Executive Summary

Topic: Deep Learning for Automated Thyroid Nodule Delineation

Authors: Palash Mahata (C25007)
Institution: Praxis Tech School

1. Introduction and Problem Statement

The diagnosis of thyroid diseases heavily relies on ultrasound imaging, where the accurate boundary definition (segmentation) of the thyroid gland or nodules is crucial. Traditionally, this segmentation is performed manually by radiologists. This process is not only time-consuming and inefficient but also prone to significant inter-observer variability, leading to inconsistent diagnoses.

This capstone project addresses this limitation by developing and comparatively analysing advanced Deep Learning (DL) models for automated, high-accuracy thyroid ultrasound image segmentation.

This problem is particularly acute in rural or underserved areas, where patients who receive imaging reports often face long delays before consulting a specialist. This "diagnostic gap" creates anxiety and can postpone urgent care. An automated system provides a critical first-line analysis, bridging the gap between imaging and expert consultation.

The primary objective was to rigorously compare three state-of-the-art Convolutional Neural Network (CNN) architectures—U-Net, SegNet, and Residual U-Net—to select a model that not only achieves high accuracy but also demonstrates the consistent and stable performance required for clinical integration.

2. Methodology and Technical Approach

2.1. Data Preparation and Preprocessing

- **Dataset:** The project utilized a comprehensive public Thyroid Dataset consisting of **3,585** high-quality image-mask pairs.
- **Data Split:** The dataset was partitioned into a **training set of 2,509 samples (70%)** and a **testing set of 1,076 samples (30%)** to ensure unbiased evaluation.
- **Preprocessing:** All input images and masks underwent a standardized preprocessing pipeline, including conversion to grayscale, pixel intensity normalization, and resizing to 256x256 pixels to standardize inputs for all models.

2.2. Deep Learning Architectures

Three distinct semantic segmentation architectures were implemented in PyTorch to evaluate the trade-offs between accuracy, efficiency, and complexity.

1. **U-Net (The Baseline):** The industry standard for biomedical segmentation. Its **skip connections** concatenate high-resolution features from the encoder to the decoder, enabling precise localization and boundary definition.
2. **SegNet (The Efficient Alternative):** An architecture designed for computational efficiency. It uses **max-pooling indices** for upsampling, drastically reducing memory usage by avoiding the need to store entire feature maps.
3. **Residual U-Net (The Advanced Performer):** An enhanced U-Net that incorporates **Residual Blocks**. These blocks mitigate the vanishing gradient problem, enabling the stable training of a much deeper network capable of learning more complex features.

2.3. Training Environment and Evaluation

- **Environment:** All models were trained for 25 epochs on Kaggle and Google Colab GPU environments.
- **Training:** An **Adam Optimizer** (learning rate of 1e-4) and **Binary Cross-Entropy (BCELoss)** were used to train the models.
- **Evaluation:** The core performance metric was the **Dice Coefficient**, a standard for segmentation tasks that measures the spatial overlap between the predicted mask and the ground truth mask. A score of 1.0 represents a perfect match.

3. Key Findings and Model Selection

3.1. Final Model Performance Comparison

The comparative evaluation on the independent test set yielded the following quantitative results:

Model Architecture	**Final Test Dice Coefficient**	**Comparative Performance & Stability**
U-Net	**0.9765**	Highest peak accuracy, but training showed some stability fluctuations.
Residual U-Net	**0.9732**	Extremely high accuracy with the most consistent and stable performance.
SegNet	**0.7783**	Significantly lower performance, indicating poor feature reconstruction for this task.

3.2. Selection Rationale: Residual U-Net

While the standard U-Net achieved a marginally higher peak Dice score, the **Residual U-Net was ultimately selected as the optimal model for deployment.**

The rationale for this decision is rooted in reliability: the integration of Residual Blocks in the ResU-Net yielded the most **stable and consistent training curves** and highly reproducible results. In a capstone project aiming for a deployable clinical solution, the inherent **robustness and reliability** of the Residual U-Net were deemed superior to the marginal, less stable accuracy gain of the standard U-Net. This ensures consistent and trustworthy results in a real-world production environment.

The SegNet model's performance was insufficient for this specific clinical task, demonstrating that its efficiency trade-off came at too high a cost to accuracy.

4. Conclusion and Future Scope

This project successfully implemented, trained, and evaluated three distinct deep learning architectures for automated thyroid segmentation. The findings validate that DL, specifically advanced models like Residual U-Net, can achieve human-level accuracy in a stable and reproducible manner.

The **Residual U-Net** was chosen as the superior model due to its excellent balance of high accuracy (0.9732 Dice Coefficient) and unparalleled training stability, making it the most reliable candidate for clinical integration.

This work serves as a strong foundation for a deployable decision-support tool. Future work based on this project will involve:

- **Advanced Architecture Search:** Integrating pre-trained backbones (e.g., ResNet-50) into the Residual U-Net to further improve feature extraction.
- **Clinical Integration:** Deploying the trained model as a service that can interface with a hospital's medical imaging system (PACS) for real-time validation by radiologists.
- **Extension to Multi-Nodule:** Expanding the model to perform multi-class segmentation, enabling it to distinguish between different *types* of thyroid nodules (e.g., benign vs. malignant).