 2023.

[24] Y. A. Ayalew, K. A. Fante, and M. A. Mohammed, "Modified u-net for liver cancer segmentation from computed tomography images with a new class balancing method," *BMC Biomedical Engineering*, vol. 3, pp. 1–13, 2021.

[25] Z. Xue, P. Li, L. Zhang, *et al.*, "Multi-modal co-learning for liver lesion segmentation on pet-ct images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3531–3542, 2021.

[26] L. Han, Y. Chen, J. Li, B. Zhong, Y. Lei, and M. Sun, "Liver segmentation with 2.5 d perpendicular unets," *Computers & Electrical Engineering*, vol. 91, p. 107 118, 2021.

[27] J. D. L. Araújo, L. B. da Cruz, J. O. B. Diniz, *et al.*, "Liver segmentation from computed tomography images using cascade deep learning," *Computers in biology and medicine*, vol. 140, p. 105 095, 2022.

[28] A. S. Podda, R. Balia, S. Barra, S. Carta, G. Fenu, and L. Piano, "Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images," *Journal of computational science*, vol. 63, p. 101 816, 2022.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[30] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Applied Sciences*, vol. 12, no. 18, p. 8972, 2022.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[33] M. S. Hossain, "Microc alcification segmentation using modified u-net segmentation network from mammogram images," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 2, pp. 86–94, 2022.

[34] F. J. Pérez-Benito, F. Signol, J.-C. Perez-Cortes, *et al.*, "A deep learning system to obtain the optimal parameters for a threshold-based breast and dense tissue segmentation," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105 668, 2020.

[35] Q. Wang, H. Chen, G. Luo, *et al.*, "Performance of novel deep learning network with the incorporation of the automatic segmentation network for diagnosis of breast cancer in automated breast ultrasound," *European radiology*, vol. 32, no. 10, pp. 7163–7172, 2022.

[36] P. Pramanik, R. Pramanik, F. Schwenker, and R. Sarkar, "Dbu-net: Dual branch u-net for tumor segmentation in breast ultrasound images," *Plos one*, vol. 18, no. 11, e0293615, 2023.

[37] Y. Luo, Q. Huang, and X. Li, "Segmentation information with attention integration for classification of breast tumor in ultrasound image," *Pattern Recognition*, vol. 124, p. 108 427, 2022.

[38] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional convlstm u-net with densley connected convolutions," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0.

[39] A. Albishri, S. J. H. Shah, Y. Lee, and R. Wang, "Ocu-net: A novel u-net architecture for enhanced oral cancer segmentation," *arXiv preprint arXiv:2310.02486*, 2023.

[40] A. Pennisi, D. D. Bloisi, D. Nardi, S. Varricchio, and F. Merolla, "Multi-encoder u-net for oral squamous cell carcinoma image segmentation," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2022, pp. 1–6.

[41] S. P. Sunny, A. I. Khan, M. Rangarajan, *et al.*, "Oral epithelial cell segmentation from fluorescent multichannel cytology images using deep learning," *Computer methods and programs in biomedicine*, vol. 227, p. 107 205, 2022.

[42] B. Koonce and B. Koonce, "Resnet 34," *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, pp. 51–61, 2021.

[43] X. Zhang, Q. Li, W. Li, *et al.*, "Fd-net: Feature distillation network for oral squamous cell carcinoma lymph node segmentation in hyperspectral imagery," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[44] A. J. Shephard, S. Graham, S. Bashir, *et al.*, "Simultaneous nuclear instance and layer segmentation in oral epithelial dysplasia," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 552–561.

[45] D. F. dos Santos, P. R. de Faria, B. A. Travencolo, and M. Z. do Nascimento, "Automated detection of tumor regions from oral histological whole slide images using fully convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 69, p. 102 921, 2021.

[46] S.-Y. Huang, C.-Y. Chiou, Y.-S. Tan, C.-Y. Chen, and P.-C. Chung, "Deep oral cancer lesion segmentation with heterogeneous features," in *2022 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, IEEE, 2022, pp. 1–8.

[47] A. V. Matias, A. Cerentini, L. A. B. Macarini, J. G. A. Amorim, F. P. Daltoé, and A. von Wangenheim, "Segmentation, detection, and classification of cell nuclei on oral cytology samples stained with papanicolaou," *SN Computer Science*, vol. 2, no. 4, p. 285, 2021.

[48] J. Folmsbee, L. Zhang, X. Lu, *et al.*, "Histology segmentation using active learning on regions of interest in oral cavity squamous cell carcinoma," *Journal of Pathology Informatics*, vol. 13, p. 100 146, 2022.

[49] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, "Fat-net: Feature adaptive transformers for automated skin lesion segmentation," *Medical image analysis*, vol. 76, p. 102 327, 2022.

[50] K. Hu, J. Lu, D. Lee, D. Xiong, and Z. Chen, "As-net: Attention synergy network for skin lesion segmentation," *Expert Systems with Applications*, vol. 201, p. 117 112, 2022.

[51] H. Basak, R. Kundu, and R. Sarkar, "Mfsnet: A multi focus segmentation network for skin lesion segmentation," *Pattern Recognition*, vol. 128, p. 108 673, 2022.

[52] D. Dai, C. Dong, S. Xu, *et al.*, "Ms red: A novel multi-scale residual encoding and decoding network for skin lesion segmentation," *Medical image analysis*, vol. 75, p. 102 293, 2022.

[53] Y. Jiang, H. Qiao, Z. Zhang, M. Wang, W. Yan, and J. Chen, "Mdsc-net: A multi-scale depthwise separable convolutional neural network for skin lesion segmentation," *IET Image Processing*, vol. 17, no. 13, pp. 3713–3727, 2023.

[54] Z. Yu, L. Yu, W. Zheng, and S. Wang, "Eiu-net: Enhanced feature extraction and improved skip connections in u-net for skin lesion segmentation," *Computers in Biology and Medicine*, p. 107 081, 2023.

[55] Q. Ma, K. Mao, G. Wang, L. Xu, and Y. Zhao, "Lcaunet: A skin lesion segmentation network with enhanced edge and body fusion," *arXiv preprint arXiv:2305.00837*, 2023.

[56] E. M. Elmeslimany, S. S. Kishk, and D. A. Altantawy, "ΨNet: A parallel network with deeply coupled spatial and squeezed features for segmentation of medical images," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 24 045–24 082, 2024.

[57] *Liver segmentation 3d-ircadb-01*, `https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/`, 2010.

[58] L. Soler, A. Hostettler, V. Agnus, *et al.*, "3d image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database," *IRCAD, Strasbourg, France, Tech. Rep*, vol. 1, no. 1, 2010.

[59] Y. D. Won, M. K. Na, C. H. Kim, *et al.*, "The frontal skull hounsfield unit value can predict ventricular enlargement in patients with subarachnoid haemorrhage," *Scientific Reports*, vol. 8, no. 1, p. 10 178, 2018.

[60] H. Ishihara, F. Oka, R. Kawano, *et al.*, "Hounsfield unit value of interpeduncular cistern hematomas can predict symptomatic vasospasm," *Stroke*, vol. 51, no. 1, pp. 143–148, 2020.

[61] T. D. DenOtter and J. Schubert, "Hounsfield unit," 2019.

[62] J. J. Schreiber, P. A. Anderson, H. G. Rosas, A. L. Buchholz, and A. G. Au, "Hounsfield units for assessing bone mineral density and strength: A tool for osteoporosis management," *JBJS*, vol. 93, no. 11, pp. 1057–1063, 2011.

[63] R. Molteni, "Prospects and challenges of rendering tissue density in hounsfield units for cone beam computed tomography," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 116, no. 1, pp. 105–119, 2013.

[64] V. Thomsen, D. Schatzlein, and D. Mercuro, "Tutorial: Attenuation of x-rays by matter," *Spectroscopy*, vol. 20, no. 9, 2005.

[65] J. J. DeMarco and P. Suortti, "Effect of scattering on the attenuation of x rays," *Physical Review B*, vol. 4, no. 4, p. 1028, 1971.

[66] V. Artem'ev, "Attenuation of x rays by ultradisperse media," *Technical Physics Letters*, vol. 23, pp. 212–213, 1997.

[67] A. E. Kavur, N. S. Gezer, M. Barış, *et al.*, "Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation," *Medical Image Analysis*, vol. 69, p. 101 950, 2021.

[68] *Chaos - combined (ct-mr) healthy abdominal organ segmentation*, `https://chaos.grand-challenge.org/`, 2019.

[69] P. Bilic, P. Christ, H. B. Li, *et al.*, "The liver tumor segmentation benchmark (lits)," *Medical Image Analysis*, vol. 84, p. 102 680, 2023.

[70] I. 2017 and M. 2017, *Liver and liver tumor segmentation*, `https://www.kaggle.com/datasets/andrewmvd/lits-png`, 2017.

[71] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104 863, 2020.

[72] P. Iakubovskii, *Segmentation models*, `https://github.com/qubvel/segmentation_models`, 2019.

[73] S. Ghosal, M. Sarkar, and R. Sarkar, "Nofed-net: Nonlinear fuzzy ensemble of deep neural networks for human activity recognition," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 526–17 535, 2022.

[74] J. L. Rocha, D. Fournier-Prunaret, and A.-K. Taha, "Big bang bifurcations and allee effect in blumberg's dynamics," *Nonlinear Dynamics*, vol. 77, pp. 1749–1771, 2014.

[75] R. Kundu, H. Basak, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar, "Fuzzy rank-based fusion of cnn models using gompertz function for screening covid-19 ct-scans," *Scientific reports*, vol. 11, no. 1, p. 14 133, 2021.

[76] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[77] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[78] R. A. Horn, "The hadamard product," in *Proc. Symp. Appl. Math*, vol. 40, 1990, pp. 87–169.

[79] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[80] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[81] A. Tato and R. Nkambou, "Improving adam optimizer," 2018.

[82] Y. Sha, *Keras-unet-collection*, `https://github.com/yingkaisha/keras-unet-collection`, 2021. DOI: `10.5281/zenodo.5449801`.

[83] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "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, pp. 3–11.

[84] F. Milletari, N. Navab, and S.-A. Ahmadi, *V-net: Fully convolutional neural networks for volumetric medical image segmentation*, 2016.

[85] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, *Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation*, 2018.

[86] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.