

An Active Contour Model Based on Fuzzy Superpixel Centers and Nonlinear Diffusion Filter for Instance Segmentation

Yiyang Chen¹, Member, IEEE, Fuzheng Zhang², Guina Wang², Guirong Weng²,
and Daniele Fontanelli³, Fellow, IEEE

Abstract—The significant weaknesses of the active contour model (ACM) are the manual setting of contour and the inability to process images with complex information, which limits its efficiency and application scope. In this article, an ACM, called FSC&NDF, is combined with fuzzy superpixel centers (FSCs) and nonlinear diffusion filter (NDF) to solve the above two problems simultaneously. YOLOv9 is adopted to locate the superpixels of interest; the joint boundaries of these superpixels are set as the initial contour, which is close to the morphological features of the target. Improved fuzzy superpixel clustering is applied to extract image features and yield superpixel centers, and the clusters are integrated into the main body of the energy function, NDF module further enhances boundary positioning and suppresses noise. In addition, the proposed connection mechanism makes it possible to convert object detection to instance segmentation. Experimental results show that FSC&NDF overcomes the limitations of previous ACMs in all aspects and its FPS, AP, AP₅₀, and AP_M are higher than mainstream deep learning algorithms. The platform experiment based on the telecentric lens further proves the practicality of FSC&NDF.

Index Terms—Active contour, automatic contour generation, fuzzy superpixel centers (FSCs), image segmentation, nonlinear diffusion filter (NDF).

I. INTRODUCTION

A VARIETY of image segmentation algorithms are designed to solve different image problems, including particle swarm optimization, wavelet analysis, genetic algorithm, watershed transformation (WT), active contour evolution, and deep learning model. These algorithms improve the efficiency of segmentation to a certain extent and are accepted in a wide range of industrial applications [1], [2], [3]. Active contour model (ACM) is a comprehensive image segmentation method, which is frequently used in natural image segmentation [4], [5], [6], medical image

processing [7], [8], [9], remote sensing [10], [11], [12]. The geometric active contour (GAC) model [13] is a classic ACM which is also known as the level set method. The contour is regarded as the intersection of a continuous 3-D surface and the 2-D image, and the core energy function is constructed based on this surface, which is an implicit expression of the 2-D contour. According to the Euler equation and calculus of variations [14], the derivative of the energy function with respect to the surface equal to the inverse of the derivative of the surface with respect to time, which is a brachistochrone problem. Therefore, the evolution of the 2-D contour is transformed into the iteration of the 3-D surface, which is equivalent to solve low-dimensional problems with high-dimensional functions. The drawback is that the initial state of the contour is decisive. While the initial contour is set manually, inappropriate ones can lead to undersegmentation.

The bias field theory and clustering method are applied to reform the original image in [15] and [16], which extracts the grayscale information of each position. Since the clustering method is sensitive to the original image, the results may be severely distorted by high-level noise interference. Another mainstream method is to convert image fitting into distribution approximation [17], [18]. Classic divergence such as Jeffreys divergence and K-L divergence are suitable to describe the grayscale distribution of the image, providing a reasonable channel for quantification of image texture and regional similarity. When the number of targets in the image increases significantly, it is difficult to accurately describe the distribution of each target. Fuzzy *c*-means (FCMs) clustering is combined with different modules of ACM to yield promising results [19], [20]. By clustering the preprocessed image, key information are obtained and integrated into the energy function. However, each iteration requires repeated clustering operations, which increases the time cost and computational overhead.

Generating a suitable initial contour for ACM depends on the feature information of the image. Superpixel segmentation is a commonly used preprocessing method in image processing tasks, which uses the feature information of the image, such as texture, color, and brightness, to group the pixels of the original image, and each of these groups is called a superpixel. Based on this, the basic information of the image is classified and simplified, providing a guarantee for subsequent accurate

Received 11 October 2024; revised 18 January 2025; accepted 16 May 2025. Date of publication 2 June 2025; date of current version 10 June 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62103293 and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20210709. The Associate Editor coordinating the review process was Dr. Pritam Paral. (Corresponding author: Guina Wang.)

Yiyang Chen, Fuzheng Zhang, Guina Wang, and Guirong Weng are with the School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215137, China (e-mail: yychen90@suda.edu.cn; fzzhang1128@stu.suda.edu.cn; wangguina@suda.edu.cn; wgr@suda.edu.cn).

Daniele Fontanelli is with the Department of Industrial Engineering, University of Trento, 38123 Trento, Italy (e-mail: daniele.fontanelli@unitn.it).

Digital Object Identifier 10.1109/TIM.2025.3573369

segmentation. Simple linear iterative clustering (SLIC) [21] is one of the most well-known algorithms, which implements clustering by calculating the color distance and spatial distance between pixels. Its classification results are presented as regular matrices, and for irregular targets, the generalization results are not ideal. Kernel function similarity measurement is added in LSC [22] to optimize the comprehensive distance between pixels and achieve high memory efficiency. Adaptive morphological reconstruction (AMR) is applied in [23] to screen seed regions from multiple scales and obtain all boundaries that may contain target information.

Superpixels target the entire image domain, to extract superpixel subblocks in key areas, it is necessary to locate these areas in advance. Object detection algorithms are tailored for this purpose, which are based on deep learning and have gained widespread attention due to their excellent target positioning capabilities. They can be divided into two categories, one is to set numerous candidate boxes on the image and filter them to obtain the final detection box [24], [25], [26], [27], and the other is to predict the center point of the target object, and obtain the four vertices of the detection box through the regression of the center point [28], [29]. Generally, the detection box holds the complete target area, so a novel point is that the rough contour of the target object can be acquired by extracting the superpixels within the box. With the support of ACM, the rough contour evolves toward the ground truth, so that instance segmentation is realized. This is a win-win design: the object detection algorithm is converted into instance segmentation without adding complex segmentation modules, and the initial contour of ACM is set automatically.

More in depth, we propose an ACM combined with fuzzy superpixel centers (FSCs) and nonlinear diffusion filter (NDF) to solve the manual contour initialization problem of ACM, transforming the object detection algorithm into instance segmentation. First, the object detection algorithm locates the area of interest, while the original image is divided into superpixels. Then, the superpixels of interest are extracted, and their joint boundary is obtained as the initial contour of ACM. Furthermore, the FCM clustering based on regularization entropy on superpixels is applied, while the clustered regional information is integrated into the energy function of ACM to achieve an optimized iterative process. In addition, a module based on a nonlinear diffusion function is designed to enhance the boundary information of the target object and suppress background interference. The proposed model, called FSC&NDF, achieves high-precision segmentation compared with traditional and deep learning algorithms. The main innovations are as follows.

- 1) The object detection algorithm and superpixels work together to extract the rough contour of key objects, avoiding manual contour initialization of ACM.
- 2) A new energy description method is proposed, which incorporates the information of superpixels extracted by FSCs and optimizes the iteration direction.
- 3) NDF module is designed to enhance boundary gradient and reduce background noise, further improving the efficiency of contour evolution.

II. MOTIVATIONS

Previous ACMs rely on manually drawn initial contours to achieve a stable segmentation process, which limits the application scenarios of the model. These models are short of feature information extraction, resulting in the inability to handle complex images with multiple targets and scenes. In addition, their antinoise and boundary positioning capabilities need to be further improved.

A. Necessity of Automatic Contour Generation

Although the segmentation effect of ACM is commendable, it is deeply affected by the state of the initial contour. Three models, RSF [15], GADF [30], and OLPFI [31], are adopted to present the problem. As shown in Fig. 1, the initial contours are all set manually. When the contours present the state of Case 1, the three models yield ideal segmentation effects on single-target images, while they are not up to par with multitarget images. When the contours change to the state of Case 2, all segmentation results become worse, with over-segmentation and under-segmentation appearing. Taking RSF model as an example, the structure of its energy equation is similar to that of most ACMs, which is

$$E = \sum_{i=1}^2 c_i \int_{\Omega_i} \left(\int K_{\lambda}(I(p) - f_i(q))^2 M_i(p) dp \right) dq. \quad (1)$$

Notice that i takes values between 1 and 2, which means that the model treats the entire image domain as Ω_1 inside the contour and Ω_2 outside the contour, $I(p) - f_1(q)$ and $I(p) - f_2(q)$ represent the energy of these two regions, respectively. This is equivalent to ignoring all other contours when one contour evolves. Therefore, when processing multitarget images, it is necessary to manually draw a contour for each target, and the same mechanism applies to GADF and OLPFI. Technically speaking, the location and shape of the initial contour are not specified by the model, meaning that the initial contour can be set arbitrarily in the image domain.

Let the target area set be Ω_T , and the intersection of the closed contour and the target area be Ω_{IS} ; then, there are three cases: $\Omega_{IS} = \emptyset$, $\Omega_{IS} \subset \Omega_T$, and $\Omega_{IS} = \Omega_T$. When $\Omega_{IS} = \emptyset$, $K_{\lambda}(I(p) - f_1(q))^2$ has no relationship with the grayscale information of the ground-truth target, while the probability distribution it obeys may be similar to the ground-truth target, which may lead to model misjudgment such as falling into the local optimum and failing to overcome the noise. When $\Omega_{IS} \subset \Omega_T$, $K_{\lambda}(I(p) - f_1(q))^2$ is partially related to the ground-truth information, hence the contained edge gradient makes the transformation of the contour close to the real boundary. However, the existence of uncertainties may interfere with the evolution direction of the contour, inappropriate learning rate may force the contour to cross the real boundary and enter inside the target, causing missegmentation. Finally, when $\Omega_{IS} = \Omega_T$, the initial contour, and the ground-truth boundary completely overlap; therefore, $K_{\lambda}(I(p) - f_1(q))^2$ and the real target area follow the same distribution. This fact reveals an important shortcoming of the existing model: the setting of the prior contour is not only labor-intensive but also meaningless

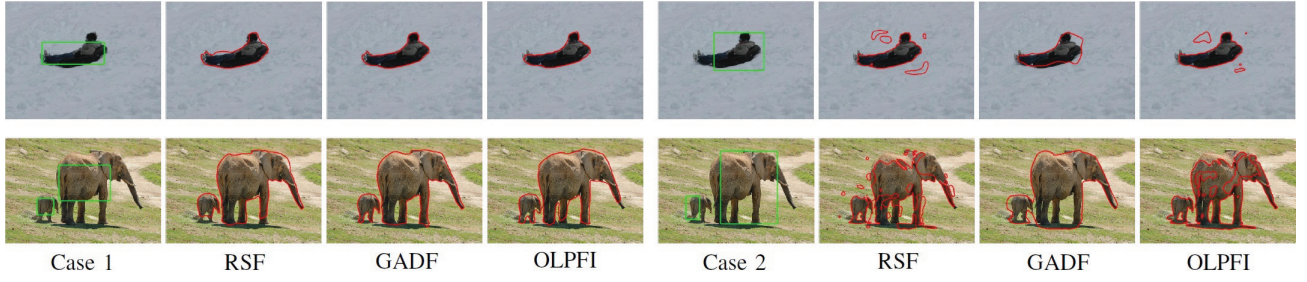


Fig. 1. Segmentation problems caused by manual calibration of the initial contour. The green line is the initial contour, and the red line is the final contour. Row 1: results of single target image under different initial contours. Row 2: results of multitarget images under different initial contours.

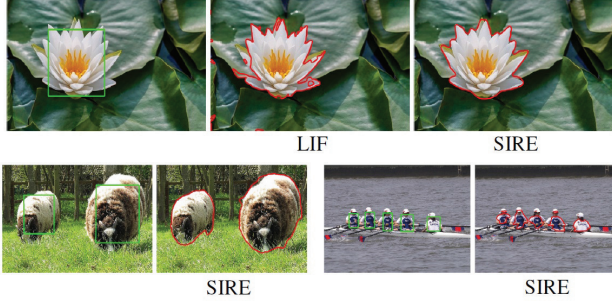


Fig. 2. Boundary positioning results of LIF and SIRE. The green line is the initial contour, and the red line is the final contour. Row 1: polygon target. Row 2: multiple targets and complex scenes.

when the contour is set extremely close to the ground-truth boundary.

B. Boundary Positioning and Denoising

A common phenomenon is that ACM works well when dealing with simple targets, once the target or scene becomes intricate, i.e., the targets are polygonal or various types of objects appear in the scene, the segmentation effect becomes poor. The key to solving this problem is to improve boundary positioning capability. SIRE [32] introduces the Softsign function to enhance the boundary information; the corresponding iteration function is as follows:

$$\begin{cases} \frac{\partial \phi}{\partial t} = -\nu (\partial S_1(\phi) / \partial \phi) \text{Softsign}((d_1(p) - d_2(p)) / U) \\ S_1(x) = 1 / (1 + e^{-x}) \\ \text{Softsign}(x) = x / (1 + |x|) \end{cases} \quad (2)$$

where ν is a constant, S_1 is the Sigmoid function, ϕ is the level set function, and d_1 and d_2 represent the energy description of the region inside and outside the contour, respectively. U is the regularization term, that is,

$$U = \sqrt{\frac{1}{P} \sum_{p \in \Omega} (I(p) - \bar{I})^2}. \quad (3)$$

P is the total number of pixels in the image and \bar{I} is the average gray value of the image. Since the gradient near the boundary is large, the gradient is further amplified by (2), and the smaller grayscale approaches 0, thereby enhancing the boundary gradient. In Fig. 2, LIF [33] is chosen to compare

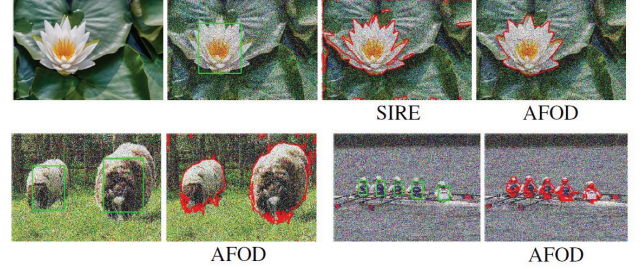


Fig. 3. Noise segmentation results of SIRE and AFOD. The green line is the initial contour, and the red line is the final contour. Gaussian noise with a variance of 0.01 is applied.

with SIRE, it is difficult for LIF to segment images with complex shapes, while SIRE suffers when multiple types of objects appear in the scene.

AFOD [34] proposes an ACM based on adaptive fractional order differentiation, which quantifies the intensity frequency of all image information. Its level set iteration function is

$$\frac{\partial \phi}{\partial t} = -H'(\phi) \left(b_1 \int_{\Omega} K_{\lambda} (I - (f_1 + f_2) / 2) dp + b_2 \int_{\Omega} K_{\lambda} (D^{\theta} I - (d_1 + d_2) / 2) dp \right) \quad (4)$$

where H is the Heaviside function, b_1 and b_2 are adjustable parameters, Ω refers to the whole image domain, K_{λ} is the kernel function, f_1 and f_2 are fitting intensities of I inside and outside the contour, d_1 and d_2 are fitting intensities of $D^{\theta} I$ inside and outside the contour, D^{θ} means fractional differential operation, and θ is defined as follows:

$$\theta(x) = \frac{|\nabla I(x)|}{\max_{x \in \Omega} \{|\nabla I(x)|\}} \quad (5)$$

where ∇ is the gradient operator. In the case of noise interference, each domain is given a specific differential process, which improves the positioning effect of the target object. Fig. 3 shows the comparison between SIRE and AFOD in dealing with noisy images. For single-target images, noise is overcome by AFOD, when faced with complex images, the segmentation effect is still poor. Due to the discussed limitations, YOLOv9 combined with AMR-WT is adopted in this work to solve the initial contour problem, and instance segmentation based on YOLOv9 is achieved.

III. METHODOLOGY

This section elaborates on the theoretical basis of FSC&NDF framework. The feature information of superpixels, such as texture, regional color, and shape, is merged by FSCs, boosting the target positioning effect. In addition, the boundary gradient is enhanced and the noise is suppressed by the proposed NDF module, leading to faster segmentation speed and greater boundary processing capability.

A. Construction of NDF Module

Diffusion is a common physical process that balances the concentration of a substance without changing its mass. For images, pixel grayscale can be regarded as concentration, and the diffusion of grayscale implies gradient change. By using gradient information to drive the local grayscale to diffuse toward the target boundary, while keeping the overall grayscale stable, diffusion filtering is designed. Let $\Omega \subset \mathbb{R}^2$ be the grayscale image, and each pixel in the image is denoted by p , i.e., $p \in \Omega$. Inspired by Fick's law, a relational expression between the potential function Ψ , flux Φ , and diffusivity g are given as follows:

$$\nabla(\Psi(|\nabla u|)) = \Phi(\nabla u) = g(|\nabla u|^2) |\nabla u| \quad (6)$$

where $u(p)$ is the grayscale image, $\nabla u = [\partial u_x, \partial u_y]$ is the gradient of each point in the image, and $|\nabla u| = (\partial u_x^2 + \partial u_y^2)^{1/2}$ is the modulus of the gradient. The rightmost expression of (6) is a variant of Fick's law, and g determines the strength of diffusion. To protect the original information of the image, the value of Φ must remain stable, as described in the following proposition.

Proposition 1: The relationship between g and ∇u should meet the following requirements.

- 1) $g(|\nabla u|^2)$ is a decreasing function with $|\nabla u|$ as the independent variable, satisfying $0 < g(|\nabla u|^2) < 1$.
- 2) $\lim_{|\nabla u| \rightarrow \infty} g(|\nabla u|^2) = 0$.
- 3) $\lim_{|\nabla u| \rightarrow 0} g(|\nabla u|^2) = 1$.

Proof: The detailed proof is described in [35].

In this sense, when the gradient is rather small, the diffusion rate approaches the maximum value of 1, and grayscale speeds up its movement and crosses the nonboundary area. When the gradient is rather large, the diffusion rate decreases, indicating that it approaches the boundary. Thus, the value of $g(|\nabla u|)|\nabla u|$ is kept at a reasonable level, retaining uniform grayscale. Based on the above requirements and references in [35], this article proposes the following diffusion coefficient function:

$$g(s) = \frac{2}{1 + e^s} \quad (7)$$

that, in light of (6), yields the following flux function:

$$\Phi(s) = \frac{2s}{1 + e^{s^2}} \quad (8)$$

and the potential function

$$\Psi(s) = s^2 - \ln\left(\left(1 + e^{s^2}\right)/2\right) \quad (9)$$

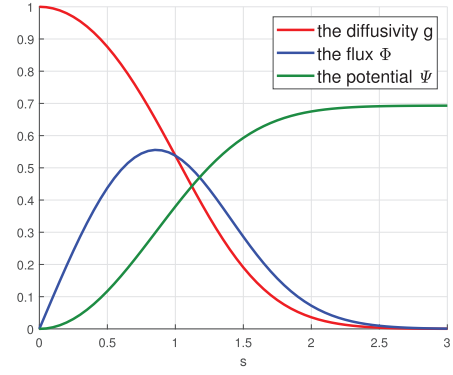


Fig. 4. Schematic of the three basic functions.

the function graphs are shown in Fig. 4. The role of Ψ is to constrain the image gradient which fluctuates in a large range to a small interval so that it is convenient to fix parameters to control the change of image information. For instance, when $|\nabla u|$ is too large, $\Psi(|\nabla u|) < 0.7$ to prevent fluctuations in the iterative process, when $|\nabla u|$ is small, $\Psi(|\nabla u|) \rightarrow 0^+$ to eliminate minor disturbances.

In simple terms, (9) is derived by s to obtain (8), to ensure that (9) is close to 0 when $s \rightarrow 0$, $\ln(2)$ is added at the end of the equation to control its sign change. As can be seen from Fig. 4, the diffusion coefficient equation $g(s)$ satisfies the three conditions mentioned above, and the flux function $\Phi(s)$ does reflect the change rate of the potential function $\Psi(s)$ with s .

Next, the potential function Ψ and the image gradient $|\nabla u|$ are combined to form a local energy function as follows:

$$E^{\text{PG}}(x) \approx \int_{\Omega} \Psi(|\nabla u(x)|) dx. \quad (10)$$

E^{PG} contains the normalized gradient information of the entire image domain, whose optimization process is equivalent to find an optimal grayscale distribution that minimizes the sum of the normalized gradients. In fact, minimizing the normalized gradients increases the contrast between large gradients and small gradients, thereby enhancing the foreground and suppressing the background. In the variational method, the standard way to solve (10) is to find a steady-state solution to the gradient flow equation, i.e.,

$$-\frac{\partial E^{\text{PG}}}{\partial u} = \frac{\partial u}{\partial t} = \text{div}(g(|\nabla u|^2) \nabla u) \quad (11)$$

where $g(|\nabla u|^2)$ is the diffusion coefficient function as in (11). Since the image gradient reflects the image characteristics, $g(|\nabla u|^2)$ adjusts the diffusion coefficient through $|\nabla u|^2$ to reach the excellent performance of edge protection, which reduces the noise and maintains the image edge information. If (11) is employed directly to iterate the initial contour, the edge information, and noise may be unified into a flat area, making it unsuitable for contour to distinguish boundaries and noise. Therefore, a threshold should be set in the image iteration process to fine-tune the gap between the target edge and noise. Thus, (11) is modified as follows:

$$\frac{\partial u^{\text{PG}}}{\partial t} = \text{div}(g(|\nabla u|^2 / \eta^2) \nabla u) \quad (12)$$

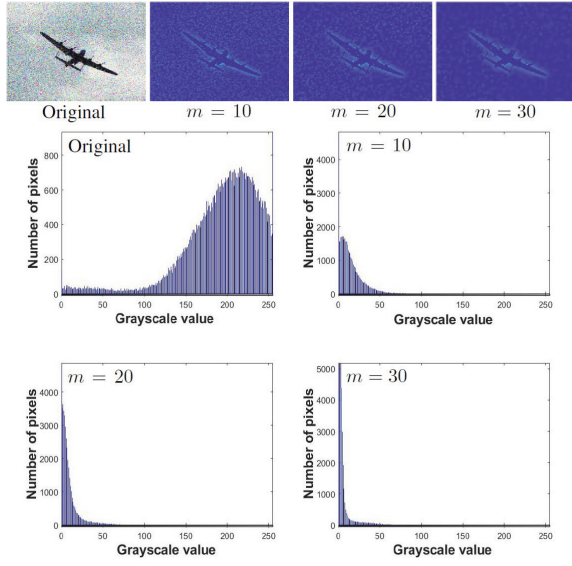


Fig. 5. Effect of NDF module. Row 1: original image with Gaussian noise, and the processing effect under $m = 10$, $m = 20$, and $m = 30$. Rows 2 and 3: color histograms in different states.

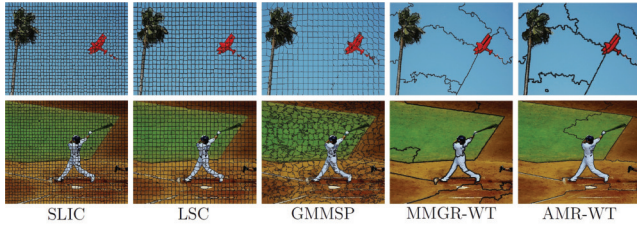


Fig. 6. Effects of five superpixel segmentation models.

where $\eta = \sqrt{\text{std}2(I(x))}$ is the edge threshold and $\text{std}2$ represents the calculation of standard deviation. Its role is to reduce the energy fluctuation range while maintaining the significance of the edge and flat area, which better ensures the distinction between boundaries and noise.

Letting $g(|\nabla u|^2/\eta^2)$ be g^* , and the solution of (12) is approximated by the discrete equation

$$u^{m+1} = u^m + \Delta t_1 (\text{div}(g^* \cdot |\nabla u^m|)) \quad (13)$$

where m is the current iteration number, Δt_1 is the iteration step size, which is a constant, and g^* is updated in time with u^m . Depending on m , the degree of grayscale diffusion is different, the grayscale distribution under different m is shown in Fig. 5. As m increases, the number of pixels with larger grayscale values remains stable, while those with grayscale values close to 0 increase significantly.

B. Contour Generation Based on YOLOv9 and AMR-WT

Morphological reconstruction (MR) is frequently used in WT algorithms such as GMMSP [36] and MMGR-WT [37], while adaptive MR (AMR) further improves the segmentation effectiveness and yields AMR-WT [23]. Traditional superpixel segmentation requires prespecifying the number of superpixels, which may lead to the individual characteristics loss of each target. However, AMR is able to automatically filter out

meaningless regions in the original superpixels, so that the shape and size of each superpixel is different, and it is closer to the appearance of the real subregion in the image, as shown in Fig. 6. The area of interest can be fused from these superpixels, so the key task is to reasonably select the target superpixels.

This article adopts YOLOv9 [25] to locate the area of interest, calculates the intersection between the detection box and each superpixel, and retains the complete superpixels within the box. Then, the nonpublic boundaries of these superpixels are connected to form the initial contour. More in depth, denote the output of AMR-WT as S , having its size equal to the original image. Let the i th superpixel region be S_i , the mask of the original image be M_I , and the region within the deformed detection box be S_b , the regional correlation equation is defined as follows:

$$M_I = 0, \quad \text{if and only if } (S_i \cap S_b = S_i) \wedge (S_i \cup S_b = S_b) \\ M_I = 1 \quad (14)$$

the regional boundary with a value of 1 in M_I is the initial contour of a single target. This depicted process determines the initial value of the level set. The generation of the initial contour is illustrated in Fig. 7.

C. Energy Function Combined With FSCs

The initial contour C is specified in the image domain Ω , which divides the image into two nonoverlapping regions, with the initial energy function defined as follows:

$$E^I(C) = \lambda_1 \int_{\text{outside}(C)} (I(x) - I_1^f(x)) dx \\ + \lambda_2 \int_{\text{inside}(C)} (I(x) - I_2^f(x)) dx \quad (15)$$

in which $(I(x) - I_1^f(x))$ and $(I(x) - I_2^f(x))$ describe the image information outside and inside C , while I_1^f and I_2^f are the fitting functions of the image to be defined.

Usually, the contour C is implicit in the level set function ϕ . To embed the presence of ϕ , the Heaviside function $H_\varepsilon(\phi)$ [15] is introduced into (15) to have

$$\begin{cases} E^I(\phi) = \lambda_1 \int_{\Omega} (I(x) - I_1^f(x)) H_\varepsilon(\phi) dx \\ \quad + \lambda_2 \int_{\Omega} (I(x) - I_2^f(x)) (1 - H_\varepsilon(\phi)) dx \\ H_\varepsilon(\phi) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan \left(\frac{\phi}{\varepsilon} \right) \right) \\ H_\varepsilon'(\phi) = \delta_\varepsilon(\phi) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + \phi^2}. \end{cases} \quad (16)$$

Similar to the approach followed for (16), the same method to find the minimum value of (16) is computed with its partial differential equation, i.e.,

$$-\frac{\partial E^I}{\partial \phi} = \frac{\partial \phi}{\partial t} = -\lambda_1 (I(x) - I_1^f(x)) \delta_\varepsilon(\phi) \\ + \lambda_2 (I(x) - I_2^f(x)) \delta_\varepsilon(\phi). \quad (17)$$

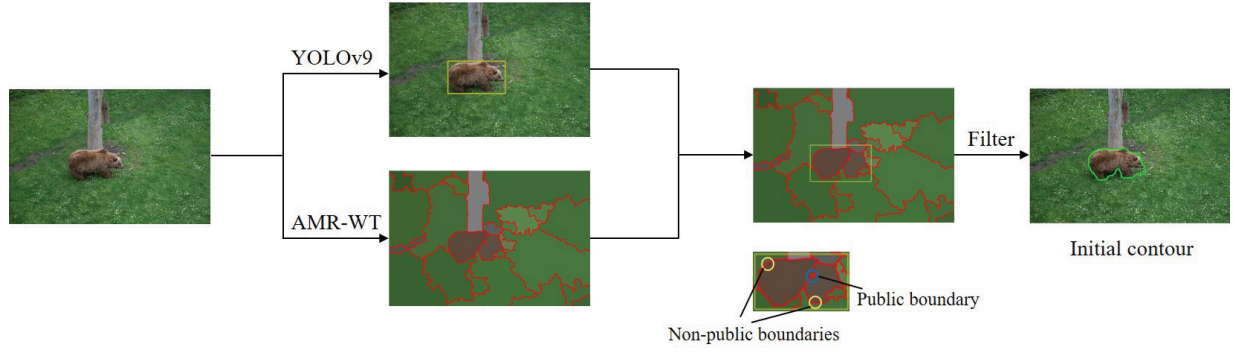


Fig. 7. Contour generation process.

To simplify the model computation and the parameter control, let $\lambda_1 = \lambda_2 = \lambda$, i.e., the parameter λ is a single signed constant. Then, (17) is rewritten as follows:

$$\frac{\partial \phi}{\partial t} = \lambda \left(I_1^f(x) - I_2^f(x) \right) \delta_\varepsilon(\phi). \quad (18)$$

In (15), $I(x) - I_1^f(x)$ and $I(x) - I_2^f(x)$ represent the high frequency components outside and inside the contour C , respectively. While superpixels contain global information of the entire image domain, the classification of these superpixels is considered to obtain approximate features inside and outside the contour. Wu et al. [37], this article performs an optimized FCM clustering on all superpixels, with the number of clusters being 2, representing the two regions inside and outside the contour, respectively. Its objective function is

$$J = \sum_{r=1}^{N^S} \sum_{k=1}^2 S_r b_{kr} \|X_r - c_k\|^2 + \sum_{r=1}^{N^S} \sum_{k=1}^2 S_r b_{kr} \cdot \ln \left(\frac{b_{kr}}{1 - \xi_k} \right) \quad (19)$$

where N^S is the number of superpixels, S_r represents each superpixel region, b_{kr} is the membership of the r th superpixel region with respect to the k th cluster center, which satisfies $\sum_{k=1}^2 b_{kr} = 1$, c_k refers to the k th cluster center, and, finally, $1 - \xi_k$ is the regularization coefficient, which refers to the prior probability of X_r belonging to c_k , with the constraint $\sum_{k=1}^2 (1 - \xi_k) = 1$. The second term of (19) is the regularized entropy, which computes the correlation between each superpixel and the other superpixels. The closer the predicted b_{kr} is to the prior probability, the higher the correlation between the two superpixels is; hence, a higher weight is associated during the clustering. In this way, the similarity of different regions is taken into account during clustering, making the classification results more effective. X_r is the representative element of each superpixel, which is defined as follows:

$$X_r = \underbrace{\arg \min}_{x_{ri}} \left\{ \sum_{j=q, q \neq i}^{N'} |x_{ri} - x_{rj}| \right\}, \quad i = 1, 2, \dots, N' \quad (20)$$

having x_{ri} as the current pixel in the r th superpixel region, x_{rj} denoting all the pixels in the region except the current pixel and N' is the number of pixels in each region. Hence, X_r is the pixel in the r th superpixel region that has the smallest $L1$ norm among all the pixels. Since regions with lots of feature information are clustered, variations in membership during the

iterative process are unstable. Thus, the second term in (19) aims to control the change amplitude of the membership b_{kr} , which is called regularization entropy.

Proposition 2: According to the Lagrange multiplier method, the update equations of c_k , b_{kr} , and ξ_k are

$$c_k = \frac{\sum_{r=1}^{N^S} S_r b_{kr} X_r}{\sum_{r=1}^{N^S} S_r b_{kr}} \quad (21)$$

$$b_{kr} = \frac{(1 - \xi_k) \exp(\xi_k - \|X_r - c_k\|^2)}{\sum_{k=1}^2 (1 - \xi_k) \exp(\xi_k - \|X_r - c_k\|^2)} \quad (22)$$

$$\xi_k = 1 - \frac{\sum_{r=1}^{N^S} S_r b_{kr}}{\sum_{r=1}^{N^S} S_r}. \quad (23)$$

Proof: See the Appendix.

The clustered superpixel regions obtained by (19) are recorded as S_{c1} and S_{c2} . The intersection ratio of the two regions with the area inside the contour is then computed, which is denoted by IOU_{c1} and IOU_{c2} , respectively. If $\text{IOU}_{c1} > \text{IOU}_{c2}$, S_{c1} is embedded in I_2^f ; otherwise, S_{c1} is embedded in I_1^f , and the other area is assigned to the remaining fitting function. The clustered regions embedded in I_1^f and I_2^f are recorded as I_{S1} and I_{S2} , respectively. Therefore, the fitting functions are constructed as follows:

$$\begin{cases} I_1^f = I + I_{S1} - I_{\sigma_1} \\ I_2^f = I + I_{S2} - I_{\sigma_2} \\ I_{\sigma_i} = I * G_{\sigma_i}, \quad i = 1, 2 \end{cases} \quad (24)$$

where I_{S1} is the clustered superpixel region that belongs outside the contour, I_{S2} is that belongs inside the contour, and G_{σ_i} is the Gaussian kernel with size $\omega_i \times \omega_i$. Then, (18) is rewritten as follows:

$$\frac{\partial \phi}{\partial t} = \lambda (I_{S1} - I_{S2} - I * (G_{\sigma_1} - G_{\sigma_2})) \delta_\varepsilon(\phi) \quad (25)$$

which is the original state of the global energy function.

D. Evolution of the Final Iterative Equation

A positive peak and a negative valley appear on both sides of a zero point when the target boundary is second-order differentiated, which detects the sudden change in the gray level of the image. Therefore, the model has a strong ability to locate target edges, the downside of which is that it is more

sensitive to noise. Due to the edge enhancement characteristic of the NDF module, a combination with the energy function is foreseen. Let $e(x, t = 0) = I_{S_1} - I_{S_2} - I * (G_{\sigma_1} - G_{\sigma_2})$ replace u with e in (13) for iteration. Record e after n iterations as $e(x, t = n)$, and then, by substituting it into (25), the final iterative equation

$$\frac{\partial \phi}{\partial t} = \lambda \cdot e(x, t = n) \cdot \delta_\varepsilon(\phi) \quad (26)$$

is obtained. Under the action of the diffusion coefficient function, the target boundary information is enhanced, while the information are strengthened due to the presence of η .

In (26), $e(x, t = n)$ is the data driven term, and its size is highly dependent on the gray value of the image. To improve the parameter robustness, an energy uniform function is defined as follows:

$$\begin{cases} \rho(x) = \frac{2}{\pi} \text{sign}(x) \cdot \arctan(x^2) \\ \text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \end{cases} \quad (27)$$

and (26) is further rewritten as follows:

$$\frac{\partial \phi^{\text{FSC\&NDF}}}{\partial t} = \lambda \rho(e(x, t = n) / \tau) \cdot \delta_\varepsilon(\phi) \quad (28)$$

its corresponding discretization equation is

$$\phi^{n+1} = \phi^n + \Delta t (\lambda \rho(e(x, t = n) / \tau) \cdot \delta_\varepsilon(\phi)) \quad (29)$$

where $\tau = 2\eta$. Let the initial level set function be

$$\phi_0 = \begin{cases} -c_0, & x \in \Omega_c - \partial\Omega_c \\ 0, & x \in \partial\Omega_c \\ c_0, & x \in \Omega - \Omega_c \end{cases} \quad (30)$$

where c_0 is a constant, Ω is the image domain, $\partial\Omega_c$ is the target boundary, and Ω_c is the target domain without its boundary. Obviously, the value range of e is constrained between 0 and 1, which means that the value of λ can be fine-tuned within a wider range. The fault tolerance rate of λ is improved and a satisfactory iteration effect is achieved.

Since the level set function becomes bumpier as the number of iterations increases, this article defines a new regularization term as follows:

$$\phi_R = \rho(k_s \cdot \phi^n), \quad k_s = 11 \quad (31)$$

where ρ is the energy uniform function in (27) and ϕ^n represents the level set function after the iteration process is completed. In addition, k_s is the slope improvement factor, which is set to 11 empirically for stability. The level set function retains the properties of the signed distance function while being differentiable.

For the energy minimization, the contour C derives some redundant curves under the background interference, which affects the final segmentation result. To account for this detrimental effect, we construct a curve trimming function as follows:

$$\phi_T(x_c) = \frac{1}{N''} \sum_{i=1}^{N''} \phi_R(x_i) \quad (32)$$

Algorithm 1 Implementation of FSC&NDF

Input: original image I , c_0 , ϕ_0 , m , Δt_1 , ε , λ , ω_1 , ω_2 , σ_1 , σ_2 , Δt , N , γ .

Output: ϕ_T and target contours.

- 1: **Default settings:** $c_0 = \varepsilon = \Delta t_1 = \Delta t = 1$, $m = 30$, $\lambda = 10$, $\omega_1 = \omega_2 = 15$, $\sigma_1 = 0.5$, $\sigma_2 = 3.5$, $N = 50$, and $\gamma = 7$.
- 2: Input the image into YOLOv9 and AMR-WT to obtain detection boxes and superpixels.
- 3: Initialize ϕ_0 by (30).
- 4: Compute I_{S_1} and I_{S_2} by ((21) - (23)).
- 5: Use I_{S_1} , I_{S_2} , $I(x)$, G_{σ_1} , and G_{σ_2} to compute $e(x, t = 0)$, execute (13) to calculate $e(x, t = n)$.
- 6: Compute $\rho(e(x, t = n) / \tau)$ by (27).
- 7: **for** i **from** 1 **to** N **do**
- 8: Calculate $\delta_\varepsilon(\phi)$.
- 9: Update ϕ^i using (29).
- 10: **if** $|\phi^{i+1} - \phi^i| \leq 0.001$ Break iteration; **end**
- 11: Regularize ϕ^{i+1} and obtain ϕ_R by (31).
- 12: Trimming ϕ_R and obtain ϕ_T by (32).
- 13: **end**
- 14: **return** ϕ_T and the zero level set.

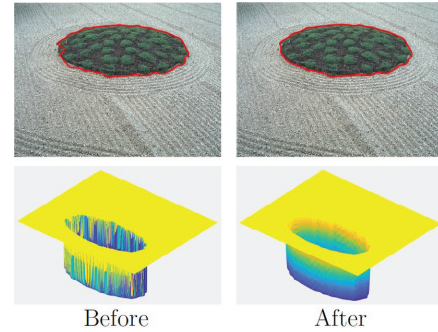


Fig. 8. Effects of curve trimming function. Row 1: contours before and after trimming. Row 2: ϕ before and after trimming.

a truncated window of size $\gamma \times \gamma$ is utilized to traverse each point in the image, and $N'' = \gamma \times \gamma$. Notice that (32) plays the role of the convolution kernel, which performs mean filtering on the entire image domain. The grayscale value of each pixel is obtained by averaging the grayscale values of the surrounding pixels. In this way, the positions where the grayscale value is more prominent, such as in the presence of noise and blur, can be filtered out. For predicted target boundaries, curve trimming reduces the fluctuation of ϕ , so that the parts on both sides of the ground truth boundary move closer to it at the same time. As can be noted from Fig. 8, the predicted contour is closer to the ground-truth boundary after trimming. In addition, ϕ becomes smoother, which is similar to the initial binary state, providing a prerequisite for the next iteration. Algorithm 1 depicts the overall described solution.

IV. MODEL EXPERIMENTS

In this section, the segmentation ability of FSC&NDF is verified based on the COCO benchmark. Comparative experi-

¹Registered trademark.

²Trademarked.

TABLE I

METRICS VALUES AND SEGMENTATION TIME T (s) FOR NATURAL IMAGES. A~ REFERS TO THE AVERAGE VALUE OF EACH METRIC

Images	LKLD				APFJD				GADF				SIRE				FSC&NDF			
	DSC	JSC	CSC	T	DSC	JSC	CSC	T	DSC	JSC	CSC	T	DSC	JSC	CSC	T	DSC	JSC	CSC	T
1	0.846	0.733	0.635	0.38	0.881	0.786	0.728	1.02	0.858	0.814	0.796	0.12	0.912	0.903	0.898	0.14	0.948	0.900	0.889	0.07
2	0.856	0.748	0.663	0.44	0.873	0.775	0.710	1.41	0.923	0.914	0.911	0.23	0.932	0.908	0.894	0.46	0.903	0.823	0.785	0.1
3	0.959	0.923	0.916	0.40	0.962	0.927	0.921	1.19	0.867	0.838	0.829	0.32	0.842	0.798	0.784	0.58	0.972	0.946	0.943	0.09
4	0.906	0.829	0.793	0.39	0.891	0.803	0.755	1.02	0.886	0.849	0.816	0.73	0.924	0.903	0.896	0.26	0.935	0.878	0.861	0.09
5	0.874	0.776	0.711	0.43	0.897	0.814	0.771	1.15	0.946	0.938	0.929	0.41	0.876	0.823	0.818	0.29	0.971	0.944	0.941	0.10
6	0.943	0.892	0.879	0.54	0.947	0.899	0.888	1.22	0.942	0.938	0.926	0.68	0.945	0.929	0.913	0.45	0.956	0.916	0.908	0.17
7	0.756	0.608	0.354	0.43	0.878	0.783	0.724	1.03	0.852	0.822	0.798	0.52	0.869	0.834	0.823	0.47	0.878	0.784	0.724	0.11
8	0.872	0.773	0.706	0.44	0.959	0.923	0.916	1.13	0.926	0.873	0.854	0.67	0.913	0.896	0.821	0.56	0.912	0.893	0.886	0.09
a~	0.876	0.785	0.707	0.43	0.911	0.838	0.801	1.14	0.900	0.873	0.857	0.46	0.901	0.874	0.855	0.40	0.934	0.885	0.867	0.10

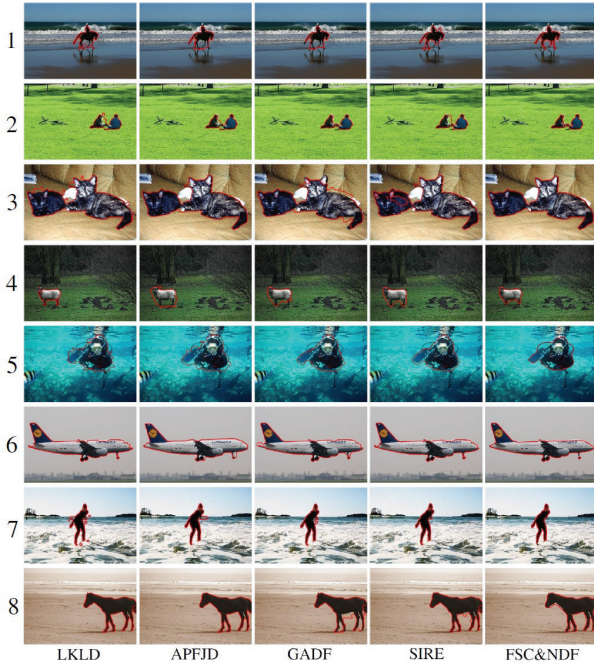


Fig. 9. Comparison results with ACMs. FSC&NDF achieves better segmentation effects on simple and complex objects.

ments of the model with conventional and noisy images from previous ACMs are conducted. Comparisons with Mask R-CNN [24], BlendMask [38], PolarMask [39], DeepSnake [40], E2EC [41], and SAM [42] is discussed. Ablation and platform experiments are also conducted. For experiments on traditional ACMs and robustness, we adopt dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), and confidence similarity coefficient (CSC) as evaluation metrics, which are defined as follows:

$$\text{DSC} = \frac{2(|M_a| \cap |M_b|)}{|M_a| + |M_b|} \quad (33)$$

$$\text{JSC} = \frac{|M_a| \cap |M_b|}{|M_a| \cup |M_b|} \quad (34)$$

$$\text{CSC} = 3 - \frac{|M_a| + |M_b|}{|M_a| \cap |M_b|} \quad (35)$$

where M_a represents the segmented area and M_b represents the ground-truth area. For experiments with deep learning models, the standard AP, AP₅₀, AP₇₅, AP_S, AP_M, and AP_L are

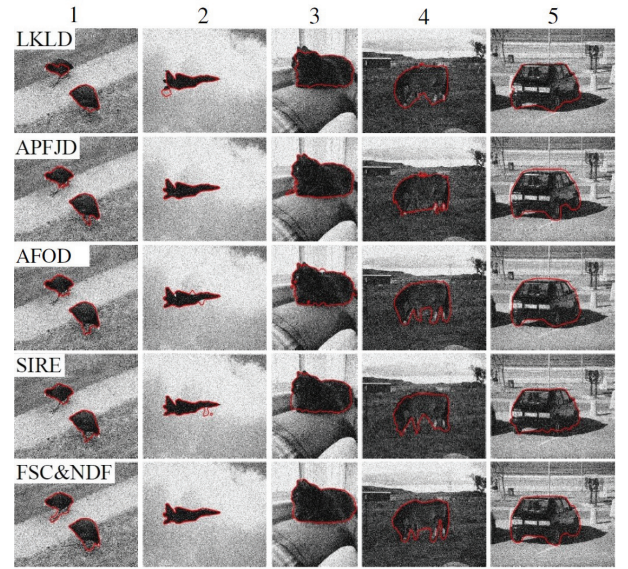


Fig. 10. Noise segmentation results. Gaussian noise with a variance of 0.05 is applied, which is a common strong noise.

used to evaluate the results. All experiments are conducted on Python 3.8 with PyTorch 1.13.0 on an Intel¹ core² i9-13900KF 3.00-GHz NVIDIA RTX 4090 workstation. The parameters are fixed according to Algorithm 1.

A. Comparative Experiments With Previous ACMs

Four recently proposed ACMs, LKLD [18], APFJD [17], GADF [30], and SIRE [32], are adopted to conduct comparative experiments against the proposed FSC&NDF. The segmentation results on natural images are shown in Fig. 9, LKLD falls into local optimal solutions, so that many noise points appear when segmenting the target area. For the other three models, there are evident under-segmentation and over-segmentation. For image 1, only FSC&NDF recognizes the leg contour, for image 3, the result of APFJD is the closest to the proposed model, and the other models have large deviations. The contrast between the target and the background in image 5 is low, and the grayscale information is complex, thus significant errors are noticeable, while FSC&NDF still accurately segments the target. The metrics values of each model are given in Table I: all the values of FSC&NDF

TABLE II
METRICS VALUES AND SEGMENTATION TIME T (s) FOR NOISY IMAGES. A~ REFERS TO THE AVERAGE VALUE OF EACH METRIC

Images	LKLD				APFJD				AFOD				SIRE				FSC&NDF			
	DSC	JSC	CSC	T	DSC	JSC	CSC	T	DSC	JSC	CSC	T	DSC	JSC	CSC	T	DSC	JSC	CSC	T
1	0.840	0.725	0.621	1.62	0.841	0.735	0.624	0.77	0.888	0.798	0.748	0.82	0.874	0.777	0.714	0.79	0.923	0.898	0.876	0.12
2	0.919	0.851	0.825	1.28	0.939	0.885	0.870	0.68	0.934	0.877	0.859	0.68	0.925	0.883	0.864	0.53	0.946	0.894	0.879	0.08
3	0.867	0.766	0.695	1.44	0.863	0.830	0.825	1.23	0.865	0.843	0.819	1.22	0.945	0.896	0.884	1.29	0.965	0.933	0.928	0.22
4	0.839	0.723	0.617	1.60	0.885	0.794	0.741	1.67	0.868	0.766	0.696	1.67	0.825	0.762	0.739	0.83	0.926	0.898	0.876	0.14
5	0.847	0.800	0.789	1.60	0.922	0.887	0.881	1.71	0.941	0.888	0.875	1.52	0.937	0.899	0.891	0.76	0.957	0.918	0.912	0.14
a~	0.862	0.773	0.709	1.51	0.890	0.826	0.788	1.21	0.899	0.834	0.799	1.18	0.901	0.843	0.818	0.84	0.943	0.908	0.894	0.14

are at the forefront, with the only exception of image 8, APFJD has a higher accuracy than FSC&NDF. However, its segmentation time 1.13 lags behind FSC&NDF by 0.09, and the accuracy of FSC&NDF remains acceptable. Moreover, FSC&NDF shows the highest average DSC, JSC, CSC, and the shortest segmentation time T . In general, the proposed model has the best segmentation performance on most images. Although the performance of previous ACMs on a few images is similar to that of the proposed model, its greatest advantage lies in the automatic contour initialization, which is a unique feature of the proposed solution.

The antinoise experimental results are given in Fig. 10, where strong Gaussian noise is applied to these images. Notice how the noise has different impacts on different methods: LKLD has errors when segmenting images 1–3; APFJD has over-segmentation when dealing with image 3; the same problem appears when AFOD and SIRE segment images 2 and 3. For images 4 and 5, all the algorithms have difficulty in processing low-contrast details, except for FSC&NDF. This fact is collaborated by the metrics in Table II: FSC&NDF outperforms all the other models in each metric, having all the DSC, JSC, and CSC values constantly above 0.85, while its average segmentation time is shortened to 0.14, which is deemed satisfactory for the application at hand. The DoG operator is designed to suppress the noise, while, due to the presence of Gaussian convolutions, the real boundary information is also smoothed. With the help of the NDF module, the grayscale value diffuses to the boundary; hence, the total grayscale value of the image remains unchanged, thereby achieving boundary enhancement and reducing the background interference. In addition, the precise description of the key area by the FSCs enhances the contour guidance ability: indeed, the force component that drives the contour in the right direction becomes larger, thus enabling noise crossing at a faster speed.

B. Comparative Experiments With Deep Learning Models

The segmentation ability of FSC&NDF is comparable to other instance segmentation algorithms, the comparison results are shown in Fig. 11. Images 2 and 3 are relatively easy to segment, and the segmentation results of the six models are ideal. Images 1 and 5 are challenging, with multiple overlapping objects: FSC&NDF successfully segments all the objects and has comparable results compared to all the other methods. As a large-vision model, SAM shows excellent performance except for image 2, with missegmentation of

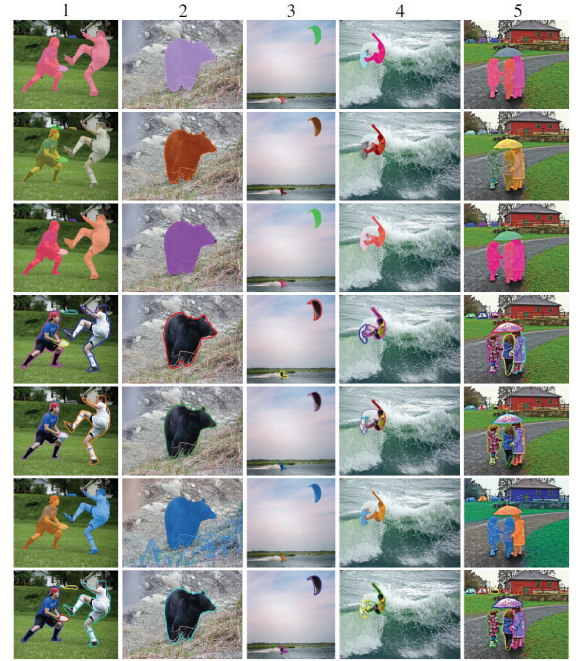


Fig. 11. Qualitative results of several models on the COCO dataset. Row 1: Mask R-CNN. Row 2: BlendMask. Row 3: PolarMask. Row 4: DeepSnake. Row 5: E2EC. Row 6: SAM. Row 7: FSC&NDF.

TABLE III
RESULTS ON THE COCO DATASET. AP^{val} REFERS TO AP VALUE OF THE VALIDATION SET. THE REMAINING METRICS VALUES ARE OBTAINED FROM THE TESTING SET

	Backbone	FPS	AP^{val}	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Mask R-CNN	ResNet-101	3.5	37.4	32.3	50.2	39.3	18.2	40.6	53.6
BlendMask	ResNet-101	4.1	38.9	34.5	55.6	42.6	22.4	43.7	52.8
PolarMask	ResNet-101	6.7	36.2	32.1	56.4	37.7	17.8	39.6	51.5
DeepSnake	DLA-34	5.8	37.6	31.8	58.2	45.3	18.6	38.2	51.9
E2EC	DLA-34	6.8	39.1	33.5	59.3	41.9	19.6	40.3	53.4
SAM	ViT Encoder	3.6	43.6	39.8	60.1	45.8	25.7	42.8	62.7
FSC&NDF	GELAN + Conv	7.1	39.6	40.1	59.9	42.3	18.8	44.6	50.1

* The bold values indicate that FSC&NDF surpasses all models except SAM in this metric, while the bold and italic values indicate that FSC&NDF surpasses all models in this metric.

background information. Table III lists the comparison of the metric values of all models, where FPS refers to the number of images that the model can process per second. FSC&NDF performs better than the other models except SAM in AP^{val} and AP_{50} and performs the best in FPS, AP, and AP^M . Under the transformer architecture, which is based on a large-pretraining dataset, the segmentation accuracy

TABLE IV
RESULTS OF DIFFERENT SUPERPIXEL SEGMENTATION
METHODS ON THE COCO DATASET

Methods	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
Baseline + SLIC	33.6	52.1	35.4	13.2	37.6	42.5
Baseline + LSC	34.7	50.4	35.5	14.2	37.1	44.8
Baseline + GMMSP	35.3	54.6	37.5	12.8	38.4	45.2
Baseline + MMGR-WT	38.2	55.9	37.8	13.4	40.2	47.4
Baseline + AMR-WT (FSC&NDF)	40.1	59.9	42.3	18.8	44.6	50.1

of SAM is higher than the proposed model. Nevertheless, FSC&NDF has a simpler model architecture, fewer parameters, and application simplicity. Although SAM has higher segmentation accuracy at different scales, FSC&NDF has greater FPS, AP, and AP_M, which shows that the proposed model has faster segmentation speed and better segmentation effect for medium objects. As aforementioned, the superiority of SAM is based on a large dataset, hence, when there is a limited available set of data, it is hard for SAM to achieve ideal results. In general, SAM is superior when the task is difficult and there is sufficient data, otherwise, FSC&NDF is the correct choice to make.

C. Ablation Experiments

1) *Effects of Different Superpixel Segmentation Methods:* Since superpixels are crucial components of the proposed model, experiments based on superpixels with different shapes are reported. The proposed solution thus implicitly demonstrates that it is compatible with different morphological superpixels. SLIC [21], LSC [22], GMMSP [36], and MMGR-WT [37] are adopted, whose results are qualitatively depicted in Fig. 6. Results based on these four state-of-the-art superpixel segmentation methods are quantified and reported in Table IV. The effectiveness based on different choices is acceptable, which indicates that alternative superpixel generation processes can be adopted. Notice that the overall performance of the proposed model with AMR-WT attains the best results.

2) *Effects of Different Regional Representative Elements:* In order to empirically prove the superiority of (20) in computing the representative feature of each superpixel region denoted by X_r -L1, experiments based on different feature computational methods are conducted. Namely, X_r -L1 is compared against mean, variance, standard deviation (std), entropy, and (20) based on L2 norm (X_r -L2), that is,

$$X_r = \underbrace{\arg \min}_{x_{ri}} \left\{ \sum_{j=q, q \neq i}^{N'} (x_{ri} - x_{rj})^2 \right\}, \quad i = 1, 2, \dots, N' \quad (36)$$

where the meaning of the variables is the same as of (20). From Table V, which reports the quantitative comparison, the segmentation effects based on linear feature extraction operators are better than those based on nonlinear operators, which is due to the magnifying characteristics of the nonlinear operator, e.g., small (or large) distances become more extreme after being processed by such operators, thus leading to an

TABLE V
RESULTS OF DIFFERENT SUPERPIXEL FEATURE CALCULATION
METHODS ON THE COCO DATASET

Methods	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
Baseline + mean	38.3	56.7	40.4	16.5	42.9	48.8
Baseline + variance	35.9	54.8	39.2	14.1	41.5	46.6
Baseline + std	38.1	55.7	40.2	17.3	43.9	49.4
Baseline + entropy	36.7	56.5	38.6	14.7	41.2	47.1
Baseline + X_r -L2	37.4	55.3	37.6	15.0	40.2	48.4
Baseline + X_r-L1 (FSC&NDF)	40.1	59.9	42.3	18.8	44.6	50.1

accuracy loss. Instead, among all the linear operators, X_r -L1 retains more information on the distance, so the extracted feature is more stable.

3) *Effects of Different Parameter Values:* The solution parameters reported in Algorithm 1 have been chosen by fine-tuning on the experimental data. More precisely, when the values of m and N increase to a certain extent, (13) and (29) remain convergent, so the impact of a further increase on the segmentation accuracy becomes negligible. Instead, the impact of different parameter choices for λ , ω_1 , ω_2 , σ_1 , σ_2 , and γ is more involved and asks for experimental fine tuning. As shown in Table VI (for λ , ω_1 , and ω_2) and Table VII (for σ_1 , σ_2 , and γ). The segmentation accuracy under different figures of merit (namely, DSC, JSC, and CSC) reaches the maximum around the reported choices.

D. Complexity Analysis

To further analyze the proposed approach, the complexity of the overall solution is reported, with different metric values according to the different compared models. For ACMs, the standard complexity representation $O(f(n))$ is adopted (Table VIII); for deep learning models, giga floating point operations per second (GFLOPs) and parameters (Params) are computed to quantify time and space complexity (Table IX). In addition, the inference time of all models is also recorded, and all statistics are yielded based on the COCO benchmark.

Note that the proposed model is compared with previous ACMs without the contour generation module (w/o CG), and it is compared with deep learning models with the contour generation module (w/ CG), for the sake of fairness. The time complexity of GADF and SIRE is higher than that of the other three models, with inference speed faster than LKLD and APFJD. LKLD, APFJD, and the proposed model have the same complexity, while the proposed model is the fastest among all models. In Table IX, the metrics values of FSC&NDF are calculated based on YOLOv9, which can be replaced by other models. The point is that with a lightweight model like YOLOv9 as the front-end network, our method is able to handle most of the tasks. Notice that the inference time of FSC&NDF ranks among the best, while the complexity is the lowest.

E. Validation Based on Experimental Platform

The proposed FSC&NDF can be deployed on hardware platforms to assist industrial production, Fig. 12 shows our gantry

TABLE VI
RESULTS OF DIFFERENT λ , ω_1 , AND ω_2 ON THE COCO DATASET

# λ / ω_1 / ω_2	2 / 3 / 3	5 / 6 / 6	8 / 10 / 10	10 / 15 / 15	13 / 20 / 20	16 / 24 / 24	18 / 27 / 27
DSC	0.756 / 0.823 / 0.785	0.812 / 0.855 / 0.813	0.857 / 0.895 / 0.858	0.934 / 0.934 / 0.934	0.906 / 0.935 / 0.917	0.854 / 0.878 / 0.856	0.799 / 0.854 / 0.817
JSC	0.738 / 0.811 / 0.747	0.782 / 0.792 / 0.825	0.835 / 0.881 / 0.835	0.885 / 0.885 / 0.885	0.882 / 0.857 / 0.875	0.847 / 0.863 / 0.821	0.783 / 0.839 / 0.794
CSC	0.696 / 0.745 / 0.671	0.785 / 0.833 / 0.766	0.822 / 0.804 / 0.824	0.867 / 0.867 / 0.867	0.874 / 0.834 / 0.849	0.823 / 0.842 / 0.776	0.740 / 0.828 / 0.756

TABLE VII
RESULTS OF DIFFERENT σ_1 , σ_2 , AND γ ON THE COCO DATASET

# σ_1 / σ_2 / γ	0.1 / 3.1 / 3	0.3 / 3.3 / 5	0.5 / 3.5 / 7	0.7 / 3.7 / 9	1 / 4 / 11
DSC	0.538 / 0.675 / 0.918	0.607 / 0.793 / 0.924	0.934 / 0.934 / 0.934	0.783 / 0.827 / 0.916	0.675 / 0.785 / 0.897
JSC	0.505 / 0.636 / 0.856	0.588 / 0.688 / 0.861	0.885 / 0.885 / 0.885	0.762 / 0.809 / 0.869	0.642 / 0.746 / 0.845
CSC	0.479 / 0.618 / 0.848	0.562 / 0.731 / 0.849	0.867 / 0.867 / 0.867	0.749 / 0.784 / 0.842	0.619 / 0.712 / 0.829

TABLE VIII
COMPARISON OF ALGORITHM COMPLEXITY WITH ACMs

Methods	Time complexity	Space complexity	Inference time (s)
LKLD	$O(n)$	$O(n^2)$	0.673
APFJD	$O(n)$	$O(n^2)$	1.246
GADF	$O(n^2)$	$O(n^2)$	0.478
SIRE	$O(n^2)$	$O(n^2)$	0.562
FSC&NDF (w/o CG)	$O(n)$	$O(n^2)$	0.095

TABLE IX
COMPARISON OF ALGORITHM COMPLEXITY WITH
DEEP LEARNING MODELS

Methods	Params (M)	GFLOPs (G)	Inference time (s)
Mask R-CNN	44.18	275.5	0.195
BlendMask	38.96	264.3	0.111
PolarMask	34.46	252.7	0.122
DeepSnake	36.28	223.6	0.105
E2EC	38.72	245.8	0.097
SAM	636	3851.4	0.456
FSC&NDF (w/ CG)	25.3	102.3	0.137

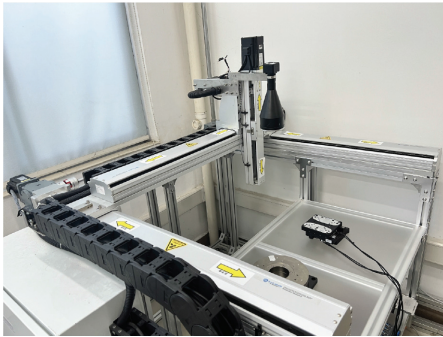


Fig. 12. Laser processing platform based on gantry robot.

robot for laser processing. The platform combines a CMOS camera and a telecentric lens to achieve target segmentation and positioning. First, the camera is moved directly above

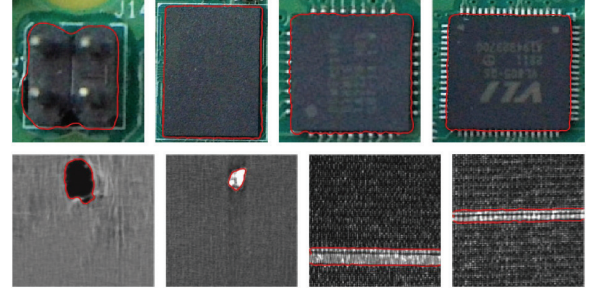


Fig. 13. Platform segmentation results. Row 1: chips segmentation. Row 2: fabric defect segmentation.

the processing plane manually. At this time, the processing plane displays the source pattern to be engraved on the material. The digital image is acquired by the camera, which is processed by the FSC&NDF to obtain the pattern contour. Then, the pixel coordinates of the contour are converted to real world coordinates using the intrinsic and extrinsic calibration matrices of the camera. The position information of this coordinate are transmitted to the gantry robot to guide the laser to the starting position, where the original material is correctly placed. Finally, the laser moves along the coordinates of each contour vertex, which engraves the target pattern on different materials. To verify the practicality of the algorithm, two materials are selected for real-time segmentation under the telecentric lens. In Fig. 13, Row 1 shows the segmentation results of the chip, and Row 2 shows the segmentation results of the fabric defect. It is obvious that the proposed model performs well at the microscopic level.

V. CONCLUSION AND FUTURE WORK

In this article, an instance segmentation algorithm, called FSC&NDF, combined with YOLOv9 and advanced ACM is proposed. The novel innovations are: the combination of YOLOv9 and AMR-WT realizes initial contour positioning; the NDF module is designed to enhance the target information and suppress the background noise; the superpixel clustering boosts the feature information extraction. FSC&NDF frees tra-

ditional ACMs from manual contour initialization and shows segmentation effects that are not inferior to mainstream deep learning methods. AMR-WT acts as a glue to make YOLOv9 and ACM work together harmoniously. The proposed NDF module and superpixel fuzzy clustering improve the noise resistance effect and feature capture ability.

In the future work, we will extend the applicability of the approach to various object detection algorithms so as to solve segmentation tasks in different fields, i.e., transferability. Instead, to increase its effectiveness, the algorithm will be updated to improve the evolutionary ability of ACMs together with their information description on the inner and outer contour areas, which can be modified to better guide the contour iteration. Finally, we are planning to use the feature map output by a CNN as the input to the energy function, leading to the improvement of its feature analysis ability.

APPENDIX PROOF OF PROPOSITION 2

From the constraints, the Lagrange function of (19) is

$$L = \sum_{r=1}^{N^S} \sum_{k=1}^2 S_r b_{kr} \|X_r - c_k\|^2 + \sum_{r=1}^{N^S} \sum_{k=1}^2 S_r b_{kr} \cdot \ln \left(\frac{b_{kr}}{1 - \xi_k} \right) + \sum_{r=1}^{N^S} \alpha_r \left(1 - \sum_{k=1}^2 b_{kr} \right) + \sum_{r=1}^{N^S} \beta_r \left(1 - \sum_{k=1}^2 (1 - \xi_k) \right)$$

$$\text{s.t. } \sum_{k=1}^2 b_{kr} = 1, \sum_{k=1}^2 (1 - \xi_k) = 1$$

where α_r and β_r are the Lagrange multipliers. Then, the partial derivative of each variable is taken, and we have

$$\frac{\partial L}{\partial c_k} = \sum_{r=1}^{N^S} 2 \cdot S_r b_{kr} \|X_r - c_k\| = 0$$

the solution of the above equation yields

$$c_k = \frac{\sum_{r=1}^{N^S} S_r b_{kr} X_r}{\sum_{r=1}^{N^S} S_r b_{kr}}.$$

The next PDE is

$$\frac{\partial L}{\partial b_{kr}} = S_r \left(\|X_r - c_k\|^2 + \left(\ln \left(\frac{b_{kr}}{1 - \xi_k} \right) + 1 - \xi_k \right) \right) - \alpha_r = 0.$$

By shifting the terms, there exists

$$b_{kr} = (1 - \xi_k) \exp \left(\frac{\alpha_r}{S_r} + \xi_k - 1 - \|X_r - c_k\|^2 \right)$$

and substitute the above equation into $\sum_{k=1}^2 b_{kr} = 1$

$$b_{kr} = \frac{(1 - \xi_k) \exp (\xi_k - \|X_r - c_k\|^2)}{\sum_{k=1}^2 ((1 - \xi_k) \exp (\xi_k - \|X_r - c_k\|^2))}.$$

The final PDE is

$$\frac{\partial L}{\partial \xi_k} = \sum_{r=1}^{N^S} \left(\frac{S_r b_{kr}}{1 - \xi_k} + \beta_r \right) = 0$$

since $\sum_{k=1}^2 (1 - \xi_k) = 1$, which gives rise to

$$\xi_k = 1 - \frac{\sum_{r=1}^{N^S} S_r b_{kr}}{\sum_{r=1}^{N^S} S_r}.$$

REFERENCES

- [1] H. Bi, L. Xu, X. Cao, Y. Xue, and Z. Xu, "Polarimetric SAR image semantic segmentation with 3D discrete wavelet transform and Markov random field," *IEEE Trans. Image Process.*, vol. 29, pp. 6601–6614, 2020.
- [2] B. Chen et al., "Attention-guided and noise-resistant learning for robust medical image segmentation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024.
- [3] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2021.
- [4] B. Wang, D. Tao, R. Dong, Y. Tang, and X. Gao, "A contour co-tracking method for image pairs," *IEEE Trans. Image Process.*, vol. 30, pp. 5402–5412, 2021.
- [5] G. Wang, F. Zhang, Y. Chen, G. Weng, and H. Chen, "An active contour model based on local pre-piecewise fitting bias corrections for fast and accurate segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.
- [6] H. Li, J. Song, and L. Dai, "Superpixel segmentation based on feature fusion and boundary constraint for ferrograph image segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.
- [7] Z. Liu, X. Bai, H. Liu, and Y. Zhang, "Multiple-surface-approximation-based FCM with interval memberships for bias correction and segmentation of brain MRI," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2093–2106, Sep. 2020.
- [8] C. J. J. Sheela and G. Suganthi, "Brain tumor segmentation with radius contraction and expansion based initial contour detection for active contour model," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 23793–23819, Sep. 2020.
- [9] S. Zhao et al., "Attractive deep morphology-aware active contour network for vertebral body contour extraction with extensions to heterogeneous and semi-supervised scenarios," *Med. Image Anal.*, vol. 89, Oct. 2023, Art. no. 102906.
- [10] H. Qi, X. Kong, L. Cheng, J. Hu, and J. Gu, "Addressing fine-grained lake water body extraction: A hybrid approach combining vision transformer and geodesic active contour," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4204614.
- [11] P. Polewski, J. Shelton, W. Yao, and M. Heurich, "Instance segmentation of fallen trees in aerial color infrared imagery using active multi-contour evolution with fully convolutional network-based intensity priors," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 297–313, Aug. 2021.
- [12] Y. Chen, P. Ge, G. Wang, G. Weng, and H. Chen, "An overview of intelligent image segmentation using active contour models," *Intell. Robot.*, vol. 3, no. 1, pp. 23–55, 2023.
- [13] V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische Math.*, vol. 66, no. 1, pp. 1–31, Dec. 1993.
- [14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
- [15] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.
- [16] G. Wang, Z. Li, G. Weng, and Y. Chen, "An optimized denoised bias correction model with local pre-fitting function for weak boundary image segmentation," *Signal Process.*, vol. 220, Jul. 2024, Art. no. 109448.
- [17] P. Ge, Y. Chen, G. Wang, and G. Weng, "An active contour model driven by adaptive local pre-fitting energy function based on Jeffreys divergence for image segmentation," *Expert Syst. Appl.*, vol. 210, Dec. 2022, Art. no. 118493.
- [18] C. Yang, G. Weng, and Y. Chen, "Active contour model based on local Kullback–Leibler divergence for fast image segmentation," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106472.
- [19] R. Jin and G. Weng, "Active contours driven by adaptive functions and fuzzy c-means energy for fast image segmentation," *Signal Process.*, vol. 163, pp. 1–10, Oct. 2019.

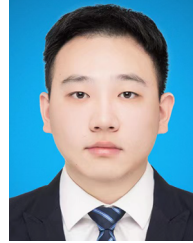
- [20] G. Liu, Y. Jiang, B. Chang, and D. Liu, "Superpixel-based active contour model via a local similarity factor and saliency," *Measurement*, vol. 188, Jan. 2022, Art. no. 110442.
- [21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [22] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1356–1363.
- [23] T. Lei, X. Jia, T. Liu, S. Liu, H. Meng, and A. K. Nandi, "Adaptive morphological reconstruction for seeded image segmentation," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5510–5523, Nov. 2019.
- [24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [25] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2024, pp. 1–21.
- [26] W. Liu et al., "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, Jan. 2016, pp. 21–37.
- [27] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
- [28] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 642–656, Mar. 2020.
- [29] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6568–6577.
- [30] D. Chen, J. Spencer, J.-M. Mirebeau, K. Chen, M. Shu, and L. D. Cohen, "A generalized asymmetric dual-front model for active contours and image segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 5056–5071, 2021.
- [31] X. Yan and G. Weng, "Hybrid active contour model driven by optimized local pre-fitting image energy for fast image segmentation," *Appl. Math. Model.*, vol. 101, pp. 586–599, Jan. 2022.
- [32] B. Dong, G. Weng, Q. Bu, Z. Zhu, and J. Ni, "An active contour model based on shadow image and reflection edge for image segmentation," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122330.
- [33] K. Zhang, H. Song, and L. Zhang, "Active contours driven by local image fitting energy," *Pattern Recognit.*, vol. 43, no. 4, pp. 1199–1206, Apr. 2010.
- [34] M.-M. Li and B.-Z. Li, "A novel active contour model for noisy image segmentation based on adaptive fractional order differentiation," *IEEE Trans. Image Process.*, vol. 29, pp. 9520–9531, 2020.
- [35] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [36] Z. Ban, J. Liu, and L. Cao, "Superpixel segmentation using Gaussian mixture model," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4105–4117, Aug. 2018.
- [37] C. Wu, L. Zhang, H. Zhang, and H. Yan, "Improved superpixel-based fast fuzzy C-means clustering for image segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1455–1459.
- [38] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan, "BlendMask: Top-down meets bottom-up for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8570–8578.
- [39] E. Xie et al., "PolarMask: Single shot instance segmentation with polar representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12190–12199.
- [40] S. Peng, W. Jiang, H. Pi, X. Li, H. Bao, and X. Zhou, "Deep snake for real-time instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8530–8539.
- [41] T. Zhang, S. Wei, and S. Ji, "E2EC: An end-to-end contour-based method for high-quality high-speed instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4433–4442.
- [42] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2023, pp. 4015–4026.



Yiyang Chen (Member, IEEE) received the M.Eng. degree from the Imperial College London, London, U.K., in 2013, and the Ph.D. degree from the University of Southampton, Southampton, U.K., in 2017.

After that, he worked as a Research Fellow in control systems from 2017 to 2018 and in traffic signal control from 2018 to 2020 with the University of Southampton. He joined the School of Mechanical and Electrical Engineering, Soochow University, Suzhou, China, in 2020, as an Associate Professor.

He has published several articles in top control conferences and journals. His research interests include iterative learning control, optimization, artificial intelligence, image processing, and robotic systems.



Fuzheng Zhang received the B.S. degree in vehicle engineering from Shanghai University of Engineering Science, Shanghai, China, in 2022. He is currently pursuing the M.S. degree in image processing with the School of Mechanical and Electrical Engineering, Soochow University, Suzhou, China.

His research interests include image segmentation, artificial intelligence, deep learning, and LLM.



Guina Wang received the M.S. and Ph.D. degrees from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2007 and 2021, respectively.

She is currently an Associate Professor with the School of Mechanical and Electric Engineering, Soochow University, Suzhou, China. Her current research interests include optical sensing technology and image processing techniques.



Guirong Weng received the M.S. degree from Nanjing University of Science and Technology, Nanjing, China, in 1985.

He is currently a Professor with the School of Mechanical and Electric Engineering, Soochow University, Suzhou, China, and an Executive Director of Jiangsu Province Automation Society. His research interests include image processing and biological information processing.



Daniele Fontanelli (Fellow, IEEE) received the Ph.D. degree in automation, robotics, and bioengineering from the University of Pisa, Pisa, Italy, in 2006.

He is the Co-Founder of Polytec Intralogistics Srl and IDRA Labs. He is the PI of the EU Project MAGICIAN-iMmersive leArninG for ImperfeCtion detection and repAir through human-robot interaction, and he was the Co-Founder and the PI of the EIT-Digital international master's on "Autonomous Systems" from 2017 to 2023. He is currently a

Full Professor with the Department of Industrial Engineering, University of Trento, Trento, Italy. He has authored or co-authored more than 250 scientific articles in peer-reviewed top journals and conference proceedings. His research interests include distributed and real-time estimation and control, localization algorithms, synchrophasor estimation, clock synchronization algorithms, resource aware control, wheeled mobile robots, service robotics, and human-robot interaction and estimation.

Dr. Fontanelli is a member of the IMEKO TC17-Measurement in Robotics. He is a Senior Area Editor of IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT and an Associate Editor of IEEE ROBOTICS AND AUTOMATION LETTERS.