PAMANTASAN NG LUNGSOD NG MAYNILA
(University of the City of Manila)
Intramuros, Manila

Elective 3 (LAB)

Final Project

K-Bliver: Utilization of K-means Clustering in Liver Ailment Detection

Submitted by:
Cacal, Chad R.
Dela Cruz, Alfonso Rafael C.
Palabay, Joven Carl B.
Suyu, Chester T.

Instructor:
Engr. Maria Rizette H. Sayo

Aug, 10, 2024

INTRODUCTION

Image segmentation is a pivotal task in the field of computer vision, enabling the partitioning of an image into distinct regions that are more meaningful and easier to analyze. This task is especially relevant in applications such as medical imaging, object detection, and image editing, where accurate segmentation is crucial for further analysis.

K-means clustering is a widely used image segmentation method that groups data into clusters based on similarity. The user specifies the number of clusters (K) in advance. The algorithm then categorizes the data into groups where the similarity within the same cluster is high, and the similarity between different clusters is low (Wu, 2021). This makes K-means effective for segmenting images into distinct regions based on pixel intensities or colors.

While the K-means algorithm is effective, it is computationally demanding, especially for high-resolution medical images. The iterative processes of data assignment and centroid recalculation require substantial computational resources. Recent advancements have focused on using Graphics Processing Units (GPUs) to accelerate K-means clustering. By leveraging the Compute Unified Device Architecture (CUDA), the segmentation process becomes more efficient, enabling faster analysis of large medical image datasets and enhancing the diagnostic workflow. This GPU-based approach divides the problem into sub-problems processed independently on Streaming Multiprocessors (SMs) of one or more GPUs, resulting in quicker and more efficient image segmentation (Fakhi et al., 2017).

The K-means clustering algorithm is a popular method for image segmentation in medical imaging, particularly for analyzing liver ailments. It groups image data into clusters based on similarity, which is essential for distinguishing between healthy and diseased liver tissue. By segmenting liver scans, K-means helps identify abnormalities such as tumors or fibrosis. Users must specify the number of clusters (K) in advance, and the algorithm then organizes the image data into groups with high similarity within clusters and low similarity between them. This makes K-means effective in segmenting liver images based on pixel intensity and texture variations, significantly aiding the diagnostic process.

Image segmentation is essential in medical imaging as it enables the division of medical images into distinct regions, which facilitates the identification and evaluation of pathological changes in liver tissues. Accurate segmentation supports effective diagnosis, treatment planning, and monitoring of liver diseases, ensuring that clinicians can address liver ailments more effectively.

Objectives

This project aims to implement liver image segmentation using K-means clustering in Python, exploring both traditional CPU-based methods and modern GPU-accelerated techniques. The goal is to provide a comprehensive understanding of how K-means clustering can be utilized for medical image segmentation and to demonstrate its practical application with a focus on optimizing performance.

In this project, we will utilize Python and various image processing libraries to implement a system that performs the following:

- **Understand the Role of K-means Clustering in Liver Image Segmentation:**
  - Investigate how K-means clustering segments liver images based on pixel similarity, enhancing the analysis of liver ailments.
- **Analyze the Effectiveness of K-means Clustering for Liver Image Types:**
  - Evaluate the performance of K-means clustering across various liver image types, assessing its strengths and limitations in detecting different tissue textures and abnormalities.
- **Compare K-means Clustering with Other Medical Image Segmentation Techniques:**
  - Conduct a comparative analysis of K-means clustering and alternative methods like Gaussian Mixture Models (GMM) or Mean Shift to assess the relative effectiveness of K-means in medical imaging.
- **Evaluate the Computational Efficiency of K-means Clustering in Liver Image Segmentation:**
  - Measure the computational requirements and performance bottlenecks of K-means clustering in the context of liver image segmentation, identifying opportunities for optimization in real-world clinical applications.

By focusing on these objectives, this project seeks to enhance liver ailment diagnosis and treatment planning through improved image segmentation techniques, leveraging the power of modern computational resources to achieve accurate and efficient results.

REVIEW OF RELATED LITERATURE

This section provides an overview of the key research studies relevant to the implementation of image segmentation using K-means clustering. The studies reviewed highlight various aspects of K-means clustering, including its application in image segmentation, performance optimization techniques, and the broader implications for computational efficiency in complex tasks.

K-means Clustering in Image Segmentation

K-means clustering is a widely used method for partitioning data into distinct groups, and it has been extensively applied in image segmentation tasks. In their work, Wu (2021) discusses how K-means, a distance-based clustering algorithm, effectively aggregates data into multiple categories by minimizing intra-cluster variance while maximizing inter-cluster variance. The algorithm's strength lies in its ability to categorize image pixels into clusters based on similarity, making it a suitable tool for segmenting images into meaningful regions.

GPU-Accelerated K-means Clustering

While K-means is effective, it is also computationally intensive, especially when applied to high-resolution images. Fakhi et al. (2017) address this challenge by proposing a GPU-based implementation of K-means clustering to accelerate its performance. Their study focuses on offloading the compute-intensive tasks of data assignment and centroid recalculation to GPUs, leveraging the parallel processing capabilities of modern GPUs. The proposed method significantly reduces computation time by dividing the problem into sub-problems, which are processed independently on Streaming Multiprocessors (SM) using the Compute Unified Device Architecture (CUDA). The efficiency gains from this approach are crucial for large-scale image segmentation tasks.

Enhanced Clustering Techniques

Pandey and Sharma (2023) explore the broader impact of clustering algorithms in enhancing learning experiences within blended-learning environments. Although their study is focused on educational technology, it underscores the importance of optimizing clustering techniques to improve outcomes. This research provides a conceptual framework that can be applied to various fields, including image segmentation, where the effectiveness of clustering algorithms like K-means is critical. Their findings emphasize the necessity of adapting and refining clustering methods to meet specific application requirements.

Liver disease detection using machine learning methods

Chronic liver disease is a major global health issue, caused by factors like undiagnosed hepatitis, obesity, contaminated food, hazardous fumes, and excessive alcohol use. Early identification is crucial but challenging due to the subtle early symptoms. This study aims to enhance liver disease forecasting using AI. A two-step approach was employed: first, predicting the presence of liver disease with supervised learning algorithms (K-Nearest Neighbors, XGBoost, Logistic Regression, Naive Bayes, Random Forest, Decision Tree, Support Vector Machine); second, determining disease severity using unsupervised learning with K-Means Clustering. The study evaluated classification methods using metrics such as Precision, Recall, Accuracy, F1 Score, F2 Score, ROC-AUC, and R2 Score, aiming to provide advanced insights into ML/DL approaches for liver disease prediction (2023).

The literature underscores K-means clustering's effectiveness in image segmentation and the importance of computational efficiency. Transitioning from CPU to GPU acceleration has significantly improved performance. This project will implement and optimize K-means clustering in Python, using these insights to develop efficient segmentation techniques for various image processing tasks.

METHODOLOGY

The development of the Python code for image segmentation using the K-means clustering algorithm will follow an Agile methodology, structured into several focused sprints. Initially, the project will begin with setting up the development environment, including necessary libraries like NumPy, SciPy, and OpenCV, followed by implementing the basic K-means clustering algorithm. The first phase will involve testing the basic implementation on a simple image to ensure the algorithm correctly segments the image based on pixel similarity.

Subsequently, the focus will shift to integrating image preprocessing steps, such as resizing, normalization, and color space conversion, to enhance the segmentation process. During this sprint, the cluster initialization process will also be refined to improve accuracy, with tests conducted to evaluate the impact of different preprocessing techniques on segmentation outcomes.

Next, the code will undergo optimization, particularly by implementing GPU acceleration using CUDA to speed up the computationally intensive K-means algorithm. A comparative analysis will be conducted to measure the performance differences between the CPU-based and GPU-accelerated versions, with iterative refinements made based on the results.

The project will then proceed to extensive testing on a diverse set of images with varying complexities to ensure robustness and accuracy. Based on these tests, the code will be further refined, addressing edge cases and improving overall performance. Finally, the project will culminate in the documentation and finalization phase, where comprehensive documentation will be prepared to explain the code's functionality and usage. This will include detailed comments within the code, a summary report of the project’s outcomes, and a final review to ensure completeness and accuracy. This Agile approach ensures continuous feedback and iterative improvements, leading to a robust and efficient solution for image segmentation using K-means clustering.

Image Segmentation with K-means Clustering

Image 1 represents the input: a high-resolution photograph of a butterfly resting on a flower amidst a background of green foliage. This image contains a rich variety of colors and textures, including the bright orange and black of the butterfly's wings, the green leaves, and the colorful flowers.

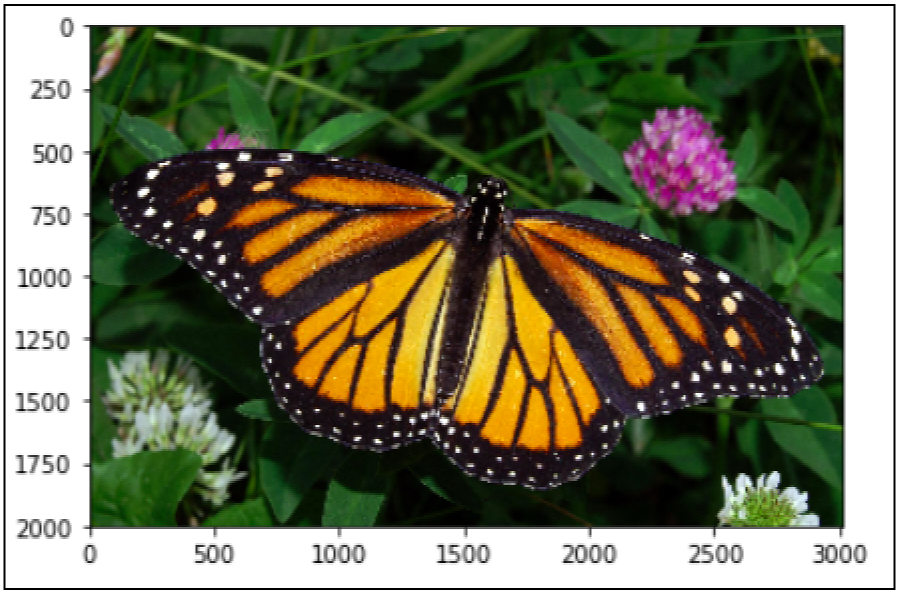

Image 1: Original Image

Image 2 represents the output after applying K-means clustering for image segmentation. In this processed image, the algorithm has reduced the original colors to a limited number of clusters, thereby simplifying the image into distinct regions based on color similarity. The butterfly's wings are grouped into two primary color clusters, with the background foliage and flowers similarly reduced to a few dominant color clusters. This transformation highlights the butterfly while simplifying the overall image, making it easier to identify distinct regions within the image.

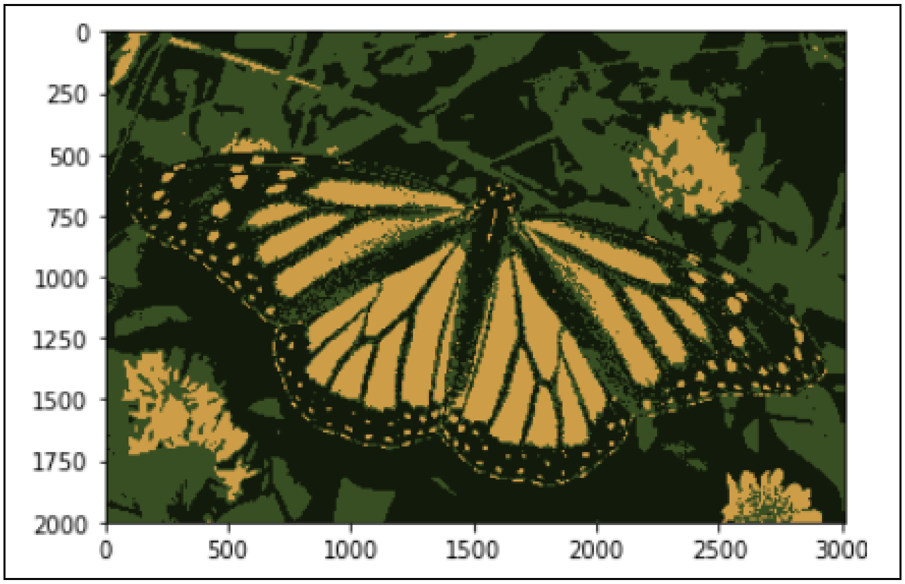

Image 2: Image Output

K-means clustering in image segmentation works by partitioning an image into several clusters, where each pixel in the image is assigned to the cluster with the closest mean color value. The number of clusters KKK is specified before running the algorithm. The algorithm then iteratively assigns pixels to clusters and recalculates the cluster centroids (the average color of all pixels in the cluster) until the centroids stabilize, meaning the clusters no longer change significantly.

In this example, the K-means algorithm has effectively reduced the image complexity by clustering pixels with similar colors into the same group. This results in a segmented image where the butterfly and the background are simplified into distinct color regions. This process is beneficial in various applications such as object recognition, where distinguishing between different parts of an image based on color can be crucial. The K-means clustering algorithm has thus transformed the original image into a more abstract representation, making it easier to analyze specific regions of interest.

BIBLIOGRAPHY

Liver disease detection using machine learning methods. (2023). IET Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10401218/authors#authors

Liu, L., Wang, L., Xu, D., Zhang, H., Sharma, A., Tiwari, S., Kaur, M., Khurana, M., & Shah, M. A. (2021). CT image segmentation method of liver tumor based on artificial intelligence enabled medical imaging. Mathematical Problems in Engineering, 2021, 1–8. https://doi.org/10.1155/2021/9919507

B. Wu, "K-means clustering algorithm and Python implementation," 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering https://ieeexplore.ieee.org/abstract/document/9543260

T. Kansal, S. Bahuguna, V. Singh and T. Choudhury, "Customer Segmentation using K-means Clustering," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems
https://ieeexplore.ieee.org/document/8769171

Aladžuz, A., Delalić, A., & Šćeta, L. (2022). Cluster analysis in Python: An example of market Segmentation. In Lecture notes in networks and systems (pp. 1032–1041).
https://link.springer.com/chapter/10.1007/978-3-031-05230-9_122

Image preprocessing algorithm based on K-Means. (2019, December 1). IEEE Conference Publication | IEEE Xplore.
https://ieeexplore.ieee.org/abstract/document/9029154

Pandey, M., & Sharma, A. (2023). Image segmentation techniques using python and deep learning. International Journal of Communication and Information Technology
https://www.computersciencejournals.com/ijcit/archives/2023.v4.i1.A.55
https://www.computersciencejournals.com/ijcit/article/55/4-1-2-146.pdf

H. Fakhi, O. Bouattane, M. Youssfi and O. Hassan, "New optimized GPU version of the k-means algorithm for large-sized image segmentation," *2017 Intelligent Systems and Computer Vision (ISCV)*, Fez, Morocco, 2017, pp. 1-6, doi: 10.1109/ISACV.2017.8054924.

Jardim, S., António, J., & Mora, C. (2022). Graphical Image Region Extraction with K-Means Clustering and Watershed. Journal of Imaging, 8(6), 163. https://doi.org/10.3390/jimaging8060163

Improved fuzzy C-Means and K-Means algorithms for texture and boundary segmentation. (2018, October 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8751905

Et-Taleby, A., Boussetta, M., & Benslimane, M. (2020). Faults Detection for Photovoltaic Field Based on K-Means, Elbow, and Average Silhouette Techniques through the Segmentation of a Thermal Image. International Journal of Photoenergy, 2020, 1–7. https://doi.org/10.1155/2020/6617597

Image segmentation using K-means color quantization and density-based spatial clustering of applications with noise (DBSCAN) for hotspot detection in photovoltaic modules. (2016, November 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7848290

Naous, T., Sarkar, S., Abid, A., & Zou, J. (2022). Clustering plotted data by image segmentation. https://openaccess.thecvf.com/content/CVPR2022/html/Naous_Clustering_Plotted_Data_by_Image_Segmentation_CVPR_2022_paper.html

Hung, P. D., Ngoc, N. D., & Hanh, T. D. (2019). K-means Clustering Using R A Case Study of Market Segmentation. https://doi.org/10.1145/3317614.3317626

Dayang, P., KOUYIM MELI, A. Evaluation of Image Segmentation Algorithms for Plant Disease Detection https://www.researchgate.net/profile/Armandine-Sorel-Kouyim-Meli/publication/356055094_Evaluation_of_Image_Segmentation_Algorithms_for_Plant_Disease_Detection/links/619a50e6d7d1af224b142d6a/Evaluation-of-Image-Segmentation-Algorithms-for-Plant-Disease-Detection.pdf

Cheng, G., & Wei, J. (2019). Color Quantization application based on K-Means in remote sensing image processing. Journal of Physics Conference Series, 1213(4), 042012. https://doi.org/10.1088/1742-6596/1213/4/042012
.
Paoletti1, M., Haut1, J., Plaza1, J., Plaza1, A. Yinyang K-means clustering for hyperspectral image analysis. https://www2.umbc.edu/rssipl/people/aplaza/Papers/Conferences/2017.CMMSE.Yingyang.pdf