

Cellular Vision

Enhancing Medical Insights through Image Analysis

Kyrollos, Norhan, Mohamed

University of Nebraska-Lincoln

1 INTRODUCTION

Cellular analysis is fundamental to medical diagnostics and research, but traditional detection methods face limitations in scalability and accuracy. Addressing this issue, our project aims to revolutionize cell detection methodologies, enhancing precision and speed.

Applications span medical diagnostics, pharmaceutical research, pathology, and environmental studies. Our approach is to use traditional circle detection and to use YOLO for more efficiency in cellular detection. The image dataset we had was limited, so we worked on refining the datasets for a wider one. Through these efforts, we provide a tool for researchers, propelling the field and facilitating breakthroughs in medical insights.

2 OBJECTIVES

- 1) Utilizing Circle Detection By Traditional Methods (Hough Transform)
- 2) Dataset Refinement for Our Machine Learning Model (Data Augmentation)
- 3) Pushing Boundaries with YOLO for Circle Detection
- 4) Unveiling Special Properties Post-Detection

3 RELATED WORK

The field of medical image processing, crucial for accurate diagnosis, has seen substantial growth with the advancement of computer vision techniques applied to images from X-ray radiography, CT scans, and MRI. One of the related works is concerned with White blood cells (WBCs)[1].

White blood cells (WBCs) are crucial for disease diagnosis, prompting the application of computer vision methods for improved cell analysis. Existing approaches include boundary support vectors, iterative Otsu methods, and fuzzy cellular neural networks [1]. However, challenges persist due to the high variability in cell characteristics and inconsistent lighting conditions during image capture.

Ellipse detection, a well-established research problem, offers a potential solution for approximating WBCs. Traditional methods categorize into symmetry-based, Hough transform-based (HT), and random sampling approaches. Symmetry-based techniques consider ellipse geometry, while HT quantized parameters in parameter space. Random sampling methods, such as McLaughlin's approach [1], demonstrate improvements in accuracy and computational complexity compared to HT.

An alternative approach to ellipse detection involves optimization methods, showing superior results in accuracy and robustness [1]. However, such methods are underutilized in WBC detection, with exceptions like Karkavitsas and Rangoussi's work [2] using genetic algorithms (GAs).

This paper reviews these existing methodologies and introduces a novel approach to WBC detection, treating it as an optimization problem using the differential evolution (DE) algorithm. DE, known for its simplicity, fast convergence, and robustness, is applied to evolve candidate ellipses for accurate WBC detection. The contribution lies in introducing an efficient WBC detector, leveraging the DE algorithm and ellipse detection, a novel application in medical image processing.

Our work addresses a wide array of cell types for medical cell analysis through the lens of circular detection methods. In contrast to existing research that often concentrates on specific cell types like white blood cells using various techniques, our approach focuses on the general characteristics of cells represented by circles. This choice not only enhances detection accuracy by aligning with the spherical or circular nature of many cell types but also provides a more versatile solution for comprehensive cellular analysis. In summary, our work offers a distinctive and efficient approach to understanding cells in diverse domains through the application of circular detection methods.

4 PROBLEM DEFINITION

The problem addressed in this project is the detection and analysis of cells within clusters. Specifically, the inputs to our system consist of images containing cell clusters, and the goal is to develop a cell detection method that accurately identifies and isolates individual cells within these clusters. The outputs of our system include images where cells are detected, and further, we extract specific properties from these detected cells.

Inputs:

Images containing clusters of cells.

Outputs:

1. Processed images with accurately detected individual cells.
2. Extracted properties from the detected cells, such as size, location, and distances between cells.

Importance and Interest:

Accurate cell detection and analysis are vital in various fields, including medical diagnostics, pharmaceutical research, and environmental studies. The ability to precisely identify and understand cell structures contributes to advancements in the medical field. The extracted properties provide valuable insights for researchers and clinicians, facilitating a more in-depth understanding of cellular structures and their implications across diverse domains.

5 IMAGES

Data Augmentation Methodology

Initial Dataset:

Our dataset initially comprised a limited set of 6 images provided by Professor Ashok Samel, restricting the diversity and quantity necessary for robust model training in cellular analysis.

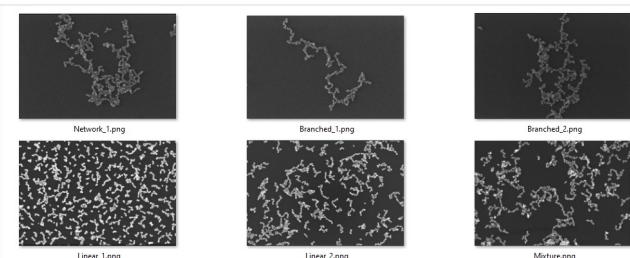


Fig. 1. Initial Dataset of 6 images.

Data Augmentation Approaches Implemented

Cell Separation Technique:

- Employed cell detection methodologies to identify and isolate individual cells within the images.

- Each detected cell was separated into an individual image file, preserving its original location and size attributes.
- This approach significantly expanded the dataset, resulting in the generation of over 4000 images.

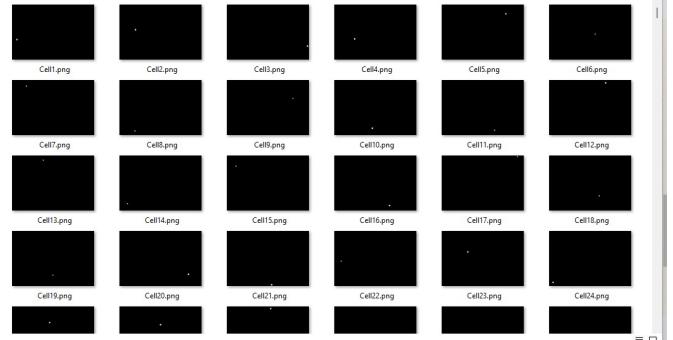


Fig. 2. More than 4000 images of separated cells.

Slicing and Cropping:

- Acknowledging the limitations of the initial dataset, additional steps were taken.
- Utilized MATLAB scripts to perform slicing and cropping on the original 6 images.
- This process resulted in the creation of an additional 12000 images.



Fig. 3. 12000 Sliced and Cropped Images

Total Augmented Dataset Size:

- **Separated Cells:** >4000 images
- **Sliced and Cropped Images:** 12000 images
- **Overall Augmented Dataset:** >16000 images

Rationale for Data Augmentation

The data augmentation was essential to address the initial dataset's scarcity and lack of diversity. By significantly expanding the dataset to over 16000 images, our aim was to enable the machine learning model to learn a broader spectrum of cellular characteristics and variations, thereby enhancing its robustness and accuracy in cell detection.

This augmented dataset, encompassing diverse cellular attributes and configurations, substantially improves the model's training potential, empowering it to recognize various cellular patterns and structures. Ultimately, this extensive dataset enables the development of more precise and scalable cell detection methodologies, significantly contributing to the versatility and accuracy of our proposed approaches in cellular analysis.

6 APPROACH, IMPLEMENTATION, and ANALYSIS

First Objective: Traditional Circle Detection using Hough Transform Algorithm

Image Preprocessing:

- Initiated the process by preparing the provided medical images for cell detection.
- Performed grayscale conversion to simplify image data and facilitate subsequent analysis.
- Implemented noise reduction techniques utilizing Gaussian filters to enhance image clarity and reduce interference.

Parameter Optimization for Hough Transform:

- Experimented with fine-tuning the parameters of the built-in Hough transform algorithm in MATLAB.
- Adjusted parameters such as sensitivity, edge threshold, and acceptable radius range to optimize detection accuracy.
- Executed the code with modified parameters to identify cells within the medical images.

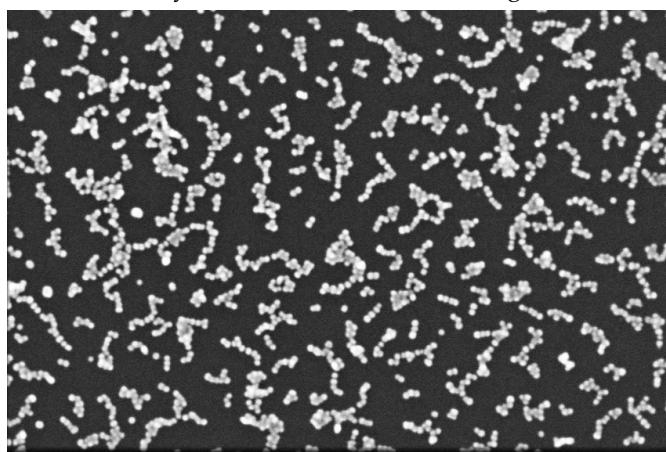


Fig. 4. Tested Image.

Addressing Overlapping Cells:

- Observed numerous cases of overlapping cells that were not initially detected.
- Implemented a strategy to handle the issue by modifying the program.

- Changed the color of the detected cells to match the background color (e.g., turning the cells black).

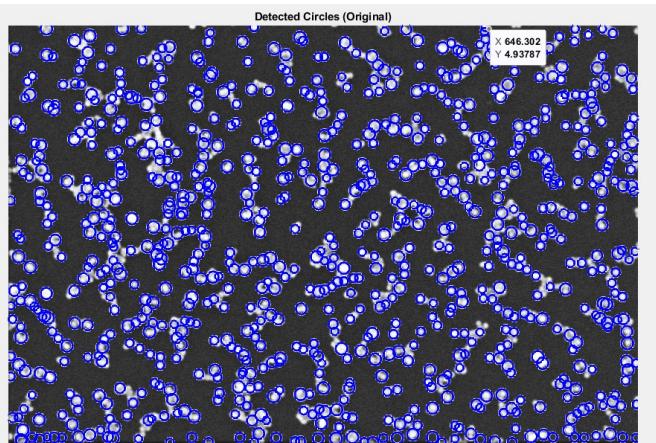


Fig. 5. Result After First Detection.

Refinement for Accurate Detection:

- Modified the Hough transform parameters to differentiate between bright cells and the newly turned black cells, thereby enabling the detection of all cells.
- Adjusted the parameters to recognize only bright cells while ignoring the black cells, which were previously detected.

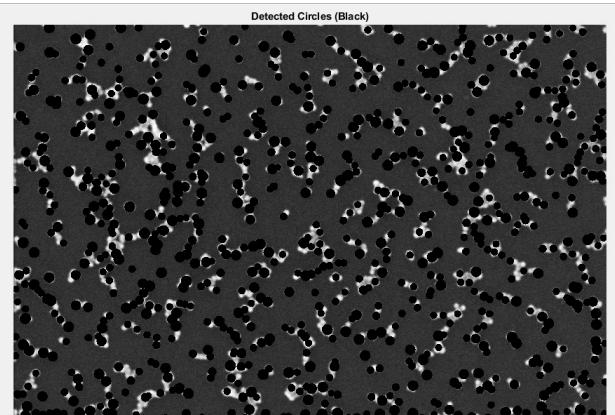


Fig. 6. Changing the Color of Detected Cells.

Outcome and Result:

- After these adjustments and refinements, successfully achieved the detection of all cells within the medical images.
- By fine-tuning parameters and employing specific strategies to handle overlapping cells, attained comprehensive and accurate cell detection using the Hough transform algorithm.

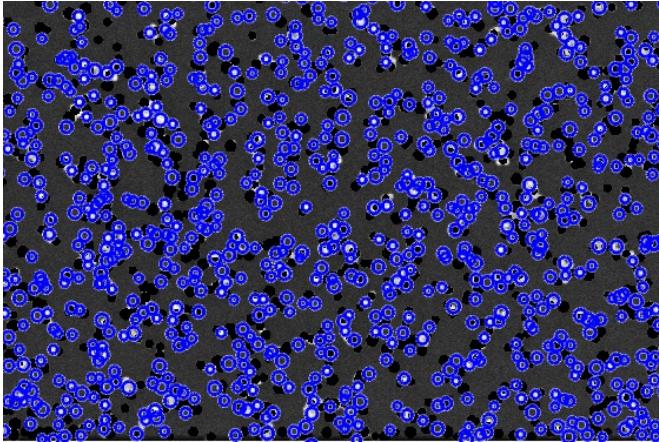


Fig. 7. Final Result (All Cell Get Detected)

Second Objective:

Data Augmentation for Enhanced Dataset

The second objective centered on augmenting the dataset to overcome limitations posed by the initial scarcity of images. Employing two distinct methodologies, we significantly expanded the dataset size from a mere 6 images to over 16000 images. The first approach involved separating each detected cell into individual images, resulting in over 4000 images. Additionally, utilizing slicing and cropping techniques on the original 6 images contributed an extra 12000 images to the dataset. This augmentation process aimed to diversify and amplify the dataset, as detailed in the image section earlier, ensuring a more comprehensive training set for our proposed methodologies in cell detection.

Third Objective: Cell Detection using YOLOv5

Dataset Labeling with Roboflow YOLO Labeler:

- Due to the absence of labeled data in our initial dataset, initiated the process by labeling cells in the dataset.
- Utilized the Roboflow YOLO Labeler tool for efficient and accurate annotation of cells within the images.

Finally, we had a dataset containing the following:



Fig. 8 . Dataset components

Training the DL Model with Python Code:

- Developed a custom Python code to facilitate the training of the YOLOv5 machine learning model.
- Utilized the labeled dataset to train the model, allowing it to learn the characteristics and features of cells.

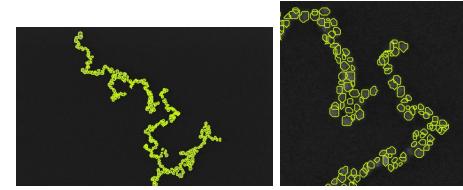


Fig.9. Part of the training split in the dataset.

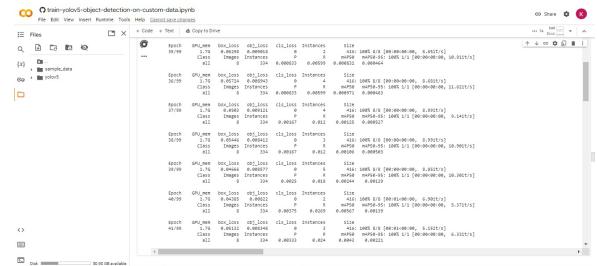


Fig.10. Finetuning the model over 100 Epochs

Testing the DL Model with Reference Images:

- Faced the challenge of lacking reference images for evaluating the trained model.
- Addressed this limitation by generating reference images using Roboflow to assess the performance and accuracy of the trained ML model.



Fig.11. Visualization of the model performance over the val dataset

Results

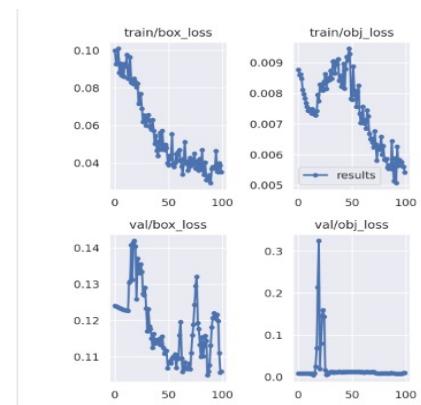


Fig.12. The training/validating loss results over 100 Epochs

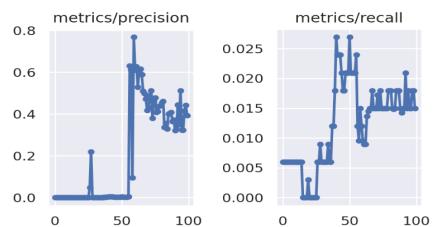


Fig.13. The precision/recall over 100 Epochs

A1	ImageName	Radius
1	Cell1.png	7.895470722
2	Cell2.png	7.182144888
3	Cell3.png	7.408407032
4	Cell4.png	5.047525655
5	Cell5.png	7.488338886
6	Cell6.png	7.763462089
7	Cell7.png	7.734219181
8	Cell8.png	7.933568276
9	Cell9.png	7.378292584
10	Cell10.png	7.818315204
11	Cell11.png	7.649332911
12	Cell12.png	7.690870143
13	Cell13.png	5.052597459
14	Cell14.png	7.735950667
15	Cell15.png	7.837137017
16	Cell16.png	7.363225127
17	Cell17.png	6.48631457
18	Cell18.png	7.179372554
19	Cell19.png	7.92961204
20	Cell20.png	7.804431495
21	Cell21.png	5.012261051

Fig. NO14. A snippet of the datasheet that collects the radii of detected cells on one of the dataset images.

To sum up

- The process involved labeling the dataset using Roboflow, training the YOLOv5 model with Python code using the labeled dataset, and generating reference images to evaluate the model's performance.
- This strategy aimed to overcome the absence of labeled data initially and validate the efficacy of the trained DL model for cell detection purposes.
- By employing these steps and methodologies, we aimed to leverage the YOLOv5 model for accurate and efficient cell detection, despite the initial lack of labeled reference data. This approach allowed us to train and test the model, enhancing our capability to detect cells within medical images and furthering our objectives in advancing cell detection methodologies.

Fourth Objective: Extracting properties from the detected cells

Size Analysis:

- Defined a range of radii and adjusted sensitivity and edge threshold parameters for accurate size estimation.
- Extracted and saved radii information to quantify the sizes of individual cells systematically.
- Provided a quantitative measure of cell sizes within the image.

Location Analysis:

- Captured x, y coordinates of detected circle centers for each cell in the image.
- Utilized visualizations with red crosses on the original image to qualitatively represent cell locations.
- Stored center coordinates in text and Excel files, ensuring structured data for further spatial analysis.
- Offered a comprehensive approach to understanding the spatial distribution of cells within the cluster

A1	B	C	D	E	F	G	H
1	X	Y	Distance				
2	254	57	1320.729997				
3	1167	899	617.511762				
4	80.30406147	957.8735123	1553.283754				
5	747	772	866.8953837				
6	277.0698083	115.0904759	1284.376197				
7	1090.115237	750.2199804	562.4975777				
8	1413.912008	23.20641939	393.3813261				
9	1184	224	387.4929537				
10	706.3238197	301.4952885	829.0841373				
11	1468.428973	287.8053672	127.0851952				
12	1445.989409	592	210.1786442				
13	1004.63096	97.92941757	605.2933816				
14	558	236.5052949	985.15692				
15	232	98	1332.126743				
16	991.6323062	934	758.6007268				
17	1026	867	687.3862766				
18	58.66700375	588	1483.035045				
19	1037	929.2287333	723.6275584				
20	177.4431249	197.9068197	1367.087718				
21	669.3099014	329.3569481	863.1518858				
22	624	832.5339254	1003.974574				

Fig. NO15. A part of the datasheet that collects the x-y coordinate locations of each detected cell of the image.

Distance Between Cells:

- Computed distances between a randomly chosen cell and all other detected cells using Euclidean distance calculations.
- Enhanced interpretation through a color map visualization, providing a visual representation of distance variations.
- Saved distances to an Excel file for structured data, facilitating spatial organization analysis.
- Facilitated exploration of spatial relationships, aiding in the identification of clusters, patterns, or irregularities within the cell population.

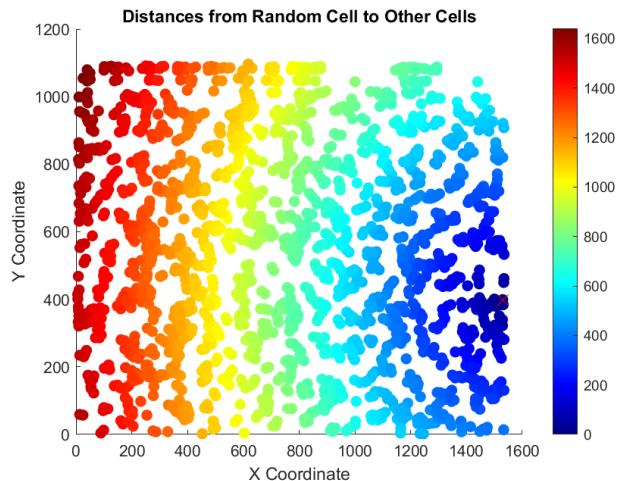


Fig. NO16. A color map that shows the distances between a randomly selected cell (marked at the red cross) and all other detected cells.

These methods collectively provide a thorough analysis of cellular understanding, offering both quantitative and qualitative insights for applications in biomedical research.

Briefly describe what approach you are planning to use. What software systems, components do you plan to use? Use of schematics, flowcharts, or pseudocode often makes it easier to follow an algorithm/approach.

REFERENCES

- [1] Erik Cuevas, Margarita Díaz, Miguel Manzanares, Daniel Zaldivar, Marco Perez-Cisneros, "An Improved Computer Vision Method for White Blood Cells Detection", Computational and Mathematical Methods in Medicine, vol. 2013, Article ID 137392, 14 pages, 2013. <https://doi.org/10.1155/2013/137392>
- [2] G. Karkavitsas and M. Rangoussi, "Object localization in medical images using genetic algorithms," World Academy of Science, Engineering and Technology, vol. 2, pp. 6–9, 2005.