

A Bipolar Fuzzy Approach to Image Segmentation: Enhancing Similarity Measures and Entropy Computation

Stephen Macharia Gathigi, Moses Nderitu Gichuki, Kewamoi Chesire Sogomo

Department of Mathematics, Egerton University, Nakuru, Kenya
Email: sgathigi@egerton.ac.ke, gichuki@egerton.ac.ke, kesecho@yahoo.com

How to cite this paper: Gathigi, S.M., Gichuki, M.N. and Sogomo, K.C. (2025) A Bipolar Fuzzy Approach to Image Segmentation: Enhancing Similarity Measures and Entropy Computation. *Advances in Pure Mathematics*, 15, 651-666.
<https://doi.org/10.4236/apm.2025.1510034>

Received: July 15, 2025

Accepted: September 25, 2025

Published: September 28, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Image segmentation is a fundamental process in digital image analysis, with applications in object recognition, medical imaging, and computer vision. Traditional segmentation techniques often struggle with uncertainty, imprecise boundaries, and misclassified regions due to their inability to effectively model both positive and negative information. This study introduces a bipolar fuzzy-based computational approach that enhances segmentation accuracy by incorporating dual membership functions to represent both the presence and absence of image features. To further improve segmentation robustness, we extend classical similarity measures by formulating a Bipolar Fuzzy Jaccard Similarity, which quantifies both positive and negative membership interactions, leading to more precise region classification. Additionally, a novel Bipolar Rényi Entropy (BRE) measure is developed to capture uncertainty in segmentation by integrating bipolar fuzzy probability distributions, allowing for adaptive sensitivity to dominant and rare features. Experimental validation on grayscale image datasets demonstrates the superiority of the proposed approach over conventional fuzzy and graph-based segmentation methods, particularly in applications requiring high precision, such as medical imaging and AI-driven pattern recognition. The integration of bipolar fuzzy similarity and entropy measures provides a powerful computational framework for more accurate and interpretable image segmentation.

Keywords

Fuzzy, Bipolar Fuzzy, Similarity Measure, Entropy

1. Introduction

Image segmentation plays a crucial role in digital image processing, enabling the

extraction of meaningful structures from complex image data. Traditional segmentation methods, including thresholding, region-growing, and edge detection, often suffer from challenges in handling imprecise boundaries and noise.

The limitations of crisp set theory become evident when attempting to model real-world phenomena with inherent vagueness, ambiguity, and subjectivity. As [1] and [2] suggest, rigid, dichotomous logic often fails to capture the complexity of human thought, perception, and decision-making, which are rarely confined to absolute categories. Traditional mathematical models struggle to balance precision and interpretability, often requiring an unequivocal structure while also needing to reflect the fluid nature of real-world uncertainties.

To bridge this gap, fuzzy set theory offers a more adaptive and expressive framework by allowing for gradual membership, enabling models to handle partial truths and uncertainty more effectively. Unlike crisp logic, fuzzy logic preserves stability even under minor variations in assumptions, making it particularly suitable for complex systems such as artificial intelligence, decision support systems, and natural language processing. By incorporating fuzzy logic, mathematical models gain the ability to represent reality more accurately, ensuring that they remain both mathematically rigorous and practically relevant in domains where strict boundaries fail to capture nuanced interpretations.

In many real-life domains, it is necessary to be able to deal with bipolar information – information possessing both what is possible and what is not possible. It is noted in [3] that positive information represents what is granted to be possible, while negative information represents what is granted to be impossible. According to [3], dual Boolean logic lacks the representational and reasoning capabilities for directly modeling the coexistence and interaction of bipolar relationships. While fuzzy logic may model uncertainty, it is limited when it comes to describing polarity. To resolve these two features, bipolar fuzzy set theory combines both the fuzziness and polarity into a unified model, hence providing a theoretical basis for bipolar clustering, conflict resolution and coordination [4].

This paper presents a bipolar fuzzy-based computational approach that enhances classical fuzzy segmentation models by incorporating dual membership functions that represent both the presence and absence of image features. The proposed method improves segmentation accuracy and potentially facilitates precise region classification and boundary detection outcomes. Additionally, the approach integrates Bipolar Fuzzy Jaccard Similarity to refine similarity assessments and Bipolar Rényi Entropy to quantify uncertainty, making it highly effective for complex image segmentation tasks. Experimental validation demonstrates its superior performance in applications such as medical imaging, remote sensing, and AI-driven pattern recognition, where precise segmentation is crucial for analysis and decision-making.

2. Theoretical Framework

This section presents the foundational concepts of bipolar fuzzy sets, similarity

measures, and entropy functions, which form the basis of the proposed computational approach.

2.1. Fundamental Concepts of BFS

A few basic definitions are given here below as found in [5].

Let X be a nonempty set. A pair $\mu = (\mu^+, \mu^-)$ is called a bipolar-valued fuzzy set (or bipolar fuzzy set) in X if $\mu^+ : X \rightarrow [0, 1]$ and $\mu^- : X \rightarrow [-1, 0]$ are mappings.

In particular, the bipolar fuzzy empty set (resp. the bipolar fuzzy whole set), denoted by $0_{bp} = (0_{bp}^+, 0_{bp}^-)$ [resp. $1_{bp} = (1_{bp}^+, 1_{bp}^-)$], is a bipolar fuzzy set in X defined by

$$\forall x \in X, 0_{bp}^+(x) = 0 = 0_{bp}^-(x) \quad [\text{resp. } 1_{bp}^+(x) = 1 \text{ and } 1_{bp}^-(x) = -1]$$

The collection of all bipolar fuzzy sets of the set X will be denoted by $BPF(X)$.

2.2. Entropy and Similarity Measures

A fundamental question in the realm of fuzzy sets involves the use of relative comparisons to calculate the similarities and distances between fuzzy sets [6] [7]. Similarity involves recognizing patterns and making associations which enable one to classify objects and concepts, particularly in the domain of classification and clustering. In these applications, an unknown object is assigned to a particular category if its similarity measure to objects within that category is higher than its similarity to objects in other categories [8].

2.2.1. Similarity

In essence, the similarity of two fuzzy sets is 1 if they are identical *i.e.*, if they both contain the same values with the same degree of membership, and 0 if they have nothing in common, *i.e.*, they do not contain any of the same values. These two properties are referred to as reflexivity and overlapping respectively.

Definition (Similarity Measure)

Intuitively, the similarity of two fuzzy sets is 1 if they are identical *i.e.*, if they both contain the same values with the same degree of membership, and 0 if they have nothing in common, *i.e.*, they do not contain any of the same values. These two properties are referred to as reflexivity and overlapping respectively.

A function $S : A \times B \rightarrow [0, 1]$ is called a similarity measure if

1) Reflexivity: $S(A, A) = 1$, meaning an element always has full similarity to itself

2) Symmetry: $S(A, B) = S(B, A) \quad \forall A, B \in X$. This property ensures that similarity is bidirectional

3) Overlapping: $S(A, B) \geq \max\left(\frac{|A \cap B|}{|A|}, \frac{|A \cap B|}{|B|}\right)$. This property ensures that

shared characteristics contribute strongly to the similarity score

4) Transitivity: if $S(A, B) \geq \alpha$ and $S(B, C) \geq \alpha$ then $S(A, C) \geq \alpha$. This

property is important for clustering and classification, ensuring that if A is similar to B and B is similar to C , then A should be similar to C .

It is not necessary for a similarity measure to have all these properties [9]. There are situations when symmetry does not need to be satisfied and in other situations, it has been argued whether transitivity is necessary or even useful.

2.2.2. The Jaccard Similarity Measure

The Jaccard similarity/index compares the similarity between two fuzzy sets by calculating the size of their intersection divided by the size of their union. According to [10], the fuzzy Jaccard similarity between two fuzzy sets A and B in a universal set X is defined as

$$J(A, B) = \frac{\sum_{x \in X} \min(\mu_A(x), \mu_B(x))}{\sum_{x \in X} \max(\mu_A(x), \mu_B(x))}$$

Instead of simply counting shared elements, fuzzy Jaccard similarity is computed using the ratio of the sum of the minimum membership values (fuzzy intersection) to the sum of the maximum membership values (fuzzy union).

2.3. Entropy

Entropy, a fundamental concept in information theory, measures the uncertainty or randomness of a system. Shannon entropy, introduced in 1948, is the most well-known measure of information uncertainty and forms the basis of modern information theory [2]. However, Shannon entropy has limitations when dealing with higher-order moments of probability distributions, long-tailed distributions, or multi-fractal systems. In such scenarios, a more generalized entropy measure is required.

To address these limitations, Alfred Rényi introduced the Rényi entropy in 1961 as a one-parameter generalization of Shannon entropy [11]. Rényi entropy provides flexibility in quantifying information by introducing an order parameter, α , which allows fine-tuning sensitivity to different probabilities within a distribution. This feature makes it highly useful in fields like machine learning, cryptography, statistical mechanics, quantum computing, and image processing.

2.3.1. Classical Shannon and Rényi Entropies

The Rényi entropy of order α for a probability distribution $P = \{p_1, p_2, \dots, p_n\}$ is defined as $H_\alpha(P) = \frac{1}{1-\alpha} \ln \sum_{i=1}^n p_i^\alpha$ where $\alpha > 0, \alpha \neq 1$ is the order of the entropy, controlling the weighting of probabilities.

Note

- 1) When $\alpha \rightarrow 1$ Rényi entropy converges to Shannon entropy

$$H_1(P) = \sum_{i=1}^n p_i \ln p_i$$

- 2) When $\alpha \rightarrow \infty$ Rényi entropy focusses only on the most probable event.

The motivation behind Rényi entropy arises from its ability to:

1) Extend Shannon Entropy: While Shannon entropy assumes an equal weighting of probabilities, Rényi entropy allows adjustable sensitivity to rare or dominant events based on the value of α .

2) Enhance Information Divergence Measures: The Rényi divergence, a generalization of Kullback-Leibler (KL) divergence, provides a more robust way to measure differences between probability distributions [12].

3) Improve Uncertainty Quantification in Fuzzy and Quantum Systems: Many real-world problems involve uncertainty beyond probabilistic randomness—for instance, fuzzy logic and quantum mechanics, where entropy measures must be adapted accordingly [13].

4) Enable Robust Feature Selection and Learning Algorithms: In machine learning and deep learning, Rényi entropy helps in feature selection, adversarial learning, and information-theoretic optimization [12].

5) Enhance Image Processing Methods: Rényi entropy is particularly effective in image segmentation and edge detection, where different levels of entropy sensitivity are required to enhance contrast and extract meaningful structures [14].

Given these advantages, Rényi entropy has become a fundamental tool in information science, artificial intelligence, and complex system analysis.

2.3.2. Entropy on Fuzzy Sets

In the context of fuzzy sets, the concept of the entropy of a fuzzy set was first introduced by [6] as a means of quantifying the degree of fuzziness within a fuzzy set. This laid the foundation for measuring the uncertainty or ambiguity present in fuzzy systems. The formal axiomatization of fuzzy entropy was later developed by [15], who drew parallels between fuzzy entropy and Shannon probability entropy, interpreting it as a measure of the amount of information contained in a fuzzy set.

Further advancements were made by [16] who proposed that the entropy of a fuzzy set could be calculated based on its distance from the nearest crisp (non-fuzzy) set, providing a geometric perspective on fuzziness measurement. Yager [17] [18] offered an alternative interpretation, defining entropy as the distance between a fuzzy set and its complement. According to [16], any meaningful measure of fuzziness should reflect the lack of distinction between a fuzzy set and its negation, leading to the proposal of a metric for quantifying the distance between a fuzzy set and its complement.

2.3.3. Formulation of Entropy on Fuzzy Sets

Formally, in [8] the entropy $H(A)$ of a fuzzy set A defined on a universal set X with membership function is expressed as:

$$H(A) = - \sum_{x \in X} [\mu_A(x) \log \mu_A(x) + (1 - \mu_A(x)) \log (1 - \mu_A(x))]$$

where $\mu_A(x)$ represents the membership degree of element x in the fuzzy set A .

Illustrative example

Define $d_p(A, A^c) = \left[\sum_{i=1}^n |\mu_A(x_i) - \mu_{A^c}(x_i)|^p \right]^{\frac{1}{p}}$, $p = 1, 2, 3, \dots$. Then the measure of fuzziness of A can be defined as

$$f_p(A) = 1 - \frac{d_p(A, A^c)}{\|\text{sup}(A)\|}$$

For $p = 1$, $d_p(A, A^c)$ yields the Hamming metric.

$$d_1(A, A^c) = \sum_{i=1}^n |\mu_A(x_i) - \mu_{A^c}(x_i)|$$

Since $\mu_{A^c}(x) = 1 - \mu_A(x)$, this becomes $d_p(A, A^c) = \sum_{i=1}^n |2\mu_A(x_i) - 1|$.

For $p = 2$ we have the Euclidean metric

$$d_2(A, A^c) = \left[\sum_{i=1}^n |\mu_A(x_i) - \mu_{A^c}(x_i)|^2 \right]^{\frac{1}{2}} \text{ and}$$

since $\mu_{A^c}(x) = 1 - \mu_A(x)$, then this becomes

$$d_2(A, A^c) = \left[\sum_{i=1}^n (2\mu_A(x_i) - 1)^2 \right]^{\frac{1}{2}}$$

2.4. Min-Max Normalization in Image Processing

Min-max normalization is a widely used technique in image processing, machine learning, and data preprocessing for transforming data into a predefined range, typically $[0, 1]$ or $[-1, 1]$. This method is particularly effective in scenarios where pixel intensity values or feature distributions need to be scaled for consistency across different datasets.

This sub-section reviews min-max normalization from the perspective of fuzzy set theory, highlighting its applications, limitations, and recent advancements.

2.4.1. Grayscale Images

In digital topology, a grayscale image can be represented as a digital image where pixel intensities are defined on a discrete grid, and adjacency relations determine connectivity. According to [19], a grayscale image is a function $I: X \rightarrow [0, L]$ where

- 1) $X \subseteq \mathbb{Z}^n$ is a finite subset of the integer lattice representing the image domain.
- 2) L is the maximum intensity level.
- 3) $I(x)$ represents the grayscale intensity value at pixel x , where 0 corresponds to black, L corresponds to white, and intermediate values represent shades of gray [20].

2.4.2. Graph-Based Representation of a Grayscale Image

In Graph-based approach, a grayscale image can be modeled as an undirected

graph $G = (V, E)$ where the vertices V represent pixels in the image, E define the adjacency relations.

Each vertex will possess a grayscale intensity value $I(v) \in [0, 255]$ with 0 corresponding to a low intensity pixel (black) and 255 corresponding to the highest intensity pixel (white). All intermediate values represent shades of gray [20].

Illustrative example

A 3×3 grayscale image

(50)	(120)	(200)
(80)	(150)	(220)
(90)	(180)	(250)

will have a graphical representation with 4-adjacency relation as shown below

(50)	---	(120)	---	(200)
(80)	---	(150)	---	(220)
(90)	---	(180)	---	(250)

For the same image, the graphical representation with 8-adjacency relation will be

(50)	---	(120)	---	(200)
\		/		/
(80)	---	(150)	---	(220)
/		\		\
(90)	---	(180)	---	(250)

2.4.3. Mathematical Formulation

The min-max normalization ensures that data is linearly scaled within $[0, 1]$, allowing direct interpretation as fuzzy membership values [21]. This transformation is defined by

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where;

- 1) x is the original data value,
- 2) x' is the normalized value in the range $[0, 1]$ and,
- 3) $\min(x)$ and $\max(x)$ are the minimum and maximum values in the dataset.

2.4.4. Applications in Fuzzy-Based Image Processing

Min-max normalization is widely applied in fuzzy-based image processing, where it ensures that data is transformed into a standardized range suitable for fuzzy logic operations. By normalizing image intensities, feature values, or clustering inputs, it enhances the effectiveness of fuzzy techniques in image segmentation, similarity measurement, and contrast enhancement. Two of the areas where this formulation finds application are:

- 1) Fuzzy clustering techniques, such as Fuzzy C-Means (FCM), rely heavily on min-max normalization to standardize input features before assigning fuzzy

memberships to clusters. Without proper normalization, feature scales may distort clustering results, leading to misclassification [21] [22].

2) In fuzzy-based image retrieval, similarity measures compute distances between images in a feature space. Min-max normalization ensures that feature values fall within a comparable range, enhancing the accuracy of fuzzy histogram similarity measures and fuzzy correlation techniques [23].

2.4.5. Limitations of Min-Max Normalization in Applications

Despite its effectiveness in fuzzy-based image processing, min-max normalization has several limitations that can affect its performance in certain applications. One major drawback is its sensitivity to outliers, as extreme values in the dataset disproportionately influence the scaling range, leading to distorted fuzzy membership values [24]. This can be particularly problematic in medical imaging or low-contrast image analysis, where small variations in intensity carry significant meaning. Additionally, min-max normalization assumes a linear transformation, which may not always be suitable for complex or non-linearly distributed data, such as images with varying lighting conditions or textures [23]. The technique also depends on the fixed minimum and maximum values within the dataset, which can make it ineffective for dynamically changing environments, such as real-time image processing or adaptive fuzzy systems [25]. Moreover, by strictly mapping values between 0 and 1, min-max normalization can lead to loss of contrast and information, particularly when applied to images with already narrow intensity distributions [26].

As a result, alternative normalization methods, such as sigmoid-based transformations which uses a logistic function to transform data, preserving fine-grained differences in middle-range values [27] or adaptive fuzzy normalization, which dynamically adjusts min-max scaling based on data distributions [28] are often preferred for applications requiring greater robustness and flexibility.

3. Results

3.1. Conceptual Framework for Bipolar Fuzzy Similarity Measures

In the context of fuzzy image representation, each pixel intensity $I(p_i)$ in an image is transformed into a single-valued membership function. To extend this idea into the bipolar fuzzy context, each pixel will be assigned two membership values,

$$B(p_i) = (\mu^+(p_i), \mu^-(p_i))$$

where $\mu^+(p_i)$ and $\mu^-(p_i)$ represent the degree of relevance to the desired feature and undesired features respectively.

For grayscale images, the bipolar fuzzy memberships functions may be defined as

$$\mu^+(p_i) = \frac{I(p_i) - \min(I)}{\max(I) - \min(I)}$$

Where:

- $I(p_i)$ is the grayscale intensity of the pixel p_i .
- $\min(I)$ is the lowest intensity in the image (typically 0 for black intensity).
- $\max(I)$ is the highest intensity value in the image (usually 255 for white intensity).
- $\mu^+(p_i) \in [0, 1]$.

This proposed transformation can be shown by taking $I(p_i) = 200$. In this case

$$\mu^+(p_i) = \frac{200 - 0}{255 - 0} = 0.784$$

(a pixel with higher intensity will have a higher positive membership).

The negative membership function $\mu^-(p_i)$ determines how strongly a pixel belongs to the background, noise, or undesired features. By definition this will be given by

$$\mu^-(p_i) = -(1 - \mu^+(p_i)) \in [-1, 0]$$

This formulation ensures that brighter pixels have higher μ^+ - values and lower μ^- values and vice versa.

In example above,

$$\text{if } \mu^+(p_i) = 0.784 \text{ then}$$

$$\mu^-(p_i) = -(1 - 0.784) = -0.216$$

This corresponds perfectly with the fact that since the positive membership measures the presence of a desired feature, the negative membership naturally represents the presence of the counter property of that feature. Therefore, a pixel of higher intensity will have a lower negative membership, representing higher presence of the low intensity characteristics.

3.1.1. Bipolar Fuzzy Similarity Measures

To compare two bipolar fuzzy images B_1 and B_2 we extend classical fuzzy similarity measures to account for both positive and negative memberships. The general form of a bipolar fuzzy similarity measure may be expressed as:

$$S_B(B_1, B_2) = \alpha S^+(F_1, F_2) - \beta S^-(F_1, F_2)$$

where:

- $S^+(F_1, F_2)$ is the similarity between positive memberships.
- $S^-(F_1, F_2)$ is the similarity between negative memberships.
- α, β are weighting factors that determine the importance of positive and negative similarities.

In Section 4.2.3, a characterization of Bipolar Jaccard similarity for Bipolar fuzzy sets was proposed. This was defined for two bipolar fuzzy sets A and B may as

$$s(A, B) = \sum_{i=1}^n \left(\frac{A^+(x_i) \wedge B^+(x_i), A^-(x_i) \wedge B^-(x_i)}{A^+(x_i) \vee B^+(x_i), A^-(x_i) \vee B^-(x_i)} \right)$$

This characterization is equivalent to

$$s(A, B) = \frac{\sum \min(A^+(x_i), B^+(x_i)) - \sum \min(A^-(x_i), B^-(x_i))}{\sum \max(A^+(x_i), B^+(x_i)) + \sum \max(A^-(x_i), B^-(x_i))}$$

In this formulation, the numerator captures the overlap in positive memberships while taking care of dissimilarities in negative memberships, while the denominator normalizes the measure to ensure that $s(A, B) \in [-1, 1]$.

3.1.2. Bipolar Jaccard Similarity for Grayscale Images Using Min-Max Normalization

In this section, we present a systematic step-by-step approach for computing the bipolar Jaccard similarity between two grayscale images for experimental validation. The utility of this approach can be seen in domains where intensity differences carry critical information and therefore the bipolar approach is potentially more effective in smoothening out minor variations thereby making similarity computation more robust against unwanted variations in pixel intensity.

This step-by-step approach is presented as a means of validating the effectiveness of the proposed Bipolar Fuzzy Jaccard Similarity, demonstrating its ability to enhance similarity computations.

Suppose we have to 4×4 grayscale images as below

Image 1

$$A = \begin{matrix} 50 & 80 & 120 & 200 \\ 30 & 90 & 130 & 210 \\ 60 & 100 & 140 & 220 \\ 70 & 110 & 150 & 230 \end{matrix}$$

Image 2

$$B = \begin{matrix} 55 & 85 & 125 & 205 \\ 35 & 95 & 135 & 215 \\ 65 & 105 & 145 & 225 \\ 75 & 115 & 155 & 235 \end{matrix}$$

Step 1: Compute the minimum and maximum pixel intensities

Find the minimum and maximum pixel values across both images in order to normalize the images

$$\min(A, B) = \min(30, 35) = 30$$

$$\max(A, B) = \max(230, 235) = 235$$

Step 2: Compute positive membership function μ^+ using min-max normalization

For every pixel p_i , we can then by min-max normalization

$$\mu^+(p_i) = \frac{I(p_i) - \min(I)}{\max(I) - \min(I)}$$

$$\mu^+(p_i) = \frac{A(p_i) - 30}{235 - 30}$$

In A we have

$$\mu_A^+ = \begin{bmatrix} \frac{50-30}{205} & \frac{80-30}{205} & \frac{120-30}{205} & \frac{200-30}{205} \\ \frac{30-30}{205} & \frac{90-30}{205} & \frac{130-30}{205} & \frac{210-30}{205} \\ \frac{60-30}{205} & \frac{100-30}{205} & \frac{140-30}{205} & \frac{220-30}{205} \\ \frac{70-30}{205} & \frac{110-30}{205} & \frac{150-30}{205} & \frac{230-30}{205} \end{bmatrix}$$

$$\mu_A^+ = \begin{bmatrix} 0.0976 & 0.3439 & 0.4390 & 0.8293 \\ 0 & 0.2978 & 0.4878 & 0.8780 \\ 0.1463 & 0.3415 & 0.5366 & 0.9268 \\ 0.1951 & 0.3902 & 0.5854 & 0.9756 \end{bmatrix}$$

Similarly for B

$$\mu_B^+ = \begin{bmatrix} 0.1220 & 0.2683 & 0.4634 & 0.8537 \\ 0.0244 & 0.3171 & 0.5122 & 0.9024 \\ 0.1707 & 0.3659 & 0.5610 & 0.9512 \\ 0.2195 & 0.4146 & 0.6098 & 1 \end{bmatrix}$$

Step 3: Compute negative membership function μ^-

The negative membership function $\mu^-(p_i)$ is defined by $\mu^-(p_i) = -(1 - \mu^+(p_i)) \in [-1, 0]$.

In A we have

$$\mu_A^- = \begin{bmatrix} -0.9024 & -0.7561 & -0.5610 & -0.1707 \\ -1 & -0.7073 & -0.5122 & -0.1220 \\ -0.8537 & -0.6585 & -0.4634 & -0.0732 \\ -0.8049 & -0.6098 & -0.4146 & -0.0244 \end{bmatrix}$$

For B

$$\mu_B^- = \begin{bmatrix} -0.8780 & -0.7317 & -0.5366 & -0.1463 \\ -0.9756 & -0.6829 & -0.4878 & -0.0976 \\ -0.8293 & -0.6341 & -0.4390 & -0.0488 \\ -0.7805 & -0.5854 & -0.3902 & 0 \end{bmatrix}$$

Step 4: Compute bipolar Jaccard similarity

The bipolar Jaccard similarity is given by

$$s(A, B) = \frac{\sum \min(A^+(x_i), B^+(x_i)) - \sum \min(A^-(x_i), B^-(x_i))}{\sum \max(A^+(x_i), B^+(x_i)) + \sum \max(A^-(x_i), B^-(x_i))}$$

It can be verified that

$$s(A, B) = \frac{-0.5584}{16.9884} = -0.077$$

The Bipolar Jaccard Similarity value of -0.077 suggests that the images have slightly more dissimilarities than similarities. This means that while some pixels

have similar intensity values, there are still noticeable variations. The negative membership values indicate a higher degree of dissimilarity than similarity (if the similarity score were closer to 1, it would suggest that the images are nearly identical, whereas a score approaching -1 would indicate significant differences between them).

3.1.3. Bipolar Fuzzy Jaccard Similarity Using Sigmoid-Based Membership

Instead of using linear min-max normalization, we define sigmoid-based fuzzy membership functions by

$$\mu^+(p) = \frac{1}{1 + e^{-k(I(p)-T)}}$$

$$\mu^-(p) = -(1 - \mu^+(p))$$

Where:

- $I(p)$ is the pixel intensity.
- T is the threshold intensity.
- k controls the steepness of the function (chosen as 0.1 for this example).

Suppose we have to 4×4 grayscale images as below

Image 1

$$A = \begin{matrix} 50 & 80 & 120 & 200 \\ 30 & 90 & 130 & 210 \\ 60 & 100 & 140 & 220 \\ 70 & 110 & 150 & 230 \end{matrix}$$

Image 2

$$B = \begin{matrix} 55 & 85 & 125 & 205 \\ 35 & 95 & 135 & 215 \\ 65 & 105 & 145 & 225 \\ 75 & 115 & 155 & 235 \end{matrix}$$

Then computing this in Python using the Bipolar fuzzy Jaccard formula,

$$s(A, B) = \frac{\sum \min(A^+(x_i), B^+(x_i)) - \sum \min(A^-(x_i), B^-(x_i))}{\sum \max(A^+(x_i), B^+(x_i)) + \sum \max(A^-(x_i), B^-(x_i))}$$

we obtain a Bipolar Fuzzy Jaccard Similarity -0.1120 .

See Appendix xx for the Full Python Implementation of the Bipolar Fuzzy Jaccard Similarity Algorithm.

The BJS with sigmoid transformation is lower (-0.112) than with the min-max transformation (-0.077) implying the two functions assigns bipolar fuzzy values differently. This difference could be as a result of:

- 1) Min-max transformation preserves relative pixel differences, making it more sensitive to slight variations.
- 2) Sigmoid-based transformation compresses extreme values, emphasizing mid-range intensities, which affects how differences in brightness levels contribute to similarity.

3) The penalty for mismatches (negative membership function) behaves differently in sigmoid-based transformation, leading to a more negative similarity score in some cases.

Key differences between the two approaches

Normalization method	Effect on Membership Values	Effect on Similarity Score
Min-Max Normalization (linear)	Preserves relative intensity differences	More sensitive to small pixel changes, leading to lower similarity if pixel values differ
Sigmoid-Based Membership (non-linear)	Sigmoid-Based Membership (non-linear)	Less sensitive to small differences but may yield negative similarity scores due to contrast compression

3.2. Extending Rényi Entropy to Bipolar Fuzzy Sets

To extend Rényi entropy to bipolar fuzzy sets, we define a new entropy function that considers both the positive and negative membership degrees of elements in a BFS.

3.2.1. Bipolar Fuzzy Probability Distribution

Given a bipolar fuzzy set A , the bipolar fuzzy probability distribution may be defined as

$$p_i^+ = \frac{\mu^+(x_i)}{\sum_{j=1}^n \mu^+(x_j)}$$

where p_i^+ represents the positive probability contribution.

$$p_i^- = \frac{|\mu^-(x_i)|}{\sum_{j=1}^n |\mu^-(x_j)|}$$

where p_i^- represents the negative probability contribution.

Using these probability distributions, we can define the bipolar Rényi entropy (BRE) by

$$H_\alpha^B(A) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n (p_i^+)^{\alpha} + \sum_{i=1}^n (p_i^-)^{\alpha} \right)$$

This formulation ensures that the definition encapsulates both the supporting and opposing membership degrees.

3.2.2. Properties of the Proposed BRE

1) If $\mu^-(x_i) \forall i$ then $H_\alpha^B(A)$ reduces to the classical Rényi entropy. The proposed BRE for bipolar fuzzy sets generalizes the standard Rényi entropy.

2) The parameter α enables dynamic control over the entropy computation, with higher α emphasizing dominant elements.

To rigorously examine the proposed BRE, we analyze its mathematical proper-

ties, ensuring its validity as a generalization of classical Rényi entropy.

Non-negativity

A fundamental property of entropy measures is that it must be non-negative *i.e.*, $H_\alpha^B(A) \geq 0$.

Proof

$\forall i, 0 \leq p_i^+ \leq 1$ and $0 \leq p_i^- \leq 1$. Since $\alpha > 0$ then $0 \leq (p_i^+)^{\alpha} \leq 1$ and $0 \leq (p_i^-)^{\alpha} \leq 1 \forall i$.

$$\text{Therefore } H_\alpha^B(A) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n (p_i^+)^{\alpha} + \sum_{i=1}^n (p_i^-)^{\alpha} \right) \geq 0.$$

Monotonicity

BRE, like classical entropy measures, is expected to be monotonically decreasing in α meaning that as the probability contribution increases with increase in α , the entropy associated with it decreases, signifying a reduction in uncertainty.

Proof

Consider the function

$$S_\alpha = \sum_{i=1}^n (p_i^+)^{\alpha} + \sum_{i=1}^n (p_i^-)^{\alpha}$$

Taking the derivative with respect to α .

$$\frac{d}{d\alpha} S_\alpha = \sum_{i=1}^n (p_i^+)^{\alpha} \ln(p_i^+) + \sum_{i=1}^n (p_i^-)^{\alpha} \ln(p_i^-)$$

Since $0 \leq p_i^+ \leq 1$ and $0 \leq p_i^- \leq 1 \forall i$ then their logarithms will be negative

$$\frac{d}{d\alpha} S_\alpha < 0$$

which implies S_α is monotonically decreasing in α . Therefore,

$$H_\alpha^B(A) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n (p_i^+)^{\alpha} + \sum_{i=1}^n (p_i^-)^{\alpha} \right) \text{ is monotonically decreasing}$$

Regarding the monotonicity of entropy, it is important to note that maximum entropy occurs when the probability distribution is uniform, and the entropy decreases as the probability distribution becomes more skewed towards a few specific events. Therefore, intuitively, with more information about a system, the uncertainty about its state decreases, leading to a decrease in entropy.

Concavity

Entropy is generally expected to be concave. To establish that BRE is concave, we need to compute its second derivative and show that it is non-positive *i.e.*,

$$\frac{d^2}{d\alpha^2} (H_\alpha^B(A)) \leq 0.$$

Proof

Differentiating $S_\alpha = \sum_{i=1}^n (p_i^+)^{\alpha} + \sum_{i=1}^n (p_i^-)^{\alpha}$ a second time

$$\frac{d}{d\alpha} \left(\frac{d}{d\alpha} S_\alpha \right) = \frac{d}{d\alpha} \left(\sum_{i=1}^n (p_i^+)^{\alpha} \ln(p_i^+) + \sum_{i=1}^n (p_i^-)^{\alpha} \ln(p_i^-) \right)$$

$$\begin{aligned}
&= \sum_{i=1}^n (p_i^+)^{\alpha} (\ln p_i^+)^2 + \sum_{i=1}^n (p_i^-)^{\alpha} (\ln p_i^-)^2 - \frac{\sum_{i=1}^n (p_i^+)^{\alpha} (\ln p_i^+) + \sum_{i=1}^n (p_i^-)^{\alpha} (\ln p_i^-)^2}{\sum_{i=1}^n (p_i^+)^{\alpha} + \sum_{i=1}^n (p_i^-)^{\alpha}} \\
&\leq 0, \text{ since } 0 \leq p_i^+ \leq 1 \text{ and } 0 \leq p_i^- \leq 1 \quad \forall i
\end{aligned}$$

4. Conclusion

A key contribution of this work is the formulation of the Bipolar Fuzzy Jaccard Similarity, which refines similarity assessments by considering both shared and opposing characteristics in segmented regions. Additionally, the paper extends Rényi entropy to the bipolar fuzzy domain, enhancing the quantification of uncertainty in image segmentation tasks. The integration of bipolar fuzzy similarity and entropy measures provides a robust and interpretable computational framework for image segmentation, particularly in domains requiring comprehensive image analysis.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Popper, K.R. (1959) *The Logic of Scientific Discovery*. Hutchinson.
- [2] Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [3] Zhang, B. (1998) Bipolar Fuzzy Sets and Relations: A Computational Framework for Cognitive Modeling and Decision Analysis. *Proceedings of the IEEE International Conference on Fuzzy Systems*, Anchorage, 4-9 May 1998, 835-840.
- [4] Zhang, B. (2016) Bipolar Fuzzy Clustering and Its Applications. *Information Sciences*, **366**, 23-39.
- [5] Kim, H., Park, J. and Rhee, P.-K. (2018) A Study on Bipolar Fuzzy Metric Spaces and Applications. *Fuzzy Sets and Systems*, **347**, 95-115.
- [6] Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338-353. [https://doi.org/10.1016/s0019-9958\(65\)90241-x](https://doi.org/10.1016/s0019-9958(65)90241-x)
- [7] Dubois, D. and Prade, H. (2012) Bridging Gaps between Fuzzy Sets and Belief Functions. *International Journal of Approximate Reasoning*, **52**, 311-330.
- [8] Zimmermann, H.J. (2001) *Fuzzy Set Theory and Its Applications*. Springer.
- [9] Tversky, A. (1977) Features of Similarity. *Psychological Review*, **84**, 327-352. <https://doi.org/10.1037/0033-295x.84.4.327>
- [10] Dubois, D. and Prade, H. (1980) *Fuzzy Sets and Systems: Theory and Applications*. Academic Press.
- [11] Rényi, A. (1961) On measures of Entropy and Information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 547-561.
- [12] van Erven, T. and Harremoës, P. (2014) Rényi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory*, **60**, 3797-3820. <https://doi.org/10.1109/tit.2014.2320500>
- [13] Müller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S. and Tomamichel, M. (2013) On

- Quantum Rényi Entropies: A New Generalization and Some Properties. *Journal of Mathematical Physics*, **54**, Article 122203. <https://doi.org/10.1063/1.4838856>
- [14] Yang, X., Zhang, D. and Yu, Z. (2011) A Rényi Entropy-Based Thresholding Method for Image Segmentation. *Pattern Recognition Letters*, **32**, 2109-2118.
- [15] De Luca, A. and Termini, S. (1972) A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory. *Information and Control*, **20**, 301-312. [https://doi.org/10.1016/s0019-9958\(72\)90199-4](https://doi.org/10.1016/s0019-9958(72)90199-4)
- [16] Kaufmann, A. (1975) Introduction to the Theory of Fuzzy Subsets. Academic Press.
- [17] Yager, R.R. (1979) On the Measure of Fuzziness and Negation. *Information Sciences*, **19**, 39-61.
- [18] Yager, R.R. (1982) A New Approach to the Summarization of Data. *Information Sciences*, **28**, 69-86. [https://doi.org/10.1016/0020-0255\(82\)90033-0](https://doi.org/10.1016/0020-0255(82)90033-0)
- [19] Kong, T.Y. and Rosenfeld, A.A. (1991) Digital Topology: A Comparison of the Graph-Based and Topological Approaches. In: Brookes, S.D., Scedrov, A. and Main, M., Eds., *Topology and Category Theory in Computer Science*, Oxford University Press, 273-290.
- [20] Rosenfeld, A. (1979) Fuzzy Digital Topology. *Information and Control*, **40**, 76-87. [https://doi.org/10.1016/s0019-9958\(79\)90353-x](https://doi.org/10.1016/s0019-9958(79)90353-x)
- [21] Bezdek, J.C. (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Springer.
- [22] Dunn, J.C. (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, **3**, 32-57. <https://doi.org/10.1080/01969727308546046>
- [23] Van der Weken, D., Scheunders, P. and De Witte, S. (2004) A Fuzzy Histogram-Based Approach to Texture Similarity. *Pattern Recognition*, **37**, 1923-1933.
- [24] Patro, S.G.K. and sahu, K.K. (2015) Normalization: A Preprocessing Stage. *International Journal of Computer Applications*, **116**, 7-10. <https://doi.org/10.17148/iarjset.2015.2305>
- [25] Zhang, X., Zhang, L. and Zhang, X. (2021) Adaptive Fuzzy Normalization for Robust Image Processing. *IEEE Transactions on Image Processing*, **30**, 4175-4189.
- [26] Tang, J. (1996) Contrast Enhancement Using Clustered Patterns. *Pattern Recognition*, **29**, 23-32.
- [27] Pal, N.R., Pal, S.K. and Keller, J.M. (2002) A Possibility Measure Framework for Fuzzy Classification. *IEEE Transactions on Fuzzy Systems*, **10**, 578-587.
- [28] Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J. (2001) Color Image Segmentation: Advances and Prospects. *Pattern Recognition*, **34**, 2259-2281. [https://doi.org/10.1016/s0031-3203\(00\)00149-7](https://doi.org/10.1016/s0031-3203(00)00149-7)