

Assignment 4
AIL862 SPECIAL TOPICS IN COMPUTER VISION
Kashish Srivastava (2024AIB2289)
26th March 2025

Unsupervised Image Segmentation using DINOv2 and K-Means Clustering

1. Introduction

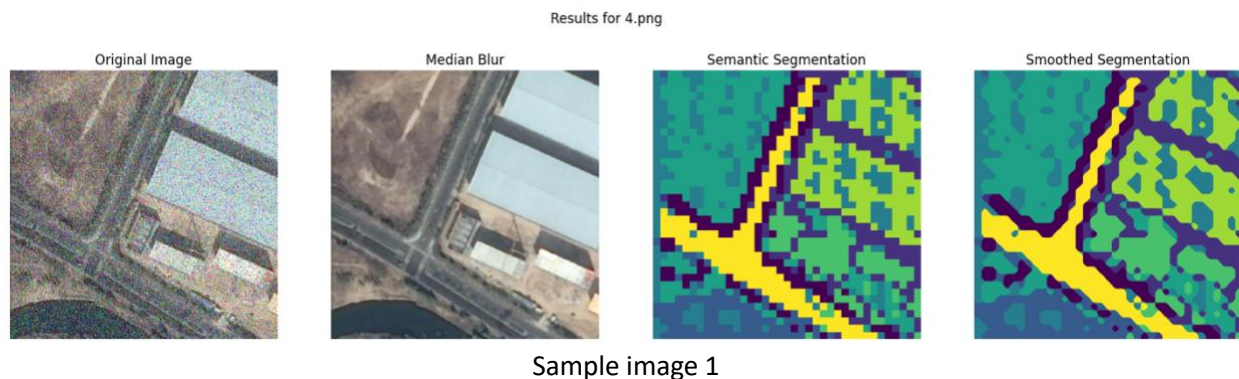
This report presents an image segmentation approach using Facebook's DINO-v2 (Vision Transformer) model combined with K-Means clustering. The method leverages self-supervised learning and unsupervised clustering to perform semantic segmentation on images without labeled training data.

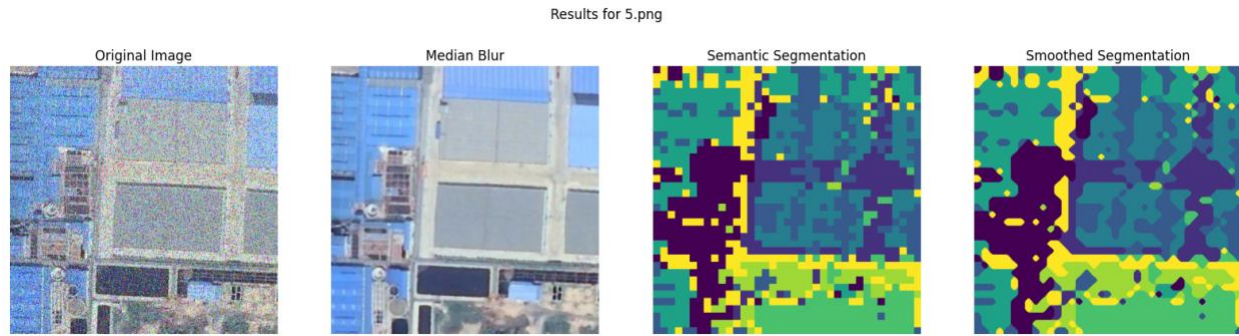
2. Methodology

The image segmentation process involves several key steps:

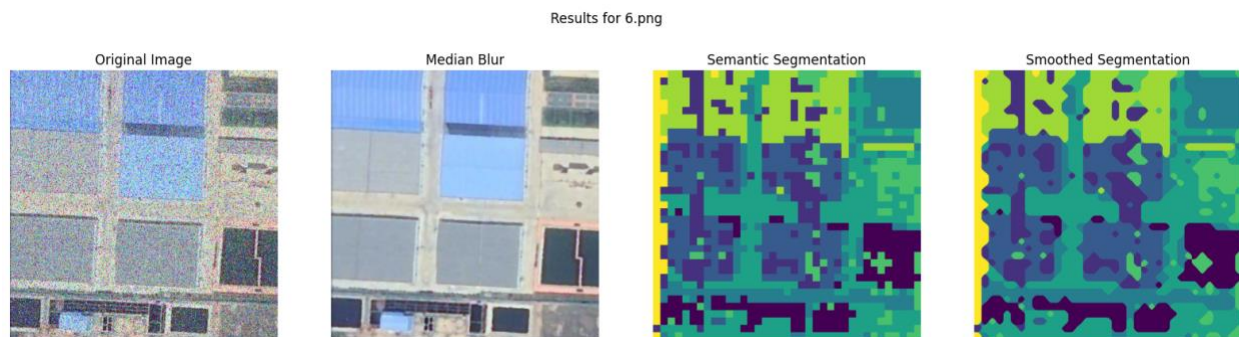
- a. **Model Initialization:** Loading the DINO-v2 Vision Transformer (ViT-L/14).
- b. **Preprocessing:** Resizing, applying median blur, and normalizing images.
- c. **Feature Extraction:** Passing preprocessed images through DINO-v2 and extracting patch tokens.
- d. **Dimensionality Reduction:** Applying PCA to reduce feature dimensionality. 99% variance captured.
- e. **Clustering:** Performing K-Means clustering on reduced features.
- f. **Post-processing:** Resizing and smoothing the segmentation map.

3. Results





Sample image 2



Sample image 3

The figure above illustrates the results of the semantic segmentation process for some input images. The following observations can be made:

1. Original Image: It contains visible noise that may affect segmentation quality.
2. Median Blur: After applying a median blur filter, the noise in the image is significantly reduced, resulting in a smoother version of the original image. This preprocessing step helps reduce Salt-and-pepper noise.
3. Semantic Segmentation: The segmentation map identifies distinct regions in the image using different colors, representing clusters of similar features.
4. Smoothed Segmentation: The smoothed segmentation map refines the boundaries between regions, producing a more coherent and visually appealing result. This step reduces noise and irregularities in the initial segmentation.

4. Conclusion

This approach demonstrates the potential of combining self-supervised learning with unsupervised clustering for semantic segmentation. It offers a flexible solution applicable to various image types without requiring labeled data.

5. **Python code link:** <https://owncloud.iitd.ac.in/nextcloud/index.php/s/5TCzRg42kKQE8D4>