ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.657389

RESEARCH ARTICLE

Total Variation Diffusion-Guided Fuzzy Active Contour Model for Noisy Image Segmentation

Meng Zhang[1,2], Yi Yang[1,2,*], Sixian Zhang[1,2], Pengbo Mi[1,2] and Deqiang Han[3]

[1] SKLSVMS, Xi'an Jiaotong University, Xi'an 710049, China
[2] School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3] School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Abstract

Image segmentation is an important task in computer vision and plays a critical role in many fields. Fuzzy Active Contour Model (FACM) has been widely applied in image segmentation because it can handle complex shape changes. However, it is difficult for current FACMs to obtain ideal performance when segmenting noisy images. Therefore, this paper proposes a Total Variation Diffusion-Guided Fuzzy Active Contour Model (TVDGFACM), which formulates noisy image segmentation as a hierarchical fusion process. Specifically, this model introduces total variation and adaptively fuses anisotropic and isotropic diffusion mechanisms to suppress noise interference while preserving image edges. Moreover, TVDGFACM fuses local intensity evidence according to regional dependencies and noise-aware pixel reliability, which helps adjust pixel memberships to generate a smooth and coherent segmentation result. Furthermore, a balance factor is constructed based on noise density, which is used to fuse global and local intensity information, thereby ensuring segmentation efficiency and stability. Experiments on synthetic, natural, remote sensing images, and large-scale benchmark dataset show that TVDGFACM is rational and effective.

Keywords: image segmentation, fuzzy active contour model, noisy images, total variation, regional dependencies.

1 Introduction

Image segmentation is a fundamental and important task in computer vision, aiming to divide an image into disjoint regions [1–3]. Pixels within the same region typically exhibit similar features, whereas pixels in different regions show significant differences [4, 5]. In recent years, image segmentation has been widely applied in various fields, such as the scene understanding, object tracking, and remote sensing image analysis [6].

Among existing image segmentation methods, Fuzzy Active Contour Model (FACM) is widely used because it can handle complex shape changes [7]. FACM employs a closed and smooth evolving curve to divide

Academic Editor:
Weiping Ding

Submitted: 12 September 2025
Accepted: 11 May 2026
Published: 29 May 2026

Vol. 3, **No.** 2, 2026.
10.62762/CJIF.2025.657389

***Corresponding author:**
✉ Yi Yang
jiafeiyy@mail.xjtu.edu.cn

Citation
Zhang, M., Yang, Y., Zhang, S., Mi, P., & Han, D. (2026). Total Variation Diffusion-Guided Fuzzy Active Contour Model for Noisy Image Segmentation. *Chinese Journal of Information Fusion*, 3(2), 93–124.

an image into object and background regions, where the 0.5 level set of the membership function is used to represent the evolving curve. Existing FACMs can generally be categorized into two types. The first type is FACMs based on global intensity information. For example, Krinidis and Chatzis [8] proposed a fuzzy energy-based active contour model. Wu et al. [9] proposed a fuzzy active contour model with kernel metric. These models make segmentation decisions according to the intensity values of individual pixels. However, noise causes random fluctuations in pixel intensities, so it is difficult for FACMs based on global intensity information to accurately segment noisy images.

The second type is FACMs based on local intensity information. For example, Krinidis, S. and Krinidis, M. [10] proposed a fuzzy energy-based active contour model exploiting local information. Mondal et al. [11] proposed a global and local fuzzy energy based active contour model. Fang et al. [12] proposed a region-edge-based active contour model driven by hybrid and local fuzzy region based energy. These models fuse local intensity information by measuring the different contributions (i.e., excitation weights) of neighboring pixels to the membership of the center pixel, thereby smoothing pixel memberships and reducing the impact of noise. Existing excitation weight estimation methods include methods based on relative spatial distance and methods based on intensity similarity. However, it is difficult for them to accurately measure the noise distributions and the degrees to which pixels are polluted by noise. This easily leads to estimation bias and affects segmentation performance.

Furthermore, as the noise density increases, the intensity means of different regions in an image tend to approach, and the intensity variances significantly increase. This makes their intensity distribution patterns become approximately identical and lack clearly separable peaks. To suppress noise interference, Zhang et al. [13] introduced total variation into active contour model. Total variation is a representative image denoising method, which can transform a noisy image into a diffusion domain to restore its intensity distribution characteristics [14]. The anisotropic diffusion mechanism used in traditional total variation can protect image edges to some extent. However, this diffusion mechanism is prone to misinterpreting noise in homogeneous regions as true edges [15], which will lead to staircase artifacts and reduce segmentation accuracy. Therefore, it is desirable to construct an energy function that combines different diffusion mechanisms in a data-driven manner, so that the denoising performance and boundary preservation capability can be jointly optimized in an information fusion framework.

From an information fusion viewpoint, noisy image segmentation can be regarded as a problem of integrating complementary and uncertain information. Global intensity statistics provide coarse region-level discrimination, local neighborhood information reflects spatial contextual consistency, and diffusion-based structural information contributes to noise suppression and boundary preservation. However, these information sources have different reliability under different noise densities. Therefore, an effective segmentation model should not simply use them independently, but should adaptively fuse them within a unified energy minimization framework.

To achieve the above purposes, this paper proposes a Total Variation Diffusion-Guided Fuzzy Active Contour Model (TVDGFACM). This model introduces total variation into FACM, and constructs an energy function based on global and local intensity information. The main contributions in this paper are as follows:

1. A diffusion-mechanism fusion strategy is designed within the total variational framework. It adaptively integrates anisotropic diffusion and isotropic diffusion according to structural information, thereby jointly optimizing boundary preservation and noise suppression. This helps overcome the drawback that traditional total variation easily produces staircase artifacts in homogeneous regions.

2. A noise-aware local contextual information fusion strategy is constructed. Specifically, local structure entropy is used to characterize pixel reliability, and the excitation weight estimation method is reconstructed according to the regional dependencies between noisy pixels and their neighboring pixels. Compared with existing estimation methods, the reconstructed method can more reliably fuse local intensity information by measuring noise distributions and the degrees to which pixels are polluted by noise, thereby helping to generate a smooth and coherent segmentation result.

3. A global-local energy fusion strategy is developed based on noise density. The balance factor

adaptively adjusts the contribution of the global energy term and local energy term, thereby achieving a balance between computational efficiency and segmentation stability under different noise conditions.

Experiments on synthetic, natural, remote sensing images, and large-scale benchmark datasets show that TVDGFACM outperforms most existing active contour models and deep learning methods. Moreover, TVDGFACM exhibits stronger noise resistance, spatial smoothness, and boundary preservation capabilities.

The rest of this paper is organized as follows. Section 2 introduces preliminaries. Section 3 presents the TVDGFACM's overall framework and its main components in detail. Section 4 verifies TVDGFACM using synthetic, natural, remote sensing images, and large-scale benchmark dataset. Section 5 draws conclusions.

2 Preliminaries

Image segmentation is a classic task in computer vision. It aims to classify each pixel into a specific region based on features such as intensity information. Current image segmentation methods include thresholding [16], edge detection [17], clustering [18], and Fuzzy Active Contour Model (FACM) [7]. Among them, FACM stands out due to its convex energy function design and its ease in handling complex shape changes. However, existing FACMs are susceptible to the noise density and have difficulty in accurately estimating the excitation weights. Therefore, it is difficult for current FACMs to achieve satisfactory performance when segmenting noisy images.

2.1 Basics of Image Segmentation

The basic problem of image segmentation is to divide an image domain Ω into N disjoint regions $\Omega_1, \Omega_2, \ldots, \Omega_N$, and satisfies the following conditions:

1) $\bigcup_{i=1}^{N} \Omega_i = \Omega$;

2) Ω_i is a connected set, $i = 1, 2, \ldots, N$;

3) For any i and j, $i \neq j$, $\Omega_i \cap \Omega_j = \varnothing$;

4) $Q(\Omega_i) =$ TRUE, $i = 1, 2, \ldots, N$, i.e., the pixels in the region Ω_i must satisfy a predefined logical property Q;

5) $Q(\Omega_i \cup \Omega_j) =$ FALSE, where Ω_i and Ω_j are two adjacent regions, i.e., adjacent regions should not satisfy the same logical property Q.

In the above conditions, $\varnothing$ represents the empty set. Q is a composite expression. When the intensity mean of region Ω_i is less than μ_i and its standard deviation is less than σ_i, $Q(\Omega_i) =$ TRUE is satisfied.

2.2 Basics of Fuzzy Active Contour Model

FACM uses the membership function $u(x)$ to represent a closed and smooth evolving curve C. It guides the curve evolution based on an energy function E. The evolving curve always moves in the direction that minimizes the energy function E. In other words, the smaller the energy function, the better the segmentation result. Let the given image be $I(x) : x \rightarrow \Omega$, where x is a point in the bounded open region Ω. The evolving curve can divide an image $I(x)$ into the inside region $\Omega_1 = \text{inside}(C)$ and outside region $\Omega_2 = \text{outside}(C)$ of the evolving curve. The correspondence between different regions and the membership functions can be expressed as:

$$\begin{cases} C & = \{x \in \Omega : u(x) = 0.5\} \\ \text{inside}(C) & = \{x \in \Omega : u(x) > 0.5\} \\ \text{outside}(C) & = \{x \in \Omega : u(x) < 0.5\} \end{cases} \tag{1}$$

2.2.1 *Existing Fuzzy Active Contour Model*

Fuzzy Energy-based Active Contour Model (FEBACM) [8]: This model innovatively uses the 0.5 level set of the membership function to represent the evolving curve. Its energy function can be formulated as:

$$\begin{aligned} E_{FEBACM} = & \int [u(x)]^m (I(x) - c_1)^2 dx \\ & + \int [1 - u(x)]^m (I(x) - c_2)^2 dx \end{aligned} \tag{2}$$

where $u(x)$ denotes the membership function. $I(x)$ is the image to be segmented. x is the pixel coordinates. m represents the weighting exponent, which is usually set to 2. c_1 and c_2 are the intensity means of inside regions and outside region of the evolving curve, respectively.

As can be seen from Eq. (2), FEBACM constructs a convex energy function using global intensity information. It can obtain the global optimal solution and enhance the robustness to the initial evolving curve. However, FEBACM assumes that the given image is composed of two homogeneous regions. However, noise will destroy this homogeneity

assumption, so it is difficult for FEBACM to accurately segment noisy images.

Fuzzy Energy-based Active Contour Model Exploiting Local Information (LFEBACM) [10]: This model introduces local intensity information into Eq. (2) to enhance its robustness to noise, as follows:

$$\begin{aligned} E_{LFEBACM} = & \int [u(x)]^m (I(x) - c_1)^2 dx \\ & + \int [1 - u(x)]^m (I(x) - c_2)^2 dx \\ & + \int [u(x)]^m \int w_{xy} [1 - u(y)]^m (I(y) - c_1)^2 dy dx \\ & + \int [1 - u(x)]^m \int w_{xy} [u(y)]^m (I(y) - c_2)^2 dy dx \end{aligned} \tag{3}$$

where $I(y)$ represents the intensity value of neighboring pixel y in the local region centered at the pixel x. w_{xy} is the excitation weight, which is used to measure the different contributions of neighboring pixel y to the membership $u(x)$ of the center pixel x.

The last two terms in Eq. (3) are the local energy term. It uses the excitation weight w_{xy} to measure the dependencies between the center pixel x and its neighboring pixel y. The larger the excitation weight of neighboring pixel y, the higher the possibility that this pixel belongs to the same class as the center pixel x. The greater the contribution of neighboring pixel y to the membership $u(x)$ of the center pixel x. When the center pixel x is polluted by noise, then the neighboring pixel y that is not polluted by noise will be assigned a larger excitation weight w_{xy}. This helps to adjust the membership $u(x)$ of the center pixel x to reduce the impact of noise. Therefore, the excitation weight estimation method is a key to constructing the local energy term. LFEBACM uses the relative spatial distance between pixels to estimate the excitation weight, which can be formulated as:

$$w_{xy} = \frac{1}{1 + d_{xy}} \tag{4}$$

where d_{xy} represents the spatial Euclidean distance between pixels x and y.

Global and Local Fuzzy Energy-based Active Contour Model (GLFEBAM) [11]: This model has the same energy function form as LFEBACM. It introduces the intensity similarity between pixels into Eq. (4), as follows:

$$w_{xy} = \frac{1}{1 + d_{xy}} \cdot \exp\left(- \frac{(I(x) - I(y))^2}{\sum_{y \in N_x} (I(x) - I(y))^2} \right) \tag{5}$$

where N_x represents the neighboring window centered at the pixel x.

2.2.2 *Drawbacks of Existing Fuzzy Active Contour Model*

The above FACMs have been applied to noisy image segmentation. However, they rely heavily on the intensity distribution of different regions to construct the energy function. Noise destroys intensity statistical characteristics. This makes it difficult for current FACMs to achieve satisfactory segmentation performance.

Figure 1. Test image.

This section uses Figure 1 to illustrate the impact of noise on the intensity distribution characteristics. When the test image is not polluted by noise, the intensity distributions of the object and background regions present a clear bimodal pattern, as shown in Figure 2 (a). In this case, an energy function constructed based on the intensity domain can effectively guide curve evolution. However, when the test image is polluted by noise, the random fluctuations of noise destroy the original intensity distributions. As a result, the intensity means of the object and background regions become closer, and the intensity variances increase significantly. This causes their intensity distributions overlap, as shown in Figure 2 (b). As the noise density further increases, the statistical characteristics of intensity distribution become increasingly blurred and may even degenerate into an approximately unimodal distribution, as shown in Figure 2 (c). Therefore, under the impact of high

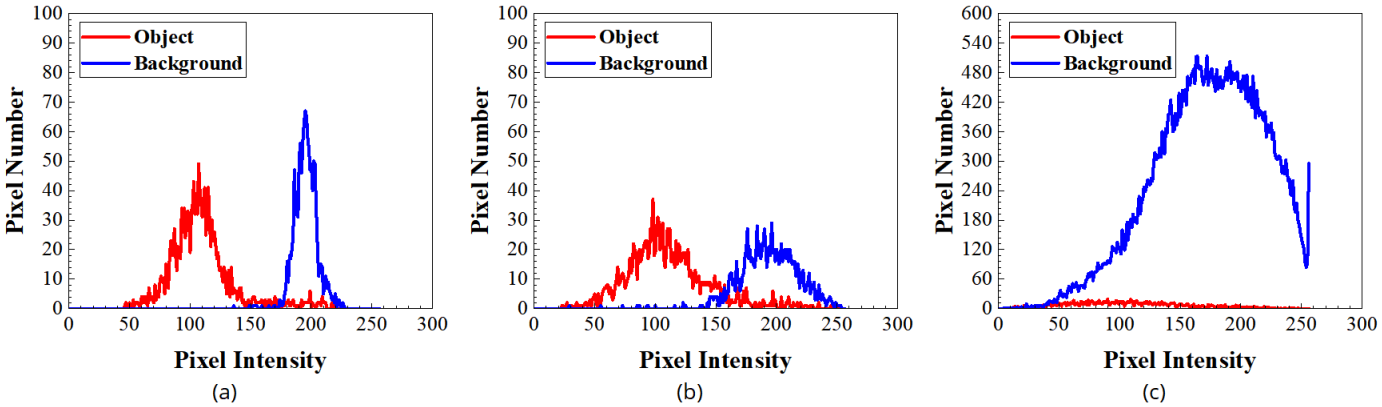


Figure 2. Impact of noise on the intensity distribution characteristics. (a) Without noise; (b) 1% mixed noise; (c) 5% mixed noise.

noise, it is difficult for FACMs that rely on the intensity distribution of different regions to construct the energy functions to achieve ideal performance.

To reduce the impact of noise and enhance noise resistance, Zhang et al. [13] introduced total variation into the active contour model for collaborative optimization, which helps to achieve both image denoising and segmentation. Total variation is a classic image denoising method, and its energy function can be formulated as:

$$E_{TV} = \int (I(x) - I_{TV}(x))^2 dx + \gamma \int |\nabla I_{TV}(x)| dx \quad (6)$$

where $I_{TV}(x)$ represents the corrected image obtained by total variation. γ denotes the Lagrange multiplier. ∇ is the derivative operator.

The first term in Eq. (6) is the data fidelity term, which is primarily used to prevent image distortion. The second term in Eq. (6) is the regularization term, which enforces smoothness of the corrected image. Total variation usually uses the $L1$ norm to penalize the gradient magnitude. Its optimization effect can be understood as promoting a sparse distribution of gradients. Taking the variation for E_{TV} yields its Euler-Lagrange equation:

$$2(I(x) - I_{TV}(x)) - \gamma \nabla \cdot \left(\frac{\nabla I_{TV}(x)}{|\nabla I_{TV}(x)|} \right) \quad (7)$$

The corresponding gradient descent flow can be expressed as:

$$\frac{\partial I_{TV}(x)}{\partial t} = -2(I(x) - I_{TV}(x)) + \gamma \nabla \cdot \left(\frac{\nabla I_{TV}(x)}{|\nabla I_{TV}(x)|} \right) \quad (8)$$

The diffusion function $\nabla \cdot (\nabla I_{TV}(x)/|\nabla I_{TV}(x)|)$ can be viewed as an edge-adaptive anisotropic diffusion:

- When $|\nabla I_{TV}(x)|$ is small, the normalization factor $1/|\nabla I_{TV}(x)|$ makes the diffusion effect relatively stronger, thereby quickly suppressing the fine oscillations caused by noise in the homogeneous regions.
- When $|\nabla I_{TV}(x)|$ is large, the normalization factor $1/|\nabla I_{TV}(x)|$ makes diffusion relatively weaker, thereby suppressing the smoothing effect to preserve image edges and other detail information.

From the above description, the total variation constructed based on the $L1$ norm with the anisotropic diffusion mechanism helps to protect image edges, which makes it widely used in image denoising. However, it may still misinterpret some noise as edges and retain them, thereby inducing staircase artifacts. The root cause is that the total variational regularization term relies solely on the gradient magnitude $|\nabla I_{TV}(x)|$, which makes it difficult to effectively distinguish between structural edges and local gradient spikes caused by random noise.

Specifically, noise may create large instantaneous gradients in local regions, which makes data fidelity term $(I(x) - I_{TV}(x))^2$ tend to retain a certain intensity variation at these locations. Furthermore, the regularization term constructed based on the $L1$ norm tends to concentrate intensity changes in fewer locations rather than transitioning slowly over a wider range with smaller gradients. Therefore, during energy minimization, if a trade-off is to be made between maintaining consistency with $I(x)$ and reducing the regularization term, the total variation usually chooses a small number of jumps to explain the local intensity differences. This makes it easier for the local spikes caused by noise to be solidified into edge-like jumps, thereby triggering staircase artifacts.

Therefore, directly introducing total variation into FACM will affect segmentation accuracy and stability.

In addition, existing FACMs usually use relative spatial distance or intensity similarity to estimate excitation weights. Although these methods are simple in structure and easy to implement, they still have some limitations. This section analyzes these limitations using two local regions, N_1 and N_2, in a noisy image shown in Figure 3 (a).

Specifically, methods based on relative spatial distance between pixels assign larger excitation weights to neighboring pixels that are closer to the center pixel, and smaller excitation weights to those that are farther away. However, the relative spatial distance between pixels is fixed. This makes it difficult for methods based on relative spatial distance to dynamically adjust the excitation weights according to the noise distribution.

As shown in Figure 3 (b), in the local region N_1, the center pixel and neighboring pixel A are not polluted by noise, while neighboring pixels B and C are polluted, with pixel B being more severely affected than pixel C. In this case, the excitation weight of pixel A should be larger than those of pixel B and C, and the excitation weight of pixel B should be smaller than that of pixel C. However, as shown in Figure 3 (d), the excitation weights obtained according to the relative spatial distance between pixels do not conform to the above analysis.

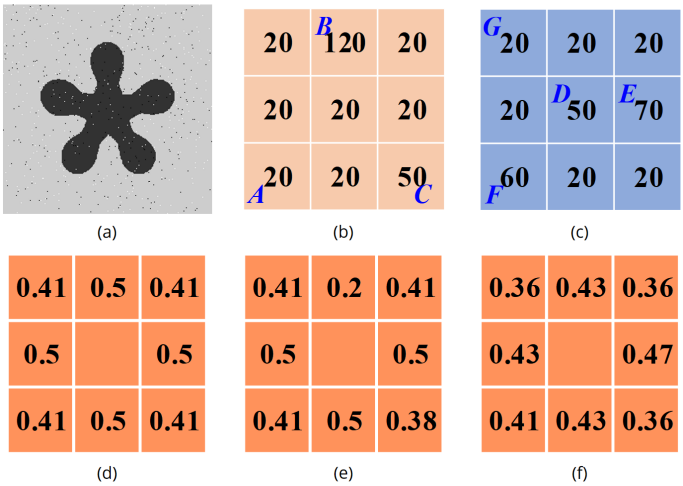


Figure 3. Excitation weights obtained by existing FACM. (a) Noisy image; (b) Local region N_1; (c) Local region N_2; (d) Excitation weights of pixels in the local regions N_1 and N_2 calculated using Eq. (4); (e) Excitation weights of pixels in the local region N_1 calculated using Eq. (5); (f) Excitation weights of pixels in the local region N_2 calculated using Eq. (5).

In contrast, methods based on intensity similarity between pixels can address this problem to some extent, as shown in Figure 3 (e). Such methods assign larger excitation weights to neighboring pixels whose intensity values are closer to that of the center pixel, and smaller weights otherwise. However, since the intensity value of an individual pixel ignores spatial context information, it is difficult for methods based on intensity similarity to effectively measure whether a pixel is polluted by noise and to what degree it is affected.

As shown in Figure 3 (c), in the local region N_2, the center pixel D and neighboring pixel E and F are polluted by noise, whereas neighboring pixel G is not. In this case, the excitation weight of pixel G should be larger than those of pixels E and F, and the excitation weights of pixel E and F should be smaller than those of other neighboring pixels. Since pixel E is more severely polluted by noise than pixel F, the excitation weight of pixel E should be smaller than that of pixel F. However, as shown in Figure 3 (f), the excitation weights obtained according to the intensity similarity between pixels do not conform to the above analysis.

3 Total Variation Diffusion-Guided Fuzzy Active Contour Model

To enhance the anti-noise capability and optimize the excitation weight estimation method, this section proposes a Total Variation Diffusion-Guided Fuzzy Active Contour Model (TVDGFACM), which fuses structural diffusion information, global intensity information, and local contextual information. As shown in Figure 4, it consists of three stages: image denoising, energy function construction and energy function minimization.

1. In the image denoising stage, TVDGFACM constructs a total variation diffusion-guided energy term to obtain the corrected image, which uses nonlinear structural tensors to identify image edges and homogeneous regions, thereby guiding total variation to execute different diffusion mechanisms to achieve both boundary preservation and noise suppression. This design retains the advantage that total variation can protect image edges while overcoming its drawback that traditional total variation easily causes staircase artifacts in homogeneous regions due to misinterpreting noise for true edges.

2. In the energy function construction stage, TVDGFACM first uses the intensity information of the corrected image in the inside and outside

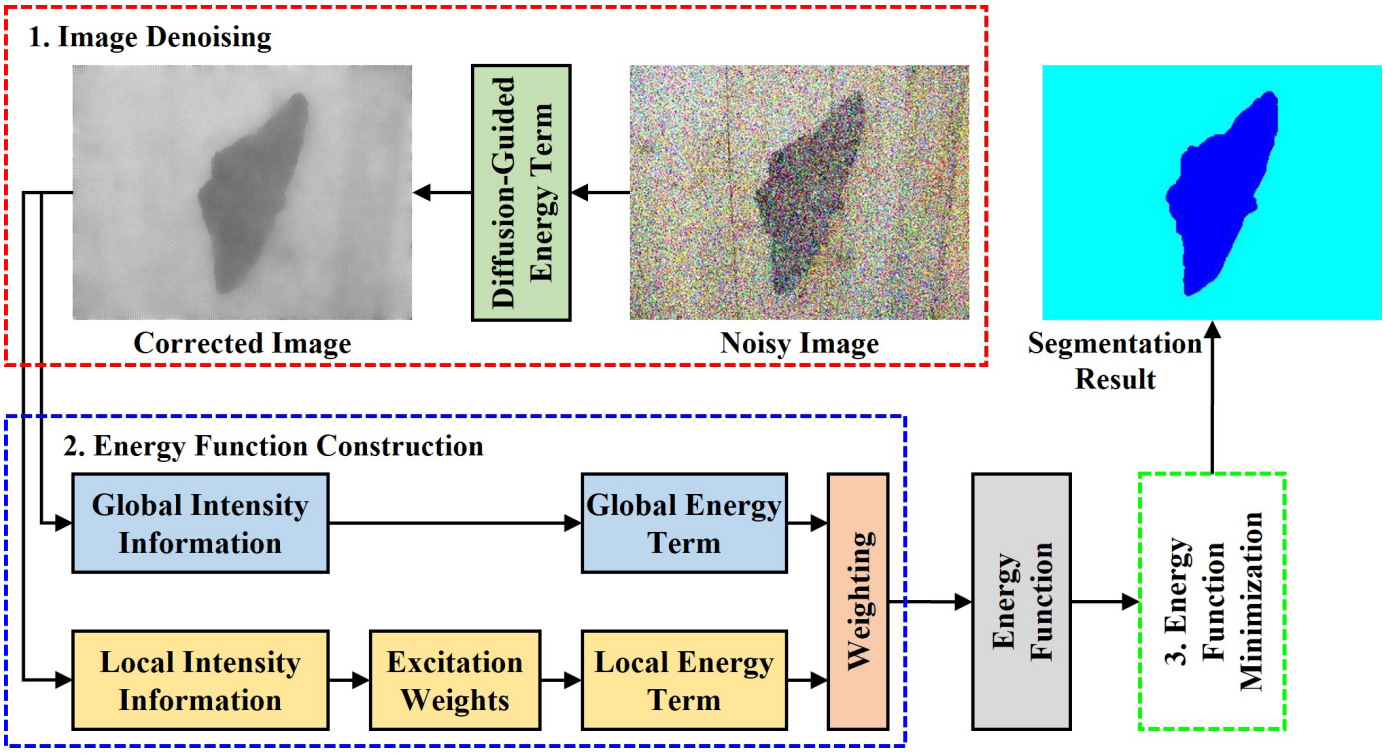


Figure 4. TVDGFACM's overall framework.

regions of the evolving curve to construct a global energy term. Then, TVDGFACM designs a local structure entropy to detect noisy pixels, and reconstructs the excitation weight estimation method based on the regional dependencies between noisy pixels and their neighboring pixels. Different from existing estimation methods, the reconstructed method can measure the noise distributions and the degrees to which pixels are polluted by noise. Finally, a local energy term is built using the reconstructed excitation weight estimation method and local intensity information.

3. In the energy function minimization stage, TVDGFACM obtains the update equations for each variable by solving the derivatives of energy function with respect to these variables. Furthermore, the energy function is then minimized through iteration until convergence is achieved.

3.1 Total Variation Diffusion-Guided Energy Term

In the first stage shown in Figure 4, to remove intensity statistical characteristics while avoiding staircase artifacts, this section designs a total variation diffusion-guided energy term based on the total variation, which can adaptively fuse anisotropic and isotropic diffusion mechanisms. This energy term can be formulated as:

$$E_D = \int (I(x) - I_C(x))^2 dx + \frac{\gamma}{\theta(x)} \int |\nabla I_C(x)|^{\theta(x)} dx \tag{9}$$

where $I_C(x)$ is the corrected image obtained by the total variation diffusion-guided energy term. $\theta(x)$ is the adaptive structure detector. Its core logic and implementation mechanism are elaborated as follows:

- When pixel x lies on an image edge, $\theta(x) = 1$. In this case, the total variation diffusion-guided energy term uses the $L1$ norm with the anisotropic diffusion mechanism to penalize the gradient magnitude, thereby helping to preserve image edge and other detail information.
- When pixel x lies in a homogeneous region, $\theta(x) = 2$. In this case, the total variation diffusion-guided energy term uses the $L2$ norm with the isotropic diffusion mechanism to penalize the gradient magnitude. Compared with the $L1$ norm, the $L2$ norm does not depend on the gradient direction. It can uniformly smooth high-frequency components in all directions. Therefore, the $L2$ norm can effectively remove noise in homogeneous regions without causing staircase artifacts.

To achieve the above purpose, this section uses the nonlinear structure tensors to identify image edges and homogeneous regions. The nonlinear structure tensor for the pixel x can be expressed as:

$$T(x) = \begin{bmatrix} T_{11}(x) & T_{12}(x) \\ T_{21}(x) & T_{22}(x) \end{bmatrix} \tag{10}$$

where $T_{mn}(x)$ is the balanced forward-backward diffusion of linear structure tensor $U_{mn}(x)$, as follows:

$$T_{mn}(x) = div\left(g\left(\|\nabla U_{mn}(x)\|^2\right)\right) \cdot \nabla U_{mn}(x) \tag{11}$$

where $div(\cdot)$ represents the divergence operator. $g(x) = 1/(\|x\|^2 + \varepsilon^2)$ is the diffusion function. ε denotes a non-zero minimum to avoid the denominator being zero.

As can be seen from Eq. (10), $T(x)$ is a positive semi-definite matrix. It has two non-negative eigenvalues $\lambda_1(x)$ and $\lambda_2(x)$, which correspond to the strength of the principal and secondary gradient directions in a local region, respectively. When $\lambda_1(x) \approx \lambda_2(x) \approx 0$, the gradient near pixel x is weak and there is no dominant gradient direction. This indicates that pixel x lies in a homogeneous region. When $\lambda_1(x) \gg \lambda_2(x) \approx 0$, pixel x has a single dominant gradient direction. This indicates that pixel x lies on an image edge. According to this property, we define a direction consistency metric $D(x) = (\lambda_1(x) - \lambda_2(x))^2$. When pixel x lies in a homogeneous region, $D(x)$ is close to 0. On the contrary, when the pixel x lies on an image edge, $D(x)$ takes a large value. Based on the direction consistency metric, we define an adaptive structure detector, which can be formulated as:

$$\theta(x) = \text{Round}\left(2 - \exp\left(-\frac{1 - D_N(x)}{D_N(x) + \epsilon^2}\right)\right) \tag{12}$$

where Round$(\cdot)$ represents the rounding function. $D_N(x)$ denotes the normalized direction consistency metric.

Based on the above descriptions, the total variation diffusion-guided energy term can drive different diffusion mechanisms by identifying image edges and homogeneous regions. It transforms noisy images from intensity domain to a total variation diffusion-guided domain, thereby restoring intensity distribution characteristics while avoiding staircase artifacts. Compared with the diffusion domain obtained by traditional total variation, the total variation diffusion-guided domain has more favorable intensity distribution characteristics. As shown in Figure 5, under the impact of 5% mixed noise, it is difficult for the corrected image obtained by traditional total variation to smooth homogeneous regions, and its intensity distribution still contains substantial overlap between regions. In contrast, the corrected image obtained by the total variation diffusion-guided energy term can effectively remove noise while protecting image edges, and its intensity distributions are almost completely separated.

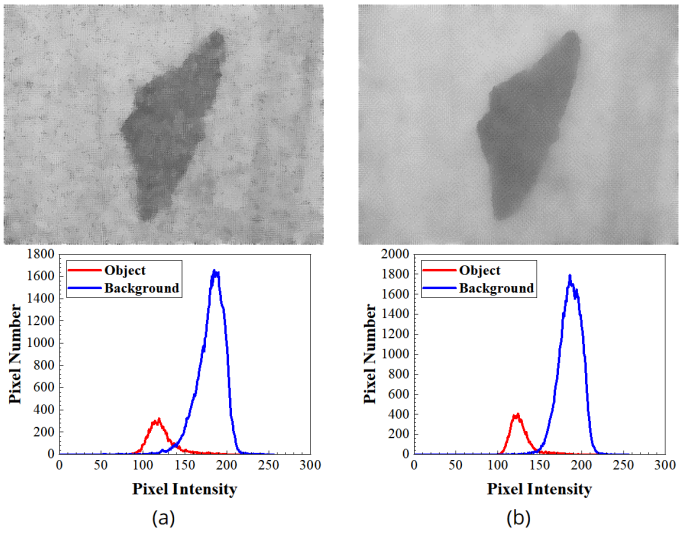


Figure 5. Comparison between traditional total variation and the total variation diffusion-guided energy term. (a) Corrected image obtained by traditional total variation and its intensity distribution; (b) Corrected image obtained by the total variation diffusion-guided energy term and its intensity distribution.

3.2 Excitation Weight Estimation Method

In the second stage shown in Figure 4, this paper continues the energy function framework of LFEBACM [10], and designs a global energy term and a local energy term. To more accurately estimate the excitation weights of neighboring pixels to the center pixel, this section reconstructs the excitation weight estimation method. It uses information entropy to detect noisy pixels and calculates the excitation weights based on the regional dependencies between noisy pixel and its neighboring pixels. In this sense, the excitation weight estimation process can be regarded as reliability-guided local information fusion. Information entropy can be formulated as:

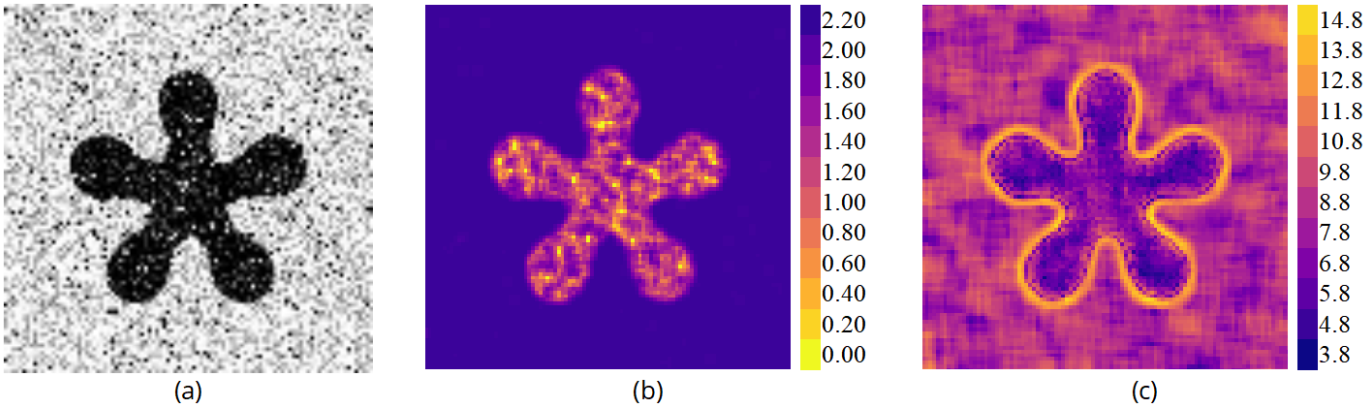


Figure 6. Visualization for different information entropy. (a) Noisy image; (b) Traditional information entropy; (c) Local structure entropy.

$$L_r(x) = -\sum_{y \in N_x} p(y) \log p(y) \tag{13}$$

where $L_r(x)$ represents the information entropy of pixel x. $p(y)$ denotes the probability distribution function of neighboring pixel y, which can be formulated as:

$$p(y) = \frac{I(y)}{\sum_{y \in N_x} I(y)} \tag{14}$$

The information entropy is related to the intensity variation within a local region. It takes larger values in homogeneous regions and smaller values at image edges or in regions polluted by noise. This property helps to detect noisy pixels. As shown in Figure 6 (b), we use Eq. (13) to calculate the information entropy of the noisy image shown in Figure 6 (a) and visualize the result. The visualization results indicate that traditional information entropy can identify noisy pixels in the object region, but it is difficult to detect noisy pixels in the background region and image edges. The main reason is that its probability distribution function is easily affected by noise.

To address the above problem, a local structure entropy is designed, which can be formulated as:

$$L_s(x) = -\sum_{y \in N_x} p_s(y) \log p_s(y) \tag{15}$$

where $p_s(y)$ represents the probability distribution function, which can be expressed as:

$$p_s(y) = \frac{\mathrm{mod}(I(y), \mathrm{mean}(N_x))}{\mathrm{mean}(N_x)} \tag{16}$$

where $\mathrm{mean}(N_x)$ is the intensity mean of all pixels in the local region N_x. $\mathrm{mod}(\cdot)$ is the modulus operator.

The local structure entropy uses statistical variables (i.e., the intensity mean) and the modulus operator to reduce the impact of noise to a certain extent and enhance the contrast between pixels. Similarly, this section uses Eq. (15) to calculate the local structure entropy of the noisy image shown in Figure 6 (a) and visualizes the result. As shown in Figure 6 (c), the local structure entropy can accurately identify noisy pixels and image edges. Moreover, compared with traditional information entropy, the local structure entropy has a wider values range. This enables noisy pixels and image edges detected by the local structure entropy more discriminative.

The local structure entropy takes larger values at image edges or in region polluted by noise, and smaller values in homogeneous regions. This indicates that the local structure entropy can be used to distinguish the region types in which the center pixel and its neighboring pixels are located. When the center pixel and its neighboring pixels are located at the same type of region, such as a homogeneous region or a region polluted by noise, the values of their local structural entropies are very close. On the contrary, the values of their local structural entropies are significantly different. Compared with the relative spatial distance and intensity similarity, the local structure entropy can reveal the noise distribution and the degree to which pixels are polluted by noise. Because the local structure entropy is calculated in a neighboring window, it can capture more context information. The reconstructed excitation weight estimation method based on the local

structure entropy can be formulated as:

$$h_{xy} = \begin{cases} 1 - \dfrac{\exp\left((L_s(y) - \bar{L}_s(x))^2\right)}{\sum_{y \in N_x} \exp\left((L_s(y) - \bar{L}_s(x))^2\right)}, & L_s(y) \geq \bar{L}_s(x) \\ 1 + \dfrac{\exp\left((L_s(y) - \bar{L}_s(x))^2\right)}{\sum_{y \in N_x} \exp\left((L_s(y) - \bar{L}_s(x))^2\right)}, & L_s(y) < \bar{L}_s(x) \end{cases} \tag{17}$$

where $L_s(y)$ represents the local structure entropy of neighboring pixel y. $\bar{L}_s(x)$ denotes the average local structure entropy in the local region N_x.

According to Eq. (17), the reconstructed excitation weight estimation method is designed as a piecewise function. It takes the relationship between $L_s(y)$ and $\bar{L}_s(x)$ as the judgment basis to quantify the degree to which neighboring pixel y is polluted by noise. The core logic and implementation mechanism are elaborated as follows:

- When $L_s(y) \geq \bar{L}_s(x)$, this indicates that the neighboring pixel y is seriously polluted by noise, so a relatively small excitation weight is assigned to it to reduce its interference on the membership of the center pixel x. The greater the difference between $L_s(y)$ and $\bar{L}_s(x)$, the smaller the excitation weight.
- When $L_s(y) < \bar{L}_s(x)$, this indicates that the neighboring pixel y is only slightly polluted by noise, so a relatively large excitation weight is assigned to it to enhance its role in correcting the membership of the center pixel x. The greater the difference between $L_s(y)$ and $\bar{L}_s(x)$, the larger the excitation weight.

To validate the effectiveness of the reconstructed excitation weight estimation method, this section uses Eq. (17) to calculate the excitation weights of neighboring pixels in the two local regions shown in Figure 3. As shown in Figure 7, in the local region N_1, the excitation weight of pixel A is larger than those of pixels B and C, and the excitation weight of pixel B is smaller than that of pixel C. Furthermore, in the local region N_2, the excitation weight of pixel E is smaller than that of pixel F, and the excitation weights of pixels E and F is smaller than that of pixel G. The above results are consistent with the corresponding analyses in Section 2.2.2.

However, although pixel H is not affected by noise, its excitation weight was smaller than those of pixels E and F. This is mainly because the reconstructed excitation weight estimation method explicitly accounts for spatial context information.

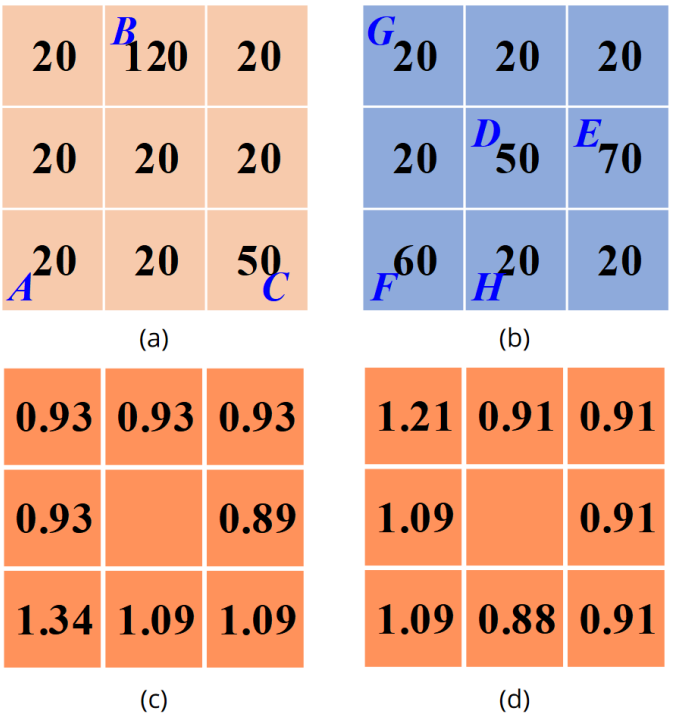


Figure 7. Excitation weights obtained by TVDGFACM. (a) Local region N_1; (b) Local region N_2; (c) Excitation weights of pixels in the local region N_1 calculated by Eq. (17); (d) Excitation weights of pixels in the local region N_2 calculated by Eq. (17).

The excitation weights obtained by Eq. (17) is related to the noise distribution. Specifically, the local region centered at pixel H contains more noisy pixels, whereas the neighboring window centered at pixel G contains fewer noisy pixels, so the excitation weight of pixel H is smaller than that of pixel G. Although pixels E and F are polluted by noise, their excitation weights are larger than that of pixel H. This is mainly because the local regions centered at pixels E and F contain fewer noisy pixels. Compared with excitation weight estimation methods based on the relative spatial distance and intensity similarity between pixels, the reconstructed excitation weight estimation method can adaptively calculate excitation weight according to noise distributions and better reveal the degrees to which pixels are polluted by noise.

3.3 Energy Function and Its Minimization

In the third stage shown in Figure 4, TVDGFACM obtains the update equations for each variable by solving the derivatives of energy function with respect to these variables. It achieves the energy function minimization and curve evolution through step-by-step iteration. The energy function of TVDGFACM is composed of a total variation

diffusion-guided energy term, a global energy term and a local energy term, which can be formulated as:

$$\begin{aligned} E_T = &\int (I(x) - I_c(x))^2 dx + \frac{\gamma}{\theta(x)} \int |\nabla I_c(x)|^{\theta(x)} dx \\ &+ \int (1-\beta(x))[u(x)]^m (I_c(x) - v_1)^2 dx \\ &+ \int (1-\beta(x))[1-u(x)]^m (I_c(x) - v_2)^2 dx \\ &+ \int \beta(x)[u(x)]^m h_{xy} [1-u(y)]^m (I_c(y) - v_1)^2 dy dx \\ &+ \int \beta(x)[1-u(x)]^m h_{xy} [u(y)]^m (I_c(y) - v_2)^2 dy dx \end{aligned} \tag{18}$$

where $\beta(x)$ represents the balance factor. v_1 and v_2 denote the intensity means of corrected image $I_c(x)$ in the inside and outside regions of evolving curve respectively.

Eq. (18) provides a unified energy-level fusion formulation. The total variation diffusion-guided term, global energy term, and local energy term contribute complementary information for the segmentation decision.

The first two terms in Eq. (18) are the total variation diffusion-guided energy term. The third and fourth terms are the global energy term. The last two terms are the local energy term. When the noise density is small, the global energy term plays a major role. This can speed up the curve evolution. When the noise density is large, the local energy term should dominate. This is conducive to generating a smooth and coherent segmentation result. Based on the above analyses, this section defines a balance factor $\beta(x)$ according to the noise density, as follows:

$$\beta(x) = 1 - \exp\left(-(I(x) - I_c(x))^2\right) \tag{19}$$

According to Eq. (19), when the noise density is low, $(I(x) - I_c(x))^2$ is small, $\beta(x)$ takes a small value. In this case, the global energy term plays a major role. When the noise density is large, $(I(x) - I_c(x))^2$ is large, $\beta(x)$ takes a large value. In this case, the local energy term dominates. Therefore, the balance factor acts as an adaptive fusion coefficient between global and local intensity information.

Assume the neighboring window sizes for calculating the local energy term and local structure entropy are $K \times K$ and $P \times P$, respectively. For an image of size $H \times W$, the computational complexity of the proposed model in a single iteration is $O(HW(K^2 + P^2))$, and the computational complexity in N iterations is $O(NHW(K^2 + P^2))$.

Keeping $u(x)$ and $I_c(x)$ unchanged, minimizing Eq. (18) can obtain the update equations for v_1 and v_2:

$$\begin{aligned} v_1 &= \frac{\int (1-\beta(x))[u(x)]^m I_c(x) dx}{\int (1-\beta(x))[u(x)]^m dx}, \\ v_2 &= \frac{\int (1-\beta(x))[1-u(x)]^m I_c(x) dx}{\int (1-\beta(x))[1-u(x)]^m dx} \end{aligned} \tag{20}$$

Keeping v_1, v_2 and $I_c(x)$ unchanged, minimizing Eq. (18) can obtain the update equations for $u(x)$:

$$u(x) = \left(1 + \left(\frac{(1-\beta(x))(I_c(x) - v_1)^2 + \beta(x) b_1}{(1-\beta(x))(I_c(x) - v_2)^2 + \beta(x) b_2}\right)\right)^{-1} \tag{21}$$

where

$$\begin{aligned} b_1 &= \int h_{xy} [1-u(y)]^m (I_c(y) - v_1)^2 dy, \\ b_2 &= \int h_{xy} [u(y)]^m (I_c(y) - v_1)^2 dy. \end{aligned} \tag{22}$$

Keeping $u(x), v_1$ and v_2 unchanged, minimizing Eq. (18) can obtain the update equations for $I_c(x)$:

$$\begin{aligned} \frac{\partial I_c(x)}{\partial t} =& \gamma \nabla \left(\frac{\nabla I_c(x)}{|I_c(x)|^{1-\theta(x)}}\right) \\ &- 2(I(x) - I_c(x)) \\ &- 2(1-\beta(x))[u(x)]^m (I_c(x) - v_1) \\ &- 2(1-\beta(x))[1-u(x)]^m (I_c(x) - v_2) \end{aligned} \tag{23}$$

Traditional active contour models typically employ gradient descent flow to minimize the energy function, but this method is computationally inefficient. Therefore, this paper uses a fast numerical method invented by Song and Chan [19] and developed by Krinidis and Chatzis [8] to minimize the energy function. This method does not solve the Euler-Lagrange equations, while directly calculates whether the energy function decreases when the membership value of any pixel changes. If the energy function decreases, the membership function is updated. Otherwise, it is not updated.

To illustrate the convergence of the proposed model, we adopt an alternating minimization strategy. The total variation diffusion-guided energy term E_D is minimized with respect to $I_c(x)$ via gradient descent flow (Eq. (22)), while the global and local energy terms are convex with respect to $u(x)$ as proved in

Appendix A. The alternating minimization scheme ensures that the overall energy function decreases monotonically. Combined with the fast numerical method, the global optimum is obtained by gradually reducing the energy function, and thus achieving model convergence. Moreover, the fast numerical method does not rely on the gradient descent step, which avoids the convergence oscillations caused by inappropriate step settings in gradient descent flow methods. Combined with the energy descent criterion, the optimization process exhibits a more stable energy decrease numerically and eventually tends to plateau.

According to the above descriptions, the optimization process of TVDGFACM is shown in Algorithm 1.

Algorithm 1: The optimization process of TVDGFACM.

Input : The noisy image $I(x)$, stopping criteria η and maximum number of iterations N.

Output: The membership function $u^n(x)$.

Initialize the membership function $u(x)$ and the corrected image $I_c(x) = I(x)$;

for n *from* 1 *to* N **do**

- Calculate the balance factor $\beta(x)$ using Eq. (19);
- Calculate the intensity means of different regions v_1^n and v_2^n using Eq. (20);
- Calculate the membership function $u^n(x)$ using Eq. (21);
- Calculate the corrected image $I_c^n(x)$ using Eq. (22);
- Calculate the energy function E^n using Eq. (18);
- **if** $E^n < E^{n-1}$ **then**
  - update the membership function to $u^n(x)$;
- **end**
- **else**
  - do not update;
- **end**
- **if** *maximum number of iterations or* $|E^n - E^{n-1}| < \eta$ *is satisfied* **then**
  - break;
- **end**

end

Output: The membership function $u^n(x)$.

3.4 Illustrative Example

This section gives an illustrative example of TVDGFACM. The color noisy image $I_r(x)$ and given initial evolving curve C are shown in Figure 8. The row and column coordinates of upper left corner of initial evolving curve are $(15, 100)$. The row and column coordinates of lower right corner of initial evolving curve are $(210, 215)$. The initial evolving curve is a rectangular box formed by these two points. In the inside region of initial evolving curve, the membership function $u(x) = 1$. In the outside region of evolving curve, the membership function $u(x) = 0$.

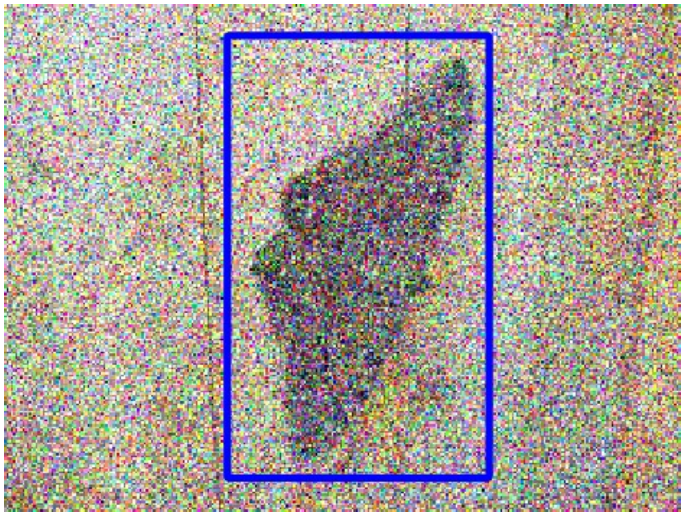


Figure 8. Color noisy image and given initial evolving curve.

Before the iteration begins, the color noisy image $I_r(x)$ is converted to the grayscale image $I(x)$. The corrected image $I_c(x)$ is initialized to the grayscale noisy image $I(x)$. Next, during the iteration, the balance factor $\beta(x)$ is first calculated using Eq. (19). Then, the intensity means v_1 and v_2 in the inside and outside regions of the evolving curve is calculated using Eq. (20). Furthermore, the membership function $u(x)$ is calculated according to Eq. (21), and its corresponding evolving curve is shown in Figure 9.

Then, the corrected image $I_c(x)$ is obtained and updated using Eq. (22), as shown in Figure 10. Finally, the above steps are iterated sequentially until the convergence condition is reached. This can obtain the final evolving curve and corrected image. They are shown in Figures 11 and 12, respectively.

4 Experimental Results and Analyses

To validate the advancement and effectiveness of the proposed method, this section conducts experiments using a synthetic image, two natural images from the BSD500 dataset, and three remote sensing images from the AID dataset. Among them, synthetic images are structurally and statistically controllable, which facilitates the systematic verification of the

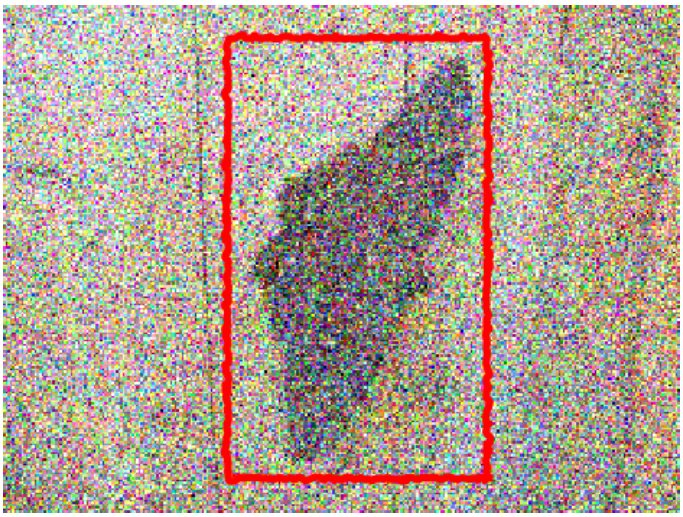


Figure 9. Evolving curve obtained in the first iteration.

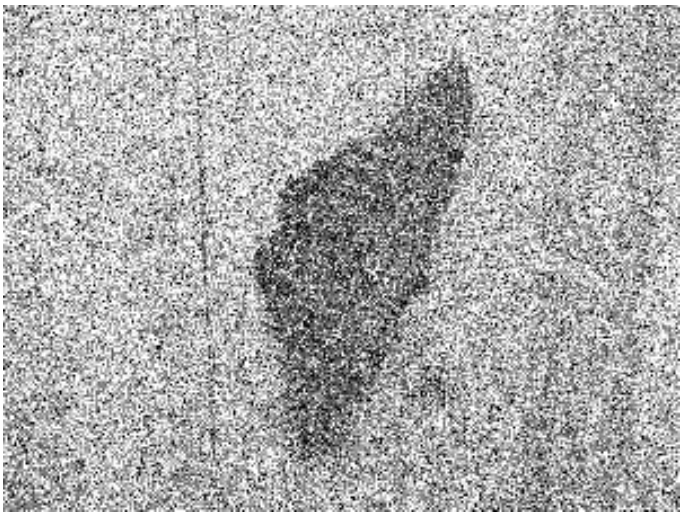


Figure 10. Corrected image obtained in the first iteration.

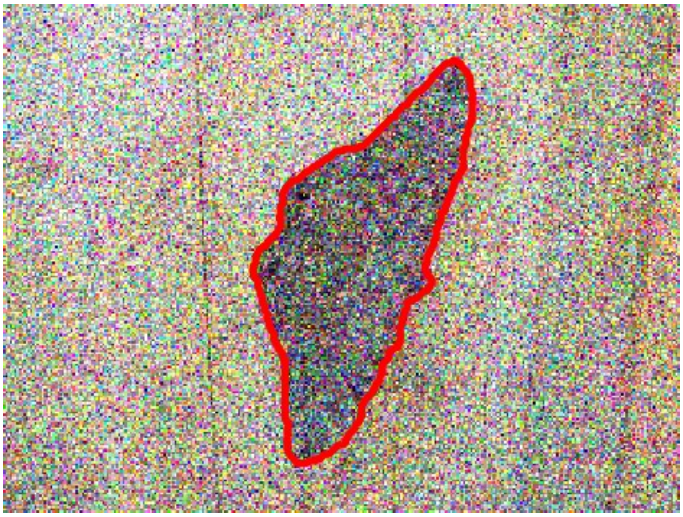


Figure 11. Final Evolving curve.

boundary preservation and noise resistance of the segmentation method under different noise densities. Natural images have complex textures and unstable

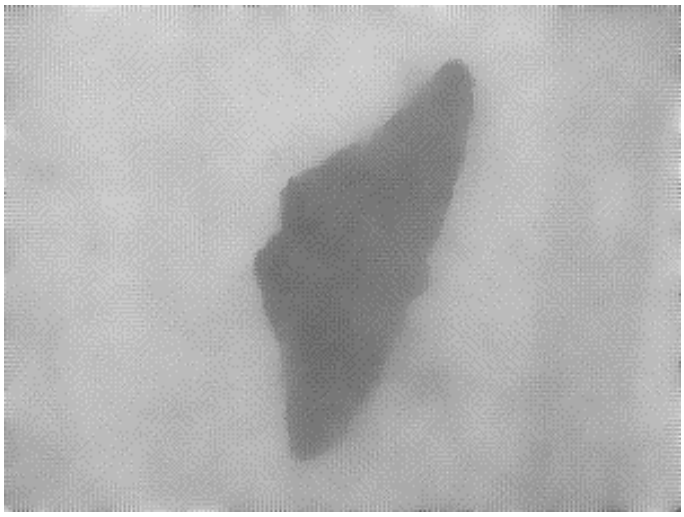


Figure 12. Final corrected image.

foreground-background contrasts, which can be used to verify the effectiveness of local information fusion and edge localization in real-world scenarios. Remote sensing images typically exhibit larger-scale variations and more complex background structures, which are primarily used to examine the stability and applicability of the segmentation method under high-resolution and complex texture conditions. In addition, we further conduct comparison experiments on large-scale benchmark datasets to strengthen the support for generalization of the proposed model.

The experimental CPU configuration is a 13th Gen Intel(R) Core (TM) i7-13700KF CPU at 3.40GHz. The experimental GPU configuration is a NVIDIA RTX3060ti 8GB. The compilation environment of MATLAB code is MATLAB 2018b. The compilation environment of Python code is python 3.8, PyTorch 2.4 and CUDA v11.8. Moreover, the segmentation accuracy (SA) and mean intersection over union (MIoU) are used to evaluate the segmentation performance. Furthermore, we further introduce global Moran's I (GMI) and boundary F1 (BF1) scores to measure segmentation quality from the spatial smoothness and boundary preservation, respectively. Overall, SA and MIoU focus more on pixel-level classification accuracy and the degree of region overlap. GMI can reflect the spatial smoothness of the segmentation results. BF1 directly measures the matching degree between the predicted boundary and the true boundary within a certain tolerance.

In subsequent experiments, mixed noise with different densities is added to the test images. Mixed noise with density d consists of Gaussian noise with mean 0 and variance d, salt and pepper noise with density d, and

speckle noise with density d. For ease of description, noise density is expressed as a percentage, i.e., $(d \times 100)\%$.

4.1 Discussions on Parameter Setting

In the proposed model, some key parameters significantly affect segmentation performance, including the Lagrange multipliers, neighboring window size, entropy window size, stopping criteria and weighting exponent m.

The Lagrange multiplier γ is a key parameter that affects the performance of the total variation diffusion-guided energy term. A small Lagrange multiplier makes it difficult for the total variation diffusion-guided energy term to effectively remove noise. A large Lagrange multiplier leads to an over-smoothed corrected image and causes the loss of image edges. To determine the optimal Lagrange multiplier, this section selects different values from the set $[50, 500]$ with a step size of 50 for experiments. A synthetic image, a natural image, and two remote sensing images polluted by 5% mixed noise are used to evaluate the impact of the Lagrange multiplier on segmentation performance. The test images and their initial evolving curves are shown in Figure 13.

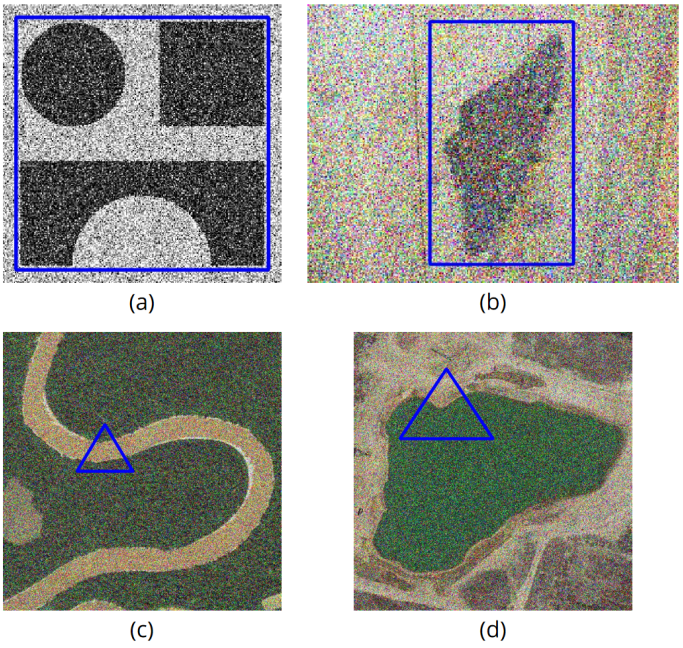


Figure 13. Test images polluted by 5% mixed noise and their initial evolving curves. (a), (b), (c), and (d) are the synthetic image, natural image, remote sensing image 1, and remote sensing image 2, respectively.

As can be seen from the experimental results shown in Figure 14, different Lagrange multipliers have little impact on the segmentation performance of the synthetic image. The main reason is that the intensity distributions of different regions in the synthetic image are relatively homogeneous. Although noise disturbs the intensity statistical characteristic, the local energy term alone can still achieve a better segmentation performance. In contrast, the intensity distributions of different regions in natural images and remote sensing images are more complex, and noise further aggravates this complexity. As a result, the Lagrange multiplier has a greater impact on the segmentation performance for these images. According to Figure 14, the natural image achieves the best segmentation performance when $\gamma = 300$ and $\gamma = 350$. In addition, when γ lies in the interval $[200, 350]$, the segmentation performance of remote sensing Image 1 remains almost unchanged and is better than that obtained with other values. As γ increases, the segmentation performance of remote sensing Image 2 gradually improves. When $\gamma \geq 250$, its segmentation performance becomes nearly stable. Based on the above analysis, the Lagrange multiplier γ is uniformly set to 300 in the subsequent experiments for all test images. Although the optimal γ varies slightly across different images, setting $\gamma = 300$ provides a balanced performance for all test scenarios, ensuring fair comparison and avoiding image-specific parameter tuning.

The neighboring window size is an important parameter that affects the performance of the local energy term. A small neighboring window makes it difficult for the proposed model to accurately estimate the dependencies between pixels, thereby resulting

Table 1. Impact of neighboring window size on segmentation performance. The optimal results are in bold.

Metrics	SA (%)					MIoU (%)				
	3×3	5×5	7×7	9×9	11×11	3×3	5×5	7×7	9×9	11×11
Synthetic image	99.86	**99.92**	98.78	99.74	99.52	99.71	**99.84**	99.04	97.58	99.47
Natural image	98.89	98.91	**98.96**	98.80	98.67	95.40	95.49	**95.67**	95.00	94.47
Remote sensing image 1	98.93	**99.02**	98.74	98.49	98.15	97.02	**97.26**	96.51	95.85	94.94
Remote sensing image 2	99.06	**99.31**	99.27	99.09	98.89	98.02	**98.52**	98.45	98.08	97.65

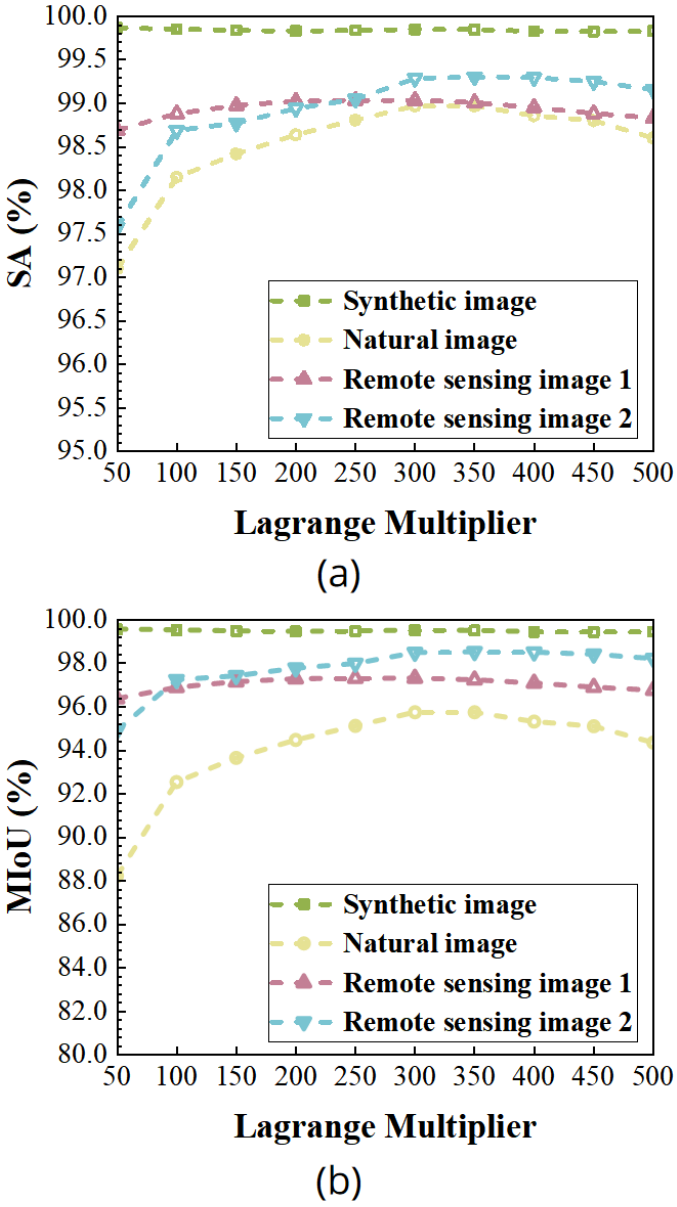


Figure 14. Impact of Lagrange multipliers on segmentation performance. (a) Segmentation accuracy; (b) Mean intersection over union.

in misclassified pixels in the segmentation results. A large neighboring window will cause image edges and details to become blurred due to excessive fusion of neighborhood information, thereby resulting in segmentation boundary shifts. As shown in Table 1, to determine the optimal neighborhood window size, this section selects different values from the set $[3, 11]$ with a step size of 2 for experiments. The test images and their initial evolving curves are shown in Figure 13. According to the experimental results, when the neighboring window size is 5×5, the synthetic image and two remote sensing images reach the optimal segmentation performance. When the neighboring window size is 7×7, the natural image obtains the best evaluation metrics. Considering that computational complexity increases with the neighboring window size, this paper sets the neighboring window size to 5×5 in subsequent experiments.

The entropy window size is a key parameter that affects the performance of the excitation weight estimation method. It is difficult for a small entropy window to capture sufficient local statistical information, which makes modulo operations and mean calculations susceptible to noise, thereby hindering accurate identification for noisy pixels and image edges. A large entropy window will integrate irrelevant intensity information across regions and weaken local structural differences, which makes it difficult to accurately measure the degree to which pixels are polluted by noise. As shown in Figure 15, to determine the optimal entropy window size, this section selects different values from the set $[3, 11]$ with a step size of 2 for experiments. The test images and their initial evolving curves are shown in Figure 13. Experimental results show that as the entropy window increases, the segmentation performance of synthetic images gradually decreases, while the segmentation performance of natural and remote sensing images shows a gradual improvement. This is because the intensity distribution of synthetic images is usually relatively homogeneous, and an excessively large entropy window can easily introduce interference information from neighboring pixels unrelated to the center pixel, thereby causing bias in the excitation weights estimation. In contrast, the intensity distribution of natural and remote sensing images is more complex and spatially varied. An excessively small entropy window cannot adequately characterize the contextual statistical features. Therefore, appropriately increasing the window size helps improve the model's representation ability for local structures and noise patterns. Therefore, in subsequent experiments, we set the entropy window for synthetic images to 5×5 and the entropy window for natural and remote sensing images to 9×9.

The stopping criterion is a key parameter affecting the model's convergence speed and segmentation performance. A small stopping criterion often significantly increases the number of iterations and computational cost. Meanwhile, in the later stages of iteration, the energy function decrease is usually quite limited, and the segmentation performance improvement tends to saturate. Continuing iteration not only yields little benefit but may also introduce unnecessary over-smoothing. A large stopping

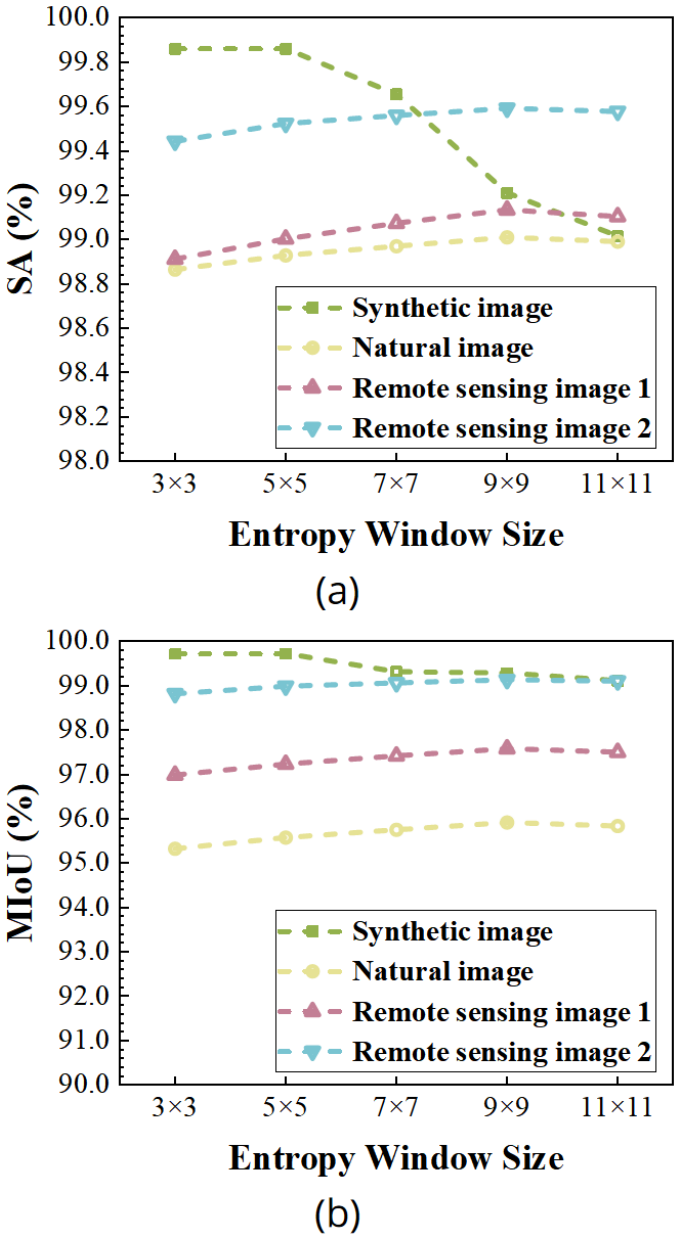


Figure 15. Impact of entropy window size on segmentation performance. (a) Segmentation accuracy; (b) Mean intersection over union.

criterion may cause the model to terminate the iteration prematurely. This will result in the energy function not decreasing sufficiently, and the membership function and the correction image still being adjusted, thereby leading to problems such as unstable boundaries and local under-segmentation. As shown in Table 2, to determine the optimal stopping criterion, this section selects different values from the set $[5, 1, 0.1, 0.01, 0.001]$ for experiments. The maximum number of iterations is set to 100. The test images and their initial evolving curves are shown in Figure 13. Experimental results show that when the stopping criterion is 0.1, the synthetic and remote sensing images achieve the best segmentation performance. When the stopping criterion is 0.01, the natural image obtains the best evaluation index. When the stopping criterion is 0.1, the SA and MIoU of the natural image differ from the optimal SA and MIoU by only 0.01% and 0.02%, respectively, while the number of iterations differs by 23. Considering computational efficiency, this paper uniformly sets the stopping criterion to 0.1 in subsequent experiments.

The weighting exponent m is primarily used to adjust the model's constraint strength on the membership function. To analyze the impact of the weighting exponent on segmentation performance, this section selects different values from the set $[2, 6]$ with a step size of 1 for experiments. The test images and their initial evolving curves are shown in Figure 13. The experiment results are shown in Figure 16. Experimental results show that as the weighting exponent increases, the segmentation performance continuously decreases. This is because the increase in the weighted exponent causes the membership values to tend towards the extreme states of 0 or 1, thereby significantly reducing the flexibility of the membership function. This makes it difficult for the proposed model to adapt to regional dependencies between pixels, and it cannot accurately respond to the differences in noise distribution and the degree to which pixels are polluted by noise characterized by the local structure entropy, thereby affecting the rationality of the excitation weight estimation. Furthermore, excessive

Table 2. Impact of stopping criterion on convergence speed and segmentation performance. The optimal results are in bold.

	Synthetic image			Natural image			Remote sensing image 1			Remote sensing image 2		
	SA (%)	MIoU (%)	Iter.	SA (%)	MIoU (%)	Iter.	SA (%)	MIoU (%)	Iter.	SA (%)	MIoU (%)	Iter.
5	99.92	99.85	26	98.64	94.46	27	98.96	97.12	26	98.71	97.29	28
1	99.92	99.85	32	98.70	94.69	31	99.00	97.22	31	98.76	97.40	34
0.1	**99.93**	**99.86**	34	99.04	96.02	37	**99.02**	**97.26**	33	**99.46**	**98.86**	40
0.01	99.93	99.86	100	**99.05**	**96.04**	60	99.01	97.25	44	99.44	98.81	48
0.001	99.93	99.86	100	99.04	96.00	100	98.96	97.10	100	99.42	98.77	77

constraints will weaken the model's capability to capture complex edges and details, especially for the heterogeneous boundary regions in natural images and remote sensing images, which may result in loss of edge information or deviation of the segmentation boundary. Even for synthetic images with relatively uniform intensity distribution, local misclassification may occur due to insufficient smoothness of the membership degree, which leads to a gradual decline in overall segmentation performance as the weighting index increases. Therefore, in subsequent experiments, this paper uniformly sets the weighting exponent m to 2.

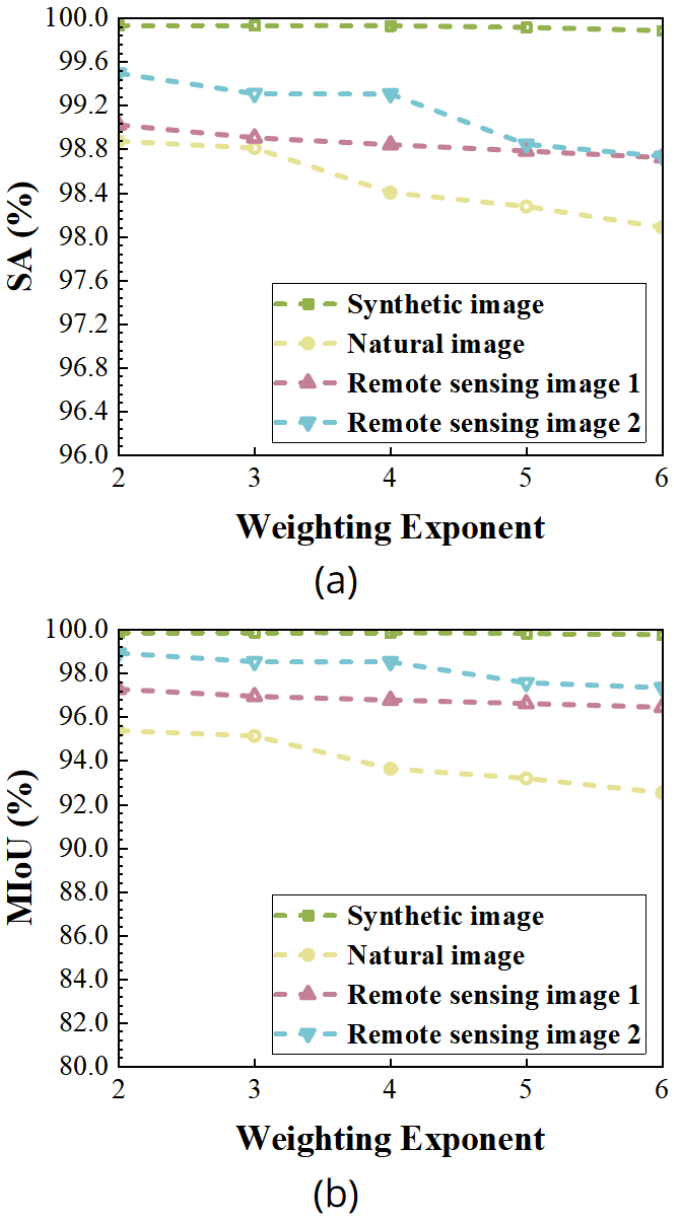


Figure 16. Impact of weighting exponent on segmentation performance. (a) Segmentation accuracy; (b) Mean intersection over union.

4.2 Comparison Experiment

To comprehensively demonstrate the advantages of TVDGFACM, this section compares the proposed method with seven existing active contour models (FEBACM [8], GLFEBACM [11], MSR [13], PLPF [20], ACFDI [1], AWMAD [21], and DoGEED [22]) and two deep learning methods (HAUS [23] and DIP [24]), respectively. To evaluate the robustness of different methods to noise, we added mixed noise with densities of 2%, 4%, 6%, 8% and 10% to the test images, respectively. Moreover, we report the average segmentation performance±standard deviation.

4.2.1 Comparison with Active Contour Models

The segmentation results obtained by different active contour models on six test images polluted by 6% mixed noise are shown in Figures 17- 22, respectively. Moreover, SA (%), MIoU (%), GMI (%), and BF1 (%) obtained by different active contour models are shown in Table 3, 4, 5 and 6, respectively.

The experimental results show that FEBACM is the worst-performing segmentation model. Its segmentation results contain a large number of misclassified pixels, which leads to a decrease in spatial coherence and a shift in the predicted boundary. Meanwhile, the segmentation performance of FEBACM has a large standard deviation. This indicates that it is not robust to noise interference. The main reason is that FEBACM is a segmentation model based on global intensity information. It typically assumes that images are composed of homogeneous regions. However, mixed noise destroys this homogeneity assumption. This makes it difficult for FEBACM to accurately segment noisy images. In contrast, GLFEBACM introduces local intensity information into FEBACM. This improves its noise resistance to a certain extent. Compared with FEBACM, GLFEBACM produces fewer misclassified pixels, which makes its segmentation results have better spatial coherence and boundary preservation. Moreover, the segmentation performance of GLFEBACM has a smaller standard deviation. However, GLFEBACM lacks an explicit denoising mechanism and its excitation weight estimation method is limited. As a result, its segmentation performance is inferior to the proposed method. MSR is essentially an active contour model based on global intensity information. Unlike FEBACM, MSR introduces total variation to remove image noise. It achieves a better segmentation performance, spatial coherence and boundary preservation on

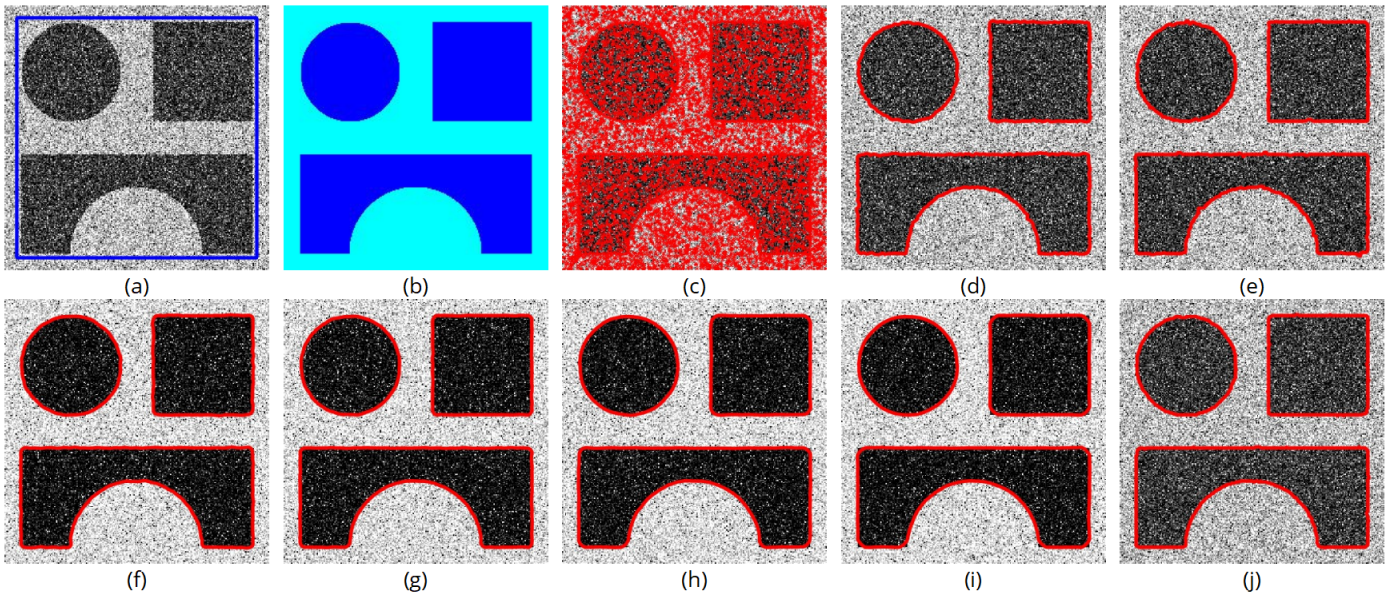


Figure 17. Segmentation results obtained by different active contour models on the synthetic image polluted by 6% mixed noise. (a) Initial evolving curve; (b) Ground truth; (c) FEBACM; (d) GLFEBACM; (e) MSR; (f) PLPF; (g) ACFDI; (h) AWMAD; (i) DoGEED; (j) TVDGFACM.

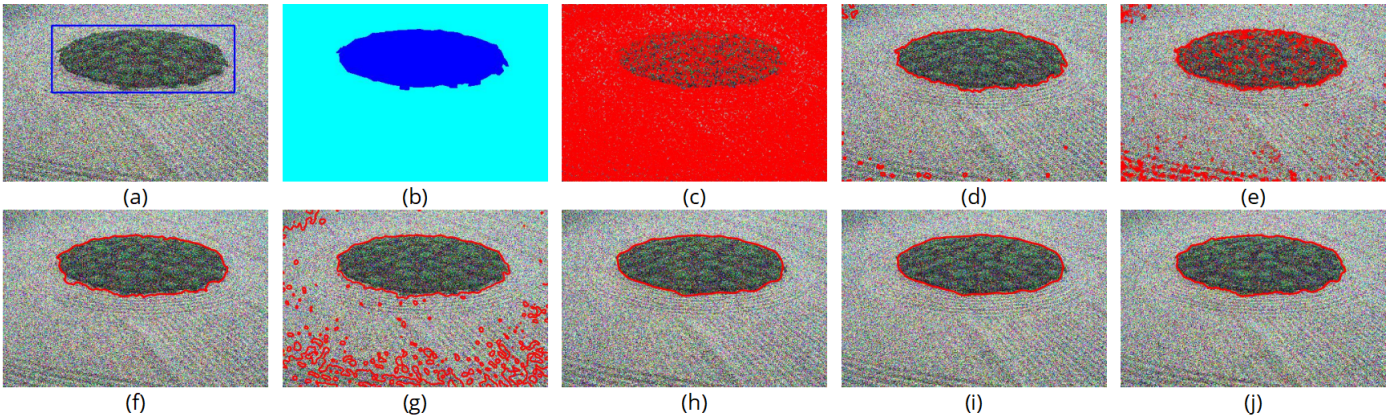


Figure 18. Segmentation results obtained by different active contour models on the natural image 1 polluted by 6% mixed noise. (a) Initial evolving curve; (b) Ground truth; (c) FEBACM; (d) GLFEBACM; (e) MSR; (f) PLPF; (g) ACFDI; (h) AWMAD; (i) DoGEED; (j) TVDGFACM.

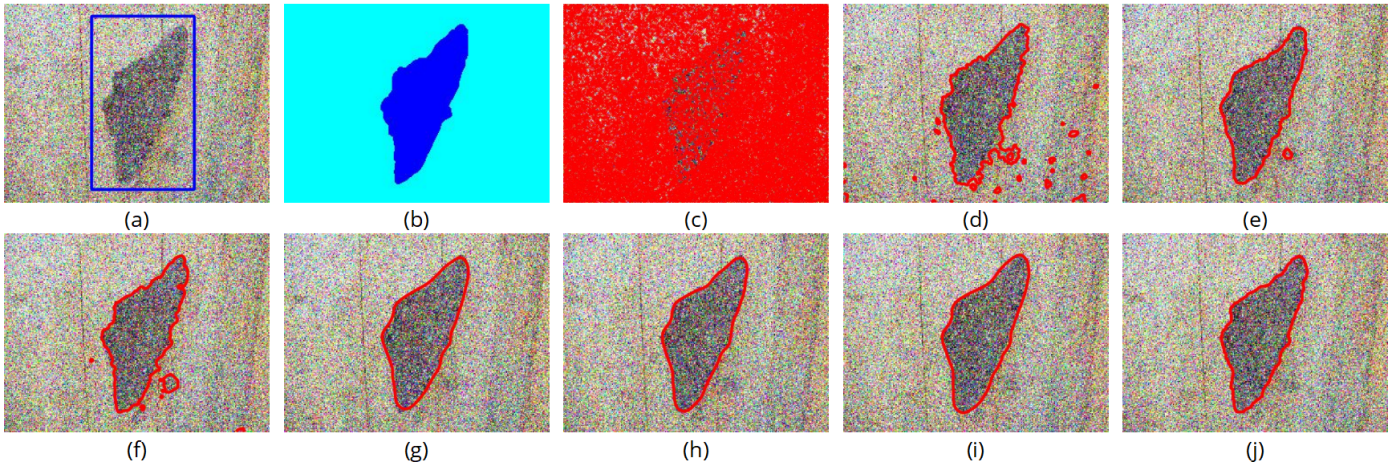


Figure 19. Segmentation results obtained by different active contour models on the natural image 2 polluted by 6% mixed noise (a) Initial evolving curve; (b) Ground truth; (c) FEBACM; (d) GLFEBACM; (e) MSR; (f) PLPF; (g) ACFDI; (h) AWMAD; (i) DoGEED; (j) TVDGFACM.

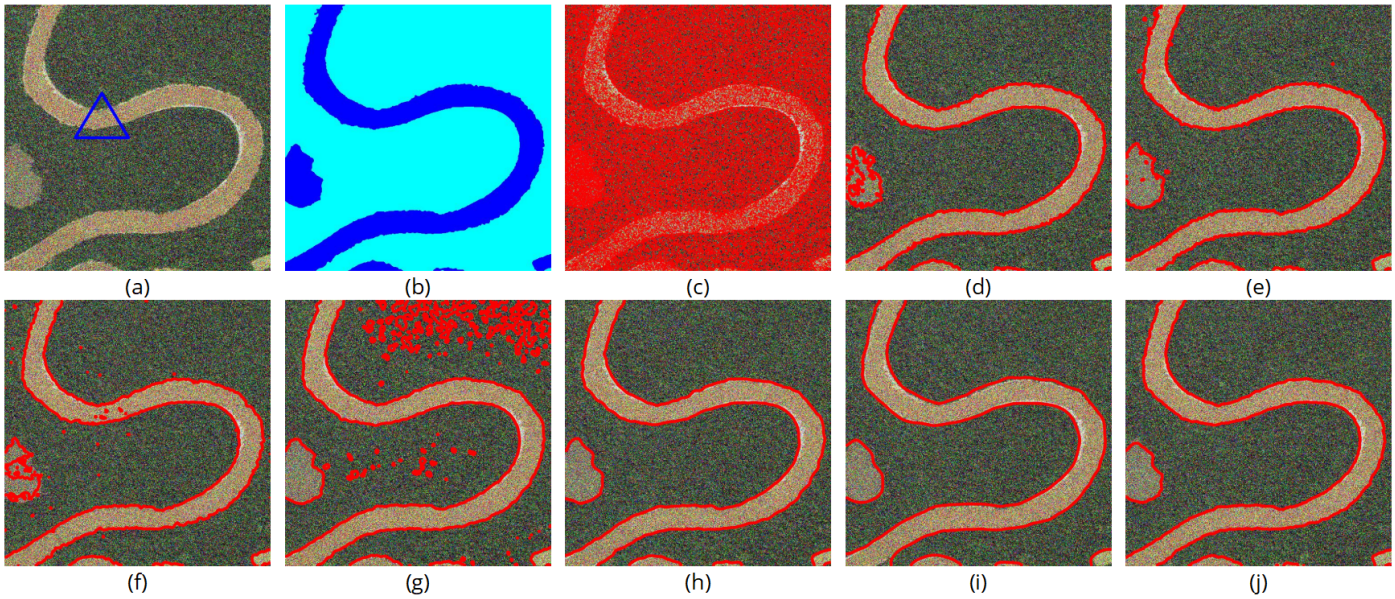


Figure 20. Segmentation results obtained by different active contour models on the remote sensing image 1 polluted by 6% mixed noise. (a) Initial evolving curve; (b) Ground truth; (c) FEBACM; (d) GLFEBACM; (e) MSR; (f) PLPF; (g) ACFDI; (h) AWMAD; (i) DoGEED; (j) TVDGFACM.

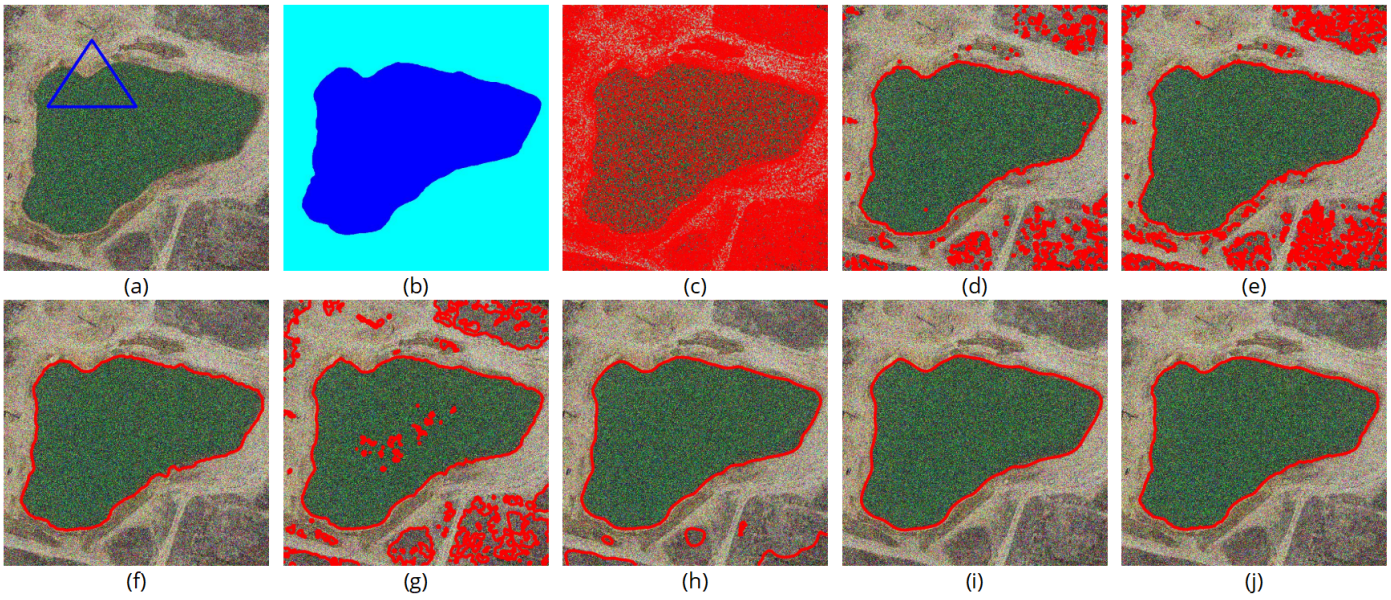


Figure 21. Segmentation results obtained by different active contour models on the remote sensing image 2 polluted by 6% mixed noise. (a) Initial evolving curve; (b) Ground truth; (c) FEBACM; (d) GLFEBACM; (e) MSR; (f) PLPF; (g) ACFDI; (h) AWMAD; (i) DoGEED; (j) TVDGFACM.

Table 3. SA (%) obtained by different active contour models. The optimal results are in bold.

Image	FEBACM	GLFEBACM	MSR	PLPF	ACFDI	AWMAD	DoGEED	TVDGFACM
Synthetic image	83.40±8.22	99.28±0.33	99.46±0.32	99.71±0.18	99.65±0.21	99.61±0.10	99.48±0.05	**99.80±0.10**
Natural image 1	61.72±4.23	98.00±1.38	93.34±4.30	98.97±0.15	86.58±2.11	99.39±0.08	99.34±0.12	**99.42±0.02**
Natural image 2	61.76±2.83	91.28±8.16	91.55±7.89	98.34±0.16	98.33±1.16	98.84±0.23	98.75±0.29	**99.02±0.21**
Remote sensing image 1	76.42±6.55	98.19±1.09	98.75±0.35	98.15±1.07	95.35±0.48	98.97±0.19	97.27±0.29	**99.23±0.15**
Remote sensing image 2	69.72±6.31	95.83±1.29	89.84±0.68	98.94±0.07	82.04±1.82	95.56±3.79	99.06±0.06	**99.47±0.11**
Remote sensing image 3	81.13±6.33	96.67±1.13	92.55±0.46	97.27±0.84	93.44±2.27	97.26±0.74	95.59±1.35	**98.75±0.25**

synthetic images, but performs poorly on natural and remote sensing images. The reason is that, compared with synthetic image, the intensity distributions of natural and remote sensing images are more complex.

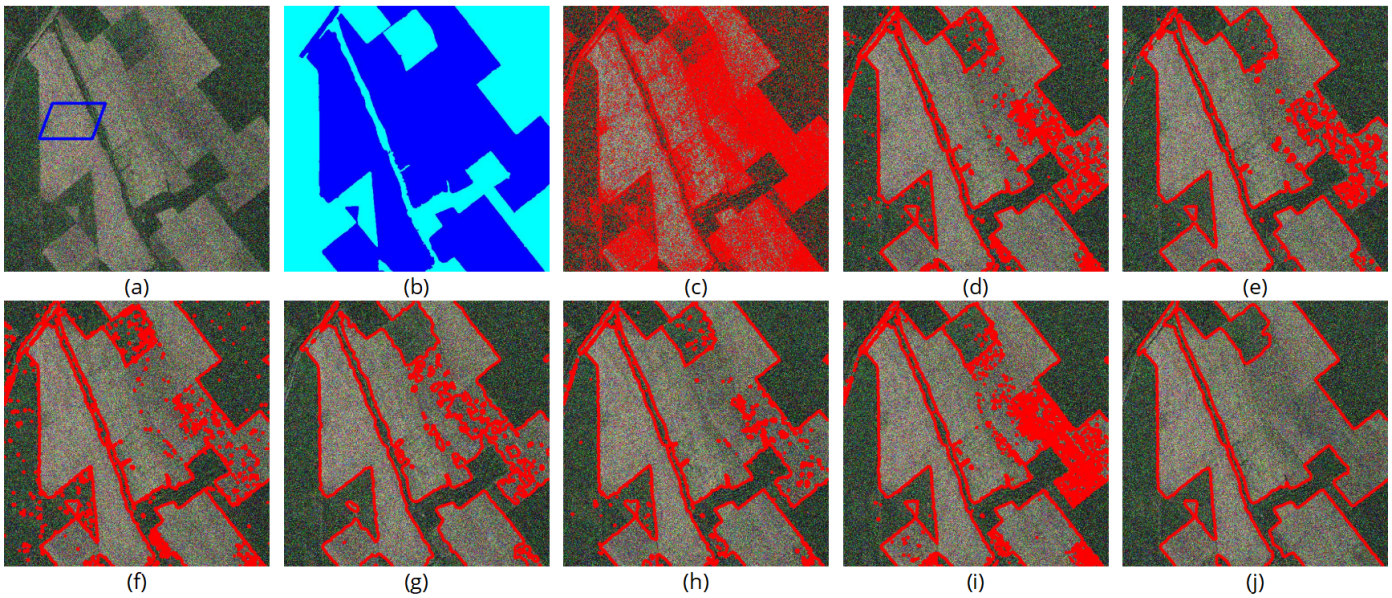


Figure 22. Segmentation results obtained by different active contour models on the remote sensing image 3 polluted by 6% mixed noise. (a) Initial evolving curve; (b) Ground truth; (c) FEBACM; (d) GLFEBACM; (e) MSR; (f) PLPF; (g) ACFDI; (h) AWMAD; (i) DoGEED; (j) TVDGFACM.

Table 4. MIoU (%) obtained by different active contour models. The optimal results are in bold.

Image	FEBACM	GLFEBACM	MSR	PLPF	ACFDI	AWMAD	DoGEED	TVDGFACM
Synthetic image	72.24±12.6	98.57±0.64	98.92±0.63	99.43±0.36	99.31±0.42	99.22±0.20	98.97±0.09	**99.59±0.21**
Natural image 1	41.10±4.11	93.50±3.88	80.53±11.2	96.40±0.50	69.53±3.15	97.81±0.29	97.63±0.42	**97.93±0.09**
Natural image 2	39.36±2.89	78.48±15.5	78.92±15.8	93.30±0.61	93.57±4.05	95.18±0.94	95.00±1.08	**95.87±0.89**
Remote sensing image 1	57.65±9.01	94.98±3.01	96.59±0.94	94.85±2.96	88.56±1.05	97.17±0.49	92.92±0.69	**97.87±0.42**
Remote sensing image 2	53.67±7.61	91.66±2.44	81.19±1.10	97.78±0.15	69.45±2.57	91.42±7.02	98.02±0.12	**98.88±0.23**
Remote sensing image 3	68.56±9.10	93.46±2.13	86.05±0.79	94.56±1.64	87.66±4.01	94.57±1.42	91.47±2.50	**97.46±0.50**

Table 5. GMI (%) obtained by different active contour models. The optimal results are in bold.

Image	FEBACM	GLFEBACM	MSR	PLPF	ACFDI	AWMAD	DoGEED	TVDGFACM
Synthetic image	46.03±22.4	97.31±0.15	97.43±0.17	97.61±0.06	97.65±0.03	97.67±0.01	97.67±0.01	**97.78±0.08**
Natural image 1	13.91±7.68	98.29±0.32	65.99±22.9	99.16±0.02	94.11±1.41	99.16±0.03	**99.22±0.01**	**99.22±0.01**
Natural image 2	9.659±6.38	97.54±0.95	93.91±6.01	97.92±0.20	98.64±0.09	98.55±0.01	98.63±0.04	**98.66±0.06**
Remote sensing image 1	24.83±14.9	97.80±0.14	97.90±0.39	97.57±0.60	96.36±0.49	98.34±0.05	98.29±0.03	**98.48±0.01**
Remote sensing image 2	28.12±15.5	97.76±0.36	95.88±0.59	99.45±0.01	96.97±0.47	99.02±0.49	99.38±0.01	**99.49±0.01**
Remote sensing image 3	48.23±16.2	95.54±1.22	97.10±0.40	96.28±0.84	97.26±0.62	97.08±0.58	95.23±1.00	**98.25±0.02**

Table 6. BF1 (%) obtained by different active contour models. The optimal results are in bold.

Image	FEBACM	GLFEBACM	MSR	PLPF	ACFDI	AWMAD	DoGEED	TVDGFACM
Synthetic image	24.26±5.97	99.75±0.41	99.13±0.17	99.88±0.02	99.78±0.06	98.81±0.07	97.03±0.23	**99.90±0.12**
Natural image 1	4.048±0.18	53.29±15.5	17.47±8.72	68.32±4.44	14.48±4.39	82.27±6.93	86.81±1.14	**86.90±3.09**
Natural image 2	5.988±0.15	35.13±28.1	40.48±27.9	62.93±5.07	75.07±10.7	73.75±2.53	70.63±8.64	**76.08±7.82**
Remote sensing image 1	10.87±1.54	87.11±5.18	85.41±9.59	79.85±13.2	52.75±7.26	91.28±3.44	43.72±7.92	**94.14±1.95**
Remote sensing image 2	4.940±0.77	38.68±4.27	21.57±2.90	64.87±3.48	25.40±4.00	65.84±20.5	63.97±3.00	**86.04±6.44**
Remote sensing image 3	17.46±4.24	61.29±10.9	54.59±1.66	67.39±9.74	51.66±15.2	73.26±9.08	52.09±10.0	**90.98±3.00**

Although relying on total variation can reduce the impact of noise, it is difficult to achieve ideal performance using only global intensity information.

PLPF and AWMAD introduce median filtering and mean filtering to suppress noise interference, respectively. PLPF linearly weights the filtered image and the noisy image to enhance its robustness to noise. AWMAD designs a fitting function based on adaptive mean filtering to calculate low-frequency components, thereby reducing the impact of noise on edge feature extraction. Although PLPF and AWMAD achieved satisfactory segmentation performance,

spatial coherence and boundary preservation on multiple test images, the median filter and mean filter are only effective against single types of noise and exhibit poor suppression capabilities for mixed noise. Therefore, there is still a certain gap between their performance and that of the proposed model. ACFDI uses Fourier transform to extract low-frequency and high-frequency components from noisy images. The evolving curve is initialized using the low-frequency region position matrix to improve computational efficiency. The high-frequency region position matrix is used as weights to weight the intensity information within local regions, thereby enhancing noise resistance. However, high-frequency weights may misinterpret strong salt and pepper noise as true edges. This will result in false edges and impacting segmentation performance. DoGEED introduces anisotropic edge enhancement diffusion and the difference of Gaussian operator for image denoising and edge enhancement. This improves its robustness to noise interference to some extent. However, under mixed noise conditions, DoGEED's edge enhancement performance deteriorates. This makes its segmentation performance, spatial coherence and boundary preservation are worse than those of the proposed model.

Unlike the comparison models, TVDGFACM designs a total variation diffusion-guided energy term based on total variation. It can transform the noisy image to a total variation diffusion-guided domain to enhance the anti-noise capability. Moreover, TVDGFACM designs a more stable excitation weight estimation method. It adaptively adjusts pixel membership according to the regional dependencies between noisy pixel and its neighboring pixels. This helps to generate a smooth and coherent segmentation result. Therefore, compared with the comparison model, TVDGFACM exhibits better segmentation performance, spatial coherence and boundary preservation.

These results demonstrate the effectiveness of the proposed information fusion strategy. By fusing structural diffusion information, local contextual evidence, and global intensity statistics, TVDGFACM can make more reliable segmentation decisions under mixed noise conditions than methods relying on a single type of information.

To validate the convergence of the proposed model, we visualize its convergence curves on six test images, as shown in Figure 23. It can be seen from the visualization result that the energy function of the proposed model gradually decreases with the increase of the number of iterations. When the number of iterations is greater than 25, the energy function of the proposed model remains essentially unchanged on all test images.

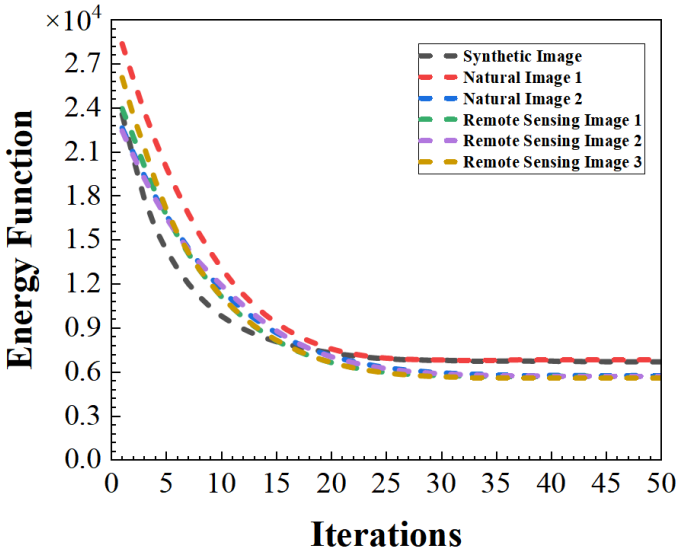


Figure 23. Convergence curves on six test images.

4.2.2 *Comparison with Deep Learning Methods*

Deep learning methods have been widely used in image segmentation tasks. To further validate the proposed model, we compared TVDGFACM with two deep learning methods, including HAUS [23] and DIP [24].

HAUS takes a pre-trained DINOv2-ViT-B14-REG as backbone to extract features. It treats superpixels as graph nodes and constructs a graph based on feature similarity. HAUS partitions the graph using recursive spectral clustering, where each connected component can be considered a subgraph, and further recursively partitioning on each subgraph until the stopping condition is met. Moreover, HAUS refines image edges using a conditional random field. Since HAUS uses a pre-trained network, it does not require retraining. For input preprocessing, HAUS resizes the input image according to the backbone downsampling ratio to extract 60×60 feature tokens.

DIP uses variational energy to initialize convolutional neural networks and minimizes the energy function by iteratively optimizing the network weights, rather than relying on pre-trained models or training sets. DIP uses a U-Net with skip connections as its core network structure. Moreover, it introduces a total variational regularization term to penalize boundary oscillations and pseudo-details. For the training setting, DIP takes

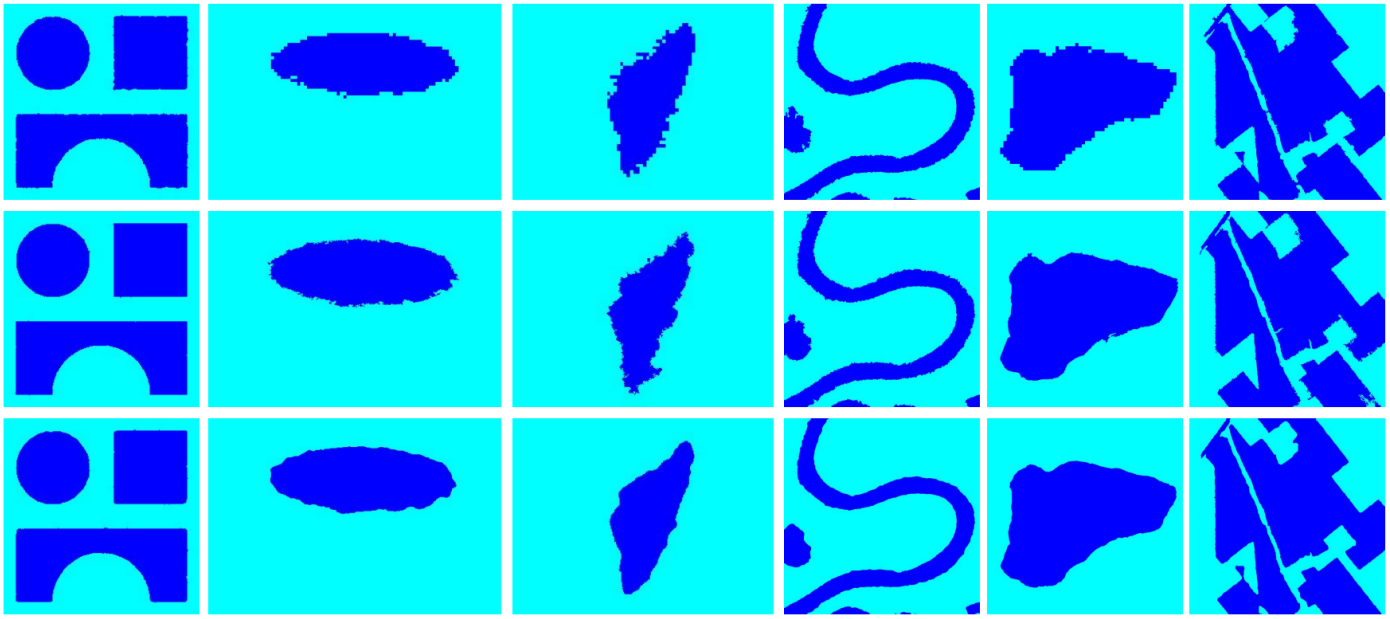


Figure 24. Segmentation results obtained by different deep learning methods. The first to last rows are the segmentation results obtained by HAUS, DIP, and TVDGFACM, respectively. (a) Synthetic image; (b) Natural image 1; (c) Natural image 2; (d) Remote sensing image 1; (e) Remote sensing image 2; (f) Remote sensing image 3.

an Adam as the optimizer and adds a zero-mean Gaussian perturbation to the input in each iteration to improve generalization. Moreover, DIP requires no preprocessing or training data, and directly segments a single noisy input image.

To ensure the fairness and objectivity of the comparison experiments, this paper fine-tunes some key hyperparameters of HAUS and DIP to achieve their optimal segmentation performance. HAUS's hyperparameters include the number of superpixels, the maximum number of subgraphs, and the smoothing threshold. The impacts of these hyperparameters on segmentation performance are reported in Table B1, Table B2, and Table B3, respectively. DIP's hyperparameters include the learning rate, total variation regularization parameter, and input perturbation standard deviation. Their impacts are reported in Table B4, Table B5, and Table B6, respectively. Based on the fine-tuned parameters, the segmentation results obtained by different deep learning methods on six test images polluted by 6% mixed noise are shown in Figure 24. Moreover, SA (%), MIoU (%), GMI (%) and BF1 (%) obtained by different deep learning methods are shown in Tables 7, 8, 9 and 10, respectively.

Compared with the above active contour models, deep learning methods, relying on deep features and global context, can obtain smoother and more coherent segmentation results. However, the optimization objectives of HAUS and DIP focus more on overall regional consistency and semantic similarity. They lack explicit and noise-adaptive constraints on precise edge locations and geometric shapes. As a result, deep learning methods tend to produce locally unstable clustering or excessive smoothing near object boundaries, which manifests as jagged edges and small fragmented regions. Although HAUS uses a conditional random field to refine image edges, it operates on the final segmentation result and cannot compensate for edge loss during the optimization process. In contrast, TVDGFACM can suppress noise interference and preserves image edges by guiding different total variational diffusion mechanisms during model iteration. Therefore, its segmentation performance, spatial coherence and boundary preservation are significantly better than those of deep learning methods.

4.2.3 Comparison on Large-Scale Benchmark Dataset

To verify TVDGFACM's segmentation performance on large-scale benchmark datasets, this section compares TVDGFACM with deep learning methods on Crack500 dataset [25]. Crack500 is a dataset for road crack detection. We crop each image into four non-overlapping image regions, and retain only the regions containing more than 2000 crack pixels. Moreover, we added 5% mixed noise to each image. The partial experimental results obtained by different

Table 7. SA (%) obtained by different deep learning methods. The optimal results are in bold.

Method	Synthetic image	Natural image 1	Natural image 2	Remote sensing image 1	Remote sensing image 2	Remote sensing image 3
HAUS	99.48±0.33	98.80±0.23	98.42±0.39	99.19±0.33	98.16±0.24	98.37±0.13
DIP	**99.91±0.02**	99.17±0.22	98.31±0.35	99.05±0.32	99.10±0.02	97.54±1.50
TVDGFACM	99.80±0.10	**99.42±0.02**	**99.02±0.21**	**99.23±0.15**	**99.47±0.11**	**98.75±0.25**

Table 8. MIoU (%) obtained by different deep learning methods. The optimal results are in bold.

Method	Synthetic image	Natural image 1	Natural image 2	Remote sensing image 1	Remote sensing image 2	Remote sensing image 3
HAUS	98.97±0.65	95.72±0.71	93.70±1.61	97.76±0.92	96.15±0.51	96.72±0.25
DIP	**99.83±0.04**	97.07±0.75	93.04±1.42	97.36±0.90	98.10±0.04	95.15±2.84
TVDGFACM	99.59±0.21	**97.93±0.09**	**95.87±0.89**	**97.87±0.42**	**98.88±0.23**	**97.46±0.50**

Table 9. GMI (%) obtained by different deep learning methods. The optimal results are in bold.

Method	Synthetic image	Natural image 1	Natural image 2	Remote sensing image 1	Remote sensing image 2	Remote sensing image 3
HAUS	97.45±0.16	99.07±0.01	97.91±0.37	97.95±0.20	99.47±0.02	97.82±0.10
DIP	97.61±0.04	98.24±0.33	97.01±0.38	97.57±0.44	99.41±0.02	97.38±0.70
TVDGFACM	**97.78±0.08**	**99.22±0.01**	**98.66±0.06**	**98.48±0.01**	**99.49±0.01**	**98.25±0.02**

Table 10. BF1 (%) obtained by different deep learning methods. The optimal results are in bold.

Method	Synthetic image	Natural image 1	Natural image 2	Remote sensing image 1	Remote sensing image 2	Remote sensing image 3
HAUS	99.89±0.16	61.93±8.87	68.78±11.9	93.65±4.25	62.22±7.17	87.20±1.66
DIP	**100.0±0.00**	77.92±9.23	65.69±9.03	93.24±4.76	73.13±0.94	81.74±12.4
TVDGFACM	99.90±0.12	**86.90±3.09**	**76.08±7.82**	**94.14±1.95**	**86.04±6.44**	**90.98±3.00**

methods are shown in Figure 25, and the average segmentation performance is shown in Table 11.

It can be seen form the experiment results that TVDGFACM obtains smoother and more continuous object boundaries, and improves BF1 by 24.67% and 6.50% compared with HAUS and DIP, respectively. This advantage is mainly due to the fact that deep learning methods usually lack precise edge localization constraints, while TVDGFACM effectively fuses image edge information into optimization process by introducing a total variation diffusion-guided energy term, which enhances the boundary characterization capability.

Table 11. Average segmentation performance obtained by different methods. The optimal results are in bold.

Methods	SA (%)	MIoU (%)	GMI (%)	BF1 (%)
HAUS	96.36±1.38	80.51±4.53	95.29±0.70	29.77±9.69
DIP	97.39±1.25	84.53±2.03	96.07±0.99	47.94±13.4
TVDGFACM	**97.78±1.54**	**86.99±3.75**	**97.10±0.86**	**54.44±20.0**

To further validate the convergence of the proposed model on large-scale benchmark datasets, this section visualizes the average convergence curve of the proposed model on the Crack500 dataset, as shown in Figure 26. It can be seen from the visualization results that after approximately 30 iterations, the energy function of the proposed model no longer decreases, which indicated that the proposed model has converged.

4.2.4 Computational Cost and Scalability Analysis

To analyze the computational efficiency and applicability of the proposed model, this section provides a quantitative discussion from three aspects: running time in a single iteration, memory consumption, and scalability with image size, as shown in Table 12. It can be seen from the experimental results that the test images have different spatial sizes, which indicates that the proposed model is not constrained by the input image size and can adapt to images with any size. Moreover, the number of iterations on different test images does not change significantly with image size, which indicates that the proposed model has strong convergence. It is worth noting that the running time and memory consumption increase with image size. Currently, the proposed model can meet the engineering feasibility requirements for processing standard resolution images on ordinary CPU platforms. For larger-scale or real-time applications, throughput and response speed can be further improved through parallelization and vectorization, or by employing multi-scale and hardware acceleration strategies.

4.3 Ablation Experiment

To evaluate the effectiveness of some important components in the proposed model, this section

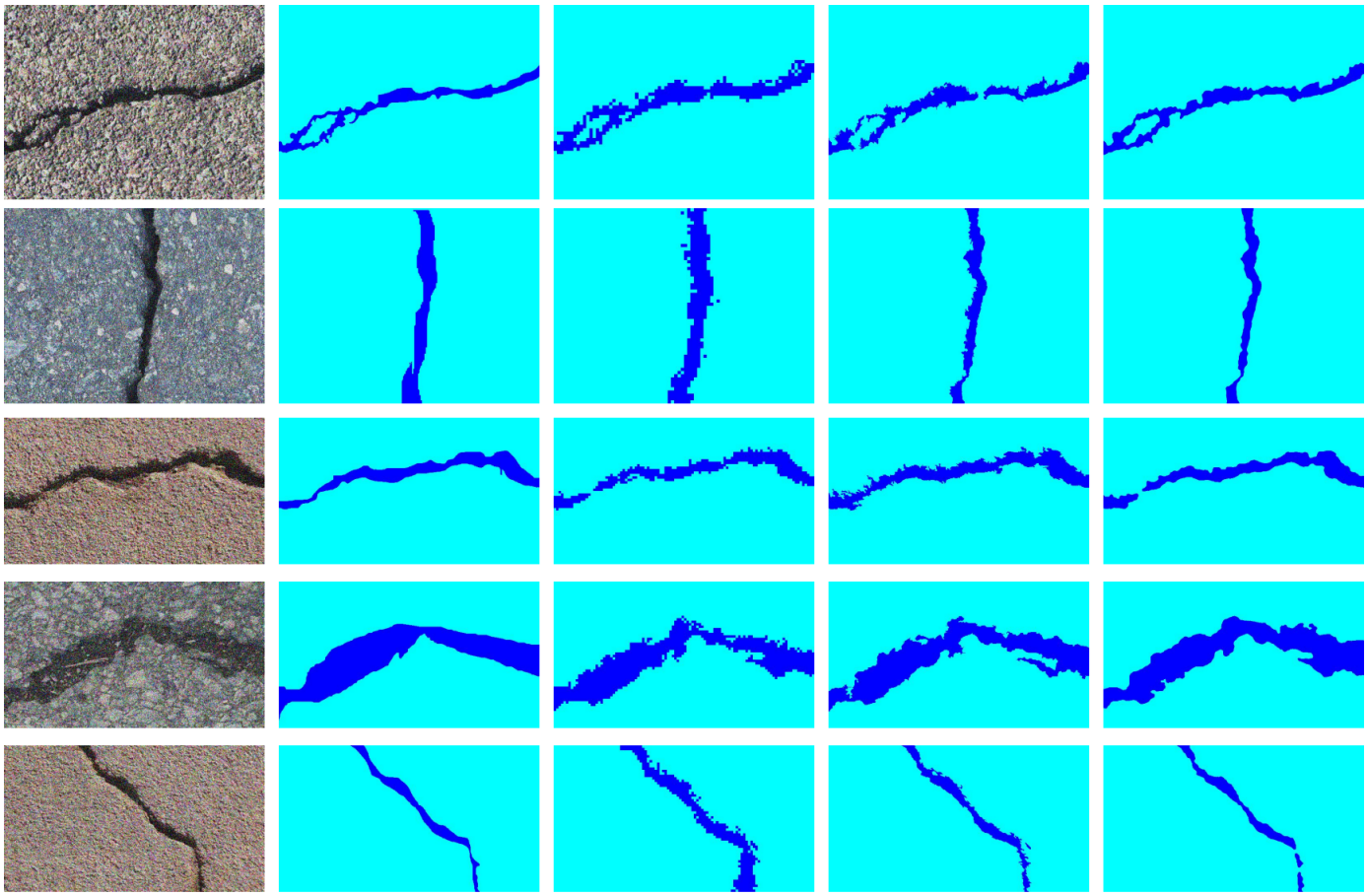


Figure 25. Segmentation results obtained by different methods. The first to last rows are the segmentation results obtained by HAUS, DIP, and TVDGFACM, respectively. (a) Noisy images; (b) Ground truth; (c) HAUS; (d) DIP; (e) TVDGFACM.

Table 12. Quantitative discussion on running time, memory consumption and scalability with image size.

Images	Image size	Iterations	Time (s)	Memory (MB)
Synthetic image	256×256	34	0.59	7.73
Natural image 1	321×481	35	1.40	7.75
Natural image 2	225×300	37	0.61	7.77
Remote sensing image 1	600×600	33	3.37	7.83
Remote sensing image 2	600×600	40	3.35	7.88
Remote sensing image 3	600×600	38	3.27	7.93

conducts the ablation experiments on the local energy term and total variation diffusion-guided energy term respectively. This helps reveal the contributions of different fused information sources to the final segmentation decision.

The local energy term smooths the pixel membership by measuring the noise distribution and the degree to which pixels are polluted by noise. This helps generate a smooth and coherent segmentation result. To validate its effectiveness, this section removes the local energy term from TVDGFACM. Mixed noise with different densities is added to the test images. The segmentation results obtained on six test images polluted by 6% mixed noise after removing the local energy term are shown in Figure 27. The segmentation performance obtained on six test images polluted by mixed noise with different densities after removing the local energy term is shown in Table 13. "x" represents that the local energy term is removed, and "✓" indicates the presence of the local energy term.

As can be seen from Figure 27, after removing the local energy term, the segmentation result on synthetic image is basically not affected. TVDGFACM can still accurately capture object edges. This is because the

Table 13. Segmentation performance obtained on six test images after removing the local energy term.

Images	Metric / Noisy Density	Segmentation Accuracy (%)					Mean Intersection over Union (%)				
		2	4	6	8	10	2	4	6	8	10
Synthetic Image 1	x	99.65	99.61	99.48	99.26	99.13	99.30	99.23	98.95	98.53	98.28
	✓	99.89	99.80	99.66	99.55	99.36	99.80	99.33	98.93	98.91	98.49
Natural Image 1	x	99.27	99.10	98.82	98.89	98.77	97.42	96.84	95.89	96.12	95.72
	✓	99.47	99.46	99.27	99.26	99.23	98.11	98.07	97.39	97.35	97.25
Natural Image 2	x	98.98	98.81	98.27	98.59	97.79	95.82	95.12	93.05	94.28	91.47
	✓	99.12	99.02	98.71	98.79	98.56	96.33	95.93	94.65	94.93	94.05
Remote Sensing Image 1	x	99.27	99.06	98.75	98.43	98.31	97.97	97.40	96.55	95.71	95.39
	✓	99.43	99.34	99.22	99.10	99.07	98.40	98.17	97.85	97.51	97.42
Remote Sensing Image 2	x	98.67	98.38	98.04	97.70	97.07	97.32	96.73	96.07	95.39	94.19
	✓	99.61	99.56	99.45	99.41	99.34	99.17	99.07	98.82	98.75	98.60
Remote Sensing Image 3	x	97.71	96.80	96.51	95.71	95.39	95.28	93.48	92.93	91.40	90.79
	✓	99.05	98.94	98.71	98.54	98.48	98.08	97.86	97.38	97.06	96.93

Table 14. Segmentation performance obtained on six test images after removing the total variation diffusion-guided energy term.

Images	Metric / Noisy Density	Segmentation Accuracy (%)					Mean Intersection over Union (%)				
		2	4	6	8	10	2	4	6	8	10
Synthetic Image 1	x	99.68	99.24	98.90	98.34	97.54	99.36	98.50	97.83	96.72	95.20
	✓	99.89	99.80	99.66	99.55	99.36	99.80	99.33	98.93	98.91	98.49
Natural Image 1	x	97.16	96.01	94.52	92.11	90.19	90.83	87.67	84.00	78.83	75.20
	✓	99.47	99.46	99.27	99.26	99.23	98.11	98.07	97.39	97.35	97.25
Natural Image 2	x	96.77	96.08	94.96	93.92	93.63	88.08	85.87	82.52	79.85	78.81
	✓	99.12	99.02	98.71	98.79	98.56	96.33	95.93	94.65	94.93	94.05
Remote Sensing Image 1	x	99.12	98.44	97.56	96.50	95.27	97.57	95.69	93.36	90.66	87.56
	✓	99.43	99.34	99.22	99.10	99.07	98.40	98.17	97.85	97.51	97.42
Remote Sensing Image 2	x	90.06	87.91	86.65	85.16	83.35	81.54	78.10	76.16	73.90	71.24
	✓	99.61	99.56	99.45	99.41	99.34	99.17	99.07	98.82	98.75	98.60
Remote Sensing Image 3	x	97.17	95.53	94.14	91.91	89.62	94.32	91.11	88.46	84.28	80.09
	✓	99.05	98.94	98.71	98.54	98.48	98.08	97.86	97.38	97.06	96.93

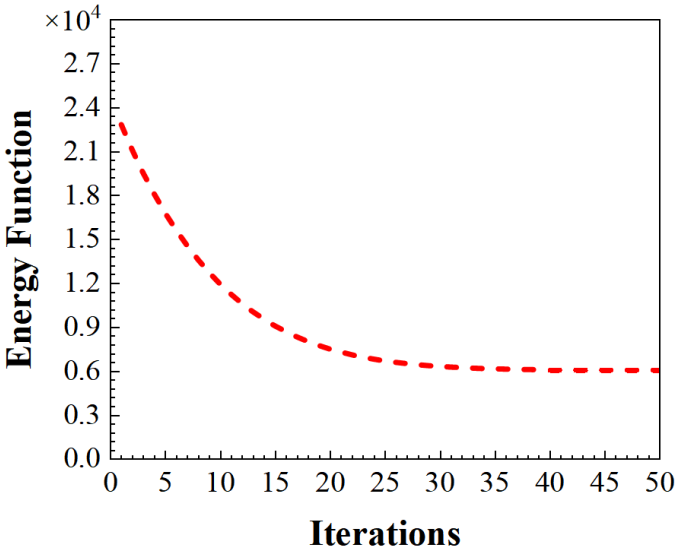


Figure 26. Average convergence curve on Crack500 dataset.

intensity distribution of different regions in synthetic image is homogeneous. Although mixed noise destroys its intensity statistical characteristics, the total variation diffusion-guided energy term can transform noisy image into a total variation diffusion-guided domain to restore its intensity distribution. Therefore, in synthetic images, using only the global energy term can achieve ideal segmentation performance. However, in the natural and remote sensing images, removing the local energy term significantly impacts segmentation result. This results in a large number of misclassified pixels in the segmentation result. As can be seen from Table 13, this impact becomes increasingly significant as the noise density increases. The main reason for this phenomenon lies in the fact that the intensity distribution characteristics of different regions in natural and remote sensing images are rather complex. Although the total variation

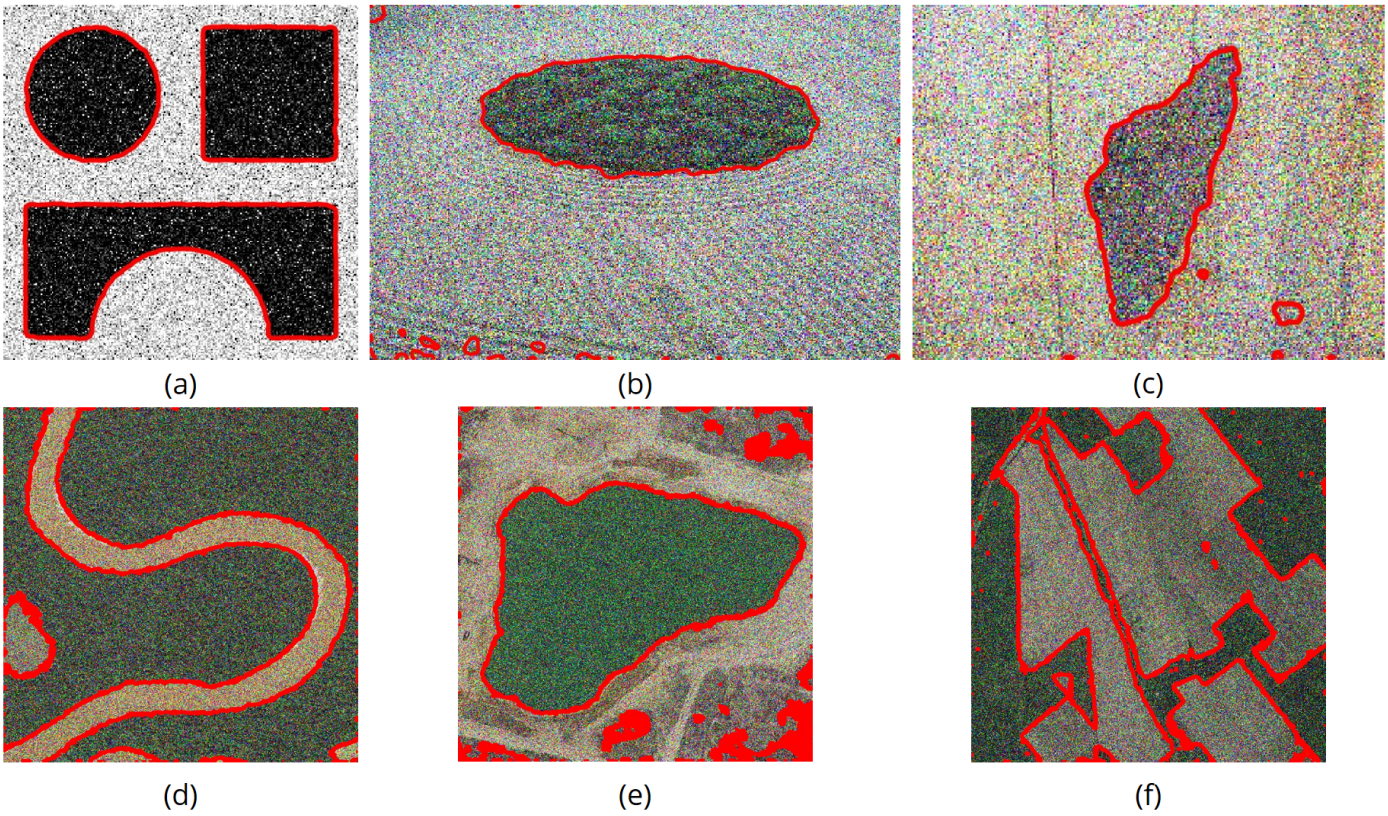


Figure 27. Segmentation results obtained on six test images after removing the local energy term. (a) Synthetic image; (b) Natural image 1; (c) Natural image 2; (d) Remote sensing image 1; (e) Remote sensing image 2; (f) Remote sensing image 3.

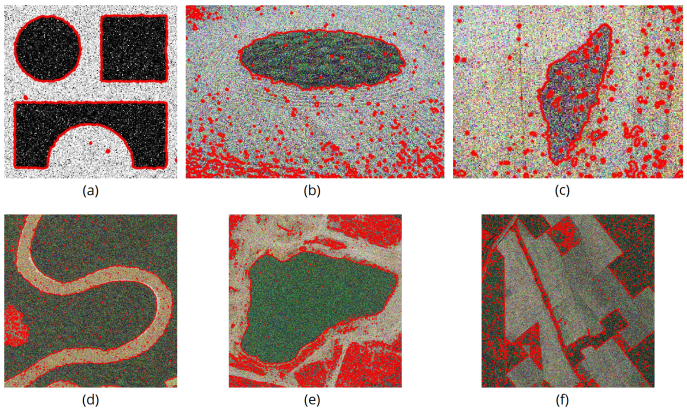


Figure 28. Segmentation results obtained on six test images after removing the total variation diffusion-guided energy term. (a) Synthetic image; (b) Natural image 1; (c) Natural image 2; (d) Remote sensing image 1; (e) Remote sensing image 2; (f) Remote sensing image 3.

diffusion-guided energy term reduces the impact of noise on intensity distributions, it is still difficult to accurately segment natural and remote sensing images by using only the global energy term.

The total variation diffusion-guided energy term transforms the noisy image from the intensity domain to a total variation diffusion-guided domain. It can restore intensity distribution characteristics while preserving image details such as edges. To validate its effectiveness, this section removes the total variation diffusion-guided energy term from TVDGFACM. Meanwhile, mixed noise with different densities is added to the test images. The segmentation results obtained on six test images polluted by 6% mixed noise after removing the total variation diffusion-guided energy term are shown in Figure 28. The segmentation performance obtained on six test images polluted by mixed noise with different densities after removing the total variation diffusion-guided energy term is shown in Table 14. Among them, "x" represents that the total variation diffusion-guided energy term is removed, and "✓" indicates the presence of the total variation diffusion-guided energy term.

According to Figure 28, after removing the total variation diffusion-guided energy term, there will be a large number of misclassified pixels in the segmentation result. As can be seen from Table 14, after removing the total variation diffusion-guided energy term, the segmentation performance will rapidly decline as the noise density increases. This is because noise destroys the image structure and

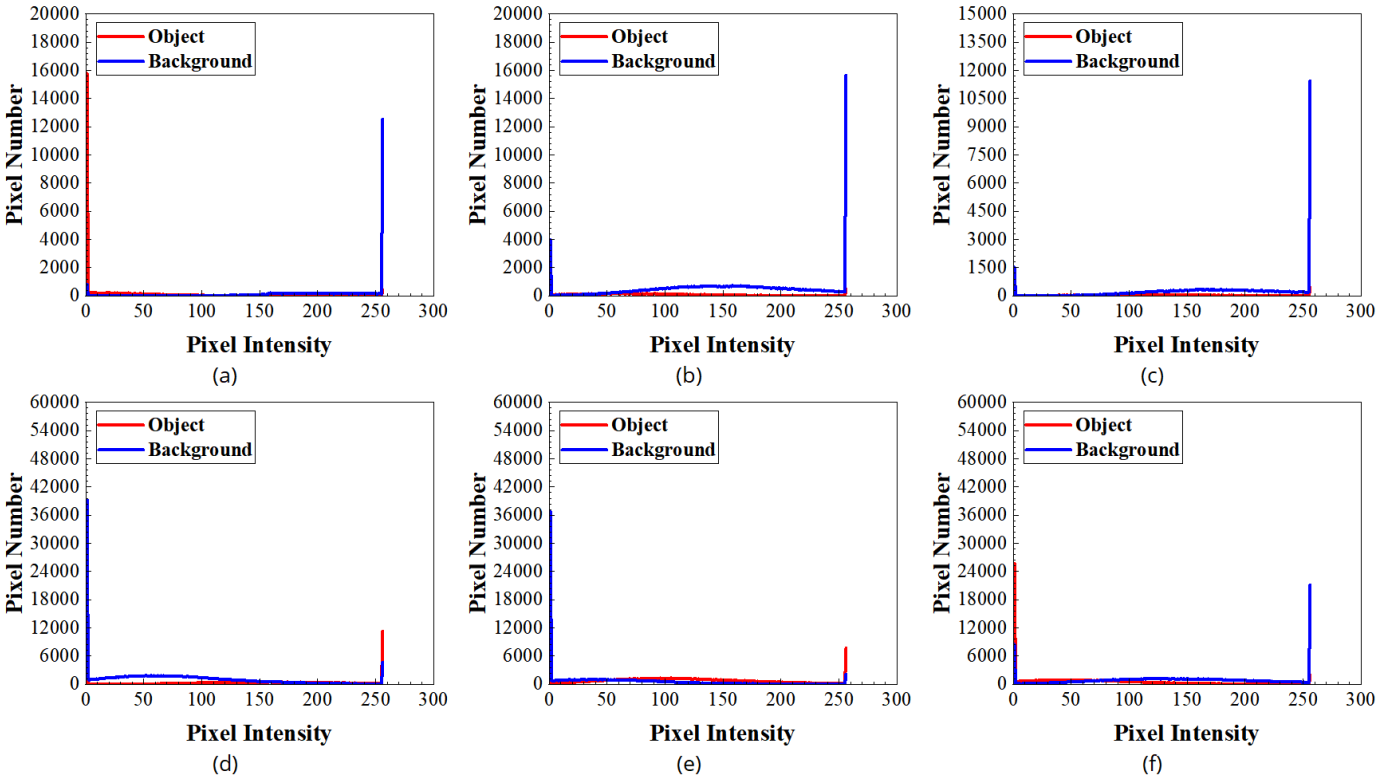


Figure 29. Intensity distributions of different regions in noisy images. (a) Synthetic image; (b) Natural image 1; (c) Natural image 2; (d) Remote sensing image 1; (e) Remote sensing image 2; (f) Remote sensing image 3.

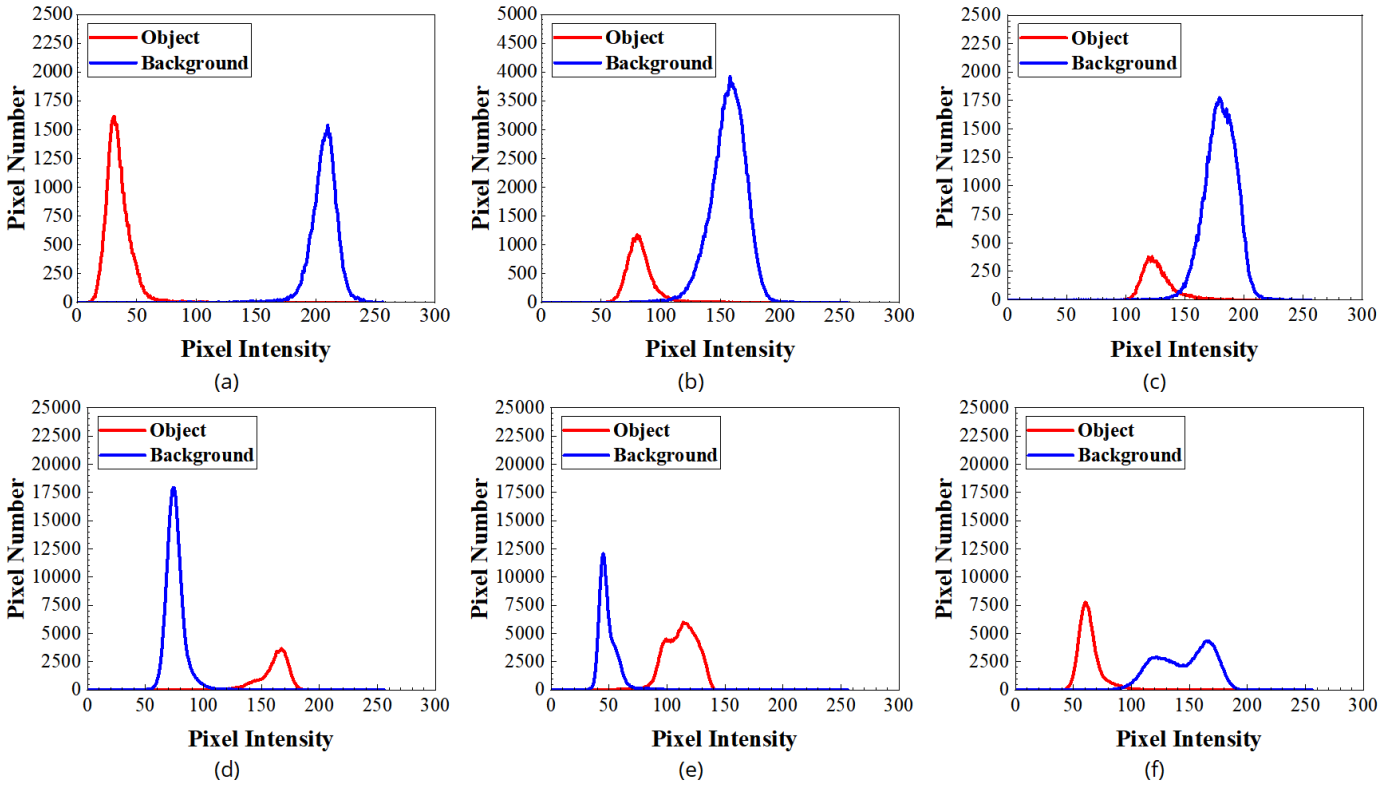


Figure 30. Intensity distributions of different regions in corrected images. (a) Synthetic image; (b) Natural image 1; (c) Natural image 2; (d) Remote sensing image 1; (e) Remote sensing image 2; (f) Remote sensing image 3.

intensity distribution characteristics. Under the impact of high noise, it is difficult to effectively distinguish objects and background regions solely relying on the local energy term. As can be seen from Figures 29

and 30, the pixel intensities in noisy image are mainly distributed around 0 and 255. In contrast, the intensity distributions of different regions in corrected image have two clearly separable peaks, which helps segmentation methods to obtain the optimal decision boundary.

5 Conclusion

To further improve the robustness of FACM to noise and optimize the excitation weight estimation method, this paper proposes TVDGFACM. TVDGFACM first introduced total variation into FACM. It can restore intensity distribution characteristics and protect image edges while avoiding staircase artifacts by guiding different diffusion mechanisms. Then, TVDGFACM designs a local structure entropy to detect noisy pixels, and reconstructs the excitation weight estimation method based on the regional dependencies between noisy pixel and its neighboring pixels. This helps to generate a smooth and coherent segmentation result. Finally, a balance factor based on noise density is constructed to adjust the weights between the global energy term and the local energy term, thereby enhancing the stability during curve evolution. Overall, TVDGFACM provides a concise information fusion framework that integrates diffusion-based structural information, global intensity statistics, and local contextual dependencies for robust noisy image segmentation. Under the impact of mixed noise with different densities, experiments are conducted using the synthetic, natural, remote sensing images and large-scale benchmark dataset. Experimental results show the advancement and effectiveness of TVDGFACM.

In the future work, we will continue to explore more effective segmentation model frameworks. For example, we will comprehensively analyze the damage to the image quality from the noise formation mechanism. We will consider the impact of other interference factors (such as the intensity inhomogeneity) on the intensity distribution characteristics. Furthermore, we will also attempt to explore how to introduce deep learning techniques into FACM to enhance feature representation capability and segmentation performance.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the National Natural Science Foundation of China under Grant 62473304 and Grant U22A2045.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Jiang, X., Feng, C., Han, T., Wu, B., & Yang, Y. (2025). Active contour model combining frequency domain information for noisy vessel image segmentation. *Pattern Recognition*, 112126. [CrossRef]

[2] Li, Z., Zhang, F., Wang, G., Weng, G., & Chen, Y. (2026). An active contour model based on Kullback–Leibler divergence and morphology for image segmentation with edge leakage. *Signal Processing, 238*, 110143. [CrossRef]

[3] Li, Z., Tang, S., Zeng, Y., Chai, S., Ye, W., Yang, F., & Huang, K. (2024). A level-set method with a multiplicative–additive constraint model for image segmentation and bias correction. *Knowledge-Based Systems, 297*, 111972. [CrossRef]

[4] Vijai, A., Padmavathi, S., & Venkataraman, D. (2025). Artificial intelligence based semantic segmentation on aerial images with variational mode decomposition. *Engineering Applications of Artificial Intelligence, 156*, 111140. [CrossRef]

[5] Ge, P., Wan, M., Xu, Y., Kong, X., Kang, Y., Weng, G., ... & Chen, Q. (2025). A hybrid active contour model using local region-based K-medoids for infrared image segmentation. *Expert Systems with Applications, 290*, 128325. [CrossRef]

[6] Zhao, W., Sang, J., Shu, Y., & Li, D. (2024). Robust image segmentation and bias field correction model based on image structural prior constraint. *Expert Systems with Applications, 251*, 123961. [CrossRef]

[7] Dickson, A. J., Linsely, J. A., Daniel, V. A. A., & Rahul, K. (2024). Sparse deep belief network coupled with extended local fuzzy active contour model-based liver cancer segmentation from abdomen CT images. *Medical & Biological Engineering & Computing, 62*(5), 1361-1374. [CrossRef]

[8] Krinidis, S., & Chatzis, V. (2009). Fuzzy energy-based

active contours. *IEEE Transactions on Image Processing*, *18*(12), 2747-2755. [CrossRef]

[9] Wu, Y., Ma, W., Gong, M., Li, H., & Jiao, L. (2015). Novel fuzzy active contour model with kernel metric for image segmentation. *Applied Soft Computing*, *34*, 301-311. [CrossRef]

[10] Krinidis, S., & Krinidis, M. (2012, September). Fuzzy energy-based active contours exploiting local information. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 175-184). Berlin, Heidelberg: Springer Berlin Heidelberg. [CrossRef]

[11] Mondal, A., Ghosh, S., & Ghosh, A. (2016). Robust global and local fuzzy energy based active contour for image segmentation. *Applied Soft Computing*, *47*, 191-215. [CrossRef]

[12] Fang, J., Liu, H., Zhang, L., Liu, J., & Liu, H. (2021). Region-edge-based active contours driven by hybrid and local fuzzy region-based energy for image segmentation. *Information Sciences*, *546*, 397-419. [CrossRef]

[13] Zhang, H., Tang, L., & He, C. (2019). A variational level set model for multiscale image segmentation. *Information Sciences*, *493*, 152-175. [CrossRef]

[14] Rabelo, R. A. L., Ribeiro, P. H. A., Santos, W. M. S., Silva, R. C. C., & Souza, J. C. O. (2025). A non-monotone proximal point method for image reconstruction using non-convex total variation models. *Computers and Electrical Engineering*, *126*, 110491. [CrossRef]

[15] Liu, X. (2025). Hyperspectral mixed noise removal using nonconvex low-rank and total generalized variation. *Signal Processing: Image Communication*, *138*, 117344. [CrossRef]

[16] Abualigah, L., Almomani, M. H., Alomari, S. A., Zitar, R. A., Snasel, V., Saleem, K., ... & Ezugwu, A. E. (2025). A control-driven transition strategy for enhanced multi-level threshold image segmentation optimization. *Egyptian Informatics Journal*, *30*, 100646. [CrossRef]

[17] Ren, J., Shang, R., Chen, J., Zhang, W., Feng, J., Liu, M., ... & Stolkin, R. (2024). Sar image segmentation based on complicated region-sensitive adaptive superpixel generation and hybrid edge correction. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1-17. [CrossRef]

[18] Cui, H., Xie, Z., Zeng, W., Ma, R., Zhang, Y., Yin, Q., & Xu, Z. (2024). Intuitionistic fuzzy local information C-means algorithm for image segmentation. *Information Sciences*, *681*, 121205. [CrossRef]

[19] Song, B., & Chan, T. (2002). A fast algorithm for level set based optimization. *UCLA Cam Report*, *2*(68). https://ww3.math.ucla.edu/camreport/cam02-68.pdf

[20] Huang, K., Ouyang, J., & Weng, G. (2025). Active contour model based on fuzzy C-means and local pre-fitting energy for image segmentation. *Signal, Image and Video Processing*, *19*(2), 193. [CrossRef]

[21] Dong, B., Bu, Q., Zhu, Z., & Ni, J. (2025). An active contour model with adaptive weighted mean filtering and anisotropic diffusion filtering. *Signal Processing*, *237*, 110071. [CrossRef]

[22] Zhang, F., Bi, X., Wang, G., Weng, G., & Chen, Y. (2025). Anisotropic edge-enhanced active contour model with Gaussian difference for robust multi-category image segmentation. *The Visual Computer*, *41*(12), 10153-10170. [CrossRef]

[23] Rossetti, S., & Pirri, F. (2024). Hierarchy-agnostic unsupervised segmentation: parsing semantic image structure. *Advances in Neural Information Processing Systems*, *37*, 98898-98935.

[24] Benfenati, A., Catozzi, A., Franchini, G., & Porta, F. (2025). Unsupervised noisy image segmentation using Deep Image Prior. *Mathematics and Computers in Simulation*. [CrossRef]

[25] Zhang, L., Yang, F., Zhang, Y. D., & Zhu, Y. J. (2016, September). Road crack detection using deep convolutional neural network. In *2016 IEEE international conference on image processing (ICIP)* (pp. 3708-3712). IEEE. [CrossRef]

Appendix A

As shown in Eq. (18), the energy function $E(u)$ of the proposed model can be formulated as:

$$E_T(u) = E_D + E_G^A(u) + E_G^B(u) + E_L^A(u) + E_L^B(u) \quad \text{(A1)}$$

where

$$E_D = \int (I(x) - I_C(x))^2 dx + \frac{\gamma}{\theta(x)} \int |\nabla I_C(x)|^{\theta(x)} dx \quad \text{(A2)}$$

$$E_G^A(u) = \int (1 - \beta(x))[u(x)]^m (I_C(x) - v_1)^2 dx \quad \text{(A3)}$$

$$E_G^B(u) = \int (1-\beta(x))[1-u(x)]^m (I_C(x)-v_2)^2 dx \quad \text{(A4)}$$

$$E_L^A(u) = \int \beta(x)[u(x)]^m \int h_{xy}[1 - u(y)]^m (I_C(y) - v_1)^2 dy dx \quad \text{(A5)}$$

$$E_L^B(u) = \int \beta(x)[1 - u(x)]^m \int h_{xy}[u(y)]^m (I_C(y) - v_2)^2 dy dx \quad \text{(A6)}$$

Since E_D does not involve u, this section only needs to prove the convexity of $E_G^A(u)$, $E_G^B(u)$, $E_L^A(u)$ and $E_L^B(u)$ with respect to u. Among them, $E_G^A(u)$ in Eq. (A3) can be written as:

$$E_G^A(u) = \int e_G^A(u) dx \quad \text{(A7)}$$

where

$$e_G^A(u) = (1 - \beta(x))[u(x)]^m (I_C(x) - v_1)^2 \quad \text{(A8)}$$

Let $x_1, x_2 \in \Omega$, for $\forall \theta \in [0,1]$, we have

$$\theta x_1 + (1-\theta)x_2 = \theta(x_1 - x_2) + x_2 \in \Omega \tag{A9}$$

Therefore, the domain of $e_G^A(u)$ is convex with respect to u. The second order derivative of Eq. (A8) with respect to u can be formulated as:

$$\begin{aligned}\frac{\partial^2 e_G^A}{\partial u^2} &= \frac{\partial}{\partial u}\left(\frac{\partial e_G^A}{\partial u}\right) \\ &= (1-\beta(x))m(m-1)[u(x)]^{m-2}(I_C(x)-v_1)^2\end{aligned} \tag{A10}$$

Since $m > 1$, $u(x) \in [0,1]$, and $(I_c(x) - v_1)^2 \geq 0$, we have:

$$\frac{\partial^2 e_A^4}{\partial u^2} \geq 0 \tag{A11}$$

Since the domain of $e_A^4(u)$ is convex and $\partial^2 e_A^4/\partial u^2 \geq 0$, $e_A^4(u)$ is convex. Therefore, $\forall x_1, x_2 \in \Omega$ and $\theta \in [0,1]$, we have:

$$e_A^4(\theta x_1 + (1-\theta)x_2) \leq \theta e_A^4(x_1) + (1-\theta)e_A^4(x_2) \tag{A12}$$

Calculating the integral of the two sides of (A10), we have:

$$\begin{aligned}\int e_A^4(\theta x_1 + (1-\theta)x_2)dx \leq \theta \int e_A^4(x_1)dx \\ + (1-\theta)\int e_A^4(x_2)dx\end{aligned} \tag{A13}$$

With $E_A^4(u) = \int e_A^4(u)dx$, Eq. (A11) can be rewritten as:

$$\begin{aligned}E_A^4(\theta x_1 + (1-\theta)x_2) \leq \theta E_A^4(x_1) \\ + (1-\theta)E_A^4(x_2)\end{aligned} \tag{A14}$$

Hence, $E_A^4(u)$ is convex with respect to u. In the same way, we can also prove that $E_B^4(u)$ is convex with respect to u.

Again let:

$$E_L^4(u) = \int \beta(x)[u(x)]^m E_{L_1}^4(u)dx \tag{A15}$$

$$E_{L_1}^4(u) = \int h_{xy}[1-u(y)]^m (I_c(y)-v_1)^2 dy \tag{A16}$$

$$e_{L_1}^4(u) = h_{xy}[1-u(y)]^m (I_c(y)-v_1)^2 \tag{A17}$$

Let $y_1, y_2 \in \Omega$, for $\forall \theta \in [0,1]$, we have:

$$\theta y_1 + (1-\theta)y_2 = \theta(y_1 - y_2) + y_2 \in \Omega \tag{A18}$$

Therefore, the domain of $e_{L_1}^4(u)$ is convex with respect to u. The second order derivative of Eq. (A17) with respect to u can be formulated as:

$$\begin{aligned}\frac{\partial^2 e_{L_1}^4}{\partial u^2} &= \frac{\partial}{\partial u}\left(\frac{\partial e_{L_1}^4}{\partial u}\right) \\ &= h_{xy}m(m-1)[1-u(y)]^{m-2}(I_c(y)-v_1)^2\end{aligned} \tag{A19}$$

Since $h_{xy} \geq 0$, $m > 1$, $u(y) \in [0,1]$, and $(I_c(y) - v_1)^2 \geq 0$, we have:

$$\frac{\partial^2 e_{L_1}^4}{\partial u^2} \geq 0 \tag{A20}$$

Since the domain of $e_{L_1}^A(u)$ is convex and $\partial^2 e_{L_1}^A/\partial u^2 \geq 0$, $e_{L_1}^A(u)$ is convex. Therefore, $\forall y_1, y_2 \in \Omega$ and $\theta \in [0,1]$, we have:

$$e_{L_1}^A(\theta y_1 + (1-\theta)y_2) \leq \theta e_{L_1}^A(y_1) + (1-\theta)e_{L_1}^A(y_2) \tag{A21}$$

Calculating the integral of the two sides of (A21), we have:

$$\begin{aligned}\int e_{L_1}^A(\theta y_1 + (1-\theta)y_2)dy \leq \theta \int e_{L_1}^A(y_1)dy \\ + (1-\theta)\int e_{L_1}^A(y_2)dy\end{aligned} \tag{A22}$$

With $E_{L_1}^A(u) = \int e_{L_1}^A(u)dx$, Eq. (A22) can be rewritten as:

$$E_{L_1}^A(\theta y_1 + (1-\theta)y_2) \leq \theta E_{L_1}^A(y_1) + (1-\theta)E_{L_1}^A(y_2) \tag{A23}$$

Hence, $E_{L_1}^A(u)$ is convex with respect to u. Again let

$$\bar{E}_L^A(u) = \beta(x)[u(x)]^m E_{L_1}^A(u) \tag{A24}$$

Since $E_{L_1}^A(u)$ is convex and $\beta(x)[u(x)]^m \geq 0$, $\bar{E}_L^A$ is convex. Therefore, $\forall x_1, x_2 \in \Omega$ and $\theta \in [0,1]$, we have:

$$\bar{E}_L^A(\theta x_1 + (1-\theta)x_2) \leq \theta \bar{E}_L^A(x_1) + (1-\theta)\bar{E}_L^A(x_2) \tag{A25}$$

Calculating the integral of the two sides of (A25), we have:

$$\begin{aligned}\int \bar{E}_L^A(\theta x_1 + (1-\theta)x_2)dx \leq \theta \int \bar{E}_L^A(x_1)dx \\ + (1-\theta)\int \bar{E}_L^A(x_2)dx\end{aligned} \tag{A26}$$

Since $E_{L_1}^A(u) = \int \bar{E}_L^A(u)dx$, Eq. (A26) can be rewritten as:

$$E_{L_1}^A(\theta x_1 + (1-\theta)x_2) \leq \theta E_{L_1}^A(x_1) + (1-\theta)E_{L_1}^A(x_2) \tag{A27}$$

Hence, $E_{L_1}^A(u)$ is convex with respect to u. In the same way, we can also prove that $E_L^B(u)$ is convex with respect to u.

Based on the above description, since $E_G^A(u)$, $E_G^B(u)$, $E_L^A(u)$, and $E_L^B(u)$ are all convex with respect to u, the energy function $E_T(u)$ of the proposed model is also convex with respect to u.

Appendix B

To ensure fairness and objectivity of the comparison experiments with deep learning methods, this section fine-tunes some key hyperparameters of HAUS and DIP based on experimental images polluted by 6% mixed noise to achieve their optimal segmentation performance. HAUS's hyperparameters include the number of superpixels (Table B1), the maximum number of subgraphs (Table B2), and the smoothing threshold (Table B3). DIP's hyperparameters include the learning rate (Table B4), total variation regularization parameter (Table B5), and input perturbation standard deviation (Table B6). The optimal values are highlighted in bold in each table.

Table B1. Impact of the number of superpixels on segmentation performance. The optimal results are in bold.

Images	Metrics	20	40	60	80	100
Synthetic image	SA (%)	98.37	99.41	**99.75**	99.68	99.64
	MIoU (%)	95.98	97.94	**98.58**	98.45	98.37
Natural image 1	SA (%)	98.85	98.87	**98.91**	98.82	98.80
	MIoU (%)	95.89	95.99	**96.11**	95.85	95.74
Natural image 2	SA (%)	98.58	98.45	98.68	**98.79**	98.43
	MIoU (%)	92.62	92.22	92.95	**93.34**	92.02
Remote sensing image 1	SA (%)	98.88	**98.98**	98.08	98.75	98.44
	MIoU (%)	96.69	**96.73**	94.88	96.31	95.06
Remote sensing image 2	SA (%)	97.14	97.90	98.09	**98.19**	97.88
	MIoU (%)	94.06	95.58	95.99	**96.19**	95.54
Remote sensing image 3	SA (%)	96.20	94.02	**97.55**	96.02	97.31
	MIoU (%)	93.05	89.52	**95.11**	92.68	94.35

Table B2. Impact of the maximum number of subgraphs on segmentation performance. The optimal results are in bold.

Images	Metrics	10	15	20	25	30
Synthetic image	SA (%)	99.55	**99.58**	99.39	99.37	99.00
	MIoU (%)	98.29	**98.38**	98.21	98.18	97.86
Natural image 1	SA (%)	98.73	**98.99**	98.92	98.85	98.67
	MIoU (%)	95.53	**96.43**	96.14	95.89	95.23
Natural image 2	SA (%)	96.88	97.22	**97.79**	97.70	97.73
	MIoU (%)	92.65	93.18	**93.42**	93.26	93.29
Remote sensing image 1	SA (%)	97.64	98.00	98.42	**98.88**	98.54
	MIoU (%)	95.02	95.42	96.20	**96.69**	96.33
Remote sensing image 2	SA (%)	97.79	97.61	**97.82**	97.14	96.95
	MIoU (%)	95.34	95.06	**95.48**	94.06	93.68
Remote sensing image 3	SA (%)	97.08	97.58	97.40	**97.89**	97.83
	MIoU (%)	94.96	95.38	95.21	**95.81**	95.32

Table B3. Impact of smoothing threshold on segmentation performance. The optimal results are in bold.

Images	Metrics	0	0.05	0.1	0.15	0.2
Synthetic image	SA (%)	99.26	99.61	99.41	**99.79**	99.33
	MIoU (%)	97.66	98.31	98.11	**98.67**	97.80
Natural image 1	SA (%)	98.86	98.77	**98.79**	98.71	98.78
	MIoU (%)	95.90	95.61	**95.73**	95.43	95.67
Natural image 2	SA (%)	98.36	**98.66**	98.62	97.91	97.06
	MIoU (%)	92.71	**93.69**	93.34	90.56	87.91
Remote sensing image 1	SA (%)	97.19	98.53	**98.65**	98.14	97.75
	MIoU (%)	94.96	96.19	**96.64**	96.23	95.64
Remote sensing image 2	SA (%)	97.13	97.25	97.34	**97.45**	97.34
	MIoU (%)	95.05	95.31	95.49	**95.72**	95.51
Remote sensing image 3	SA (%)	97.10	**97.77**	97.32	97.55	97.39
	MIoU (%)	95.35	**95.81**	95.58	95.75	95.44

Table B4. Impact of learning rate on segmentation performance. The optimal results are in bold.

Images	Metrics	0.001	0.005	0.01	0.05	0.1
Synthetic image	SA (%)	99.20	99.84	**99.87**	99.85	99.84
	MIoU (%)	98.41	99.68	**99.74**	99.70	99.68
Natural image 1	SA (%)	98.96	98.66	**99.30**	98.76	99.10
	MIoU (%)	96.35	95.39	**97.51**	95.73	96.84
Natural image 2	SA (%)	98.41	98.66	98.64	**98.80**	98.73
	MIoU (%)	93.52	93.75	93.68	**94.15**	93.85
Remote sensing image 1	SA (%)	98.18	99.03	**99.08**	98.87	98.55
	MIoU (%)	95.11	97.30	**97.44**	96.88	95.99
Remote sensing image 2	SA (%)	98.96	**98.98**	98.88	98.32	97.42
	MIoU (%)	97.82	**97.86**	97.65	96.51	94.72
Remote sensing image 3	SA (%)	97.35	**98.10**	97.93	97.82	97.73
	MIoU (%)	94.73	**96.19**	95.86	95.65	95.48

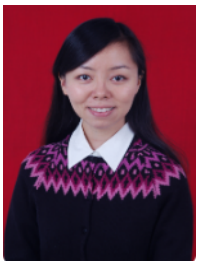
Table B6. Impact of input perturbation standard deviation on segmentation performance. The optimal results are in bold.

Images	Metrics	0.001	0.005	0.01	0.05	0.1
Synthetic image	SA (%)	99.85	99.87	99.86	**99.89**	99.88
	MIoU (%)	99.70	99.74	99.71	**99.79**	99.76
Natural image 1	SA (%)	99.17	99.05	**99.18**	99.07	98.63
	MIoU (%)	97.08	96.67	**97.11**	96.79	96.21
Natural image 2	SA (%)	98.65	**98.77**	98.71	98.58	98.33
	MIoU (%)	93.75	**94.03**	93.84	93.22	93.17
Remote sensing image 1	SA (%)	99.06	99.05	**99.11**	99.08	99.03
	MIoU (%)	97.39	97.35	**97.53**	97.51	97.42
Remote sensing image 2	SA (%)	98.92	98.91	98.97	**98.99**	98.93
	MIoU (%)	97.72	97.71	97.85	**97.88**	97.74
Remote sensing image 3	SA (%)	98.08	**98.15**	97.73	97.60	97.82
	MIoU (%)	96.15	**96.29**	95.47	95.22	95.66

Table B5. Impact of total variation regularization parameter on segmentation performance. The optimal results are in bold.

Images	Metrics	0.001	0.005	0.01	0.05	0.1
Synthetic image	SA (%)	99.86	99.85	99.87	**99.89**	99.88
	MIoU (%)	99.71	99.70	99.74	**99.79**	99.76
Natural image 1	SA (%)	99.07	99.05	**99.18**	99.17	98.84
	MIoU (%)	96.99	96.67	**97.11**	97.08	96.81
Natural image 2	SA (%)	98.57	**98.77**	98.38	98.68	98.33
	MIoU (%)	93.65	**94.03**	93.22	93.82	93.17
Remote sensing image 1	SA (%)	99.06	99.05	**99.11**	99.11	99.11
	MIoU (%)	97.39	97.35	**97.53**	97.51	97.51
Remote sensing image 2	SA (%)	98.91	98.92	**98.99**	98.93	98.97
	MIoU (%)	97.71	97.72	**97.88**	97.74	97.85
Remote sensing image 3	SA (%)	98.08	**98.15**	97.73	97.60	97.82
	MIoU (%)	96.15	**96.29**	95.47	95.22	95.66

Meng Zhang was born in Heze, China, in 1995. He received the bachelor's degree in Automation from Harbin University of Science and Technology, Harbin, in 2018, and the master's degree in Aeronautical and Astronautical Science and Technology from Xi'an Jiaotong University, Xi'an, in 2020. He is currently pursuing the Ph.D. degree in Aeronautical and Astronautical Science and Technology at Xi'an Jiaotong University. His research interests include image processing, information fusion, and pattern recognition. (Email: zhangmeng726@stu.xjtu.edu.cn)



Yi Yang was born in Xi'an, China, in 1980. She received the bachelor's degree in automation from the Xi'an University of Technology, Xi'an, China, in 2002, and the master's and Ph.D. degrees in control science and engineering from Xi'an Jiaotong University, Xi'an, in 2005 and 2010, respectively. She is currently a Lecturer with the School of Aerospace, Xi'an Jiaotong University. Her research interests include evidence theory, pattern classification, and image processing. (Email: jiafeiyy@mail.xjtu.edu.cn)

Sixian Zhang was born in Jingdezhen, China, in 1999. He received the bachelor's degree in Aircraft Design and Engineering from Xi'an Jiaotong University, Xi'an, in 2019, and the master's degree in Aeronautical and Astronautical Science and Technology from Xi'an Jiaotong University, Xi'an, in 2021. He is currently pursuing the Ph.D. degree in Aeronautical and Astronautical Science and Technology at Xi'an Jiaotong University. His research interests include image processing, information fusion, and object tracking. (Email: touchmeteor@stu.xjtu.edu.cn)

Pengbo Mi was born in Xingtai, China, in 1998. He received the bachelor's degree in Mechanical Design, Manufacturing and Automation from North China University of Technology, Beijing, in 2020, and the master's degree in Aeronautical and Astronautical Science and Technology from Xi'an Jiaotong University, Xi'an, in 2022. He is currently pursuing the Ph.D. degree in Aeronautical and Astronautical Science and Technology at Xi'an Jiaotong University. His research interests include multi-source image classification and fusion. (Email: PengboMi@stu.xjtu.edu.cn)

Deqiang Han was born in Xi'an, China, in 1980. He received the bachelor's degree in communication and control engineering and the master's and Ph.D. degrees in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively. He is currently a Professor with Xi'an Jiaotong University. His research interests include evidence theory, information fusion, and pattern classification. Dr. Han is an International Society of Information Fusion Member and a Technical Program Committee Member for the 1st–12th Chinese Conference on Information Fusion in 2009–2023. (Email: deqhan@mail.xjtu.edu.cn)