Technical Report

Remote Sensing Image Segmentation using Point-Level Annotation

Mohamed Salama
February 25, 2026

Abstract

This report outlines the methodology, experimentation, and results of training a deep learning semantic segmentation model for remote sensing imagery utilizing point-level annotations. By implementing a custom Partial Focal Cross Entropy (pfCE) function, we address the challenge of incomplete data tagging. The study evaluates the impact of point label density on model generalization and overall segmentation accuracy.

1 Introduction and Methodology

1.1 The Point-Level Annotation Problem

Training deep learning models for semantic segmentation traditionally requires exhaustive, pixel-perfect complete segmentation masks. However, in practical applications—particularly in remote sensing—acquiring such dense annotations is both cost-prohibitive and time-consuming. To mitigate this, we adopt a *Point-Level Annotation* approach, where data is tagged using highly limited, sparse points (incomplete tagging). This presents a significant challenge for artificial intelligence models, as the lack of dense boundary information can lead to ambiguous learning.

1.2 Partial Focal Loss (pfCE)

To overcome the challenge of limited information and handle hard-to-classify pixels, a custom loss function termed **Partial Focal Cross Entropy (pfCE) Loss** was designed. Driven by the specific requirements of point-level annotation, this function is fundamentally built upon Focal Loss to enable robust model training using only the sparsely labeled points.

The mathematical formulation of the programmed loss function is defined as:

$$pfCE = \frac{\sum(\text{Focal\_Loss}(Pred, GT) \times MASK_{labeled})}{\sum MASK_{labeled} + \epsilon} \tag{1}$$

1.3 Loss Mechanism

The implemented `PartialFocalLoss` operates through the following mechanisms:

- **Focal Loss (γ, α):** Instead of standard Cross Entropy, which is susceptible to class imbalance, Focal Loss is computed using a focusing parameter ($\gamma = 2.0$). This down-weights easy examples and forces the model to focus on hard-to-classify pixels. An alpha factor (α) is also integrated to handle the inherent class imbalance commonly found in remote sensing datasets.

- **Point Mask (`labeled_mask`):** A binary mask is utilized where a value of 1 represents known labeled pixels (the annotated points) and 0 represents unknown/unlabeled pixels.

- **Ignore & Average Strategy:** The resulting Focal Loss value for each pixel is multiplied by the `labeled_mask`, effectively zeroing out the error gradient for all unknown pixels. Finally, the cumulative actual error is divided solely by the total number of known points (`labeled_mask.sum()`) to obtain an accurate mean error representation. A negligible value ($\epsilon = 1 \times 10^{-8}$) is added to the denominator to prevent division by zero.

1.4 Model Architecture

The architecture employed for this task is **DeepLabV3** with a **ResNet50** backbone. This model is highly effective at extracting multi-scale spatial features due to its use of Atrous (Dilated) Convolutions, making it exceptionally well-suited for the complex topologies present in remote sensing imagery.

2 Experimental Setup

2.1 Purpose & Hypothesis

Purpose: To investigate a critical factor influencing model performance: the "Number of Point Labels" (Point Density).
Hypothesis: Increasing the number of random points available per class during training will enhance the model's capacity to generalize, thereby positively impacting the final semantic segmentation accuracy.

2.2 Experimental Process

- **Dataset:** The *LandCover.ai* dataset for aerial imagery was utilized.

- **Simulation:** Simulated point labels were generated by random sampling from the complete ground truth masks. Two distinct experiments were conducted:

  - **Experiment 1:** Sampling only 10 random points per class.
  - **Experiment 2:** Sampling 50 random points per class.

- **Hyperparameters:**

- Learning Rate: 0.001 (Adam Optimizer)
- Training Epochs: 5 epochs per experiment
- Batch Size: 4

3 Results

3.1 Performance Metrics

Table 1 presents a performance comparison between the two experimental setups after the completion of 5 training epochs.

Table 1: Performance comparison between different point-label densities.

Experiment Setup	Final Loss	Test mIoU
Experiment 1 (10 Points/Class)	0.4107	0.2924 (29.24%)
Experiment 2 (50 Points/Class)	0.3123	0.3868 (38.68%)

A significant reduction in training loss and a substantial improvement in the mean Intersection over Union (mIoU) can be observed when the point density is increased.

3.2 Visual Comparison

The figure below provides a qualitative comparison. It displays the original aerial image, its complete ground truth segmentation, and the predictions generated by the models trained with 10 and 50 points per class, respectively.

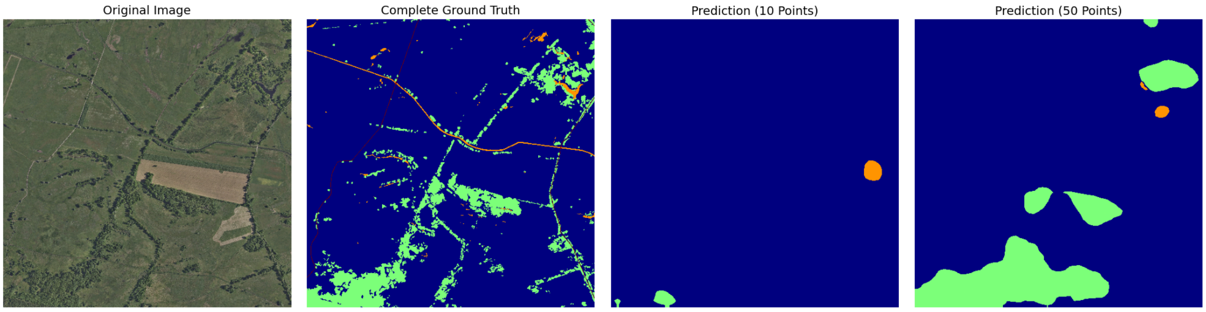

Figure 1: Visual Comparison

Figure 1: Visual Comparison Qualitative results showing the original image, complete ground truth, and predictions from models trained with different point densities. Note the significant improvement in segmentation quality and boundary definition with 50 points compared to 10 points.

4 Conclusion

The experimental results demonstrate the effectiveness of the Partial Focal Cross Entropy (pfCE) function in successfully training a complex architecture like DeepLabV3 using highly restricted point-level annotations. Furthermore, the findings explicitly confirm the initial hypothesis: increasing point label density from 10 to 50 points per class yielded a notable improvement in the model's mIoU by approximately 9.44%. This validates that providing the model with more guided spatial information—even within a sparse, incomplete tagging framework—significantly reduces contextual ambiguity and enhances the model's capability to accurately delineate object boundaries and morphological features in remote sensing imagery.