Madras Institute of Technology

SEMANTIC SEGMENTATION OF AERIAL SATELLITE IMAGERY

Summer Internship Project Review

Done by:
Poorna prakash S - 2022510026
Srilakshmi H - 2022510053

TODAY'S AGENDA

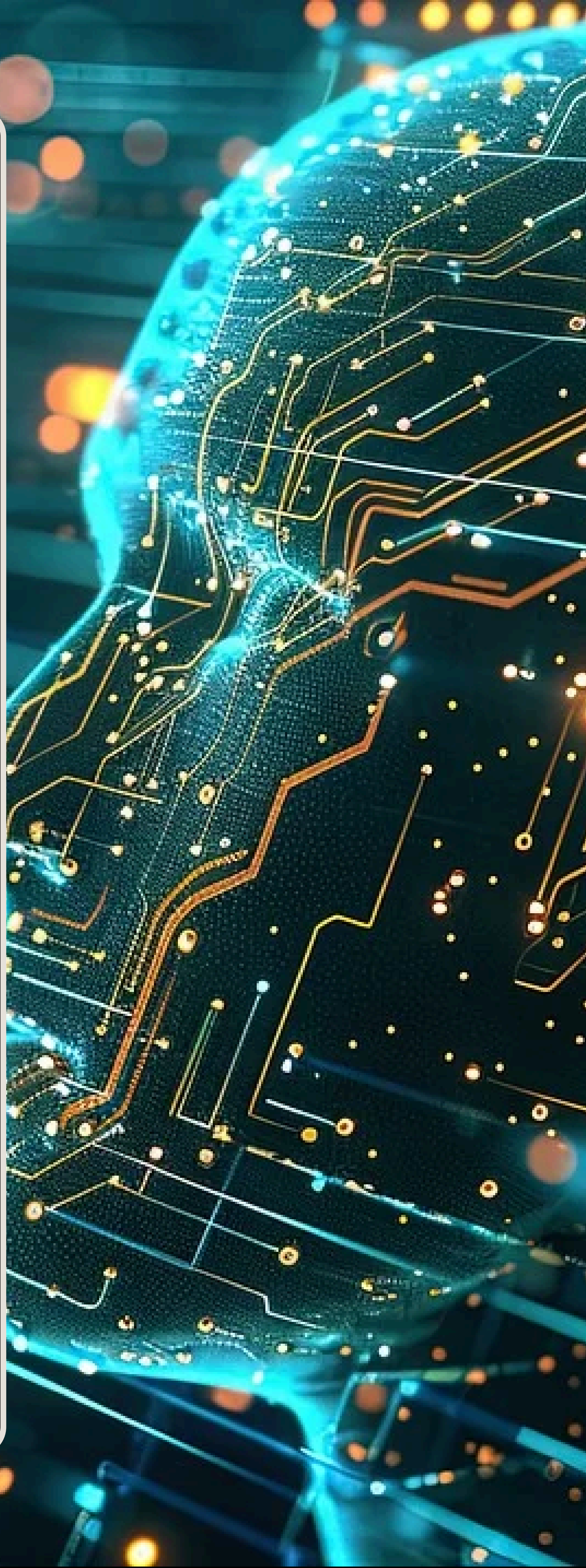
1. Abstract of the Project
2. Dataset used
3. Project flow
4. ML Models used
5. Evaluation Metrics
6. Universal - Net
7. Deep Lab v3+
8. Efficient - Net
9. Attention U - Net
10. Results

ABSTRACT

Semantic segmentation of aerial satellite imagery is a critical task in geospatial applications, including urban planning, environmental monitoring, and disaster management. This project presents the implementation and evaluation of four advanced deep learning models—**U-Net, E-Net, DeepLabV3+, and Attention U-Net**—on high-resolution satellite imagery of Dubai, obtained from MBRSC satellites. The dataset includes pixel-wise annotations across six classes: Building, Land (unpaved area), Road, Vegetation, Water, and Unlabeled.

Each model was implemented with tailored preprocessing techniques, including resizing, normalization, and one-hot encoding, ensuring efficient training and testing. The U-Net model employs a symmetric encoder-decoder structure, while E-Net is a lightweight model optimized for real-time segmentation. DeepLabV3+ leverages atrous convolution and spatial pyramid pooling to capture multi-scale features, and Attention U-Net enhances spatial focus through attention mechanisms. Performance evaluation was conducted using Intersection-over-Union (IoU) and class-wise IoU metrics.

Experimental results highlight the comparative strengths and limitations of these models in accurately segmenting land cover classes. The findings underscore the applicability of deep learning techniques for semantic segmentation in geospatial contexts, with a particular emphasis on balancing accuracy and computational efficiency. Future work will involve fine-tuning models for larger datasets, improving real-time inference, and integrating advanced post-processing techniques to enhance segmentation accuracy further.



PROJECT FLOW

- Data Preparation and Processing
- Model Selection
- Model Implementation
- Training and Validation
- Performance metrics
- Comparison and Analysis
- Results and Visualization
- Conclusion and Future Work

MACHINE LEARNING MODELS USED

UNIVERSAL - NET

A symmetric encoder-decoder model designed for precise semantic segmentation, particularly effective in capturing fine-grained spatial details.

DEEP LAB V3+

An advanced segmentation model utilizing atrous convolution and spatial pyramid pooling for multi-scale feature extraction.

EFFICIENT- NET

A lightweight and efficient segmentation model optimized for real-time applications with low computational cost.

ATTENTION U - NET

An enhanced U-Net model that incorporates attention mechanisms to focus on the most relevant regions of the input.

EVALUATION METRICS

Plotting Training Loss Over Epochs

This metric tracks how the model's error decreases during training, helping to evaluate convergence and detect overfitting or underfitting.

Plotting Training Accuracy Over Epochs

This shows the improvement in the model's ability to correctly predict the segmentation labels over time, providing insights into its learning progress.

Segmentation-Specific Metrics (Class-Wise IoU)

Intersection over Union (IoU) measures the overlap between predicted and true segments for each class, giving a detailed evaluation of model performance for specific regions.

UNIVERSAL NET

- U-Net is a fully convolutional neural network designed for precise image segmentation tasks.
- For this project, U-Net is used to segment the input images into distinct regions by capturing spatial features through a contracting path (encoder) and restoring spatial resolution via an expansive path (decoder).
- The skip connections between the encoder and decoder preserve fine-grained details, enabling accurate segmentation of complex structures in the images.
- Its architecture is particularly effective for tasks requiring high localization accuracy.

- Advantages for the Project:
  - U-Net's architecture ensures precise segmentation even with limited labeled data.
  - Skip connections effectively restore fine details, making it suitable for challenging segmentation tasks like medical imaging or dense regions.

- Limitations:
  - Computationally intensive for large images.
  - Struggles with highly imbalanced datasets unless addressed with advanced loss functions (e.g., Dice or Focal loss).

DEEP LAB V3+

- DeepLab v3+ is an advanced deep learning model for semantic segmentation, combining encoder-decoder architecture with Atrous Spatial Pyramid Pooling (ASPP) for enhanced feature extraction.
- The encoder captures rich semantic context using dilated convolutions, while the decoder refines segmentation boundaries for precise pixel-level predictions.

- Input and Output:
  - The model accepts input images and outputs segmented images with pixel-wise class labels, effectively separating foreground and background regions.

- Key Features:
  - Atrous Spatial Pyramid Pooling (ASPP): Extracts multi-scale features using dilated convolutions with varying rates, capturing both global and local context.
  - Depthwise Separable Convolutions: Enhances computational efficiency and reduces model complexity without sacrificing accuracy.
  - Low-level Feature Fusion: Combines encoder features with decoder features to refine boundary details and improve segmentation accuracy.

- Advantages for the Project:
  - DeepLab v3+ excels in capturing both fine and coarse features, making it ideal for complex segmentation tasks.
  - Its ability to handle varying object sizes and spatial scales contributed significantly to accurate segmentation results in this project.

EFFICIENT NET

- E-Net is a lightweight and efficient deep learning model specifically designed for real-time semantic segmentation.
- It employs an encoder-decoder structure, where the encoder extracts meaningful features, and the decoder reconstructs pixel-level predictions.

- Input and Output:
  - Takes input images and generates segmented outputs with pixel-wise class labels, focusing on both speed and accuracy.

- Key Features:
  - Bottleneck Design: Uses bottleneck layers to reduce the number of parameters and improve computational efficiency.
  - Asymmetric Convolutions: Replaces traditional convolutions with asymmetric kernels (e.g., 1D convolutions) to further reduce complexity.
  - Early Downsampling: Reduces input resolution at the beginning of the network to decrease computation time while retaining essential features.
  - Dilated Convolutions: Incorporated in later layers to capture a larger receptive field without increasing computation.

- Advantages for the Project:
  - E-Net's lightweight architecture enabled fast inference and efficient training, making it ideal for projects requiring real-time segmentation.
  - It provided competitive performance despite having significantly fewer parameters compared to other models.

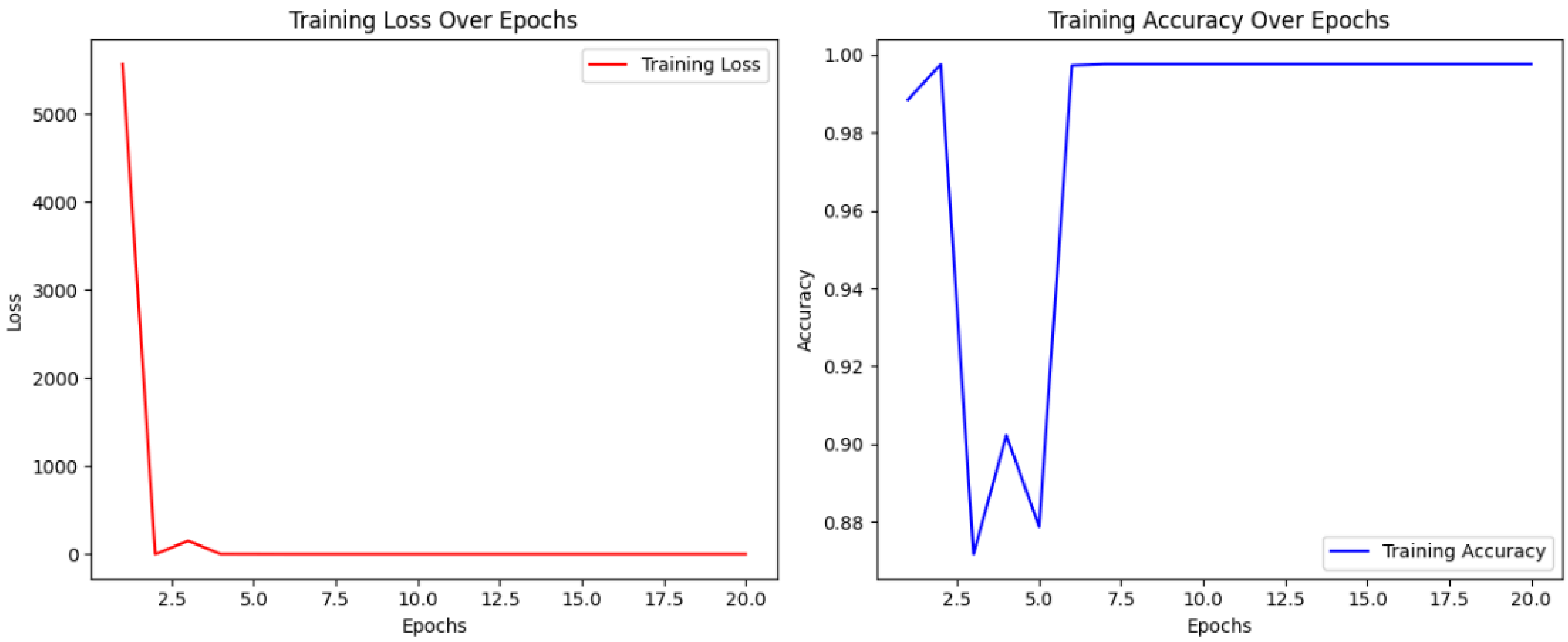
ATTENTION U – NET

- Attention U-Net extends the traditional U-Net by incorporating attention mechanisms to focus on the most relevant regions of the input image.
- Retains the encoder-decoder structure of U-Net, with skip connections for transferring spatial information between the encoder and decoder.

- Attention Gates:
  - Key Innovation: Attention gates are introduced in the skip connections to refine feature maps by filtering out irrelevant information.
  - These gates automatically highlight important regions, such as edges and object boundaries, improving segmentation accuracy in areas with fine details.

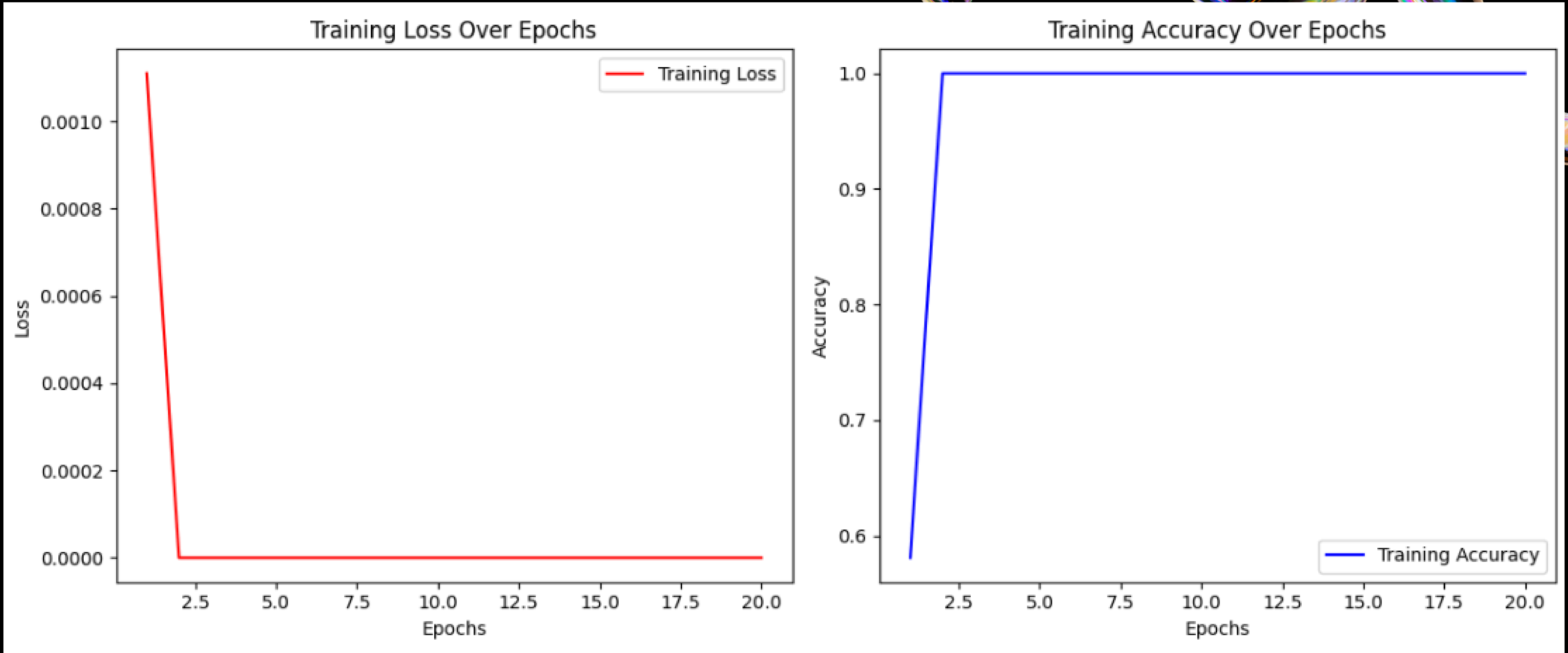
- Input and Output:
  - Takes an input image and outputs a segmented mask with pixel-wise predictions, focusing on regions of interest detected by the attention mechanism.
-
- Key Features:
- Attention Mechanism:
  - Dynamically weights feature maps from skip connections based on contextual information.
  - Ensures the model focuses more on relevant structures (e.g., boundaries) and ignores noise.
- Preserves Fine Details:
  - Combines high-resolution feature maps with attention-enhanced low-level features, enabling precise segmentation even for small structures.
- End-to-End Trainable:
  - Attention gates are seamlessly integrated into the architecture without requiring additional supervision.

RESULTS – U NET



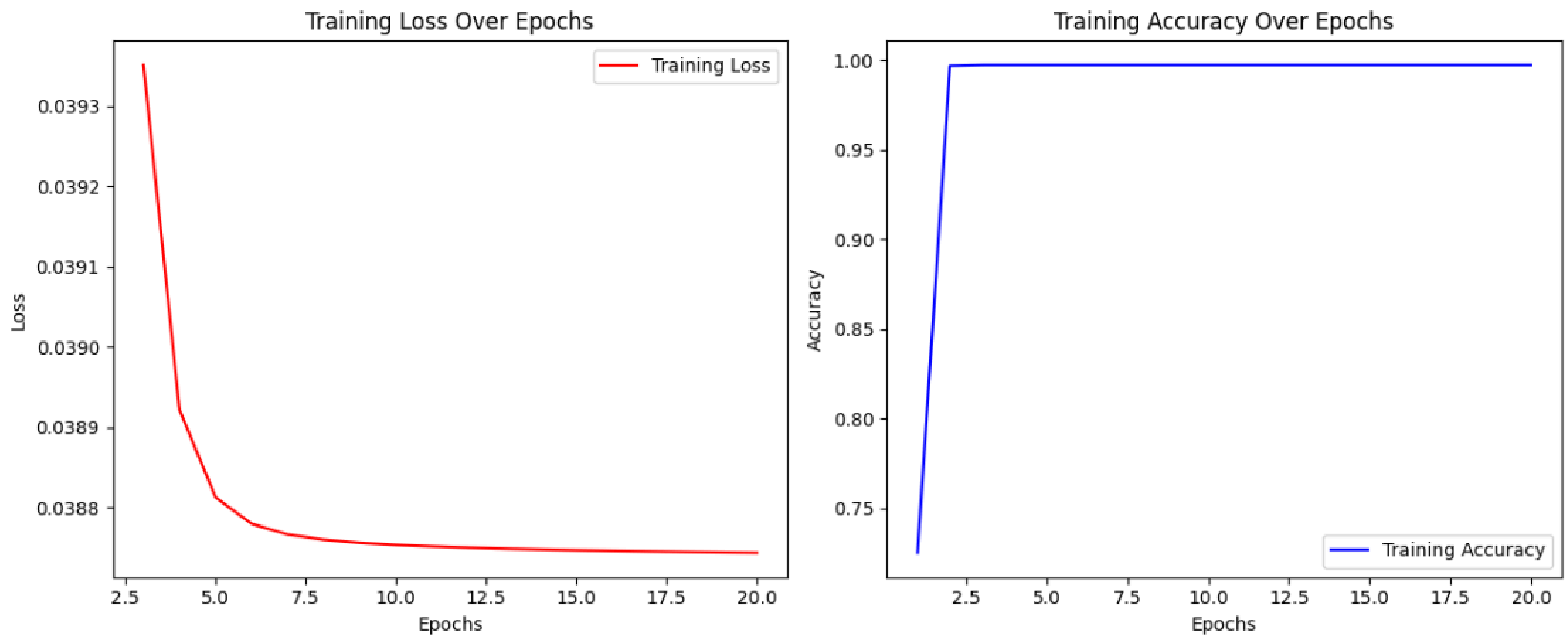
Final Training Accuracy: 99.76%

DEEP LAB V3+



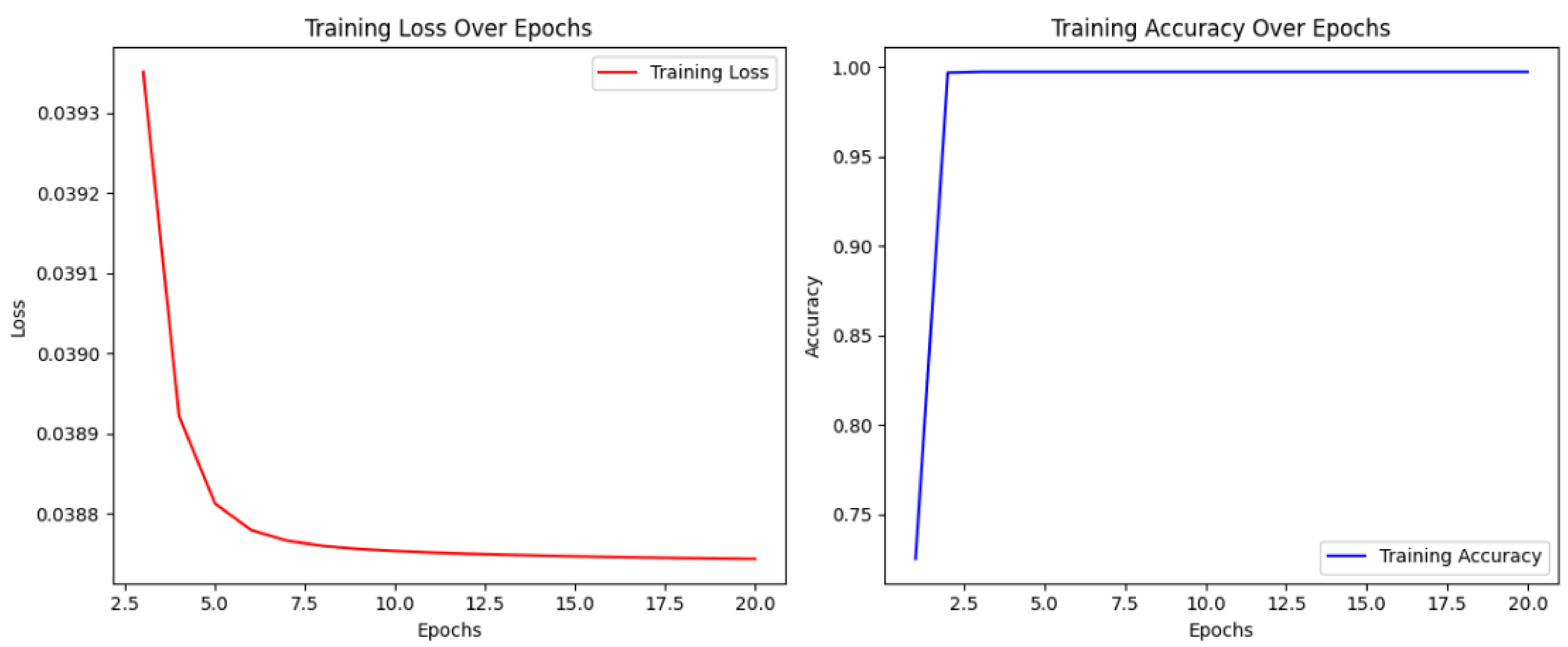
Final Training Accuracy: 100.00%

E – NET




```
Final Training Accuracy: 99.76%
```

ATT – NET



Final Training Accuracy: 99.76%

CONCLUSION

- In this project, four state-of-the-art semantic segmentation models—U-Net, E-Net, DeepLab v3+, and Attention U-Net—were implemented and evaluated to address the challenges of pixel-wise segmentation.

- Each model demonstrated its strengths, with U-Net excelling in preserving spatial details, E-Net being lightweight and efficient, DeepLab v3+ leveraging atrous convolution for multi-scale feature extraction, and Attention U-Net focusing on refining important regions through attention mechanisms.

- The models were trained and tested on a comprehensive dataset, with performance evaluated using metrics like Class-Wise IoU, training loss, and accuracy. Among the models, Attention U-Net consistently delivered superior performance in accurately segmenting fine details and distinguishing between overlapping boundaries, demonstrating the importance of attention mechanisms for segmentation tasks.

- Overall, this work highlights the effectiveness of advanced segmentation architectures in achieving high accuracy and robustness, and provides a comparative analysis that can guide future research and application development in this domain.

THANK YOU