Article

Dual-Fusion Active Contour Model with Semantic Information for Saliency Target Extraction of Underwater Images

Shudi Yang, Jiaxiong Wu * and Zhipeng Feng

School of Mechanical Engineering, University of Science and Technology, Beijing 100083, China; b20170278@xs.ustb.edu.cn (S.Y.); fengzpz@ustb.edu.cn (Z.F.)

* Correspondence: b20170277@xs.ustb.edu.cn

Abstract: Underwater vision research is the foundation of marine-related disciplines. The target contour extraction is significant for target tracking and visual information mining. Aiming to resolve the problem that conventional active contour models cannot effectively extract the contours of salient targets in underwater images, we propose a dual-fusion active contour model with semantic information. First, the saliency images are introduced as semantic information and salient target contours are extracted by fusing Chan–Vese and local binary fitting models. Then, the original underwater images are used to supplement the missing contour information by using the local image fitting. Compared with state-of-the-art contour extraction methods, our dual-fusion active contour model can effectively filter out background information and accurately extract salient target contours. Moreover, the proposed model achieves the best results in the quantitative comparison of MAE (mean absolute error), ER (error rate), and DR (detection rate) indicators and provides reliable prior knowledge for target tracking and visual information mining.

Citation: Yang, S.; Wu, J.; Feng, Z. Dual-Fusion Active Contour Model with Semantic Information for Saliency Target Extraction of Underwater Images. *Appl. Sci.* **2022**, *12*, 2515. <https://doi.org/10.3390/app12052515>

Academic Editor: Antonio Fernández

Received: 13 January 2022

Accepted: 21 February 2022

Published: 28 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: underwater image; target contour extraction; active contour model; semantic information; saliency target

1. Introduction

In recent years, the development and utilization of the ocean have gradually become an important development direction. Since underwater vision research is the basis of marine-related disciplines, the rapid development of underwater image processing technology is inevitable [1,2]. Image segmentation is a basic method of target extraction, which aims to partition an image into several constituent regions. Each region has coherent intensities, colors, and textures [3]. Furthermore, image segmentation can provide technical support for target tracking, image restoration [4–6], and other tasks.

Now, some results have been achieved in underwater image segmentation. Liu et al. [7] proposed an improved level set algorithm based on the gradient descent method and applied it to segment underwater biological images. Wei et al. [8] improved the K-means algorithm to segment underwater image backgrounds and addressed the issue of improper K value determination. Moreover, this algorithm can minimize the impact of the initial centroid position of a grayscale image. SM et al. [9] used a canny edge detection algorithm to segment underwater images, whereas background noise greatly affected the canny edge detection algorithm. Sun et al. [10] and Li et al. [11] used fuzzy C-means to segment underwater images. Rajeev et al. [12] used the K-means algorithm to segment underwater images. However, the clustering algorithms mentioned above are greatly affected by the local gray unevenness of underwater images. Moreover, clustering algorithms contain local convergence errors and are only suitable for underwater images with a single background gray level.

Some investigators have segmented underwater images based on their optical properties and achieved good results. For example, Chen et al. [13] proposed an optical feature

extraction, calculation, and decision method to identify the collimated region of artificial light and employed a level set method to segment the objects within the collimated region. This method could better identify the target region, but the level set method could not filter out background noise when the target region contained background information. Xuan [14] et al. proposed an RGB (red, green, blue) color channel fusion segmentation method for underwater images. The proposed method obtained the grayscale image with high foreground-background contrast and employed the thresholding segmentation method to conduct fast segmentation. However, the disadvantage of this method is that when the color of the background region is similar to the foreground region, the target cannot be segmented.

Active contour models have also been used for underwater image segmentation. Zhu et al. [15] used the cluster-based algorithm for co-saliency detection and highlighted salient regions in the underwater images. Then, the local statistical active contour model was used to extract the target contours of underwater images. Qiao et al. [16] proposed an improved method based on the active contour model. The method used the RGB color space and contrast limited adaptive histogram equalization (CLAHE) to increase the contrast of the sea cucumber thorns and body. Then, the method extracted the edge of the sea cucumber thorns using the active contour model. Li et al. [17] improved the traditional level set methods by avoiding the calculation of the signed distance function (SDF) to segment underwater images. The improved method could speed up the computational complexity without re-initialization. Bai et al. [18] proposed a method based on morphological component analysis (MCA) and adaptive level set evolution to segment underwater images. The MCA decomposes the image into texture and cartoon parts sparsely. The new adaptive level set evolution method combines the threshold piecewise function with the evariable right coefficient and halting speed function and was used to obtain the edges of the cartoon part. Shelei et al. [19] segmented underwater grayscale images by fusing the geodesic active contour model (GAC) and the Chan–Vese (CV) model. However, this method required the target region of the underwater image to have a uniform grayscale. Chen et al. [20] integrated the transmission map and the saliency map into a unified level set formulation to extract the salient target contours of the underwater images. Antonelli et al. [21] proposed some spatially varying regularization methods by using local image features (such as textures, edge, noise, etc.) to prevent excessive regularization in smooth regions and preserve spatial features in nonsmooth regions. However, it is difficult for this method to obtain accurate regularization parameters for underwater images with low contrast and blurred texture information due to blurred local features, such as textures and edges. Nawal et al. [22] presented an efficient approach for unsupervised segmentation based on image feature extraction to segment natural and textural images. Unfortunately, this method is unsuitable for natural underwater images with weak texture information.

As a new technology of image processing, the neural network has also been used for underwater image segmentation. O’Byrne et al. [23] proposed the use of photorealistic synthetic imagery to train deep encoder-decoder networks. This method synthesized virtual underwater images, and each rendered image had a corresponding ground truth per-pixel label map. The mapping relationship between the underwater images and the segmented images was established by training the encoder-decoder network. Zhou et al. [24] proposed a deep neural network architecture for underwater scene segmentation. The architecture extracted features by pre-training VGG-16 and learned to expand the lower resolution feature maps using the decoder. The neural network achieved certain results in underwater image segmentation, but the lack of underwater data sets with corresponding functions is still a problem.

Most of the existing underwater image segmentation methods are used to segment images with high foreground-background contrast and single background grayscale. When the underwater images with varying background grayscale and the targets have complex textures, the segmentation results of the above methods are not satisfactory. To address the above problem, we propose a novel dual-fusion model with semantic information for

salient object segmentation of underwater images with complex backgrounds. In summary, the contributions of our model are as follows:

1. We introduce saliency maps as semantic information to segment foreground information and background information;
2. The dual-fusion energy equation is proposed to extract the contours of saliency targets by integrating the local and global intensity fitting term;
3. For the missing saliency target information, we propose the correction module to correct the saliency target contour error by introducing the original image contour information.

This paper is organized as follows: Section 1 reviews related works. In Section 2, we introduce in detail the derivation process of the dual-fusion model. Section 3 shows the experimental process, and we compare the proposed method with state-of-the-art segmentation methods, and the results demonstrate the superiority of the proposed methods. The conclusion of this paper is shown in Section 4.

2. Related Works

2.1. The Chan–Vese (CV) Model

The Chan–Vese (CV) model [25] was initially derived from the Mumford–Shah (MS) functional [26]. The MS functional aims to find an optimal contour C that divides the image domain into disjoint subregions and a piecewise smooth approximation image $I : \Omega_i \subset \mathbb{R}$ from the original image $I_0 : \Omega \subset \mathbb{R}^2$, and the energy functional of MS can be expressed as follows:

$$E^{\text{MS}}(I, C) = \int_{\Omega} (I_0 - I)^2 dx + \mu \int_{\Omega \setminus C} |\nabla I|^2 dx + v|C| \quad (1)$$

where $\mu, v \geq 0$ are positive weighting constants, $|C|$ is the length of the contour C , and ∇I is the derivative of the function I . However, the non-convexity of the above energy functional makes it difficult to be minimized. Besides, the MS functional is restricted to piecewise constant equations, so the CV model has been proposed to simplify and modify the MS functional. The basic idea of the CV model is to find a particular partition to separate a given image into foreground and background. The energy functional of the CV model can be defined as follows:

$$E^{\text{CV}}(C, c_1, c_2) = \lambda_1 \int_{\text{in}(C)} (I_0 - c_1)^2 dx + \lambda_2 \int_{\text{out}(C)} (I_0 - c_2)^2 dx + v \cdot \text{len}(C) + \mu \cdot \text{area}(\text{in}(C)) \quad (2)$$

where $\mu, v, \lambda_1, \lambda_2 \geq 0$ are positive parameters, $\text{in}(C)$ and $\text{out}(C)$ represent the region inside and outside of the contour C , and c_1 and c_2 are two constants that approximate the image intensity in $\text{in}(C)$ and $\text{out}(C)$, respectively. The Euclidean length term $\text{len}(C)$ is used to regularize the contour. The first 2 terms in Equation (2) are the global binary fitting energy. This energy can be represented by a level set formulation, and then the energy minimization problem can be converted by solving a level set evolution equation. The evolution equation can be expressed as follows:

$$c_1 = \frac{\int_{\Omega} I_0 \cdot H(\phi) dx}{\int_{\Omega} H(\phi) dx}, c_2 = \frac{\int_{\Omega} I_0 \cdot (1 - H(\phi)) dx}{\int_{\Omega} (1 - H(\phi)) dx} \quad (3)$$

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[v \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \mu - \lambda_1 (I_0 - c_1)^2 + \lambda_2 (I_0 - c_2)^2 \right] \quad (4)$$

where $H(\cdot)$ is the Heaviside function, and $\delta(\cdot)$ is the Delta function, which is derivative of the Heaviside function. For Equation (4), v is a scaling parameter. If v is small enough, small targets are likely to be extracted while if v is large, large targets can be detected.

Whereas if the image intensities are inhomogeneous, the global fitting will not be accurate. Therefore, the CV model is not suitable for inhomogeneous images, and the

segmentation results are affected by the position of the initial level set [3]. However, the CV model has better robustness to noise.

2.2. The Local Image Fitting (LIF) Model

Different from the CV model's segmentation of the foreground and background by evolving the level set curve, the local image fitting (LIF) [27] utilizes the local image information to construct a local image fitting energy functional, which can be viewed as a constraint of the differences between the fitting image and the original image. So, the LIF model can ignore the influence of the intensity inhomogeneity, and its energy functional is defined as follows:

$$\frac{\partial \phi}{\partial t} = (I(x) - I^{\text{LFI}}(x))(m_1 - m_2)\delta_\epsilon(\phi), \quad (5)$$

where $I^{\text{LFI}}(x)$ is a local fitted image:

$$I^{\text{LFI}}(x) = m_1 H_\epsilon(\phi(x)) + m_2 (1 - H_\epsilon(\phi(x))), \quad (6)$$

where m_1 and m_2 are the averages of the image intensities of the Gaussian window inside and outside the contour, respectively. m_1, m_2 can be expressed as follows:

$$\begin{cases} m_1 = \text{mean}(I \in (\{x \in \Omega | \phi(x) < 0\} \cap W_k(x))) \\ m_2 = \text{mean}(I \in (\{x \in \Omega | \phi(x) > 0\} \cap W_k(x))) \end{cases}, \quad (7)$$

where $W_k(x)$ is a truncated Gaussian window or a constant window.

Then, the LIF model uses the variation calculus and the steepest descent method to minimize $E^{\text{LIF}}(\phi)$, and the level set evolution equation can be expressed as follows:

$$\frac{\partial \phi}{\partial t} = (I(x) - I^{\text{LFI}}(x))(m_1 - m_2)\delta_\epsilon(\phi), \quad (8)$$

3. Dual-Fusion Active Contour Model

In this section, we propose a dual-fusion active contour model with semantic information to extract the target contours of underwater images. The existing methods cannot extract the target contour from the background without semantic information. So, it is necessary to introduce semantic information and roughly extract the saliency target contour from the complex background. To avoid the extraction error of the saliency target, we introduce the original image contour to correct and supplement the missing contour information. The proposed model can accurately extract the saliency target contour from the complex background using the semantic information and correction module.

3.1. Saliency Image Fitting Energy

This paper uses the pyramid feature attention network [28] to acquire the saliency images. However, due to the low contrast of underwater images, there were some errors in the saliency detection results, such as a local inhomogeneous intensity, background noise, and missing contour information. In view of the local inhomogeneous intensity of the saliency images, we preliminarily employ the local binary fitting to construct the energy functional E_{sal} :

$$E_{\text{sal}}(C, f_1(x), f_2(x)) = \lambda_1 \int_{\text{in}(C)} (S - f_1(x))^2 dx + \lambda_2 \int_{\text{out}(C)} (S - f_2(x))^2 dx, \quad (9)$$

where S is the saliency images, C is a contour in the image domain Ω , and f_1 and f_2 are the image local fitting intensities near the point x . The local fitting intensities f_1, f_2 can be expressed as follows [29,30]:

$$f_1(x) = \frac{K_\sigma(x) \cdot [H_\epsilon(\phi(x))S]}{K_\sigma(x) \cdot H_\epsilon(\phi(x))} \quad (10)$$

$$f_2(x) = \frac{K_\sigma(x) \cdot [(1 - H_\varepsilon(\phi(x)))S]}{K_\sigma(x) \cdot [1 - H_\varepsilon(\phi(x))]} \quad (11)$$

where $K_\sigma(x)$ is the Gaussian kernel, S is the saliency images, and H_ε is the Heaviside function $H(\cdot)$ can be expressed as:

$$H_\varepsilon(x) = \frac{1}{2} \left[1 + \frac{2}{\pi} \arctan\left(\frac{x}{\varepsilon}\right) \right], \quad (12)$$

However, the local binary fitting may introduce some local minimums and is sensitive to noise. Affected by the accuracy of saliency detection, the saliency map of underwater images will inevitably have background noise. Moreover, the initialization curve greatly affects the segmentation results. To solve the aforesaid problems, we introduce the global fitting term from the CV model into the energy functional E_{sal} . The local-global fitting intensities can be expressed as follows:

$$\begin{cases} I_1 = \omega c_1 + (1 - \omega)f_1 \\ I_2 = \omega c_2 + (1 - \omega)f_2 \end{cases} \quad (13)$$

where I_1 and I_2 are the mixed intensity, c_1 and c_2 are two constants derived from Equation (3), and ω is a weight coefficient ($0 \leq \omega \leq 1$). According to the test images in this paper, the value of ω can be taken from 0.5 to 0.9. Moreover, the more inhomogeneous the image intensity, the smaller the value of ω .

With the level set representation, the energy functional can be expressed as follows:

$$E_{\text{sal}}(\phi, I_1(x), I_2(x)) = \lambda_1 \int_{\Omega} (S - I_1(x))^2 H(\phi(x)) dx + \lambda_2 \int_{\Omega} (S - I_2(x))^2 (1 - H(\phi(x))) dx \quad (14)$$

The improved fitting energy E_{sal} not only takes the local intensity information into account but also avoids the local minimization. Therefore, for the saliency images of underwater images, the improved energy functional can extract the contour of the inhomogeneous images more accurately.

3.2. Original Image Fitting Energy

The local inhomogeneous intensity and noise problems can be solved by fusing the local intensity fitting and CV model. However, the missing contour information of the saliency image still needs to be solved. Therefore, the original underwater images are used to make up the missing contour information.

In this paper, we use the local image fitting model (LIF) [27] to extract the contour of original underwater images. The energy functional E_{org} can be expressed as:

$$E_{\text{org}}(\phi) = \frac{1}{2} \int_{\Omega} |I(x) - I^{\text{LFI}}(x)|^2 dx, x \in \Omega, \quad (15)$$

where $I^{\text{LFI}}(x)$ is a local fitted image, as shown in Equation (6). Although the models, such as LBF [29,30], LGIF [31], and RMPCM [3] can extract the target contours of underwater images very well, as shown in Figure 1, the LIF model has higher efficiency. This higher efficiency is because the energy functional of the LIF model does not include a kernel function. Moreover, the LIF model can fit the original image well while reducing the noise significantly by minimizing the difference between the fitted image and the original image.

Figure 1. (a–d) shows the segmentation results of the LBF, LGIF, LIF, and RMPCM model, respectively.

As shown in Figure 1, the LBF, LGIF, and LIF models could extract the target contour better, but LBF was more sensitive to the initial contour curve (as shown in the green dashed area). The energy functional of LGIF and RMPCM both involved the kernel function. The kernel function performs more than one convolution operation for each iteration step, so the evolution speed is slow. The running time of the above models is shown in Table 1.

Table 1. Iterations and CPU time (in seconds).

	LBF	LGIF	LIF	RMPCM
		1239 × 731 pixels		
Iterations	200	200	200	200
Time (s)	93.2969	55.5938	38.5469	63.3052

Figure 1 and Table 1 intuitively show that the LIF model has advantages regarding both the speed and contour extraction results. So, we use the LIF model to extract the original image contour to correct the contour information of the salient target.

3.3. Dual-Fusion Active Contour Model

To use less fitting energy at the target contours than at other locations, we use an edge indicator function [32,33] to indicate target contours. The function can be expressed as follows:

$$g \triangleq \frac{1}{1 + |\nabla G_{\sigma} \cdot I|^2}, \quad (16)$$

Then, we define the dual-fusion intensity fitting energy functional as follows:

$$E^{\text{DFIF}}(\phi) = g[\omega_1 E_{\text{org}} + (1 - \omega_1) E_{\text{sal}}], \quad (17)$$

where ω_1 is a weight coefficient ($0 \leq \omega_1 \leq 1$), and E_{org} and E_{sal} are the original images' fitting energy functional and the saliency images' fitting energy functional, respectively.

Finally, the dual-fusion intensity fitting energy functional $E^{\text{DFIF}}(\phi)$ can be obtained by combining Equations (14)–(17):

$$E^{\text{DFIF}}(\phi, I_1, I_2) = g[\omega_1 E_{\text{org}} + (1 - \omega_1) E_{\text{sal}}] = g\left[\omega_1 \frac{1}{2} \int_{\Omega} |I(x) - I^{\text{LFI}}(x)|^2 dx + (1 - \omega_1) \left(\lambda_1 \int_{\Omega} (S - I_1(x))^2 H_{\epsilon}(\phi(x)) dx + \lambda_2 \int_{\Omega} (S - I_2(x))^2 (1 - H_{\epsilon}(\phi(x))) dx \right)\right] \quad (18)$$

Then, we minimize $E^{\text{DFIF}}(\phi, I_1, I_2)$ with respect to ϕ to obtain the corresponding gradient descent flow [29–31]:

$$\frac{\partial \phi}{\partial t} = g\delta_\varepsilon(\phi)[\omega_1 e_1 + (1 - \omega_1)e_2], \quad (19)$$

where:

$$\begin{cases} e_1 = (I - m_1 H_\varepsilon(\phi(x)) - m_2(1 - H_\varepsilon(\phi(x))))(m_1 - m_2) \\ e_2 = -\lambda_1(S - I_1(x))^2 + \lambda_2(S - I_2(x))^2 \end{cases}, \quad (20)$$

where I and S are the original images and the saliency images, respectively. $I_1(x)$ represents the integrated local and global intensities, and m_1 and m_2 are averages of the image intensities in a Gaussian window inside and outside the contour.

3.4. Regularize the Level Set Function

As pointed out by Zhang's method [27], Gaussian filtering can replace the traditional regularized term to regularize the level set function. Therefore, the smoothing process of the level set function can be expressed as:

$$\phi^{k+1} = G_\eta \cdot \phi^k, \eta > \sqrt{\Delta t}, \quad (21)$$

where η is the standard deviation, and Δt is the time-step.

In fact, the smoothing effect of the level set function by Gaussian filtering is slightly worse than the traditional regularized term and is greatly affected by the time-step. However, the computing efficiency of Gaussian filtering is much higher than the traditional regularized term.

4. Results and Discussion

This section tested the proposed method on intensity-heterogeneous underwater images captured from underwater videos downloaded from the NATURE FOOTAGE website and Fish Dataset. Moreover, the method was compared with some state-of-the-art contour extraction methods in terms of its efficiency and accuracy. All contour extraction results were produced on the same computer to ensure fairness. The computer was configured as Intel(R) Core(TM) i7-8650U CPU @ 2.11 GHz, 16.00 GB memory, Windows 10 system, and x64 processor. MatlabR2017a was the software platform. We used the same parameters $\eta^2 = 6, \sigma = 2, \varepsilon = 1, \lambda_1 = 3, \lambda_2 = 1$ and time-step $\Delta t = 0.1$. The initial level set function was defined by:

$$\phi(x, t = 0) = \begin{cases} -c_0, & x \in \text{in}(C) \\ 0, & x \in C \\ c_0, & x \in \text{out}(C) \end{cases}, \quad (22)$$

where $c_0 > 0$ is a constant (in our experiments, $c_0 = 1$) and $\text{in}(C)$ and $\text{out}(C)$ represent the region inside and outside of the contour C , respectively.

Moreover, the parameter ω_1 is a constant, which controls the influence of the saliency image fitting energy and original image fitting energy. When the missing information of the saliency target contour is severe, ω_1 should be relatively larger; otherwise, ω_1 should be a small value. Moreover, ω should be smaller when the intensity inhomogeneity of the saliency image is severe. This is because the local intensity fitting can better segment the target in the intensity inhomogeneity region, and the results of contour extraction rely on the local intensity fitting. Otherwise, ω should be larger to suppress the noise interference. In the experiment, we need to choose appropriate values for ω and ω_1 according to the degree of inhomogeneity and the degree of saliency detection deviation. In the experiment of this paper, the value of ω was from 0.5 to 0.9, and the value of ω_1 was from 0.1 to 0.8.

4.1. The Benefits of Local-Global Intensity Fitting

A comparative experiment was performed to prove the effectiveness of the local-global intensity fusion in Section 3.1. We conducted different experiments, as shown in Table 2.

Table 2. The comparative experiment of the local-global intensity.

Experiments	local Intensity	Global Intensity
A	✓	
B		✓
C (our fusion intensity)	✓	✓

In experiment A, the fitting intensity of the energy functional is the local intensity. In experiment B, the fitting intensity of the energy functional is the global intensity. Moreover, the energy functional with the fusion local-global intensity is shown in Experiment C. The contour extraction results of the experiments are shown in Figure 2.

Figure 2. The contour extraction results. (a) Original saliency image. (b) The result of the local intensity fitting. (c) The result of the global intensity fitting. (d) The result of our method.

As shown in Figure 2, experiment A extracted the target contour in the intensity inhomogeneity region, but the result was greatly affected by the initial contour curve (blue circled area) and was sensitive to noise (green circled area). Moreover, the method of experiment A also extracted the contours of the non-boundary regions. Experiment B could extract the target contour in the intensity homogeneity region and was not disturbed by noise, but the target contour in the intensity inhomogeneity region could not be extracted. By comparing the segmentation results in Figure 2a–c, it can be seen that the fused energy functional (experiment C) can not only effectively eliminate the influence of the initial curve and noise interference but also effectively segment intensity-inhomogeneous regions.

4.2. The Effect of Original Image Correction

Figure 3 shows the result of our method in the underwater image segmentation. As shown in Figure 3, the coordinate points $\{[X, Y] : [77, 41]\}$ and $\{[X, Y] : [152, 57]\}$ are located at the saliency target edge in Figure 3b. However, in Figure 3c, the coordinate point at the same position is inside the target instead of on the target edge. This error is caused by the deviation in the saliency detection. Therefore, it is necessary to use the original image to supplement the missing information. This paper used the local image fitting method to extract the contour information of the original image. It then used the contour information to correct the deviation caused by saliency detection. The result of the correction is shown in Figure 3e. As shown in Figure 3e, the missing contour information of the saliency image is accurately supplemented, and the background information is filtered out.

Figure 3. The results of our method. (a) The original underwater image with an initial zero level contour. (b) the contour extraction result of the saliency target. (c) The contour extraction result without correction. (d) The final level set function. (e) The result of our method.

4.3. Performance of the Dual-Fusion Active Contour Model

Figure 4 shows the performance of our method. It can be seen from Figure 4d that our method can filter out the background information and accurately extract the target contour. Figure 4b shows the saliency images of the original underwater images. The red circles represent the intensity inhomogeneity region, the yellow circles represent the noise region, and the green circles represent the missing region of the target. For the regions of the intensity inhomogeneity and noise, our method can still extract the target contour well using the local-global intensity fitting term. Moreover, the saliency image of the first image obviously lacks part of the target information (green circle region). Our method can still extract the complete target contour by integrating the original image contour information.

Figure 4. The salient object segmentation results of the proposed model. (a) Original underwater images with an initial zero level contour. (b) Saliency images (c) Final level set function. (d) The results of our method.

4.4. Qualitative Comparison

4.4.1. Comparison of the Segmentation Results with Other Models

To verify the effectiveness of the proposed method, we compared the segmentation results with other classic models, such as LBF [29,30], LGIF [31], LIF [27], and RMPCM [3], respectively. The comparison results are shown in Figure 5.

Figure 5. Comparison of our method with LBF, LGIF, LIF, RMPCM. (a) Results of the LBF model. (b) Results of the LGIF model. (c) Results of the LIF model. (d) Results of the RMPCM model. (e) Results of our method.

It can be seen from Figure 5 that the LBF model is limited by the initial contour curve and cannot completely extract the target contour. The LGIF model is minimally affected by local background noise due to the fusion of the global intensity fitting, but it still cannot accurately extract the target contour. The LIF and RMPCM models can completely extract the target contour, but they are greatly affected by background noise and local target features. Our model introduces semantic information to filter out background noise very well. Moreover, because of the global-local intensity fitting, our method can handle local inhomogeneous regions without interfering in the local target features. In addition, the target contour of the original image perfectly complements the missing semantic information.

Furthermore, we compared the segmentation results with Zhu's method [15] and Chen's method [20], which also introduced saliency images as semantic information. Since we could not obtain the source codes of Zhu's method [15] and Chen's method [20], to ensure the fairness of the comparison results, we adapted the segmentation images from

Zhu's method [15] (2017,IEEE) and Chen's method [20] (2019,IEEE) as the comparison images. The comparison results are shown in Figure 6.

Figure 6. Comparison of our method with Zhu's method and Chen's method [20]. (a) Comparison of our method with Zhu's method. (b) Comparison of our method with Chen's method. The first row is the original underwater images; the second row is the segmentation results in Zhu's method [15] and Chen's method [20]. The third row shows the results of our method.

As can be seen in Figure 6, even though our method, Zhu's method [15], and Chen's method [20] all introduce semantic information, our method can extract the target contour more accurately than Zhu's method [15] and Chen's method [20]. As shown in the blue circle region of Figure 6a, our method extracted the target contour in the detail region more accurately. This is because we added the local-global fitting term to better extract the contours of local inhomogeneous regions, and the original image correction module can correct the errors in semantic information. As shown in the green circle region of Figure 6b, our method can filter out background noise better than Chen's method [20] and is more robust.

4.4.2. Comparison of the Saliency Segmentation Results with Other Models

To further verify the superiority of the proposed method, we also compared the contour extraction results of the underwater image with the saliency image as the input of several classic models. To test the robustness of the proposed method, we only selected low-quality saliency images (inhomogeneous local intensity and incomplete saliency information) for the comparison experiments. As shown in Figure 7, the segmentation results of LBF are severely affected by the initial contour curve and are disturbed by the inhomogeneous regions inside the target. The LGIF model can avoid the influence of the initial contour curve but cannot extract complete contour information, as shown in the green dotted region in Figure 7(2). The LIF model can extract the target contour relatively completely, but it easily falls into the local optimum and is also affected by the initial contour curve. The RMPAM model avoids the local optimum error, but it also has the problem that the contour information cannot be extracted completely, as shown in the green dotted region in Figure 7(2). Our method can effectively avoid the local optimum and supplement the missing contour information through the original image. Hence, the results of our method are more accurate and complete than other methods.

Figure 7. Comparison of our method with LBF, LGIF, LIF, RMPCM. (a–e) Results of LBF, LGIF, LIF, RMPCM, and our method. The upper rows of (1), (2), and (3) are the segmentation results of the saliency images, and the lower rows are the segmentation results of the corresponding original images.

4.5. Quantitative Comparison

In the following experiment, we compare the proposed method with the aforementioned methods using several evaluation indexes to conduct a quantitative analysis. Here, three evaluation indicators, namely the mean absolute error (MAE), the error rate (ER), and the detection rate (DR), are employed for quantitative comparison. The MAE, ER, and DR can be expressed by the following equations:

$$\text{MAE} = \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^n |Det_{(x,y)} - gt_{(x,y)}|, \quad (23)$$

$$\text{ER} = \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^n |Det_{(x,y)} - gt_{(x,y)}| / \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^m Det_{(x,y)} \cdot gt_{(x,y)}, \quad (24)$$

$$\text{DR} = \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^m Det_{(x,y)} \cdot gt_{(x,y)} / \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^m (Det_{(x,y)} + gt_{(x,y)}), \quad (25)$$

where m and n represent the length and width of the image, Det is the result of image segmentation, and gt is the hand-crafted ground truth. So, $Det_{(x,y)} \cdot gt_{(x,y)}$ represents the contour that is accurately extracted by the model and $Det_{(x,y)} + gt_{(x,y)}$ represents the union of the image segmentation result and the ground truth. The larger the $Det_{(x,y)} \cdot gt_{(x,y)}$, the more contour that is correctly extracted. $Det_{(x,y)} - gt_{(x,y)}$ represents the pixels that are incorrectly extracted, so the larger $Det_{(x,y)} - gt_{(x,y)}$ is, the more pixels are incorrectly extracted. So, the smaller the value of ER, the more accurate the result of contour extraction. The smaller the value of MAE, the more accurate the contour extraction result. Moreover, a large value of DR can indicate that the contour extraction result of the model is accurate. The evaluation results of the aforementioned 5 methods are shown in Tables 3–5.

Table 3. The MAE results of LBF, LGIF, LIF, RMPCM, and our method in Figure 5.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
LBF	9.5723	3.6588	3.0737	12.8131	2.1789	4.4882	6.9886	5.3132
LGIF	7.9119	3.7481	3.4620	9.6148	2.2546	4.1299	7.4210	6.3514
LIF	6.0811	4.0945	2.5343	7.4206	2.9782	4.2036	10.2057	4.9874
RMPCM	10.5081	4.3048	5.1594	12.9181	2.3604	6.0406	7.9683	3.9875
Our	2.3695	3.2604	1.9161	4.5302	1.2455	2.7715	5.9417	2.9702

Table 4. The ER results of LBF, LGIF, LIF, RMPCM, and our method in Figure 5.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
LBF	0.7434	0.5195	0.0510	0.7863	0.2585	0.2769	0.3403	0.2386
LGIF	1.0012	0.5978	0.1414	0.7304	0.1237	0.2283	0.2092	0.3109
LIF	0.8355	0.4713	0.0761	0.3578	0.1209	0.2938	0.5013	0.1483
RMPCM	1.1478	0.3300	0.0962	1.0167	0.1167	0.2137	0.3200	0.9286
Our	0.2709	0.2649	0.0452	0.3411	0.0776	0.2007	0.2060	0.0571

Table 5. The DR results of LBF, LGIF, LIF, RMPCM, and our method in Figure 5.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
LBF	1.3148	1.7739	13.9989	1.2499	3.5065	3.4029	2.7005	3.9709
LGIF	0.9778	1.5494	6.0806	1.3423	6.7686	4.0834	4.1945	3.0709
LIF	1.1707	1.9462	10.1875	2.6965	6.9244	3.2240	1.8872	6.1722
RMPCM	0.8569	2.7046	8.6031	0.9690	7.0274	4.2727	2.8823	1.0500
Our	3.4725	3.3199	15.3967	2.7586	10.3096	4.6516	4.3053	14.0682

A smaller value of MAE represents a higher contour extraction accuracy. According to Table 3, the contour extracted by the proposed model obtained the smallest MAE value, which shows that the proposed model can extract target contours more accurately than the other 4 models. Table 4 shows the error rate (ER) of the 5 methods. The ER values between the target contour extracted by the proposed method and ground truth are the smallest, so the proposed method has the highest accuracy. Table 5 shows the detection rates of the above five methods. The detection rate represents how many contour pixels are correctly extracted. Therefore, our model with the highest detection rate can extract the target contour more accurately.

5. Conclusions

Aiming to resolve the problem that conventional active contour models cannot effectively extract the contours of the salient object in underwater images, we proposed a dual-fusion active contour model with semantic information. The proposed method extracted the saliency target contour by fusing the local and global intensity and extracted the original image contour information by the local image fitting model to correct the saliency information deviation. We verified the superiority of the dual-fusion active contour model with semantic information by qualitative comparison and quantitative comparison. The qualitative comparison results show that the proposed model can effectively suppress the interference of noise and can more accurately extract the contour of the intensity inhomogeneity region. Moreover, we also verified that the missing saliency target contour can be effectively corrected by the contour information of the original image. The quantitative comparison results show that the salient object contour extraction results of the proposed model achieve the highest accuracy and lowest error. Therefore, the proposed model can effectively solve the problem that conventional active contour models cannot extract the salient object contour due to the lack of semantic information and provides support for underwater target tracking, underwater image restoration, and other technologies.

Author Contributions: Conceptualization, methodology and software, S.Y.; writing—review and editing, J.W.; supervision, funding acquisition, Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R&D Program of China (2018YFC0810500).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xiao, Z.L.; Zhang, M.; Chen, L.S.; Jin, H.Y. Detection and segmentation of underwater CW-like signals in spectrum image under strong noise background. *J. Vis. Commun. Image Represent.* **2019**, *60*, 287–294. [\[CrossRef\]](#)
2. Hou, G.; Li, J.; Wang, G.; Yang, H.; Huang, B.; Pan, Z. A novel dark channel prior guided variational framework for underwater image restoration. *J. Vis. Commun. Image Represent.* **2020**, *66*, 102732. [\[CrossRef\]](#)
3. Wu, Y.F.; Li, M.; Zhang, Q.F.; Liu, Y. A retinex modulated piecewise constant variational model for image segmentation and bias correction. *Appl. Math. Model.* **2018**, *54*, 697–709. [\[CrossRef\]](#)
4. Abas, P.E.; De Silva, L.C. Review of underwater image restoration algorithms. *IET Image Processing* **2019**, *13*, 1587–1596.
5. Wang, R.; Wang, Y.; Zhang, J.; Fu, X. Review on underwater image restoration and enhancement algorithms. In Proceedings of the 7th International Conference on Internet Multimedia Computing and Service, Zhangjiajie, China, 19–21 August 2015; pp. 1–6.
6. Schettini, R.; Corchs, S. Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 746052. [\[CrossRef\]](#)
7. Liu, Y.; Li, H. Design of Refined Segmentation Model for Underwater Images. In Proceedings of the 5th International Conference on Communication, Image and Signal Processing (CCISP), Chengdu, China, 13–15 November 2020; pp. 282–287. [\[CrossRef\]](#)
8. Chen, W.; He, C.Y.; Ji, C.L.; Zhang, M.Y.; Chen, S.Y. An improved K-means algorithm for underwater image background segmentation. *Multimed. Tools Appl.* **2021**, *80*, 21059–21083. [\[CrossRef\]](#)
9. SM, A.R.; Jose, C.; Supriya, M.H. Hardware realization of canny edge detection algorithm for underwater image segmentation using field programmable gate arrays. *J. Eng. Sci. Technol.* **2017**, *12*, 2536–2550.
10. Sun, Y.T.; Luan, X.L. An Underwater Optical Image Segmentation Algorithm Based on Fuzzy C-means Model. *J. Phys. Conf. Ser.* **2018**, *1087*, 052007. [\[CrossRef\]](#)
11. Li, X.; Song, J.D.; Fan, Z.; Ouyang, X.G.; Khan, S.U. Map Reduce-based fast fuzzy c-means algorithm for large-scale underwater image segmentation. *Future Gener. Comput. Syst.* **2016**, *65*, 90–101. [\[CrossRef\]](#)
12. Rajeev, A.A.; Hiranwa, S.; Sharma, V.K. Improved Segmentation Technique for Underwater Images Based on K-means and Local Adaptive Thresholding. In *Information and Communication Technology for Sustainable Development*; Springer: Singapore, 2018; pp. 443–450.
13. Chen, Z.; Zhang, Z.; Bu, Y.; Dai, F.; Fan, T.; Wang, H. Underwater object segmentation based on optical features. *Sensors* **2018**, *18*, 196. [\[CrossRef\]](#)
14. Xuan, L.; Zhang, M.J. Underwater color image segmentation method via RGB channel fusion. *Opt. Eng.* **2017**, *56*, 023101. [\[CrossRef\]](#)
15. Zhu, Y.; Hao, B.; Jiang, B.; Nian, R.; He, B.; Ren, X.; Lendasse, A. Underwater image segmentation with co-saliency detection and local statistical active contour model. In Proceedings of the OCEANS 2017-Aberdeen, Aberdeen, UK, 19–22 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5.
16. Qiao, X.; Bao, J.H.; Zeng, L.H.; Zhou, J.; Li, D.L. An automatic active contour method for sea cucumber segmentation in natural underwater environments. *Comput. Electron. Agric.* **2017**, *135*, 134–142. [\[CrossRef\]](#)
17. Li, Y.J.; Xu, H.L.; Li, Y.; Lu, H.M.; Serikawa, S. Underwater image segmentation based on fast level set method. *Int. J. Comput. Sci. Eng.* **2019**, *19*, 562–569. [\[CrossRef\]](#)
18. Bai, J.S.; Pang, Y.J.; Zhang, Q.; Zhang, Y.H. Underwater Image Segmentation Methods Based on MCA and Adaptive Level Set Evolution. In Proceedings of the 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), Beijing, China, 8–10 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 734–738.
19. Li, S.L.; Mengxing, H. Research of Underwater Image Segmentation Algorithm Based on the Improved Geometric Active Contour Models. In Proceedings of the 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), Shanghai, China, 3–7 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 44–50.
20. Chen, Z.; Sun, Y.; Gu, Y.; Wang, H.; Qian, H.; Zheng, H. Underwater object segmentation integrating transmission and saliency features. *IEEE Access* **2019**, *7*, 72420–72430. [\[CrossRef\]](#)
21. Antonelli, L.; De Simone, V.; di Serafino, D. Spatially adaptive regularization in image segmentation. *Algorithms* **2020**, *13*, 226. [\[CrossRef\]](#)

22. Houhou, N.; Thiran, P.J.; Bresson, X. Fast texture segmentation based on semi-local region descriptor and active contour. *Numer. Math. Theory Methods Appl.* **2009**, *2*, 445–468. [[CrossRef](#)]
23. O’Byrne, M.; Pakrashi, V.; Schoefs, F.; Ghosh, B. Semantic Segmentation of Underwater Imagery Using Deep Networks Trained on Synthetic Imagery. *J. Mar. Sci. Eng.* **2018**, *6*, 93. [[CrossRef](#)]
24. Zhou, Y.; Wang, J.; Li, B.; Meng, Q.; Rocco, E.; Saiani, A. Underwater scene segmentation by deep neural network. In Proceedings of the 2nd UK Robotics and Autonomous Systems Conference, (UK-RAS 2019), Loughborough University, Loughborough, UK, 24 January 2019.
25. Chan, T.; Vese, L. Active contours without edges. *IEEE Trans. Image Process.* **2001**, *10*, 266–277. [[CrossRef](#)]
26. Mumford, D.; Shah, J. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **1989**, *42*, 577–685. [[CrossRef](#)]
27. Zhang, K.H.; Song, H.H.; Zhang, L. Active contours driven by local image fitting energy. *Pattern Recognit.* **2010**, *43*, 1199–1206. [[CrossRef](#)]
28. Zhao, T.; Wu, X. Pyramid Feature Attention Network for Saliency Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3085–3094.
29. Li, C.M.; Kao, C.Y.; Gore, J.C.; Ding, Z.H. Implicit active contour driven by local binary fitting energy. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; Volume 1, pp. 1–7.
30. Li, C.M.; Kao, C.Y.; Gore, J.C.; Ding, Z.H. Minimization of region-scalable fitting energy for image segmentation. *IEEE Trans. Image Process.* **2008**, *17*, 1940–1949.
31. Wang, L.; Li, C.M.; Sun, Q.S.; Xia, D.S.; Kao, C.Y. Active contours driven by local and global intensity fitting energy with application to brain MR image segmentation. *Comput. Med. Imaging Graph.* **2009**, *33*, 520–531. [[CrossRef](#)]
32. Li, C.M.; Xu, C.Y.; Gui, C.F.; Martin, D. Fox Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **2010**, *19*, 3243–3254.
33. Li, C.M.; Xu, C.Y.; Gui, C.F.; Martin, D. Fox Level set evolution without re-initialization: A new variational formulation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 430–436.