

NUCLEI DETECTION AND SEGMENTATION IN COLORECTAL TISSUE SAMPLES

BACKGROUND

Whole slide image (WSI) annotation is currently time-consuming, inconsistent, and prone to errors due to its manual nature. The large image size and tissue variability often result in incomplete or inaccurate labels, making automated models less reliable. Automated methods also struggle to detect small objects like nuclei while over-segmenting larger regions.

ABSTRACT

Our research presents a program to automate cell nuclei annotation in colon tissue Whole Slide Images (WSI), a task currently done manually by medical professionals. The model addresses challenges like noise and overlapping structures, offering accurate detection and strong generalization. It enhances diagnostic efficiency and accuracy, reducing manual workload in digital pathology.

METHODOLOGY

Part 1: Setting the image

Input Image Loading and Conversion

- Converts to a gray and colored image

Preprocessing

- Removes the darkest and lightest values to reduce the impact of high contrast values
- Scales image to an 8-bit range

Extracting Specific Color Clusters

- Extracts the darkest cluster and isolates it

Otsu's Thresholding

- Implements Otsu's method which separates the foreground from the background based on the intensity values.

Part 2: Processing the image

Applying KMeans Clustering

- Groups image pixels into clusters based on intensity values.
- Targets dark pixels to potentially identify nuclei.

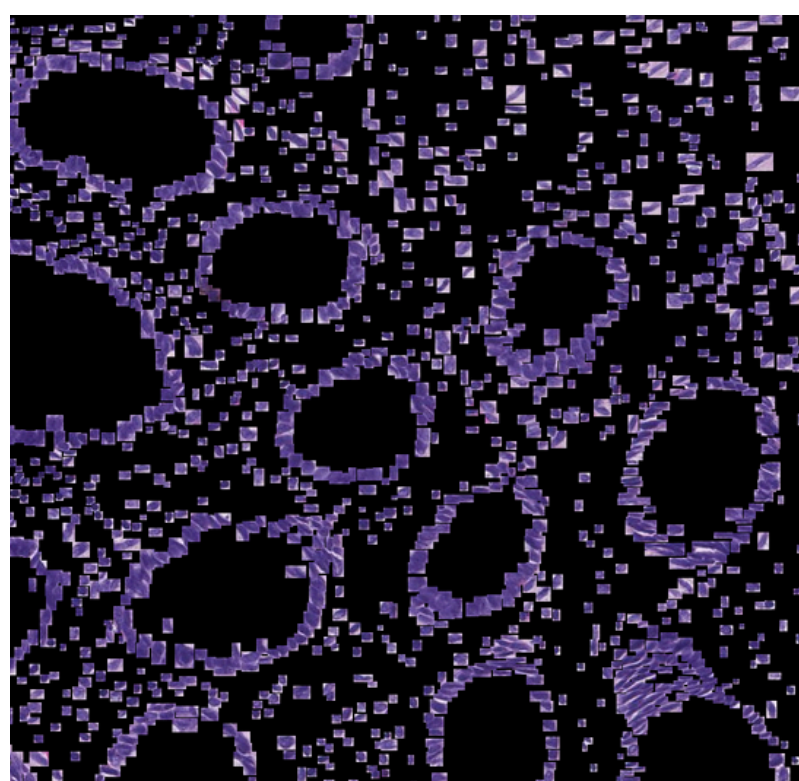
Filtering Unwanted Objects

- Identifies non-nuclei objects by size and appearance.
- Develops code to group pixels into distinct objects, classifying them as nuclei or otherwise.

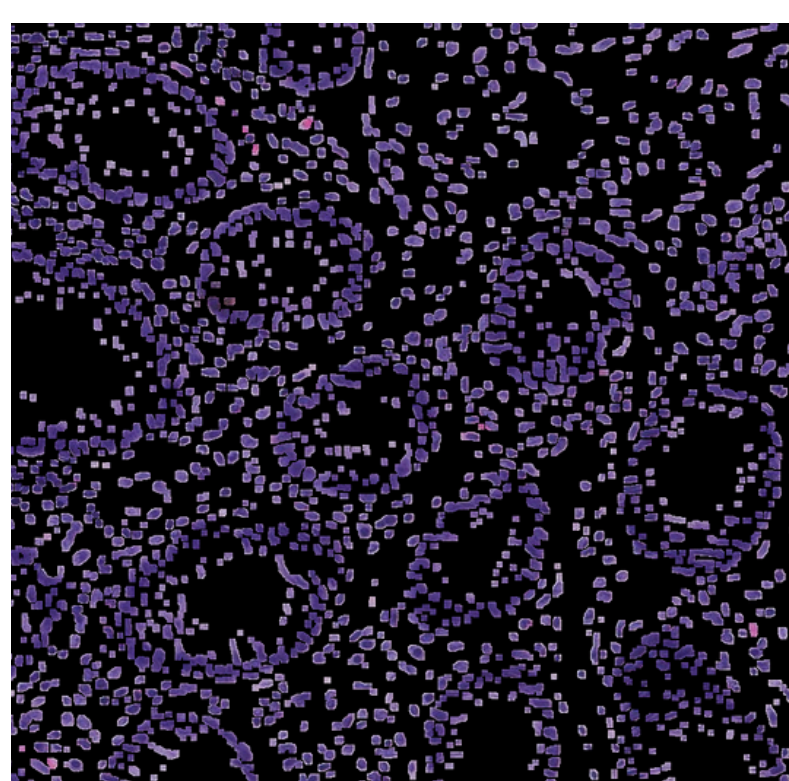
Object Size-Based Selection

- Filters objects based on size to isolate smaller, likely nuclei objects.
- Returns many small, dark regions classified as nuclei.

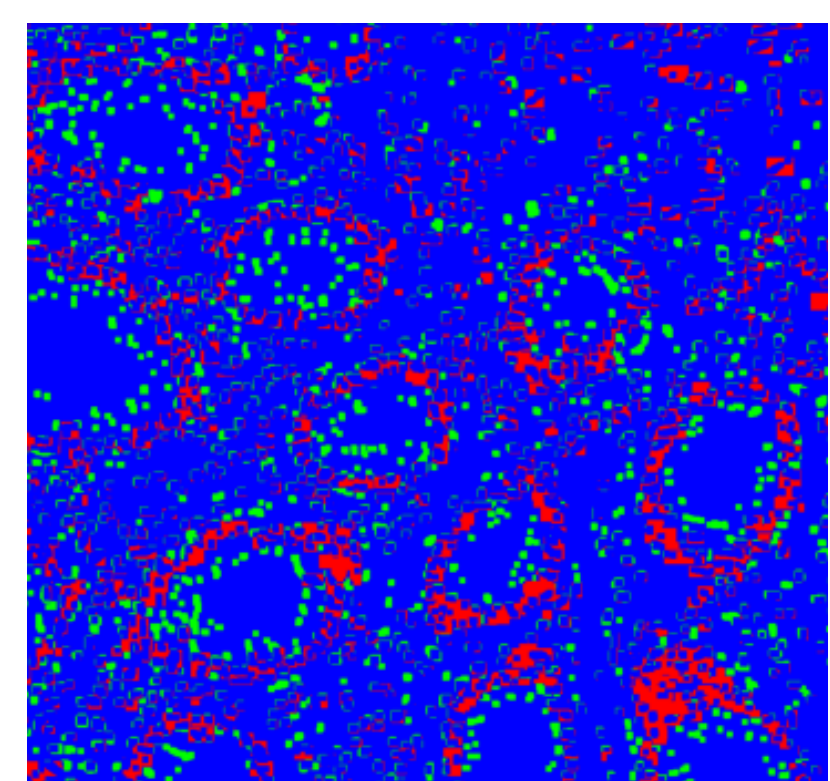
ANALYSIS



Masked



Segmented



Overlap accuracy

RESULTS

Our program achieved a 78% accuracy rate in detecting and classifying objects within the processed images. This indicates that the majority of objects were correctly identified and categorized, though some misclassifications or missed detections occurred. The program's performance can be considered effective for general use, but there is room for improvement, particularly in refining the detection of more challenging or less distinct objects.

