DL-OBIA: Merging Deep Learning with Object-based Image Analysis

Vihang Pancholi
vpancho1@asu.edu
1229382416

Abstract

Traditionally, object-based image analysis (OBIA) has concentrated on pixel-level loss computation, which can be problematic close to object edges or in dimly illuminated areas, potentially resulting in inaccuracies. The purpose of this study is to determine whether object-level loss calculation in OBIA can improve image analysis's accuracy and efficiency. The study uses the Land Cover Challenge dataset as a baseline and uses deep learning methods for pixel-based segmentation, including DeepLabV3+ and U-Net++. To achieve object-level semantic segmentation, a unique method combines pretrained DetCon weights and a ResNet50 encoder with DeepLabV3+. The findings show that the new strategy performs better than baseline techniques, especially when it comes to mean Intersection-over-Union (IoU). The improved mIoU point to the possibility of further improvements through improved model parameters and improved DetCon pretraining on remote sensing tasks. This study opens the door for improvements in remote sensing and natural resource management by highlighting the potential of using deep learning techniques into OBIA for more reliable land cover classification.

I. Introduction

Traditional object-based image analysis (OBIA) divides images into distinct objects or areas, which are then labeled for further analysis. It starts with feature extraction from each object, using this information to train classifiers for distinguishing different object types. However, conventional OBIA can encounter challenges in pixel-level loss computation due to ambiguity near object edges or in poorly lit areas, leading to potential errors. Despite these challenges, OBIA models have shown resilience to pixel-level errors, which is beneficial for various image processing tasks. Integrating deep learning techniques into OBIA can enhance efficiency and performance, improving object recognition, feature extraction, and classification for more accurate and robust analyses.

A. Objective

This study seeks to examine the effects of using object-level loss calculation in OBIA rather than pixel-level loss calculation. The study intends to solve the shortcomings of pixel-level loss and investigate how this technique may enhance the precision and effectiveness of image analysis by moving the focus to object-level metrics. The objective of this study is to provide important new understandings into the possible advantages of using object-level loss computations in OBIA and the ways in which deep learning methods can improve the procedure as a whole.

B. Background

Through the use of segmentation-based techniques, object-based image analysis (OBIA) (Blaschke, 2010) enabled more efficient analysis of high-resolution pictures. By taking into account spatial relationships and contextual information, it improves classification accuracy.

OBIA is a vital tool for remote sensing applications since it makes activities like change detection, feature extraction, and accuracy assessment easier [1]. Using low-level cues, DetCon (Henaff et al., 2021) propose a self-supervised learning technique that divides images into objects and background regions. This allows for faster pretraining on large datasets and improves performance on future tasks. Top-performing models among self-supervised methods pretrained on ImageNet display state-of-the-art performance, matching recent approaches training larger models on far larger datasets. DetCon's performance significantly depends on mask alignment with object boundaries, indicating the possibility of improved unsupervised segmentations and sophisticated scene understanding via repetitive iterations [2]. Blaschke et al. [2013] trace the history of Geographic Object-based Image Analysis (GEOBIA) (Chen et al., 2018) back to its roots in image segmentation and spatial analysis, the paper investigates this emerging trend in remote sensing and GIScience literature. It explores if GEOBIA represents a new paradigm by evaluating how it differs from per-pixel techniques, talking about fundamental ideas like objects and ontologies, and examining theoretical underpinnings. GEOBIA constitutes a unique and developing paradigm in remote sensing and GIScience based on Kuhn's definition of a paradigm and a review of scientific literature [3, 4].

II. Methodology

The approach to the study has been outlined in this section. The details, features, and preprocessing methods of the employed datasets are described. The machine learning models that were tested are described in depth, along with the rationale for their choice and any performance-enhancing modifications. Finally, a summary of the evaluation measures used to gauge the performance of the approach is given, conveying a sense of how the success of the methods was determined. The pipeline for the project has been described below in Figure 1.

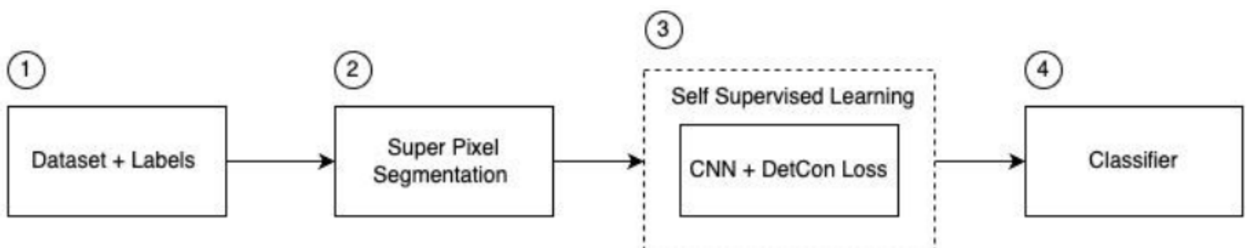

Figure 1: Project pipeline.

A. Data

1. Dataset Description

The research employed the Land Cover Challenge dataset (Ilke et al., 2018), which comprised 803 RGB satellite images with a pixel resolution of 50 cm and a size of 2448x2448 pixels [8]. The images, which show a range of land cover types in the Chesapeake Bay watershed, were taken by a satellite operated by DigitalGlobe. The collection also contains 172 test photos and 171 validation images; however, the test images do not possess masks. For the purpose of labeling land cover, each satellite image is matched with a mask image. This allows for the identification of seven different groups of land cover types: water, barren land, urban land, agricultural land, rangeland, forest land, and unknown. Urban land (0,255,255), agricultural land (255,255,0), rangeland (255,0,255), forest land (0,255,0), water (0,0,255), barren land (255,255,255), and unknown (0,0,0) are the color-coded classes in RGB.

2. *Data Augmentation*

A range of transformations were applied to the training dataset in order to increase the diversity of the data and boost the generalization capacity of the model. Each image in the train data was subject to random cropping and horizontal and vertical flipping, which mimicked various viewing angles and orientations. Center Cropping was utilized in the validation dataset to obtain a centrally centered portion from every image.

B. Model Setup

1. *Traditional OBIA Method*

The conventional approach to OBIA involves performing superpixel segmentation of the image, applying a mask to the segmented image, obtaining features from the objects, and using a random forest classifier to categorize the objects. This architecture has been detailed in Figure 2 below.

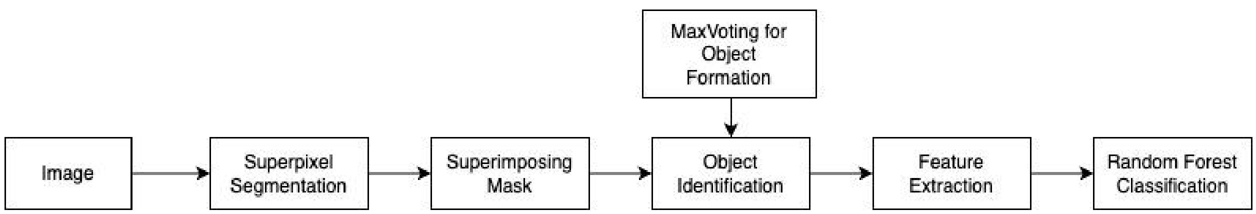

Figure 2: Traditional OBIA approach.

However, because of complexity of methodology and paucity of labeled data which is crucial to the classifier's effectiveness for both training and validation, the efforts of this experiment have mostly been unsuccessful. The initial experiments resulted in an underperforming model.

2. *Baseline Setup: Pixel-based Approach*

Two models—DeepLabV3+ and U-Net++—were used to segment the image pixel by pixel in order to create the baseline experiment. U-Net++ incorporates a number of U-shaped sub-networks into a nested, dense skip connection architecture. In contrast, DeepLabV3 uses a spatial pyramid pooling module in conjunction with dilated convolutions to capture fine features and multi-scale context. By computing loss at the pixel level in this experiment, the effectiveness of the models in differentiating between various forms of land cover can be precisely measured.

3. *Novel Approach Setup*

The novel approach used a custom ResNet50 with an additional sequential layer before the head FCN performing semantic segmentation. Another model uses DeepLabV3+ with a ResNet50 encoder to perform semantic segmentation and classify images at the object level. DetCon, a technique for learning representations via object recognition and contrastive learning, improves model performance. This structure has been described in Figure 3 below.

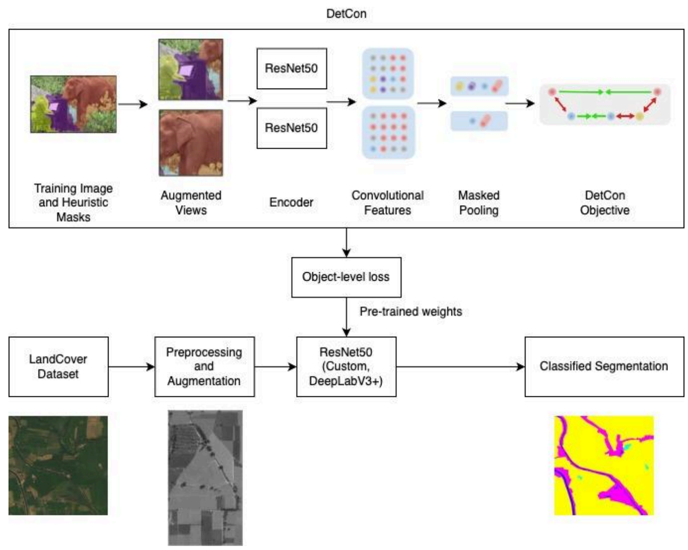

Figure 3: Novel approach.

4. *Ablation Study*

Gamma correction was applied to the dataset as a preprocessing step to preserve a consistent brightness and contrast level throughout the dataset by adjusting the image luminance. Two learning rate values, 0.00008 and 0.0001, were used to examine the novel method and determine how these rates affected the model's performance.

5. *Evaluation Metrics*

The Intersection-over-Union (IoU) and precision metrics, and dice loss were used to evaluate the performance of the model. The overlap between the ground truth and expected masks is measured by dice loss. The intersection of the actual and expected masks with respect to their union is quantified by IoU.

III. Results

The results of the experiments conducted above have been detailed in this section. Table I below shows the performance comparison of the baseline models and the novel approach models with respect to mean IoU.

Table I: Performance comparison of baseline and novel method.

	Model	**mIoU**
Baseline	U-Net++	0.442
	DeepLabV3+	0.464
Novel Approach	DetCon + ResNet50	0.502
	DetCon + DeepLabV3+	0.579

Among the baseline methods, it was found that DeepLabV3+ performed better than U-Net++ for semantic segmentation. In the novel models, DetCon combined with DeepLabV3+ showed better results compared to DetCon combined with ResNet50. It can be observed that the novel approach is more performant than the baseline, albeit by a small margin.

The results of the ablation studies have been tabulated below in Table II.

Table II: mIoU when novel approaches are subject to ablation studies.

Model	**Ablation Study (mIoU)**		
	Gamma Correction	**Learning Rate**	
		0.0001	**0.00008**
DetCon + ResNet50	0.496	0.502	0.480
DetCon + DeepLabV3+	0.568	0.579	0.501

Gamma correction did not contribute to a better performance; on the contrary, a slight dip in mIoU was observed with its application in both models. With respect to learning rate, a larger value resulted in an increased performance in both the models, with the DetCon + DeepLabV3+ model combination being superior in all cases. The precision was observed to be uniform across all models, and was therefore omitted as a helpful metric in the evaluation process.

IV. Discussion

The novel approach exhibits a modest improvement over traditional pixel-based segmentation and loss calculation methods, indicating progress in the development of more effective models for object-level semantic segmentation. This improvement, albeit minor, suggests that the approach is on the right track. The minority of the improvement may be due to the use of DetCon weights pretrained on the ImageNet dataset. Training DetCon on a dataset more closely aligned with remote sensing tasks could potentially lead to further enhancements in model performance, offering a promising avenue for future research and development.

V. Conclusion

The research investigated novel approaches in object-based image analysis (OBIA) and showed how object-level loss computation could improve model performance. The new method produced a powerful image segmentation and classification model by using complex architectures like DeepLabV3+ with a ResNet50 encoder and pretrained DetCon weights. Further substantial progress might be made by matching DetCon pretraining with remote sensing tasks and fine-tuning model parameters. This study opens up new avenues for advancement in remote sensing and natural resource management in the future by advancing image analysis methods for more precise and effective land cover classification.

References

[1] Blaschke, Thomas. "Object based image analysis for remote sensing." ISPRS journal of photogrammetry and remote sensing 65.1 (2010): 2-16.

[2] Hénaff, Olivier J., et al. "Efficient visual pretraining with contrastive detection." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[3] Blaschke, Thomas, et al. "Geographic object-based image analysis–towards a new paradigm." ISPRS journal of photogrammetry and remote sensing 87 (2014): 180-191.

[4] Chen, Gang, et al. "Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities." GIScience & Remote Sensing 55.2 (2018): 159-182.

[5] Robinson C, Hou L, Malkin K, Soobitsky R, Czawlytko J, Dilkina B, Jojic N. Large Scale High-Resolution Land Cover Mapping with Multi-Resolution Data. Proceedings of the 2019 Conference on Computer Vision and Pattern Recognition (CVPR 2019).

[6] Gupta, Ritwik, et al. "Creating xBD: A dataset for assessing building damage from satellite imagery." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019.

[7] Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." IEEE transactions on pattern analysis and machine intelligence 34.11 (2012): 2274-2282.

[8] Demir, Ilke, et al. "DeepGlobe 2018: A Challenge to Parse the Earth Through Satellite Images." In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.

[9] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." Proceedings of the European conference on computer vision (ECCV). 2018.