Object Detection in Satellite Imagery

Author: Suparna Reddy
M.Tech in Artificial Intelligence, Jain University (2022–2024)
Date: July 14, 2025

1. Introduction

This report documents my implementation and analysis of an unsupervised image segmentation pipeline for high-resolution satellite TIFF images. The primary goal of this project was to identify semantically meaningful regions—such as vegetation, water bodies, and built-up areas—without using ground truth labels. The approach integrates classical image processing techniques, deep learning-based feature extraction using a pretrained ResNet18 model, and unsupervised clustering using KMeans.

2. Approach

The entire process is modular and includes the following key steps:

Image Preprocessing
- I used rasterio to load .tif satellite images.
- Normalized pixel values to the range [0, 1].
- Images were divided into 32x32 non-overlapping patches, ensuring consistent division across all dimensions.

Feature Extraction
- Each 32x32 patch was resized to 224x224 to match the input size expected by ResNet18.
- The final fully connected layer of the pretrained ResNet18 was replaced with an identity function to extract 512-dimensional features.
- To incorporate texture information, I also calculated the standard deviation of pixel values within each patch.
- The final feature vector for each patch was a concatenation of the 512 ResNet features and 1 standard deviation value → 513-dimensional.

Clustering and Label Mapping
- All patch features were passed to a KMeans clustering algorithm (k=4).
- The resulting labels were reshaped into a 2D label map based on the original patch layout and resized to the original image dimensions.

Visualization
- I visualized and saved the cluster maps using Matplotlib and OpenCV.
- An overlay of cluster maps on the original satellite image was created using cv2.addWeighted for better interpretation.

3. Assumptions

- Input images are 3-band RGB TIFFs, compatible with the ResNet18 architecture.
- The number of clusters (k=4) should ideally correspond to semantically meaningful classes like vegetation, buildings, roads, and water.
- Patch size (32x32) is a reasonable compromise between capturing fine-grained detail and maintaining spatial coherence.
- Nearby patches with similar features likely belong to the same semantic region.
- Due to the lack of ground truth, I rely solely on unsupervised evaluation metrics.

4. Challenges and Limitations

- Choosing k in KMeans: Small changes in k significantly impacted results.
- Roads were difficult to detect: Their thin, linear nature made it hard to isolate them using patch-based methods.
- No ground truth labels meant I couldn't use conventional evaluation metrics like IoU, Precision, or Recall.
- Memory and compute costs were high. Images produced over 140,000 patches, resulting in large matrices (e.g., 147,000 x 513) for clustering.
- Domain mismatch: ResNet18 is trained on natural RGB images (ImageNet), which might not be optimal for satellite imagery. Ideally, fine-tuning on domain-specific datasets like SpaceNet or BigEarthNet would improve results.

5. Results and Output Analysis

I tested the pipeline on three .tif satellite images: 1_1.tif, 1_2.tif, and 1_3.tif.

Metadata Summary:
- All images had 3 channels, used the EPSG:4326 CRS, and had consistent affine transforms.
- The patches extracted per image ranged from 129k to 147k.
- Noise was minimal in all images, as indicated by low standard deviation values.

Evaluation Metrics

Image	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index
1_2.tif	0.247	46083.30	1.6407
1_3.tif	0.216	39595.36	1.8181
1_1.tif	0.129	38961.33	2.1447

6. Analysis of Evaluation Metrics

To better interpret the clustering outcomes, I analyzed the three unsupervised evaluation metrics used in the report: Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index. Each of these metrics provides a different perspective on the quality of clustering.

Silhouette Score:
This metric measures how similar each patch is to its own cluster compared to other clusters. It ranges from -1 to +1, where values closer to 1 indicate that patches are well-matched to their assigned cluster and poorly matched to neighboring clusters. Values near 0 suggest overlapping or ambiguous cluster boundaries. Negative values indicate misclassified patches.
- 1_2.tif (0.247) has the highest Silhouette Score, suggesting it has the most distinct and compact clusters.
- 1_1.tif (0.129) has the lowest score, indicating weak cluster boundaries and potential overlap.

Calinski-Harabasz Index:
This score evaluates the ratio of between-cluster dispersion to within-cluster dispersion. Higher values indicate better clustering, as they reflect more compact and well-separated clusters.
- 1_2.tif (46083.30) again leads with the best-defined clusters.
- 1_1.tif has the lowest score, showing weaker definition between clusters.

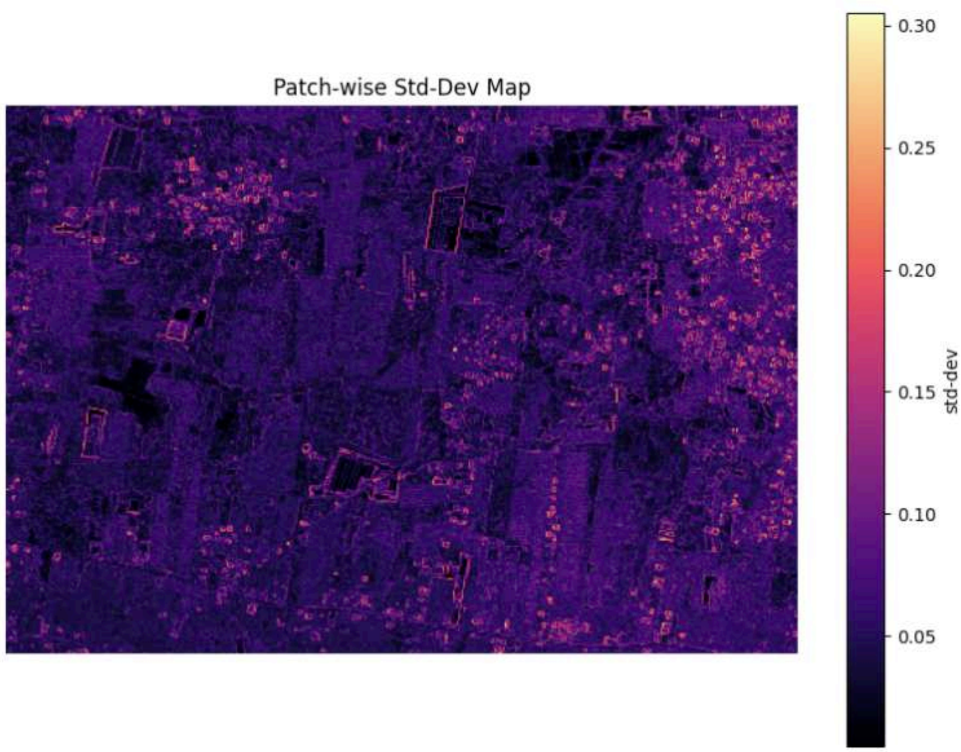
Davies-Bouldin Index:
This index measures the average similarity between each cluster and its most similar cluster. Lower values are preferred as they represent better separation and more compact clusters.
- 1_2.tif (1.6407) has the best Davies-Bouldin score.
- 1_1.tif (2.1447) indicates poorer separation and more within-cluster variance.

Overall, 1_2.tif consistently outperforms the other images across all metrics. The results also suggest that clustering quality is sensitive to the content of the input image and that incorporating both deep features and texture cues enhances segmentation performance.

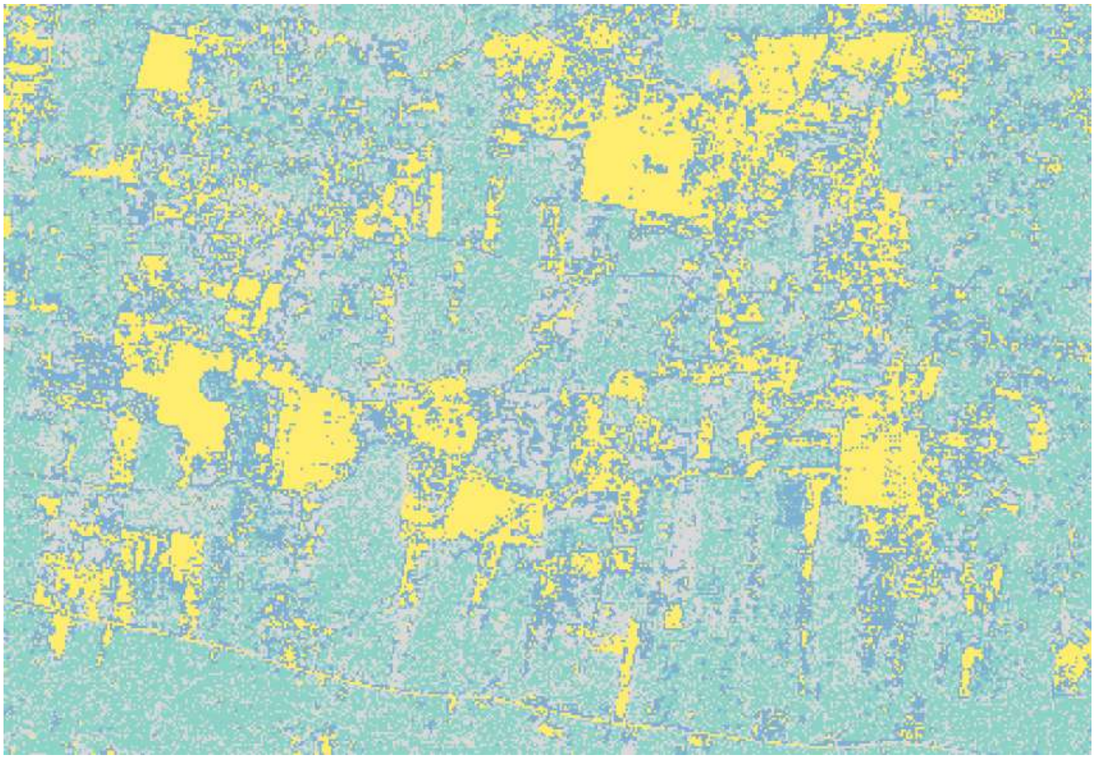
7. Output Images

Input image: 1_1.tif

Standard Deviation MAP



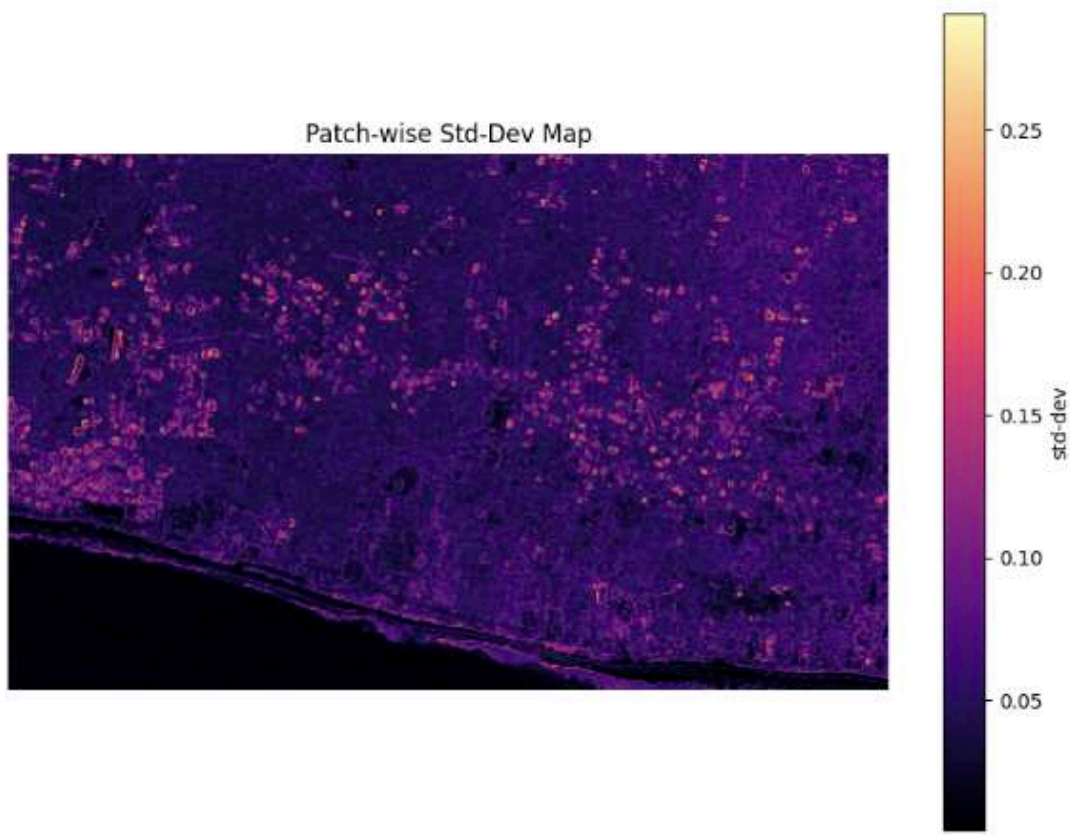
Cluster image of 1_1



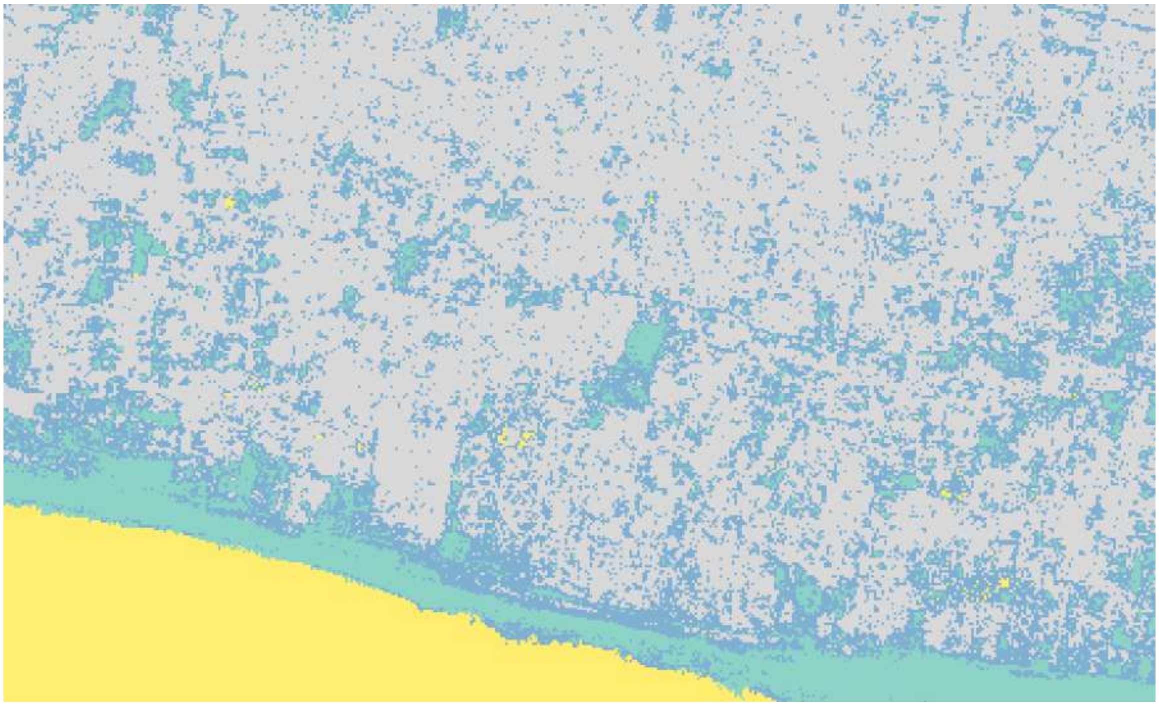
Overlay MAP of 1_1

Input image: 1_2.tif

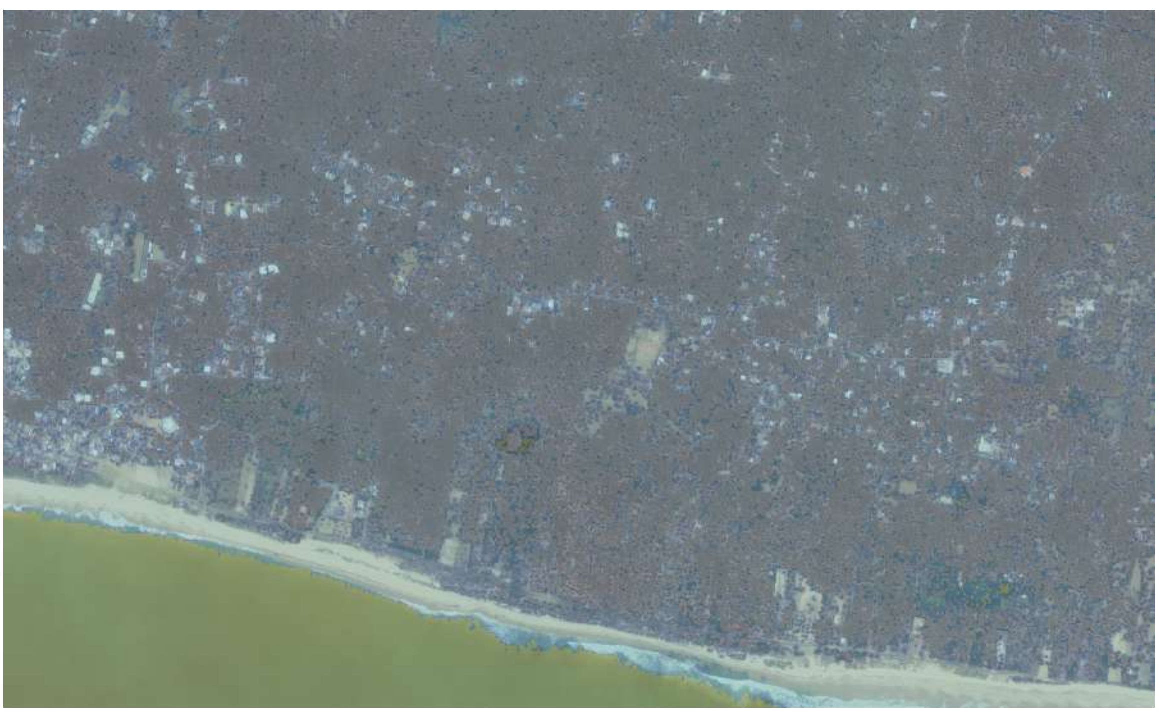
Standard Deviation MAP



Cluster image of 1_2

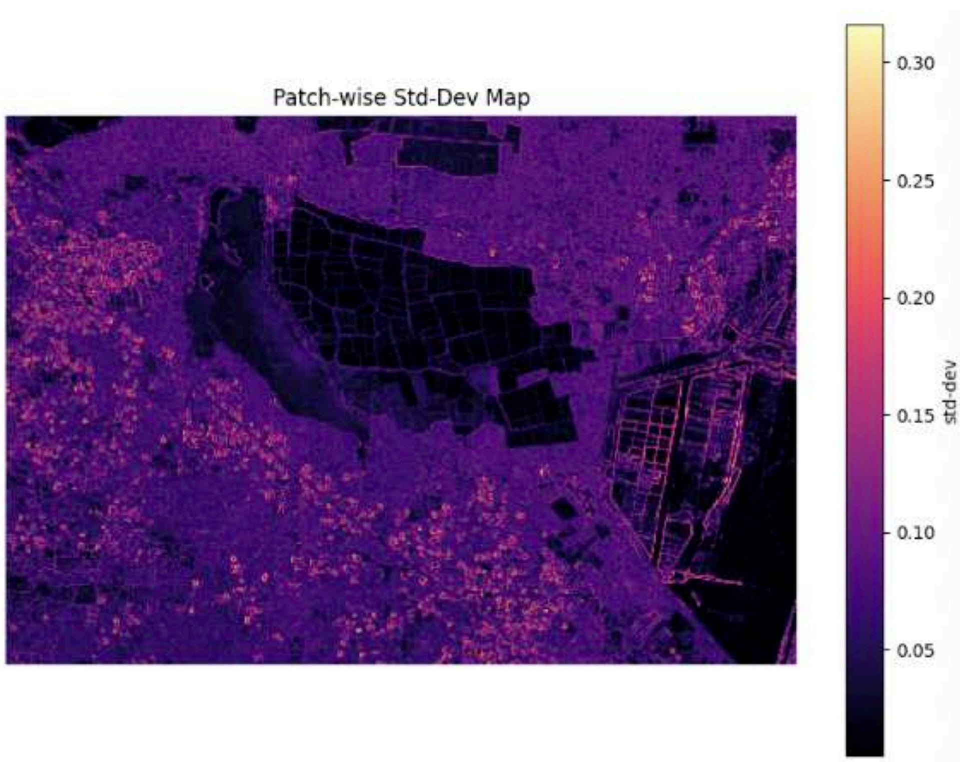


Overlay MAP of 1_2

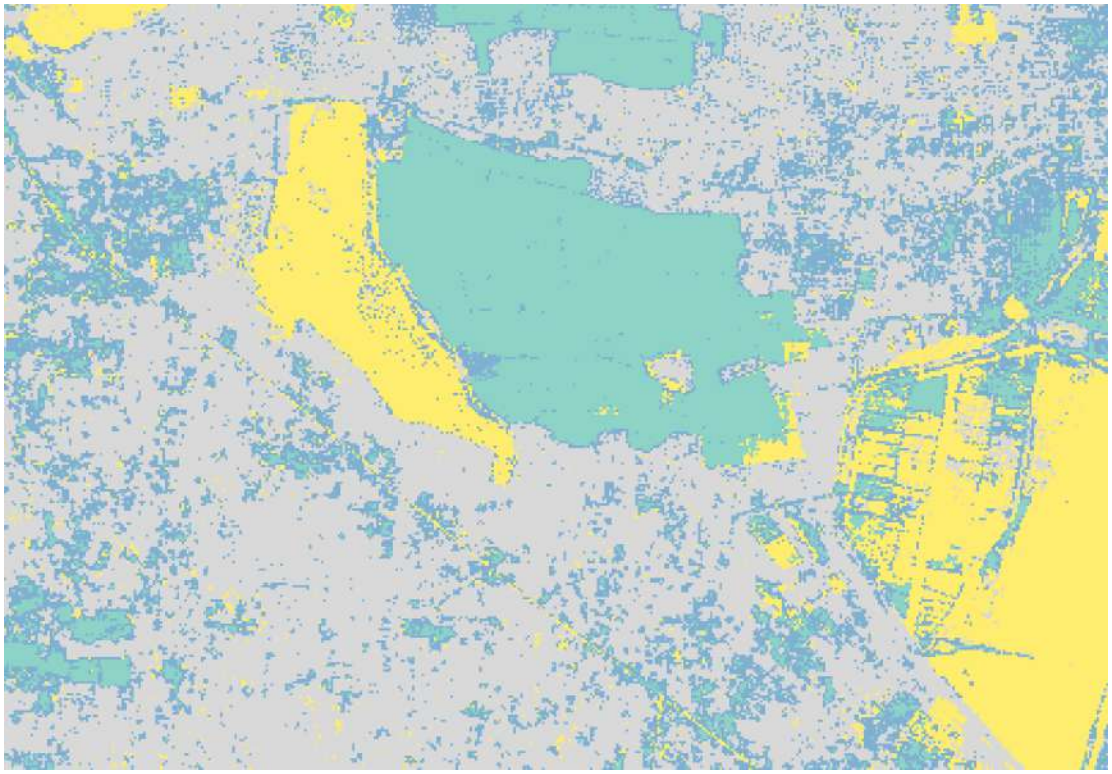


Input image: 1_3.tif

Standard Deviation MAP



Cluster image of 1_3



Overlay MAP of 1_3