



# Fuzzy C-Means Clustering with Region Constraints for Superpixel Generation

Xiaohong Jia<sup>1,2</sup> · Yao Zhao<sup>3</sup> · Bin Zhang<sup>1</sup> · Xuejun Zhang<sup>1</sup> · Guanghui Yan<sup>1</sup>

Received: 17 October 2024 / Revised: 23 January 2025 / Accepted: 17 February 2025  
© The Author(s) 2025

**Abstract** Superpixels are becoming increasingly essential in computer vision applications, as they significantly reduce the number of computational primitives required for subsequent tasks. However, most existing superpixel algorithms primarily focus on color intensity and position coordinates, often ignoring local neighborhood factors and region constraints. This oversight reduces their effectiveness in noisy and cluttered environments. To address this issue, we propose a seminal and novel Fuzzy C-Means clustering with Region Constraints for Superpixel generation (RCFCMS). First, RCFCMS employs watershed transform to establish region constraints, effectively preventing boundary crossings. Second, relative position relationship is utilized to initialize superpixel centers within each region, offering an adaptive grid layout. Third, RCFCMS adopts spatial information to perceive local similarity, significantly mitigating noise interference. Finally, connected components are

applied to eliminate isolated pixels, further optimizing pixel assignment. The experimental results on two benchmark datasets demonstrate that the proposed RCFCMS achieves competitive performance across varying numbers of superpixels. Specifically, when the number of superpixels is 500, RCFCMS achieves the following metrics: UE = 0.102, ASA = 0.948, FM = 0.401, and CO = 0.572 on BSDS500; and UE = 0.124, ASA = 0.935, FM = 0.296, and CO = 0.608 on SBD. These results further highlight the effectiveness of RCFCMS for superpixel generation.

**Keywords** Fuzzy C-means clustering · Superpixel generation · Region constraints · Noise interference

## 1 Introduction

Superpixels exploit low-level local properties of images to obtain meaningful atomic regions [1], which can effectively abstract the main structure and significantly reduce the unessential redundancy. Therefore, superpixels are closer to the human visual perception of objects. Since superpixels can capture critical features of images while improving computational efficiency, it has been widely used in computer vision, such as image segmentation [2], quality assessment [3], saliency estimation [4], and 3D reconstruction [5], to name a few.

The good superpixels are expected to have the following properties [6–8]: (1) Compactness [6]: the distribution of superpixels is regular and uniform in the homogenous region; (2) Connectivity [6]: each pixel belongs to only one superpixel with a connected region; (3) Controllability [7]: the number of superpixels can be flexibly set; (4) Stability [7]: the superpixels can overcome interference of different textures; (5) Performance [8]: the superpixels can adhere well to the

✉ Xiaohong Jia  
jiaxhm@163.com

Yao Zhao  
yzhao@bjtu.edu.cn

Bin Zhang  
zhangbin87@mail.lzjtu.cn

Xuejun Zhang  
xuejunzhang@mail.lzjtu.cn

Guanghui Yan  
ghyan@mail.lzjtu.cn

<sup>1</sup> School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China

<sup>2</sup> The Key Laboratory of Big Data and Artificial Intelligence in Transportation, Ministry of Education (Beijing Jiaotong University), Beijing 100044, China

<sup>3</sup> Institute of Information Science, Beijing Jiaotong University, Beijing 100091, China

boundaries of the image; and (6) Efficiency [8]: the superpixels can be quickly generated while being memory efficient. Over the past few decades, many superpixel algorithms have been designed in this direction. They can be broadly categorized as graph-based algorithms, gradient-based algorithms, and cluster-based algorithms.

The graph-based algorithms utilize edge weights connecting nodes on a graph to measure the similarity of pixels, thereby creating superpixels. The pioneering example is Normalized cuts (Ncuts) [9] by globally minimizing a cost function. Since Ncuts considers all the nodes of the graph [10], it is time-consuming. To speed Ncuts up, the researchers introduced constraints [11] to optimize computing strategies [12]. However, they still cannot meet the practical requirements. Afterward, Moore et al. [13] described superpixel lattices keeping the regular topology of the grid. Yet, its performance is unstable. On this basis, entropy rate superpixel (ERS) [14] was proposed, which employs entropy rate and balancing term to surpass the previous quality. Further, Li et al. [15] proposed linear spectral clustering (LSC), which applies weighted k-means to optimize the cost function of normalized cuts. Focusing on random walk (RW), lazy RW [16] and dynamic RW [17] were proposed, respectively. They not only recognize regions with intensity variability, but also have linear complexity.

The gradient-based algorithms map the distribution of image gradient as the altitude information, then apply water immersion to obtain over-segmentation, that is, superpixels. However, the typical watershed transform (WT) [18] is sensitive to gradient magnitude, causing it to produce unusable results of superpixels. Some researchers employ various morphological operators to remove useless local minima so as to overcome the mentioned problem. Among them, adaptive morphological reconstruction (AMR) proposed by Lei et al. [19] is more practical. Unfortunately, these algorithms cannot control the number of superpixels and corresponding shapes. The TurboPixels [20] algorithm introduces compactness constraints to create a lattice-like structure for adhering to local image boundaries. Subsequently, Machairas et al. [21] proposed waterpixels, which uses a regularized gradient to balance superpixel regularity and boundary respect. The waterpixels can achieve promising results. However, its time complexity is significant. Utilizing the gradient derived from structured forest edges [22], Wei et al. [23] proposed superpixel hierarchy (SH) through a region-merging approach. Recently, Yuan et al. [24] introduce watershed-based superpixels, employing a novel flooding priority and boundary marching strategy (WSBM) to enhance both speed and boundary adherence. Similarly, Yuan et al. [25] proposed the superpixels with content-adaptive criteria (SCAC), which further considers the gradient, color, and texture to reduce the segmentation error.

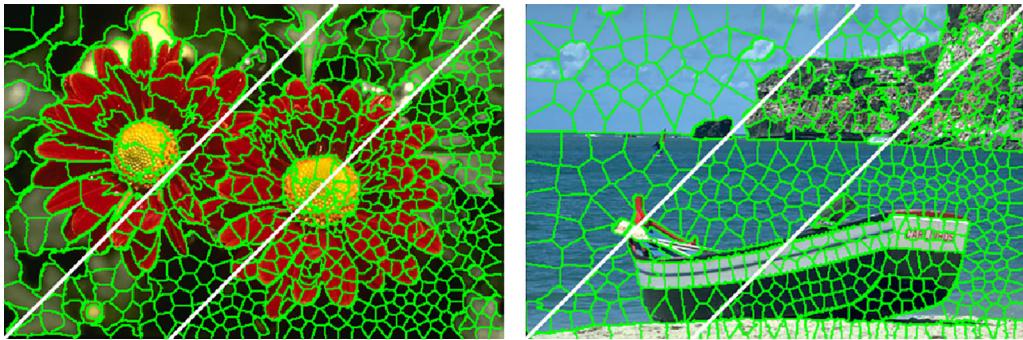
The cluster-based algorithms use local feature differences with intensity and coordinates in limited search region to construct superpixels. The mean shift and quick shift find a convergence of seeds via local maximum density. Insufficiently, they produce superpixels with uneven size and lack of compactness. To solve the problems, simple linear iterative clustering (SLIC) based on restricted k-means was proposed by Achanta et al. [26]. The SLIC not only offers flexibility and compactness but also efficiently respects boundaries. Inspired by SLIC, density-based spatial clustering of applications with noise (DBSCAN) [27], and Gaussian mixture model (GMMSP) [28] are also adapted to generate comparable superpixels.

The aforementioned algorithms primarily rely on intensity features and gradient information to generate superpixels, often neglecting the importance of feature selection [29–31] and feature optimization [32, 33]. To explore richer feature representations, researchers have adopted a two-stage strategy [34], where convolutional neural networks (CNNs) are first employed to extract features, followed by traditional methods for recognition and classification tasks [35–37]. However, this approach lacks the efficiency of end-to-end deep learning models. To address this limitation, fully convolutional networks have been proposed to streamline the process and directly generate superpixels. While CNNs excel in capturing local features, they face challenges in modeling global representations effectively. As an alternative, Transformer architectures [38] have been introduced, offering improved capabilities in capturing long-range dependencies.

Although deep learning-based methods [39, 40] often demonstrate superior performance compared to conventional algorithms, they typically depend on a supervised learning framework, requiring extensive training on a GPU. Additionally, these methods struggle to generate a precise number of superpixels. In this paper, we primarily analyze unsupervised superpixel algorithms without the use of labeled training data.

Since its intuitiveness and simplicity, clustering strategies are widely favored by researchers. The paper mainly explores the fuzzy C-means (FCM) to create superpixels. As is well known, FCM with soft takes into account ambiguity attributes by assigning each data point a degree of membership to multiple clusters. This characteristic enables FCM to effectively manage uncertainty and overlapping samples, making it particularly suitable for complex datasets. However, FCM ignores spatial context information in images, it can be sensitive to noise and artifacts. In order to compensate for the defect of FCM, fuzzy local information C-means clustering (FLICM) [41] was proposed, which integrates local intensity and spatial information, resulting in more robust performance.

By researching SLIC based on k-means and FLICM based on neighborhood information, we propose a seminal and



**Fig. 1** Superpixel results with 200/500/1000 superpixels using the proposed RCFCMS, where each region represents similar colors

novel fuzzy C-means clustering with region constraints for superpixel generation (RCFCMS). The proposed RCFCMS employs local spatial information to capture neighborhood differences, which can help suppress interference within images. Meanwhile, the proposed RCFCMS adopts fuzzy measurement to distinguish ambiguity, which can perceive the membership variety of each pixel. Additionally, the proposed RCFCMS considers the region constraints to avoid forced boundary crossings, which can keep the contour structure of the target. Figure 1 shows the superpixel results of RCFCMS. The main contributions of the proposed RCFCMS algorithm are summarized as follows:

- The RCFCMS is a novel attempt to utilize fuzzy C-means clustering with region constraints to generate superpixels.
- The RCFCMS employs pixel-level local spatial information to increase the homogeneous property from polluting noise.
- The RCFCMS adopts the region-level contour structure to prevent the superpixel diffusion from crossing edges.

The remainder of the paper is structured as follows. Section 2 briefly describes the motivation. In Sect. 3, we carefully analyze the methodology of the proposed RCFCMS. In the next Section, experimental results will be fully demonstrated. Finally, we deeply conclude our work in Sect. 5.

## 2 Motivation

The SLIC [26] employs  $k$ -means with hard attributes to create superpixels. It is a surprisingly simple and strikingly effective method in practice. However, the membership values of  $k$ -means only allow 0 or 1, ignoring the fact that ambiguity exists in the dataset. In order to overcome the drawback, the FCM was proposed. Unfortunately, FCM is sensitive to outliers because it does not consider the spatial information of an image. Thus, many researchers incorporated neighbor constraints into the original FCM to enhance the applicabil-

ity of image data. Among them, the FLICM [41] proposed by Krinidis and Chatzis is free of parameter selection, and it has gained more attention. In addition, SLIC restricts the search region around each cluster center but overlooks the consistency of pixel assignments within the regions. Conversely, AMR-WT [19] generates regions with larger receptive fields, facilitating more effective capture of target information.

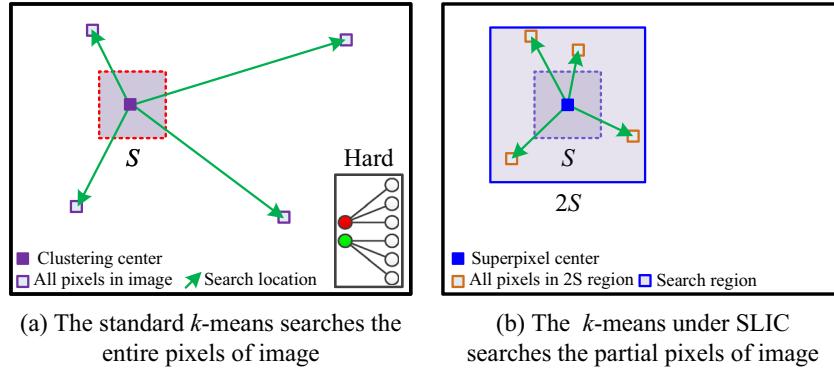
Motivated by the limited search region from SLIC, local information from FLICM, and adaptive region constraints from AMR-WT, we propose a superpixel generation method based on FCM with region constraints. This approach combines the advantages of soft measurement and neighborhood information, resulting in more desirable superpixels.

### 2.1 SLIC

Since superpixels can extract the main features of an image, it is usually used as a preprocessing technique for visual tasks. The SLIC is one of the most novel and popular superpixel algorithms. It can easily and quickly execute while producing uniformly and compactly distributed a set of super-regions.

The SLIC utilizes  $k$ -means to assign each pixel to one of the superpixels in the two-dimensional position  $[p, q]^T$  and three-dimensional CIELAB color space  $[\ell, a, b]^T$ . It is worth noting that the conventional  $k$ -means needs to measure the similarity between all tested pixels and different cluster centers. When the number of cluster centers tends to be large, the time-consuming obviously increases. In order to speed up the compute efficiency of  $k$ -means, the SLIC limits the size of the search region and reduces the calculation of redundant distance, as shown in Fig. 2. This approach makes complexity of SLIC independent of the number of superpixels.

The SLIC crucially operates as follows. First, initializing cluster centers with 5-dimensional space by the regular grid interval  $S = \sqrt{N/K}$ , where  $N$  and  $K$  are the number of pixels and superpixels, respectively. Second, adopting the local perturbations to avoid seed locations at the highest gradient position. Subsequently, the SLIC mainly utilizes the combined sum distance of position  $[p, q]^T$  and color



**Fig. 2** The comparison of the search region on standard  $k$ -means and limited  $k$ -means. The conventional  $k$ -means traverses every pixel in image. However, the  $k$ -means under SLIC only focus on local pixels

within a  $2S \times 2S$  region ( $S \times S$  stands for expected superpixel size). This limited approach can help SLIC to reduce distance calculation

$x \in [\ell, a, b]^T$  to measure the similarity between each pixel and the nearest superpixel center as follows:

$$d_c = \sqrt{\|x_j - c_i\|^2}, \quad (1)$$

$$d_s = \sqrt{\|p_j - v_i\|^2 + \|q_j - h_i\|^2}, \quad (2)$$

$$D_{ij} = \sqrt{d_c^2 + \left(\frac{d_s^2}{S}\right) w^2}, \quad (3)$$

where  $x_j$  is color intensity of the  $j$ th pixel and  $(p_j, q_j)$  are its position coordinates.  $c_i$  denotes the  $i$ th clustering center of superpixel color and  $(v_i, h_i)$  denotes the  $i$ th clustering center of superpixel position.  $\|\cdot\|$  represents the Euclidean distance. According to [26],  $D_{ij}$  is a simplified distance measure. The parameter  $w$  is used to control the relative importance between color intensity similarity  $d_c$  and spatial position proximity  $d_s$ . By minimum distance of  $D_{ij}$ , compute superpixel assignment is defined as follows:

$$u_{ij} = \begin{cases} 1 & D_{ij} = \min_{i \in \{1:k\}} (D_{ij}), \\ 0 & \text{Otherwise} \end{cases}, \quad (4)$$

where  $u_{ij}$  stands for the hard membership of the  $i$ th target pixel with 5-dimensional space belonging to the  $j$ th superpixel center, and  $k$  is the number of clusters in the overlapping region. Then  $u_{ij}$  is employed to update the superpixel centers of color intensity and spatial position. By iterating hard membership and superpixel centers, the superpixels are created.

Inspired by SLIC, fuzzy superpixels (FS) [42] and fuzzy simple linear iterative clustering (fuzzy SLIC) [43] have been proposed consecutively. FS utilizes the membership degree from FCM in overlapping regions to assign image classifications, but it neglects spatial structural information. To address this limitation, fuzzy SLIC introduces a linear

combination of local membership to generate more robust superpixels. Recently, Wu et al. [44] proposed a robust joint learning approach for superpixel generation and superpixel-based image segmentation, which builds upon the concept of fuzzy SLIC. The above algorithms all adopt the limited search region strategy from SLIC, highlighting that this strategy remains a significant source of influence for many researchers.

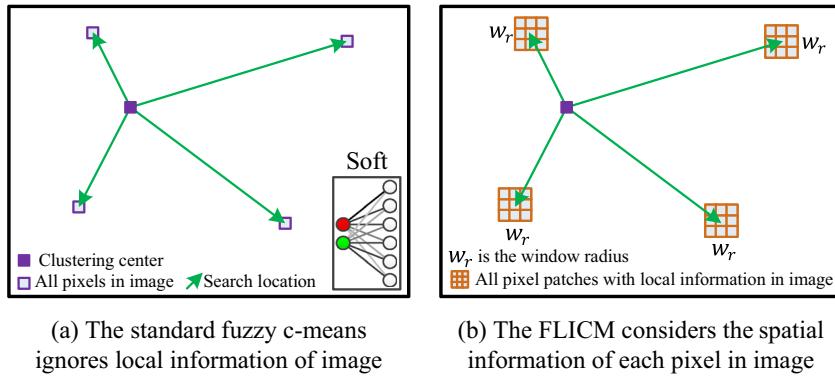
## 2.2 FLICM

The  $k$ -means assumes that each sample only belongs to one of the categories, resulting in clear partition boundaries. The process embodies the characteristic of ‘either this or that.’ However, there is no definite property of partial image data. So, the  $k$ -means may mis-divide some pixels. The FCM with the characteristic of ‘both this and that’ overcomes the mentioned above shortcomings. Unfortunately, the FCM is sensitive to intensity non-uniformity because it does not consider the spatial information of an image. To improve the performance of FCM for image, many researchers have incorporated local neighborhoods into the original FCM.

In this Section, we focus on the well-known fuzzy local information C-means clustering (FLICM). The FLICM defines a novel fuzzy factor with spatial information to suppress noise or outliers. Meanwhile, the FLICM is free of parameter selection except for the number of clusters. Compared with FCM, FLICM utilizes the neighboring pixels within a window to correct target pixel classification, as depicted in Fig. 3. It is not difficult to find that FLICM has better immunity to noise and cluttered pixels.

The objective function of FLICM proposed by Krinidis and Chatzis is defined as follows:

$$J = \sum_{i=1}^{\tau} \sum_{j=1}^N \left( u_{ij}^m \|x_j - c_i\|^2 \right)$$



**Fig. 3** The difference between FCM and FLICM in the clustering process, where  $w_r$  is the window radius, it is usually set to 1 or 2

$$+ \sum_{\substack{r \in \mathcal{N}_j \\ r \neq j}} \frac{1}{d_{jr} + 1} (1 - u_{ir}^m)^2 \|x_r - c_i\|^2 \Big), \quad (5)$$

where  $\tau$  is the number of clusters,  $N$  is the number of all pixels in an image,  $u_{ij}$  denotes the membership degree of the  $j$ th pixel  $x_j$  with respect to the  $i$ th clustering center  $c_i$ , and  $u_{ij}$  satisfies the constraint that  $\sum_{i=1}^{\tau} u_{ij} = 1$ ,  $m$  denotes a fuzzy exponent. The second term of objective function represents fuzzy factor,  $\mathcal{N}_j$  represents neighbor index within full window,  $d_{jr}$  represents spatial Euclidean distance of  $x_j$  and its neighboring pixel  $x_r$ .

According to the optimization process of FCM, in FLICM, the calculation of  $u_{ij}$  and  $c_i$  is obtained as follows:

$$u_{ij} = \frac{\sum_{\substack{r \in \mathcal{N}_j \\ r \neq j}} \frac{1}{d_{jr} + 1} (1 - u_{ir}^m)^2 \|x_r - c_i\|^2 \frac{-1}{m-1}}{\sum_{t=1}^{\tau} (\|x_j - c_t\|^2 + \sum_{\substack{r \in \mathcal{N}_j \\ r \neq j}} \frac{1}{d_{jr} + 1} (1 - u_{ir}^m)^2 \|x_r - c_t\|^2 \frac{-1}{m-1})}, \quad (6)$$

$$c_i = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}. \quad (7)$$

By iteratively updating Eqs.(6) and (7), FLICM can achieve the image segmentation. Since FLICM traverses every pixel of image.

Building upon FLICM, the deviation-sparse fuzzy C-means (DSFCM) algorithm [45] was proposed, which leverages estimated deviations to suppress outliers. Subsequently, researchers embedded multi-objective information [46] and rich spatial information [47] into FCM, thereby enhancing its segmentation performance. To improve FCM's ability to capture details, Yang et al. [48] proposed the fuzzy c-multiple-means clustering algorithm, which constructs a sub-clustering strategy and explores intra-cluster disparity attributes. Recently, Wang et al. [49] conducted a comparative analysis of DSFCM and residual-driven FCM in terms of noise evaluation. These studies all incorporate the local

information strategy inspired by FLICM, underscoring the ongoing influence of this strategy in the field.

### 2.3 AMR-WT

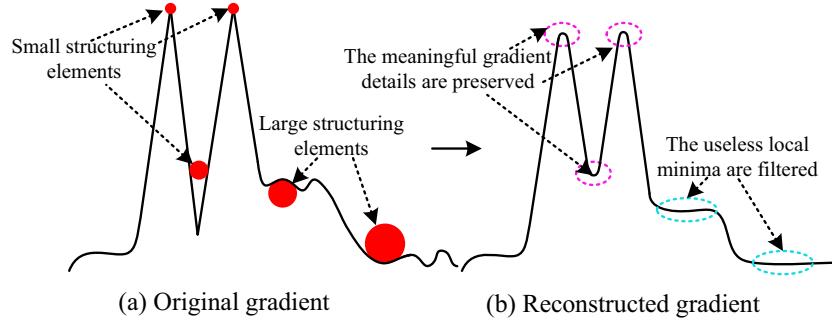
The SLIC algorithm is a typical method for generating a controllable number of superpixels, but it struggles to produce satisfactory segmentation results when the number of superpixels is small. To capture target contour information of an image, watershed transform (WT) [18] based on morphological reconstruction has been proposed. However, MR uses a single-scale structuring element to adjust all gradient magnitudes, which may lead to under-segmentation or over-segmentation. To address this limitation, Lei et al. [19] introduced an adaptive morphological reconstruction (AMR) with operation that is able to generate a better seed image.

Compared to MR, AMR employs multi-scale structuring elements to adaptively reconstruct the gradient image, i.e., small-scale structuring elements smooth larger gradient magnitudes, while large-scale elements smooth smaller gradient magnitudes, as shown in Fig. 4. AMR effectively removes interfering regional minima while preserving key seed points, which is defined as follows:

$$\psi(g, s, z) = V_{s \leq e \leq z} \left\{ R_g^C(f)_{b_e} \right\}, \quad (8)$$

where  $g$  is the gradient image corresponding to the input image,  $f$  is the marker image, defined as  $f = \varepsilon_{b_e}(g)$ ,  $\varepsilon$  represents the erosion operation,  $R_g^C$  represents the morphological closing reconstruction,  $b_s \subseteq \dots \subseteq b_e \subseteq b_{e+1} \dots \subseteq b_t$  denotes a series of nested structuring elements, where  $s \leq e \leq z$ ,  $s \geq 1$ , and  $s, e, z \in N^+$ ,  $V$  stands for the pointwise maximum.

AMR can offer a significant gradient image, then utilizes WT to generate a boundary-adherent segmentation result. Therefore, we can use the results of AMR-WT as prior knowledge, and on this basis, we construct a controllable number of superpixels.



**Fig. 4** The gradient reconstruction process of AMR

### 3 Methodology

As illustrated above, the smaller limited search region promotes faster superpixel generation, while the local information effectively improves segmentation accuracy. Guided by these strategies, we propose fuzzy C-means clustering with region constraints for superpixel generation (RCFCMS). The proposed RCFCMS incorporates constrained region and local information in its objective function. It is worth noting that the constrained region is not limited to  $2S \times 2S$  but extends to a more extensive and receptive region derived from prior contour structures. Meanwhile, the spatial information can discriminate neighborhood differences. Therefore, the RCFCMS has better performance. Figure 5 shows the critical differences between SLIC and RCFCMS.

#### 3.1 The Framework of RCFCMS

We utilize the AMR-WT to provide adaptive region constraints, effectively preventing forced cross-boundary segmentation, as depicted in Fig. 6b. In each sub-region, superpixel centers are initialized to maximize the distance from both the boundaries and other superpixel centers, resulting in a more uniform distribution of initialized centers, as shown in Fig. 6c. Following this, FCM incorporating local information is employed to assign pixels within each sub-region. Figure 6d illustrates the heatmap of the highest membership values obtained from the FCM. Based on the membership intensities, the final superpixels are generated, as demonstrated in Fig. 6e. It can be observed that RCFCMS inherits the prior contours from AMR-WT and exploits the fuzzy properties of FCM. Similar to SLIC, the region constraints enhance the execution efficiency of RCFCMS, while, like FLICM, the local information improves the noise immunity of RCFCMS.

The proposed RCFCMS leverages the advantages of both region constraints and local information. Its objective function within the sub-region mask is defined as follows:

$$J = \sum_{i=1}^{k_l} \sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in N_j} \frac{1}{d_{jr} + 1} \times \left( \|x_r - c_i\|^2 + \left( \frac{\|p_r - v_i\|^2 + \|q_r - h_i\|^2}{S_l} \right) w^2 \right), \quad (9)$$

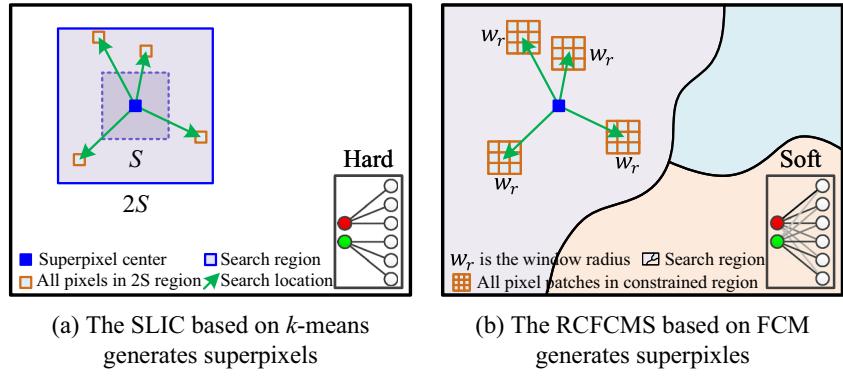
where  $k_l$  is the number of cluster within the  $l$ th regions-of-target, and  $\sum_{l=1}^L k_l = K$ ,  $L$  represents the total number of regions obtained by AMR-WT, while  $K$  denotes the number of superpixels set by the user,  $n_l$  is the number of the pixels falling into the  $l$ th irregular mask, and  $\sum_{l=1}^L n_l = N$ ,  $N$  is the number of all pixels input image,  $k_l$  is obtained by  $k_l = \frac{K \times n_l}{N}$ ,  $S_l = \sqrt{n_l/k_l}$  represents the approximate grid interval in the  $l$ th regions,  $w$  represents weight factor of compactness,  $x_r$ ,  $(p_r, q_r)$  are neighboring intensity and corresponding coordinates of the  $j$ th pixel  $x_j$ ,  $(v_i, h_i)$  are the vertical and horizontal coordinates of the  $i$ th superpixel center  $c_i$ .

In order to minimize revised objective function Eq. (9), we embed the membership constraint into objective function using the Lagrange multiplier  $\lambda_j$  as follows:

$$\tilde{J} = \sum_{i=1}^{k_l} \sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in N_j} \frac{1}{d_{jr} + 1} (\|x_r - c_i\|^2 + \left( \frac{\|p_r - v_i\|^2 + \|q_r - h_i\|^2}{S_l} \right) w^2) + \sum_{j=1}^{n_l} \lambda_j (\sum_{i=1}^{k_l} u_{ij} - 1). \quad (10)$$

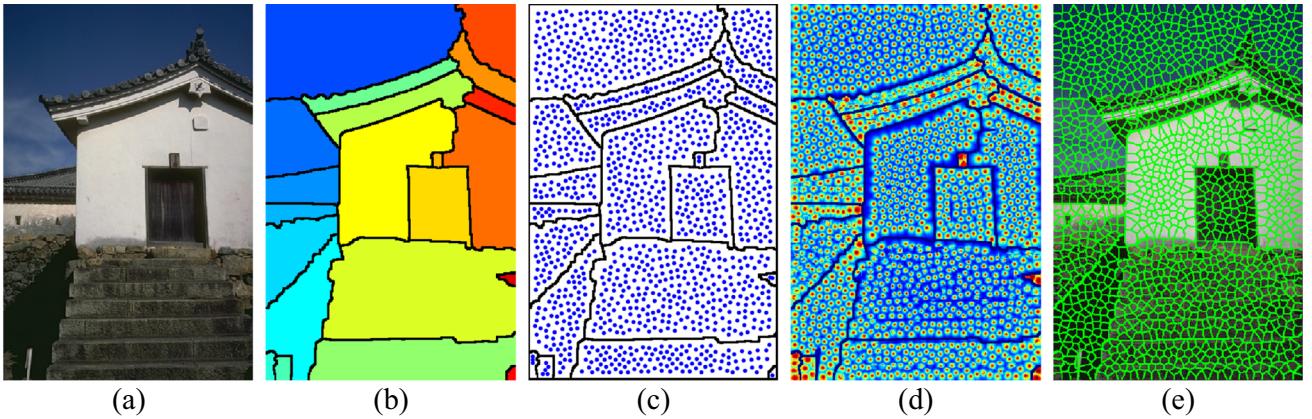
By taking the partial derivative of  $\tilde{J}$  with respect to  $u_{ij}$ , we have the following:

$$\frac{\partial \tilde{J}}{\partial u_{ij}} = mu_{ij}^{m-1} \left( \sum_{r \in N_j} \frac{1}{d_{jr} + 1} (\|x_r - c_i\|^2 + \left( \frac{\|p_r - v_i\|^2 + \|q_r - h_i\|^2}{S_l} \right) w^2) \right) + \lambda_i = 0. \quad (11)$$



**Fig. 5** Comparison of the architectures of SLIC and RCFCMS. SLIC employs  $k$ -means within a  $2S \times 2S$  search region to generate superpixels, which have a small receptive field and disregard the spatial

information of the image. In contrast, RCFCMS utilizes FCM within a constrained region to generate superpixels, which have a larger receptive field and incorporate the spatial information of the image



**Fig. 6** The framework of the proposed RCFCMS. **a** Input image. **b** Region constraints from AMR-WT [19]. **c** Initialize centers. **d** The highest membership heatmap. **e** The superpixel result

By substituting  $\sum_{i=1}^{k_l} u_{ij} = 1$  into Eq. (11), the updating  $u_{ij}$  is obtained as follows:

$$u_{ij} = \frac{\left( \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1} (\|x_r - c_i\|^2 + (\frac{\|p_r - v_i\|^2 + \|q_r - h_i\|^2}{S_l}) w^2) \right)^{-\frac{1}{m-1}}}{\sum_{t=1}^{k_l} \left( \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1} (\|x_r - c_t\|^2 + (\frac{\|p_r - v_i\|^2 + \|q_r - h_t\|^2}{S_l}) w^2) \right)^{-\frac{1}{m-1}}}. \quad (12)$$

By taking the partial derivative of  $\tilde{J}$  with respect to  $c_i$ , we get the following:

$$\frac{\partial \tilde{J}}{\partial c_i} = \sum_{j=1}^{n_l} 2 \left( u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1} (x_r - c_i) \right) = 0. \quad (13)$$

Thus, the updating  $c_i$  can be obtain as follows:

$$c_i = \frac{\sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1} x_r}{\sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1}}. \quad (14)$$

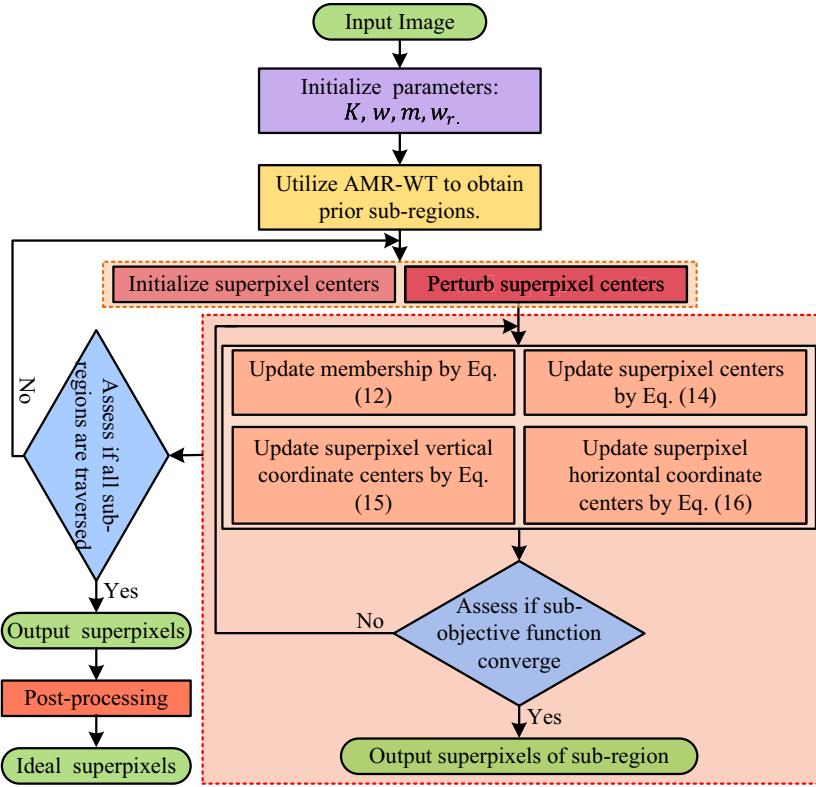
Since  $c_i$ ,  $v_i$ , and  $h_i$  adopt the similar calculation strategy, it can be easily inferred that updating  $v_i$  and  $h_i$  is as follows:

$$v_i = \frac{\sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1} p_r}{\sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1}}, \quad (15)$$

$$h_i = \frac{\sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1} q_r}{\sum_{j=1}^{n_l} u_{ij}^m \sum_{r \in \mathcal{N}_j} \frac{1}{d_{jr}+1}}. \quad (16)$$

The systematic flowchart of RCFCMS is summarized in Fig. 7. The corresponding pseudocode of Fig. 7 is presented in Algorithm 1. Based on the calculation steps outlined in the framework and pseudocode, it is evident that RCFCMS is both simple and intuitive.

Compared to both SLIC and FS, the RCFCMS breaks the limitation of local search and introduces prior contours. Simultaneously, the objective function of RCFCMS is more intuitive. The above description also verifies the superiority of RCFCMS.



**Fig. 7** Systematic flowchart of RCFCMS

### Algorithm 1 RCFCMS superpixel generation

- 1: Initialize superpixel number  $K$ , compactness coefficient  $w$ , fuzzy exponent  $m$ , window radius  $w_r$ , sub-region provided by AMR-WT.
- 2: **for** each sub-region **do**
- 3: Initialize superpixel centers  $C_i = [c_i, v_i, h_i]^T$  within each sub-region, where  $i \in [1, k_l]$ .
- 4: Perturb superpixel centers to the lowest gradient position in a  $3 \times 3$  neighborhood.
- 5: **repeat**
- 6:     Update membership values  $U$  using Eq. (12);
- 7:     Update superpixel intensity centers  $C$  using Eq. (14);
- 8:     Update superpixel vertical coordinate centers  $V$  using Eq. (15);
- 9:     Update superpixel horizontal coordinate centers  $H$  using Eq. (16);
- 10:   **until** convergence of sub-objective function Eq. (9);
- 11:   Assign pixel categories using the highest membership  $\max\{U\}$ .
- 12: **end for**
- 13: Merge disturbed superpixels to their neighbors by connected components.

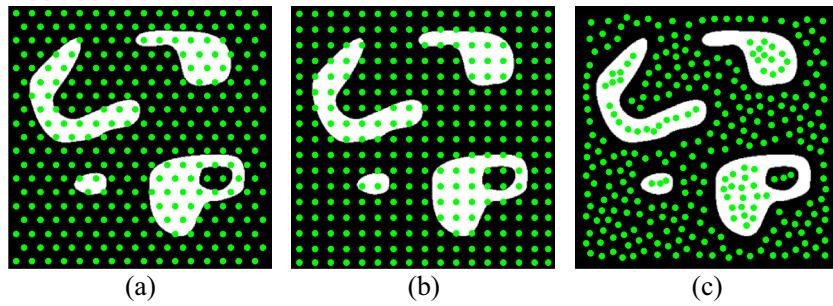
### 3.2 The Initialization-Center

In traditional superpixel approaches, the initialization of superpixel centers is commonly performed through uniform sampling over a grid structure. They neglect region-boundary constraints. To leverage the advantages of homogeneous region-level context, AMR-WT is employed to obtain regional segmentation that serves as prior constraints

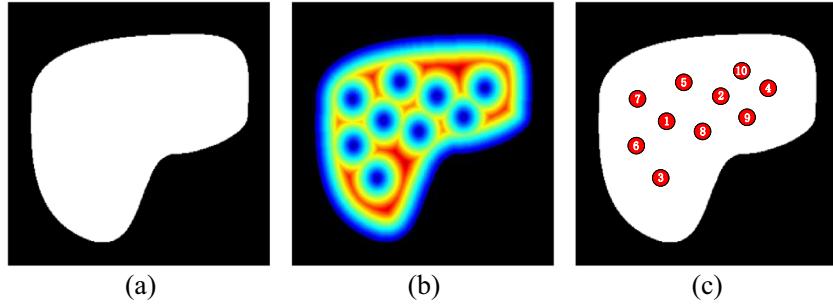
in RCFCMS. Figure 8 compares the initialization of superpixel centers. As observed, grid-based initialization fails to effectively place centers to adequately cover object regions, leading to suboptimal superpixel generation. The centers initialized by our strategy, as shown in Fig. 8, are not only well-distributed across all object regions but also located at a sufficient distance from the boundaries. This lays the foundation for recognizing complex objects by iteratively updating the superpixel centers. Hence, effectively initializing superpixel centers within each sub-region presents a challenging issue.

We expand the idea of MaskSLIC [50] to generate superpixel centers with respect to the shape of the different subregions. Its basis is the assumption that the superpixel centers are spaced at the maximum distance from the boundaries and that they are at the maximum distance from any other superpixel centers. For RCFCMS,  $R_l$  represents the  $l$ th region in the AMR-WT result, which can be regarded as the  $l$ th mask.  $B_l$  denotes the label set of background (non-mask) corresponding to  $R_l$ .  $P_l$  denotes the label set of initialized superpixel points. Assume  $G_l = B_l \cup P_l$ , and initially  $P_l = \phi$ , place the superpixel point  $p^*$  in region  $R_l$ , which is farthest from  $G_l$ , using the following equation:

$$p^* = \operatorname{argmax} \sqrt{(p_i - p_j)^2 + (q_i - q_j)^2}, \quad (17)$$



**Fig. 8** The comparison of 300 initialization centers. **a** SLIC adopts the hexagonal grids. **b** Fuzzy SLIC adopts the square grids. **c** RCFCMS adopts the region constraints



**Fig. 9** The initialization of the superpixel centers inside a mask. **a** Irregular region mask. **b** The distance heatmap of superpixel centers. **c** The final superpixel centers and their placement order

where  $(p_i, q_i) \in R_l$ ,  $(p_j, q_j) \in G_l$ . After  $P_l$  is determined, implementing  $R_l = R_l \wedge p^*$ ,  $P_l = p^* \cup P_l$ ,  $\wedge$  represents set difference,  $\cup$  represents the set union, respectively. Repeat the above steps to find the positions set  $P_l$  in region  $R_l$ . Proceeding in this manner, initialize all superpixel centers across the different regions.

Figure 9 shows an illustrative example of initializing superpixel centers in a target region. It is clear that the placement process of superpixel centers follows the assumption mentioned above. For more details, please refer to [50].

### 3.3 The Post-processing

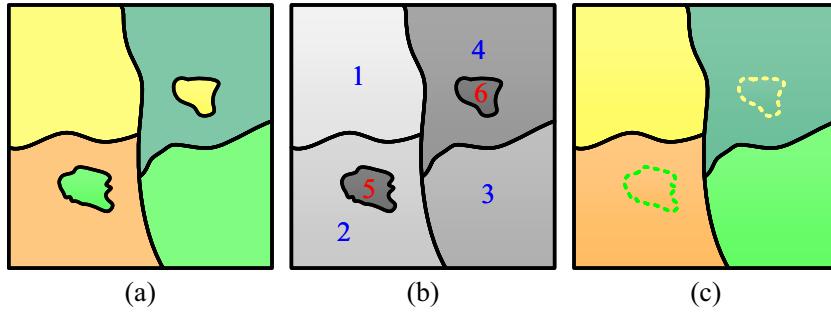
Like other clustering-based superpixel algorithms, the RCFCMS also adopts a post-processing strategy for connectivity enforcement. It is a remarkable benefit that the final procedure does not change the number of superpixels, only assigning ‘orphaned’ pixels to the nearest superpixel centers.

The post-processing employs a connected components algorithm to reassign the superpixel labels, then merge interference regions using the inclusion relation of area. The operation can eliminate small isolated regions and obtain the more pure superpixel result, which improves the segmentation accuracy, as shown in Fig. 10.

## 4 Experiments

In this Section, we conduct experiments on the Berkeley segmentation dataset (BSDS500) [51], and the Stanford background dataset (SBD) [52]. Simultaneously, we also analyze the potential effectiveness of the proposed RCFCMS on both the Skin Melanoma (SM) dataset [53] and the Microsoft Research Cambridge (MSRC) dataset [54]. Table 1 provides detailed information about the different datasets. It is important to note that MSRC did not define the division of training and testing sets on the official website; however, this can be determined based on the specific context. Additionally, the proposed RCFCMS is an unsupervised clustering process, independent of training set, and achieves the superpixels solely through iterative strategy.

To demonstrate the superiority of RCFCMS, we compare it against the following state-of-the-art superpixel algorithms, namely, ERS [14], SLIC [26], GMMSP [28], SH [23], FS [42], WSGL [24], SCAC [25], and Fuzzy SLIC [43], by presenting both visual and quantitative comparisons. These algorithms employ implementations either publicly available or provided by the authors’ source code, while using their suggested parameters. Furthermore, except for FS, all other comparative algorithms have been validated on the BSDS500. Some of them were also analyzed on the MRSC and SBD. Meanwhile, we utilize the SM with domain shifts



**Fig. 10** The illustration of the post-processing. **a** The example of superpixels with isolated regions. **b** Reassign the superpixel labels. **c** Merge interference regions

**Table 1** Descriptions of benchmark datasets, where “–” represents “not applicable”

Dataset	Samples	Size	Training set	Validation set	Testing set
BSDS500	500	321×481 481×321	200	100	200
SBD	715	Multi-sizes	238	–	477
SM	1279	Multi-sizes	900	–	379
MSRC	591	320×213 213×320	–	–	–

to demonstrate the generalization capability and potential attributes of RCFCMS.

For RCFCMS, three key parameters are considered, fuzzy exponent  $m$ , window radius  $w_r$ , compactness factor  $w$ . To better justify their selection, we evaluate all metrics on Fig. 11, as shown in Tables 2, 3, and 4. As demonstrated in these tables, the performance of RCFCMS is influenced by these parameters. When  $m = 2$ , RCFCMS performs optimally in four of the metrics. When  $w = 15$ , and  $w_r = 1$ , RCFCMS achieves the best results across two metrics. Based on these findings, we set  $m = 2$ ,  $w_r = 1$ , and  $w = 15$  in this paper.

Additionally, all experiments are conducted on a workstation equipped with a 13th Gen Intel(R) Core(TM) i9-13900K CPU operating at 3 GHz and 32 GB of RAM, using MATLAB R2023b. All comparative algorithms are unsupervised, without the need for training costs.

#### 4.1 Evaluation Metrics

To evaluate quantitatively the quality of superpixels obtained by different algorithms against the ground truth, we adopt widely used four standard metrics: achievable segmentation accuracy, under-segmentation error, F-measure, and compactness.

**Achievable Segmentation Accuracy (ASA)** measures the upper-bound fraction of ground truth segment labeled rightly

by the superpixels. We compute ASA as follows:

$$ASA = \frac{\sum_{t=1}^K \max_{i=1:M} |\mathcal{S}_t \cap \mathcal{G}_i|}{\sum_{i=1}^M |\mathcal{G}_i|}, \quad (18)$$

where  $K$  is the number of superpixels,  $M$  is the number of ground truth segment,  $\mathcal{S}_t$  represents the  $t$ th superpixel region, and  $\mathcal{G}_i$  represents the  $i$ th region from ground truth. In other words, the ASA is the ratio between the correctly classified superpixels and all pixels.

**Under-segmentation Error (UE)** measures the percentage of the superpixels leak across ground truth boundaries. The UE is expressed as follows:

$$UE = \frac{\sum_{t=1}^K \sum_{i=1}^M \min \{|\mathcal{S}_t \cap \mathcal{G}_i|, |\mathcal{S}_t - \mathcal{G}_i|\}}{\sum_{i=1}^M |\mathcal{G}_i|}, \quad (19)$$

UE and ASA exhibit a strong correlation. But  $UE$  is not equal to  $1-ASA$ . If a superpixel covers multiple ground truth segment borders, the value of  $UE$  will increase.

**F-Measure (FM)** measures the degree of matching between ground truth boundaries and superpixels. The FM can be calculated as follows:

$$FM = \frac{2(BR \times BP)}{BR + BP}, \quad (20)$$

where BR and BP stand for boundary recall and boundary precision, respectively. The BR computes the fraction of correctly recalled ground truth boundaries by the superpixel boundaries. The BP computes the fraction of correctly recognized ground truth boundaries by the detected boundaries. The FM effectively balances recall and precision at decision boundaries, offering greater intuitiveness than the BP-BR curve, especially in noisy environments.

**Compactness (CO)** measures the degree of proximity between each superpixel and corresponding circle. The CO is defined as follows:

$$CO = \frac{1}{N} \frac{\sum_{t=1}^K |\mathcal{S}_t| 4\pi A(\mathcal{S}_t)}{P(\mathcal{S}_t)}, \quad (21)$$



**Fig. 11** The visual comparison of 300 superpixels generated by different algorithms on BSDS500. The upper-left region of each image displays superpixel results without noise interference, while the lower-right region of each image displays superpixel results with noise interference

**Table 2** The impact of fuzzy exponent  $m$  on RCFCMS, with other parameters fixed at their final selected values. The best values are in bold

$m$	UE↓	ASA ↑	FM ↑	CO↑
1.5	0.0903	0.9543	0.3755	0.4499
2	<b>0.0846</b>	<b>0.9572</b>	<b>0.3873</b>	<b>0.4680</b>
2.5	0.1008	0.9488	0.3855	0.4550
3	0.1102	0.9437	0.3830	0.4334

**Table 3** The impact of window radius  $w_r$  on RCFCMS, with other parameters fixed at their final selected values. The best values are in bold

$w_r$	UE↓	ASA ↑	FM ↑	CO↑
1	<b>0.0846</b>	<b>0.9572</b>	0.3873	0.4680
2	0.0874	0.9558	0.3973	0.5451
3	0.0858	0.9566	<b>0.4088</b>	0.5682
4	0.0860	0.9565	0.4076	<b>0.5829</b>

**Table 4** The impact of compactness factor  $w$  on RCFCMS, with other parameters fixed at their final selected values. The best values are in bold

$w$	UE↓	ASA ↑	FM ↑	CO↑
10	0.0982	0.9500	0.3530	0.3354
15	<b>0.0846</b>	<b>0.9572</b>	0.3873	0.4680
20	0.0874	0.9558	<b>0.4016</b>	0.5451
25	0.0883	0.9553	0.3973	<b>0.5852</b>

where  $N$  stands for the number of pixels,  $|\mathcal{S}_t|$  represents the number of pixels in region  $\mathcal{S}_t$ ,  $A(\mathcal{S}_t)$  denotes the area of region  $\mathcal{S}_t$ , and  $P(\mathcal{S}_t)$  denotes the perimeter corresponding to the area of a circle with the same area as  $A(\mathcal{S}_t)$ .

The four metrics are plotted against the number of different superpixels. As more superpixels adhere to the object boundaries, the value of UE decreases, while the values of ASA and FM increase. Simultaneously, the superpixels tend more toward a circle, the higher the value of CO.

#### 4.2 Performance Comparison

Most existing superpixel algorithms work well on clean or noise-free images, they are sensitive to noise due to ignoring the importance of spatial information. The proposed RCFCMS considers local similarity and priori boundary, aiming to guarantee noise insensitiveness and promote detail preservation. To further verify the performances of proposed RCFCMS and state-of-the-art superpixel algorithms with default parameters on noisy images, we corrupt the BSDS500 by Gaussian noise (30%) and Salt & Pepper (SP) noise (10%). The mixed noise is utilized not only to simulate potential

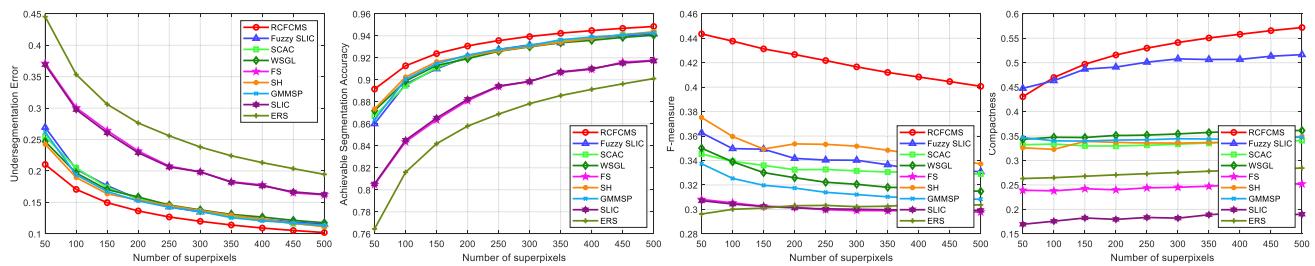
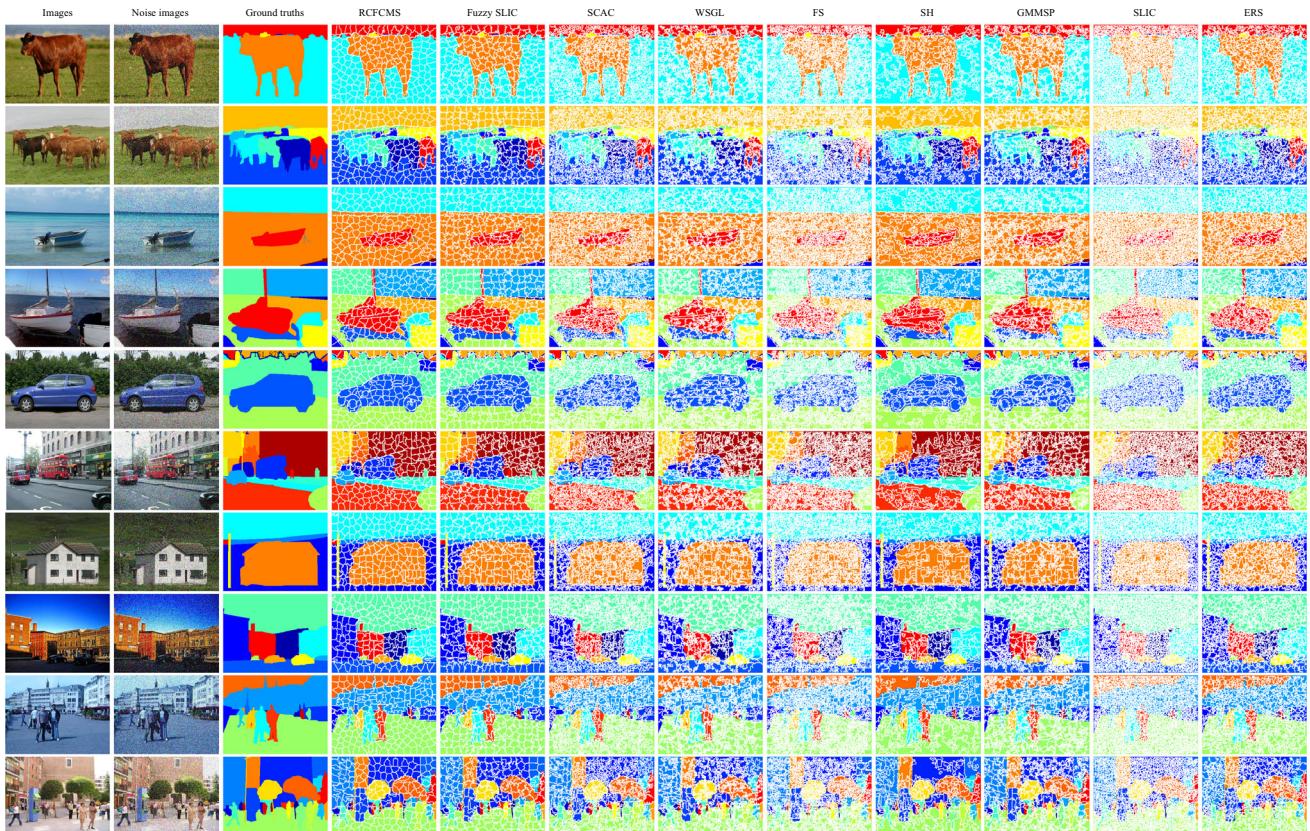
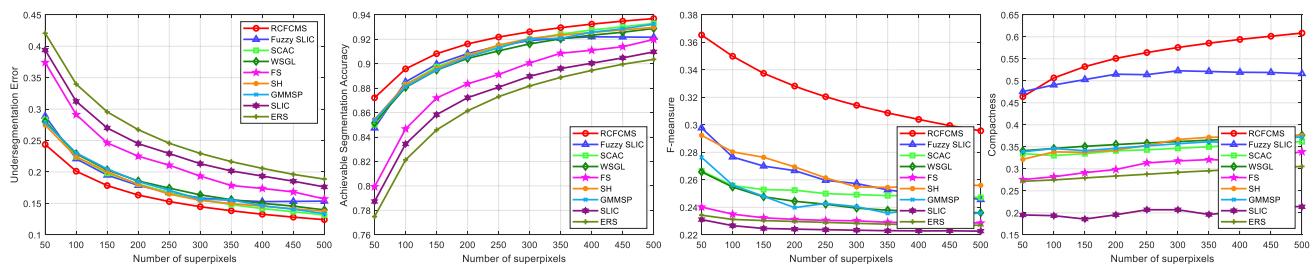
interferences in test images but also to evaluate the generalization capability of test algorithms. The superpixel results of different algorithms, both on clean and noisy images, are depicted in Fig. 11.

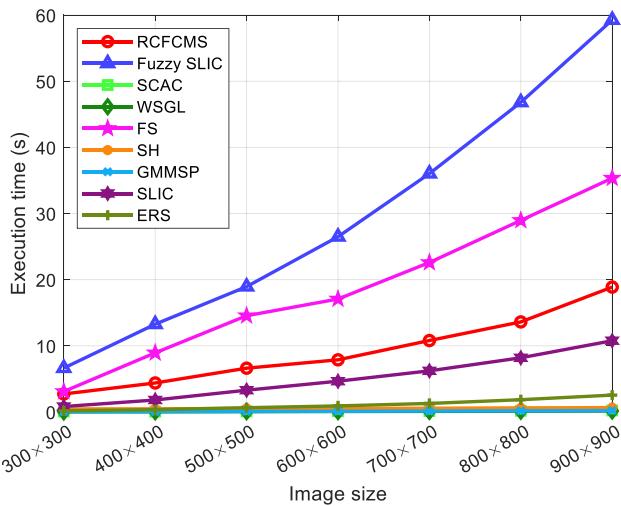
As illustrated in Fig. 11, both SLIC and FS lack sufficient robustness to mixed noise due to their disregard for the spatial information of the image. The ERS, GMMSP, SH, WSGL, and SCAC are, respectively, affected by the noise to varying degrees, especially in the regularity of superpixel shapes. The fuzzy SLIC combines neighborhood membership filtering and onion peeling optimization to improve the performance of superpixels. It can obtain good compactness, but its results are not entirely satisfactory. By comparing the red grid curves of the upper-left and lower-right parts in each image, it is evident that the proposed RCFCMS produces more compact and regular superpixels than the other algorithms. This is because the RCFCMS utilizes region constraints obtained by AMR-WT [19] to prevent boundary crossing and incorporates spatial constraints based on local similarity to suppress the influence of noise.

Figure 12 illustrates comprehensively the UE, ASA, FM, and CO quantitative results of all tested algorithms on BSDS500 with mixed noise. It is clearly indicated that ERS, SLIC, and FS show the worst performance in terms of boundary adherence, and their compactness is also concerning. The GMMSP, SH, WSGL, SCAC, and fuzzy SLIC have similar indexes of UE and ASA. In the FM, their performance shows a stepwise pattern, with SH and fuzzy SLIC achieving better scores. Focusing on CO, it is evident that GMMSP, SH, WSGL, and SCAC yield similar results. Since fuzzy SLIC considers local spatial information, its CO stands out more prominently by sacrificing FM. Figure 12 indicates that in noisy environments, the proposed RCFCMS not only achieves optimal boundary adherence ( $UE = 0.102$ ,  $ASA = 0.948$ , and  $FM = 0.401$ , when  $K = 500$ ), but also maintains perceptual satisfaction ( $CO = 0.572$ , when  $K = 500$ ).

To verify the stability of the RCFCMS, similar experiments are conducted on SBD. The SBD contains more semantic target information compared to BSDS500. Therefore, it can provide better feedback on the performance of the test algorithm. The visual comparison of the nine superpixel algorithms discussed above is presented in Fig. 13. We can clearly see that the proposed RCFCMS achieves superior adherence to object edges in weak boundaries, while maintaining homogeneity in complex texture. The fuzzy SLIC obtains the second-best visual effect. Although other algorithms can fit target contour, due to noise interference, their visual results of chaotic superpixel grid are suboptimal.

We further conduct experiments to evaluate four metrics, namely, UE, ASA, FM, and CO, using varying numbers of superpixels with state-of-the-art superpixel algorithms, as illustrated in Fig. 14. According to Fig. 14, it can be seen that ERS, SLIC, and FS exhibit poor values of UE, ASA, FM,

**Fig. 12** The quantitative comparison of different algorithms on BSDS500 corrupted by mixed noise**Fig. 13** The visual comparison of 200 superpixels generated by different algorithms on SBD with mixed noise**Fig. 14** The quantitative comparison of different algorithms on SBD corrupted by mixed noise



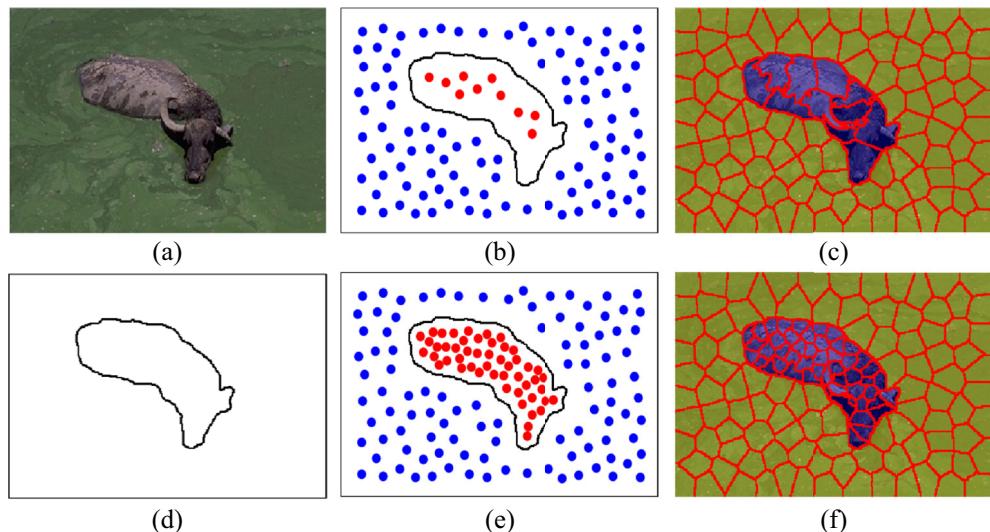
**Fig. 15** Average execution time of all algorithms with  $K = 300$  on testing set from BSDS500, where each image is sampled at progressively higher resolutions, i.e.,  $300 \times 300, \dots, 900 \times 900$

and CO. Except for mentioned three algorithms, other algorithms obtain similar metrics of UE and ASA. For FM, the SH based on edge detectors and fuzzy SLIC based on local information have comparable performance. However, fuzzy SLIC obtains the second-best values for CO because of membership constraints. Clearly, the proposed RCFCMS achieves the best performance under varying numbers of superpixels, when  $K = 500$ , with the metrics reaching  $UE = 0.124$ ,  $ASA = 0.935$ ,  $FM = 0.296$ , and  $CO = 0.608$ . This is mainly attributed to the region constraints and spatial filtering of RCFCMS.

### 4.3 Complexity Analysis

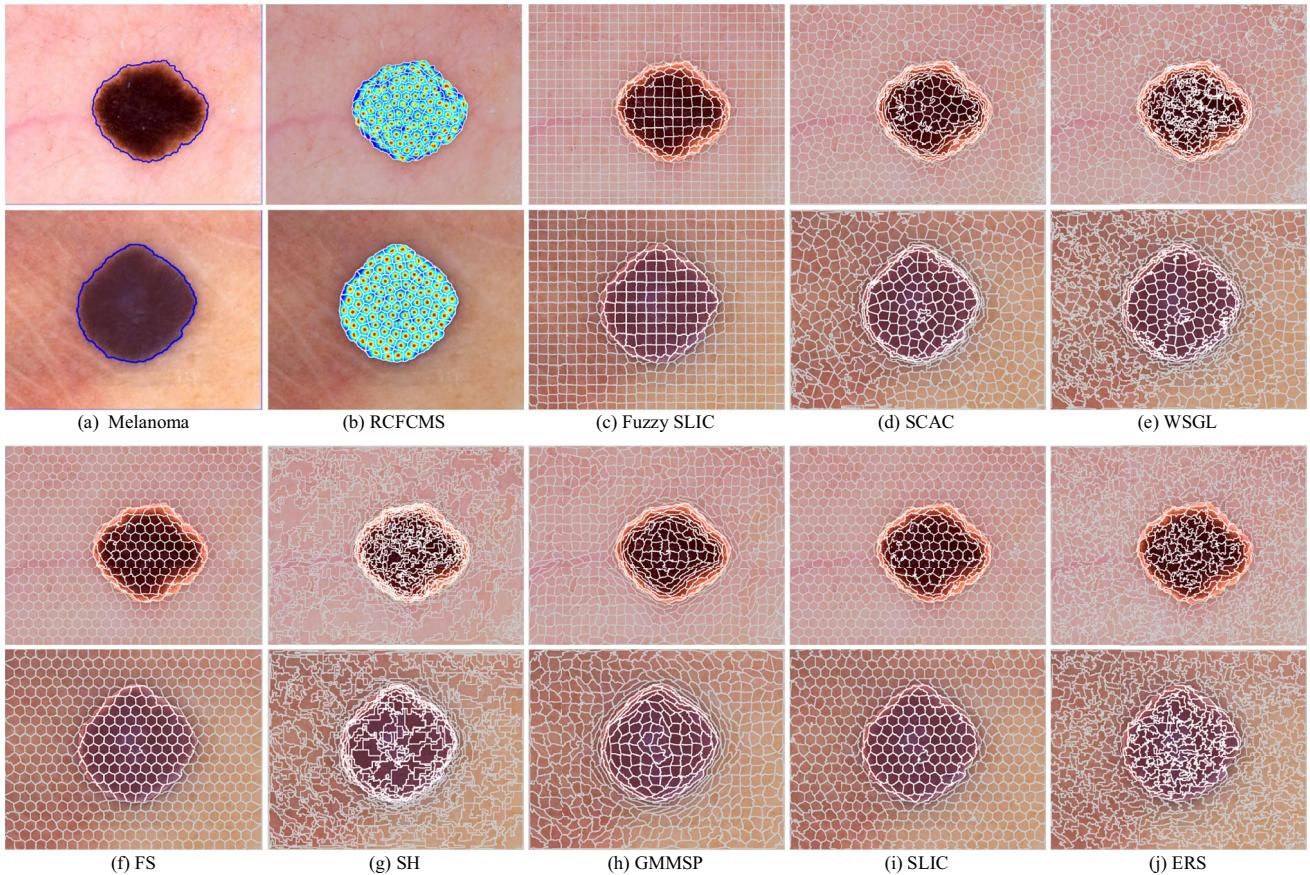
The computational complexity is critical for estimating the efficiency of superpixel algorithms. Suppose an image includes  $N$  pixels, and it is expected to generate  $K$  superpixels,  $2S \times 2S$  denotes a smaller search region, and  $S$  is the size of superpixels. Due to  $K \times S = N$ , therefore the computational complexity of most cluster-based superpixel algorithms is  $O(N)$ . This is explicitly stated in algorithms such as SLIC, GMMSP, FS, and Fuzzy SLIC. According to [23], we can get that ERS and SH have computational complexities of  $O(N)$  and  $O(1)$ , respectively. For WSGL and SCAC, the authors do not explicitly state computational time complexities. However, since WSGL uses a gradient image and the flooding process to identify superpixels, and SCAC employs a strategy similar to that of SLIC for superpixel generation, their computational complexities are also approximately  $O(N)$ .

Since the AMR-WT and post-processing have low computational complexity, the computational complexity of RCFCMS mainly lies in the initialization of superpixel centers utilizing maximum distance and the fuzzy clustering process updating of  $\mathbf{U}$ ,  $\mathbf{C}$ ,  $\mathbf{V}$ , and  $\mathbf{H}$ . The computational complexity of the first part is  $\sum_{l=1}^L O(n_l \times k_l^2)$ , while the computational complexity of the other part is  $\sum_{l=1}^L O(n_l \times k_l)$ . Therefore, the overall computational complexity of our algorithm is  $\sum_{l=1}^L O(n_l \times (k_l^2 + k_l))$ . Since  $\sum_{l=1}^L k_l = K$ , and  $\sum_{l=1}^L n_l = N$ , we can infer that  $\sum_{l=1}^L O(n_l \times k_l (k_l + 1)) < O(N \times K (K + 1))$ . For RCFCMS, a local search strategy is employed to avoid redundant distance calculations. Addi-

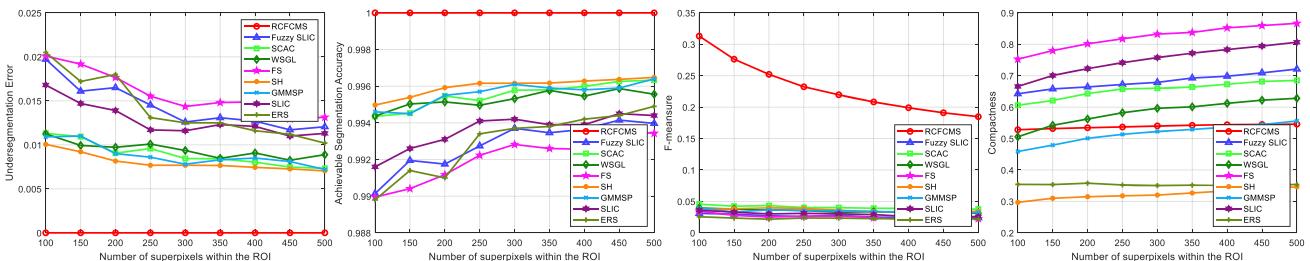


**Fig. 16** The potential attributes of RCFCMS. **a** Test image. **d** Boundaries of ground truth. **b** Initialize superpixel centers, where algae includes 100 points, and buffalo consists of 10 points. **c** Result of

RCFCMS. **e** Initialize superpixel centers, where algae includes 100 points, and buffalo consists of 10 points. **f** Result of RCFCMS



**Fig. 17** The visual comparison of superpixels generated by different algorithms on SM images



**Fig. 18** The quantitative comparison of different algorithms on Fig. 17

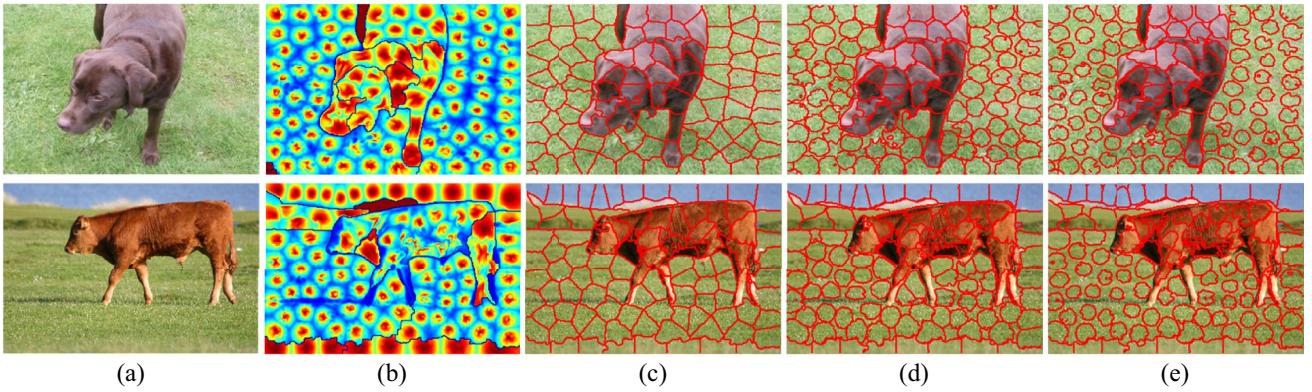
tionally, in practice, a pixel may fall within the search regions of multiple centers. Thus,  $1 \ll K \ll N$ , computational complexity of RCFCMS is also approximately  $O(N)$ . It is evident that the initialization step, which involves a constrained region with larger receptive field, is time-consuming, whereas the clustering process incurs relatively low computational cost.

To further validate the above analysis, we compare the execution times of all algorithms across different images, as shown in Fig. 15. Due to differences in programming styles, the execution times exhibit some variations, despite hav-

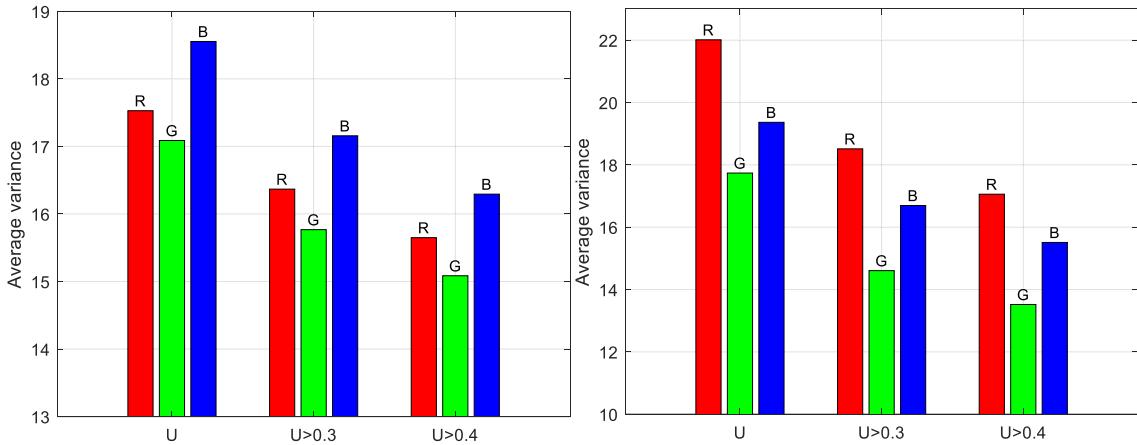
ing similar complexities. The execution time of RCFCMS is slightly higher than that of SLIC and ERS, but lower than that of Fuzzy SLIC and FS. The execution times of the other comparative algorithms are relatively small. The proposed RCFCMS still has significant room to improve the execution performance.

#### 4.4 More Discussions

The proposed RCFCMS not only takes into account spatial information but also incorporates region constraints.



**Fig. 19** The process of generating sparse superpixels using the RCFCMS. **a** The test images. **b** The highest membership  $U$  heatmap. **c** The superpixel results with  $U$ . **d** The superpixel results with  $U > 0.3$ . **e** The superpixel results with  $U > 0.4$



**Fig. 20** The average pixel variance of different enclosed areas in R, G, and B channels. The results from left to right show the cow and dog from MSRC [54], respectively

This enhances the robustness and flexibility of RCFCMS. As shown in Fig. 16, a potential attribute of RCFCMS is demonstrated, indicating that the spatial information in the RCFCMS can mitigate interference from algal texture, thereby obtaining relatively regular superpixels. When the buffalo is identified as a focal target, the region constraints can flexibly adjust the number of superpixels, thereby increasing the emphasis on the buffalo. Compared to the situation where seed points are globally allocated in traditional superpixel algorithms, the RCFCMS algorithm is more practical.

We use data from the ISIC 2018 challenge [53] to further demonstrate the superiority of RCFCMS. Figure 17a presents the SM images along with their corresponding meaningful masks delineated by blue lines. We set the number of superpixels within masks to 100. Using the area ratio, we can infer the total number of superpixels. Figure 17b-j shows the visual comparison of different algorithms. It is evident that only RCFCMS has the capability to independently calculate superpixels within target masks. This leads to faster processing of RCFCMS for region constraints.

Similarly, we evaluate the performance of different algorithms on SM images using four metrics: UE, ASA, FM, and CO, as shown in Fig. 18. Since RCFCMS incorporates constraint of target mask, it achieves the best results in terms of UE and ASA. In contrast, the UE and ASA of other algorithms are influenced to varying degrees by pixels outside the region of interest (ROI). Furthermore, RCFCMS better distributes the initialized centers within the target region, resulting in a higher FM, albeit at the cost of CO. In summary, compared to other algorithms, RCFCMS generates more meaningful superpixels while maintaining superior label consistency.

Unlike other superpixel algorithms, the proposed RCFCMS can generate sparse superpixels by controlling the highest membership. Figure 19 displays 100 conventional superpixels and the corresponding sparse superpixels generated by RCFCMS on MSRC. Comparing Fig. 19c–e, it can be concluded that RCFCMS can divide conventional superpixels into sparse superpixels, consisting of two parts, namely, pure superpixels and undetermined pixels. For the pure superpixels part, each pixel is assigned to definitely label. Indeed,

the smaller the area of superpixels, the higher the purity of the superpixels. For the undetermined pixels part, each pixel is allocated to ambiguous labels. Therefore, RCFCMS can capture the homogeneity properties of the image. The average pixel variance of each enclosed area in different channels also supports the abovementioned, as shown in Fig. 20.

## 5 Conclusion

In the paper, we develop a novel and effective RCFCMS algorithm that can be applied to various computer vision tasks. The promising performance of RCFCMS can be primarily attributed to its initialization of superpixel centers, incorporation of spatial information, and integration of region constraints. The groundbreaking initialization of superpixel centers efficiently places primary centers, benefiting capture target features. The innovative incorporation of spatial information effectively suppresses noise interference, ensuring the generation of regular superpixels. The pioneering integration of region constraints enforces boundary adherence, avoiding spanning multiple targets. The experimental results demonstrate that the proposed RCFCMS outperforms state-of-the-art superpixel algorithms. When the number of superpixels is set to 500, the performance of the RCFCMS is further validated by the four metrics: UE = 0.102, ASA = 0.948, FM = 0.401, and CO = 0.572 on BSDS500, and UE = 0.124, ASA = 0.935, FM = 0.296, and CO = 0.608 on SBD.

Although the potential analysis of RCFCMS validates its effectiveness, there exist some challenges worth discussing: First, RCFCMS relies on the constrained regions obtained by AMR-WT, limiting its flexibility. Second, the initialization of superpixel centers is performed sequentially, leading to high time consumption. Third, RCFCMS utilizes color and coordinate differences to generate superpixels, overlooking contributions of other features.

In future work, we plan to develop a two-stage superpixel generation framework. Employing deep learning architectures, such as Convolutional Neural Networks, Transformers, and Mamba, to extract richer image features, and then adopting an improved FCM to generate high-quality superpixels. This framework not only expands the theoretical foundation of FCM but also extends its practical applicability. Realizing them needs more research efforts.

**Funding** This work was supported in part by the National Natural Science Foundation of China (Program No. 62366029), in part by Gansu Province Youth Science and Technology Fund (Program No. 23JRRRA855), in part by Key Laboratory of Big Data & Artificial Intelligence in Transportation (Beijing Jiaotong University), Ministry of Education (Program No. BATLAB202302), in part by Key Research and Development Project of Lanzhou Jiaotong University (Program No. ZDYF2304), in part by Young Scholars Science Foundation of Lanzhou Jiaotong University (Program No. 2023006).

**Data Availability** Data are available in a publicly accessible repository that are cited in the article.

## Declarations

**Conflict of interest** The authors declare no Conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Xu, S., Wei, S., Ruan, T., Liao, L.: Learning invariant inter-pixel correlations for superpixel generation. In: AAAI Conference on Artificial Intelligence, pp. 6351–6359 (2024)
2. Kim, H., Oh, M., Hwang, S., Kwak, S., Ok, J.: Adaptive superpixel for active learning in semantic segmentation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 943–953 (2023)
3. Fang, Y., Zeng, Y., Jiang, W., Zhu, H., Yan, J.: Superpixel-based quality assessment of multi-exposure image fusion for both static and dynamic scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 2526–2537 (2021). <https://doi.org/10.1109/TIP.2021.3053465>
4. Yu, Q., Jie, M., Jing, X.: Superpixel-wise contrast exploration for salient object detection. *Knowl. Syst.* (2024). <https://doi.org/10.1016/j.knosys.2024.111617>
5. Li, Y., Liu, Y., Zhu, J., Ma, S., Niu, Z., Guo, R.: Spatiotemporal road scene reconstruction using superpixel-based Markov random field. *Inf. Sci.* (2020). <https://doi.org/10.1016/j.ins.2019.08.038>
6. Liu, Y.-J., Yu, M., Li, B.-J., He, Y.: Intrinsic manifold SLIC: a simple and efficient method for computing content-sensitive superpixels. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(3), 653–666 (2018). <https://doi.org/10.1109/TPAMI.2017.2686857>
7. Ng, T.C., Choy, S.K., Lam, S.Y., Yu, K.W.: Fuzzy superpixel-based image segmentation. *Pattern Recognit.* **134**, 109045 (2023). <https://doi.org/10.1016/J.PATCOG.2022.109045>
8. Xu, S., Wei, S., Ruan, T., Zhao, Y.: ESNet: an efficient framework for superpixel segmentation. *IEEE Trans. Circ. Syst. Vid. Technol.* **34**(7), 5389–5399 (2024). <https://doi.org/10.1109/TCSVT.2023.3347402>
9. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000). <https://doi.org/10.1109/34.868688>
10. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **59**(2), 167–181 (2004). <https://doi.org/10.1023/B:VISI.0000022288.19776.77>
11. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002). <https://doi.org/10.1109/34.1000236>
12. Tao, W., Jin, H., Zhang, Y.: Color image segmentation based on mean shift and normalized cuts. *IEEE Trans. Syst. Man Cybern. B* **37**(5), 1382–1389 (2007). <https://doi.org/10.1109/TSMCB.2007.902249>

13. Moore, A.P., Prince, S.J.D., Warrell, J., Mohammed, U., Jones, G.: Superpixel lattices. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
14. Liu, M.-Y., Tuzel, O., Ramalingam, S., Chellappa, R.: Entropy rate superpixel segmentations. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2097–2104 (2011)
15. Li, Z., Chen, J.: Superpixel segmentation using linear spectral clustering. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1356–1363 (2015)
16. Shen, J., Du, Y., Wang, W., Li, X.: Lazy random walks for superpixel segmentation. *IEEE Trans. Image Process.* **23**(4), 1451–1462 (2014). <https://doi.org/10.1109/TIP.2014.2302892>
17. Kang, X., Zhu, L., Ming, A.: Dynamic random walk for superpixel segmentation. *IEEE Trans. Image Process.* **29**, 3871–3884 (2020). <https://doi.org/10.1109/TIP.2020.2967583>
18. Kavitha, K.J., Shan, P.B.: Medical image watermarking based on novel encoding for EHR and fusion based morphological watershed segmentation algorithm for medical images. *Multimed. Tools Appl.* **83**(9), 25163–25190 (2024). <https://doi.org/10.1007/s11042-023-16490-8>
19. Lei, T., Jia, X., Liu, T., Liu, S., Meng, H., et al.: Adaptive morphological reconstruction for seeded image segmentation. *IEEE Trans. Image Process.* **28**(11), 5510–5523 (2019). <https://doi.org/10.1109/TIP.2019.2920514>
20. Levinstein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K.: Turbopixels: fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(12), 2290–2297 (2009). <https://doi.org/10.1109/TPAMI.2009.96>
21. Machairas, V., Faessel, M., Cárdenas-Peña, D., Chabardes, T., Walter, T., Decencière, E.: Waterpixels. *IEEE Trans. Image Process.* **24**(11), 3707–3716 (2015). <https://doi.org/10.1109/TIP.2015.2451011>
22. Dollar, P., Zitnick, C.L.: Structured forests for fast edge detection. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1841–1848 (2013)
23. Wei, X., Yang, Q., Gong, Y., Ahuja, N., Yang, M.-H.: Superpixel hierarchy. *IEEE Trans. Image Process.* **27**(10), 4838–4849 (2018). <https://doi.org/10.1109/TIP.2018.2836300>
24. Yuan, Y., Zhu, Z., Yu, H., Zhang, W.: Watershed-based superpixels with global and local boundary marching. *IEEE Trans. Image Process.* **29**, 7375–7388 (2020). <https://doi.org/10.1109/TIP.2020.3002078>
25. Yuan, Y., Zhang, W., Yu, H., Zhu, Z.: Superpixels with content-adaptive criteria. *IEEE Trans. Image Process.* **30**, 7702–7716 (2021). <https://doi.org/10.1109/TIP.2021.3108403>
26. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012). <https://doi.org/10.1109/TPAMI.2012.120>
27. Shen, J., Hao, X., Liang, Z., Liu, Y., Wang, W., Shao, L.: Real-time superpixel segmentation by DBSCAN clustering algorithm. *IEEE Trans. Image Process.* **25**(12), 5933–5942 (2016). <https://doi.org/10.1109/TIP.2016.2616302>
28. Ban, Z., Liu, J., Cao, L.: Superpixel segmentation using gaussian mixture model. *IEEE Trans. Image Process.* **27**(8), 4105–4117 (2018). <https://doi.org/10.1109/TIP.2018.2836306>
29. Abd El-Hafeez, T.: A new system for extracting and detecting skin color regions from pdf documents. *Int. J. Comput. Sci. Eng.* **9**(2), 2838–2846 (2010)
30. Ali, A.A., El-Hafeez, T.A., Mohany, Y.K.: An accurate system for face detection and recognition. *J. Adv. Math. Comput. Sci.* **33**(3), 1–19 (2019). <https://doi.org/10.9734/JAMCS/2019/v33i330178>
31. El-Sayed, M.A., Hafeez, T.A.-E.: New edge detection technique based on the Shannon entropy in gray level images (2012). [arXiv:1211.2502](https://arxiv.org/abs/1211.2502)
32. Saabia, A.A.-B., El-Hafeez, T., Zaki, A.M.: Face recognition based on grey wolf optimization for feature selection. In: International Conference on Advanced Intelligent Systems and Informatics 2018 (AISI 2018), pp. 273–283 (2019). [https://doi.org/10.1007/978-3-319-99010-1\\_25](https://doi.org/10.1007/978-3-319-99010-1_25)
33. Mostafa, G., Mahmoud, H., Abd El-Hafeez, T., ElAraby, M.E.: Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms. *J. Big Data* **11**(1), 88 (2024). <https://doi.org/10.1186/s40537-024-00944-3>
34. Jampani, V., Sun, D., Liu, M.-Y., Yang, M.-H., Kautz, J.: Superpixel sampling networks. In: IEEE/CVF European Conference on Computer Vision (ECCV), pp. 363–380 (2018)
35. Taha, M.E., Mostafa, T., El-Rahman, A., El-Hafeez, T.A.: A novel hybrid approach to masked face recognition using robust PCA and goa optimizer. *Sci. J. Damietta Fac. Sci.* **13**(3), 25–35 (2023)
36. Omar, A., Abd El-Hafeez, T.: Optimizing epileptic seizure recognition performance with feature scaling and dropout layers. *Neural Comput. Appl.* **36**(6), 2835–2852 (2024). <https://doi.org/10.1007/s00521-023-09204-6>
37. Eman, M., Mahmoud, T.M., Ibrahim, M.M., Abd El-Hafeez, T.: Innovative hybrid approach for masked face recognition using pre-trained mask detection and segmentation, robust pca, and knn classifier. *Sensors* **23**(15), 6727 (2023). <https://doi.org/10.3390/s23156727>
38. Elmessery, W.M., Maklakov, D.V., El-Messery, T.M., Baranenko, D.A., Gutiérrez, J., Shams, M.Y., El-Hafeez, T.A., Elsayed, S., Alhag, S.K., Moghanm, F.S., Mulyukin, M.A., Petrova, Y.Y., Elwaikel, A.E.: Semantic segmentation of microbial alterations based on SegFormer. *Front. Plant Sci.* (2024). <https://doi.org/10.3389/fpls.2024.1352935>
39. Yang, F., Sun, Q., Jin, H., Zhou, Z.: Superpixel segmentation with fully convolutional networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13961–13970 (2020)
40. Wang, Y., Wei, Y., Qian, X., Zhu, L., Yang, Y.: AINet: association implantation for superpixel segmentation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 7058–7067 (2021)
41. Krinidis, S., Chatzis, V.: A robust fuzzy local information c-means clustering algorithm. *IEEE Trans. Image Process.* **19**(5), 1328–1337 (2010). <https://doi.org/10.1109/TIP.2010.2040763>
42. Guo, Y., Jiao, L., Wang, S., Wang, S., Liu, F., Hua, W.: Fuzzy superpixels for polarimetric SAR images classification. *IEEE Trans. Fuzzy Syst.* **26**(5), 2846–2860 (2018). <https://doi.org/10.1109/TFUZZ.2018.2814591>
43. Wu, C., Zheng, J., Feng, Z., Zhang, H., Zhang, L., Cao, J., Yan, H.: Fuzzy SLIC: fuzzy simple linear iterative clustering. *IEEE Trans. Circ. Syst. Vid. Technol.* **31**(6), 2114–2124 (2021). <https://doi.org/10.1109/TCSVT.2020.3019109>
44. Wu, C., Zhao, J.: Robust joint learning of superpixel generation and superpixel-based image segmentation using fuzzy c-multiple-means clustering. *Signal Image Video Process.* **18**(3), 2345–2354 (2024). <https://doi.org/10.1007/s11760-023-02911-6>
45. Zhang, Y., Bai, X., Fan, R., Wang, Z.: Deviation-sparse fuzzy c-means with neighbor information constraint. *IEEE Trans. Fuzzy Syst.* **27**(1), 185–199 (2019). <https://doi.org/10.1109/TFUZZ.2018.2883033>
46. Zhao, F., Xiao, Z., Liu, H., Tang, Z., Fan, J.: A knee point driven kriging-assisted multi-objective robust fuzzy clustering algorithm for image segmentation. *Knowl. Syst.* (2023). <https://doi.org/10.1016/J.KNOSYS.2023.110522>
47. Lei, T., Jia, X., Xue, D., Wang, Q., Meng, H., Nandi, A.K.: Fuzzy student's t-distribution model based on richer spatial combination. *IEEE Trans. Fuzzy Syst.* **30**(8), 3023–3037 (2022). <https://doi.org/10.1109/TFUZZ.2021.3099560>

48. Yang, X., Zhu, M., Sun, B., Wang, Z., Nie, F.: Fuzzy c-multiple-means clustering for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **20**, 1–5 (2023). <https://doi.org/10.1109/LGRS.2023.3246633>
49. Wang, C., Zhou, M., Pedrycz, W., Li, Z.: Comparative study on noise-estimation-based fuzzy c-means clustering for image segmentation. *IEEE Trans. Cybern.* **54**(1), 241–253 (2024). <https://doi.org/10.1109/TCYB.2022.3217897>
50. Irving, B.: Maskslic: regional superpixel generation with application to local pathology characterisation in medical images (2016). [arXiv:arXiv:1606.09518](https://arxiv.org/abs/1606.09518)
51. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 898–916 (2011). <https://doi.org/10.1109/TPAMI.2010.161>
52. Gould, S., Fulton, R., Koller, D.: Decomposing a scene into geometric and semantically consistent regions. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1–8 (2009)
53. Codella, N.C.F., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A.: Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168–172 (2018)
54. Jamie, S., John, W., Carsten, R., Antonio, C.: Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: IEEE/CVF European Conference on Computer Vision (ECCV), pp. 1–15 (2006)



**Xiaohong Jia** has received his Ph.D. degree in Light Industrial Chemical Process Systems Engineering from Shaanxi University of Science and Technology, Xi'an, China, in 2021. He is currently a Lecturer with the School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China. His current research interests include image processing and pattern recognition.



**Yao Zhao** has received his B.S. degree from Fuzhou University, Fuzhou, China, in 1989 and M.E. degree from Southeast University, Nanjing, China, in 1992, both from the Radio Engineering Department, and Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), Beijing, China, in 1996. He became an Associate Professor at BJTU in 1998 and a full Professor in 2001. From 2001 to 2002, he worked as a Senior Research Fellow in the

Information and Communication Theory Group, Faculty of Informa-

tion Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is now the Director of the Institute of Information Science, BJTU. His research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding. He is now also leading several national research projects from the 973 Program, 863 Program, and National Science Foundation of China. He serves on the editorial boards of several international journals, including as an Area Editor of Signal Processing: Image Communication (Elsevier) and as an Associate Editor of Circuits, System & Signal Processing (Springer). He was named a Distinguished Young Scholar by the National Science Foundation of China in 2010.



**Bin Zhang** has received his Ph.D. degree in Optical Engineering from Harbin Institute of Technology, Harbin, China, in 2022. He is currently a Lecturer with the School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China. His current major interests are laser image recognition and laser transmission.



**Xuejun Zhang** has received his Ph.D. degree in Computer Science and Technology from Xi'an Jiaotong University, Xi'an, China, in 2016. He is currently a Professor with the School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China. He is a Senior Member of CCF and an ACM Member. His main research interests include data privacy and machine learning, network security, and edge computing.



**Guanghui Yan** has received his Ph.D. degree in Computer Science and Technology from Northwestern Polytechnical University, Xi'an, China, in 2009. He is currently a Professor with the School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China. His current research interests include artificial intelligence, pattern recognition, and machine learning.