**Experiment Design**

We need a consistent baseline setup first:

* **Dataset:** Choose a suitable public remote sensing dataset with full segmentation masks (e.g., ISPRS Vaihingen/Potsdam, DeepGlobe Land Cover, SEN12MS, etc. - *you'll need to select one appropriate for your specific remote sensing task*). Split it into training, validation, and test sets.
* **Base Model:** Use the U-Net with a pre-trained MobileNetV2 backbone (as developed previously), incorporating regularization techniques (Dropout, L2).
* **Training:** Train using the Partial Cross-Entropy loss, Adam optimizer, a defined learning rate, and Early Stopping based on validation loss (calculated using partial labels on the validation set).
* **Evaluation:** Evaluate all models on the *held-out test set* using the *full ground truth segmentation masks*. The primary metric will be Mean Intersection over Union (Mean IoU), calculated manually after prediction as discussed. Pixel Accuracy can be a secondary metric.

**Experiment 1: Effect of Point Label Density**

* **Purpose:** To investigate how the number of simulated point labels per object instance affects the final segmentation performance.
* **Hypothesis:** Increasing the number of point labels per object will improve segmentation accuracy (Mean IoU), likely with diminishing returns after a certain point density. Very sparse labels (e.g., 1 per object) might lead to poor results.
* **Experimental Process:**
  1. Define a set of point densities to test, e.g., N = [1, 3, 5, 10, 20] points per object instance. (*Adjust these values based on typical object sizes in your chosen dataset*).
  2. For each value of N:
     + Generate the training and validation sets by simulating point labels: For each object instance in the ground truth masks, randomly sample N pixels within that object's boundary. Mark these pixels with the object's class ID, and mark all other pixels with the IGNORE\_LABEL.
     + Train the base model using these simulated partial labels until convergence (using Early Stopping).
     + Evaluate the trained model on the test set using the full ground truth masks and calculate Mean IoU.
  3. Compare the Mean IoU results across the different values of N.
* **Expected Results:** A plot showing Mean IoU vs. N points per object. We expect IoU to increase as N increases, possibly showing a steeper rise initially and then plateauing. Qualitative results (predicted masks) might show better object completeness and fewer false positives with higher N.

**Experiment 2: Effect of Point Sampling Strategy**

* **Purpose:** To determine if the spatial strategy used for sampling point labels influences segmentation performance, particularly boundary accuracy.
* **Hypothesis:** Sampling points strategically near object boundaries might lead to better boundary definition in the final segmentation compared to purely random sampling within the object, for a fixed number of points.
* **Experimental Process:**
  1. Fix a reasonable number of points per object instance (e.g., N=5, based on results from Experiment 1 or a chosen default).
  2. Implement two (or more) sampling strategies:
     + **Strategy A (Random):** Sample N points randomly from anywhere within the object's mask (as done in Experiment 1).
     + **Strategy B (Boundary-Focused):** Sample N points randomly, but only from pixels that are within a small distance (e.g., 2-3 pixels) of the object's boundary.
     + *(Optional Strategy C: Centroid/Skeleton-based sampling)*
  3. For each strategy:
     + Generate training and validation sets using the respective point sampling method.
     + Train the base model using these simulated partial labels until convergence.
     + Evaluate the trained model on the test set using the full ground truth masks, calculating Mean IoU. Consider adding a boundary-specific metric (like Boundary F1-score) if precise boundaries are critical.
  4. Compare the Mean IoU (and Boundary F1, if used) between the different sampling strategies.
* **Expected Results:** A comparison (e.g., bar chart) of Mean IoU for Strategy A vs. Strategy B. We might observe that Strategy B yields slightly better boundary metrics or visually sharper edges in the output masks, even if the overall Mean IoU is similar to Strategy A.

**Technical Report Structure**

Based on the task description, structure your report as follows:

1. **Introduction:**
   * Briefly introduce the problem: Semantic segmentation in remote sensing.
   * State the specific challenge: Training effectively with limited, point-based annotations instead of full masks.
   * Mention the proposed approach: Using a deep learning model (e.g., U-Net w/ backbone) trained with a custom Partial Cross-Entropy loss.
   * State the purpose of the report: To detail the method, present experiments exploring factors affecting performance, and discuss the results.
2. **Method:**
   * **Dataset:** Describe the chosen remote sensing dataset, including its source, classes, resolution (if relevant), and how it was split into train/validation/test sets.
   * **Point Label Simulation:** Explain *how* the point labels were simulated from the full ground truth masks for the baseline experiments (e.g., random sampling). Mention the IGNORE\_LABEL used.
   * **Network Architecture:** Detail the segmentation model used (e.g., U-Net with MobileNetV2 backbone). Mention the use of pre-trained weights, regularization techniques (Dropout, L2), and the specific layers used. A diagram can be helpful.
   * **Partial Cross-Entropy Loss:** Explain the concept and provide the mathematical formulation or pseudocode for the implemented partial CE loss function.
   * **Training Details:** Specify the optimizer (Adam), learning rate, batch size, number of epochs (or use of Early Stopping), and the computing environment (hardware/software libraries like TensorFlow/Keras).
   * **Evaluation Metrics:** Define the metrics used for final evaluation (Mean IoU, potentially Pixel Accuracy, Boundary F1) and explain how they are calculated on the full masks of the test set.
3. **Experiments and Results:**
   * **Experiment 1: Effect of Point Label Density**
     + **Purpose & Hypothesis:** State clearly.
     + **Experimental Process:** Describe how N was varied and the training/evaluation procedure specific to this experiment.
     + **Results:**
       - Present quantitative results (e.g., a table or graph showing Mean IoU vs. N).
       - Show qualitative results (e.g., side-by-side images of input, ground truth, and predicted masks for different N values on sample test images).
       - Discuss the observed trends and whether they support the hypothesis.
   * **Experiment 2: Effect of Point Sampling Strategy**
     + **Purpose & Hypothesis:** State clearly.
     + **Experimental Process:** Describe the different sampling strategies tested and the procedure.
     + **Results:**
       - Present quantitative results comparing the strategies (e.g., bar chart of Mean IoU, table with Boundary F1).
       - Show qualitative results comparing predicted masks from different strategies, focusing on object boundaries.
       - Discuss the findings regarding the impact of sampling strategy.
4. **Discussion:**
   * Summarize the key findings from the experiments regarding point density and sampling strategy.
   * Discuss the overall effectiveness of the partial CE loss approach for point-supervised segmentation in this context.
   * Acknowledge limitations (e.g., use of simulated points vs. real-world sparse annotations, limited dataset scope, specific network choice).
   * Suggest potential future work (e.g., exploring different network backbones, incorporating semi-supervised techniques mentioned in the PDF to leverage unlabeled pixels, testing on real point annotations, experimenting with different loss function variations).
5. **Conclusion:**
   * Briefly reiterate the problem, method, and main experimental conclusions.
6. **(Optional) Appendix:**
   * Include code snippets (like the partial CE loss implementation).
   * Additional result plots or images.

This structure provides a clear framework for exploring the factors influencing your model's performance and presenting your findings professionally.