Weakly Supervised Segmentation with Point Annotations using Partial Cross-Entropy Loss

Syed Waleed Hyder
syedwaleedhyder@gmail.com

***Abstract*— This study investigates the impact of point annotation on the performance of segmentation models by exploring how varying the percentage of labeled points influences model accuracy and Intersection Over Union (IoU) compared to fully supervised learning. Employing a U-Net architecture, experiments are conducted with partial labels at varying densities—1%, 5%, and 10%—and performances are benchmarked against a fully supervised model trained with 100% label density. The study aims to contribute to optimizing training pipelines in scenarios where full labeling is impractical, thus broadening the applicability of deep learning in resource-limited settings.**

I. Introduction

In this work, we investigate a point annotation-based semantic segmentation approach, which relies only on partial ground truth segmentation points. We transform the ground truth by applying the mask on target segmentation during the calculation of partial cross-entropy loss. Since the spatial dimensions are preserved for training, we can use any semantic segmentation network for generating the masks. Our hypothesis is increasing the percentage of labeled points will improve model performance in terms of both accuracy and Intersection over Union (IoU); but can we somehow match the result of the fully supervised model using reduced label density? Experiments on a semantic segmentation dataset demonstrate that our approach achieves similar performances using reduced segmentation masks for training.

In summary, our contributions include:

- We sample the point annotations to reduce the label density of the segmentation masks.
- We further apply partial cross-entropy loss only on the sampled coordinates to train the model on the reduced labels.

II. Our Approach

Our study aims to evaluate the efficiency of training segmentation models using a partial labeling approach and compares it to traditional fully supervised learning. To this end, we use the U-Net architecture, a model well-established for image segmentation, combined with a loss function designed to handle partially labeled data effectively.

A. Setup

Utilize the U-Net architecture and train the model multiple times, each with a different percentage of labeled points i.e. 1%, 5%, 10%).

The dataset utilized remains consistent across all experiments to isolate the effect of labeled data density. Model performance is quantitatively assessed through IoU metrics.

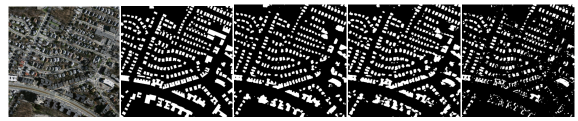

Fig. 1: Results with different label density

B. Loss Function

Given the partial nature of our dataset labels, we employ a custom loss function, Partial Cross-Entropy Loss (pCE), which is specifically tailored to handle instances where only a subset of pixels are labeled within each training sample. The pCE loss is defined as follows:

$$\text{pCE} = -\frac{1}{\sum \mathbf{M}} \sum_{i \in \mathbf{L}} y_i \log(p_i)$$

where p_i represents the predicted probability of the labeled class at pixel i, y_i is the ground truth label, and $\mathbf{M}$ is a binary mask indicating labeled pixels ($\mathbf{L}$) within the image. This formulation ensures that the loss calculation is confined exclusively to the labeled pixels, thus optimizing the model's learning focus towards the most informative parts of the input data.

III. Experiments

Datasets, and Annotations. We use the Massachusetts Buildings Dataset, which consists of 151 aerial images of the Boston area, with each of the images being 1500 × 1500 pixels for an area of 2.25 square kilometers. Hence, the entire dataset covers roughly 340 square kilometers. The data is split into a training set of 137 images, a test set of 10 images, and a validation set of 4 images. Due to computational limits, slices of size 512x512 are created, with zero padding on the edges to maintain consistent slice dimensions.

Implementation Details. Our semantic segmentation model is implemented in pyTorch. We randomly initialize our model and use ADAM optimization with a learning rate of 0.001. We train our model for 20 epochs with a batch size of 4.

Metrics. *IoU* score.

A. Robustness against reduced label density

We conduct an experiment in which label density of the segmentation mask is reduced to examine the robustness of the segmentation model. In particular, at training, we randomly generate a point mask which contains 1 with a probability p. We use this mask to calculate the loss. The

Label Density p			
1%	**5%**	**10%**	**100%**
0.660	**0.760**	**0.770**	**0.789**

TABLE I: IoU scores against different label densities. The best results are in **bold**. Second best results are *underlined*.

IoU scores of this approach with different p are illustrated in Tab. I. It can be seen that point annotation approach demonstrates a similar performance compared to fully supervised approach as shown in Fig. 1.

Limitations. Our segmentation-based approach struggles with accurately delineating object boundaries due to the limited availability of such data. This can be improved by applying different sampling algorithm which samples the pixels on the object boundaries.

IV. Conclusion

This study demonstrates that segmentation models trained with point annotations and partial cross-entropy loss achieve competitive performance with significantly reduced label requirements. Our experiments with U-Net architecture suggest that using just 1%, 5%, and 10% labeled data can approach the accuracy and IoU scores of fully supervised models. This method proves particularly effective in scenarios where extensive data labeling is impractical, such as geographical and medical imaging.

Future work will aim to improve object boundary delineation and explore more complex datasets, potentially broadening the applicability of this weakly supervised approach.