Image Segmentation Using an Evolutionary Algorithm

*An evaluation of Differential Evolution with different objective function to solve image segmentation

BS Steyn
(Department of Computer Science)
Stellenbosch University
Stellenbosch, South Africa
21740178@sun.ac.za

***Abstract*—This paper compares two Differential Evolution variants for image segmentation, using entropy and inter-class variance as fitness functions. Segmentation quality is evaluated using Dice Score, Jaccard Index, and convergence behaviour. Results show that the inter-class variance-based approach consistently achieves superior accuracy and stability. Statistical analysis confirms the significance of these performance differences, supporting the use of inter-class variance for threshold-based segmentation.**

I. Introduction

Image segmentation [2] is a fundamental task in image processing, aimed at partitioning an image into distinct regions that are homogeneous in terms of intensity or texture. Threshold-based segmentation remains a widely used approach due to its simplicity and efficiency. However, determining optimal threshold values becomes increasingly complex in multi-level segmentation tasks, particularly in the presence of noise or varying contrast. Meta-heuristic algorithms offer robust solutions to such optimization problems by performing global search without requiring gradient information. Differential Evolution (DE) [7] is one such algorithm, known for its simplicity, convergence reliability, and effectiveness in continuous search spaces. The performance of DE in segmentation depends largely on the choice of fitness function, which guides the search toward desirable segmentation outcomes.

This paper investigates the use of DE for multi-level image segmentation using two fitness formulations: entropy [4] and inter-class variance [6]. Entropy quantifies the information content within segmented regions, while inter-class variance, as used in Otsu's method, measures the separability between classes. The goal is to compare these two fitness strategies in terms of segmentation quality, convergence behaviour, and statistical significance across multiple test runs.

The remainder of this paper is structured as follows: Section II provides a background for this paper. Section III provides the details on the implementation used in this paper. Section IV provides the evaluation criteria and method used in this paper. Section V provides the findings of this paper. Section VI concludes the paper.

II. Background

This section provides background on the individual algorithms and the problem being solved.

A. Differential Evolution

Differential Evolution (DE) [7] is a population-based meta-heuristic designed for global optimization in continuous domains. The algorithm operates by evolving a population of candidate solutions through mutation, crossover, and selection. Unlike traditional evolutionary algorithms, DE utilizes scaled differences between individuals to guide the search process, which enhances convergence toward global optima.

Each candidate solution in DE is represented as a real-valued vector. Mutation is performed by generating a mutant vector as follows:

$$\boldsymbol{v}_i = \boldsymbol{x}_{r_1} + F \cdot (\boldsymbol{x}_{r_2} - \boldsymbol{x}_{r_3}) \tag{1}$$

where $\boldsymbol{x}_{r_1}$, $\boldsymbol{x}_{r_2}$, and $\boldsymbol{x}_{r_3}$ are distinct individuals randomly selected from the population, and F is a scaling factor controlling the amplification of the differential variation. The crossover step generates a trial vector $\boldsymbol{u}_i$ by combining elements from the mutant vector and the current target vector $\boldsymbol{x}_i$ based on a crossover probability CR.

The final candidate for the next generation is selected using greedy selection:

$$\boldsymbol{x}_i^{(t+1)} = \begin{cases} \boldsymbol{u}_i, & \text{if } f(\boldsymbol{u}_i) > f(\boldsymbol{x}_i) \\ \\ \boldsymbol{x}_i, & \text{otherwise} \end{cases} \tag{2}$$

where $f(\cdot)$ denotes the objective function. This selection mechanism ensures that the quality of the population does not deteriorate over generations. DE has been shown to be robust and efficient across a wide range of optimization problems. Its ability to maintain a balance between exploration and exploitation without requiring gradient information makes it suitable for solving non-linear, multi-modal, and noisy objective functions. The performance of DE is primarily governed by the choice of F, CR, and population size, which must be carefully tuned for each problem instance.

B. Otsu's Method

Otsu's method [5] is a classical thresholding technique used for grayscale image segmentation. The method aims to determine an optimal threshold value that maximizes the separability between foreground and background pixel intensities. This is achieved by maximizing the inter-class variance, which serves as a proxy for class separability in the histogram of the image.

Given a grayscale image with intensity values in the range $[0, L-1]$, where L denotes the number of grey levels, the histogram of the image can be interpreted as a probability distribution. Let the threshold be denoted by t, and let $\omega_0(t)$ and $\omega_1(t)$ represent the probabilities of the background and foreground classes, respectively. The corresponding class means are denoted as $\mu_0(t)$ and $\mu_1(t)$. The global mean intensity of the image is defined as μ_T. The inter-class variance is given by:

$$\sigma_b^2(t) = \omega_0(t) \cdot \omega_1(t) \cdot [\mu_0(t) - \mu_1(t)]^2 \tag{3}$$

Otsu's method selects the threshold t^* that maximizes $\sigma_b^2(t)$:

$$t^* = \underset{t \in [0, L-1]}{\arg\max}\, \sigma_b^2(t) \tag{4}$$

This approach assumes that the image histogram is bimodal, where two distinct peaks correspond to the background and foreground classes. The method is deterministic and computationally efficient, making it suitable for real-time segmentation tasks.

However, Otsu's method is limited in that it performs only binary thresholding by default. Although extensions such as multi-level Otsu exist, they are restricted to a small number of thresholds and are sensitive to noise, lighting variation, and non-bimodal intensity distributions. These limitations reduce the applicability of Otsu's method in scenarios requiring adaptive or multi-region segmentation.

III. Implementation

This section provides an overview of the implementation approach using Differential Evolution.

A. Problem Representation

Each candidate solution is represented as a vector of real-valued thresholds. These thresholds define the boundaries between intensity-based segments. For an image with n thresholds, each solution $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$ is constrained such that $0 < x_1 < x_2 < \ldots < x_n < 255$. The thresholds are rounded to integer values during evaluation.

B. Fitness Evaluation

Two fitness evaluation strategies are considered for threshold-based image segmentation, namely entropy-based fitness and inter-class variance-based fitness. Both approaches aim to evaluate the quality of a threshold vector by measuring how effectively it separates the grayscale intensity histogram into informative or distinct regions.

1) Entropy-Based Fitness: The entropy-based fitness function measures the information content of the segmented image. Let $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]$ denote a sorted vector of n thresholds. These thresholds partition the image histogram into $n+1$ disjoint regions. For region i, let $P_i(k)$ represent the normalized frequency of intensity level k within the region. The entropy of region i is defined as:

$$H_i = -\sum_k P_i(k) \log_2 P_i(k) \tag{5}$$

The overall entropy-based fitness is computed as a weighted average over all regions:

$$f_{\text{entropy}}(\boldsymbol{t}) = \sum_{i=0}^{n} \frac{|\mathcal{R}_i|}{|\mathcal{I}|} \cdot H_i \tag{6}$$

where $|\mathcal{R}_i|$ is the number of pixels in region i and $|\mathcal{I}|$ is the total number of pixels in the image. This fitness function favours threshold sets that maximize the diversity of pixel intensities within each region, thereby enhancing visual contrast and detail preservation.

2) Inter-Class Variance-Based Fitness: The inter-class variance-based fitness is adapted from Otsu's method. This criterion aims to maximize the separability between different regions. For a threshold vector $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]$, let ω_i and μ_i denote the probability and mean intensity of region i, respectively. Let μ_T denote the global mean intensity of the image. The inter-class variance is defined as:

$$\sigma_b^2 = \sum_{i=0}^{n} \omega_i(\mu_i - \mu_T)^2 \tag{7}$$

The fitness function is given by:

$$f_{\text{variance}}(\boldsymbol{t}) = \sigma_b^2 \tag{8}$$

This formulation promotes threshold values that yield regions with highly distinct average intensities, thereby producing segmentations with clear boundaries. This approach is particularly effective when the image histogram is multi-modal and the regions to be separated are statistically distinct.

C. Segmentation and Post-Processing

After the Differential Evolution algorithm converges, the best threshold vector $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]$ is selected to perform image segmentation. Each pixel in the grayscale image is assigned to one of the $n+1$ regions based on its intensity value relative to the thresholds. Let I_{ij} denote the intensity value of the pixel located at position (i, j) in the image. The segmented image is obtained by mapping each pixel to a corresponding region index k as follows:

$$s_{ij} = \begin{cases} 0, & \text{if } I_{ij} < t_1 \\ 1, & \text{if } t_1 \leq I_{ij} < t_2 \\ \vdots & \\ n, & \text{if } I_{ij} \geq t_n \end{cases} \tag{9}$$

To enhance visual distinctness, each region index k is mapped to an output intensity value r_k, where the values are distributed evenly across the range $[0, 255]$. The output intensity for region k is computed as:

$$r_k = \left\lfloor \frac{255 \cdot k}{n+1} \right\rfloor \tag{10}$$

The final segmented image is formed by assigning r_k to each pixel in region k. This approach ensures that the segmented regions are visually distinguishable and preserves the relative ordering of intensity levels. The result is a multi-level segmented image suitable for subsequent analysis or visualization.

IV. Evaluation Procedure

This section outlines the empirical procedure followed to evaluate the image segmentation performance of the Differential Evolution algorithm. The evaluation is conducted using three metrics: Dice Score, Jaccard Index, and average inter-individual variance per generation. Each metric captures a distinct characteristic of algorithm performance, namely segmentation accuracy, region overlap, and population diversity. The testing procedure includes multiple independent runs to account for the stochastic nature of the algorithm.

A. Ground Truth-Based Evaluation

The Dice Score [1] and Jaccard Index [3] are used to quantify the similarity between the segmented image and a reference ground truth segmentation. Let $\mathcal{S}$ denote the set of pixels identified as foreground in the segmented image and let $\mathcal{G}$ denote the corresponding set in the ground truth image. The Dice Score is calculated as:

$$\text{Dice} = \frac{2|\mathcal{S} \cap \mathcal{G}|}{|\mathcal{S}| + |\mathcal{G}|} \tag{11}$$

The Jaccard Index, also referred to as Intersection over Union (IoU), is defined as:

$$\text{Jaccard} = \frac{|\mathcal{S} \cap \mathcal{G}|}{|\mathcal{S} \cup \mathcal{G}|} \tag{12}$$

Both metrics are computed after binarising the segmented image, where all non-zero pixel values are considered foreground. These metrics are calculated for each run and then averaged over all runs to determine the overall segmentation performance.

B. Population Diversity Evaluation

The average variance between individuals in the population is used as a measure of population diversity throughout the evolution process. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ denote the M individuals in a given generation, where each individual is a threshold vector. The average pairwise Euclidean distance is computed as:

$$\text{Diversity} = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \tag{13}$$

This metric is evaluated at each generation and averaged across all generations and independent runs. The purpose of this metric is to assess the exploratory behaviour of the algorithm and to determine whether the population retains sufficient diversity to avoid premature convergence.

C. Experimental Setup

All experiments are conducted using grayscale images where the Otsu's mask is used as the ground truth mask. Each configuration is executed independently 30 times to account for the stochastic nature of the algorithm. Metrics are computed per run and aggregated using mean and standard deviation. The goal is to compare segmentation accuracy and convergence characteristics across different fitness functions and threshold counts.

D. Wilcoxon Signed-Rank Test

A Wilcoxon signed-rank test [8] was performed to determine whether the segmentation performance differed significantly between the two DE variants. The test was applied to paired Dice and Jaccard scores across 30 independent runs.

V. Results

This subsection provides the comparative results of the experiments as well as a detailed discussion of the obtained results. The results reported in this section are based on `coins.jpg`. This image is selected for discussion, as the performance trends observed across other test images followed the same pattern. Therefore, analysing the results for `coins.jpg` is considered representative of the overall algorithm behaviour.

A. Dice and Jaccard

Consider the table:

TABLE I
TABLE SHOWING THE AVERAGE DICE AND JACCARD SCORES OVER 30 INDEPENDENT RUNS

Algorithm	Dice	Jaccard
$DEInter-Class$	0.9973195171467528	0.994659361240881
$DEEntropy$	0.2847092160250243	0.17660982065639677

Table I shows that using inter-class variance as a fitness function leads to better segmentation when considering the segmentation output of the Otsu algorithm as a sudo true mask. This performance can be partially attributed to the use of inter-class variance in both Otsu's method and the corresponding DE variant. The results can further be examine in the Figure 1:

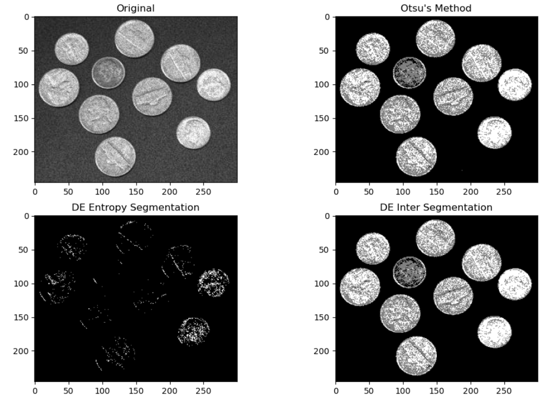

Fig. 1. Figure showing segmentation results along with the original image.

Figure 1 illustrates the segmentation outputs for `coins.jpg`. The DE variant using inter-class variance successfully captures the full shape of each coin and exhibits consistent boundary preservation. In contrast, the entropy-based segmentation produces fragmented and incomplete object masks, primarily due to its sensitivity to local texture and noise. These visual differences explain the performance gap in Dice and Jaccard scores and confirm the superior suitability of inter-class variance for object-level image segmentation tasks.

B. Convergence

This subsection looks at the convergence of the two DE variations. Figure 2 illustrates:

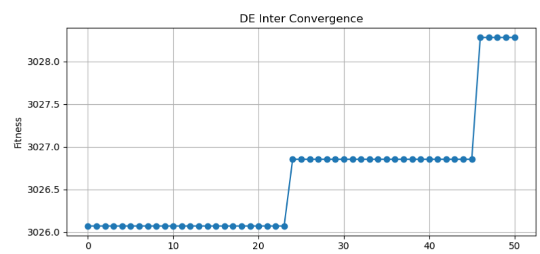

Fig. 2. Figure showing the convergence for inter-class variance

Figure 2 shows the convergence behaviour of the DE variant using inter-class variance as the fitness criterion. The fitness remains constant during the initial generations, indicating exploratory behaviour. A gradual improvement is observed in generation 22, followed by a prolonged plateau, suggesting local exploitation. A final substantial fitness increase occurs in generation 45, after which convergence is achieved. This pattern demonstrates the algorithm's capacity to escape local optima and discover improved threshold sets, albeit with

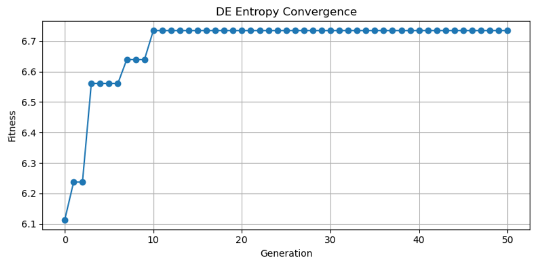

Fig. 3. Figure showing the convergence for Entropy

delayed convergence. Figure 3 shows the convergence trend for the entropy-based variant: Figure 3 displays the convergence curve for the DE variant using entropy-based fitness. A sharp improvement in fitness is observed within the first ten generations, after which the fitness value plateaus. This indicates rapid convergence toward a local optimum. The absence of further improvements suggests limited exploratory capability or premature convergence. Although entropy provides a fast optimization path, the resulting segmentations often lack spatial coherence and completeness, as confirmed by the visual and quantitative evaluation.

C. Wilcoxon Signed-Rank Test

The test results indicated that there was a statistically significant difference between both the Dice and Jaccard. This was based on the average Dice and Jaccard scores over 30 independent runs.

VI. Conclusion

This paper evaluated the performance of Differential Evolution applied to image segmentation using two distinct fitness functions: entropy and inter-class variance. Experimental results demonstrated that the inter-class variance-based approach consistently outperformed the entropy-based variant in both Dice and Jaccard scores. This improvement is attributed to the ability of inter-class variance to promote global class separability, resulting in more coherent object-level segmentation. In contrast, the entropy-based fitness was prone to overfitting local intensity variations, leading to fragmented segmentations. Convergence analysis further confirmed that the entropy-based variant converged rapidly but prematurely, whereas the inter-class variant achieved better final solutions through delayed but effective exploration. Statistical validation using the Wilcoxon signed-rank test confirmed that the performance difference was significant. These findings support the use of inter-class variance as a robust objective for threshold-based segmentation using evolutionary algorithms.

References

[1] Lee R. Dice. “Measures of the amount of ecologic association between species”. In: *Journal of Ecology* 26.3 (1945), pp. 297–302.

[2] Robert M. Haralick and Linda G. Shapiro. “Image segmentation techniques”. In: *Computer Vision, Graphics, and Image Processing* 29.1 (1985), pp. 100–132.

[3] Paul Jaccard. “Étude comparative de la distribution florale dans une portion des Alpes et des Jura”. In: *Bulletin de la Société Vaudoise des Sciences Naturelles* 37 (1901), pp. 547–579.

[4] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong. “A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram”. In: *Computer Vision, Graphics, and Image Processing* 29.3 (1985), pp. 273–285.

[5] Nobuyuki Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66.

[6] Zongbao Peng, Lei Wang, Liang Tong, Haijun Zou, Daoyong Liu, and Cheng Zhang. “Multi-threshold image segmentation of 2D OTSU inland ships based on improved genetic algorithm”. In: *PLOS ONE* 18.8 (2023), e0290750. DOI: 10.1371/journal.pone.0290750. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0290750.

[7] Rainer Storn and Kenneth Price. “Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces”. In: *Journal of Global Optimization* 11.4 (1997), pp. 341–359.

[8] Frank Wilcoxon. “Individual comparisons by ranking methods”. In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83.