ASSIGNMENT

COMPUTER VISION – 22AIE313

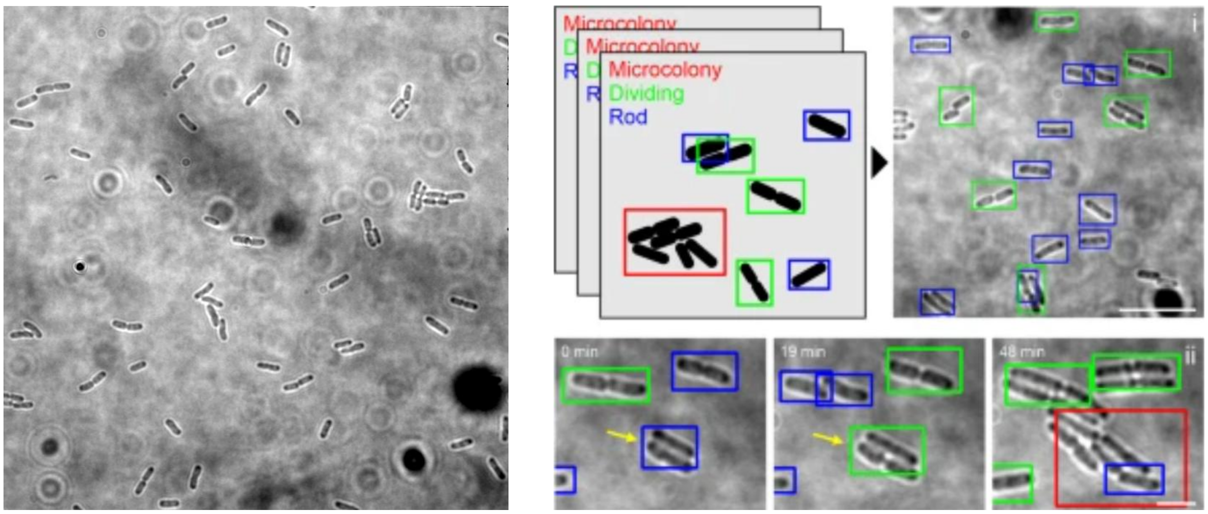
Submitted by
Pradeep raj D – CH.EN.U4AIE22042
Shyam Prasath J – CH.EN.U4AIE22042

March 14, 2025

Problem Statement

The dataset comprises *Escherichia coli* (*E. coli*) images obtained from the DeepBacs study. The objective is to employ image processing techniques to segment individual bacterial cells and classify them into three growth stages:

1. **Rod-shaped cells:** Standard morphology of individual *E. coli* bacteria.
2. **Dividing cells:** Bacteria undergoing binary fission.
3. **Microcolonies:** Clusters of four or more closely associated cells.



Accurately identifying these stages is essential for understanding bacterial growth dynamics and responses to various conditions.

Importance of the Study

Accurate segmentation and classification of bacterial growth stages are vital in microbiology and related fields. They enable:

- **Antimicrobial Research:** Monitoring morphological changes in response to antibiotics facilitates the assessment of drug efficacy and the development of new treatments.
- **Pathogenesis Studies:** Understanding growth patterns aids in elucidating infection mechanisms and bacterial proliferation within hosts.
- **Biotechnology Applications:** Optimizing bacterial cultures for industrial processes, such as fermentation, relies on precise growth stage identification.

Challenges in the Dataset

The DeepBacs *E. coli* dataset presents several challenges:

1. **Imaging Noise and Artifacts:** Bright-field microscopy images may contain background noise, uneven illumination, and artifacts, complicating accurate segmentation.
2. **Cell Overlapping and Clustering:** Bacteria often appear in clusters or overlap, making it difficult to distinguish individual cells.
3. **Morphological Variability:** *E. coli* cells exhibit diverse shapes during different growth stages, requiring adaptable segmentation algorithms.
4. **Low Contrast:** Minimal contrast between cells and the background in bright-field images poses challenges for standard segmentation techniques.
5. **Small Object Size:** The diminutive size of *E. coli* cells necessitates high-resolution imaging and precise detection methods.

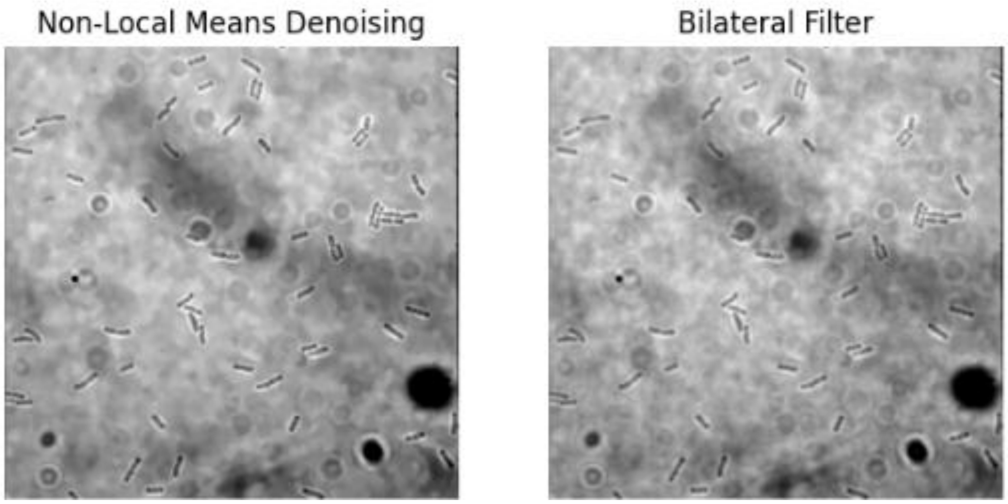
Purpose of the Problem Statement

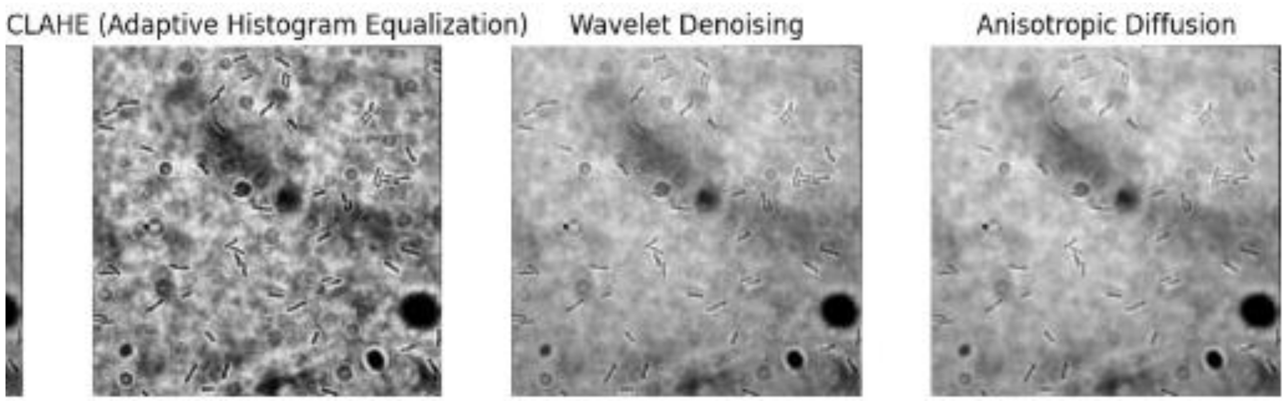
This study aims to:

- **Automate Bacterial Analysis:** Develop algorithms to efficiently and accurately segment and classify *E. coli* cells, reducing reliance on manual analysis.
- **Enhance Research Efficiency:** Provide tools for rapid assessment of bacterial growth stages, supporting studies in antimicrobial resistance and bacterial physiology.
- **Improve Image Processing Techniques:** Advance methods for handling noisy, low-contrast images with overlapping small objects, benefiting broader applications in biomedical imaging.

By addressing these objectives, the study contributes to a deeper understanding of bacterial behavior and supports the development of automated systems for microbiological research and clinical diagnostics.

Noise Reduction:





Inference from above filtering methods:
1.Guassian blur and median blur have little to no effects.

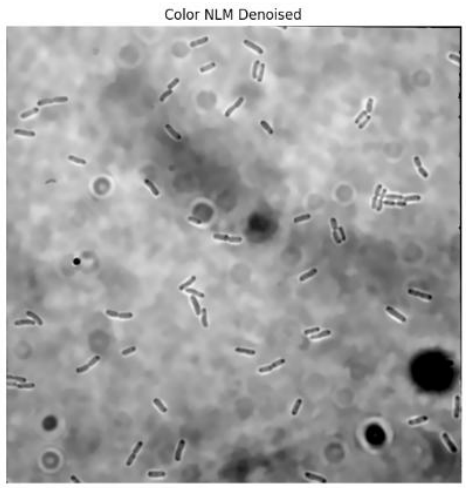
2.Non-local means and bilateral filter methods have minimal effects.

3.Frequence based methods like wavelet denoising too had minimal effects.

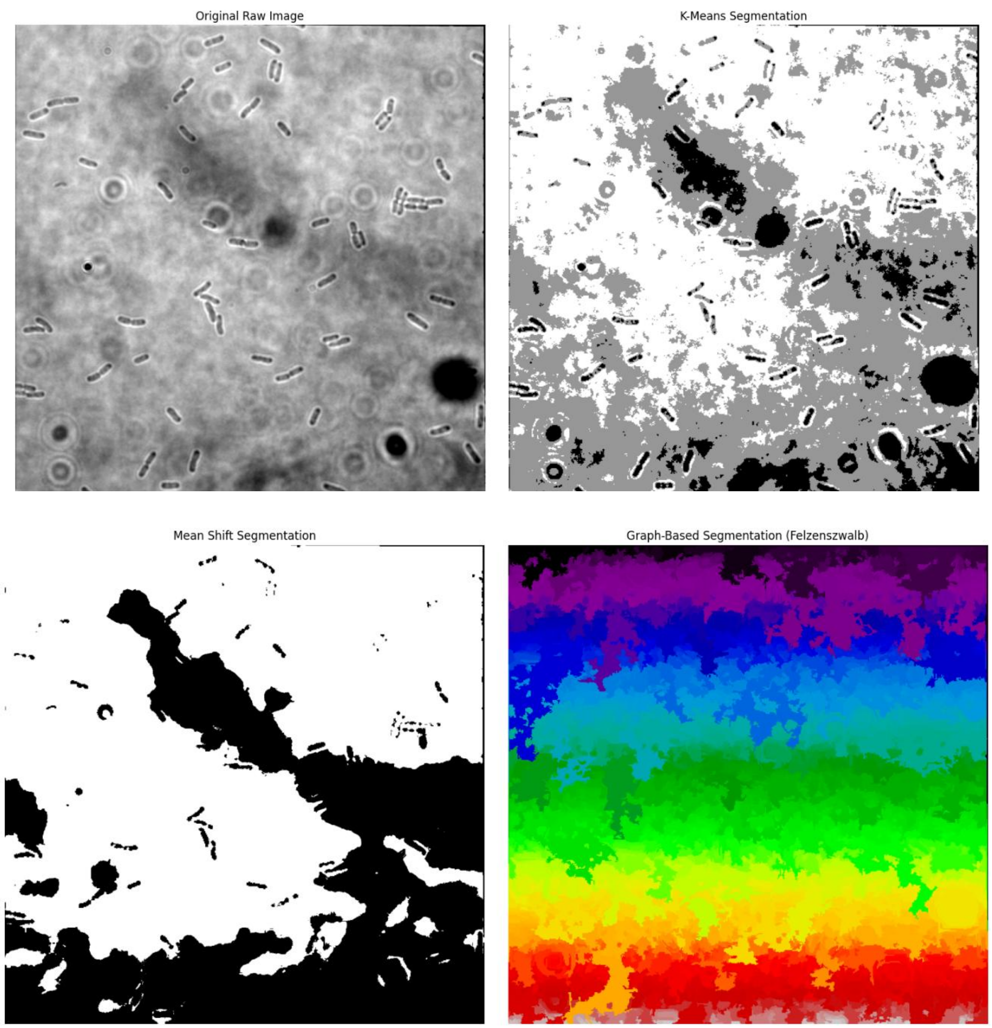
4.CLAHE - adaptive histogram equalization made the case worse as it increases the contrast overall, but since the original image has less contrast between background and the subject, it did not help to differentiate.

Best method: NLM, below is the result after adjusting the parameters of Non local means method:

This method although has applied denoising well, this does not have impact on the subject separation as it smoothens the subject also.

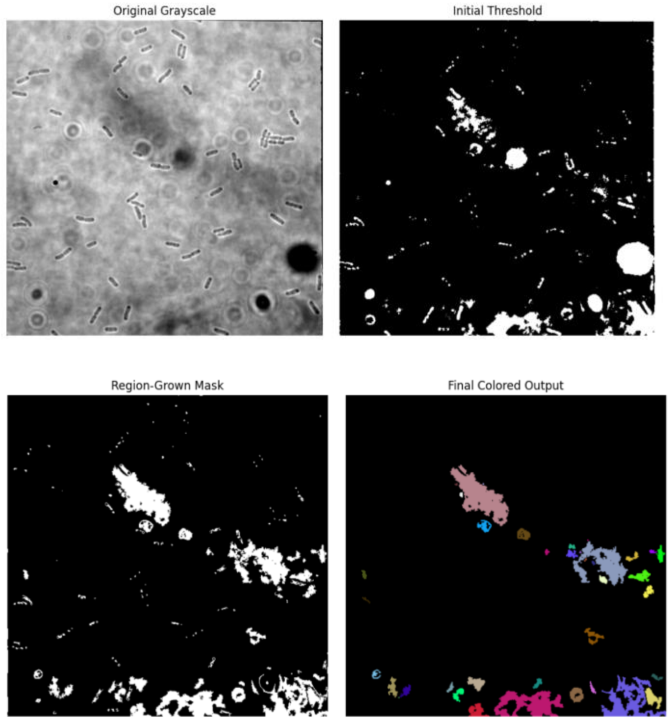


Basic Segmentation algorithms:



Inference: K-means segmentation barely worked, while meanshift and graph based segmentation could not do anything as the image has very bad subject separation and noise.

Region growing algorithm:



Inference: The images taken and the task is too complex for region growing algorithm to work. Connected Component analysis is used in below hybrid approach to refine segmentation.

Hybrid methodology 1:

1.Apply Adaptive Thresholding

2.Morphological Operations (Noise Removal)

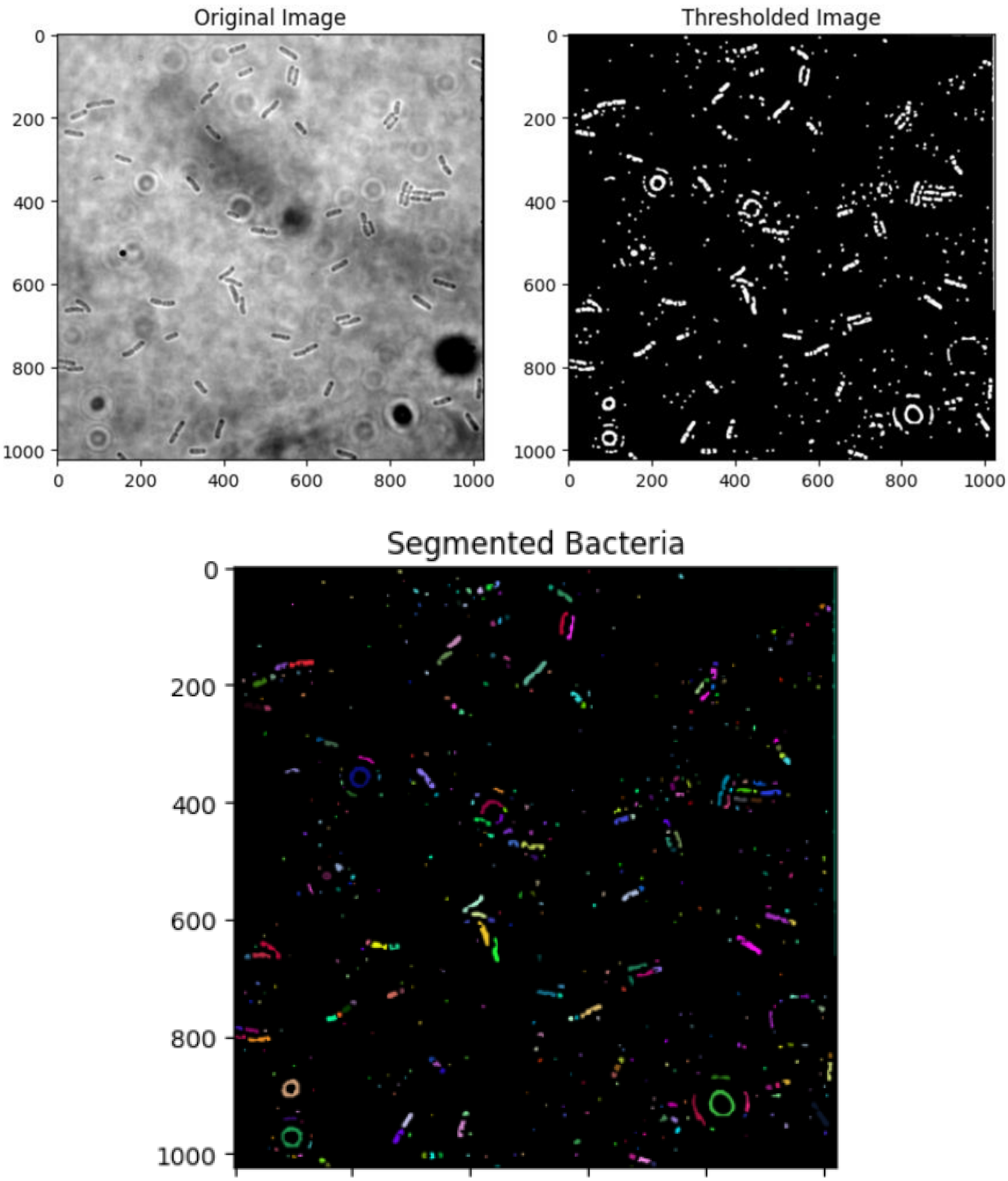
3.Distance Transform for Separation

4.Define Unknown Region

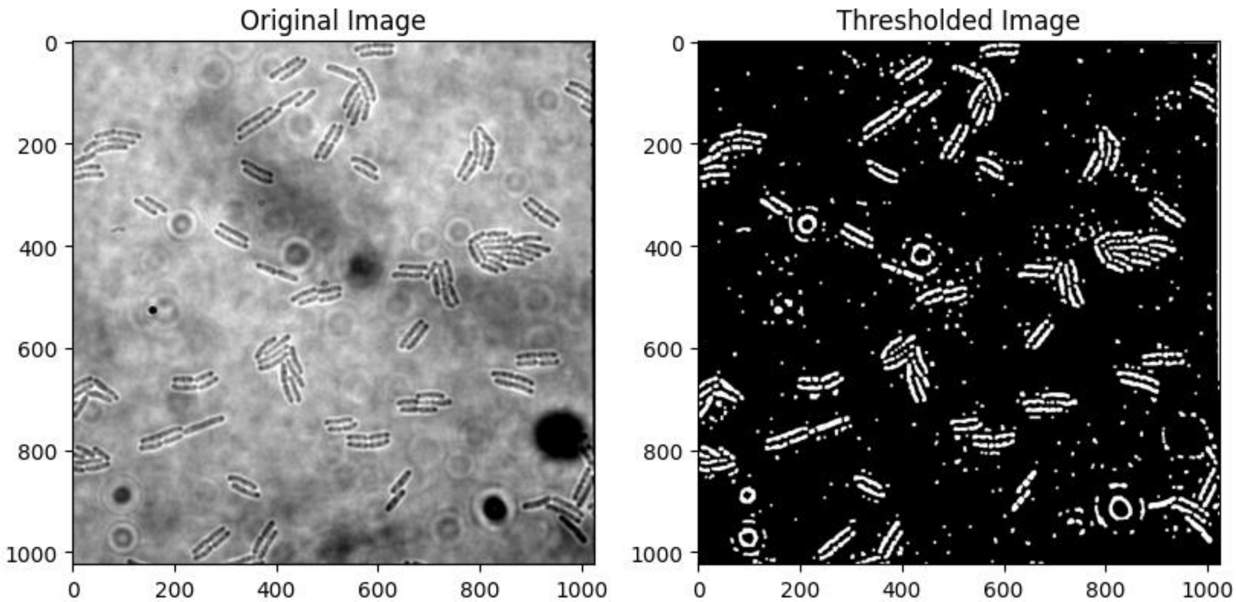
5.Marker-Based Segmentation

6.Apply Watershed Algorithm

7.Assign Unique Colors to Bacteria


Original Image
Thresholded Image
Segmented Bacteria
0
200
400
600
800
1000

Another sample with more grown stage of bacteria:



Problems in above method:

1. **Over-segmentation in Watershed Output**
    - Problem: The Watershed algorithm may create too many small, unnecessary regions.
    - Solution: Convert the watershed markers into a binary mask, keeping only bacteria regions.
2. **Noise and False Detections**
    - Problem: Tiny regions caused by noise or background irregularities can be falsely detected as bacteria.
    - Solution: Connected Component Analysis (CCA) is used to identify and label distinct bacterial regions.
3. **Small Non-Bacterial Artifacts**
    - Problem: Debris, noise, or segmentation errors create small objects that shouldn't be counted as bacteria.
    - Solution: Removing small objects by setting a minimum area threshold eliminates these unwanted regions.
4. **Inconsistent Labeling After Object Removal**
    - Problem: After filtering out small objects, gaps in label numbering can cause inconsistencies in further analysis.
    - Solution: Reassign labels sequentially to maintain consistency and correctness.

Rule based approach to solve the above issues:

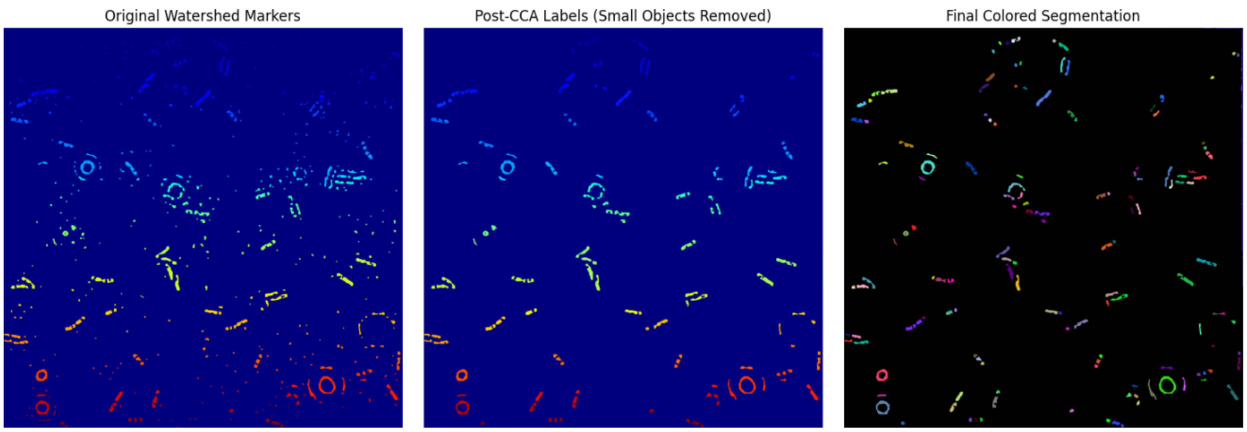
1.Create a Binary Mask from Watershed 'markers' – Converts watershed segmentation output into a binary mask where bacteria are labeled as 1, and everything else is 0.

2.Connected Component Analysis (CCA) – Identifies individual bacterial regions by labeling connected components in the binary mask.

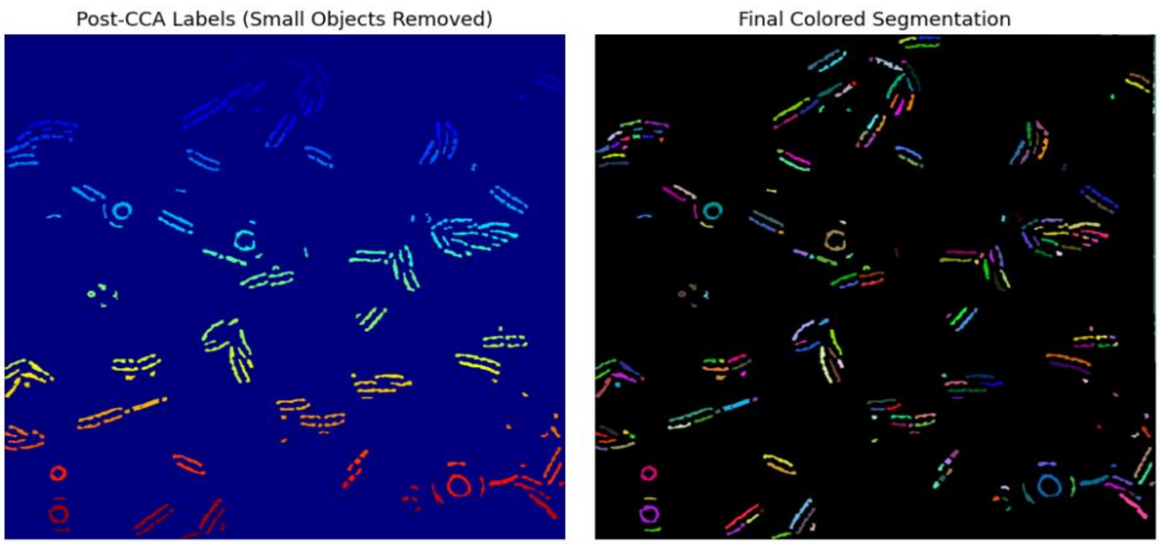
3.Remove Small Objects by Area – Filters out small noise or irrelevant objects by setting a minimum area threshold.

4.Reassign Labels After Removal – Re-indexes the remaining connected components to ensure proper labeling after removing small objects.

5.Color the Final Labels – Assigns random colors to the segmented bacterial regions for better visualization.



Another sample of mature stage:



Further refinement:

Problems:

1.Non-Rod-Shaped Objects (Circles & Irregular Shapes)

- Problem: Bacterial rods have an elongated shape, but segmentation can include round or irregular objects.
- Solution: Filter objects based on aspect ratio to keep rod-like shapes.

2.Donut-Shaped or Hollow Objects (Artifacts)

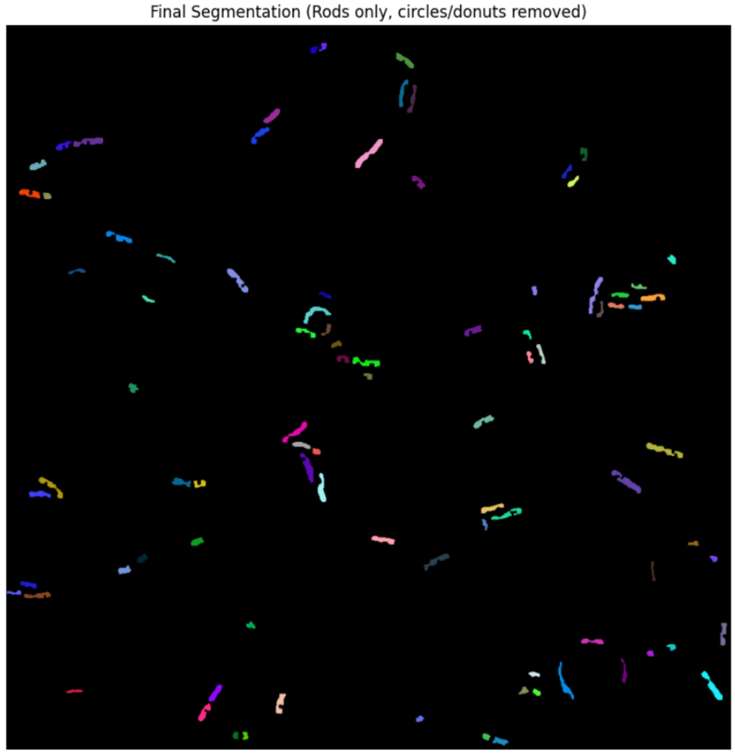
- Problem: Some artifacts appear as rings or incomplete objects, which should be removed.
- Solution: Use solidity (area/convex area ratio) to exclude objects with hollow structures.

More rule based filtering:

Steps Taken in the Code

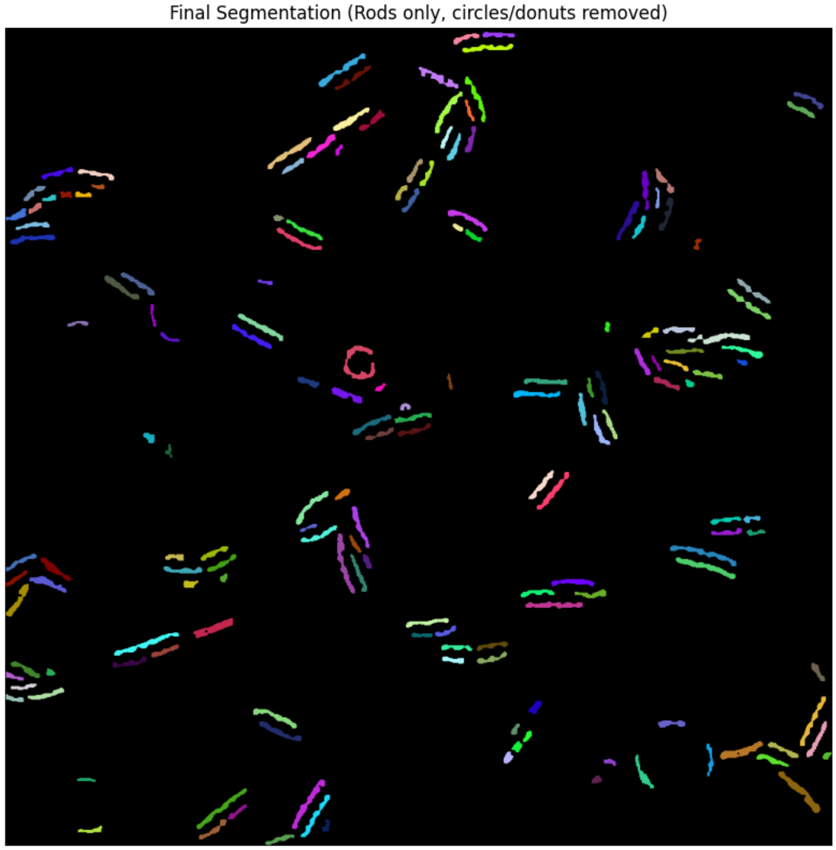
1. **Set Filtering Parameters**
   - **Define minimum area, aspect ratio, and solidity thresholds for object selection.**
2. **Filter Small Objects**
   - **Remove segmented objects smaller than MIN_AREA to eliminate noise.**
3. **Filter Non-Rod-Like Objects**
   - **Compute aspect ratio to exclude circular or compact shapes.**
4. **Filter Donut-Shaped Artifacts**
   - **Use solidity to remove hollow or ring-like objects.**
5. **Re-label the Filtered Objects**
   - **Perform Connected Components Analysis (CCA) to reassign labels after filtering.**

Final Segmentation (Rods only, circles/donuts removed)



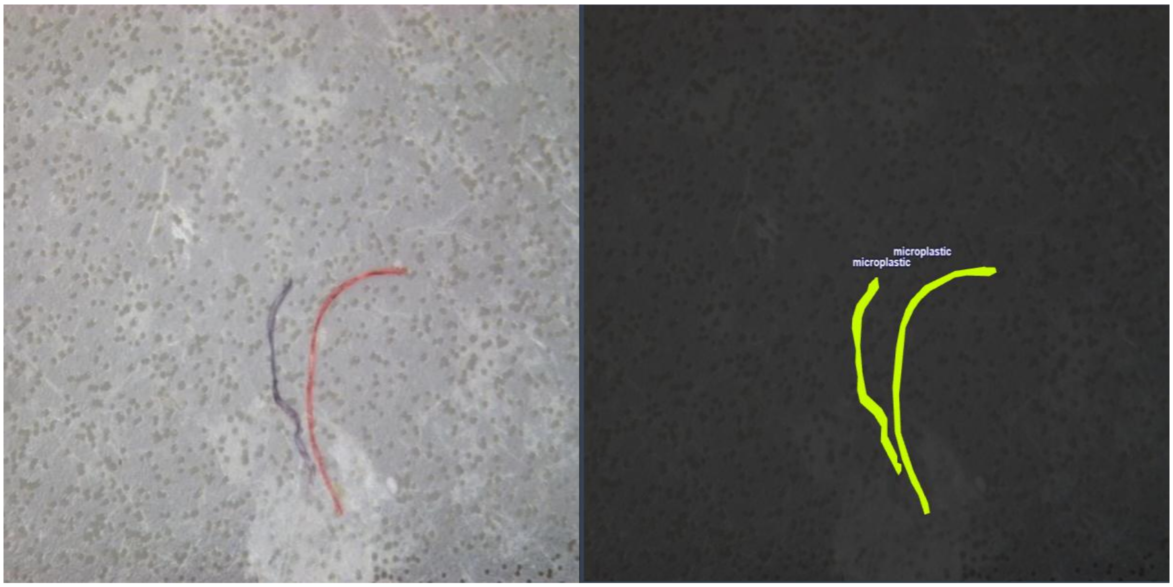
Another sample:

Final Segmentation (Rods only, circles/donuts removed)



Secondary Problem Statement: Segmenting Microplastics in Microscopic Images

Microplastics, defined as plastic particles less than 5mm in size, have become a significant environmental concern due to their widespread presence in water bodies, soil, and even the food chain. Detecting and segmenting these microscopic particles from complex backgrounds in microscopic images is a challenging task. Aim is to develop an effective segmentation pipeline that can accurately isolate microplastics from other particles and noise in microscopic images.



Importance of This Study

- **Environmental Impact:** Microplastics pose a severe threat to marine life and ecosystems, as they can be ingested by aquatic organisms, leading to bioaccumulation and toxic effects.
- **Human Health Concerns:** Studies have shown microplastics in drinking water and food, raising concerns about their long-term impact on human health.
- **Automated Analysis for Efficiency:** Current methods of microplastic detection rely on manual analysis or expensive spectroscopic techniques. An automated segmentation approach using computer vision can provide a faster, cost-effective alternative for large-scale monitoring.

Challenges in Microplastic Segmentation

- **Complex Backgrounds:** Microscopic images contain various particles, debris, and biological matter, making it difficult to distinguish microplastics from other objects.
- **Varied Shapes and Sizes:** Microplastics are irregular in shape and can be transparent, making segmentation more challenging than traditional object detection.
- **Lighting and Image Noise:** Variations in lighting conditions and the presence of noise in microscopic images can affect segmentation accuracy.

- **Overlapping Particles:** Microplastics often cluster together or overlap with other elements, requiring advanced techniques to separate them correctly.

Purpose of This Study

This study aims to develop a computer vision pipeline from scratch to automate the segmentation of microplastics from microscopic images. The key objectives include:

- Implementing denoising and enhancement techniques to improve image quality.
- Utilizing edge detection, clustering, and morphological operations to accurately segment microplastics.
- Evaluating the effectiveness of different segmentation techniques, including Frangi filters, K-means clustering, and contour-based methods.
- Creating a reliable and scalable model that can be applied to large datasets for environmental monitoring and research.

By automating microplastic segmentation, this study contributes to developing faster and more efficient tools for environmental analysis, ultimately aiding in pollution control and regulatory efforts.

Methodology Overview :

Hybrid approach 2:

1.Apply Color Non-Local Means (NLM) Denoising

- **Reduce noise while preserving texture and color features.**

2.Convert to Grayscale for Edge Detection & Frangi Filtering

- **Prepare the image for feature extraction.**

3.Enhanced Edge Detection with Canny & Dilation

- **Detect boundaries and strengthen edges using dilation.**

4.Apply Frangi Filter for Structure Enhancement

- **Highlight tubular structures like microplastics.**

5.Combine Segmentation Approaches

- **Merge edge detection and Frangi filter results.**

6.Morphological Refinement

- **Use closing operations to remove small holes and refine the mask.**

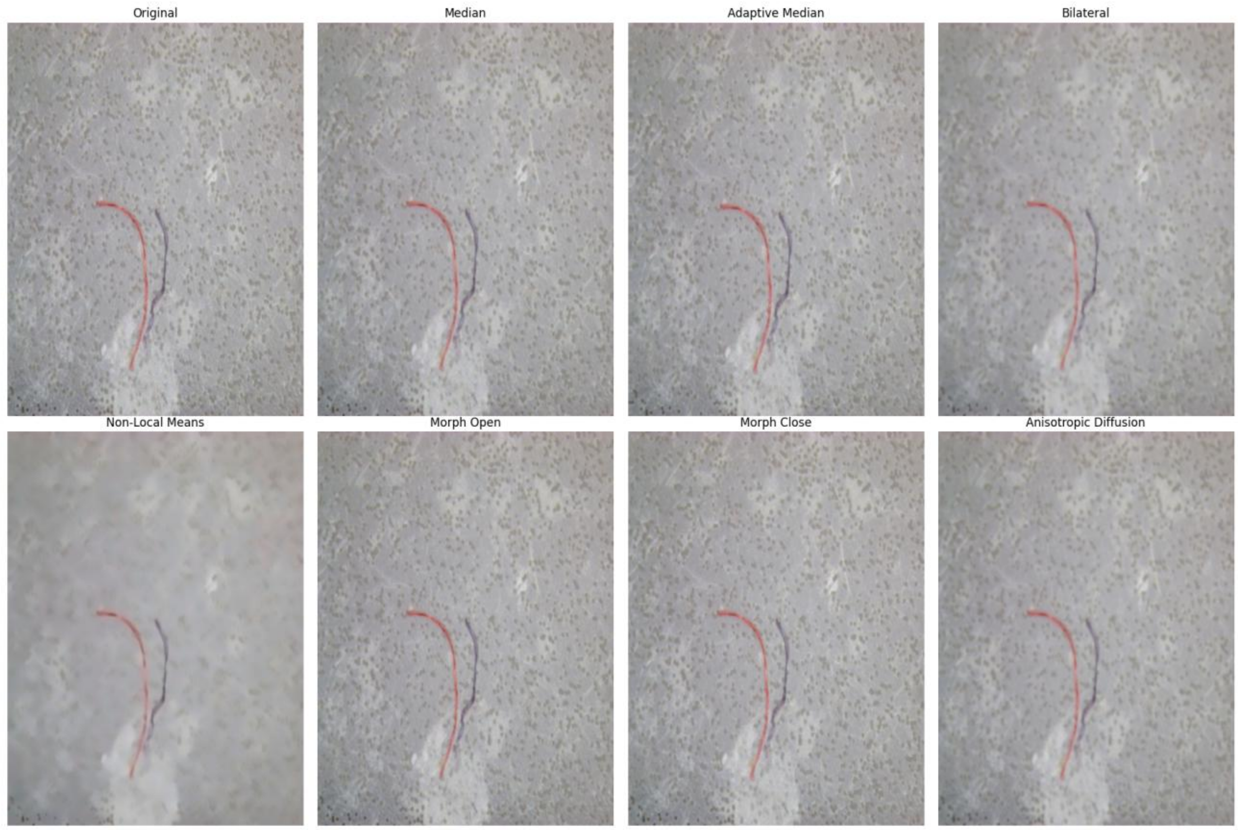
7.Final Cleanup with Connected Components Analysis

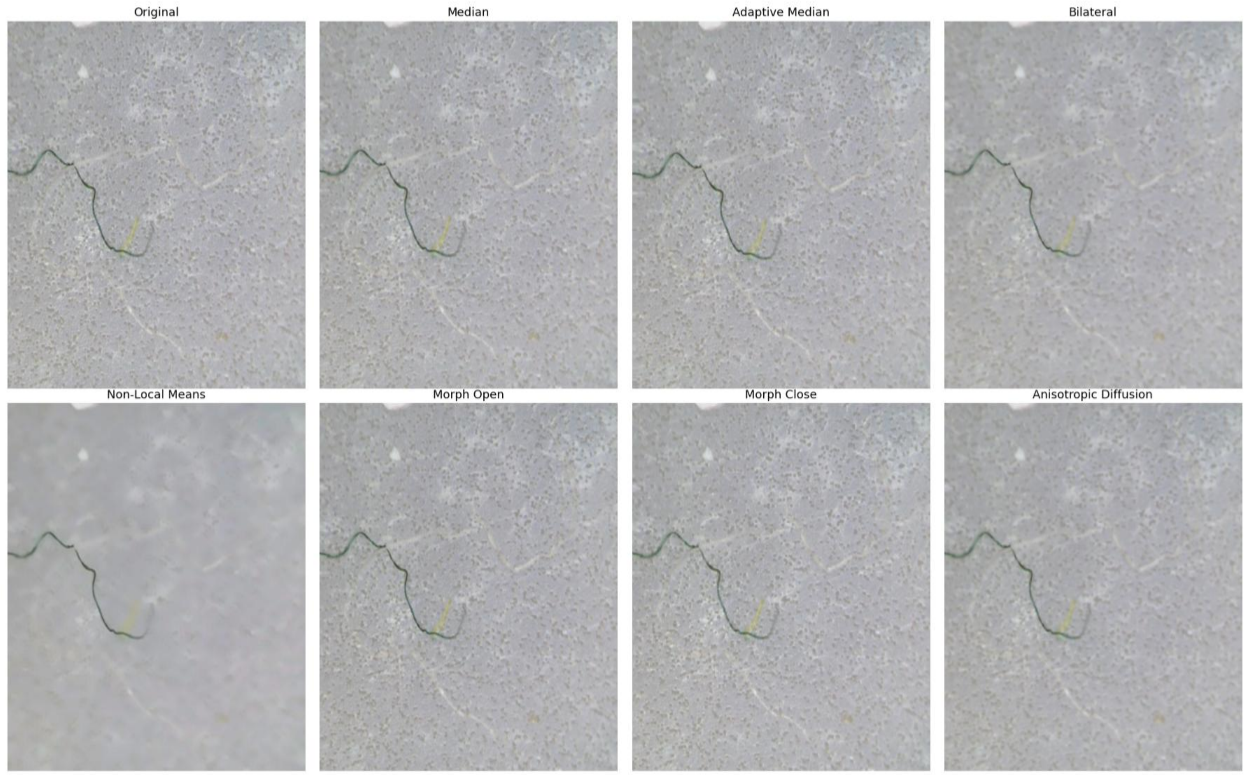
- **Label and extract significant segmented components.**

Noise Reduction algorithm experimentation :

In our experimentation with various noise reduction techniques, we evaluated multiple methods to determine their effectiveness in segmenting microplastics from microscopic images.Such as

- **Gaussian Blur :** A standard noise reduction technique**.** (Linear Filtering)
- **Median Filtering**
- **Adaptive Median Filter**: A custom implementation that processes each channel separately.
- **Bilateral Filtering**
- **Non-Local Means Denoising (NLM)**
- **Anisotropic Diffusion**





Among these approaches, Non-Local Means (NLM) denoising demonstrated superior performance in preserving essential details while effectively reducing noise.

NLM outperformed other methods by retaining the fine structures of microplastics, which is critical for accurate segmentation. The algorithm’s ability to distinguish between noise and meaningful textures made it particularly suitable for our dataset.

Extracting the Microplastics

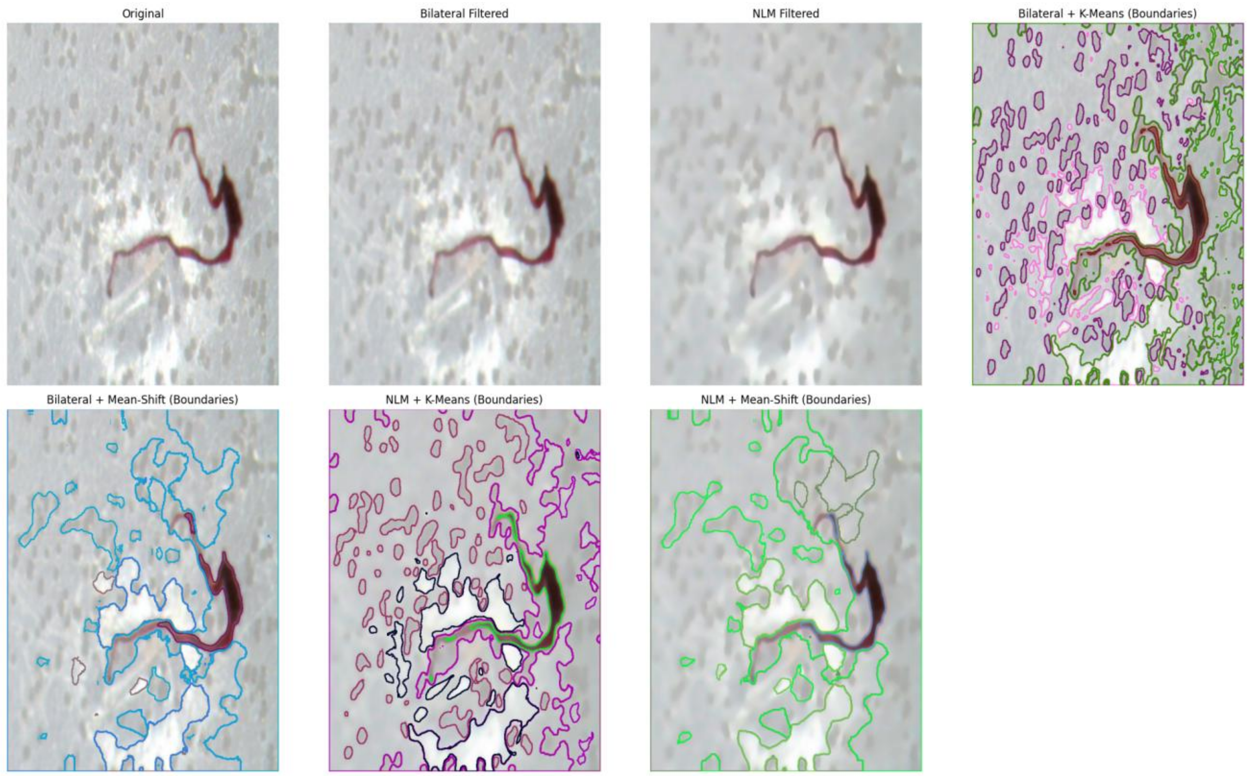
Based on Colour :

After denoising, two segmentation techniques were applied to extract microplastics from the images:

1. **K-Means Clustering-Based Segmentation:**
    - The image was reshaped into a 2D array and clustered into four segments (k=4) using the k-means algorithm.
    - The label map obtained from k-means was used to define different segmented regions.
    - To visualize the segmentation, region boundaries were outlined using contour detection.
2. **Mean-Shift Segmentation with K-Means Approximation:**

- OpenCV's `pyrMeanShiftFiltering` function was applied with spatial window size sp = 21 and color range sr = 51, which smooths color variations while preserving edges.
- To obtain labeled segments from the mean-shift result, a secondary k-means clustering (k=4) step was applied.

The segmentation was performed on both denoised images (Bilateral Filtered and NLM Filtered) to compare the impact of different noise reduction methods on segmentation quality.



K-Means Clustering successfully segmented the image into four distinct regions, but suffered from minor noise-related inconsistencies in segment boundaries.

Mean-Shift Segmentation provided smoother region boundaries but lacked precise delineation of smaller microplastic structures.

By applying contour detection on the segmentation results, the extracted microplastic regions were visually highlighted. The **NLM + Mean-Shift combination** yielded the most distinct boundaries, making it the most effective approach. But, clearly this is far from an acceptable result. So, Structured based refinement is required.

Based on the Structure :

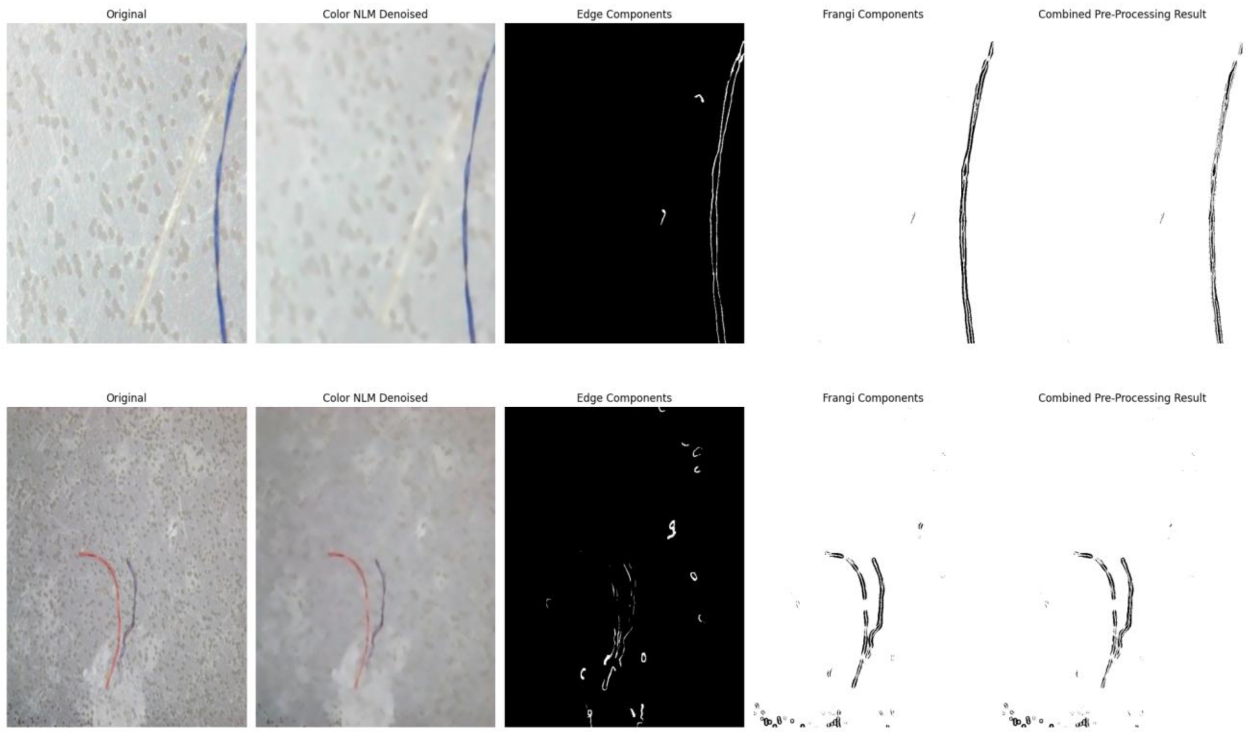
To enhance the detection of microplastic particles, structural features were extracted using edge detection, filtering, and morphological operations.

Canny Edge Detection with Dilation

- Convert the image to grayscale to simplify analysis.
- Apply **Canny edge detection**, which detects high-intensity gradient changes.
- Use a **3×3 morphological kernel** to perform dilation, which thickens and connects weak or broken edges.
- Dilation helps in forming continuous contours, which improves segmentation accuracy in the subsequent steps.

Frangi Filter for Vessel Enhancement

- The Frangi filter is commonly used for enhancing tubular or filament-like structures in an image. Since microplastic fragments often exhibit thin, elongated shapes.

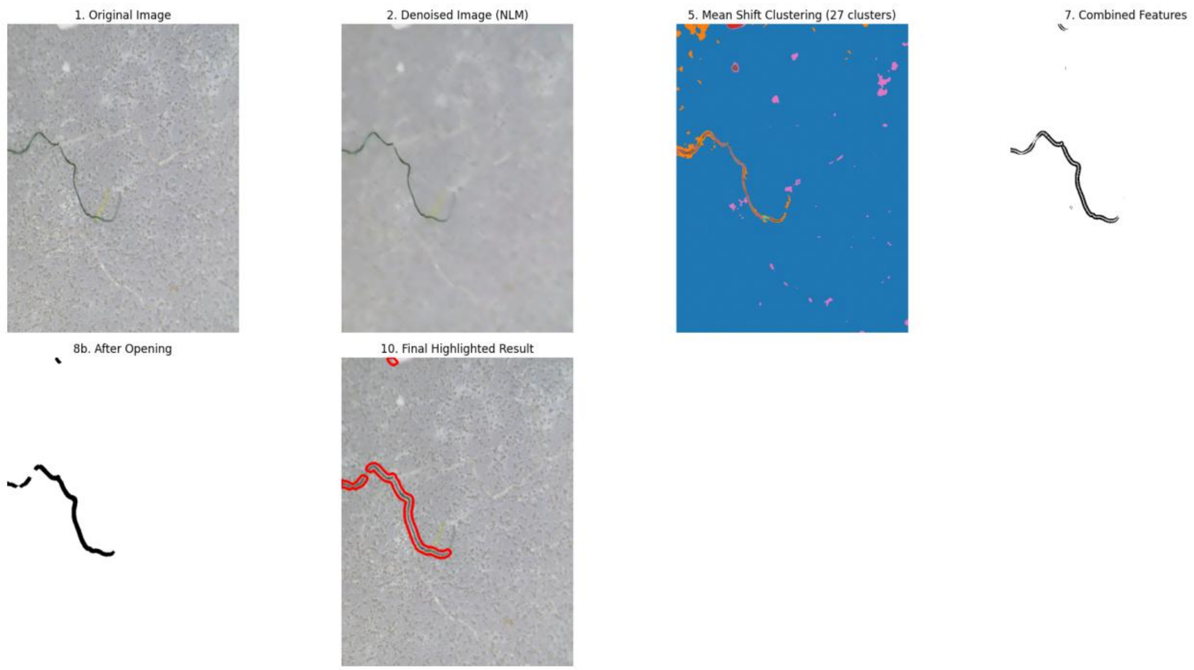


Why Use the Frangi Filter?

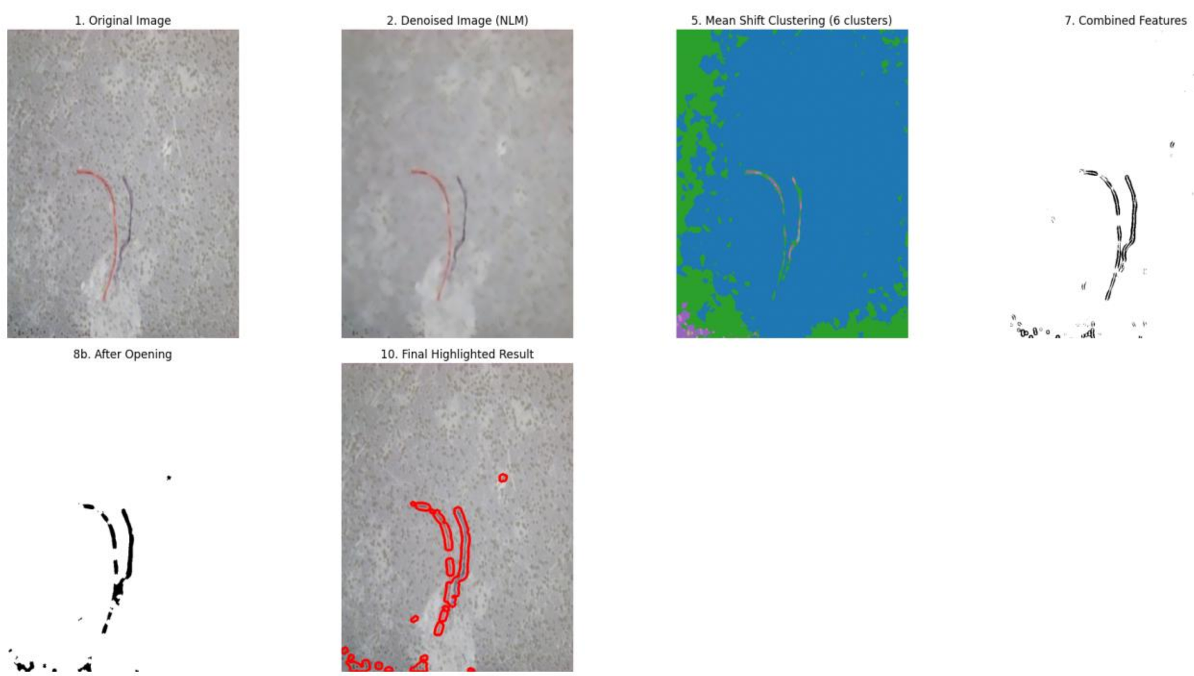
- Enhances rod-like structures such as bacteria, microplastics, or blood vessels.
- Reduces noise while preserving important structural details.
- Separates elongated objects from the background, improving segmentation.
- Works well for low-contrast images, where simple edge detection may fail.

Final Results :

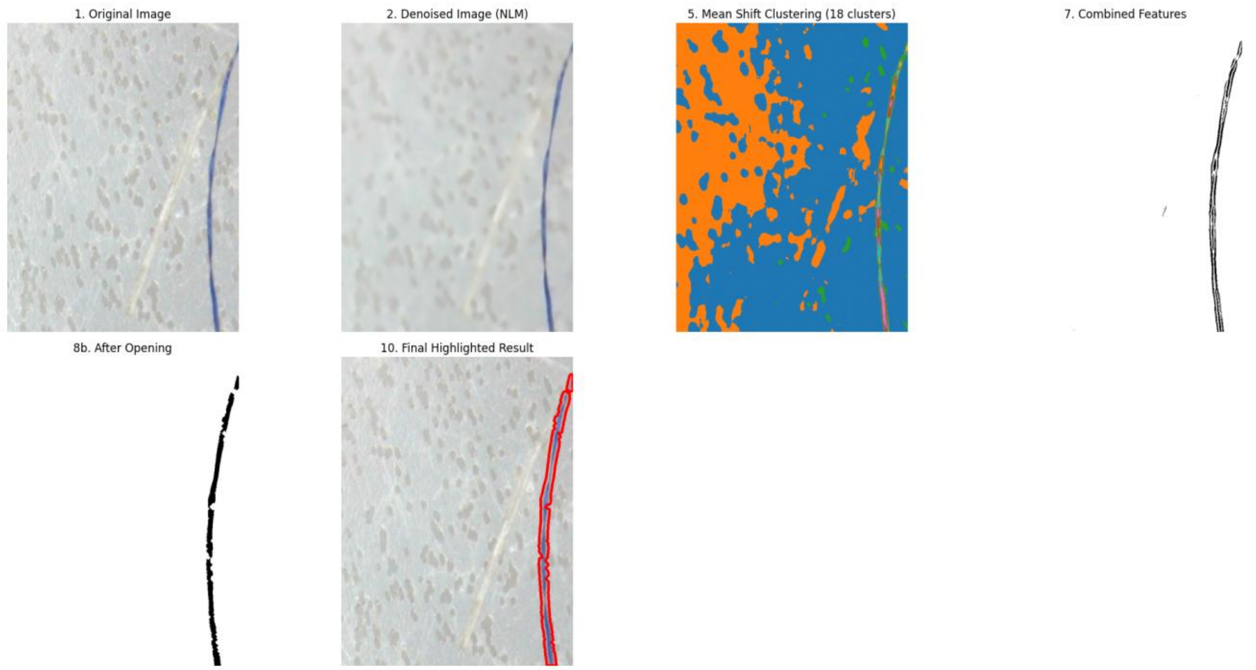
Example 1 :



Example 2 :

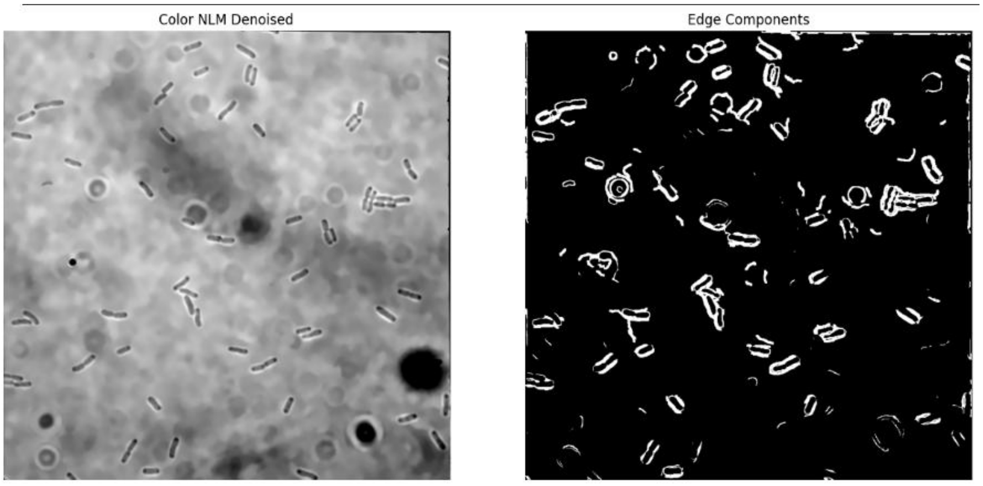


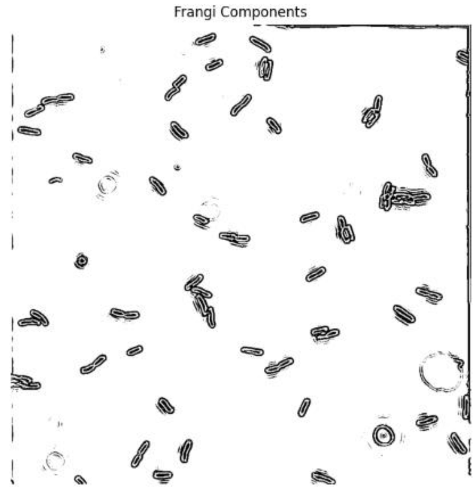
Example 3 :



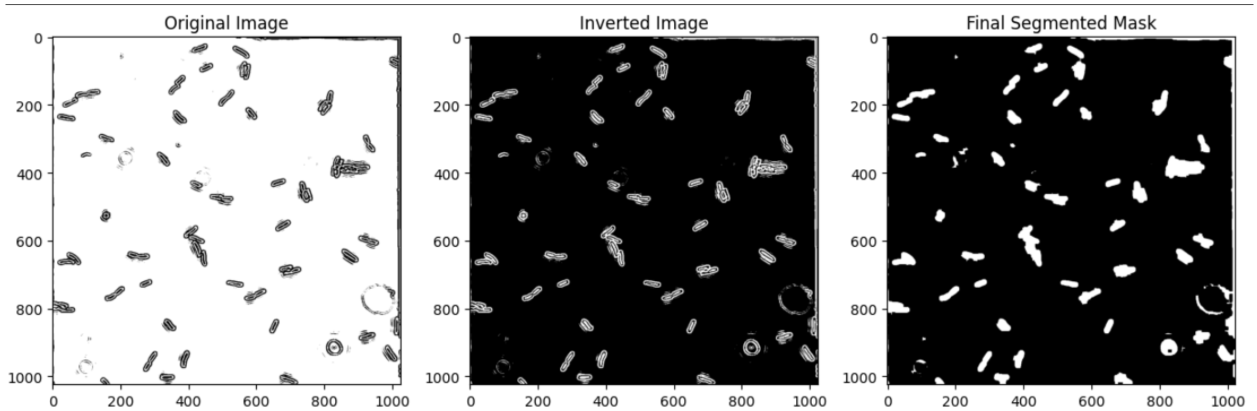
Applying this Approach on Escherichia coli (E. coli) Dataset

To further evaluate the effectiveness and generalizability of our noise reduction and segmentation approach, Applied this approach of denoising and segmenting on the previous problem statement. This dataset consists of microscopic images of E. coli cells, which has tubular-like structure which is similar to microplastics.

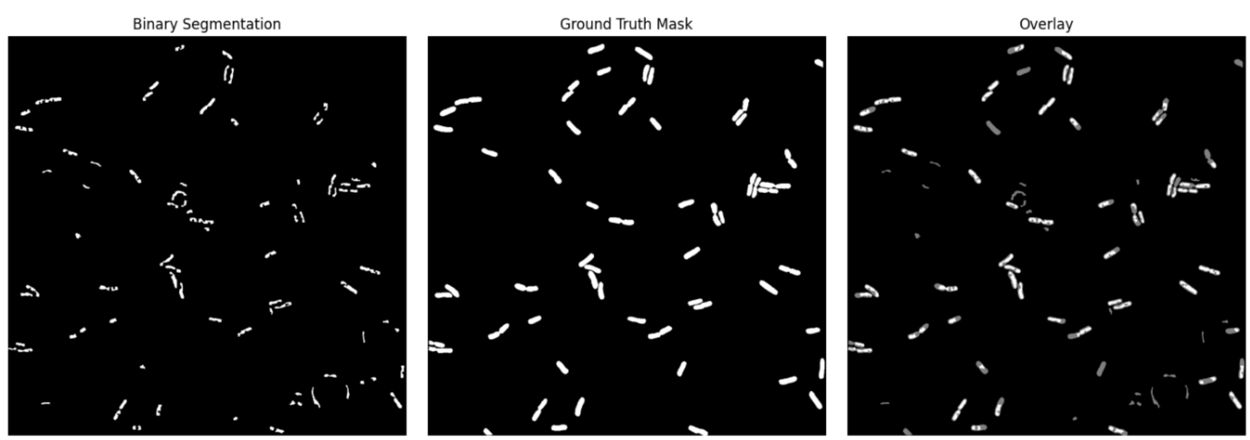




Creating segment mask from above output:



Evaluating the segmentation:



Hybrid approach 1:

```
Pixel Accuracy: 0.977
Precision: 0.833
Recall: 0.371
Dice Coefficient: 0.513
IoU (Jaccard Index): 0.345
```

Hybrid approach 2:

```
IoU: 0.4380
Dice Coefficient (F1 Score): 0.6091
Accuracy: 0.9592
Precision: 0.4395
Recall (Sensitivity): 0.9921
Specificity: 0.9581
Balanced Accuracy: 0.9751
```

Object detection phase:

1.Collecting Object Stats

Each region from `final_labels` is measured via `regionprops`: we store area, aspect ratio, centroid, etc**.**

2.Distance-Based Clustering with graph BFS connected component analysis

- We consider two objects neighbors if their centroids are within `DISTANCE_THRESHOLD` pixels.
- Using BFS, we find all connected components in this "neighbors" graph. Each connected component becomes a cluster of objects.

3.Classifying the Cluster

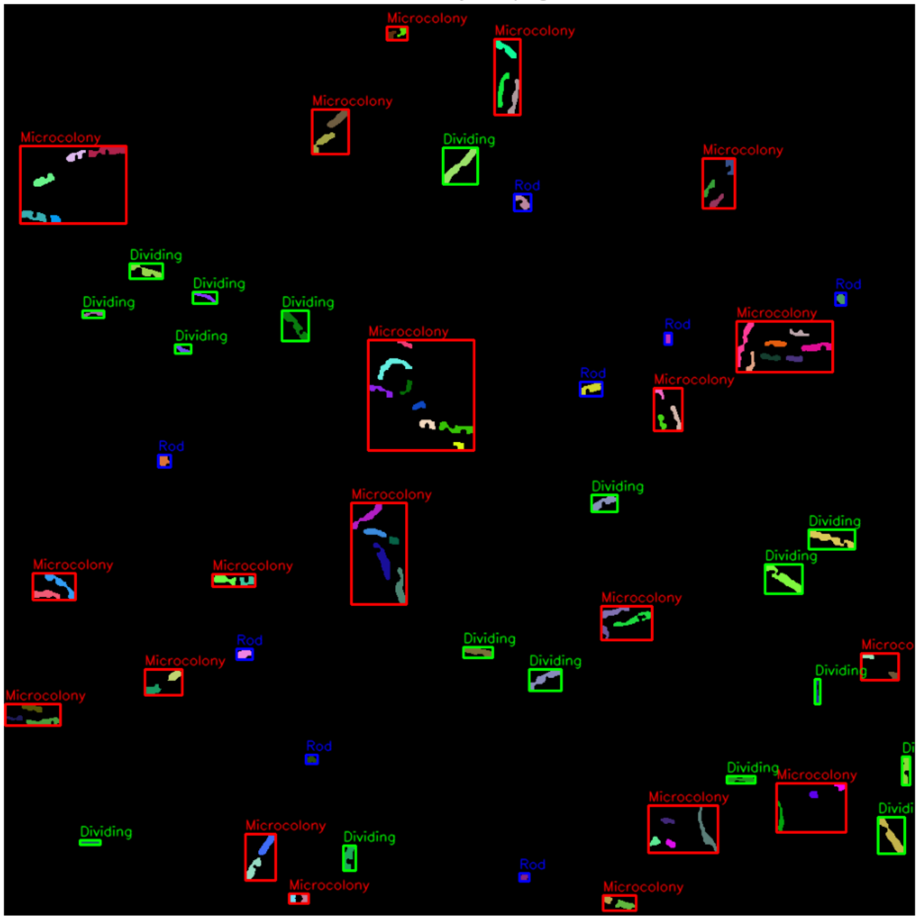
- If the cluster has multiple objects (or total area is large), we label it Microcolony.
- If it's exactly one object with a high aspect ratio, label it Dividing.
- Otherwise, label it Rod.

4.Drawing a Single Bounding Box

- We compute the bounding box that encloses all objects in the cluster.
- Draw one box and place the classification label.

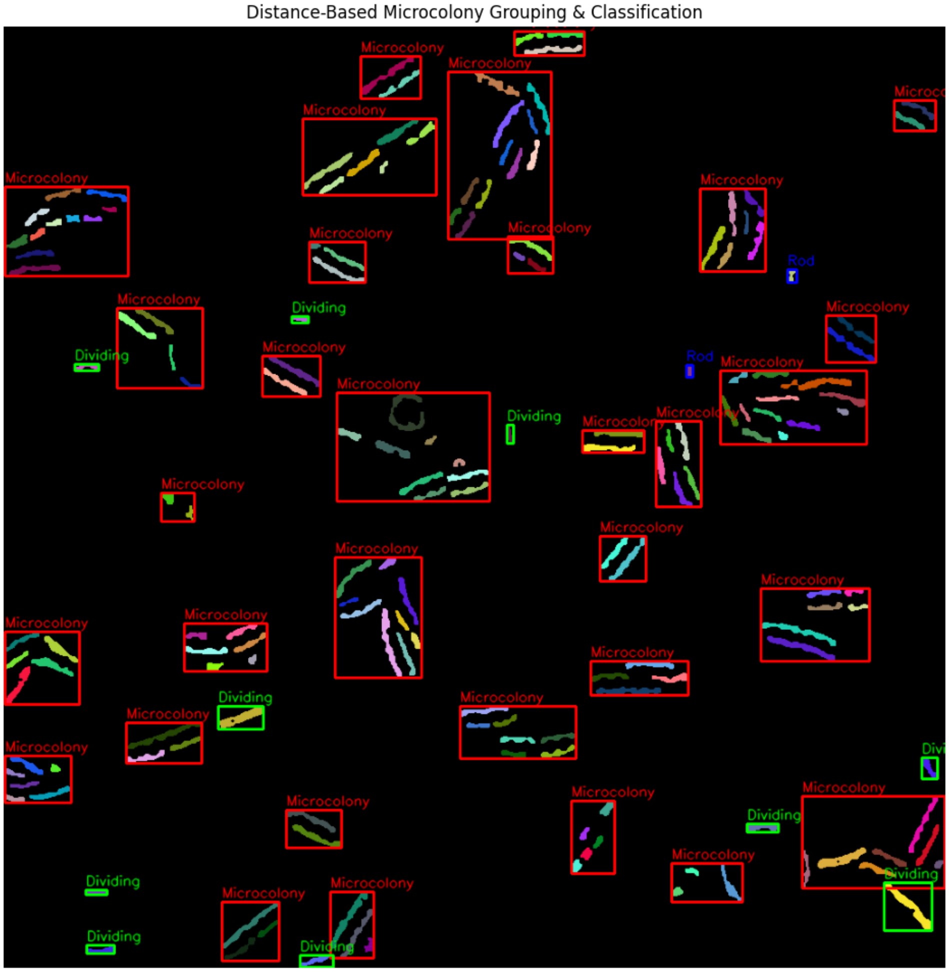
Initial result: Lot of incorrect predictions, not much reliable, machine learning or deep learning approach is required

Distance-Based Microcolony Grouping & Classification

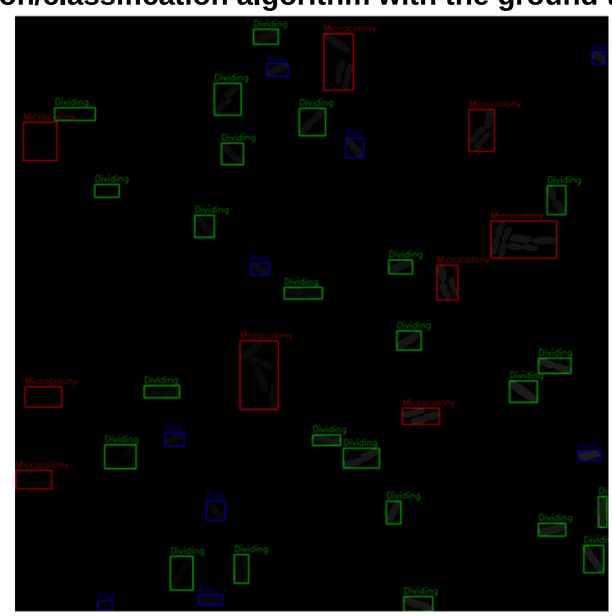


Mature stage:

Distance-Based Microcolony Grouping & Classification

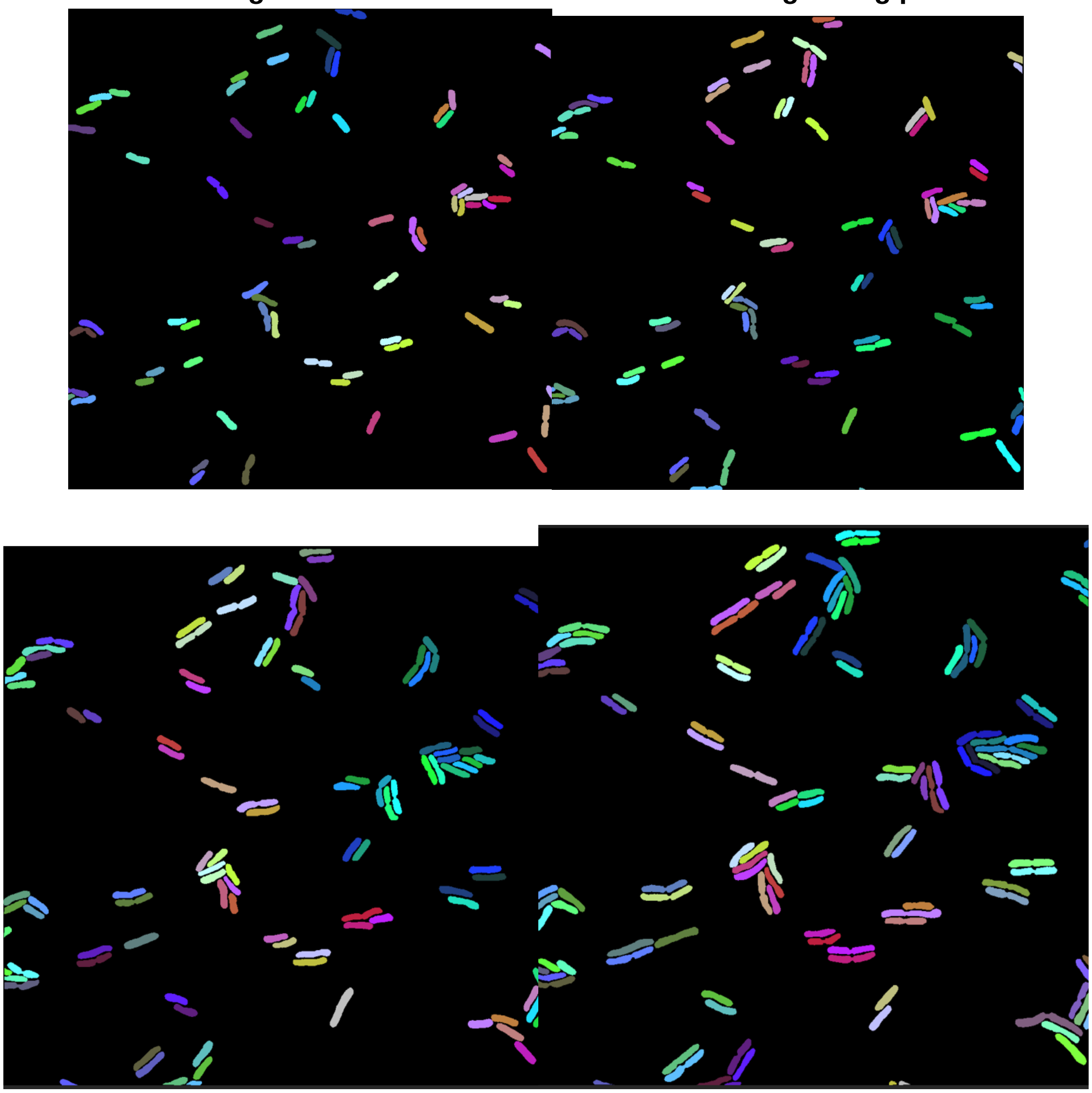


Testing the detection/classification algorithm with the ground truth for evaluation:



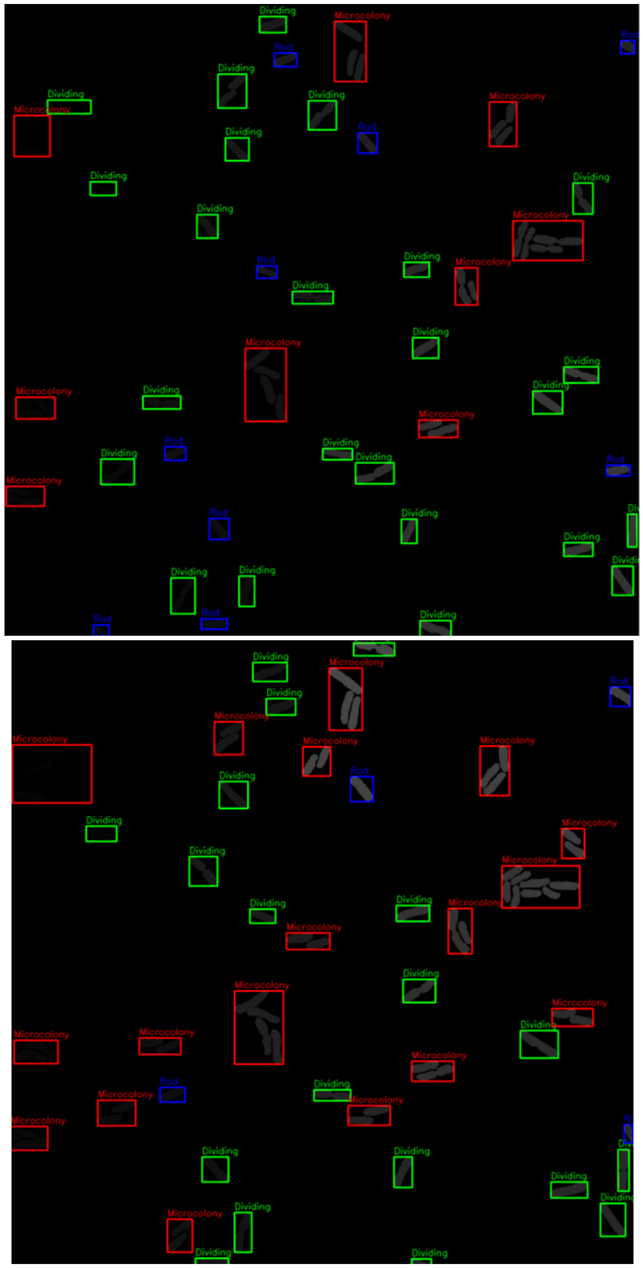
Growth phase monitoring:

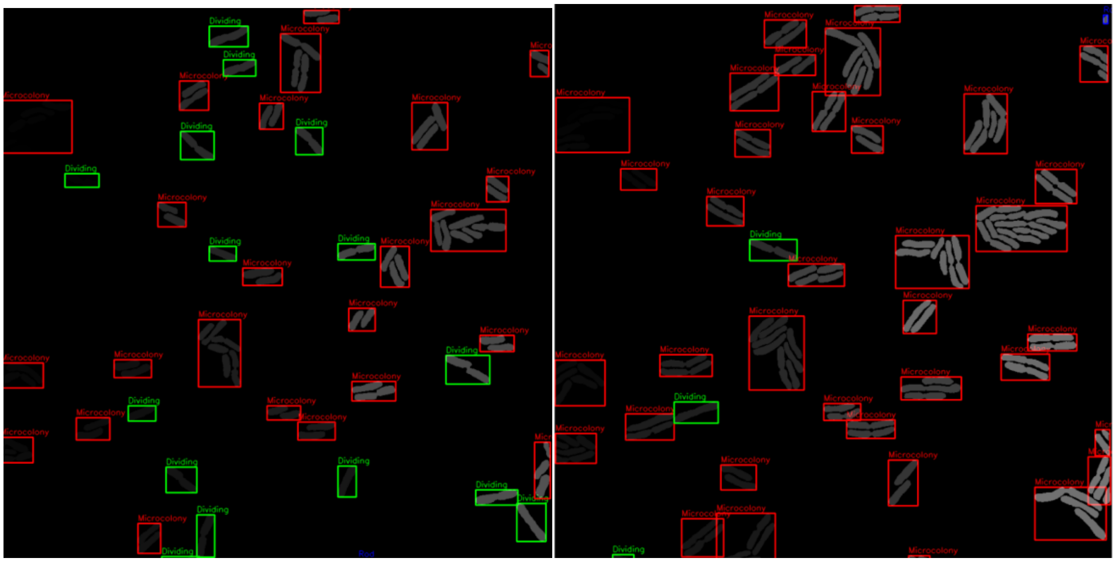
Below is the ground truth masks to demonstrate the growing process:



As seen above, the bacteria is growing and the task is the automate monitoring and analysis of this.

Entire sequence predictions:





Inference: As seen above the entire sequence with 5 frames are predicted for monitoring

Metrics for the first image alone:

- Accuracy = 0.62068
- Macro-Averaged Precision = 0.68333
- Macro-Averaged F1-score = 0.80050
- Weighted-Averaged F1-score = 0.75362

Analysis:

If we break it down by class: Microcolony: ~8 labeled, ~6 correct, ~2 incorrect

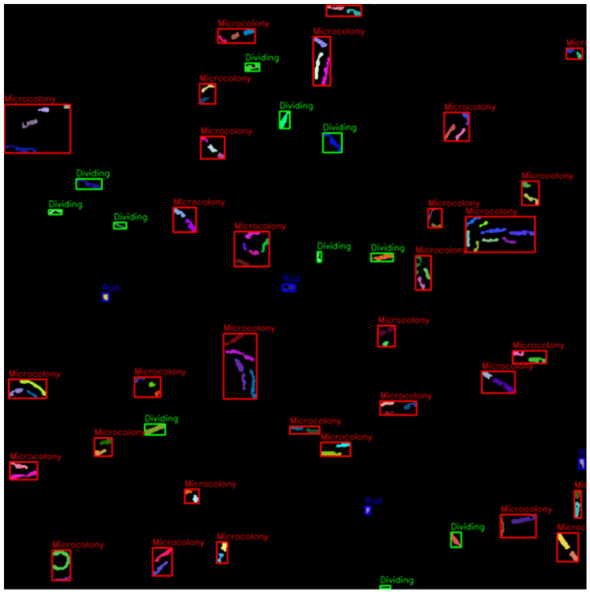
Dividing: ~15 labeled, ~7 correct, ~8 incorrect (the biggest error source)

Rod: ~6 labeled, ~5 correct, ~1 incorrect

A lot of rods are predicted as dividing. But moving on the sequences, microcolonies are predicted very accurately as seen in the last 3 images having very high accuracy almost above 90%.

Inference: This method although can predict reasonably well, for real world usage, machine learning or deep learning methods should be used as there are several challenges in the images and the features are complex for graph edge distance analysis approach to classify.

Unremovable errors during segmentation significantly affects the detection accuracy as shown below. Rule based approaches are ineffective against these tasks. One of the worst cases is shown below:



Whereas a mature stage has good predictions as shown below:

Distance-Based Microcolony Grouping & Classification