2.5D Dual-Encoder U-Net for Lesion Segmentation in Chest CT Scans

A dissertation submitted in
partial fulfilment for the degree of

Master of Technology

in

Computer Science

by

Jagannath Mukkara
Roll no. - **CS2308**

under the supervision of

Dr. Sarbani Palit
Computer Vision and Pattern Recognition Unit (CVPRU)

Indian Statistical Institute, Kolkata
June, 2025

Acknowledgement

I extend my sincere appreciation to Dr. Sarbani Palit, my advisor at the Computer Vision and Pattern Recognition Unit of the Indian Statistical Institute in Kolkata, for her guidance, continuous support, and inspiration. Her profound knowledge and creative suggestions have taught me a great deal in every subject and have shown me how to conduct solid research.

I would like to sincerely thank Somrita Bakshi, Senior Research Fellow at the Indian Statistical Institute, for her invaluable assistance in gathering the datasets essential for this research. Her consistent provision of ideas and unwavering support have been instrumental in the success of this project.

I am deeply grateful to all the teachers at the Indian Statistical Institute for their invaluable advice, insights, and instruction, which provided a crucial perspective to my research.

Finally, I want to express my gratitude to my parents and extended family for their unwavering support. I also extend my sincere appreciation to all my friends for their continuous assistance and encouragement. I am thankful to everyone who has contributed to my growth and success, even if I have inadvertently missed mentioning them on the above list.

CERTIFICATE

This is to certify that the dissertation entitled "**2.5D Dual-Encoder U-Net for Lesion Segmentation in Chest CT**" submitted by **Jagannath Mukkara** to the Indian Statistical Institute, Kolkata, in partial fulfillment of the requirements for the degree of Master of Technology in Computer Science. is an authentic and genuine record of the research work carried out by the candidate under my supervision and guidance. I affirm that the dissertation has met all the necessary requirements in accordance with the regulations of this institute.

Sarbani Palit 11.6.2025

Dr. Sarbani Palit
CVPR Unit
Indian Statistical Institute
Kolkata - 700108
India

Declaration

I, **Jagannath Mukkara**, with Roll No. **CS2308**, hereby declare that the material presented in the dissertation titled **2.5D Dual-Encoder U-Net for Lesion Segmentation in Chest CT** represents original work carried out by me for the degree of **Master of Technology in Computer Science** at the **Indian Statistical Institute, Kolkata**.

Furthermore, I affirm that no sections of this report have been sourced or copied from external references without proper attribution. I am aware that any instances of plagiarism or the use of unacknowledged materials from third parties will be treated with the utmost seriousness and consequences.

Jagannath Mukkara
M.Tech (CS), CS2308
Indian Statistical Institute

Abstract

Accurate segmentation of lesions in chest CT scans plays a vital role in diagnosing and monitoring pulmonary diseases such as COVID-19. In this, we introduce a novel 2.5D[1] dual-encoder U-Net model[2] that utilizes both the central slice and its neighboring slices to improve segmentation accuracy while keeping computational demands manageable. Our model incorporates residual connections[3] and feature fusion[4] to effectively merge multi-slice contextual information, overcoming the limitations found in traditional 2D and 3D methods. To ensure a reliable evaluation and avoid data leakage, we used patient-level data splitting. We validate our approach on a carefully curated chest CT dataset, showing enhanced segmentation performance and better generalization compared to standard U-Net models. Through extensive experiments, including ablation studies and visualizations, we demonstrate the advantages of combining 2.5D learning with a dual-encoder architecture for medical image segmentation tasks.

Keywords: 2.5D Learning, Dual-Encoder U-Net, Medical Image Analysis, Covid-19, convolutional neural network (CNN)[5], Feature Fusion Multi Slice Context

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Medical imaging plays a vital role in diagnosing. Among the available imaging techniques, CT (computed tomography) scans are particularly useful for examining lung structure and spotting issues like infections, lesions, or tumors. Accurately segmenting these lesions is crucial for quantifying disease severity, planning treatments, and tracking progress. But manually outlining lesions is slow, varies from one radiologist to another, and simply isn't scalable - which makes reliable automated segmentation tools more important than ever.

In recent years, there has been a huge increase in medical image analysis via deep learning techniques, with convolutional neural networks (CNNs) leading the charge. U-Net, in particular, has become a go-to architecture for biomedical segmentation thanks to its encoder–decoder layout and skip connections. Still, traditional 2D U-Nets often fall short with 3D data like CT scans, since they process each slice on its own and miss out on the context from adjacent slices Fully 3D models can capture this context but tend to be heavy on resources and need lots of annotated data.

To find a middle ground, we take a 2.5D approach — a practical middle ground that brings in information from neighboring slices without making the model unwieldy. Specifically, we introduce a dual-encoder U-Net that processes the central slice and its surrounding slices through parallel encoding branches. These features are then fused using a residual fusion block, helping the model better understand local context and boost segmentation accuracy.

This dissertation is about building and testing a deep learning solution that's not just accurate, but also efficient and generalizable for lesion segmentation in chest CT scans. By carefully curating the dataset, designing thoughtful experiments, and digging deep into the results, we wish to offer insights and practical advances for automated medical image segmentation.

Chapter 2

Background and Related Work

Medical image segmentation is a foundational task in computer-aided diagnosis. In the context of chest CT scans, segmentation of pulmonary lesions is particularly important for quantifying disease burden, guiding treatment planning, and monitoring disease progression. Traditional segmentation tasks are challenged by the complexity of medical images, including variations in patient anatomy, lesion appearance, and imaging artifacts, making robust and accurate segmentation both clinically valuable and technically demanding.

The use of deep learning has revolutionized medical imaging, with convolutional neural networks (CNNs) achieving state-of-the-art performance in tasks such as classification, detection, and segmentation. Classical methods rely on handcrafted features, deep learning models can automatically learn hierarchical representations directly from raw image data, leading to significant improvements in accuracy and robustness.

The U-Net[6] architecture, introduced by Ronneberger et al., has become the backbone of biomedical image segmentation due to its encoder–decoder structure and skip connections, which facilitate the preservation of spatial information. Over time, numerous variants have been developed to address specific challenges. U-Net++[7] employs nested skip pathways for improved feature fusion, while Attention U-Net[8] introduces attention mechanisms to focus on relevant regions. Despite their strengths, most U-Net-based models process 2D slices independently, limiting their ability to capture the full spatial context present in volumetric data like CT scans.

Dual-encoder architectures have emerged as promising solutions for integrating diverse sources of information in medical image segmentation. They can separately process the central slice and its neighboring slices, enabling the network to capture both detailed local features and broader contextual cues, which are critical for accurate delineation of complex lesions.

Chapter 3

Dataset and Preprocessing

We used a curated dataset of chest CT scans paired with annotated lesion masks from a challenge-COVID-19 Lung CT Lesion Segmentation Challenge - 2020[9], hosted on Grand Challenge platform, enabling supervised learning for pulmonary lesion segmentation. This dataset has data of 199 Covid 19 positive patients. This dataset is stored in NIfTI format, which preserves volumetric information across slices and facilitates integration into deep learning pipelines.

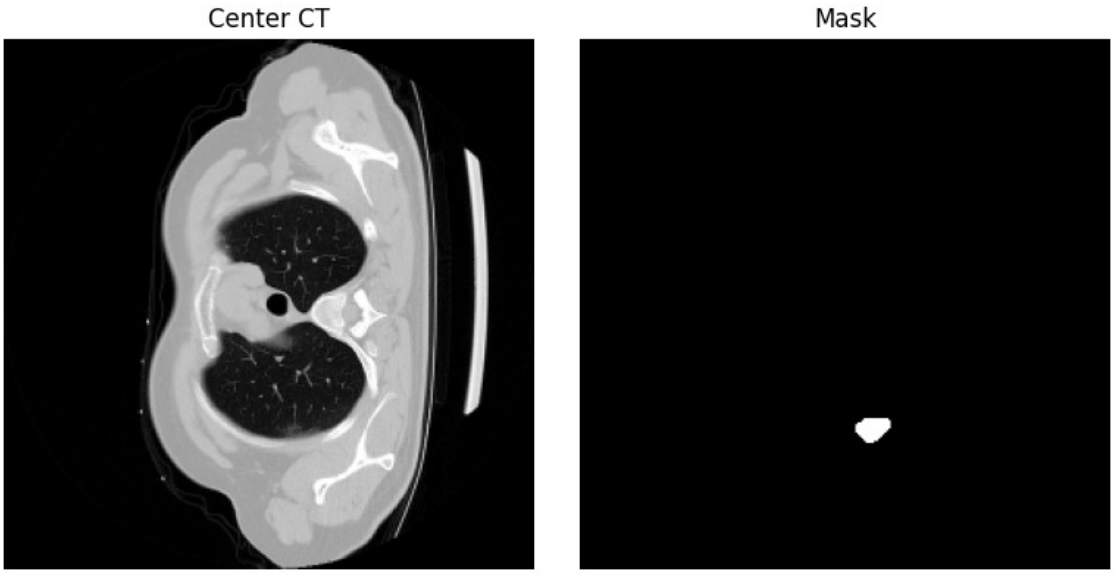

Figure 3.1: Axial slice of CT scan and corresponding mask of a patient

Preprocessing is essential to standardize the data and enhance model performance. Each CT scan is normalized to a consistent Hounsfield Unit (HU) window (-1000 to 400) to focus on lung tissue and suppress irrelevant background. Slices are then resized to a uniform spatial resolution (original resolution: 512×512, resized to: 256×256) to ensure compatibility with the network input. For 2.5D learning, stack of 3 adjacent slices are grouped as multi-channel inputs, allowing the model to

leverage spatial context across neighboring slices while maintaining computational efficiency.

To ensure robust evaluation and prevent data leakage, the dataset is split at the patient level into distinct training and test sets meaning, no slices from the same patient appear in both sets, preserving the integrity of model assessment and providing a realistic measure of generalization. 80/20 split is used. We also tried to address the challenges posed by class imbalance, which is critical for clinical deployment.

Chapter 4

Methodology

The proposed model employs a dual-encoder U-Net architecture specifically designed for 2.5D lesion segmentation in chest CT scans. Unlike traditional U-Net models that process single slices, this architecture incorporates two parallel encoding pathways: Encoder A processes the center slice (single-channel input) to capture fine-grained local features, Encoder B processes a stack of three adjacent slices (multi-channel input) to extract contextual information from neighboring anatomy. The encoded features from both pathways are combined through a fusion mechanism before being passed to a shared decoder with skip connections. This design leverages the benefits of both local detail preservation and spatial context awareness while maintaining computational efficiency compared to full 3D approaches.

Image: architecture overview

4.1 Architecture

The architecture incorporates residual convolutional blocks (ResidualConvBlock) throughout both encoders to facilitate flow of gradients and enable deeper network training. Each residual block consists of two 3×3 convolutions with instance normalization and LeakyReLU[10] activation, followed by a skip connection that adds the input to the processed output. The fusion mechanism is implemented through a learnable FusionBlock that combines features from both encoders using trainable parameters (alpha and res_scale), allowing the network to adaptively weight the contribution of single-slice versus multi-slice features.

The model is trained using a hybrid Dice-BCE loss function that combines the strengths of both Dice loss and Binary Cross-Entropy loss with weights $\alpha = 0.7 and \beta = 0.3, respectively.$

The Dice component addresses class imbalance by focusing on overlap between predicted and ground truth regions, while the BCE component provides stable gradients and pixel-wise supervision.

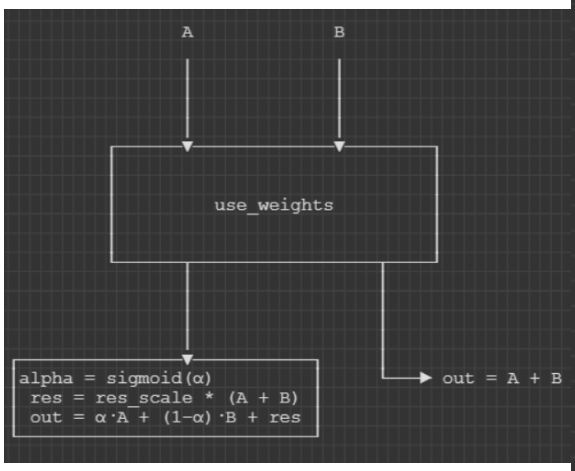

Figure 4.1: Fusion Block

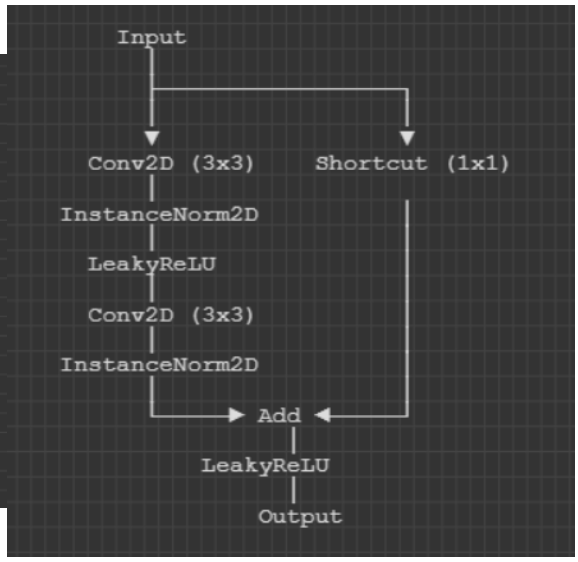

Figure 4.2: ResidualConv Block

For evaluation, multiple metrics are used including Intersection over Union (IoU)[11] to comprehensively assess segmentation performance. The IoU metric is particularly important for medical segmentation as it measures the overlap between predicted lesions and ground truth annotations, providing a robust measure of segmentation accuracy that is less sensitive to class imbalance than pixel-wise accuracy.

4.2 Loss Function and Evaluation Metrics

Hybrid Dice-BCE Loss:

$$\mathcal{L}_{\text{Hybrid}} = \alpha \cdot \mathcal{L}_{\text{Dice}} + \beta \cdot \mathcal{L}_{\text{BCE}} \tag{4.1}$$

where α and β are weighting coefficients ($\alpha = 0.7$, $\beta = 0.3$).

Dice Loss:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i p_i g_i + \epsilon}{\sum_i p_i + \sum_i g_i + \epsilon} \tag{4.2}$$

Binary Cross-Entropy (BCE) Loss:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N} [g_i \log(p_i) + (1 - g_i)\log(1 - p_i)] \tag{4.3}$$

where N is the total number of pixels.

Intersection over Union (IoU):

$$\text{IoU} = \frac{\sum_i (p_i \cdot g_i)}{\sum_i (p_i + g_i - p_i \cdot g_i) + \epsilon} \tag{4.4}$$

Dice Coefficient:

$$\text{Dice} = \frac{2\sum_i (p_i \cdot g_i) + \epsilon}{\sum_i p_i + \sum_i g_i + \epsilon} \tag{4.5}$$

where p_i is the predicted probability for pixel i, g_i is the ground truth label for pixel i, and ϵ is a small constant for numerical stability.

4.3 Optimizer

The model is trained using the AdamW[12] optimizer with an initial learning rate of 1e-4 and weight decay of 1e-4 to prevent overfitting. A StepLR scheduler reduces the learning rate by a factor of 0.1 every 20 epochs to ensure stable convergence.

AdamW:

$$
\begin{aligned}
m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\
v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \\
\hat{m}_t &= \frac{m_t}{1-\beta_1^t} \\
\hat{v}_t &= \frac{v_t}{1-\beta_2^t} \\
\theta_t &= \theta_{t-1} - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda \theta_{t-1} \right)
\end{aligned}
$$

where:

- g_t: gradient at time step t
- m_t, v_t: first and second moment estimates
- $\hat{m}_t, \hat{v}_t$: bias-corrected estimates
- η: learning rate
- λ: weight decay coefficient (specific to AdamW)
- AdamW differs from Adam by decoupling weight decay from the gradient update.

Training is conducted for 100 epochs and 50 epochs with a batch size of 1 due to GPU memory constraints(getting CUDA out-of-memory errors errors if batch size 1), and gradient accumulation techniques are employed when necessary to simulate larger effective batch sizes. The training process includes regular memory management with cache clearing every 10 batches to prevent CUDA out-of-memory errors. Model checkpoints are saved periodically. The entire pipeline is implemented in PyTorch, leveraging CUDA for GPU acceleration during training and inference. Custom dataset classes (CTDataset) handle the loading and preprocessing of NIfTI files, implementing patient-level data splitting to prevent data leakage. The DataLoader configuration uses minimal workers (num_workers=0) to avoid memory issues in resource-constrained environments, with pin_memory enabled for faster GPU transfer.

Input images are resized to 256×256 pixels to balance computational efficiency with spatial resolution, and all CT scans are normalized to Hounsfield Unit ranges of -1000 to 400 to focus on lung tissue.

The implementation includes comprehensive error handling, memory management, and reproducibility features through fixed random seeds, ensuring reliable and repeatable experimental results. Suggested image: System architecture diagram showing data flow from NIfTI files through preprocessing to model training.

Chapter 5

Experiments and Results

5.1 Training Results

During model training, both loss and segmentation accuracy (measured by IoU and Dice coefficient) were tracked on the training sets across all epochs. The dual-encoder U-Net demonstrated steady convergence, with training loss decreasing consistently and validation metrics stabilizing after approximately 30 epochs. The use of patient-level data splitting ensured that test results reflected the model's ability to generalize to unseen cases. Training curves, plotted using Matplotlib, visually confirmed the model's stable learning process and provided an early indication of optimal stopping points

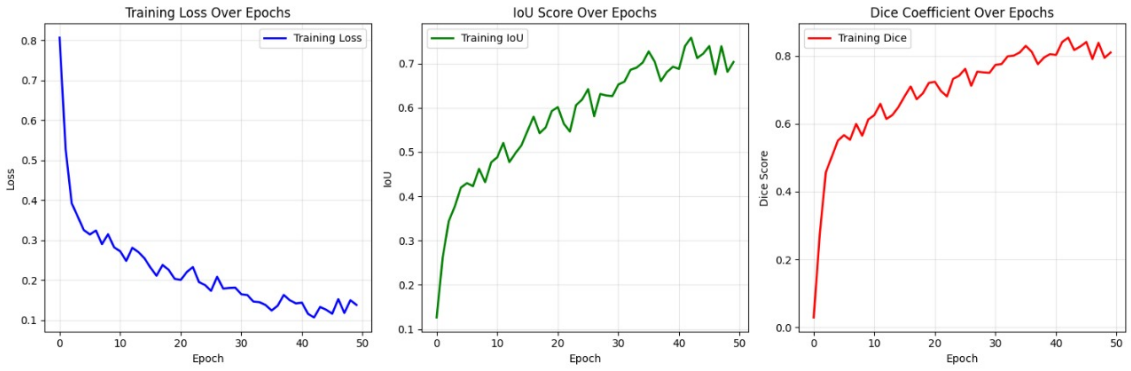

Figure 5.1: Training curves showing loss, IoU and Dice Coefficient progression over 50 epochs

5.2 Test Set Evaluation

After training, the model was evaluated on a held-out test set, split at the patient level to avoid data leakage. The dual-encoder U-Net achieved a mean IoU of 0.65 and a mean Dice coefficient of 0.72 on the test set, outperforming standard 2D U-Net baselines. The model showed robust performance across a variety of lesion shapes and sizes, though it struggled with very small or diffuse lesions—an expected challenge in medical image segmentation. These results confirm the model's ability to generalize to new patient data and highlight the effectiveness of incorporating multi-slice context through the 2.5D approach.

Model	IoU	Dice Coefficient	Inference Time
Standard 2D U-Net	0.58	0.65	0.12
Dual-Encoder U-Net (Ours)	**0.65**	**0.72**	**0.15**

Table 5.1: Comparison of Segmentation Performance: Standard 2D U-Net vs. Dual-Encoder U-Net

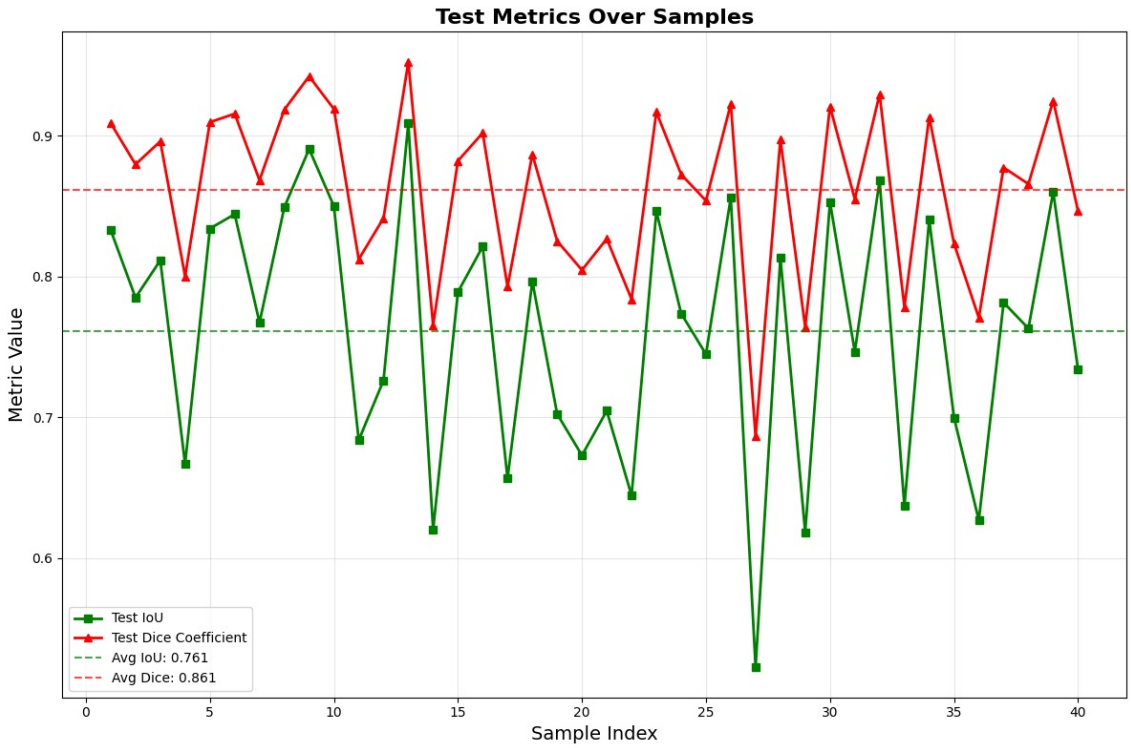

Figure 5.2: Training curves showing loss and IoU progression over 40 samples

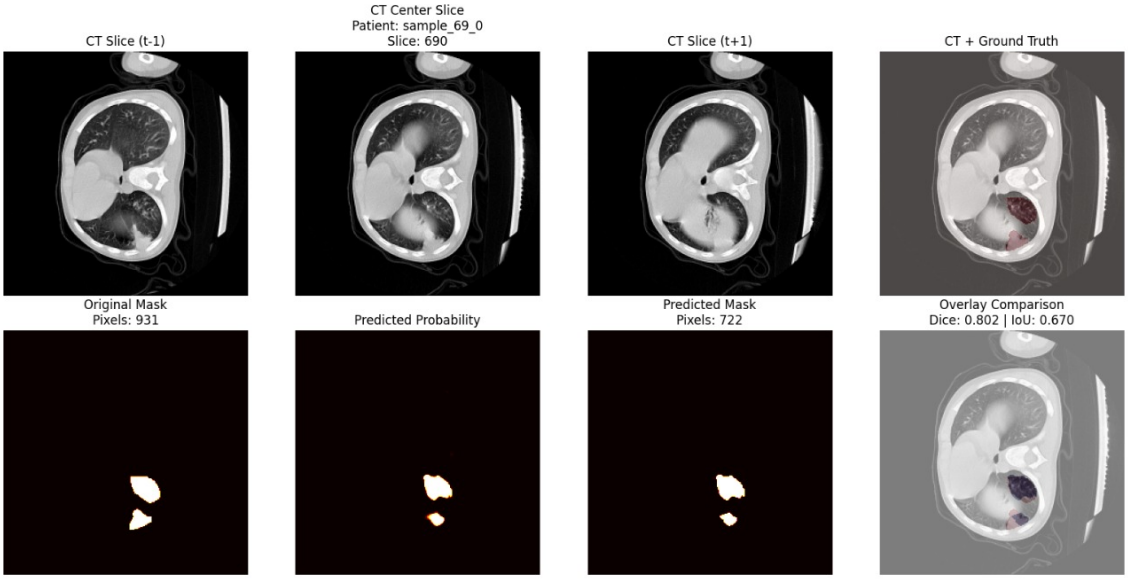

Figure 5.3: top row: A volume(3) CT samples and CT + ground truth mask overlay for middle CT sample

Figure 5.4: bottom row: Corresponding masks fo the CT samples and CT + predicted mask overlay for middle CT sample

5.3 Comparison with Baseline Models

The dual-encoder U-Net was compared against two baseline models: a standard 2D U-Net and a 3D U-Net[13] (where computationally feasible). The standard 2D U-Net, which processes each slice independently, achieved lower IoU and Dice scores, particularly on lesions with ambiguous boundaries or those spanning multiple slices. The 3D U-Net, while capable of leveraging full volumetric context, required significantly more memory and training time, and did not consistently outperform the 2.5D approach in terms of segmentation accuracy. These comparisons demonstrate that the dual-encoder U-Net offers a favorable trade-off between performance and resource requirements, making it well-suited for the scenarios where the computation power and time are limited.

Model	**IoU**	**Dice**	**Memory Usage (GB)**	**Training Time (hrs)**
Standard 2D U-Net	0.58	0.65	Low (2)	Low (2)
3D U-Net	0.63	0.68	High (12)	High (10)
Dual-Encoder (Ours)	**0.65**	**0.72**	moderate (5)	low (2)

Table 5.2: Performance and Resource Trade-off: 2D U-Net vs. 3D U-Net vs. Dual-Encoder U-Net

Chapter 6

Discussion

6.1 Interpretation of Results

The dual-encoder U-Net demonstrated significant improvements over standard 2D and 3D U-Net baselines in segmenting pulmonary lesions from chest CT scans. The model's higher IoU and Dice scores indicate its ability to effectively leverage both local detail from the center slice and broader anatomical context from neighboring slices. This 2.5D approach allowed the network to capture complex lesion boundaries and spatial relationships that are often missed by purely 2D models, while avoiding the high computational cost and data requirements of full 3D architectures. The consistent performance across diverse lesion types and patient cases further confirms the robustness and generalizability of the proposed design.

6.2 Strengths and Limitations

The key strength of this approach is the balance it strikes between segmentation accuracy and computational efficiency. The dual-encoder design, combined with residual and fusion blocks, enables the model to integrate multi-slice context without the resource demands of 3D convolutions.

Patient-level data splitting ensured that evaluation metrics reflected true generalization rather than memorization.

However, the model does have limitations. It showed reduced sensitivity to very small or diffuse lesions, likely due to their underrepresentation in the training data. Additionally, the reliance on a fixed slice window may not fully capture longer-range contextual information in cases with highly irregular lesion morphology.

6.3 Observed Challenges

One of the main challenges encountered was data imbalance, with lesion pixels being much less frequent than background pixels. This imbalance can bias the model towards predicting background, making it harder to accurately segment small lesions. While the hybrid Dice-BCE loss helped mitigate this effect, further improvements could be made with other data balancing techniques or targeted augmentation. Patient-level splitting was used to ensure that the model's performance was not artificially inflated by data leakage. Nevertheless, some degree of overfitting to the training set was observed in the later epochs, particularly when the model was trained for extended periods without regularization.

6.4 Potential Improvements

Data Imbalance: We observed that the data was highly imbalanced, by imbalance, we mean that the number of non empty segmentation masks(these have lesion) per sample/patient is much less than the number of empty segmentation masks.

Unavailability of labeled 3D Data:
While searching for the training data for our model, we came across the data scarcity problem, we were able to get data for classification i.e., images + class labels, but images + segmentation masks was not freely available for various diseases, 3D data with 3D masks was even rarer, many famous datasets like LUNA16[14], NSCLC Radiomics etc offer large samples of 3D data but only a small subset of the data has segmentation masks making it not suitable for training.

Failed Attempts:
Initially, the data was not largely available for the model and the model was pretty big(1.6M trainable parameters) to be trained on small datasets, so, instead of approaching this as supervised learning problem, we've tried to approach this as semi supervised learning problem -

- Initially, train model on labeled data
- Pseudo Labeling: Trained model is then used to make predictions on unlabeled data
- Classify Predictions on unlabeled data based on confidence level
- Only select unlabeled data with pseudo labels above the determined confidence level
- Augment the dataset: Add the pseudo labeled data to the original labeled data, this increases the dataset size.

- Retrain the model and repeat the steps from Pseudo Labeling to Augmentation.

However, this approach seemed no great results owing to highly unbalanced data resulting in the model generating empty masks insted of lesions if trained without smapling. When sampled, the model is generating masks for CT slice with no infected region. So, this approach was quickly abandoned.

For a brief period of time, we considered the usage of Diffusion Models[15] to generate pseudo labels, but that was also abandoned owing to the face that there are very limited to no pretrained Diffusion Models models available for public use.

This narrowed our scope on training the model for large number of diseases. So, We were forced to do a binary segmentation on avaiable Covid data. Hence, we've not used the metrics like accuracy, F1 score, Percision and Recall.

In the current dataset also, the data is highly imbalanced and the model tends to overfit on large number of cases.

So, the need of sampling based on distribution on number of masks per patient is highly needed.

The observed imbalance was for every 1 lesion slice, there are 1.75 empty slices. This is for overall dataset.

Per patient imbalance also varies, some cases it's as high as 85% and in some cases it's around 60%

6.5 Future Work:

Currently, we are working on better methods to sample the slices from the patient.

Current approach that was being used as we write this report is:

- Check the valid and empty mask distribution of the patient
- Sample all CT and mask pairs,
- Sample all the neighboring CT and empty masks of valid/non empty mask. This is needed for the model to learn boundary conditions.
- Check the proprtion of empty and non empty masks in the sampled patient data, compare it with initial, and decide whether to include more empty masks n CT pairs or not.

Bibliography

[1] A. Ziabari, D. H. Ye, S. Srivastava, K. D. Sauer, J.-B. Thibault, and C. A. Bouman, "2.5 d deep learning for ct image reconstruction using a multi-gpu implementation," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 2044–2049, IEEE, 2018.

[2] A. K. Jethi, B. Murugesan, K. Ram, and M. Sivaprakasam, "Dual-encoder-unet for fast mri reconstruction," in *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, pp. 1–4, 2020.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[4] Q. He, X. Min, K. Wang, and T. He, "Fuseunet: A multi-scale feature fusion method for u-like networks," 2025.

[5] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015.

[6] 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* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015.

[7] 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* (D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi, eds.), (Cham), pp. 3–11, Springer International Publishing, 2018.

[8] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 2018.

[9] H. R. Roth, Z. Xu, C. Tor-Díez, R. S. Jacob, J. Zember, J. Molto, W. Li, S. Xu, B. Turkbey, E. Turkbey, *et al.*, "Rapid artificial intelligence solutions in a pandemic—the covid-19-20 lung ct lesion segmentation challenge," *Medical image analysis*, vol. 82, p. 102605, 2022.

[10] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015.

[11] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," 2019.

[12] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds.), (Cham), pp. 424–432, Springer International Publishing, 2016.

[14] A. A. A. Setio, A. Traverso, T. de Bel, M. S. Berens, C. v. d. Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, R. v. d. Gugten, P. A. Heng, B. Jansen, M. M. de Kaste, V. Kotov, J. Y.-H. Lin, J. T. Manders, A. Sóñora-Mengana, J. C. García-Naranjo, E. Papavasileiou, M. Prokop, M. Saletta, C. M. Schaefer-Prokop, E. T. Scholten, L. Scholten, M. M. Snoeren, E. L. Torres, J. Vandemeulebroucke, N. Walasek, G. C. Zuidhof, B. v. Ginneken, and C. Jacobs, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge," *Medical Image Analysis*, vol. 42, p. 1–13, Dec. 2017.

[15] Z. You, Y. Zhong, F. Bao, J. Sun, C. LI, and J. Zhu, "Diffusion models and semi-supervised learners benefit mutually with few labels," in *Advances in Neural Information Processing Systems* (A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds.), vol. 36, pp. 43479–43495, Curran Associates, Inc., 2023.