Superpixel Segmentation Based on Anisotropic Diffusion Model for Object-Oriented Remote Sensing Image Classification

Xiaoli Li, Jinsong Chen, Longlong Zhao, Hongzhong Li, Jin Wang, Luyi Sun, Shanxin Guo, Pan Chen, and Xuemei Zhao

***Abstract*—Superpixel segmentation is an essential step of object-oriented remote sensing image classification; the accuracy of the superpixel segmentation boundary will directly affect the classification result. Most of the traditional superpixel segmentation algorithms rely on spectral similarity and spatial connectivity to construct superpixels. They cannot find the accurate boundary in the complex scenes, such as the spatial distribution of ground features being relatively broken, and large differences in the size and shape, especially long-thin shape and circular shape. Aiming at this problem, a superpixel segmentation algorithm based on an anisotropic diffusion model named ADS is proposed and applied to image classification. The anisotropic diffusion model originated in thermodynamics has excellent properties in which the diffusion is continuous and smooth and its diffusion speed depends on the medium, which provides convenience for smoothing homogeneous regions and establishing boundary constraints for different ground objects. With this advantage, the diffusion flux model is established to consider the influence of boundary factors and used to simulate the dissimilarity measure with boundary constraints between pixels and seed points by combining the traditional spectral and spatial distance. Then, the seed points of superpixel are optimized under the K-means framework. The effectiveness of the proposed algorithm is tested and verified with different spatial resolutions, such as Landsat 8 with 30 m, Sentinel-2 with 10 m, and SkySat with 0.5 m. A large number of experiments show that the proposed algorithm can better correct the superpixel boundary-fitting deviation problem in complex scenes and effectively promote the improvement of image classification accuracy.**

***Index Terms*—Anisotropic diffusion, diffusion flux, remote sensing image classification, superpixel segmentation.**

Manuscript received 13 May 2023; revised 29 July 2023 and 16 September 2023; accepted 8 October 2023. Date of publication 16 October 2023; date of current version 10 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42001286 and Grant 42001365, in part by the Fundamental Research Foundation of Shenzhen Technology and Innovation Council under Grant JCYJ20200109115637548, and Grant KCXFZ202002011006298. *(Corresponding author: Jinsong Chen.)*

Xiaoli Li, Jinsong Chen, Longlong Zhao, Hongzhong Li, Luyi Sun, Shanxin Guo, and Pan Chen are with the Center for Geospatial Information, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Shenzhen Engineering Laboratory of Ocean Environmental Big Data Analysis and Application, Shenzhen 518055, China (e-mail: xl.li2@siat.ac.cn; js.chen@siat.ac.cn; ll.zhao@siat.ac.cn; hz.li@siat.ac.cn; ly.sun@siat.ac.cn; sx.guo@siat.ac.cn; p.chen3@siat.ac.cn).

Jin Wang is with the School of Geography, South China Normal University, Guangzhou 510631, China (e-mail: egmontwj@gmail.com).

Xuemei Zhao is with the School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541000, China (e-mail: zhaoxm@guet.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3324770

I. Introduction

REMOTE sensing image interpretation is the key to intelligent Earth observation [1]. With the increase in the spatial resolution of remote sensing images, the detailed information of ground features is more abundant. In the meantime, the spatial distribution of ground features is more complex, the heterogeneity is stronger, and the boundaries are more uncertain [2], [3]. The traditional pixel-based image processing methods are extremely sensitive to noise, which will cause a "salt and pepper" phenomenon. To deeply mine the implicative knowledge to improve the processing effect, numerous new algorithms emerge endlessly.

In recent years, deep learning methods have occupied an essential position in the field of image classification [4], such as fully convolutional networks [5], U-Net [6], [7], SegNet [8], and DeepLab [9]. They enhance the receptive field of images through a series of operations, such as convolution and pooling, which can learn more deep information, such as spatial structure. However, they are often executed based on the regular grid structure of images, which is lacking in processing irregular structured data. Focus on this point, a series of graph neural networks have been extensively studied. Hong et al. [10] proposed minibatch graph convolutional networks (GCNs), which can train large-scale graph networks in a minibatch fashion. It deals with the problems of computational cost and prediction of out-of-sample data. Ding et al. [11] proposed multiscale receptive fields graph attention neural network to learn the local–global spatial context information, proposed multiadaptive receptive field-based graph neural framework to learn multiscale features [12], proposed unsupervised self-correlated learning smoothy enhanced locality preserving graph convolution embedding clustering to pay attention to the smoothy information and the nonlocal relationship [13], and proposed GCN with adaptive filters and aggregator fusion to filter noise and capture spatial node relationships from multiple adaptive aspects [14]. Although various deep learning methods have achieved good results, there are still some inevitable problems, such as high requirements for high-performance computing power, sample size, and quality, which limits its development [15]. To reduce computational complexity, superpixel segmentation, as a prestep in image processing, is widely studied [11], [14], [16].

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

The goal of superpixel segmentation is to gather the adjacent pixels with the same characteristics into subregions in image-domain space and accurately fit the boundaries of different ground features. At present, the superpixel segmentation method can be divided into three types [17], [18], [19], graph-based, gradient-based, and clustering-based. The graph-based methods express the image as an undirected graph. The edge weight connecting two nodes is used to measure the similarity between two pixels. Then, the image segmentation is equivalent to the graph partitioning. The representative algorithms are normalized cuts (NCut) [20], [21] and entropy rate superpixels (ERS) [22]. NCut [20] is an unbiased measure of disassociation between subgroups of a graph, which improves the original graph-based algorithms in noise sensitivity. ERS [22] formulates the superpixel segmentation problem as an optimization problem on graph topology based on the entropy rate of a random walk on the graph. It encourages the generation of superpixels with similar sizes and can better preserve feature boundaries. The graph-based methods strictly guarantee connectivity within the superpixels. However, most of them are with high model complexity [23]. The gradient-based methods evolve the active contours according to the image gradient. The representative algorithms are TurboPixels [24] and spatial-constrained watershed (SCoW) [25]. TurboPixels [24] is a geometric-flow-based algorithm; it evolves contours based on proximity-based boundary velocity and image-based boundary velocity. TurboPixels can respect local image boundaries and avoid undersegmentation through a compactness constraint. SCoW [25] introduces spatial constraint in the marker-controlled watershed to obtain compact and evenly distributed superpixels, which makes a balance between homogeneity and compactness. The excellent boundary-fitting characteristic of gradient-based methods is slightly inadequate in complex scenes because it is more dependent on the gradient information [26]. The clustering-based methods consider superpixel segmentation as a clustering problem with image spatial distance. The representative algorithms are simple linear iterative clustering (SLIC) [27] and linear spectral clustering (LSC) [28]. SLIC turns the image from RGB color space to LAB color space and measures the distance between pixels and seed points by normalized spatial distance and spectral distance. LSC maps each pixel to a point inside a high-dimensional space and achieves segmentation based on weighted K-means. The clustering-based methods can easily control compactness, and it is widely studied because of their simple principle and strong extensibility. A lot of improvement methods are proposed to promote the superpixel segmentation accuracy from the perspective of compactness and the goodness of fit [29], [30], [31]. However, different from natural image, remote sensing image is characterized by big data, various targets, complex spatial structure, rich feature information, and ambiguous boundary [32], [33], [34]. It is difficult to achieve the effective segmentation of various ground features at the unified scale when the essence of the model is isotropic. Suppose there are both large circular and small long-thin shapes ground features in the image. In that case, superpixels will cross the boundaries of the long-thin area when it focuses on processing the big circular areas. On the contrary, when it focuses on processing the long-thin areas, the number of superpixels will skyrocket, accompanied by a surge in computational complexity, and then lose the significance of object-oriented processing [35], [36], [37]. Although multiscale segmentation is used to deal with the problem of excessive superpixels, it will add additional new issues, such as the impact of the correctness of the merging criteria on the segmentation results. Thus, exploring flexible and anisotropic superpixel construction methods at fixed scales remains an important issue.

The classical anisotropic model is Perona–Malik (P-M). It originated in thermodynamics and developed in the image-filtering field [38]. The excellent properties of P-M in directional diffusion have been proven to have a protective effect on the boundaries between ground features [39], [40]. Therefore, it has also been attempted to be introduced into the image segmentation field. However, most of them obtain global segmentation results by first P-M filtering and then combining the existing segmentation methods. The details are easily lost in superpixel construction [41], [42], [43]. P-M is also used for postprocessing of existing superpixels to achieve image segmentation, where the boundary is a known parameter [44], [45]. In the current approach, the advantages of P-Ms boundary protection have not been fully reflected. To introduce the excellent properties of boundary protection into the superpixel segmentation field to automatically explore boundaries, the anisotropic diffusion model is incorporated into clustering-based methods, and the superpixel segmentation algorithm based on the anisotropic diffusion model named ADS is proposed in this article.

First, the initial seed points are regularly distributed in the image domain and then moved to the lowest position of the gradient in the neighborhood to avoid being at the boundary. Next, assuming that the seed points are the concentration emitter, the concentration is anisotropically diffused from the seed point to any point in space according to the concentration gradient and spectral gradient. The concentration gradient controls the diffusion direction, and the spectral gradient controls the diffusion speed. The diffusion speed modeled by the monotone-decreasing function of the spectral gradient will reach a minimum value greater than zero at the boundary between ground features, which is conducive to exploring the boundaries. To accurately describe the relationship between pixels and seed points, the diffusion flux, the spatial distance, and the spectral distance are united to model the dissimilarity that is with boundary constraints. Then, the superpixels are obtained under the K-means frame. Finally, to strictly limit the spatial continuity within the superpixel, a neighborhood block proximity criterion is designed to eliminate the isolated pixels and small blocks.

The main contributions of this article are summarized as follows.

1) A new dissimilarity measure is established by the proposed diffusion flux model, which can describe the degree of dissimilarity between pixels and seed points in a spatially anisotropic way and achieve the goal of automatically exploring the boundaries of objects in superpixel segmentation
2) The proposed superpixel segmentation algorithm improves the dependence problem of superpixel segmentation on scale parameters under fixed scale conditions,

making it more flexible to achieve the effective segmentation of different types of complex scenes.

The rest of this article is organized as follows. In Section II, the classical anisotropic diffusion model, P-M model, is reviewed. In Section III, the main idea and each component of the proposed algorithm are described, including the proposed diffusion flux model and the dissimilarity measure model with boundary constraints. In Section IV, the performance of the proposed algorithm is examined by superpixel segmentation and image classification experiments based on three spatial resolutions, such as Landsat 8 with 30 m, Sentinel-2 with 10 m, and SkySat with 0.5 m. Section V discusses the performance of the proposed method. Finally, Section VI concludes this article.

II. Related Work

The main idea of the diffusion model is originated from thermal conduction in the physical process. Thermal conduction describes the change of temperature with time. The change is mainly affected by temperature differences and medium.

Assuming that the image is expressed as $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}) = \{\boldsymbol{I}_i(x_i, y_i) : i = 1,\ldots,n\}$, where $\boldsymbol{I}_i$ is the multichannel spectral characteristics of pixel i, (x_i, y_i) is the position coordinates of pixel i, $(\boldsymbol{x},\boldsymbol{y})$ is the set of (x_i, y_i), and n is the number of pixels in the image domain $\boldsymbol{\Omega}$. The heat equation applied to the image filtering can be expressed as follows:

$$\begin{cases} \frac{\partial \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)}{\partial t} = \operatorname{div}(\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)) = \Delta \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t), \\ t \geq 0, (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{\Omega} \\ \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|_{t=0} = \boldsymbol{I}_0(\boldsymbol{x},\boldsymbol{y}) \end{cases} \tag{1}$$

where t represents the time, $\operatorname{div}(\cdot)$ is the divergence operator, ∇ is the gradient operator, Δ is the Laplace operator, and $\boldsymbol{I}_0$ is the original image at the initial time. The solution of (1) is the filtering result, and it is equivalent to the image convolution based on different scale Gaussian filters [46]

$$\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) = G_\sigma^t * \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},0) \Leftrightarrow \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) = G_\sigma^t * \boldsymbol{I}_0(\boldsymbol{x},\boldsymbol{y}) \tag{2}$$

where $*$ is the convolution operator, and $\sigma = \sqrt{2t}$, G_σ^t is the Gaussian distribution at t time with 0 as the mean and σ^2 as the variance.

According to (1) and (2), it shows that the traditional thermal equation applied to image filtering is an isotropic model. The pixels with the same spatial distance are considered to have the same smoothing effect, which can cause the boundary to be excessively smoothed and fuzzy.

To protect the boundary during filtering, P-M extended the linear thermal equation to the nonlinear thermal equation and proposed the famous anisotropic diffusion model, P-M model [47], [48]

$$\begin{cases} \frac{\partial \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)}{\partial t} = \operatorname{div}(c(|\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|)\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)), \\ t \geq 0, (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{\Omega} \\ \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|_{t=0} = \boldsymbol{I}_0(\boldsymbol{x},\boldsymbol{y}) \end{cases} \tag{3}$$

where $|\cdot|$ denotes the magnitude. $c(\cdot)$ is the diffusion coefficient used to control the diffusion speed. It expects that the diffusion speed will be larger in the smooth area and smaller in the steep site. There are two commonly used diffusion coefficients

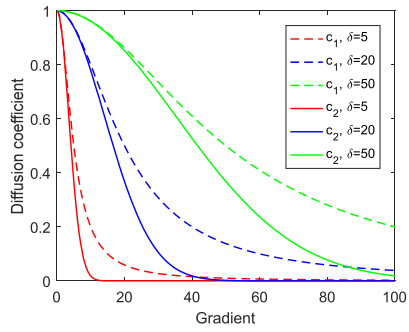

Fig. 1. Diffusion function.

[38], [49]

$$c_1(|\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|) = \frac{1}{1 + \left(\frac{|\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|}{\delta}\right)^2} \tag{4}$$

$$c_2(|\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|) = \exp\left[-\left(\frac{|\nabla \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|}{\delta}\right)^2\right] \tag{5}$$

where δ is the threshold of the gradient.

The curves of the diffusion coefficient under the assumption that the gradient range is 0–100 are shown in Fig. 1. It shows that the smaller the threshold, the faster the coefficient decreases. At the same threshold, c_1 changes more slowly than c_2. Generally, there is a high gradient at the boundary. When the gradient is far greater than the threshold, the pixel is smoothed with a minimal speed. Thus, the boundary can be protected.

P-M also gave the approximate solution of the discrete form of (3)

$$\begin{aligned} \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t+1) &= \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) \\ &+ \lambda \sum_f \left[c^f(|\nabla^f \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)|) \cdot \nabla^f \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)\right] \end{aligned} \tag{6}$$

where λ is the coefficient controlling the stable numerical solution, $0 \leqq \lambda \leqq 1/|f|$, f is the direction index, and $|f|$ represents the number of directions. Assuming that there are four directions, $f = \{(s,p) : (s,p) \in \{(0,+1),(-1,0),(0,-1),(+1,0)\}\}$. Then, $\nabla^f \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) = \boldsymbol{I}(\boldsymbol{x}+s,\boldsymbol{y}+p,t) - \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t)$, its details are as follows:

$$\begin{aligned} \nabla^{f_1} \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) &= \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}+1,t) - \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) \\ \nabla^{f_2} \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) &= \boldsymbol{I}(\boldsymbol{x}-1,\boldsymbol{y},t) - \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) \\ \nabla^{f_3} \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) &= \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}-1,t) - \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) \\ \nabla^{f_4} \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t) &= \boldsymbol{I}(\boldsymbol{x}+1,\boldsymbol{y},t) - \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},t). \end{aligned} \tag{7}$$

If $|f|=8$, $f=\{(s,p) : (s,p) \in \{(0,+1),(-1,+1),(-1,0),(-1,-1),(0,-1),(+1,-1),(+1,0),(+1,+1)\}\}$. The relation of index and direction is shown in Fig. 2.

III. Methodology

In this section, the proposed superpixel segmentation algorithm based on anisotropic diffusion model called ADS is

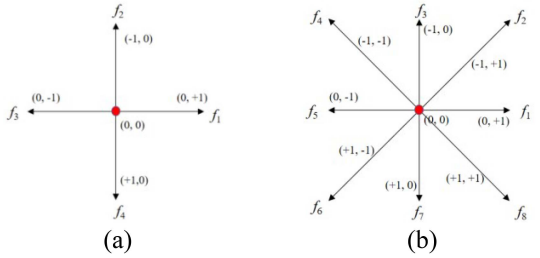

Fig. 2. Direction index. (a) $|f| = 4$. (b) $|f| = 8$.

illustrated. It changes the traditional idea of using the anisotropic diffusion model to filter the image first and then perform the segmentation but integrates it into the clustering process by the proposed diffusion flux model. Details are as follows.

A. Diffusion Flux

The anisotropic diffusion model applied to image filtering only establishes the relationship between adjacent points. In order to extend it from point to plane to describe the relationship between any pixels in the image domain and seed points, the concept of diffusion flux is proposed. Assuming that the seed points are concentration emitters, the concentration gathers at the seed points at the initial moment and diffuses to any pixels over time. The cumulative value of the diffusion amount of continuous diffusion over some time is the diffusion flux.

Given a set of seed points $\boldsymbol{S} = \{\boldsymbol{S}_j(a_j, b_j) : j = 1, \ldots, m\}$, where $\boldsymbol{S}_j$ is the multichannel spectral characteristics of seed point $j, (a_j, b_j)$ is the position coordinates of seed point j, and m is the number of seed points. For the diffusion flux $\boldsymbol{U^J}(\boldsymbol{x}, \boldsymbol{y}) = \{U^j(x_i, y_i) : i = 1, \ldots, n\}$, if and only if $(x_i, y_i) = (a_j, b_j)$, $U^j(x_i, y_i, t = 0) = 1$; otherwise, $U^j(x_i, y_i, t = 0) = 0$, where j is the dimension index of diffusion flux matrix, and $\boldsymbol{J}$ is the set of j. To accurately describe the diffusion process, the diffusion distance and direction are introduced into the diffusion flux model. As time goes by, the flux at any point in the image domain can be expressed as follows:

$$\begin{aligned}\boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t+1) &= \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t) \\ &+ \lambda \sum_f \left[\frac{1}{R} \cdot c^f(|\nabla^f \boldsymbol{I}(\boldsymbol{x}, \boldsymbol{y}, t)|) \right. \\ &\left. \cdot \nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t) \cdot \Psi[\nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t)] \right] \end{aligned} \tag{8}$$

where R is the diffusion distance, $R = 1$ when $(s, p) \in \{(0, +1), (-1, 0), (0, -1), (+1, 0)\}$, and $R = \sqrt{2}$ when $(s, p) \in \{(-1, +1), (-1, -1), (+1, -1), (+1, +1)\}$, the diffusion coefficient c is inherited from the image spectral field. The diffusion speed depends on the gradient threshold. According to the practice, the gradient image follows the Gamma distribution. It means that the big gradient occupies a small part of the image and it is probably the boundary gradient. If the sum of frequency exceeds the histogram threshold, its corresponding gradient value can be regarded as the gradient threshold. $\nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t)$ is the concentration gradient modeled by (7), and $\Psi[\nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t)]$ is the indicator function and it is used to control the diffusion directions. If and only if $\nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t) > 0$, $\Psi[\nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t)] = 1$; otherwise, $\Psi[\nabla^f \boldsymbol{U}^J(\boldsymbol{x}, \boldsymbol{y}, t)] = 0$.

Fig. 3 shows the concentration diffusion process with diffusion coefficient c_1 and $|f| = 8$, where Fig. 3(a) shows the index of pixel site, (b) is the gray value of pixels, (c) is the diffusion flux at the initial time (i.e., $t = 0$) by assuming f as the seed point, (d) is the diffusion state at $t = 0$; (e) is the diffusion result (i.e., diffusion flux at $t = 1$), (f) is the diffusion state of the representative pixel b at $t = 1$, which is close to the seed point but have significant difference in grays, (g) is the diffusion state of the representative pixel l at $t = 1$, which is far from the seed point but have similar gray, (h) is the second diffusion result (i.e., diffusion flux at $t = 2$), (i) is the diffusion flux at $t = 10$, and (j) is the diffusion flux at $t = 40$. The detailed calculation process is shown in Appendix A. The circles represent the pixels, and the red one is the seed point. The number within circles in Fig. 3(c)–(j) represents the concentration, the darker the color, the higher the concentration. The black arrow represents diffusing direction. c is the diffusion coefficient. R is the diffusion distance. It shows that the pixel with a different gray from the seed point, even if the spatial distance is short, its concentration diffusion is still well suppressed, is shown in pixel b in Fig. 3. The pixel with a similar gray from the seed point, even if the spatial distance is far, its concentration diffusion is still well protected, is shown in pixel l in Fig. 3. After several rounds of diffusion, the concentration of pixels belonging to the same class as the seed point will quickly approach 0.99, which is significantly separated from another class of 0.5, as shown in Fig. 3(j). Fig. 3 effectively illustrates the anisotropy of diffusion flux.

Fig. 4 shows the difference between isotropic diffusion and anisotropic diffusion under the unified scale, where Fig. 4(a1)–(c1) represents the isotropic diffusion; Fig. 4(a2)–(c2) represents the anisotropic diffusion, Fig. 4(a1) and (a2) are the surface plots of diffusion flux, Fig. 4(b1) and (b2) are the vertical views of diffusion flux, and Fig. 4(c1) and (c2) are the composites of flux contour and image. It can be seen that isotropic diffusion has the same diffusion effect for different types of ground features under the unified scale. Thus, it must adjust the scale parameters to adapt to different ground features. Anisotropic diffusion can control the diffusion direction and amount according to the image characteristics. It is beneficial to adapt to different types of ground features. The advantage of anisotropic diffusion is that it can help the algorithm to segment the ground features in complex scenes, such as the large difference in scale, shape, area, and texture complexity, especially narrow ground features, as shown in the tennis court and road in Fig. 4(c1) and (c2).

B. Superpixel Segmentation

In superpixel segmentation based on clustering, the dissimilarity measure plays an important role. In past studies, the spatial distance is introduced into the spectral distance to improve the spatial continuity of superpixel and overcome noises. However,

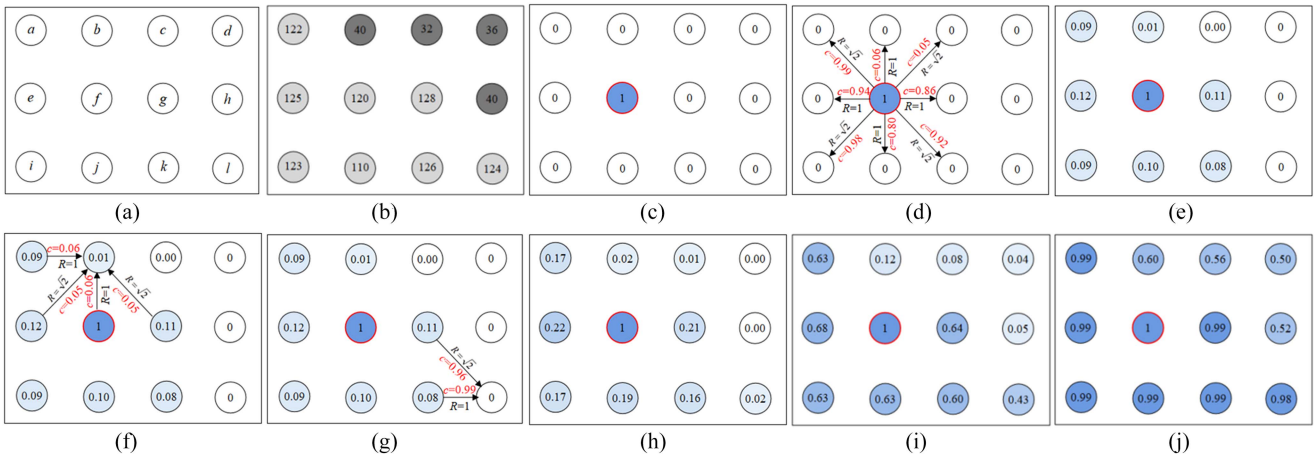

Fig. 3. Concentration diffusion process with diffusion coefficient c_1. (a) Index of pixel site. (b) Gray value of pixels. (c) Diffusion flux at the initial time (i.e., $t = 0$) by assuming f as the seed point. (d) Diffusion state at $t = 0$. (e) Diffusion result (i.e., diffusion flux at $t = 1$). (f) Diffusion state of the representative pixel b at $t = 1$, which is close to the seed point but have significant difference in gray. (g) Diffusion state of the representative pixel l at $t = 1$, which is far from the seed point but have a similar gray. (h) Second diffusion result (i.e., diffusion flux at $t = 2$). (i) Diffusion flux at $t = 10$. (j) Diffusion flux at $t = 40$.

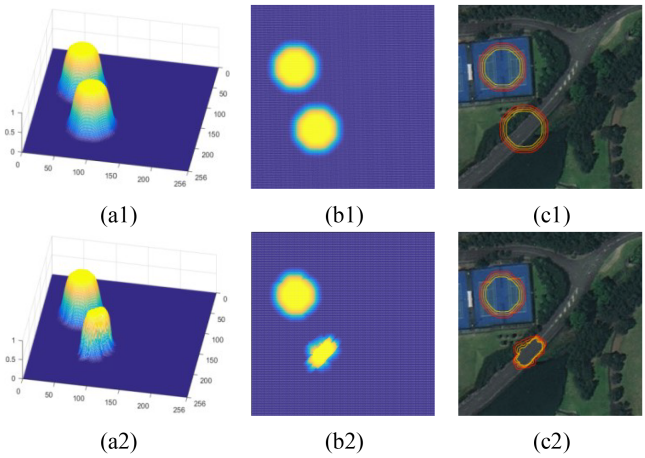

Fig. 4. Difference between isotropic diffusion and anisotropic diffusion under the unified scale. (a1)–(c1) Surface plot, vertical view, and the composite image of flux contour and image, under isotropic diffusion. (a2)–(c2) Surface plot, vertical view, and the composite image of flux contour and image, under anisotropic diffusion.

there are also many problems. For example, it is difficult to control the degree and cannot well fit the boundaries when there are many complex ground features. To deal with this problem, the diffusion flux generated based on the anisotropic diffusion model is introduced to describe the dissimilarity between any pixel i and seed point j

$$D_{ij} = \sqrt{\left(\frac{d_{ij}^{\text{Spatial}}}{N_{\text{Spatial}}}\right)^2 + \left(\frac{d_{ij}^{\text{Spectral}}}{N_{\text{Spectral}}}\right)^2 + \left(\frac{1 - d_{ij}^{\text{Flux}}}{N_{\text{Flux}}}\right)^2} \quad (9)$$

where d_{ij}^{Spatial} and d_{ij}^{Spectral} are the spatial distance and spectral distance between pixel i and seed point j, respectively. d_{ij}^{Flux} is the diffusion flux of pixel i diffused from seed point j. $N_{\text{Spatial}}, N_{\text{spectral}}, N_{\text{Flux}}$ are the corresponding normalized parameters

$$d_{ij}^{\text{Spatial}} = \sqrt{(x_i - a_j)^2 + (y_i - b_j)^2} \quad (10)$$

$$d_{ij}^{\text{Spectral}} = \|\boldsymbol{I}_i - \boldsymbol{S}_j\|_2 \quad (11)$$

$$d_{ij}^{\text{Flux}} = U^j(x_i, y_i, T) \quad (12)$$

where T is the final moment and satisfied $T > 2N_{\text{Spatial}}$.

To obtain the optimal segmentation result, a process about optimization seed points is carried out by K-means framework. First, the initialized seed points are sited at the center of the uniform grid with $N_{\text{Spatial}} \times N_{\text{Spatial}}$ size. Move each seed point to the position with the lowest gradient in the 3×3 window to avoid falling on the boundary or making noise. Then, the pixels are divided into the superpixel j by dissimilarity D_{ij} minimizing criteria. Updating the superpixel center $\boldsymbol{S}_j(a_j, b_j)$ until the superpixel center set $\boldsymbol{S}$ is convergent or the iteration reaches the maximum. Finally, eliminate the isolated pixels according to the mode of neighborhood labels to enhance the connectivity within superpixels.

C. Summary of the Proposed Algorithm

The proposed algorithm, a superpixel segmentation algorithm based on anisotropic diffusion model (ADS), is summarized in Algorithm 1, where the diffusion flux calculating method is summarized in Algorithm 2.

IV. Experiments

A. Experimental Data

The research area is located in Dapeng New District, Shenzhen. There are two tested regions, where region 1 is used to evaluate the effectiveness of algorithms in mid–low spatial resolution images, and the smaller region 2 is used for high spatial resolution images because of the consideration of the big data characteristics. As shown in Fig. 5, where Fig. 5(a) shows

Algorithm 1: ADS.

Input: Remote sensing image $\boldsymbol{I}$, the normalized parameters $N_{\text{Spatial}}, N_{\text{spectral}}, N_{\text{Flux}}$, histogram threshold η, convergence threshold θ, iteration index $\tau = 0$, maximum iterations M.
Output: Superpixel label $\boldsymbol{L}$, number of superpixels m.
Initializing the seed points $\boldsymbol{S} = \{\boldsymbol{S}_j(a_j, b_j) : j = 1, \ldots, m\}$.
Moving each seed point to the position with the lowest gradient in 3×3 window.
Repeat
for each seed point $\boldsymbol{S}_j(a_j, b_j)$ **do**
Calculating the spatial distance d_{ij}^{Spatial}.
Calculating the spectral distance d_{ij}^{Spectral}.
Calculating the diffusion flux d_{ij}^{Flux} by Algorithm 2.
Calculating the dissimilarity D_{ij} between any pixel *i* and seed point *j*.
end
The superpixel label $L_i = \text{argmin}(D_{ij})$.
Updating the seed points according to K-means.
Until max $|\boldsymbol{S}^{\tau} - \boldsymbol{S}^{\tau-1}| < \theta$ or $\tau > M$
Eliminating the isolated pixels.
Return The final superpixel label $\boldsymbol{L}$ and the number of superpixels m.

Algorithm 2: Diffusion Flux.

Input: Remote sensing image $\boldsymbol{I}$, the seed points $\boldsymbol{S}$, uniform grid size N_{Spatial}, histogram threshold η
Output: The diffusion flux d_{ij}^{Flux}
Generating the flux image $U^j(x_i, y_i, t = 0)$ at the initial time according to the seed points $\boldsymbol{S}$,
Calculating the spectrum gradient image $\nabla^f \boldsymbol{I}(x, y, t)$.
Adaptively determining the gradient threshold δ at each direction by the histogram threshold η
Calculating the diffusion coefficient by c_1 or c_2
Repeat
Updating the flux image $U^j(x_i, y_i, t)$ over time
Until $t > 2 \times N_{\text{Spatial}}$
Return The diffusion flux d_{ij}^{Flux}, i.e., $U^j(x_i, y_i, T)$.

the location of research regions in Dapeng, Fig. 5(b) shows the 30 m resolution data with R, G, and B, NIR bands, and 298×204 size clipped from Landsat 8. Fig. 5(c) shows the 10 m resolution data with R, G, and B, NIR bands, and 869×576 size clipped from Sentinel-2, and Fig. 5(d) shows the 0.5 m resolution data with R, G, and B, NIR bands, and 1284×1288 size clipped from SkySat. All images are taken in 2021. Fig. 6 shows the ground truth of them. There are six classes, namely forest, artificial, water, farmland, bare land, and grass.

B. Quantitative Evaluation Index

To quantitatively assess the effectiveness of the proposed method, the boundary recall (Recall), undersegmentation error (UE), explained variation (EV), compactness (CO), and

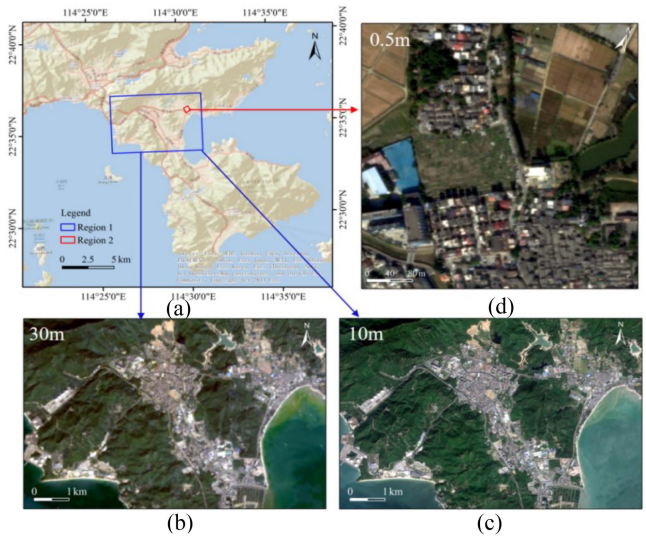

Fig. 5. Research area. (a) Location of research regions in Dapeng. (b) Research region 1 from Landsat 8 image with 30 m spatial resolution. (c) Research region 1 from Sentinel-2 image with 10 m spatial resolution. (d) Research region 2 from SkySat image with 0.5 m spatial resolution.

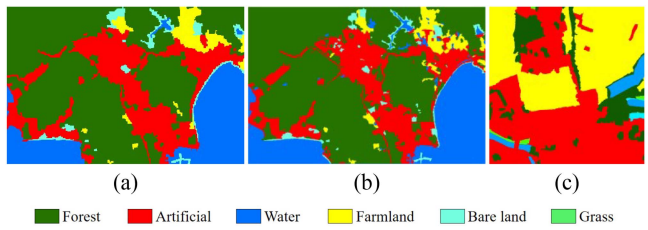

Fig. 6. Ground truth. (a) Landsat 8. (b) Sentinel-2. (c) SkySat.

achievable segmentation accuracy (ASA) are used to evaluate the boundary adherence, compactness, and quality of superpixels; the overall accuracy (OA) and Kappa (Ka) are used to evaluate the accuracy of the classification result. Let $\boldsymbol{L} = \{L_j : j = 1, \ldots, m\}$, $\boldsymbol{G} = \{G_l : l = 1, \ldots, k\}$, $\boldsymbol{C} = \{C_l : l = 1, \ldots, k\}$ be the superpixel segmentation result, ground truth, and classification result of the image $\boldsymbol{I}$, respectively, l is the index of class, and k is the number of classes. Then, the quantitative evaluation index can be calculated as follows.

1) Boundary Recall (Recall): Recall is the fraction of all boundary pixels within the ground truth segmentation *G*, which are correctly detected within the superpixel. The higher the Recall is, the better

$$\text{Rec}(\boldsymbol{G}, \boldsymbol{L}) = \frac{\text{TP}(\boldsymbol{G}, \boldsymbol{L})}{\text{TP}(\boldsymbol{G}, \boldsymbol{L}) + \text{FN}(\boldsymbol{G}, \boldsymbol{L})} \tag{13}$$

where TP($\boldsymbol{G}$, $\boldsymbol{L}$) and FN($\boldsymbol{G}$, $\boldsymbol{L}$) are the number of true positive and false negative boundary pixels in $\boldsymbol{L}$ with respect to $\boldsymbol{G}$.

2) Undersegmentation Error (UE): UE describes the leakage of superpixel with respect to a specific ground truth segment. The lower the UE is, the better. To facilitate comparison, 1-UE is used in the experimental part. The

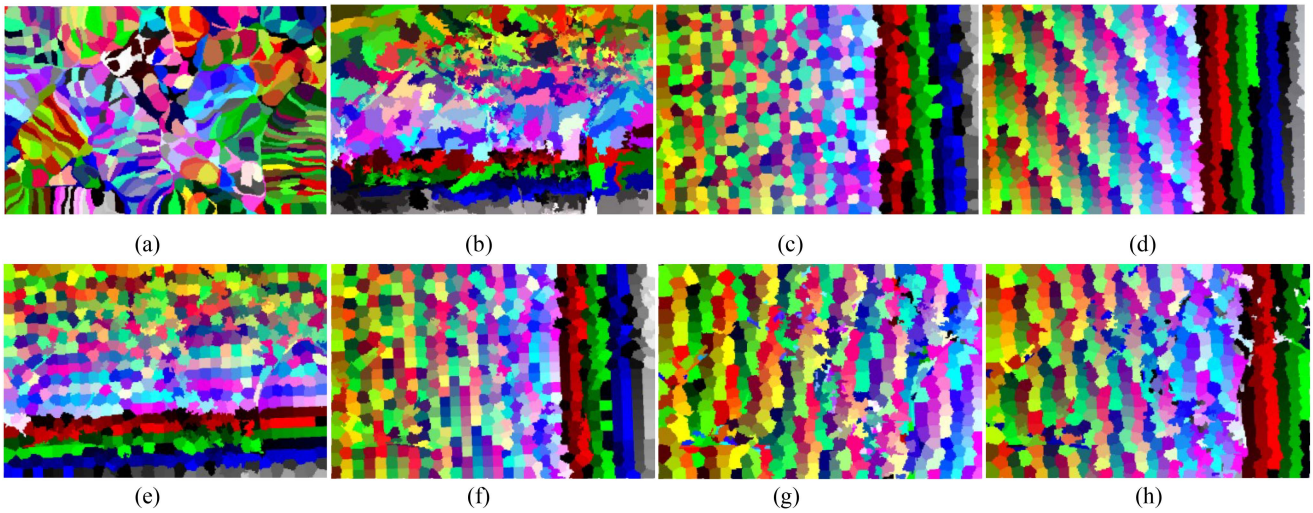

Fig. 7. Superpixel segmentation results of Landsat 8, (a) NCut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, and (h) proposed algorithm based on c_2 diffusion coefficient.

higher the 1-UE is, the better

$$\mathrm{UE}(\boldsymbol{G},\boldsymbol{L}) = \frac{1}{n}\sum_{G_l}\sum_{L_j\cap G_l\neq 0}\min\{|L_j\cap G_l|,|L_j-G_l|\}. \tag{14}$$

3) Explained Variation (EV): EV quantifies the quality of a superpixel segmentation based on image rather than ground truth. It can assess boundary adherence independent of human annotations. The higher the EV is, the better

$$\mathrm{EV}(L) = \frac{\sum_{j=1}^{m}|L_j|\,\|\boldsymbol{\mu}(L_j)-\boldsymbol{\mu}(\boldsymbol{I})\|_2^2}{\sum_{i=1}^{n}\|\boldsymbol{I}_i-\boldsymbol{\mu}(\boldsymbol{I})\|_2^2} \tag{15}$$

where $\mu(L_j)$ and $\mu(\boldsymbol{I})$ are the mean of superpixel L_j and the image $\boldsymbol{I}$, and $|\,.\,|$ is a symbol for counting.

4) Compactness (CO): CO is used to evaluate the compactness of superpixels. The higher the CO is, the better

$$\mathrm{CO}(\boldsymbol{G},\boldsymbol{L}) = \frac{1}{n}\sum_{j=1}^{m}|L_j|\frac{4\pi A(L_j)}{P(L_j)^2} \tag{16}$$

where $A(L_j)$ and $P(L_j)$ are the area and perimeter of superpixel L_j, respectively.

5) Achievable Segmentation Accuracy (ASA): ASA is the achievable accuracy for segmentation using superpixels as a preprocessing step. The higher the ASA is, the better

$$\mathrm{ASA}(\boldsymbol{G},\boldsymbol{L}) = \frac{1}{n}\sum_{L_j}\max_{G_l}\{|L_j\cap G_l|\}. \tag{17}$$

6) Overall Accuracy (OA): OA is the ratio of correctly classified pixels to the total number of pixels. The higher the OA is, the better

$$\mathrm{OA}(\boldsymbol{G},\boldsymbol{C}) = \frac{1}{n}\sum_{l=1}^{k}N_{ll} \tag{18}$$

where N_{ll} is the number of pixels that simultaneously belongs to l in both classification result and ground truth.

7) Kappa (Ka): Ka is a coefficient describing the proportion of error reduction compared with random segmentation results. The higher the Kappa is, the better

$$\mathrm{Ka} = \frac{n\sum_{l=1}^{k}N_{ll}-\sum_{l=1}^{k}N_{l+}N_{+l}}{n^2-\sum_{l=1}^{k}N_{l+}N_{+l}} \tag{19}$$

where N_{l+} is the number of pixels in class l in the ground truth, and N_{+l} is the number of pixels in class l in the classification result.

C. Superpixel Effectiveness Analysis

The number of superpixels is the main factor affecting the segmentation accuracy and time efficiency, which is usually given by human experience. To avoid the impact of fluctuation of superpixels number on algorithm evaluation, the unified number for different algorithms is determined to be the smaller value corresponding to the accuracy convergence. For the Landsat 8 image, $m = 700$. For the Sentinel-2 image, $m = 3000$. For the SkySat image, $m = 4000$.

1) Landsat 8: Fig. 7 shows the superpixel segmentation result of Landsat 8, where Fig. 7(a)–(h) shows Ncut, ERS, TurboPixels, SCoW, SLIC, LSC, and the proposed algorithm ADS with diffusion coefficients c_1 and c_2, respectively. It can be seen that the shape of superpixels based on NCut is crescent shaped, and the boundaries between superpixels are extremely smooth. ERS is with different shapes and sizes, and the boundary is zigzag. The shape of TurboPixels, SCoW, and LSC is almost regular hexagons, and they are in high compactness. SLIC improves the boundary applicability. But it is still strictly limited by compactness. Compared with the above algorithms, the proposed one can have high compactness in the homogeneous region and targeted fit various shapes in the heterogeneous region without compactness limitation because of the boundary constraint established by the anisotropic diffusion model, as shown in Fig. 7(g) and (h).

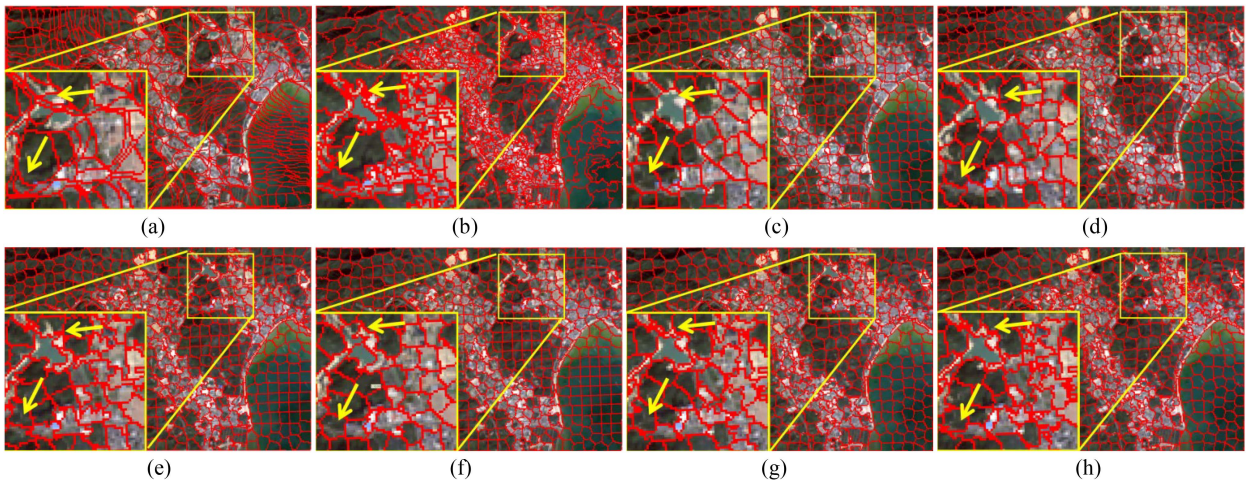

Fig. 8. Superpixel overlay images of Landsat 8, (a) NCut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, and (h) proposed algorithm based on c_2 diffusion coefficient.

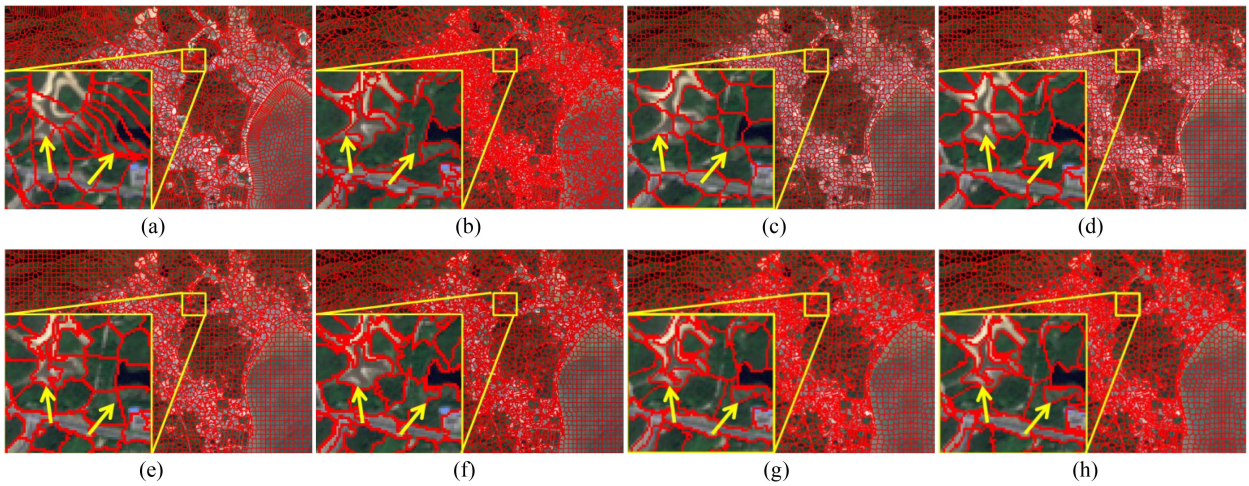

Fig. 9. Superpixel overlay images of Sentinel-2, (a) NCut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, and (h) proposed algorithm based on c_2 diffusion coefficient.

Fig. 8 shows the corresponding superpixel overlay image of Landsat 8. The yellow area is the enlarged view. It shows that NCut, TurboPixels, and SCoW cannot fit the narrow object, such as the bare land around the reservoir and the road mixed in the forest, as shown in the arrow of Fig. 8(a), (c), and (d). Although SLIC and LSC have improved performance in the reservoir, they also cannot distinguish different objects with similar colors, such as the road, as shown in Fig. 8(e) and (f). ERS and the proposed algorithm based on both c_1 and c_2 show better results, and coefficients c_1 and c_2 have a similar performance. In addition, the proposed algorithm can pertinently segment complex areas.

2) Sentinel-2: Fig. 9 shows the superpixel overlay image of Sentinel-2. It shows that NCut and TurboPixels cannot segment the water that is with obvious edges, such as the black object in enlarged view, as shown in Fig. 9(a) and (c). In addition, although other comparing algorithms effectively segment the water body, they are difficult to distinguish the tidal flat of reservoir to a certain extent, as shown in Fig. 9(b) and (d)–(f). The proposed algorithm has good ability to distinguish narrow ground objects in images with large-scale differences and broken scenes, as shown in the tidal flat in Fig. 9(g) and (h).

3) SkySat: To verify the performance of the proposed method in high-resolution remote sensing images, the SkySat data are also tested. The superpixel overlay images are shown in Fig. 10. With the increase of the spatial resolution, the number of pixels increases dramatically, which helps to improve the overall performance of the superpixels segmentation algorithms. But, there are still some thin and narrow objects that cannot be well segmented. Such as the boundaries with similar spectra shown by the arrow. NCut and Turbopixles usually cannot explore the boundaries, as shown in Fig. 10(a) and (c). The effect of SCoW is equivalent to SLIC, and they are in the second echelon, as shown in Fig. 10(d) and (e). ERS is slightly inferior to LSC and the proposed algorithm.

4) Accuracy Analysis of Superpixel: To quantitatively evaluate the effectiveness of the proposed superpixel segmentation algorithm, the superpixel evaluation index: Recall, UE, EV, CO, and ASA of different images and algorithms are drawn in

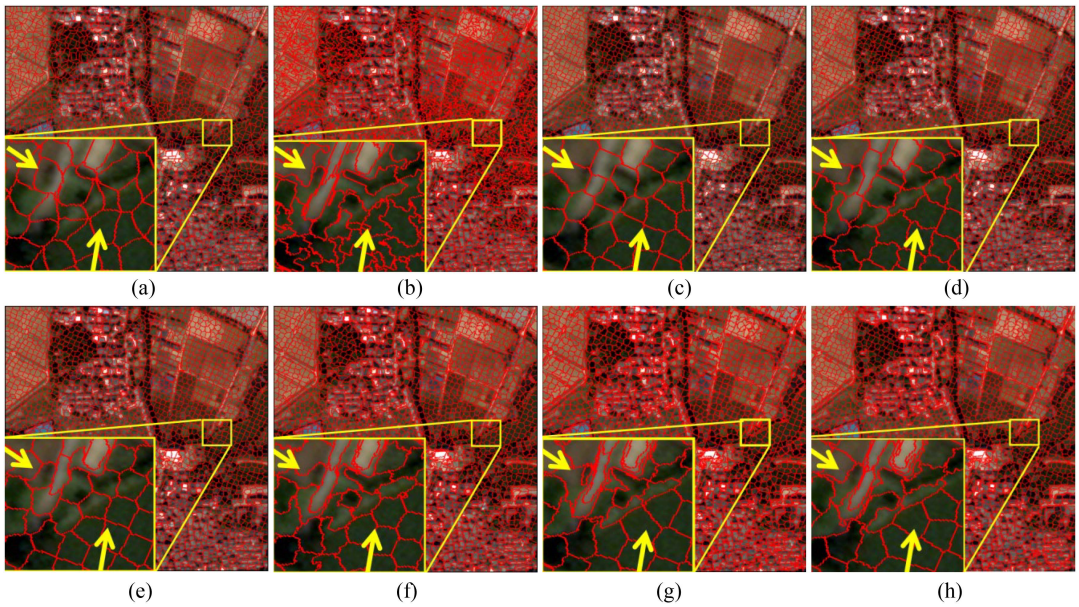

Fig. 10. Superpixel overlay images of SkySat, (a) NCut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, (h) and proposed algorithm based on c_2 diffusion coefficient.

TABLE I
SUPERPIXEL ACCURACIES

Index	Images	Algorithms							
		NCut	ERS	Turbopixels	SCow	SLIC	LSC	Proposed_c1	Proposed_c2
Recall	Landsat 8	0.95±0.01	0.97±0.02	0.89±0.01	0.92±0.01	0.93±0.03	0.94±0.03	0.98±0.00	0.98±0.01
	Sentinel-2	0.72±0.01	0.95±0.03	0.75±0.00	0.75±0.02	0.73±0.1	0.74±0.11	0.96±0.01	0.96±0.01
	Skysat	0.69±0.00	0.85±0.02	0.66±0.00	0.72±0.02	0.66±0.04	0.69±0.06	0.90±0.01	0.90±0.01
1-UE	Landsat 8	0.82±0.00	0.89±0.03	0.82±0.00	0.82±0.01	0.88±0.02	0.88±0.02	0.89±0.00	0.89±0.01
	Sentinel-2	0.83±0.01	0.95±0.03	0.88±0.00	0.87±0.01	0.87±0.02	0.88±0.03	0.92±0.01	0.92±0.01
	Skysat	0.94±0.00	0.95±0.01	0.95±0.00	0.95±0.00	0.95±0.01	0.95±0.01	0.96±0.00	0.96±0.00
EV	Landsat 8	0.993±0.00	0.995±0.00	0.994±0.00	0.994±0.00	0.995±0.00	0.995±0.00	0.995±0.00	0.995±0.00
	Sentinel-2	0.976±0.01	0.998±0.00	0.986±0.00	0.983±0.00	0.992±0.00	0.992±0.00	0.994±0.00	0.993±0.00
	Skysat	0.954±0.00	0.971±0.00	0.961±0.00	0.967±0.00	0.930±0.00	0.974±0.00	0.98±0.00	0.974±0.00
CO	Landsat 8	0.49±0.01	0.38±0.01	0.69±0.00	0.63±0.01	0.61±0.02	0.66±0.02	0.47±0.1	0.46±0.02
	Sentinel-2	0.58±0.03	0.36±0.02	0.75±0.01	0.71±0.01	0.7±0.08	0.75±0.06	0.58±0.02	0.39±0.03
	Skysat	0.51±0.00	0.28±0.02	0.63±0.00	0.65±0.00	0.57±0.03	0.62±0.07	0.40±0.02	0.42±0.04
ASA	Landsat 8	0.89±0.00	0.94±0.01	0.92±0.00	0.92±0.00	0.94±0.00	0.94±0.00	0.95±0.00	0.94±0.01
	Sentinel-2	0.93±0.00	0.96±0.01	0.94±0.00	0.93±0.01	0.95±0.02	0.94±0.01	0.96±0.00	0.96±0.00
	Skysat	0.97±0.00	0.97±0.01	0.97±0.00	0.97±0.01	0.97±0.00	0.98±0.00	0.98±0.00	0.98±0.00

Fig. 11(a1)–(c1), (a2)–(c2), (a3)–(c3), (a4)–(c4), and (a5)–(c5), respectively. To facilitate observation, the UE index is drawn by 1-UE. Then, all indices shown in the figures take maximization as the standard. The corresponding accuracy value is listed in Table I. For the commonly used indices: Recall and UE, from the perspective of superpixels algorithms, the accuracies of ERS, SLIC, and LSC are with big fluctuation ranges, and others are relatively convergent. TurboPixels has the worse Recall and 1-UE, followed by NCut and SCow. The proposed algorithm based on c_1 and c_2 can be stable in high accuracy whether Recall or 1-UE. From the perspective of different spatial resolutions, Recall decreases with the increase of resolution, and the 1-UE is reversed. It shows that there are more fine boundaries in high-resolution images. Except for ERS and the proposed algorithm, the accuracy of other algorithms has been polarized in the SkySat image. For example, the Recalls of TurboPixels, SCoW, SLIC, and LSC are only around 0.7, but 1-UEs are as high as 0.95, as shown in Table I. For the index EV, evaluating superpixel segmentation results from the perspective of images can have very high accuracy, such as 0.99. For the index CO, it indicates the degree of circularity of superpixels. ERS and the proposed algorithm have lower CO, and it indirectly demonstrates that the proposed method can achieve flexible segmentation of ground objects in an anisotropic manner without circular constraint. For the index ASA, the trend of differences between algorithms is the same as recall and UE. To sum up, the proposed one can better explore complex boundaries and improves superpixel segmentation accuracy.

D. *Influence of Superpixels on Classification*

Considering the issue of sample size for images with different resolutions, random forests are selected for discussing the influence of superpixels on classification. The spectral, texture,

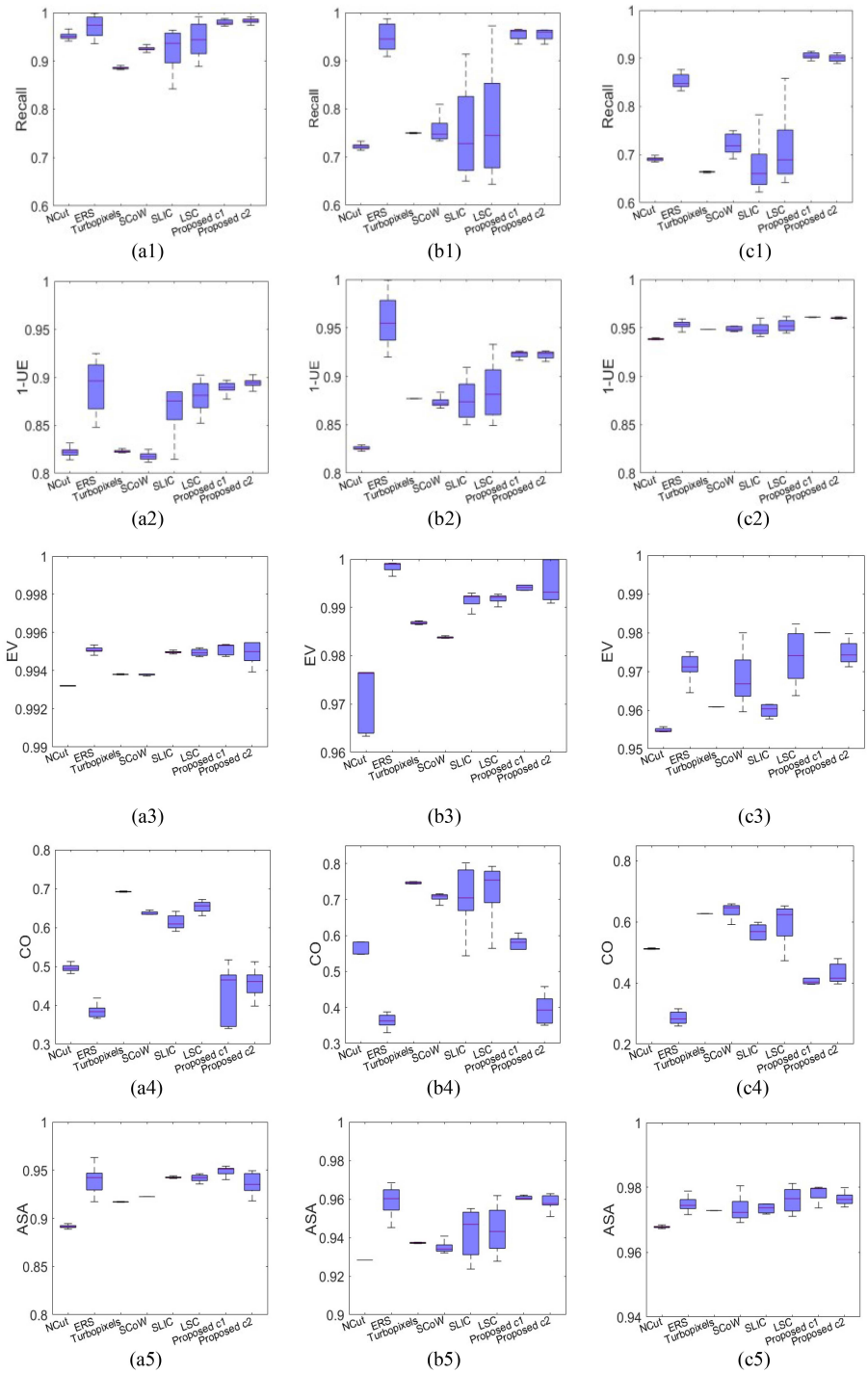

Fig. 11. Quantitative evaluation boxplot of the superpixel segmentation results of Landsat 8, Sentinel-2, and SkySat. (a1)–(c1) Recall, (a2)–(c2) 1-UE, (a3)–(c3) EV, (a4)–(c4) CO, and (a5)–(c5) ASA. (a1) Landsat 8. (b1) Sentinel-2. (c1) SkySat. (a2) Landsat 8. (b2) Sentinel-2. (c2) SkySat. (a3) Landsat 8. (b3) Sentinel-2. (c3) SkySat. (a4) Landsat 8. (b4) Sentinel-2. (c4) SkySat. (a5) Landsat 8. (b5) Sentinel-2. (c5) SkySat.

vegetation index, and water index characteristics of superpixels are input into the classifier. Take the ground truth images as sample sets, where 70% of them are used for training and 30% for testing.

1) Landsat 8: Fig. 12 shows the corresponding classification result of Landsat 8, where Fig. 12(a)–(h) shows NCut, ERS, TurboPixels, SCoW, SLIC, LSC, and the proposed algorithm with c_1 and c_2. The enlarged area of the arrow is shown in Fig. 13, where Fig. 13(a1) and (a2) shows the ground truth, and Fig. 13(b1) and (b2)–(i1) and (i2) shows the comparing and proposed algorithms. It can be seen that a lot of detail information is ignored in NCut; the refined boundary and the object with a smaller area are not classified, as shown in Fig. 13(b1) and (b2). TurboPixels, SCoW, and LSC present obvious convex structures, such as the boundary between different objects, as shown in Fig. 12(c), (d), and (f), which leads to the misclassification in

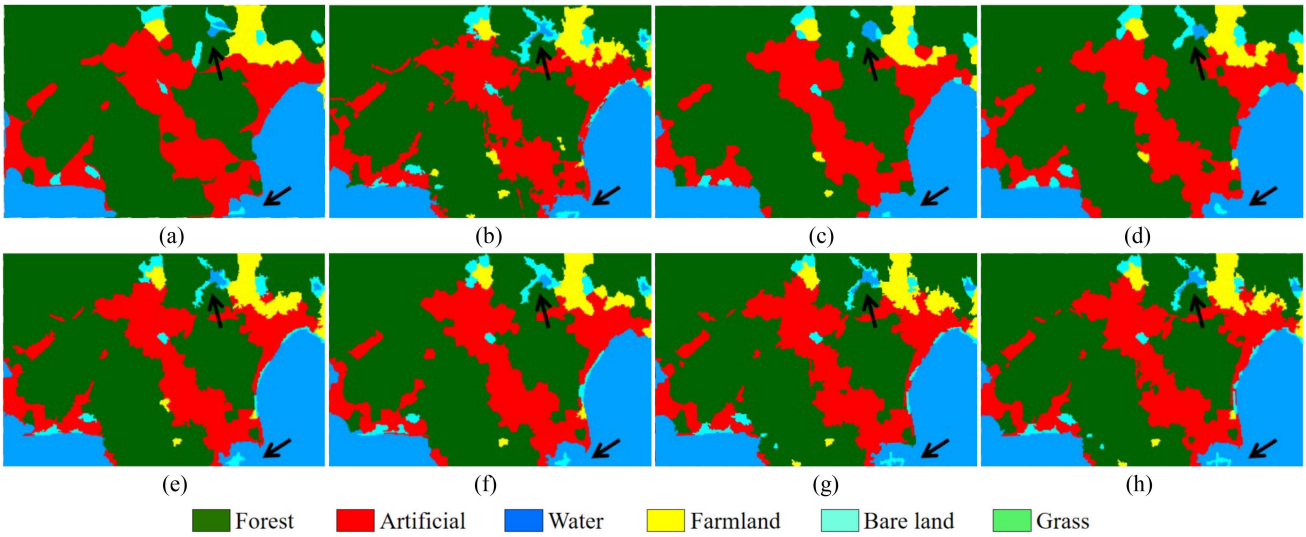

Fig. 12. Classification results of Landsat 8, (a) NCut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, and (h) proposed algorithm based on c_2 diffusion coefficient.

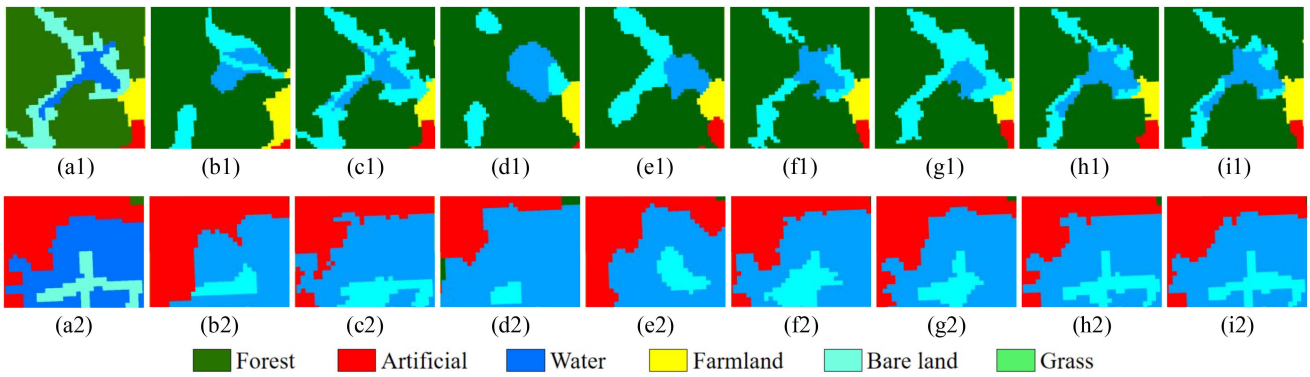

Fig. 13. Difference between classification results and ground truth of Landsat 8, (a1) and (a2) ground truth, (b1) and (b2) NCut, (c1) and (c2) ERS, (d1) and (d2) TurboPixels, (e1) and (e2) SCoW, (f1) and (f2) SLIC, (g1) and (g2) LSC, (h1) and (h2) proposed algorithm based on c_1 diffusion coefficient, and (i1) and (i2) proposed algorithm based on c_2 diffusion coefficient.

the narrow area, as shown in Fig. 13(d1), (e1), and (g1) and (d2), (e2), and (g2); SLIC is obviously improved compared with the previous algorithm but still inferior to ERS and the proposed algorithm. In addition, the proposed algorithm is more stable and adaptable.

2) Sentinel-2: Fig. 14 shows the classification result of Sentinel-2. The enlarged area of the arrow is shown in Fig. 15. Compared with the Landsat 8 classification result, some detailed information can be easily classified because of the increased pixel number, such as the road. The segmentation effect of TurboPixels, SCoW, SLIC, and LSC significantly improved, except for NCut. But there are still many misclassified pixels. ERS, the same as the proposed algorithm visually, has a satisfactory classification result. As a whole, the proposed algorithm is closer to the ground truth, as shown in Fig. 15(h1) and (i1), and (h2) and (i2).

3) SkySat: Fig. 16 shows the classification result of SkySat. The enlarged area of the arrow is shown in Fig. 17. Compared with mid–low resolution remote sensing images, the misclassification pixels are greatly reduced, and the error mainly occurs in narrow areas and borders. As shown in Fig. 17(b1) and (b2), NCut is undersegmentation. As shown in Fig. 17(e1) and (f1), SCoW and SLIC did not find some parts of the bare land, which is affected by inaccurate superpixel boundary. As shown in Fig. 17(c2), (d2), and (g2), there are some misclassifications between water and grass in ERS, TurboPixels, and LSC. The proposed algorithm is more similar to the ground truth, as shown in Fig. 17(h1) and (h2), and (i1) and (i2).

4) Accuracy Analysis of Classification: To quantitatively evaluate the influence of superpixels on classification, the classification evaluation indices OA and Ka of different images and algorithms are drawn in Fig. 18(a1)–(c1) and (a2)–(c2), respectively. The corresponding accuracy value is listed in Table II. It can be seen that the trend of OA and Ka is the same as the Recall and UE of the superpixel segmentation algorithm, and the proposed algorithm is still in the lead, followed by ERS, LSC, and SLIC, because the boundary-fitting accuracy can directly affect the internal characteristics of superpixel and then affect the judgment in the classification process. With the improvement of spatial resolution, the OA shows an overall growth trend.

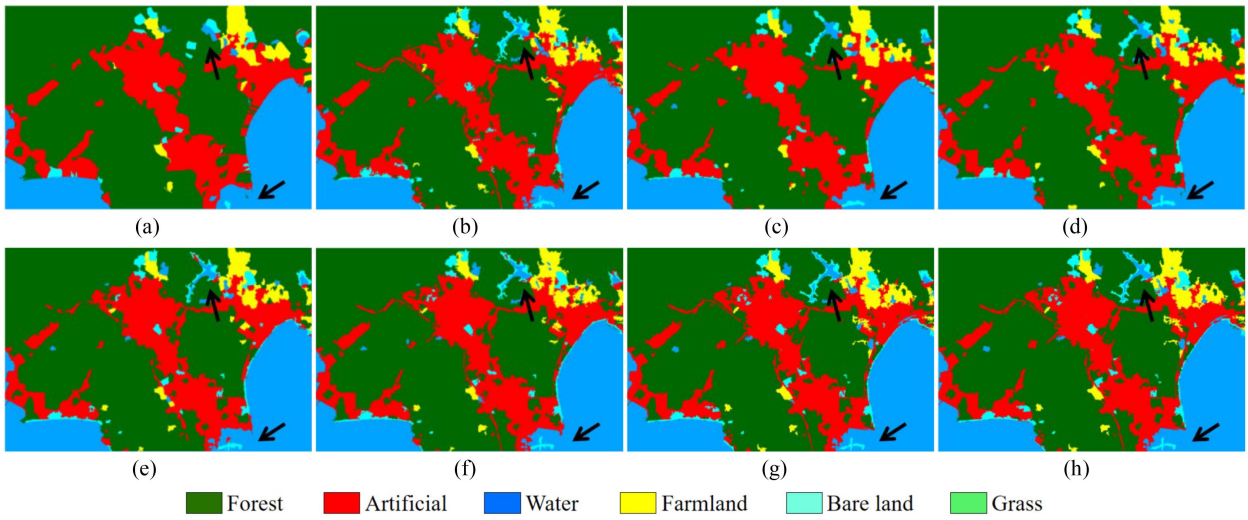

Fig. 14. Classification results of Sentinel-2, (a) NCut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, and (h) proposed algorithm based on c_2 diffusion coefficient.

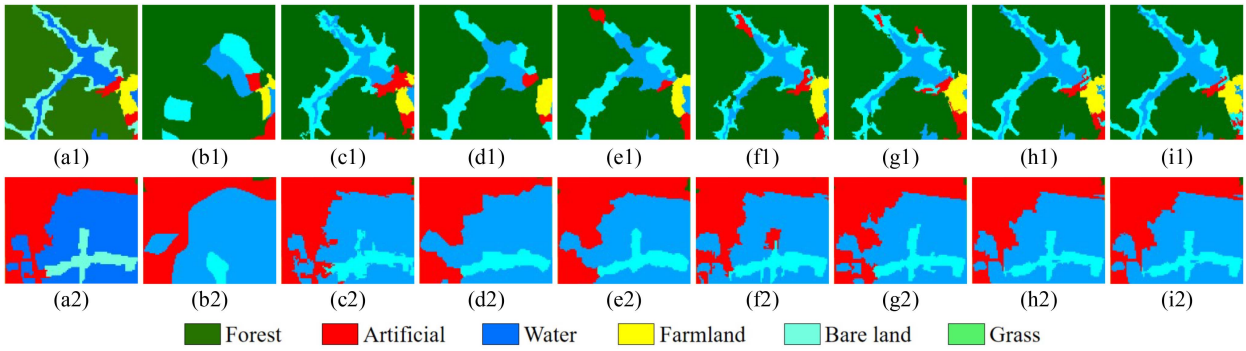

Fig. 15. Difference between classification results and ground truth of Sentinel-2, (a1) and (a2) ground truth, (b1) and (b2) NCut, (c1) and (c2) ERS, (d1) and (d2) TurboPixels, (e1) and (e2) SCoW, (f1) and (f2) SLIC, (g1) and (g2) LSC, (h1) and (h2) proposed algorithm based on c_1 diffusion coefficient, and (i1) and (i2) proposed algorithm based on c_2 diffusion coefficient.

TABLE II
CLASSIFICATION ACCURACIES

Index	Images	Algorithms							
		NCut	ERS	Turbopixels	SCow	SLIC	LSC	Proposed_c1	Proposed_c2
OA	Landsat 8	91.00±0.12	93.98±1.08	91.38±0.05	90.97±0.18	93.68±1.35	93.49±0.70	93.89±0.17	94.18±0.09
	Sentinel-2	89.14±0.11	93.95±0.85	92.07±0.04	91.65±0.31	91.81±1.14	92.28±1.52	94.66±0.17	94.66±0.32
	Skysat	94.23±0.05	95.49±0.29	95.01±0.02	95.13±0.24	94.79±0.55	95.19±0.64	96.04±0.05	96.03±0.06
Ka	Landsat 8	0.86±0.00	0.91±0.02	0.86±0.00	0.86±0.00	0.90±0.02	0.90±0.02	0.90±0.01	0.91±0.00
	Sentinel-2	0.76±0.01	0.88±0.02	0.84±0.00	0.84±0.01	0.84±0.02	0.85±0.03	0.89±0.00	0.89±0.00
	Skysat	0.91±0.00	0.93±0.01	0.92±0.00	0.92±0.01	0.92±0.01	0.92±0.01	0.94±0.00	0.94±0.00

Compared with Landsat 8 image, Sentinel-2 images have a higher spatial resolution, and some detailed information expected to be distinguished is highlighted. However, its spatial resolution is not as sufficient as the SkySat image to distinguish ground objects. Thus, the image complexity of Sentinel-2 is relatively high. Affected by image complexity, Ka of Sentinel-2 is lower than others. But the Ka of the proposed algorithm is less affected and still maintains the highest value. For example, SLIC is reduced from 0.9 to 0.84, while the proposed algorithm is only reduced from 0.9 to 0.89, almost unchanged. The effectiveness of the proposed method is further verified.

V. DISCUSSION

A. Ablation Study

To verify the effectiveness of boundary constraint terms in the proposed dissimilarity measure, an ablation study is performed. The ablated part is the third term of (9). Table III presents a comparison of the accuracy of the ablation experiment and the proposed method. It shows that the difference in Recall is relatively clear, and the boundary constraint term can increase accuracy by 7%. The increased CO accuracy after ablation also indicates the validity of boundary terms, i.e., anisotropic

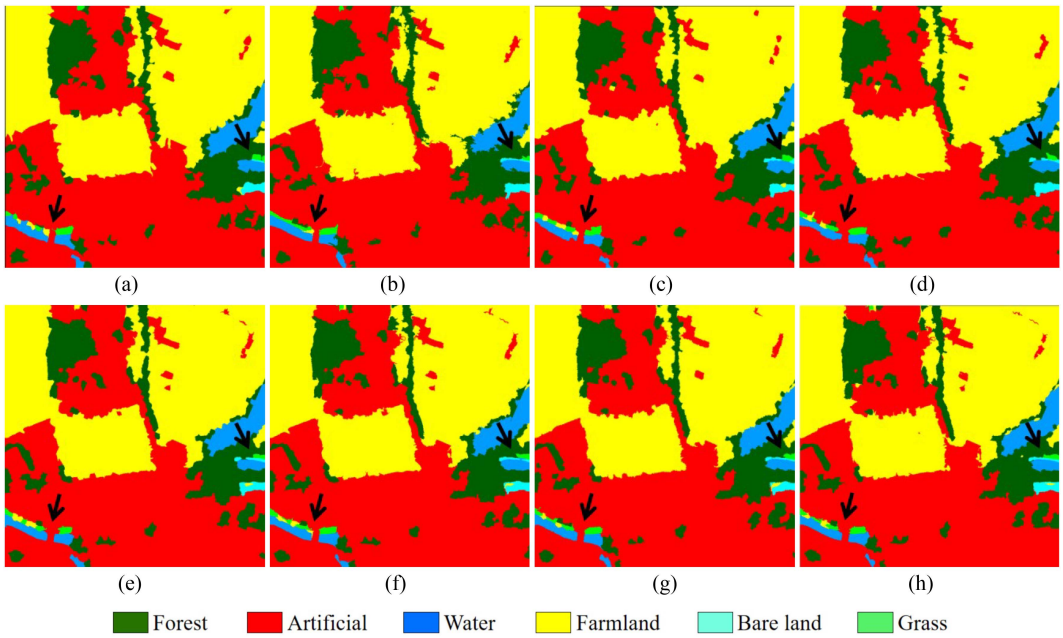

Fig. 16. Classification results of SkySat, (a) Ncut, (b) ERS, (c) TurboPixels, (d) SCoW, (e) SLIC, (f) LSC, (g) proposed algorithm based on c_1 diffusion coefficient, and (h) proposed algorithm based on c_2 diffusion coefficient.

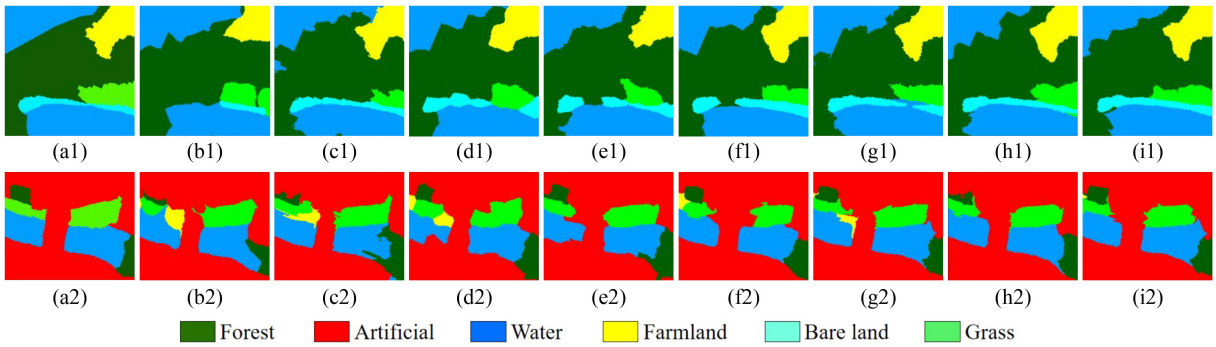

Fig. 17. Difference between classification results and ground truth of SkySat, (a1) and (a2) ground truth, (b1) and (b2) NCut, (c1) and (c2) ERS, (d1) and (d2) TurboPixels, (e1) and (e2) SCoW, (f1) and (f2) SLIC, (g1) and (g2) LSC, (h1) and (h2) proposed algorithm based on c_1 diffusion coefficient, and (i1) and (i2) proposed algorithm based on c_2 diffusion coefficient.

segmentation rather than isotropic segmentation. For the OA, there is also a certain loss of accuracy after ablation.

B. Parameter Sensitivity Analysis

To discuss the influence of parameters on model performance, the changes of OA with parameters are plotted in Fig. 19. The three columns represent the result of Landsat 8, Sentinel-2, and SkySat images. The four rows represent the characteristic normalization parameters N_{Spatial}, N_{Spectral}, N_{Flux}, and histogram threshold η. It shows that the OA decreases as N_{Spatial} and N_{Spectral} increase and increases as N_{Flux} and η increase. The OA can converge to a certain range with the change of parameters. With the improvement of the spatial resolution, the starting points of OA become larger, and the change degree of the curve is becoming more and more gentle. In addition, the proposed algorithm based on the c_2 coefficient model generally shows better performance than c_1 because the rapidly decreasing diffusion curve makes the diffusion speed stop at the boundary area faster.

The number of superpixels is a significant factor in object-oriented image processing. Too few superpixels can lead to severe undersegmentation, while too many superpixels can lead to computational redundancy and a long time. Fig. 20 shows the classification results with different numbers of superpixels by taking Landsat 8 image and $c1$ coefficient as an example. It shows that the proposed algorithm satisfies the objective law that as the number of superpixels increases, the details can be segmented better, and the accuracy of classification will also be higher. When the number of superpixels reaches 600, it can already achieve excellent classification results, as shown in Fig. 20(e). In addition, the proposed algorithm can achieve relatively good classification with fewer superpixels, as shown in Fig. 20(a). It indicates that the proposed algorithm can reduce the

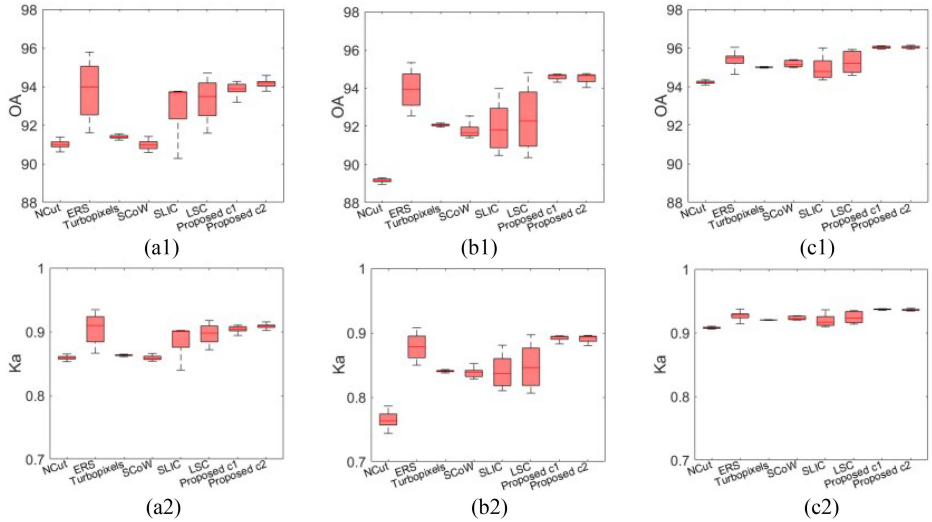

Fig. 18. Quantitative evaluation boxplot of classification results. (a1)–(c1) OA of Landsat 8, Sentinel-2, and SkySat. (a2)–(c2) Ka of Landsat 8, Sentinel-2, and SkySat. (a1) Landsat 8. (b1) Sentinel-2. (c1) SkySat. (a2) Landsat 8. (b2) Sentinel-2. (c2) SkySat.

TABLE III
ACCURACY ANALYSIS OF ABLATION STUDY

Images	Algorithms	Index						
		Recall	1-UE	EV	CO	ASA	OA	Ka
Landsat 8	Proposed_c1	0.98±0.00	0.89±0.00	0.995±0.00	0.47±0.10	0.95±0.00	93.89±0.17	0.90±0.01
	Proposed_c2	0.98±0.01	0.89±0.01	0.995±0.00	0.46±0.02	0.94±0.01	94.18±0.09	0.91±0.00
	Ablation	0.94±0.00	0.88±0.00	0.990±0.00	0.62±0.03	0.94±0.00	93.51±0.40	0.90±0.00
Sentinel-2	Proposed_c1	0.96±0.01	0.92±0.01	0.994±0.00	0.58±0.02	0.96±0.00	94.66±0.17	0.89±0.00
	Proposed_c2	0.96±0.01	0.92±0.01	0.993±0.00	0.39±0.03	0.96±0.00	94.66±0.32	0.89±0.00
	Ablation	0.87±0.01	0.97±0.00	0.993±0.00	0.68±0.00	0.95±0.00	92.89±0.20	0.88±0.00
Skysat	Proposed_c1	0.90±0.01	0.96±0.00	0.98±0.00	0.40±0.02	0.98±0.00	96.04±0.05	0.94±0.00
	Proposed_c2	0.90±0.01	0.96±0.00	0.974±0.00	0.42±0.04	0.98±0.00	96.03±0.06	0.94±0.00
	Ablation	0.76±0.02	0.95±0.00	0.980±0.00	0.52±0.03	0.98±0.00	95.38±0.10	0.93±0.01

dependence of classification results on the number of superpixels to a certain extent.

C. Sample Sensitivity Analysis

The number of samples is crucial during classifier training. To analyze the impact of different numbers of samples, the changes in accuracy from different perspectives with each algorithm are drawn in Fig. 21, where Fig. 21(a1)–(c1) shows the changes in OA of images with 30 m, 10 m, and 0.5 m resolution under different sample proportions, Fig. 21(a2)–(c2) shows the sample sensitivity rate of images with 30 m, 10 m, and 0.5 m resolution, and Fig. 21(a3)–(c3) shows the user and product accuracy of a small sample, i.e., bare land in 30 m, 10 m, and 0.5 m resolution images. For Fig. 21(a1)–(c1), it shows that the OA of each algorithm shows an increasing trend with the increase of sample proportion. Among them, the fastest increasing speed is ERS and the proposed algorithm, as shown in Fig. 21(a2) and (b2). In particular, although NCut has a higher sensitivity rate, its OA has always been low compared with other algorithms, as shown in Fig. 21(c1) and (c2). For further in-depth analysis, the accuracy of small samples is also demonstrated, as shown in Fig. 21(a3)–(c3). It shows that user accuracy consistently exceeds product accuracy. It illustrates that most pixels are missed. Compared with other algorithms, ERS and the proposed algorithm can achieve relatively high user and product accuracy simultaneously. It indicates that these algorithms have good balance and the classification results are closer to the real situation. For small samples, the accuracy can also reach over 75%.

D. Spectral Sensitivity Analysis

In the process of remote sensing image imaging, the spectrum of the image is often affected by atmosphere, lighting, and other factors, and spectral variation is ubiquitous [50]. To analyze the impact of spectral variation, the noise images with varying degrees of Gaussian white noise added are used to simulate the degree of variation of the spectrum. The OA of each algorithm and sensitivity rate are drawn in Fig. 22. It shows that the accuracy gradually decreases with the increase in noise variance. The most significant sensitivity is SLIC, next are ERS and LSC. NCut, TurboPixels, and SCoW have lower sensitivity. However, their classification accuracy is often lower than other algorithms. Through comprehensive analysis, it can be seen that the proposed algorithm can maintain low sensitivity to noise at a higher level of classification accuracy.

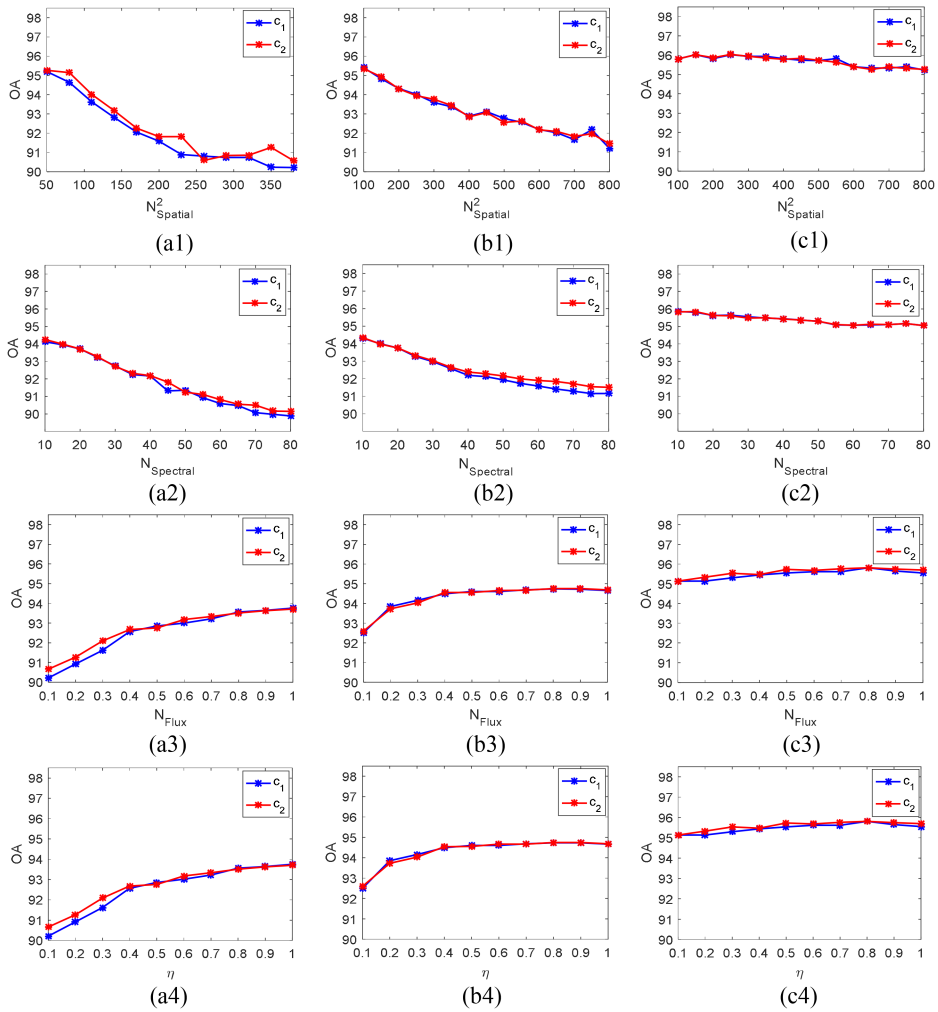

Fig. 19. Classification accuracy changing with parameters. (a1)–(c1) N_{Spatial} of Landsat 8, Sentinel-2, and SkySat. (a2)–(c2) N_{Spectral} of Landsat 8, Sentinel-2, and SkySat. (a3)–(c3) N_{Flux} of Landsat 8, Sentinel-2, and SkySat. (a4)–(c4) η of Landsat 8, Sentinel-2, and SkySat. (a1) Landsat 8. (b1) Sentinel-2. (c1) SkySat. (a2) Landsat 8. (b2) Sentinel-2. (c2) SkySat. (a3) Landsat 8. (b3) Sentinel-2. (c3) SkySat. (a4) Landsat 8. (b4) Sentinel-2. (c4) SkySat.

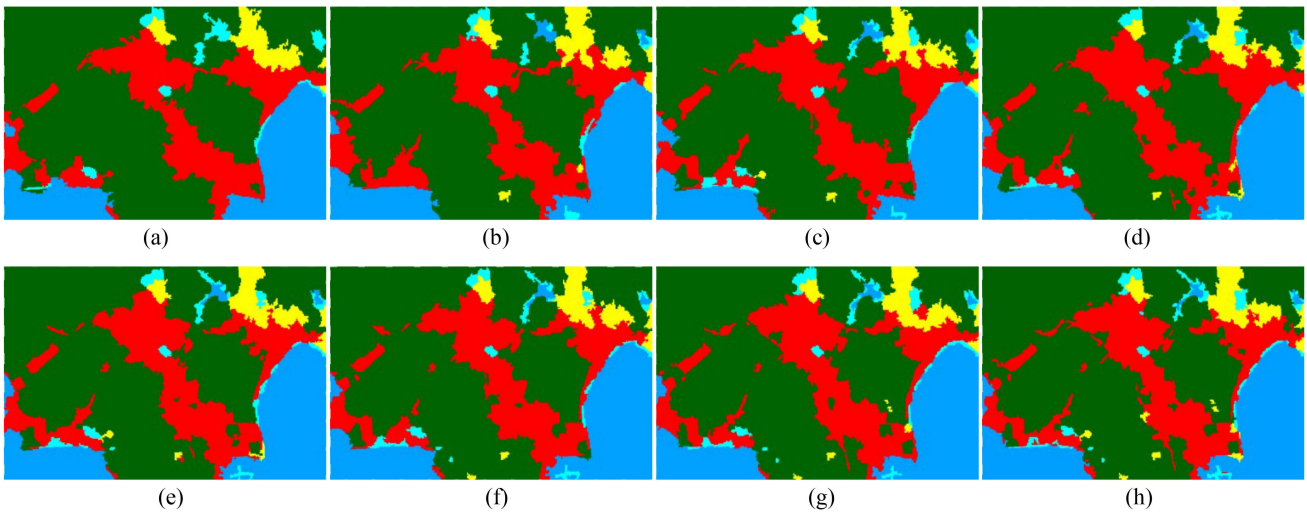

Fig. 20. Classification results with different numbers of superpixels. (a) $m = 200$ and OA $= 90.22$. (b) $m = 300$ and OA $= 90.89$. (c) $m = 400$ and OA $= 92.06$. (d) $m = 500$ and OA $= 92.83$. (e) $m = 600$ and OA $= 93.62$. (f) $m = 700$ and OA $= 93.63$. (g) $m = 900$ and OA $= 94.63$. (h) $m = 1200$ and OA $= 95.20$.

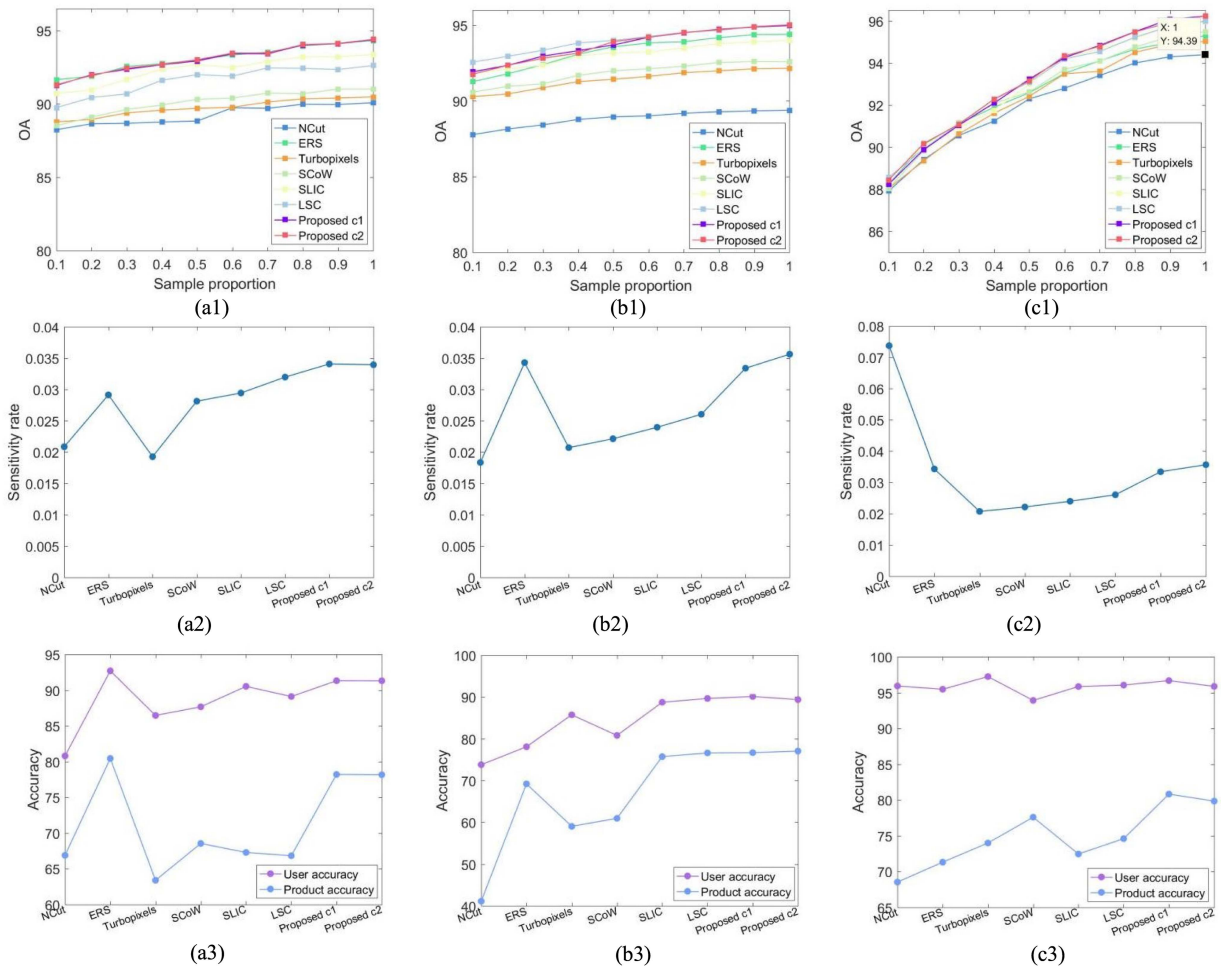

Fig. 21. Sample sensitivity analysis of each algorithm with different sample proportions. (a1)–(c1) Change in OA of 30, 10, and 0.5 m. (a2)–(c2) Sample sensitivity of 30, 10, and 0.5 m. (a3)–(c3) Small sample accuracy of 30, 10, and 0.5 m. (a1) 30 m. (b1) 10 m. (c1) 0.5 m. (a2) 30 m. (b2) 10 m. (c2) 0.5 m. (a3) 30 m. (b3) 10 m. (c3) 0.5 m.

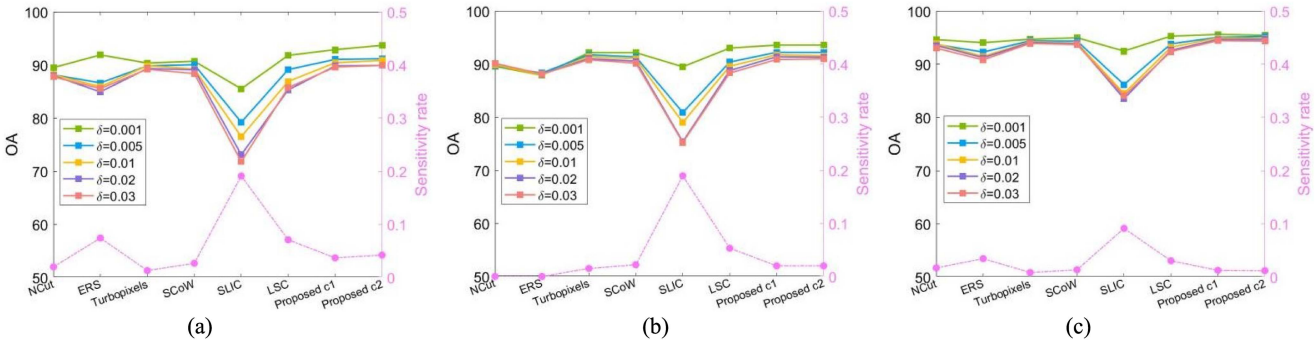

Fig. 22. Spectral sensitivity analysis of each algorithm with varying degrees of Gaussian white noise. (a) 30 m. (b) 10 m. (c) 0.5 m.

E. Time Complexity Analysis

Assuming that the number of pixels in the image is n, the proposed ADS algorithm uses double-deck loops. Same as the K-means algorithm, the time complexity of the outer loop is $O(nk\tau)$, where k is the number of superpixels, and τ is the number of iterations. The time complexity of the inner loop depends on the diffusion duration T, and it can be denoted by $O(T)$. Thus, the overall complexity of ADS is $O(nk\tau T)$ because k, τ, and T are far less than n, and ADS still can be regarded as the linear complexity, i.e., $O(n)$.

F. Remaining Challenges

The extensive experiments and analysis mentioned above indicate that the anisotropic diffusion model played an effective role in establishing boundary conditions during the clustering process, which promotes the flexible segmentation of objects

with different shapes and sizes under fixed scales. It develops a new idea for superpixel segmentation but still faces many challenges. For example, the current diffusion coefficient only considers spectral differences between pixels, which is insensitive in spectral similarity region. Focus on this problem, more features and local information should be considered in the next work. The gradient-based diffusion coefficient is currently not applicable to SAR images with inherent speckle noise. Focusing on this problem, the probability-based anisotropic diffusion model will be further studied for resistance to speckle noise.

VI. Conclusion

In this article, a novel superpixel segmentation algorithm based on an anisotropic diffusion model, named ADS, is proposed and applied to remote sensing image classification. Based on the theory of anisotropic diffusion, the diffusion flux of concentration is proposed to establish the boundary constraints during the clustering process. The concentration gradient controls the diffusion directions. The diffusion coefficient modeled by the gradient of the spectrum provides the possibility for controlling the diffusion speed at the boundary. The proposed algorithm can have high compactness in homogeneous regions and fit difficult boundaries in the complex scene without compactness limitations. The effectiveness is further verified in remote sensing image classification with three spatial resolutions, such as Landsat 8 with 30 m, Sentinel-2 with 10 m, and SkySat with 0.5 m. The experimental analysis results show that ADS can effectively improve the boundary-fitting deviation problem caused by the isotropic mechanism and increase the image classification accuracy. In the future, we will focus on the remaining challenges and further research on the modeling of diffusion coefficients to adapt to effective segmentation and classification of remote sensing images in different situations. For example, introducing antinoise mechanisms in the current model of diffusion flux would be suitable for SAR image processing.

Appendix
Diffusion Flux Calculation

In Fig. 3, taking pixel b as an example, its site (x, y) is (1, 2) assuming $\delta = 20$ and $\lambda = 1/8$; the diffusion coefficients c_1 between pixel b and other pixels are

$$c_1(b,a) = \frac{1}{1+\left(\frac{|122-40|}{20}\right)^2} = 0.06 \tag{20}$$

$$c_1(b,e) = \frac{1}{1+\left(\frac{|125-40|}{20}\right)^2} = 0.05 \tag{21}$$

$$c_1(b,f) = \frac{1}{1+\left(\frac{|120-40|}{20}\right)^2} = 0.06 \tag{22}$$

$$c_1(b,g) = \frac{1}{1+\left(\frac{|128-40|}{20}\right)^2} = 0.05 \tag{23}$$

$$c_1(b,c) = \frac{1}{1+\left(\frac{|32-40|}{20}\right)^2} = 0.86. \tag{24}$$

The diffusion distance $R(b, a) = R(b, f) = R(b, c) = 1$ and $R(b, e) = R(b, g) = \sqrt{2}$. According to Fig. 3(c) and (d), it shows that only the direction of pixel f satisfies diffusion conditions in the first diffusion in which the concentration gradient is positive, i.e., $\nabla \boldsymbol{U}(\boldsymbol{x}, \boldsymbol{y}, t) > 0$. If $\nabla \boldsymbol{U}(\boldsymbol{x}, \boldsymbol{y}, t) <= 0$, then $\Psi[\nabla \boldsymbol{U}(\boldsymbol{x}, \boldsymbol{y}, t)] = 0$; thus, pixels a, e, g, and c are ignored. Then, the first diffusion flux of pixel b is

$$\begin{aligned} U(1,2,1) &= U(1,2,2) + \frac{1}{\lambda}\left[\frac{1}{R(b,f)} \cdot c_1(b,f)\right. \\ &\quad \left. \cdot \nabla^{f_7} U(1,2,0) \cdot \Psi[\nabla^{f_7} U(1,2,0)]\right] \\ &= 0 + \frac{1}{8}[1 \cdot 0.06 \cdot (1-0) \cdot 1] = 0.01. \end{aligned} \tag{25}$$

In the second diffusion, there are four directions that can diffuse toward pixel b, i.e., pixels a, e, f, and g. Thus, the second diffusion flux of pixel b is

$$\begin{aligned} &U(1,2,2) = U(1,2,1) \\ &+ \frac{1}{\lambda}\left[\frac{1}{R(b,a)} \cdot c_1(b,a) \cdot \nabla^{f_5} U(1,2,1) \cdot \Psi\left[\nabla^{f_5} U(1,2,1)\right]\right. \\ &+ \frac{1}{R(b,e)} \cdot c_1(b,e) \cdot \nabla^{f_6} U(1,2,1) \cdot \Psi\left[\nabla^{f_6} U(1,2,1)\right] \\ &+ \frac{1}{R(b,f)} \cdot c_1(b,f) \cdot \nabla^{f_7} U(1,2,1) \cdot \Psi\left[\nabla^{f_7} U(1,2,1)\right] \\ &\left.+ \frac{1}{R(b,g)} \cdot c_1(b,g) \cdot \nabla^{f_8} U(1,2,1) \cdot \Psi\left[\nabla^{f_8} U(1,2,1)\right]\right] \\ &= 0.01 \\ &+ \frac{1}{8}\left[1 \cdot 0.06 \cdot (0.09 - 0.01) \cdot 1 + \frac{1}{\sqrt{2}} \cdot 0.05 \cdot (0.12 - 0.01) \cdot 1\right. \\ &\left.+1 \cdot 0.06 \cdot (1 - 0.01) \cdot 1 + \frac{1}{\sqrt{2}} \cdot 0.05 \cdot (0.11 - 0.01) \cdot 1\right] \\ &= 0.02. \end{aligned} \tag{26}$$

References

[1] P. Du, "Remote sensing image interpretation for urban environment analysis: Methods, system and examples," *Remote Sens.*, vol. 6, no. 10, pp. 9458–9474, Sep. 2014, doi: 10.3390/rs6109458.

[2] C. Zhang et al., "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, Oct. 2018, doi: 10.1016/j.rse.2018.06.034.

[3] S. Sun, S. Dustdar, R. Ranjan, G. Morgan, Y. Dong, and L. Wang, "Remote sensing image interpretation with semantic graph-based methods: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4544–4558, May 2022, doi: 10.1109/JSTARS.2022.3176612.

[4] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021, doi: 10.1109/TGRS.2020.3016820.

[5] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[7] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for infrared small object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 364–376, 2023, doi: 10.1109/TIP.2022.3228497.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[9] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L.Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolutional and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[10] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021, doi: 10.1109/TGRS.2020.3015157.

[11] Y. Ding et al., "Multi-scale receptive fields: Graph attention neural network for hyperspectral image classification," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119858.

[12] Z. Zhang et al., "Multireceptive field: An adaptive path aggregation graph neural framework for hyperspectral image classification," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119508, doi: 10.1016/j.eswa.2023.119508.

[13] Y. Ding et al., "Unsupervised self-correlated learning smoothy enhanced locality preserving graph convolution embedding clustering for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5536716, doi: 10.1109/TGRS.2022.3202865.

[14] Y. Ding et al., "AF2GNN: Graph convolution with adaptive filters and aggregator fusion for hyperspectral image classification," *Inf. Sci.*, vol. 602, pp. 201–219, Jul. 2022, doi: 10.1016/j.ins.2022.04.006.

[15] Z. Chen et al., "Global to local: A hierarchical detection algorithm for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5544915, doi: 10.1109/TGRS.2022.3225902.

[16] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259–272, Mar. 2012, doi: 10.1016/j.rse.2011.11.020.

[17] C. Wu et al., "Fuzzy SLIC: Fuzzy simple linear iterative clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 6, pp. 2114–2124, Jun. 2021, doi: 10.1109/TCSVT.2020.3019109.

[18] V. Machairas, M. Faessel, D. Cárdenas-Peña, T. Chabardes, T. Walter, and E. Decencière, "Waterpixels," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3707–3716, Nov. 2015, doi: 10.1109/TIP.2015.2451011.

[19] M. Wang, X. Liu, Y. Gao, X. Ma, and N. Q. Soomro, "Superpixel segmentation: A benchmark," *Signal Process., Image Commun.*, vol. 56, pp. 28–39, Aug. 2017, doi: 10.1016/j.image.2017.04.007.

[20] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000, doi: 10.1109/34.868688.

[21] L. Zhu, X. Kang, L. Ye, and A. Ming, "Explored normalized cut with random walk refining term for image segmentation," *IEEE Trans. Image Process.*, vol. 31, pp. 2893–2906, Mar. 2022, doi: 10.1109/TIP.2022.3162475.

[22] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104, doi: 10.1109/CVPR.2011.5995323.

[23] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Understanding*, vol. 166, pp. 1–27, Jan. 2018, doi: 10.1016/j.cviu.2017.03.007.

[24] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009, doi: 10.1109/TPAMI.2009.96.

[25] Z. Hu, Q. Zou, and Q. Li, "Watershed superpixel," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 349–353, doi: 10.1109/ICIP.2015.7350818.

[26] P. Márquez-Neila, L. Baumela, and L. Alvarez, "A morphological approach to curvature-based evolution of curves and surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 2–17, Jan. 2014, doi: 10.1109/TPAMI.2013.106.

[27] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[28] J. Chen, Z. Li, and B. Huang, "Linear spectral clustering superpixel," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3317–3330, Jul. 2017, doi: 10.1109/TIP.2017.2651389.

[29] R. Achanta and S. Süsstrunk, "Superpixels and polygons using simple non-iterative clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4895–4904, doi: 10.1109/CVPR.2017.520.

[30] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, and L. Shao, "Real-time superpixel segmentation by DBSCAN clustering algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5933–5942, Dec. 2016, doi: 10.1109/TIP.2016.2616302.

[31] J. Wang and X. Wang, "VCells: Simple and efficient superpixels using edge-weighted centroidal voronoi tessellations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1241–1247, Jun. 2012, doi: 10.1109/TPAMI.2012.47.

[32] J. Zhao, Y. Zhong, H. Shu, and L. Zhang, "High-resolution image classification integrating spectral-spatial-location cues by conditional random fields," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4033–4045, Sep. 2016, doi: 10.1109/TIP.2016.2577886.

[33] B. Chen, F. Qiu, B. Wu, and H. Du, "Image segmentation based on constrained spectral variance difference and edge penalty," *Remote Sens.*, vol. 7, no. 5, pp. 5980–6004, May 2015, doi: 10.3390/rs70505980.

[34] R. Qin and T. Liu, "A review of landcover classification with very-high resolution remotely sensed optical images—Analysis unit, model scalability and transferability," *Remote Sens.*, vol. 14, no. 3, Jan. 2022, Art. no. 646, doi: 10.3390/rs14030646.

[35] G. Gao, L. Zhao, J. Zhang, D. Zhou, and J. Huang, "A segmentation algorithm for SAR images based on the anisotropic heat diffusion equation," *Pattern Recognit.*, vol. 41, no. 10, pp. 3035–3043, Oct. 2008, doi: 10.1016/j.patcog.2008.01.029.

[36] M. Giacomini and S. Perotto, "Anisotropic mesh adaptation for region-based segmentation accounting for image spatial information," *Comput. Math. Appl.*, vol. 121, pp. 1–17, Sep. 2022, doi: 10.1016/j.camwa.2022.06.025.

[37] J. Yang, Z. Guo, D. Zhang, B. Wu, and S. Du, "An anisotropic diffusion system with nonlinear time-delay structure tensor for image enhancement and segmentation," *Comput. Math. Appl.*, vol. 107, pp. 29–44, Feb. 2022, doi: 10.1016/j.camwa.2021.12.005.

[38] Z. Guo, J. Sun, D. Zhang, and B. Wu, "Adaptive Perona–Malik model based on the variable exponent for image denoising," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 958–967, Mar. 2012, doi: 10.1109/TIP.2011.2169272.

[39] M. Yang et al., "Non-local means theory based Perona–Malik model for image denoising," *Neurocomputing*, vol. 120, pp. 262–267, Nov. 2013, doi: 10.1016/j.neucom.2012.08.063.

[40] Y. Wang, J. Guo, W. Chen, and W. Zhang, "Image denoising using modified Perona–Malik model based on directional Laplacian," *Signal Process.*, vol. 93, no. 9, pp. 2548–2558, Sep. 2013, doi: 10.1016/j.sigpro.2013.02.020.

[41] F. Voci, S. Eiho, N. Sugimoto, and H. Sekibuchi, "Estimating the gradient in the Perona-Malik equation," *IEEE Signal Process. Mag.*, vol. 21, no. 3, pp. 39–65, May 2004, doi: 10.1109/MSP.2004.1296541.

[42] G. Chen, "Image segmentation algorithm combined with regularized P-M denoising model and improved watershed algorithm," *J. Med. Imag. Health Inform.*, vol. 10, pp. 515–521, Feb. 2020, doi: 10.1166/jmihi.2020.2899.

[43] K. V. Rani and S. J. Jawhar, "Automatic segmentation and classification of lung tumour using advance sequential minimal optimisation techniques," *Inst. Eng. Technol. Image Process.*, vol. 14, no. 14, pp. 3355–3365, Oct. 2020, doi: 10.1049/iet-ipr.2020.0407.

[44] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade, "Distributed cosegmentation via submodular optimization on anisotropic diffusion," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 169–176, doi: 10.1109/ICCV.2011.6126239.

[45] X. Zhang, C. Xu, M. Li, and R. K. F. Teng, "Study of visual saliency detection via nonlocal anisotropic diffusion equation," *Pattern Recognit.*, vol. 48, no. 4, pp. 1315–1327, Apr. 2015, doi: 10.1016/j.patcog.2014.10.016.

[46] A. P. Witkin, "Scale space filtering," in *Proc. Int. Joint Conf. Artif. Intell.*, 1983, pp. 1019–1021.

[47] K. Shi, "Image denoising by nonlinear nonlocal diffusion equations," *J. Comput. Appl. Math.*, vol. 395, Oct. 2021, Art. no. 113605, doi: 10.1016/j.cam.2021.113605.

[48] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1260–1270, Nov. 2002, doi: 10.1109/TIP.2002.804276.
[49] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990, doi: 10.1109/34.56205.
[50] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019, doi: 10.1109/TIP.2018.2878958.

Xiaoli Li received the Ph.D. degree in photogrammetry and remote sensing from Liaoning Technical University, Fuxin, China, in 2019.

From 2019 to 2021, she was a Postdoctoral Researcher with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT, CAS). She is an Assistant Professor with SIAT, CAS, Shenzhen, China. Her research interests include remote-sensing image modeling and interpretation methods based on fuzzy theory, stochastic geometry, and information entropy.

Jinsong Chen received the Ph.D. degree in cartography and geographic information system from the Chinese Academy of Sciences, Beijing, China, in 2004.

From 2004 to 2011, he was a Postdoctoral Researcher and a Professor with The Chinese University of Hong Kong, Hong Kong. He is a Research Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include spatial and environmental big data-processing methods and multisource spatial and environmental data assimilation and information fusion methods.

Longlong Zhao received the Ph.D. degree in quaternary geology from the China University of Geosciences, Beijing, China, in 2017.

She was a Postdoctoral Fellow with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, from 2017 to 2019. She was an Associate Researcher with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. Her research interests include remote-sensing images processing, remote-sensing applications in agriculture, and ecological environment.

Hongzhong Li received the B.S. and M.S. degrees from Shandong University, Jinan, China, and the Ph.D. degree in cartography and geographic information system from the Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently an Associate Researcher with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests include polarimetric radar remote sensing and remote-sensing monitoring of the ecological environment.

Jin Wang received the B.S. and M.S. degrees from South China Normal University, Guangzhou, China, in 2011 and 2014, respectively, and the Ph.D. degree in environmental science from the University of Chinese Academy of Sciences, Beijing, China, in 2018.

He is currently a Research Professor with the School of Geography, South China Normal University. His research interests include remote sensing of urban environment, quantitative and spatial analysis of coastal areas, and land use/land cover.

Luyi Sun received the B.S. and M.Sc. degrees from the Beijing Institute of Technology, Beijing, China, in 2009 and 2012, respectively, and the Ph.D. degree in space and climate physics from University College London, London, U.K., in 2017.

She is currently with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Her research endeavors encompass the monitoring of Earth's surface deformations through the utilization of microwave remote sensing. Furthermore, she is involved in the investigation of spatiotemporal alterations in geographical features, facilitated by the integration of synthetic aperture radar and multispectral imagery.

Shanxin Guo received the master's and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2011 and 2015, respectively.

He joined a visiting Ph.D. program with the Department of Geography, University of Wisconsin Madison, Madison, WI, USA. His current research interests include digital soil mapping with remote sensing data, deep learning for land cover and land use mapping, and the Gaussian process for forest traits mapping.

Pan Chen received the Ph.D. degree in cartography and geography information system from Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2023.

He is currently a Postdoctoral Researcher with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. His research interests include remote sensing image processing, deep learning, semantic segmentation, and change detection.

Xuemei Zhao received the Ph.D. degree in photogrammetry and remote sensing from Liaoning Technical University, Fuxin, China, in 2017.

From 2017 to 2019, she was a Postdoctoral Researcher with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China. She is an Associate Professor with the Guilin University of Electronic Technology, Guilin, China. Her research interests include remote-sensing image modeling and interpretation methods based on deep learning, and information geometry.