Integration of Deep Learning (DL) in Geospatial Object-Based Image Analysis (GEOBIA) for Image Segmentation in Remote Sensing Tasks

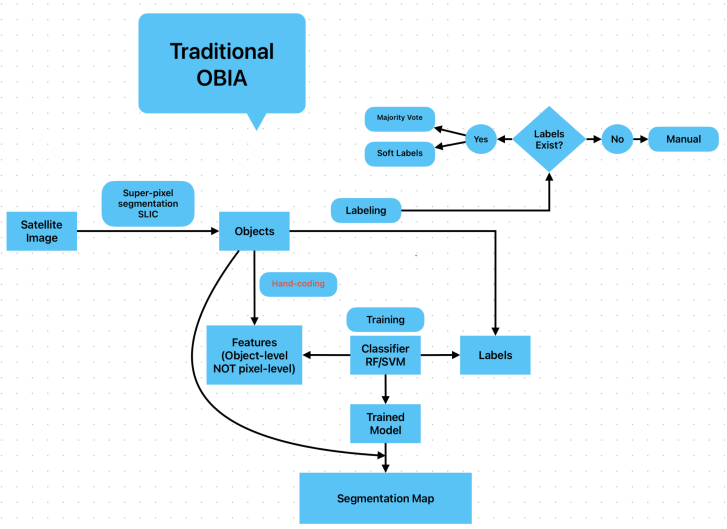
Project - CSE598: Machine Learning for Remote Sensing

Vivek Sahukar **ASU ID 1230067360**

Introduction

For semantic image segmentation, DL methods use pixel-based analysis (i.e. calculating pixel-wise loss and predictions). On the other hand, OBIA is used in traditional remote sensing for semantic segmentation, which consists of two steps: segmentation (grouping of pixels into super-pixels to generate “objects”) and classification (classifying the objects obtained from the previous steps into different required classes).

Fig.1 OBIA Workflow



The super-pixel segmentation algorithms (such as SLIC, Felsenszwalb) are used to create objects in satellite images. Those objects are labeled using existing pixel-wise labels by majority voting. Then, hand-coded features are extracted from the objects, and a machine learning classifier such as Random Forest (RF) is trained to predict the labels. Finally, a trained RF is applied to other images to get the segmentation map. OBIA incorporates spatial and spectral context, reducing noise and better reflecting real-world structures. However, OBIA's segmentation stage involves manual feature selection and parameter tuning, making it subjective and time-consuming. On the other hand, DL excels at automatic feature selection and hyperparameter tuning but has limitations in handling high-resolution imagery & capturing contextual information.

DL integration into OBIA workflow remains limited due to the following:

- challenges in adapting pixel-centric DL segmentation models to object-based frameworks
- more computational complexity and overhead
- lack of clear advantages of DL over traditional machine learning techniques in terms of accuracy (Blaschke, 2010)for OBIA classification

This project aims to bridge this gap by developing a novel framework that effectively integrates DL into OBIA pipelines for remote sensing image analysis by automating feature extraction for segmentation, learning object-level features, and improving model performance metrics.

The relevant literature review reveals some of the latest work on integrating DL with OBIA. (Zaabar et. al 2022) uses CNN models to extract the heatmaps from the images, which were later utilized as input features to perform the OBIA. The proposed method still uses the different scale parameter values obtained through trial-and-error in the OBIA step, which defeats

the purpose of integrating DL into OBIA to obviate the need for manual parameter selection in OBIA methods. (Luo et. al 2023) uses SLICO (Zero Parameter Version of Simple Linear Iterative Clustering) segmentation methods to produce superpixel objects from remote-sensing images with similar shapes and close areas. Then, ViT (Vision Transformer) is used for superpixel classification. Finally, an object-based K-nearest neighbor filtering algorithm is used as a post-processing method to reduce the pretzel phenomenon ("prediction of incorrect superpixel objects leads to many spots in the predicted image"). The main aim of integrating DL methods in OBIA is to obviate the need for manual hyperparameter selection in the segmentation stage. However, the user still has to choose the parameters for SLICO (or any other superpixel algorithm) and object-based K-nearest neighbor algorithms. (Herlawati et. al 2022) uses DL-based model viz. DeepLabV3+ for semantic segmentation for the land cover classification task, which does not incorporate features from object-level analysis. Hence, all these methods still require manual parameter selection for the segmentation algorithm, and the DL models still work with pixels instead of object-level features. Therefore, there is a need for a hybrid model that can work with object-level features and incorporate DL into the OBIA workflows.

Methodology

Three binary classification image segmentation datasets from the GeoBench paper were used. Due to time constraints, the other 3 GeoBench datasets were not included in the analysis as previously proposed in the project proposal. The details of the datasets used are provided in

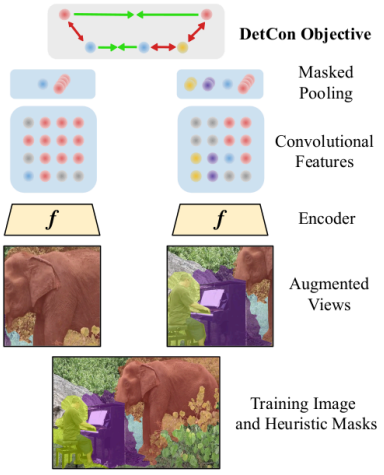
Table 1: Dataset Features

Geobench Dataset	Image Size	# Classes	Train (%)	Val (%)	Test (%)	# Bands	Resolution (m)
m-pv4ger-seg	320x320	2	80	10	10	3	0.1
m-nz-cattle	500x500	2	80	10	10	3	0.1
m-NeonTree	400x400	2	60	20	20	5	0.1

To establish the baseline, the OBIA method was used only for the nz-cattle dataset, for which mIoU was 0.39 and mean Precision was 0.72. The OBIA method did not perform well and took a lot of time to run, even for smaller datasets. Hence, it was not used in the further analysis.

Fig 2 Contrastive Detection Method

Contrastive Detection (DetCon) method proposed by (Hénaff et. al 2021) is a key component in integrating Object-Based Image Analysis (OBIA) with deep learning. The process begins with identifying object-based regions using unsupervised segmentation algorithms, which generate approximate image-computable masks. These masks facilitate the extraction of object-level features from images. Each image undergoes random data augmentations, and convolutional features are extracted for each augmented image. These features are then pooled according to the masks



to ensure that object-level features are learned independently. The pooled features from different augmentations of the same object are pulled closer in the feature space, enhancing the model's ability to recognize objects across different views. This mechanism allows the model to learn powerful, transferable object-level representations without the need for manual annotation, making it highly efficient and effective for OBIA. DetCon can use different segmentation methods, such as spatial proximity-based masks, masks generated by the Felzenszwalb-Huttenlocher algorithm, and human-annotated masks. This highlights DetCon's flexibility in utilizing both simple and complex segmentation methods to facilitate learning. The use of such diverse segmentation techniques ensures that the model can adapt to various levels of segmentation quality, which is crucial for effectively applying OBIA principles in deep learning contexts. Thus, the DetCon method was chosen to integrate DL with OBIA. By automating the generation of object-specific masks and focusing on object-level features, DetCon reduces the reliance on extensive manual annotations and improves the efficiency and scalability of learning object-centric models in diverse imaging conditions.

Images in each dataset were resized to 512x512 (power of 2), which is more suitable for downsampling and upsampling images in the model. Imagenet's mean and standard deviation were not used for normalization since it did not provide any additional gain in model performance. Only Red, Blue, and Green channels were used since the GeoBench paper also used the RGB channels and reported that the additional channels did not provide any extra gain in model performance. Train/Val/Test split according to the `default_partition` JSON file provided by the GeoBench paper. Images were counted in the train/val/test dataloader to make sure there was no data leakage across the dataset splits. Sample images and labels from the dataloader were plotted to ensure the dataloaders were working as expected.

The ResNet50 model was chosen as the final model for image segmentation. I also implemented the UNet with ResNet18 decoder for Lab2 for `nz-cattle` dataset for Lab 2. So even though UNet with ResNet architecture would probably be a better model, DetCon pre-trained weights are available only for ResNet50 and ResNet100. Resnet50 was chosen due to its lower computational cost, and the image size in the datasets is not big enough to warrant the use of a more complex Resnet100 model (which could have potentially overfitted).
The ResNet50 model was modified as follows:

1. Removed the fully connected and the average pooling layer.
2. Increased the output resolution by adding additional upsampling layers and hence redesigned the decoder.
3. Freezing the last few layers did not improve the metrics, so the whole model was trained but with fewer epochs (=10). After 10 epochs, the model began to overfit.

For Model 1 and 2, respectively, Imagenet and DetCon pre-trained Resnet50 weights were used, and training was done simultaneously on two different RTX 3090 GPUs. Training and Validation curves for loss, IoU, and Precision were plotted to determine the optimal number of epochs and check if the loss is converging and the model is not overfitting. After hyperparameter tuning, the model was trained with these final hyperparameters for each of the five different seeds: lr=0.001, batch_size=16, num_epochs=10, and Optimizer=Adam. Then, the model was run on the final test set to get the Precision and IoU. Finally, the mean for the precision and IoU was calculated for 5 seeds with 95% confidence intervals, and a t-test was performed for statistical significance.

Dice Loss was used instead of the default binary cross entropy since Dice Loss is more suited for image segmentation tasks. Mean Precision and Intersection over Union (IoU) were performance metrics. Custom functions were written to calculate the Dice Loss, Precision, and IoU. Five different seeds were used, and the mean was calculated for Precision and IoU with 95% confidence intervals. A t-test was performed to determine whether the results for the metrics for model 1 (Imagenet weights) were statistically different from the metrics for model 2 (DetCon weights). i.e., did DetCon pretraining help the model learn and perform better statistically? The model architecture and hyperparameters were kept the same across the two models to compare the effectiveness of object-level features learned by the model in the DetCon pretraining compared to other deep learning models.

Results

Table 2 shows the results for three different GeoBench datasets for binary image segmentation tasks. The table compares the results for the model trained using Imagenet vs DetCon weights and reports the mean values for Precision (mPrec) and IoU (mIoU) over 5 different seeds with 95% confidence intervals mentioned in brackets. The last two columns present the t-test statistic with p-value in the brackets, to show statistical significance. Model 1 represents the DL only method and Model 2 represents the DL+OBIA hybrid approach.

Table 2: Mean Precision and IoU with statistical significance for GeoBench Datasets

Dataset	**Imagenet (Model 1)**		**DetCon (Model 2)**		**Precision**	**IoU**
	mPrec	mIoU	mPrec	mIoU		
nz cattle	0.80 (0.75, 0.85)	0.67 (0.66, 0.67)	0.79 (0.75, 0.83)	0.68 (0.67, 0.68)	0.60 (0.57)	-2.89 (0.02)
neon tree	0.57 (0.53, 0.60)	0.17 (0.14, 0.19)	0.61 (0.57, 0.65)	0.16 (0.14, 0.19)	-2.21 (0.06)	0.46 (0.66)
pv4ger	0.98 (0.97, 0.99)	0.93 (0.90, 0.97)	0.99 (0.98, 1.00)	0.97 (0.96, 0.97)	-1.81 (0.11)	-2.47 (0.04)

The t-test precision is insignificant for any of the datasets, since the p-values are greater than the commonly used significance level of 0.05. This suggests that the differences in precision between the two models (imagenet vs DetCon) are not statistically significant at the 5% level. However, for the Neon-Tree dataset t-test precision (0.06) is quite close to the threshold, indicating a marginal case that might be considered significant in a less stringent analysis or could prompt further investigation. This shows no statistically significant difference in mean precision if DetCon weights are used compared to the Imagenet weights. Contrary to this, for the t-test IoU, the p-value are much higher than the typical significance level of 0.05 for the two datasets: nz-cattle and pv4ger. This implies that the DetCon trained model is much better than Imagenet trained model and proves the hypothesis that the DetCon pretraining helps the model to learn the object-level features and perform statistically significantly better than the usual DL models. However, it was surprising that the DetCon pre-trained model did not perform any better

for the Neon-tree dataset for mean IoU and only marginally better for mean Precision. One possible reason for this could be that the IoU for Neon-Tree is very low (0.17) compared to the other two datasets, both for Imagenet and DetCon models, which implies that both the models fail to learn from the data because the data is indeed more difficult to learn. This is more evident by analyzing the sample images; neon trees are difficult to segment, but the cows and solar panels are more easily distinguishable. The image features in the neon-tree dataset are possibly more distinct from the Imagenet dataset, so even the Deton pretraining did not help the model to learn object-level features effectively.

Discussion

These results are consistent with the mean IoU reported in the GeoBench paper. While the pv4ger-seg results exceed the GeoBench result (~0.94), the mIoU for nz-cattle is less than the GeoBench result (~0.80) for 3 models but is consistent with other 3 models mIoU (~0.68). However, mIoU for neon-tree is much less than the GeoBench result (~0.45-0.55 range). The difference could be attributed to the difference in the model architecture and hyperparameters used in the GeoBench paper. It is reiterated that a model other than the one used in GeoBench was used due to the DetCon pre-trained weights being available for ResNet only. For DetCon-pretrained weights to be more useful, we need bigger and high-resolution image data. I would test my hypothesis with much bigger datasets and multi-class labels to see whether DetCon pre-training helps to learn object-level features from multiple classes of interest. Moreover, due to time and computational resource constraints, I used the DetCon pre-trained weights provided by (Hénaff et al. 2021). For future research, I would use domain-specific unlabeled dataset and do the DetCon pre-training myself, as mentioned by (Hénaff et al. 2021), so that more object-level features, which are more specific for the downstream tasks, can be learned. I would then use those DetCon pre-trained weights for supervised learning task on similar another dataset. Additionally, I would use DetCon pre-trained weights for Resnet50 and use them for training UNet with the Resnet50 encoder so that analysis can be done, which would be more comparable with GeoBench results.

Conclusion

The above results show that the DetCon-pretraining helps the model to learn object-level features and perform better than other deep learning-based segmentation models and OBIA methods. The model seems to perform better for datasets of images with bigger sizes and higher resolution. Thus, DetCon pre-training helps integrate DL into OBIA methods and harness the benefits of both approaches: DL and OBIA. However, further experiments need to be done with much bigger datasets with multi-class labels to test whether the DetCon pretraining helps the model learn object-level features and whether this approach can be scaled up across different datasets and tasks. For future research, I propose to use the datasets from the agriculture-vision challenge, a multi-class image segmentation task for satellite images of agricultural fields. I would use their unlabeled dataset for DetCon pre-training and then fine-tune those weights for the other labeled dataset for the supervised learning task. I would also include bands apart from RGB to see if the DetCon pre-training can extract additional object-level features from the extra bands.

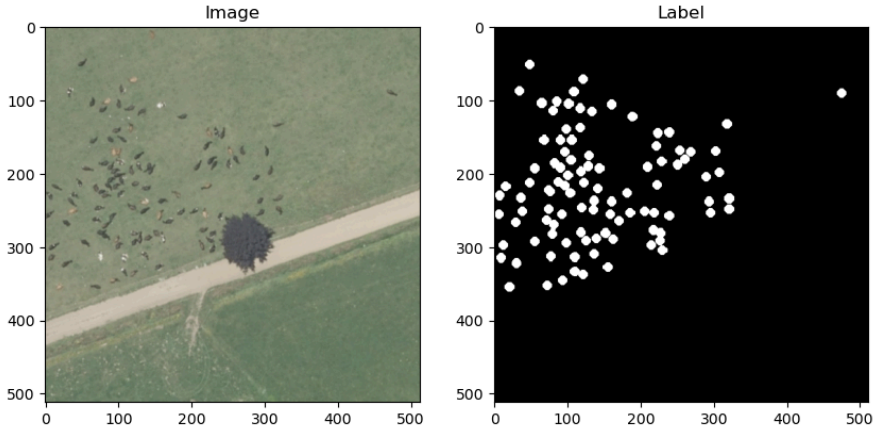
References

1. Du, S., Du, S., Liu, B., & Zhang, X. (2021). Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high-resolution remote sensing images. International Journal of Digital Earth, 14(3), 357-378.
2. Herlawati, H., Handayanto, R. T., Atika, P. D., Sugiyatno, S., Rasim, R., Mugiarso, M., ... & Purwanti, S. (2022, December). Semantic Segmentation of Landsat Satellite Imagery. In 2022 Seventh International Conference on Informatics and Computing (ICIC) (pp. 1-6). IEEE.
3. Song, A., Kim, Y., & Han, Y. (2020). Uncertainty analysis for object-based change detection in very high-resolution satellite images using deep learning network. Remote Sensing, 12(15), 2345.
4. Luo, C., Li, H., Zhang, J., & Wang, Y. (2023, July). OBViT: A high-resolution remote sensing crop classification model combining OBIA and Vision Transformer. In 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) (pp. 1-6). IEEE.
5. Wang, J., Zheng, Y., Wang, M., Shen, Q., & Huang, J. (2020). Object-scale adaptive convolutional neural networks for high-spatial resolution remote sensing image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 283-299.
6. Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2018). An object-based convolutional neural network (OCNN) for urban land use classification. Remote sensing of environment, 216, 57-70.
7. Zaabar, N., Niculescu, S., & Kamel, M. M. (2022). Application of convolutional neural networks with object-based image analysis for land cover and land use mapping in coastal areas: A case study in Ain Témouchent, Algeria. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 5177-5189.
8. Hénaff, O. J., Koppula, S., Alayrac, J. B., Van den Oord, A., Vinyals, O., & Carreira, J. (2021). Efficient visual pretraining with contrastive detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10086-10096).
9. Ibrahim, A., & El-kenawy, E. S. M. (2020). Image segmentation methods based on superpixel techniques: A survey. Journal of Computer Science and Information Systems, 15(3), 1-11.
10. Blaschke, T. (2010). Object based image analysis for remote sensing. ISPRS journal of photogrammetry and remote sensing, 65(1), 2-16.
11. Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., ... & Tiede, D. (2014). Geographic object-based image analysis–towards a new paradigm. ISPRS journal of photogrammetry and remote sensing, 87, 180-191.
12. Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote sensing of environment, 118, 259-272.
13. Chen, G., Weng, Q., Hay, G. J., & He, Y. (2018). Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities. GIScience & Remote Sensing, 55(2), 159-182.

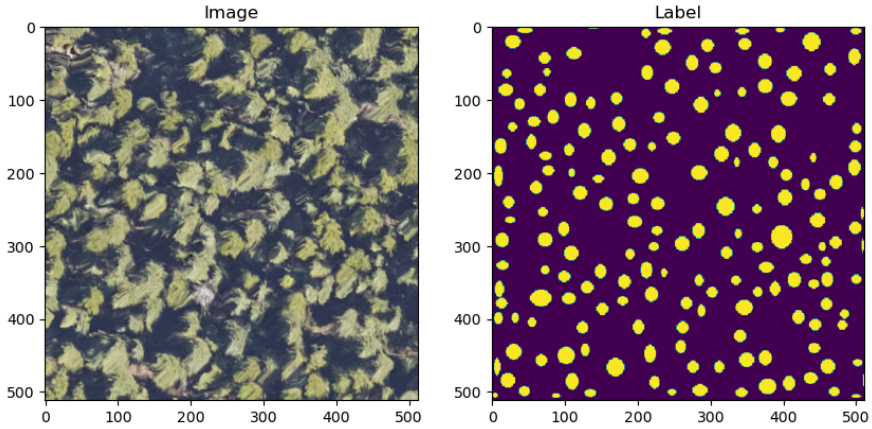
Appendices

The code files for replicating the result can be accessed at this GitHub Repository: https://github.com/viveksahukar/obia-dl

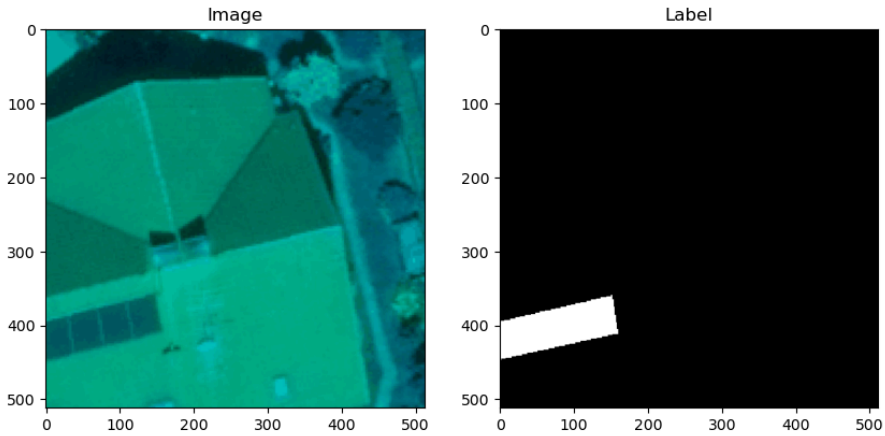
Sample Image and Label from Training Dataloader for NZ-Cattle GeoBench Dataset



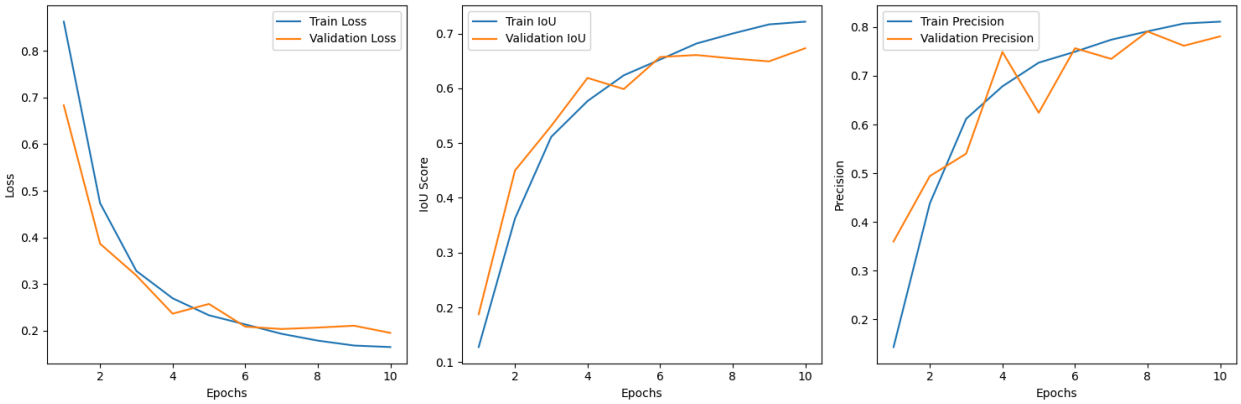
Sample Image and Label from Training Dataloader for Neon-Tree GeoBench Dataset



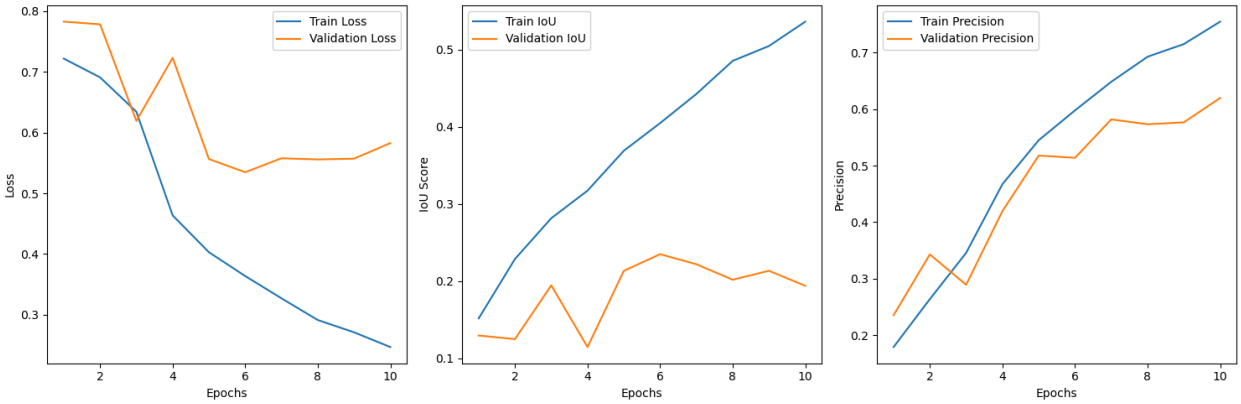
Sample Image and Label from Training Dataloader for pv4ger-seg GeoBench Dataset



Training and Validation Plots for NZ-Cattle GeoBench Dataset (Seed=5)



Training and Validation Plots for Neon-Tree GeoBench Dataset (Seed=5)



Training and Validation Plots for pv4ger-seg GeoBench Dataset (Seed=5)