*Research Article*

# An Image Segmentation Method Based on the Law of Universal Gravitation and Adaptive Affinity Propagation Spectrum Clustering

**Hua Lu[1] and Jianrong Zhu [ID] [2]**

[1]*Changsha Vocational & Technical College, Changsha 410127, Hunan, China*
[2]*Hunan University of Technology and Business, Changsha 410205, Hunan, China*

Correspondence should be addressed to Jianrong Zhu; 2375@hutb.edu.cn

There are many problems of division in natural and social sciences, and with the development of science and technology, the requirement for division is also increasing. It is difficult to divide accurately by experience and expertise alone, and the most important research branch of the division problem is the clustering algorithm. It is to group similar samples into one class and divide the elements with large differences into different classes. Due to the simplicity and efficiency of the clustering algorithm, it is widely used in image segmentation. The conventional Spectral Clustering (SC) algorithm cannot recognize nonconvex data and has the disadvantage of strong dependence on biased parameter values. To address this problem, the Gravity-based Adaptive Spectral Clustering (GASC) algorithm is proposed in this study. Based on the conventional SC algorithm, the algorithm uses gravity to calculate the similarity (gravity) between data, and uses information entropy and adaptive enhancement (AdaBoost) algorithms to obtain the weights of correct cluster sampling points and wrong cluster sampling points in each cluster, so as to reduce the dependence of the algorithm on bias parameters and reduce the number of wrongly divided sample points. Meanwhile, the GASC algorithm is applied to image segmentation. The implementation process includes three stages: image preprocessing, feature extraction, and clustering. The comparison experiments show that the mean values of normalization and accuracy of the GASC algorithm are improved compared with other clustering algorithms, and the segmentation accuracy for images is higher.

## 1. Introduction

Image segmentation is an image processing technique that divides the highly relevant regions of an image into one class and the parts of the image with widely varying attributes into different regions. This technique needs to make the contours of the target in the image continuous and the regions as complete as possible while preserving the important features of the image [1]. Clustering has an important connection with image segmentation, which in some cases can be seen as a clustering or classification problem of pixel points, and the goals of both problems solving are the same to some extent, therefore, clustering methods have an important and wide application in image segmentation. Clustering-based image segmentation can be regarded as a kind of unsupervised classification, in which pixels are automatically classified according to their intrinsic structure and features without training samples, so that pixels in the same category have similar or similar features, while the features of pixels in different categories differ greatly [2, 3]. The clustering methods commonly used for image segmentation include K-means, hierarchical clustering, spectral clustering, fuzzy C-mean clustering FCM, density clustering, fuzzy clustering, etc.

With the complexity of application scenarios, higher clustering algorithms are put forward to the demand, but the current clustering algorithms still have some problems, such as the clustering results are more influenced by the initial points, the parameters in the clustering algorithm have a great influence on the clustering results, how to handle large-

scale data sets, whether the class clusters of arbitrary shapes can be identified, and how to shorten the running time of the algorithm, which still need to be paid attention [4]. In this paper, we propose a gravity-based adaptive nearest neighbor propagation clustering algorithm (GASC). This algorithm uses the gravitational search mechanism to globally optimize the conventional SC algorithm and proposes an effective solution to the problem that the SC algorithm is not applicable to non-convex data, and on this basis, the information entropy and AdaBoost algorithms are used to adaptively adjust the parameter preference as well as reduce the missplit points to improve the clustering effect of the SC algorithm.

The main contributions of this paper can be summarized as follows: to address the shortcomings of the SC clustering algorithm, a gravity-based adaptive nearest neighbor propagation clustering algorithm (GASC) is proposed, which uses the gravity search mechanism to calculate the similarity matrix to solve the problem that the clustering algorithm is not applicable to non-convex data. Based on this, the parameter values are dynamically adjusted using information entropy and AdaBoost algorithm to improve the clustering accuracy. The GASC algorithm is applied to image segmentation, and the improved algorithm is used to process the eigenvectors of all pixels in the image, and the eigenvalue matrix of the image to be segmented is clustered to obtain the final image segmentation results. The experimental results show that the GASC algorithm improves the segmentation accuracy of images compared with the conventional clustering algorithm.

## 2. Related Work

Clustering analysis has been at the core of data mining, and it has been applied to many fields, such as pattern recognition, image processing, machine learning, and so on. Some of the more popular and commonly used clustering algorithms are the K-means clustering algorithm, FCM clustering algorithm, density clustering algorithm, spectral clustering algorithm, and density peak clustering algorithm. Image segmentation can be viewed as a pixel classification process based on feature space [5]. Both image segmentation and clustering analysis problems are unsupervised classification methods, and both have received extensive research and attention from researchers in their respective fields. The basic idea of the clustering problem is to assign classes to data points in the target dataset based on the similarity between feature attributes to obtain the optimal division or minimize the objective function to obtain the suboptimal division. It can be seen that the clustering segmentation method for images is to introduce this idea of unsupervised learning into the field of image segmentation [6]. In addition, the clustering segmentation algorithm can also be combined with other methods to accomplish the image segmentation task. When the clustering algorithm is applied to image segmentation, i.e., specific features can be applied individually or in combination with some other image segmentation methods. Classification can be divided into two categories: supervised and unsupervised. In general,

supervised classification is called classification, while unsupervised classification is called clustering [7]. Classification is the most basic statistical analysis method in the field of pattern recognition, and its purpose is to classify images by finding points, curves, and surfaces in the image feature space using labeled training samples. The advantage of classifier is that it is faster because it does not require iterative operations and it can segment multispectral images. But the disadvantage is the need to obtain training samples beforehand, which is difficult in some cases. The clustering algorithm is more similar to the classifier, but does not require training samples. Clustering-based image segmentation methods are a hot research topic in the field of image segmentation. The conventional clustering algorithms are the K-means algorithm and FCM algorithm, which generalize the K-means algorithm from the perspective of fuzzy set theory [8]. Their advantages are simple algorithms and fast execution, and the disadvantages are that they cannot guarantee global optimality, require manual input of the number of clustering categories, and are sensitive to noise and outliers. Compared with other segmentation methods, unsupervisedness, efficiency and adaptiveness are the three main features unique to image segmentation methods based on cluster analysis. The unsupervised learning strategy has its unique advantages in the image segmentation task. First, it does not need to specify a priori information such as the category of segmentation, but is self-organized according to the aggregated links between pixels, which makes the autonomy of the image segmentation method more complete [9]. Usually, clustering algorithms are very demanding in terms of their computational complexity, which makes their search strategies better and allows them to obtain more desirable segmentation results in a shorter time. The clustering algorithm does not make assumptions about the data and has simple parameter settings, which also makes it more scalable and can be easily embedded in many fields. For some image segmentation methods, the processing of specific features of the image may achieve better results in some images, but when the source image is replaced or the image changes significantly, the processing results may vary greatly, which also reflects the strong adaptive nature of the image segmentation method of ensemble clustering [10].

Koundal [11] presented an effective segmentation method that was based on neutrosophic clustering with the integration of features for images. This method can handle the indeterminacy of pixels to have strong clusters and to perform segmentation effectively with the noisy images. Experiments were performed with various types of natural and medical images to exhibit the performance of the proposed segmentation method. Zhang et al. [12] proposed an improved approach for color image segmentation, clustering-based JSEG (T-JSEG). The T-JSEG combined the texture information with the classical JSEG algorithm. They used the Gabor feature clustering to produce a texture map and combined it with color map to achieve a new map for $J$ value computing. Mathew and Simon [13] proposed an automatic approach for image segmentation based on neutrosophic set and non-subsampled contourlet transformation for natural images. This method used both color and

features for segmentation. Kim et al. [14] evaluated the proposed method on dataset consisting of texture image and limit possible number of clusters from 2 to 5. And they also evaluated the proposed method by real image containing various textures such as rock stratum. Li et al. [15] presented a novel simultaneous cartoon-texture image segmentation and image decomposition method to boost the performance of both segmentation and decomposition. Rahman and Horiguchi [16] proposed a new integrated feature distribution-based color image segmentation algorithm.

Two novel histogram-based inherent color feature extraction methods had been presented. From the histogram features, mean color texture histogram was calculated. Instead of concatenating the feature channels, a multichannel nonparametric Bayesian clustering was employed for primary segmentation. Tirandaz and Akbarizadeh [17] proposed a novel unsupervised segmentation algorithm based on Gabor filter bank and unsupervised spectral regression (USR), for SAR image segmentation. In the proposed algorithm, they used a Gabor filter bank to decompose the image into several sub-images. Features were extracted from these sub-images and further, learned, using USR. Finally, K-means clustering was employed and the image was segmented. Liu et al. [18] proposed a novel color image segmentation method based on local histograms. Starting with clustering-based color quantization, they extracted a sufficient number of representative colors. For each pixel, by counting the number of pixels with each representative color within a circular neighborhood, a local histogram was obtained. Salmi et al. [19] proposed a new semi-supervised method by combining constrained feature selection and spectral clustering (SC) to perform color-texture image segmentation. Baya et al. [20] adopted a clustering validation method, Clustering Stability (CS), to automatically segment images. CS is not limited by image dimensionality nor by the clustering algorithm. They showed clustering and validation acting together as a data-driven process able to find the optimum number of partitions according to our proposed color feature representation. Du et al. [21] proposed a multi-feature fusion method to process the feature extraction. The proposed method combined the gray level co-occurrence matrix (GLCM), Gabor wavelet transform and local binary pattern (LBP).

It had the advantages of the above three feature extraction methods. Tianet al. [22] proposed an image segmentation method based on optimized spatial information, in which the Gaussian kernel was adopted to diminish the local incorrect segmentation. The FCM clustering was spatially adjusted and optimized by the particle swarm optimization. The purpose of optimization was to obtain the appropriate control parameters influencing spatial information, which can improve segmentation results. Heshmatiet al. [23] presented an efficient scheme for unsupervised color image segmentation using neutrosophic set (NS) and non-subsampled contourlet transform (NSCT), to achieve a better segmentation result, an appropriate indeterminacy reduction operation was proposed. Catalbas [24] proposed an adaptive and robust unsupervised segmentation algorithm. One of the novelties of this proposed

algorithm was to determine optimal sub-image size by pattern analysis and another one was optimizing segmentation process by providing the most successful representation of patterns on images. Shang et al. [25] proposed a new semantic segmentation method of SAR images based on texture complexity analysis and key superpixels. Complexity analysis was performed and on this basis, mixed superpixels were selected as key superpixels.

However, there are still some problems in clustering algorithms that have not been fully solved, which affect the application of image segmentation methods based on cluster analysis in practice. For hard partition clustering algorithms, low computational complexity and high clustering accuracy are often in conflict with each other. This is particularly evident in the process of minimizing the objective function, where clustering algorithms usually improve the accuracy of clustering results at the expense of the algorithm's running time in order to avoid getting trapped in local minima. In some cases, however, the real-time nature of image segmentation limits the scope of the application of clustering algorithms. Therefore, in order to reconcile the contradiction between computational complexity and clustering accuracy, researchers focus on the search strategy of hard partitioning clustering algorithm, and then make the image segmentation method based on clustering analysis not only obtain the correct segmentation results but also have high computational efficiency in practical image segmentation applications. For the application of the soft division clustering algorithm in image segmentation, a major constraint is the ability of the algorithm to suppress the noise interference in the image. In the practical application of image segmentation, different levels of noise caused by various factors may be present in the real image causing interference. The content information of these images is already complex and highly uncertain, and the presence of noise affects the overall effect of image segmentation and the extraction of regions of interest in the images. Therefore, the soft segmentation clustering algorithm can be improved by introducing and enhancing the spatial information from the image pixels, which can improve the noise suppression ability of the algorithm in image segmentation. Under the constraint of spatial information, the image segmentation method based on clustering analysis can produce more noise-insensitive segmentation results and obtain more complete and smooth regions of interest.

## 3. SC Algorithm Theory

*3.1. SC Algorithm Overview.* Spectral clustering algorithm, which originated from spectral graph theory, is a clustering method based on graph theory. The main idea of the spectral clustering algorithm is to divide the weighted undirected graph into two or more optimal subgraphs. That is, the sample points are taken as the vertices of the graph, and the similarity between the sample points is the weight of the edges between the two points. After the spectral division, the internal similarity of the two subgraphs is the largest, and the similarity between the subgraphs is the smallest, so as to finally achieve the purpose of clustering. The optimal

clustering criterion is similar to the optimal partition criterion of graph theory. Compared with the traditional K-Means algorithm, spectral clustering algorithm has stronger adaptability to data distribution, excellent clustering effect, and much less computation.

$G(V, E)$ is used to describe a graph, $V$ represents the set of points and $E$ represents the set of edges. For any two points in $V$, there can be edge connection or no edge connection. Define the weight $W_{ij}$ as the weight between points $V_i$ and $V_j$, since discussing undirected graphs, so $W_{ij} = W_{ji}$.

### 3.2. Similarity Measurement.
The similarity measure is used to measure the degree of similarity between data. In spectral clustering, the weight of an undirected graph is expressed by the similarity between data. There are many kinds of similarity measurement methods, which generally need to be selected according to the actual situation. Common measurement methods include Euclidean distance, Gaussian kernel function, cosine similarity, etc.

### 3.2.1. European Distance.
Euclidean distance is the most commonly used measurement method, which is defined as follows:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{d} (x_{il} - x_{jl})^2}, \tag{1}$$

where $x_i$ and $x_j$ represents a d-dimensional data point vector. In the feature space, Euclidean distance has the invariance of transformation and rotation, so it is more inclined to construct spherical clustering clusters.

### 3.2.2. Cosine Similarity.
Cosine similarity is often used in text clustering, and its expression is given as follows:

$$s(x_i, x_j) = \cos \alpha = \frac{x_t^T x_j}{\|x_i\| \cdot |x_i|}. \tag{2}$$

In text clustering, the traditional Euclidean distance often cannot well describe the similarity between text objects, so it is necessary to measure the similarity without distance measurement.

### 3.2.3. Gaussian Kernel Function.
The Gaussian kernel function is often used to define the weight of edges in spectral clustering, and its calculation formula is given as follows:

$$S(x_i, x_j) = \exp\left(-\frac{d^2(x_i, x_j)}{2\sigma^2}\right), \tag{3}$$

where $x_i$ and $x_j$ represents a d-dimensional data point vector, and $D(x_i, x_j)$ is the Euclidean distance, $\sigma$ is the scale parameter of the function. The higher the similarity is, the

larger the parameter is. Therefore, in practical application, $\sigma$ it needs to be determined by multiple value experiments.

### 3.3. Similarity Matrix and Degree Matrix.
In practical applications, it is the most common to use the above Gaussian kernel function similarity measure to establish the adjacency matrix. Because the weight value between all points is greater than 0, the similarity matrix is defined as follows:

$$W_{ij} = \exp\left(-\frac{d^2(x_i, x_j)}{2\sigma^2}\right). \tag{4}$$

The degree matrix is a diagonal matrix marked as $D$, and the degree value is a diagonal element is given as follows:

$$D_{ij} = \sum_{j} W_{ij}. \tag{5}$$

### 3.4. Laplace Matrix.
The Laplace matrix can be calculated by obtaining the similarity matrix $W$ and the degree matrix $D$ through the above calculation, we can get:

$$L_{ij} = D_{ij} - W_{ij}. \tag{6}$$

There are two forms of standardized Laplace matrix, $I$ is the identity matrix:

$$\begin{aligned} L_{\text{nor}} &= D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}, \\ L_{\text{nor}} &= D^{-1}L = I - D^{-1}W. \end{aligned} \tag{7}$$

It has some good properties:

(1) Laplace matrix is a symmetric matrix, and all its eigenvalues are real numbers.

(2) For any vector $f$, we can deduce:

$$f^T L f = f^T D f - f^T W f = \frac{1}{2} \sum_{i,j=1}^{n} W_{ij}(f_i - f_j)^2. \tag{8}$$

(3) Laplacian matrix is semi-positive definite, and the corresponding $n$ real eigenvalues are greater than or equal to zero.

### 3.5. Partition Criteria of Weighted Undirected Graphs.
In general, the optimization criterion for solving spectral division is NP-hard problem, and the effective way to solve this kind of problem is to use the solution of alternative methods to approximately replace NP-hard problem. The main solution method considered is the continuous relaxation form. The optimal division criterion is to divide the two subgraphs into the largest internal similarity and the smallest similarity between subgraphs. The following are common graph segmentation criteria.

### 3.5.1. Minimum Cut Criterion (Minimum Cut).
As the first proposed general division criterion, the minimum cut set is defined as: for two disjointed subsets $A, B \subset V$, the cut is given as follows:
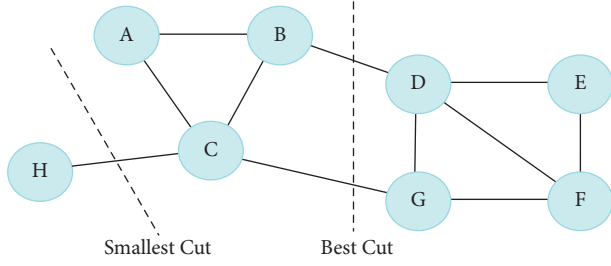
FIGURE 1: Minimum cut set criterion problem of graph.

$$\text{cut}(A, B) = \sum_{i \in A j \in B} W_{ij}. \tag{9}$$

If the graph is cut into $k$ connected subsets, the segmentation can be minimized:

$$\text{cut}(A_1, A_2, \ldots, A_k) = \sum_{i=1}^{k} \text{cut}(A_i, \overline{A_i}), \tag{10}$$

where $\overline{A_i}$ is the complement of $A_i$. When selecting the edge with the smallest weight to cut, it is easy to produce a smaller graph containing only a few vertices. Even if the cut value is minimized, it is not the desired optimal cut graph (Figure 1).

*3.5.2. Ratio Cut Criteria (Ratio-Cut).* This criterion minimizes the similarity between classes by the balance term of the number of class sizes. It is defined as follows:

$$\text{Rcut}(A_1, A_2, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{\text{cut}(A_i, \overline{A_i})}{|A_i|}, \tag{11}$$

where $|A_i|$ represents the number of points in $A_i$. The proportional cut set criterion increases the similarity between classes to reduce the over-segmentation phenomenon, but this criterion will reduce the running speed and efficiency.

*3.5.3. Standardized Cut-Set Criterion (Normalized-Cut).* The above methods do not consider the weight coefficient inside the subgraph, which is defined as follows:

$$\text{Ncut}(A_1, A_2, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{\text{cut}(A_i, \overline{A_i})}{\text{vol}(A_i)}, \tag{12}$$

where $\text{vol}(A_i)$ represents the sum of weights in $A_i$:

$$\text{vol}(A_i) = \sum_{i \in A_i} d_i. \tag{13}$$

In particular, the nonstandard Laplace matrix corresponds to the Rcut cut graph. The Rcut cut graph describes the similarity in the cluster as the number of samples $|A|$ contained in the cluster. The standardized Laplace matrix corresponds to the Ncut cut graph. The Ncut cut graph describes the similarity in the cluster as $\text{vol}(A_i)$ Since $\text{vol}(A_i)$ can better reflect the similarity within the cluster than $|A|$, the standardized Laplace matrix is selected.

*3.6. SC Algorithm Steps.* Generally speaking, the differences in spectral clustering algorithms mainly lie in the generation of the similarity matrix, the cut graph and the final clustering method. The most commonly used way to generate a similarity matrix is the Gaussian kernel function, the most commonly used cut graph method is Ncut, and the last commonly used clustering method is K-Means. The summary steps are as follows:

Step 1: Construct sample similarity matrix Wand degree matrix $D$.

Step 2: Calculate the standardized Laplace matrix $L_{\text{nor}}$.

Step 3: Calculate the eigenvector $f$ corresponding to the $k$ minimum eigenvalues of the $L_{\text{nor}}$ matrix, and form the eigenmatrix $F$ by rows.

Step 4: Row normalization matrix $F$:

$$u_{ij} = \frac{f_{ij}}{\left(\sum_k f_{ik}^2\right)^{1/2}}. \tag{14}$$

Step 5: Each row represents a sample, and K-Means clustering algorithm is applied to $n$ samples to obtain clustering result $C$.

The first problem that may be encountered in the implementation of the SC algorithm is the selection of the number of cluster classes. The common way is to use heuristic eigenvalue difference search: if the first $k$ eigenvalues are small, and the difference between the $k+1$ eigenvalue and the previous eigenvalue is relatively large, the number of cluster classes is $k$. Select the first $k$ smallest eigenvalues as the number of cluster classes, assuming that graph $G$ can be divided into $k$ connected subsets with no intersection at all, then there are $k$ eigenvalues equal to 0, and the $k+1$ eigenvalue is greater than 0. Therefore, it can be assumed that the smaller the eigenvalue, the better the clustering performance. Select the number of small eigenvalues as the number of clusters.

Spectral clustering is more widely used in small and medium-sized data sets than K-means. However, due to the changing application scenarios, the spectral clustering algorithm also faces the following problems: in the clustering process of the spectral clustering algorithm, it is necessary to solve the eigenvalues and eigenvectors of the matrix, and the complexity of solving the eigenvectors of the non-sparse matrix. Therefore, when dealing with large-scale data sets, the matrix space formed in the calculation is very large, and the solving process will not only be very time-consuming, but also require a very large memory space, facing the danger of memory overflow. Therefore, how to improve the running speed of the SC algorithm, reducing the memory space required for operation, and reducing the time and space cost of algorithm operation are the key problems faced by spectral clustering algorithm in the process of expanding their application fields. If the dimension reduction of the final clustering is not enough, the running speed and clustering effect of the spectral clustering algorithm are not good. When processing large-scale data sets, clustering may be interrupted due to the large amount of data.

## 4. GASC Algorithm Design

*4.1. Similarity of Gravity Measurement.* In order to solve the problem that the conventional SC algorithm is not applicable to non-convex data, this paper uses the gravitational search mechanism to improve the conventional SC algorithm. The formula of the law of gravity is shown in equation (15).

$$F = \frac{G(M_1 \times M_2)}{r^2}, \tag{15}$$

$M_1, M_2$ is the mass of the object, G is $6.67*10^{-11}$ N·m$^2$/kg$^2$, and r is the distance between the objects.

Gravity reflects the attraction between objects by their masses and distances. The greater the mass, the closer the objects are, the stronger the attraction. This is similar to the similarity between objects in cluster analysis.

The gravitational similarity measure is a global search of samples using a gravitational search mechanism to measure the similarity between objects. The specific principle is that the search is optimized by the interaction of mutual gravitational forces between samples, each of which is influenced by the gravitational forces of other samples in the space, resulting in accelerations close to individuals with similar characteristics. Samples with different characteristics are attracted to each other in space by the universal gravitational force, which is determined by the characteristics of the samples and the distance between the samples. Under the effect of gravity, samples approach other samples with similar characteristics and distances, i.e., they gradually approach the optimal solution of the optimization problem. The specific form is as follows.

Suppose given $X = \{x_1, x_2, \ldots, x_m\} \in R^{m \times n}$, m is the number of samples and n is the number of dimensions. The formula of the similarity matrix based on gravity is shown in equation (16).

$$S(x_i, x_j)^t = -\frac{G(t)\left(\sum_{z=1}^{d}\left(x_{iz} \times x_{jz}\right)\right)^{-1}}{\left(\sqrt{\sum_{z=1}^{d}\left(x_{iz} - x_{jz}\right)^2}\right)^2}. \tag{16}$$

In equation (16), t is the number of iterations; $(\sum_{z=1}^{d}(x_{iz} \times x_{jz}))^{-1}$ is the reciprocal of sum of the products of the sample point features. The reason for taking the inverse here is that the characteristics of the sample are different from the masses of the objects. In reality, the greater the mass of two objects, the stronger the gravitational force, but the opposite is true for the characteristics of the samples. The closer the characteristics of a sample are to other samples, the stronger the attraction with other samples, and the other samples are close to it. $(\sqrt{\sum_{z=1}^{d}(x_{iz} - x_{jz})^2})^2$ is the distance measure, which takes the value of the second power of the Euclidean distance between the data points; is the gravitational coefficient of the $t^{th}$ iteration, which decreases dynamically with the increase of the number of iterations and can better control the search process, as shown in equation (17).

$$G(t) = G_0 \times e^{-a1/t_{max}}, \tag{17}$$

$t_{max}$ is the maximum number of iteration, $G_0$ and a are constants. The values of $G(t)$ are taken to better balance the global and local search capabilities of the algorithm.

*4.2. Process of GASC Algorithm.* The AdaBoost algorithm trains the weak classifier using initial weights and updates the weights of the training samples according to the learning error rate of the weak classifier, so that the training sample points with higher learning error rates in the previous weak classifier have higher weights. Then, these points with higher error rates are given higher importance in the next weak classifier and the training set after adjusting the weights is used to train the next weak classifier. This process is repeated until the number of weak classifiers reaches a predetermined number, and finally, all weak classifiers are integrated by the established strategy to obtain the final strong learner.

In this study, use the information entropy and AdaBoost algorithms to determine the sample point weights and update the dataset with the weights in order to prepare the optimal parameter values for the next clustering. One point to note is that the sample point weights here are composed of the sample point weights of the incorrectly clustered and correctly clustered sample points from the previous clustering results. The data is updated by such weights in order to enhance the similarity of features between data in the same class.

Segmentation of images using the GASC algorithm requires three stages. Image preprocessing stage to smooth the noise in the image and enhance the segmentation effect. Feature extraction stage: The grayscale co-generation matrix is selected to extract the feature vectors of the pixels in the image to be segmented, and the image to be segmented is represented by the image composed of the feature vectors of all pixels. Segmentation stage: The feature vectors of all pixels in the image are processed using the improved algorithm, i.e., the eigenvalue matrix of the image to be segmented is clustered to obtain the final image segmentation result. In the segmentation result, the pixels divided into the same cluster are displayed with the same gray value and the pixels divided into different clusters are displayed with other different gray values to obtain the segmented image.

In this study, first we proposed the gravity-based adaptive nearest-neighbor propagation clustering algorithm (GASC) to address the problems that the SC clustering algorithm cannot identify non-convex data sets well and the algorithm results are greatly affected by the parameter values. Then the proposed GASC algorithm is applied to image segmentation. The problem that the SC algorithm is not applicable to non-convex data sets can be solved by using the gravitational search mechanism to calculate the similarity matrix; the AdaBoost algorithm selects the samples with the smallest error from all possible weak classifiers and obtains the weights of misclassified and correctly classified samples in the basic classifier from the smallest sample weights. This method can effectively identify those sample points in the cluster that are correctly classified and those that are incorrectly classified samples. Therefore, this study
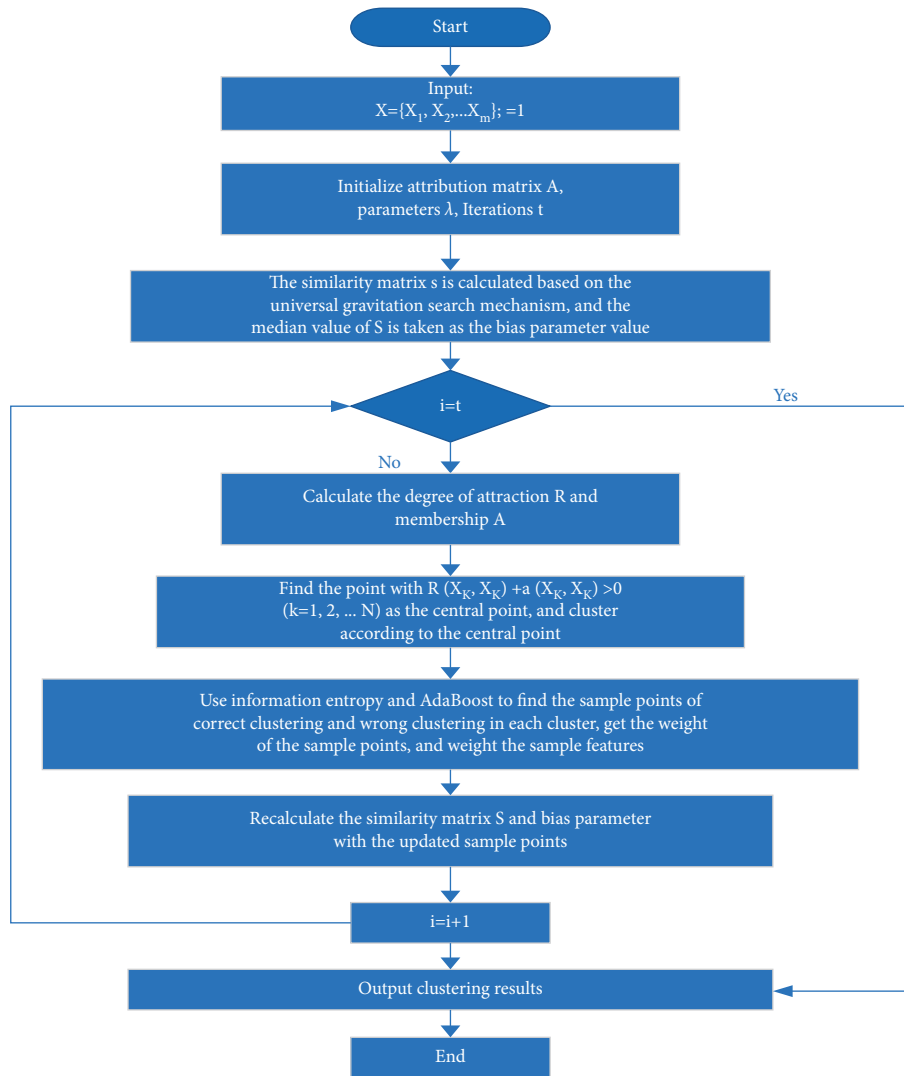
FIGURE 2: GASC algorithm.

uses the information entropy and AdaBoost algorithms to reduce the dependence of the conventional SC clustering algorithm on parameter values and to reduce the number of misclassified sample points in the clusters. The GASC algorithm flow is shown in Figure 2.

The differences between the GASC algorithm and conventional SC algorithm are: the Euclidean distance in conventional SC algorithm tends to make the nonconvex data fall into localization, so SC algorithm is not applicable to nonconvex data; GASC algorithm makes the samples globally optimized under the gravitational search mechanism. The parameter values in the conventional SC algorithm are fixed throughout the clustering process, as well as there will be a large number of misclassified sample points in the clusters obtained by the conventional SC clustering algorithm, thus making the clustering results poor; GASC algorithm first uses information entropy and AdaBoost algorithm to calculate the sample weights of correct and incorrect clusters in each cluster in the previous iteration, and then uses the obtained sample point weights update the

sample features and dynamically adjust the similarity and parameter values for the next iteration. This allows adaptive adjustment of the parameters.

## 5. Experimental Results and Analysis

In this paper, five experiments were done to evaluate the clustering effect of the algorithm using the mean value of the evaluation metrics. The number of iterations $t$ in the GASC algorithm is set to 10, the damping factor is 0.8, the value of $G_0$ is 100, and the value of $a$ is 20. The GASC algorithm is based on the theory of spectral graphs, and its essence is to transform the clustering problem into the optimal partitioning of graphs, which is a point-to-point clustering algorithm. Compared with the conventional clustering algorithm, it has the advantage of being able to cluster on the sample space of arbitrary shape and converge to the global optimal solution. In order to verify its effectiveness, four sets of data with different shapes, namely Circles, Moons, Varied, and Aniso, are selected to determine whether the GASC
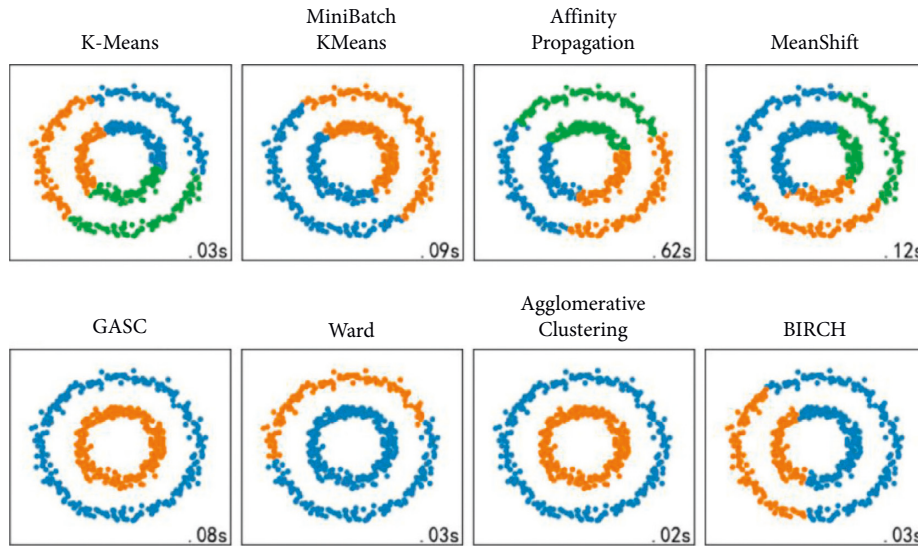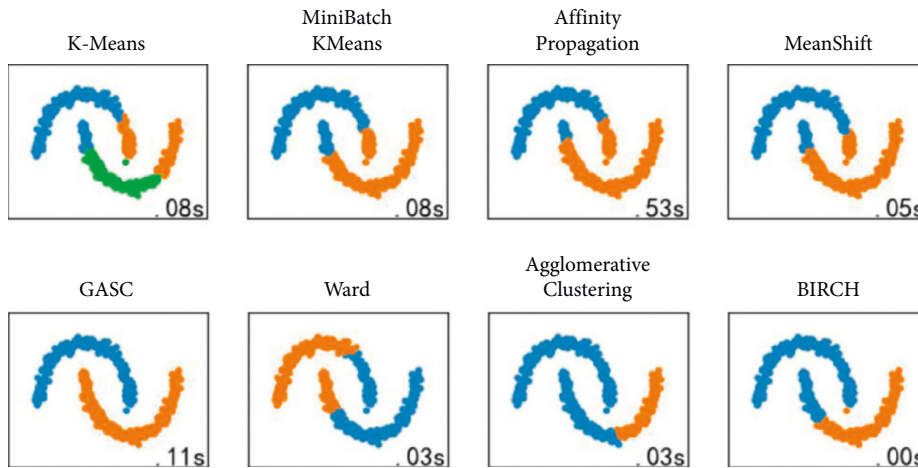
FIGURE 3: Circles clustering.
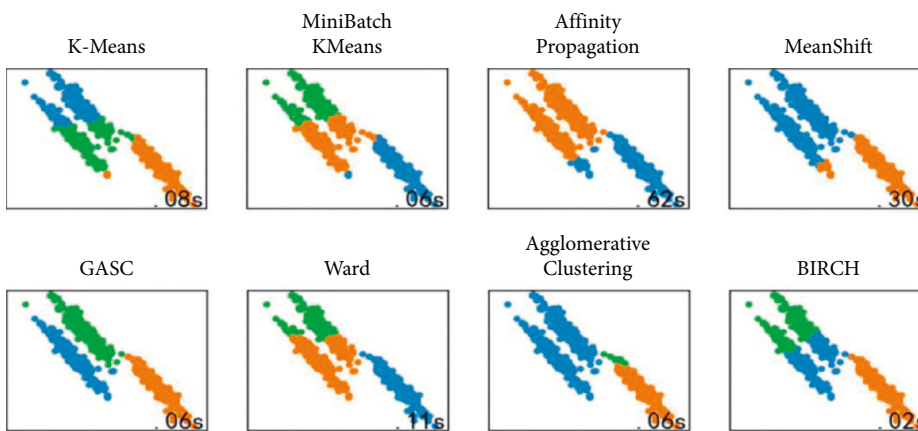


FIGURE 4: Moons clustering.



FIGURE 5: Varied clustering.

algorithm can cluster on the sample space of arbitrary shapes better than other clustering algorithms. The experimental results are shown in Figures 3–6.

The dataset shown in Figure 3 is composed of two circles; the moon-type dataset shown in Figure 4 is composed of two arcs, with the correct category being one class for each curve.
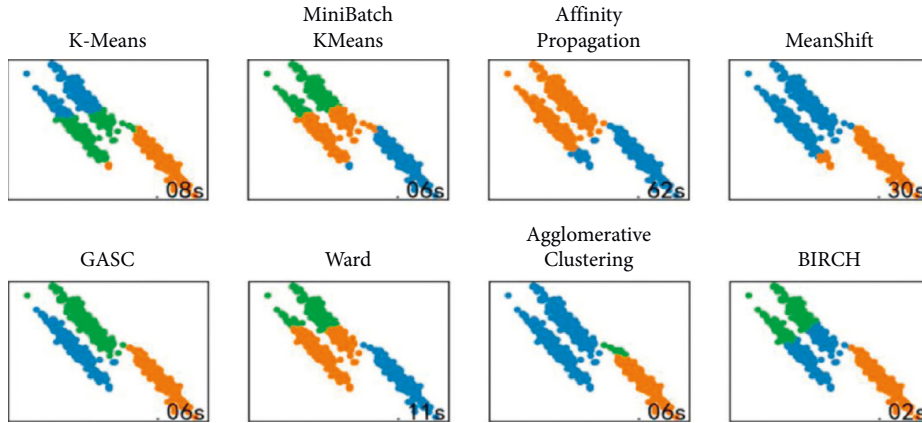
FIGURE 6: Aniso clustering.

TABLE 1: Basic information about the dataset.

| Dataset | Dataset volume | Eigenvalue | Categories |
|---|---|---|---|
| Circles | 500 | 2 | 2 |
| Moons | 500 | 2 | 2 |
| Varied | 500 | 2 | 3 |
| Aniso | 500 | 2 | 3 |
| Iris | 150 | 4 | 3 |
| Wine | 178 | 3 | 3 |
| Seeds-dataset | 210 | 7 | 3 |
| Glass | 214 | 10 | 6 |

TABLE 2: Average score and evaluation of dataset.

| Dataset | K-means | Mini batch K-means | Affinity propagation | Mean shift | GASC | Ward | AC | BIRCH |
|---|---|---|---|---|---|---|---|---|
| Circles | 0.00 | 0.00 | 0.00 | 0.00 | 0.92 | 0.26 | 0.92 | 0.00 |
| Moons | 0.36 | 0.39 | 0.42 | 0.44 | 0.92 | 0.18 | 0.26 | 0.48 |
| Varied | 0.75 | 0.78 | 0.54 | 0.80 | 0.82 | 0.84 | 0.56 | 0.53 |
| Aniso | 0.60 | 0.59 | 0.55 | 0.57 | 0.99 | 0.68 | 0.65 | 0.58 |
| Iris | 0.77 | 0.76 | 0.75 | 0.83 | 0.84 | 0.63 | 0.78 | 0.69 |
| Wine | 0.90 | 0.45 | 0.67 | 0.52 | 0.90 | 0.52 | 0.52 | 0.51 |
| Seeds-dataset | 0.72 | 0.72 | 0.51 | 0.48 | 0.74 | 0.58 | 0.74 | 0.60 |
| Glass | 0.33 | 0.33 | 0.41 | 0.47 | 0.50 | 0.16 | 0.33 | 0.25 |

The dataset shown in Figure 5 is the speckle with different variances, which is divided into three categories in total; the dataset shown in Figure 6 is the anisotropic distribution data, and each group of data is one category. The experimental results of the above dataset can visualize the correctness of the clustering results, the clustering results obtained by running the GASC algorithm on four sets of data are closer to the true class labels. The GASC algorithm only requires the similarity matrix between the data, so it is effective for handling the clustering of sparse data. And because of the use of dimensionality reduction, the complexity in dealing with high-dimensional data clustering is better than conventional clustering algorithms.

Combining the above data effect plots in Figures 3–6 clearly shows that the GASC algorithm can get better clustering results compared to other algorithms on Circles, Moons, Varied and Aniso data sets, therefore, the GASC algorithm can cluster on the sample space of arbitrary shape and converge to the global optimal solution.

In this study, we will validate the SC algorithm with UCI real dataset and artificial dataset, using the following data sets: UCI data (Iris, Seeds-dataset, Wine, Glass), artificial data (Circles, Moons, Varied, Aniso). To better evaluate the clustering performance of this algorithm, this paper compares the K-means algorithm, Mini Batch K-means, AC algorithm (affinity propagation), Mean Shift algorithm, GASC algorithm, Ward algorithm, Agglomerative Clustering algorithm, and BIRCH algorithm for clustering results with evaluation metrics adjusted rand index, mutual information based scores, and V-measure (harmonic average). Table 1 presents the basic information of the dataset. This experiment was run 10 times to take the average of all evaluation metrics to compare the clustering effect, and the average score evaluation table is shown in Table 2.

According to the evaluation metrics tables in Tables 1 and 2, it can be visualized that the overall correctness of the clustering results of the GASC algorithm for these data sets above is higher than that of the K-means algorithm, Mini Batch K-Means algorithm, Affinity propagation algorithm, Mean Shift algorithm, SC algorithm, Ward algorithm, AC algorithm, and BIRCH algorithm. Overall, the class labels obtained by the GASC algorithm are still improved compared to the other seven clustering algorithms. In summary, the GASC algorithm improves the number of correct samples for clustering, can cluster on an arbitrarily shaped sample space, and converges to the global optimal solution. Therefore, compared with other clustering algorithms, the class labels obtained by the GASC algorithm are close to the true labels.

The above result graphs and evaluation index values confirm that the image segmentation results using GASC algorithm can reduce the number of missegmented regions and obtain good segmentation results compared with the image segmentation results of other conventional algorithms.

## 6. Conclusion and Future Work

In this study, an in-depth study of the SC algorithm is done. It is found that the SC clustering algorithm cannot identify nonconvex data sets well and the results of the algorithm are greatly influenced by the parameter values. Therefore, this paper improves the shortcomings of the SC algorithm. In addition, the improved SC clustering algorithm is applied to image segmentation. To address the above problems of clustering algorithm and image segmentation, this paper proposes the Gravitational Adaptive Nearest Neighbor Propagation Spectral Clustering (GASC) algorithm by studying the shortcomings of the SC algorithm. The algorithm adopts the gravitational search mechanism to find the optimal samples based on the conventional SC algorithm, in order to solve the problem that the SC algorithm is not applicable to nonconvex data sets. The algorithm uses information entropy and AdaBoost algorithm to dynamically adjust the parameter values. It is demonstrated that the clustering results of GASC algorithm are improved compared with other clustering algorithms, and the GASC algorithm is applied to image segmentation. Experiments show that GASC algorithm can reduce the number of misclassified regions in the segmentation results and improve the quality of image segmentation compared with other traditional clustering algorithms, Although the GASC algorithm reduces the influence of parameter values and is applicable to any data set. However, this algorithm is not a perfect algorithm, because it mixes many other methods, its operation time consumption is very long, in the follow-up research will focus on how to speed up the operation of the GASC algorithm, and further improve the operation efficiency of the algorithm.

## Data Availability

The authors confirm that the data supporting the findings of this study are available within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

[1] X. P. Liang and D. S. Huang, "Image segmentation fusion using weakly supervised trace-norm multi-task learning method," *IET Image Processing*, vol. 12, no. 7, pp. 1079–1085, 2018.

[2] ZZ. Wang, "Robust segmentation of the colour image by fusing the SDD clustering results from different colour spaces," *IET Image Processing*, vol. 14, no. 13, pp. 3273–3281, 2020.

[3] M. Gong, J. H. Lan, C. L. Yang, H. T. Wu, and T. Zhi, "Adaptive image segmentation algorithm under the constraint of edge posterior probability," *IET Computer Vision*, vol. 11, no. 8, pp. 702–709, 2017.

[4] V. Sidorova, "Global segmentation of textural images on the basis of hierarchical clusterization by predetermined cluster separability," *Pattern Recognition and Image Analysis*, vol. 25, no. 3, pp. 541–546, 2015.

[5] G. Akbarizadeh and M. Rahmani, "Efficient combination of texture and color features in a new spectral clustering method for PolSAR image segmentation," *National Academy Science Letters*, vol. 40, no. 2, pp. 117–120, 2017.

[6] P. S. S. Kumar and V. S. Dharun, "Combination of fuzzy C-means clustering and texture pattern matrix for brain MRI segmentation," *Biomedical Research-India*, vol. 28, no. 5, pp. 2046–2049, 2017.

[7] J. Wu, C. X. Liu, and B. Li, "Texture-aware and structure-preserving superpixel segmentation," *Computers & Graphics*, vol. 94, pp. 152–163, 2021.

[8] X. Y. Chen, X. F. Peng, and S. A. Wang, "Superpixel segmentation based on grid point density peak clustering," *Sensors*, vol. 21, no. 19, p. 6374, 2021.

[9] A. Pugazhenthi and L. S. Kumar, "Automatic cloud segmentation from INSAT-3D satellite image via IKM and IFCM clustering," *IET Image Processing*, vol. 14, no. 7, pp. 1273–1280, 2020.

[10] M. Chu, M. Y. Zhuang, and T. H. Sun, "Image segmentation based on SOFM in camouflage design," in *Proceedings of the Eleventh International Conference on Graphics and Image Processing (ICGIP 2019)*, Hangzhou, China, October 2020.

[11] D. Koundal, "Texture-based image segmentation using neutrosophic clustering," *IET Image Processing*, vol. 11, no. 8, pp. 640–645, AUG 2017.

[12] J. Zhang, Y. W. Gao, S. W. Feng, Z. H. Chen, and Y. B. Yuan, "Image segmentation with texture clustering based jseg," *Proceedings of 2015 International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 599–603, 2015.

[13] J. M. Mathew and P. Simon, "Color texture image segmentation based on neutrosophic set and nonsubsampled contourlet transformation," *Applied Algorithms*, vol. 8321, pp. 164–173, 2014.

[14] M. Kim, J. M. Lim, H. Shin, C. Oh, and H. T. Jeong, "Estimating the number of clusters with database for texture segmentation using gabor filter," in *Proceedings of the Computer Vision Systems (ICVS 2015)*, pp. 435–444, Copenhagen, Denmark, 2015.

[15] Y. F. Li, Q. J. Zhao, W. B. Zhang et al., "A simultaneous cartoon-texture image segmentation and image decomposition method," *Chinese Journal of Electronics*, vol. 29, no. 5, pp. 906–915, 2020.

[16] M. M. Rahman and S. Horiguchi, "A fully unsupervised color textured image segmentation algorithm using weighted mean histograms features," *Signal Image and Video Processing*, vol. 6, no. 2, pp. 197–209, 2012.

[17] Z. Tirandaz and G. Akbarizadeh, "Unsupervised texture-based SAR image segmentation using spectral regression and gabor filter bank," *Journal of the Indian Society of Remote Sensing*, vol. 44, no. 2, pp. 177–186, 2016.

[18] Y. Liu, G. D. Liu, C. Y. Liu, and C. M. Sun, "A novel color-texture descriptor based on local histograms for image segmentation," *IEEE Access*, vol. 7, pp. 160683–160695, 2019.

[19] A. Salmi, K. Hammouche, and L. Macaire, "Constrained feature selection for semisupervised color-texture image segmentation using spectral clustering," *Journal of Electronic Imaging*, vol. 30, no. 1, 2021.

[20] A. E. Baya, M. G. Larese, and R. Namias, "Clustering stability for automated color image segmentation," *Expert Systems with Applications*, vol. 86, pp. 258–273, 2017.

[21] H. Du, Z. H. Wang, D. Wang, and X. L. Wang, "Multi-feature fusion method applied in texture image segmentation," in *Proceedings of the 2018 14th International Conference on Computational Intelligence and Security (Cis)*, pp. 135–139, Hangzhou, China, November 2018.

[22] X. L. Tian, L. C. Jiao, and X. H. Zhang, "A clustering algorithm with optimized multiscale spatial texture information: application to SAR image segmentation," *International Journal of Remote Sensing*, vol. 34, no. 4, pp. 1111–1126, 2013.

[23] A. Heshmati, M. Gholami, and A. Rashno, "Scheme for unsupervised colour-texture image segmentation using neutrosophic set and non-subsampled contourlet transform," *IET Image Processing*, vol. 10, no. 6, pp. 464–473, 2016.

[24] MC. Catalbas, "Unsupervised multi-label image and texture segmentation based on optimal feature representation," *Tehnicki Vjesnik-Technical Gazette*, vol. 26, no. 5, pp. 1275–1283, 2019.

[25] R. H. Shang, P. Peng, F. H. Shang, L. C. Jiao, Y. F. Shen, and R. Stolkin, "Semantic segmentation for SAR image based on texture complexity analysis and key superpixels," *Remote Sensing*, vol. 12, no. 13, p. 2141, 2020.