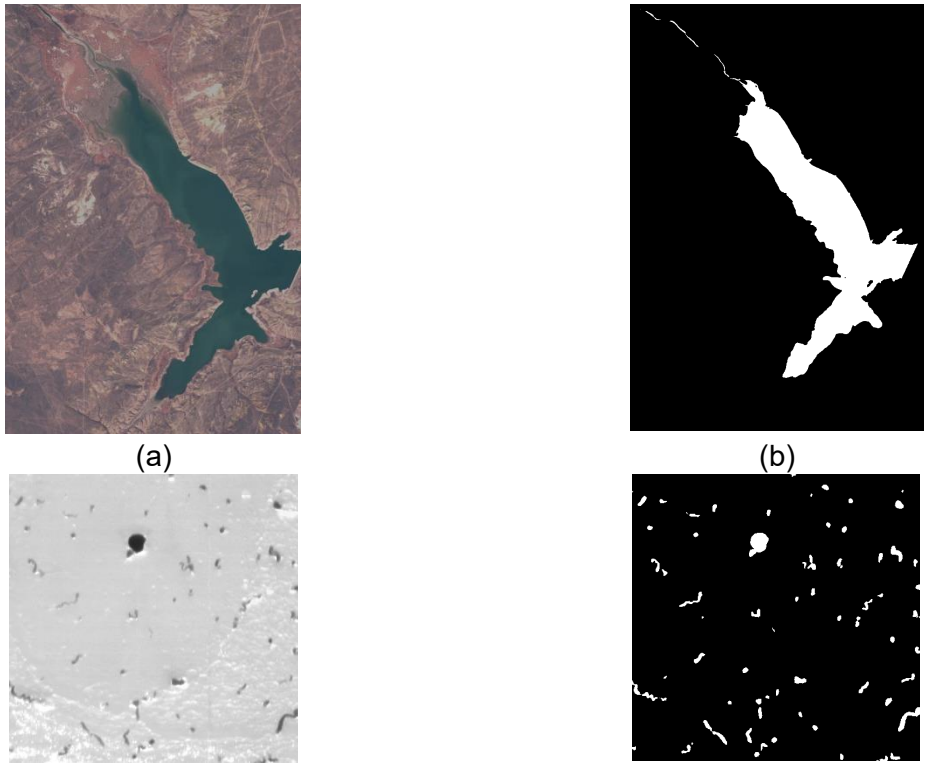
Bishop's University
Department of Computer Science
CS462/562
HW4 – Image Segmentation

Objective:
To understand and implement clustering-based image segmentation techniques.

Data: You are provided with four images accompanied by their respective ground-truth masks. Three of these images pertain to remote sensing, while the fourth resembles a microscopic image. Sample illustrations are presented in Figure 1.



(a) (b)

(c) (d)

Figure 1. Original images (left) and their Mask (right).

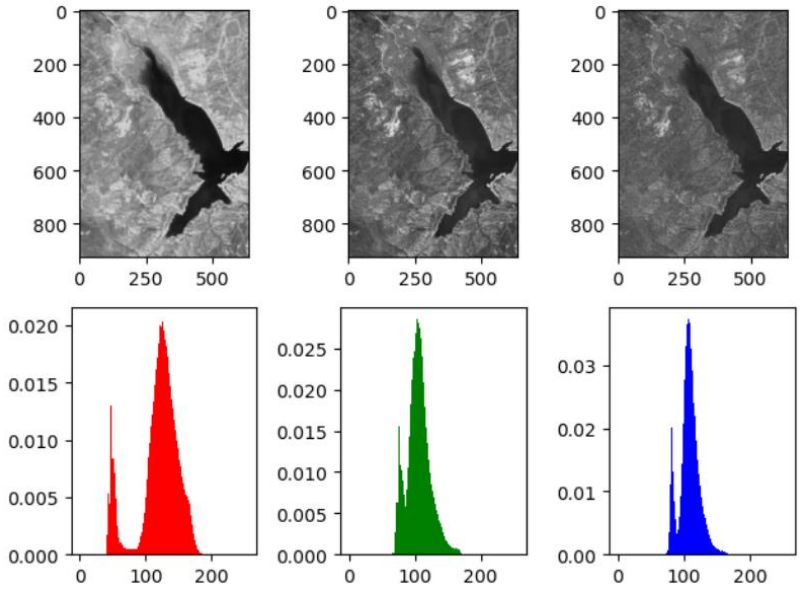


Figure 2. Top: Red, green, and blue channels of the image displayed in Figure 1(a). Bottom: Corresponding intensity histograms for each channel.

Tasks (additional details are provided in the accompanying Jupyter Notebook.**)**:

Research: Conduct research on clustering-based image segmentation techniques, focusing on algorithms like K-means clustering, Fuzzy C-means clustering. Understand how these algorithms work and their applications in image processing.

Algorithm Implementation: Choose K-means clustering and implement it using a programming language of your choice (e.g., Python). Use sample images provided to you.

Experimentation:
- Apply your implemented clustering algorithm to various images.
- Experiment with various parameter settings and analyze their impact on the segmentation results. In the case of k-means clustering, the parameter of interest is the number of clusters denoted by 'k.' For our specific task, we set k=2 to segment the input image into two clusters, highlighting the objects of interest for detection.
- The k-means algorithm requires setting initial centers, typically chosen randomly. However, for improved initialization, you can utilize the histogram to approximate the center of each mode, representing a potential cluster. For example, Figure 2 shows the histogram of intensity distribution in each channel. For the figure, the channel Red shows a better separation between two important modes, the big mode on the right represents the land (brighter) and the second shown on the left represents the water body (darker).

Mode-based analysis for better center initialization:
- Analyzing the histogram provides insight into the number of objects in the image. For instance, in the histogram of the red channel depicted in Figure 2, two modes are observed, indicating two potential objects. The left mode corresponds to the darker object, representing the water body, while the mode on the right denotes the land. By applying a threshold near 92, the

separation of these modes aids in detecting the water body, as illustrated by the mask in Figure 1(b).

- Based on this analysis, we can utilize the centers of each mode to initialize the clusters. The center of the first mode is approximately 80, while the center of the second mode is around 128. These values serve as effective initializations for the k-means clustering algorithm.

Submission Guidelines: Submit your implemented code in Pythion Notebook.

References:

- Images 1m1, im2, im3 are taken from https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies

- Image im4 is taken from https://github.com/Soumyabrata/HYTA