mathematics

Review

Review of GrabCut in Image Processing

Zhaobin Wang [1,*], Yongke Lv [1], Runliang Wu [1] and Yaonan Zhang [2]

[1] School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
[2] The National Cryosphere Desert Data Center, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China
* Correspondence: wangzhb@lzu.edu.cn; Tel.: +86-931-8912778

Abstract: As an image-segmentation method based on graph theory, GrabCut has attracted more and more researchers to pay attention to this new method because of its advantages of simple operation and excellent segmentation. In order to clarify the research status of GrabCut, we begin with the original GrabCut model, review the improved algorithms that are new or important based on GrabCut in recent years, and classify them in terms of pre-processing based on superpixel, saliency map, energy function modification, non-interactive improvement and some other improved algorithms. The application status of GrabCut in various fields is also reviewed. We also experiment with some classical improved algorithms, including GrabCut, LazySnapping, OneCut, Saliency Cuts, DenseCut and Deep GrabCut, and objectively analyze the experimental results using five evaluation indicators to verify the performance of GrabCut. Finally, some existing problems are pointed out and we also propose some future work.

Keywords: GrabCut; energy function; interactive image segmentation; graph theory

MSC: 68U10

Citation: Wang, Z.; Lv, Y.; Wu, R.; Zhang, Y. Review of GrabCut in Image Processing. *Mathematics* **2023**, *11*, 1965. https://doi.org/10.3390/math11081965

Academic Editor: Konstantin Kozlov

Received: 28 February 2023
Revised: 16 April 2023
Accepted: 18 April 2023
Published: 21 April 2023

1. Introduction

Image segmentation is the basic technology of image processing. In most applications, there will be no correct image analysis results without correct image segmentation; that is, the accuracy and efficiency of segmentation directly affect the subsequent processing results. Therefore, it is one of the hot research directions in image processing and computer vision.

Although a variety of segmentation algorithms have been proposed, there is no general method that works well for any type or any target image. There are six commonly used image-segmentation techniques [1]: level set, threshold-based segmentation, edge-based segmentation, region-based segmentation, energy functional-based segmentation and graph-based segmentation. GrabCut is a segmentation method based on graph theory.

In 2004, GrabCut was proposed by Rother et al. based on Graph Cuts [2]. GrabCut can perform image segmentation with little and simple user interaction. The user only selects the foreground and background with a rectangular region. After obtaining the color space of the foreground and background through this incomplete labeling method, the Gaussian mixture model (GMM) [3] is established to obtain the regional terms. The boundary term is obtained using the Euclidean distance between neighborhood pixels. The energy function is constructed using the regional term and the boundary term. Finally, the GMM parameter replaces the minimum estimate in Graph Cuts to achieve energy minimization.

With the development of GrabCut in the past ten years, researchers have proposed many improved GrabCut algorithms such as GrabCut based on superpixels, GrabCut based on saliency detection, constrained Markov random field (MRF) models based on modified energy functions, etc. Various improved algorithms have achieved remarkable results in different aspects. Meanwhile, because of its superiority, graph theory has been widely applied in many fields (e.g., medical image analysis, agriculture and animal husbandry, etc.)

with good performance. We searched on the Web of Science with the keyword "GrabCut". Figure 1 shows a significant growing trend for GrabCut from 2004 to 2022. Therefore, it is necessary to review the research progress and status of GrabCut.

The main contributions of this paper are as follows. First of all, the relevant important references published in the past few years are sorted out carefully. For improved GrabCut models, we mainly focus on GrabCut based on superpixel, GrabCut based on salient object segmentation, GrabCut based on modified energy function and non-interactive GrabCut. For its applications, we focus on its application in medical imaging and also summarize its application in other fields, e.g., object detection and recognition, video processing, agriculture and animal husbandry, etc. To illustrate the performance of GrabCut, we also conduct comparative experiments with existing typical methods.

The rest of the paper is organized as follows. Section 2 reviews the models of Graph Cuts and GrabCut. Section 3 reviews and summarizes the improved GrabCut algorithms that have been relatively novel or important in recent years. Section 4 classifies and summarizes the application of GrabCut. In Section 5, some typical algorithms are employed to compare their performance, some of the problems are pointed out and future work on GrabCut in discussed. In Section 6, we summarize this paper. In Section 7, we propose the future work and challenges.

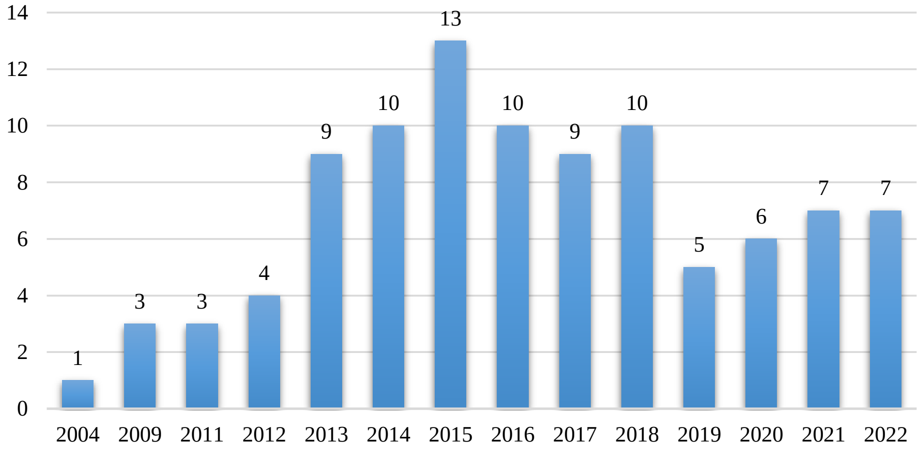


Figure 1. Number of relevant papers published from 2004 to 2022.

2. GrabCut Model

The GrabCut model is an algorithm based on graph theory. After the weighted undirected graph is obtained, the GMM is used to obtain the regional term through the points and edges of the graph, and the Euclidean distance between the pixel pairs is calculated to obtain the boundary term. The sum of the regional term and the boundary term is the energy function, which is a form of Markov random field (MRF) [4]. The optimization of the energy function is actually the optimization of Gibbs energy [5]. GrabCut is an improvement of Graph Cuts based on MRF theory and the maximum flow minimum cut algorithm. The following section introduces the GrabCut model in detail.

GrabCut is an improved algorithm for Graph Cuts through iterative methods, mainly for color images [2]. Figure 2 shows two examples of GrabCut segmentation. Input an image with n pixels, and let the set of all pixels on the image be P. A set of all unordered pairs $\{p, q\}(p \in P, q \in P)$ of neighboring elements in P in the neighborhood system is denoted as N. For example, N may contain the neighboring pixels of all unordered pairs under a standard 8- (or 26-) neighborhood system. Let the binary vector $A = \{A_1, A_2, A_3, \cdots, A_p, \cdots, A_n\}$, where A_p is the assignment of pixel p in set P, with 0 for the background and 1 for the foreground. As shown in Figure 3, an undirected graph $\varsigma = \langle v, \varepsilon \rangle$ is created from the input image. The node v of the graph corresponds to the pixel $p \in P$ of the image, and there are two additional nodes: the foreground terminal (a source S) and the background terminal (a sink T). Their relationship is as follows.

$$v = P \bigcup \{S, T\}. \tag{1}$$

Then, the soft constraints on the boundary and regional properties of the graph are described by the energy function $E(A, k, \theta, P)$,

$$E(A, k, \theta, P) = U(A, k, \theta, P) + V(A, P), \tag{2}$$

where U is a regional term, indicating a penalty that a pixel belongs to the foreground or the background (the negative logarithm of the probability that the pixel belongs to the foreground or the background). V is a boundary term indicating a penalty term between two neighboring pixels p and q. As shown in Figure 3, for the cutting process of GrabCut, the actual meaning of the energy function can be visually seen. The red edge and the blue edge represent the regional term U, and the thickness of the edge represents the weight of the edge, that is, the value of U. The yellow edge represents the boundary term V. Similarly, the thickness of the edge represents the weight of the edge, that is, the value of V. The red, yellow and blue edges are cut using energy minimization (the green dotted line is cut).

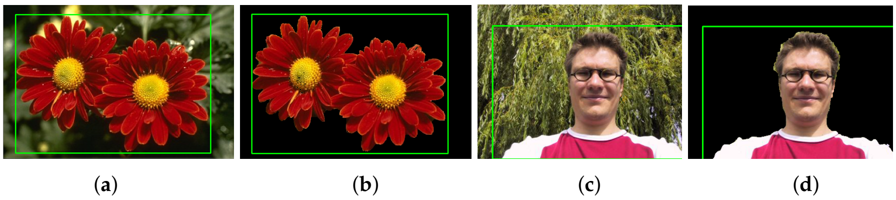


Figure 2. Two examples of GrabCut. (**a**) Original image, (**b**) Segmentation result, (**c**) Original image, (**d**) Segmentation result.

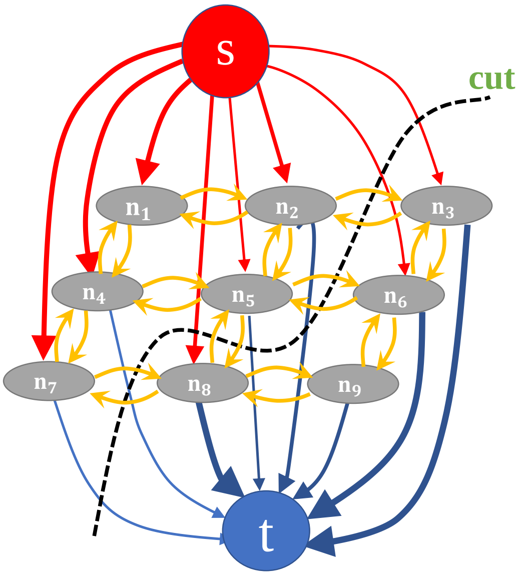


Figure 3. *S*-*T* diagram of the GrabCut segmentation process.

Among them, the energy expression of the regional term U is as shown in Equations (3)–(5), where π represents the weight of each Gaussian component, μ represents the mean vector of each Gaussian component, $\sum$ represents the covariance matrix of each Gaussian component and I_p represents the pixel of point p.

$$U(A, k, \theta, P) = \sum_{p \in P} D(A_p, k_p, \theta, p), \tag{3}$$

$$D(A_p, k_p, \theta, P) = -\log \pi(A_p, k_p) + \frac{1}{2}\log\det \sum(A_p, k_p) \\ + \frac{1}{2}\big(I_p - \mu(A_p, k_p)\big)^T \sum(A_p, k_p)^{-1}\big(I_p - \mu(A_p, k_p)\big), \tag{4}$$

$$\theta = \{\pi(A, k), \mu(A, k), \sigma(A, k)\}, \;\; A = 0, 1;\; k = 1, \cdots, K. \tag{5}$$

GrabCut models the foreground and the background in the RGB space with a full covariance GMM of K Gaussian components (typically $K = 5$) [3]. This gives an extra vector $k = \{k_1, \cdots, k_p, \cdots, k_n\}$, in which $k_p (k_p \in \{1, \cdots, K\})$ is the Gaussian component corresponding to the pixel p. All pixels belong to either the foreground or the background.

The Gaussian mixture density model is as shown in Equations (6) and (7):

$$D(x) = \sum_{i=1}^{K} \pi_i g_i(x, \mu_i, \sigma_i), \text{ where } \sum_{i=1}^{K} \pi_i = 1, \;\; 0 \le \pi_i \le 1, \tag{6}$$

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{(2\pi)^d |\sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \sigma^{-1}(x-\mu)\right). \tag{7}$$

Therefore, taking the negative logarithm is the form shown in Equation (4). Each Gaussian component in the GMM has three parameters, namely the weight π, the mean vector μ and the covariance matrix σ in Equation (5) (because there are three channels of RGB, μ is a three-element vector, σ is 3 × 3 matrix). Regardless of the foreground or the background θ, the first determination of these three parameters is achieved using the K-means algorithm [6]. The K-means algorithm clusters the foreground or the background into K kinds of pixels and finds the weighted, averaged and covariance matrix for each pixel. When these three parameters of each pixel are obtained, each Gaussian component θ is obtained. When these three parameters are determined, the RGB color values of the pixels in the image can be substituted into the foreground or the background GMM. It can find the probability that each pixel belongs to the foreground and the background and find the regional term of the energy, that is, the weight of S and T to the edge of pixel p in the image.

The weight of the edge between pixels p and q, that is, the boundary term V, is shown in Equation (8).

$$V(A, P) = \gamma \sum_{\{p,q\} \in N} [A_p \ne A_q] \exp\left(-\beta \|I_p - I_q\|^2\right). \tag{8}$$

The boundary term V is a penalty for continuity between each two neighborhood pixels p and q. In the case where the two pixels in the neighborhood have a small difference, they are likely to belong to the same foreground or the same background, so the energy is large. On the other hand, in the case where the two pixels in the neighborhood differ greatly, they are likely to belong to different categories; that is, in the edge portion, the energy is small, and it is easy to be segmented. In RGB space, the Euclidean distance between two pixels is usually calculated to objectively measure the similarity. The image contrast determines the parameter β. If the image contrast is low, the difference between the two pixels is small and the result of calculating $\|I_p - I_q\|$ is small, so the larger value of β can be used to enlarge the result. Contrarily, if the contrast of the image is relatively high, then the difference between the two pixels is very large, and the result of calculating $\|I_p - I_q\|$ is very large. It is also possible to use the method of changing the β value, using a smaller value of β to narrow the result so that the boundary term V works in any situation. Here, after many experiments, the constant $\gamma = 50$ is ideal. The weight of the yellow edge in this figure can be determined using Equation (8), the object image can be obtained, and the energy can be minimized. In order to obtain the minimum value of the energy function, the iterative process is used to optimize the GMM parameters of the foreground and the background to obtain better segmentation results.

The procedures of GrabCut are given as follows.

Step 1: Input the image. The user selects the label region U' with a rectangular region to initialize the foreground. The region inside U' is all the foreground objects F', and the region outside U' is all the background region B'.

Step 2: For each pixel p, $p \in F'$ assign a label $A_p = 1$ to the pixel p. $p \in B'$; assign a label $A_p = 0$ to pixel p.

Step 3: Using the K-means clustering algorithm, the foreground object region F' and the background region B' are respectively clustered into K kinds of pixel.

Step 4: The GMMs of the foreground and the background are initialized with the two sets of labels $A_p = 0$ and $A_p = 1$, respectively (the GMM of the foreground and the background, respectively, have K Gaussian components), and the parameters (π, μ, σ) of the two GMMs are obtained.

Step 5: Substituting each pixel p in the foreground object region F' into the two obtained GMMs, the probability that the pixel belongs to the foreground object region and belongs to the background region, respectively, are obtained (the one with the highest probability is most likely to generate the pixel p, that is, the Gaussian component k_p of the pixel p). The probability takes the form of a negative logarithm to obtain the regional term F.

Step 6: The Euclidean distance (i.e., the two norms) between every two neighboring pixels in the foreground region F' is calculated and the boundary term V is obtained.

Step 7: the minimum value of energy min $E(A, k, \theta, P)$ is obtained using the maximum flow minimum cut algorithm. The calculated result is again assigned to the set of pixels $A_p = 0$ and $A_p = 1$ in the foreground object region F'.

Step 8: Repeat steps 4 through 7 until the convergence and output image.

It is worth noting that one important aspect of the GrabCut algorithm is the use of GMMs to model the foreground and background regions. The GMMs provide a probabilistic model that allows the algorithm to estimate the likelihood of a pixel belonging to the foreground or background based on its color and texture features. This probabilistic approach improves the accuracy of the segmentation compared to traditional threshold-based approaches.

Additionally, the iterative nature of the algorithm allows it to refine the segmentation mask over multiple iterations, resulting in a more accurate segmentation of the object of interest.

3. Improved GrabCut

GrabCut may suffer from a greedy problem when optimizing, which leads to falling into a local optimum instead of the global optimum. At the same time, there are some other problems, such as the time-consuming operation of the algorithm, the need to improve the accuracy of complex image segmentation, and the limitation of the applicability of interactive segmentation. Researchers have proposed different improved algorithms for different problems of GrabCut. The main improvement based on superpixel is to improve the speed of GrabCut. The improvements based on saliency are mostly aimed at reducing errors of segmentation or achieving automatic segmentation. In order to solve the problem of interactive segmentation and improve the practicability, many non-interactive GrabCut variations are proposed. The improvement of modifying the energy function is to reduce the complexity of solving the energy function or improve the segmentation accuracy. In addition to these, there are other improvements. The following will be described in detail.

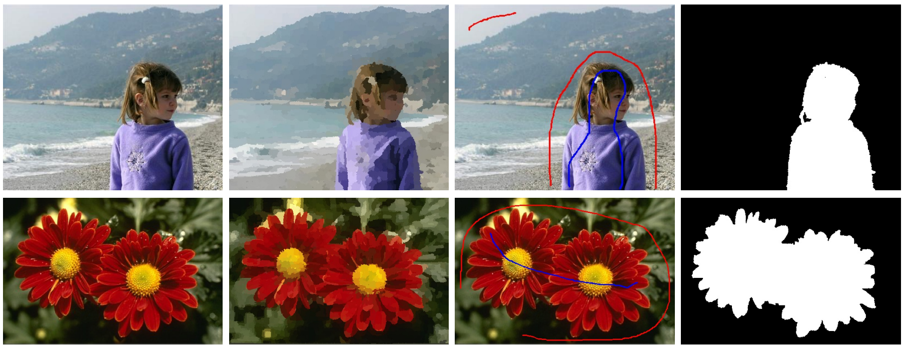
3.1. GrabCut Based on Superpixel

Superpixels divide a pixel-level map into district-level maps and extract valid information from each region, such as color histograms and texture information. The method of obtaining superpixel segmentation belongs to over-segmentation in image segmentation. Pre-segmentation using superpixel segmentation and GrabCut processing will greatly improve the efficiency of GrabCut. For example, a 400 × 300 image is divided into 400 superpixels. If we use the 400 superpixels to build the nodes of the graph, and then use GrabCut, the calculation speed will be greatly improved. So far, in the field of image processing, a variety of superpixel segmentation methods have been proposed and used for GrabCut pre-segmentation that have improved the efficiency of the algorithm, such as watershed, MeanShift, simple linear iterative clustering (SLIC), etc. [7–9].

Li et al. first proposed LazySnapping, which uses superpixels instead of pixels as GrabCut nodes [10]. The algorithm pre-segments the image with a watershed to obtain a superpixel map. The color mean of each superpixel region is then found, which will represent each superpixel. GrabCut is then used to cut the image, but only use matting (finding the foreground and background colors and the degree of fusion between them to facilitate merging the foreground onto a new background), without using the input bounding box. Because the nodes are greatly reduced, the efficiency is significantly improved and the time complexity is greatly reduced. Watershed is a mathematical morphology segmentation method based on topological theory. The gradient of the image is used as the input of the watershed segmentation, as in Equation (9), where $f(\cdot)$ is the image information and $grad(\cdot)$ is the gradient operation.

$$G(x,y) = grad(f(x,y)) = \sqrt{(f(x,y) - f(x-1,y))^2 + (f(x,y) - f(x,y-1))^2}. \quad (9)$$

Figure 4 shows two examples of the LazySnapping method. Figure 4a shows two input images. Figure 4b is the over-segmented result of watershed. Figure 4c shows the result of artificial labelling. Figure 4d shows two mask images of the segmentation results.



(**a**) (**b**) (**c**) (**d**)

Figure 4. Results of LazySnapping. (**a**) Original images, (**b**) Watershed images, (**c**) Labelled images, (**d**) Segmentation mask.

Because the watershed can effectively improve GrabCut, many scholars have used different superpixel segmentation algorithms to emulate it, and various algorithms superior to watershed improvement have been obtained. An et al. proposed an improved method for pre-segmentation using simple linear iterative clustering (SLIC) [11]. Because SLIC has a fast processing speed, the number of superpixels can be adjusted, and the size of the formed superpixels is substantially uniform, a compact superpixel map can be quickly obtained and the edge segmentation of the target is more detailed than the watershed. Because it solves the problem that the watershed segmentation region is not strong enough and the block boundary does not fit the original boundary of the object well, it becomes the main method of GrabCut pre-segmentation.

The core of the SLIC algorithm is Equations (10)–(12). The image is converted to CIELAB space to obtain the ith pixel $c_i = [l_i, a_i, b_i, x_i, y_i]^T$, where l_i, a_i and b_i are metrics on color and x_i, y_i are spatial metrics. The color distance from the point to the jth superpixel center $c_j = [l_i, a_i, b_i, x_i, y_i]^T$ is given by Equation (10). Then, the distance of the pixel space is obtained with Equation (11), and finally the distance measurement d_s with the center of the superpixel is obtained with Equation (12).

$$d_{lab} = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}, \tag{10}$$

$$d_{xy} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}, \tag{11}$$

$$d_s = d_{lab} + \frac{m}{S} d_{xy}. \tag{12}$$

Similarly, SLIC is also used for pre-processing. Ren et al. improved GrabCut by using the SLIC algorithm twice and Bayesian classifications [12]. First, this method uses SLIC to obtain the superpixel map, then performs a Bayesian classification and assigns the same pixel value to the same type of superpixel block. SLIC is performed on the classified image, then GrabCut is used and finally the boundary is optimized. The algorithm effectively integrates Bayesian classification and SLIC features, solves the segmentation degradation phenomenon when the number of superpixels is small and obtains a more robust segmentation performance.

Li et al. improved GrabCut by creating fast adaptive trimaps (FATs) after using SLIC to extract superpixels for GrabCut [13]. First, GrabCut is used on the basis of the superpixel level, and the obtained results are used to create FATs; that is, the GrabCut operation is performed again using the corrosion and expansion to obtain the unknown region. Finally, if the result does not converge, a matting processing is performed to segment more accurate results for complex images.

At present, manual monitoring of the healing process of trauma regions is very inaccurate and subjective. Silva et al. proposed a method to automatically segment ulcers in digital images [14]. Three methods of region segmentation using the superpixel strategy were evaluated from which color and texture descriptors were extracted. After the superpixel classification, the GrabCut segmentation method was applied in order to delineate the region affected by the ulcer from the rest of the image.

In order to obtain a robust segmentation under a loose bounding box, Wu et al. proposed SuperCut [15]. Instead of pre-processing with mainstream SLIC, the algorithm uses SEEDS [16] to calculate superpixels. For superpixels that pass through the bounding box, the outer pixels are treated as absolute backgrounds, and the pixels inside are considered the most likely background. The Haar-wavelet feature and pixel intensity are used to compare each pixel in the bounding box with each pixel outside the bounding box, calculate the similarity mapping factor, and design a filter. Finally, the GrabCut segmentation is performed by training the GMM model. The similarity factor is given by Equation (13). χ_F^j is the feature index of the jth foreground superpixel, χ_B^j is the feature index of the jth background superpixel and Θ is the difference of the two superpixels on the feature index χ. Even if the ratio of the foreground to the bounding box is low, which means that the bounding box contains a large background area, the method still achieves a good overlap and becomes more flexible.

$$S_\chi(F, B) = \sum_{j=1}^{j=d} \Theta\left(\chi_F^j, \chi_B^j\right). \tag{13}$$

MeanShift is also used for pre-processing. Long et al. proposed a method for pre-segmenting images using MeanShift to obtain a superpixel map and then performing subsequent processing [17]. Each of the resulting superpixel regions is represented by a color histogram, replacing the previous method using only the color mean. The improved algorithm replaces the mean value of the superpixel by using the color histogram of the superpixel, and more effectively utilizes the color information of the superpixel, thereby obtaining a more accurate segmentation result.

Combining the saliency, region growth and multi-dimensional feature based on superpixels, Zhou et al. proposed the superpixel segmentation and GrabCut-based salient object segmentation algorithm [18]. The saliency map is obtained using the minimum barrier distance transform saliency map, and the superpixel map is obtained with SLIC. Seven-dimensional features (three-channel RGB, saliency map, local binary pattern (LBP), x, y) are extracted at the superpixel level.

$$F_x^j = \frac{1}{I_c|R_S^j|} \sum_{i \in R_S^j} x_i, \tag{14}$$

$$F_y^j = \frac{1}{I_r|R_S^j|} \sum_{i \in R_S^j} y_i, \tag{15}$$

$$F_{sm}^j = \frac{1}{255|R_S^j|} \sum_{i \in R_S^j} S_i, \tag{16}$$

$$F_L^j = \frac{1}{255|R_S^j|} \sum_{i \in R_S^j} L_i, \tag{17}$$

$$F_I^j = \frac{1}{255|R_S^j|} \sum_{i \in R_S^j} I_i. \tag{18}$$

In Equations (14)–(18), $|R_S^j|$ represents the number of pixels in the jth superpixel, I_c and I_r represent the height and width of the input image, respectively, and x, y, S, L and I represent the corresponding x and y coordinates, saliency map, LBP and average of each color space channel, respectively. Therefore, the obtained F_x^j, F_y^j, F_{sm}^j, F_L^j and F_I^j are the central coordinates of x and y, the average value of the saliency map, the average value of the LBP and the average value of each color space, respectively. The purpose of this method is to maintain a high level of precision in the segmentation.

The above work enables researchers to have a clear understanding of the influence of superpixel algorithms on GrabCut. In summary, SLIC will perform better than other superpixel algorithms in general, and at the same time, the process of image segmentation is transformed from the pixel level to the superpixel level, which reduces the computational complexity, but the segmentation quality is not outstanding.

3.2. GrabCut Based on Salient Object Segmentation

Salient object segmentation refers to simulating a person's visual characteristics through an intelligent saliency-detection algorithm, detecting a salient region in the image (a region of human interest) and separating the salient regions from the background in the image. Classical saliency-detection algorithms [19] play an important role in the field of computer vision, so they have become a hot spot for many scholars.So far, a number of new and efficient saliency-detection algorithms have been proposed. It is an important direction of improving GrabCut to effectively segment the object through saliency detection and then apply it to GrabCut.

Fu et al. first proposed applying saliency detection to GrabCut, calling their method Saliency Cuts [20]. The algorithm first reduces the resolution of the input image, uses saliency detection to obtain a saliency image and binarizes the saliency image. The region where the binary image is 1 is scaled down and labeled as a foreground seed, and then expanded to form a ring area labeled as a background seed. The leftmost and rightmost pixels of the image are labelled as background seeds. After obtaining the seed, GrabCut is used to segment the object. The innovation of this algorithm is mainly to realize the automatic segmentation of GrabCut, which does not require artificially incompletely labeled seeds.

Figure 5 shows two examples of segmentation using Saliency Cuts. Figure 5a,c shows two input images. Figure 5b,d shows two mask images of the segmentation result.

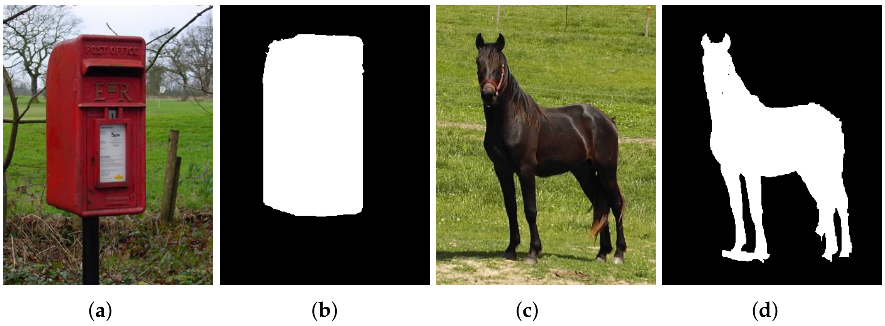


Figure 5. Results of Saliency Cuts. (**a**) Original image, (**b**) Segmentation mask, (**c**) Original image, (**d**) Segmentation mask.

Although Saliency Cuts implements automatic segmentation, which avoids the mishandling of newcomers when using GrabCut, the limitations and instability of automated segmentation remain unresolved. Kim et al. proposed an improved algorithm that also solved human errors in operation [21]. They first use saliency detection based on the superpixel level to obtain a saliency map robustly and quickly. Then, the Otsu threshold segmentation algorithm is used to merge the saliency regions outside the ROI (region of interest) selected by the user into the ROI, thereby obtaining the modified ROI. The algorithm refines the initialization information provided by the user and improves the accuracy of GrabCut.

Li et al. used an adaptive three-threshold algorithm to mark saliency images to improve Saliency Cuts [22]. Here, the salient map is divided into four kinds of seeds with three thresholds, namely, determining a foreground, determining a background, a possible foreground, and a possible background. The selection of the threshold is mainly obtained with Equations (19)–(21), where t_m is the threshold, n_t and n_b are the number of foreground and background pixels, n is the number of pixels in the entire image, i is a saliency value, n_i is the number of pixels with a saliency value of i and μ_t and μ_b are the mean saliency values of T_b and T_t, respectively, which are defined by Equation (20). Four kinds of seeds are obtained and fed to GrabCut for high-quality Saliency Cuts.

$$t_m = \arg\max \sum \omega_t \omega_b (\mu_t - \mu_b)^2, \tag{19}$$

$$\mu_k = \sum_{i \in T_k} \frac{i n_i}{n_k}, k \in \{b, t\}, \tag{20}$$

$$\begin{cases} \omega_t = \frac{n_t}{n}, n_t \in n \\ \omega_b = \frac{n_b}{n}, n_b \in n. \end{cases} \tag{21}$$

For the first time, Cheng et al. proposed a saliency model based on global region contrast (RC) and combined it with GrabCut to form a new segmentation method called SaliencyCut [23]. SaliencyCut first uses the region-based contrast (RC) algorithm to obtains a salient map, then use threshold processing to mark the seeds of the foreground, background and unknown regions and feeds it to GrabCut. Each iteration of the segmentation updates the labels using erosion and dilation. The area outside the expanded area is labeled as the background seed, the area within the corroded area is labeled as the foreground seed and the remaining areas are labeled as unknown areas. Compared with the typical saliency-detection algorithm, the saliency model proposed in this algorithm introduces spatial information and obtains a better saliency map. It is one of the better ones among the current saliency-detection algorithms. Since then, many scholars have also improved GrabCut by using RC for saliency detection.

Similarly, Gupta et al. obtained the saliency map from the saliency detection with simple priors algorithm through low-level prior information [24]. After GrabCut and edge detection, the salient text information on the natural image is extracted. For remote sensing images, Peng et al. chose the ITTI visual attention model to generate saliency maps as the initialization of GrabCut [25].

Combining superpixel, saliency and background connectivity, Niu et al. proposed a new algorithm for two-GrabCut segmentation [26]. It is different from the previous algorithm, which uses only the saliency detection as the initial marker seed. They obtained a saliency map and a background connectivity graph on a superpixel basis, and based on the results of these two graphs, the seeds were labeled using the adaptive three-threshold algorithm in ref. [22]. The seed is fed to the GrabCut at the superpixel level. The segmentation result retains the seed of the initial tag. To improve the computational efficiency, a rectangular area that contains only the foreground of the segmentation is manually cropped. New marking results are then obtained using erosion and dilation. Finally, pixel-level GrabCut is applied to refine the segmentation results.

Among them, the background connectivity uses the saliency optimization algorithm, and the main equations are Equations (22) and (23). N_s represents the number of superpixels and $\delta(\cdot) = 1$ represents the superpixel on the image boundary; otherwise, it is 0. σ_{clr} is the parameter of the Gaussian distribution, and $d_{geo}(p, q)$ is the geodesic distance of any two superpixels (p, q).

$$BndCon(p) = \frac{\sum_{i=1}^{N_s} S(p,q)\delta(q \in Bnd)}{\sqrt{\sum_{i=1}^{N_s} S(p,q)}}, \tag{22}$$

$$S(p,q) = \exp\left(-\frac{d_{geo}^2(p,q)}{\sigma_{clr}^2}\right). \tag{23}$$

In order to solve the problem that prompts are fixed in Saliency Cuts, Wang et al. proposed an adaptive Saliency Cuts [27]. This study mainly proposes a Saliency Cuts framework that can adapt to different input information, and the energy function of the framework is modified. There is no difference to Saliency Cuts when only the salient information is used as input information. For the saliency information and color as input information, the new energy function is Equation (24). For saliency information, color and depth information as input information, the new energy function is Equation (25). Among them, $E(L, K^s, \theta^s, Z^s)$ is the energy function obtained by the saliency map in GrabCut, and the boundary term is obtained using the Euclidean distance. $E(L, K^c, \theta^c, Z^c)$ is the energy function obtained by the color in GrabCut and the boundary term is obtained using the Euclidean distance. $E(L, K^d, \theta^d, Z^d)$ is the energy function obtained by the depth information in GrabCut, and the boundary term is obtained using the geodesic distance. α and β are parameters for combination.

$$E' = \alpha E(L, K^s, \theta^s, Z^s) + (1-\alpha)E(L, K^c, \theta^c, Z^c), \tag{24}$$

$$E'' = \alpha E(L, K^s, \theta^s, Z^s) + \beta E(L, K^c, \theta^c, Z^c) + (1-\alpha-\beta)E(L, K^d, \theta^d, Z^d). \tag{25}$$

In GrabCut based on salient object segmentation, the algorithm is enhanced by incorporating a saliency map that highlights the most visually distinctive parts of the image. The above algorithms can effectively solve the problem of GrabCut's interactive operation by saliency detection. The GrabCut algorithm based on salient object segmentation can produce more accurate and visually pleasing segmentations compared to the standard GrabCut algorithm, especially in images with complex backgrounds or multiple objects. However, the conversion process from the saliency map to the foreground and background seeds still requires manual intervention and is not fully automated.

3.3. GrabCut Based on Modified Energy Function

GrabCut shows excellent segmentation because it establishes GMM model for the regional term of the energy function and solves the optimization with parameter learning through iteration. However, because the energy function optimization of GrabCut is NP-hard, the disadvantages of GrabCut also appear as the complexity of the image to be processed increases. In the case where a high-resolution image needs a long time to iterate, the segmentation performance may not be ideal. Therefore, many scholars have proposed an optimized scheme for the energy function of GrabCut.

Vicente et al. proposed using a higher-order MRF to obtain a new energy function [28]. The algorithm first uses the color histogram instead of the GMM to avoid the ill-posed problem of the GMM. The energy function is reconstructed using a higher-order MRF, and the new energy function is optimized using the dual-decomposition [29] technique to achieve global optimization. Among them, the energy function formed by the higher-order MRF is given by Equations (26)–(30), where n_k^s represents the number of pixels falling into bin k and belonging to the label s.

$$E(x) = \sum_k h_k(n_k^1) + \sum_{\{p,q\}\in N} w_{pq}|A_p - A_q| + h(n^1), \tag{26}$$

$$w_{pq} = \frac{\lambda_1 + \lambda_2 exp - \beta||z_p - z_q||^2}{dist(p,q)}, \tag{27}$$

$$h_k(n_k^1) = -n_k^1 \log\left(\frac{n_k - n_k^1}{n_k^1}\right) - (n_k - n_k^1)\log\left(\frac{n_k^1}{n_k - n_k^1}\right), \tag{28}$$

$$h(n^1) = n^1 \log\left(\frac{n - n^1}{n^1}\right) - (n - n^1)\log\left(\frac{n^1}{n - n^1}\right), \tag{29}$$

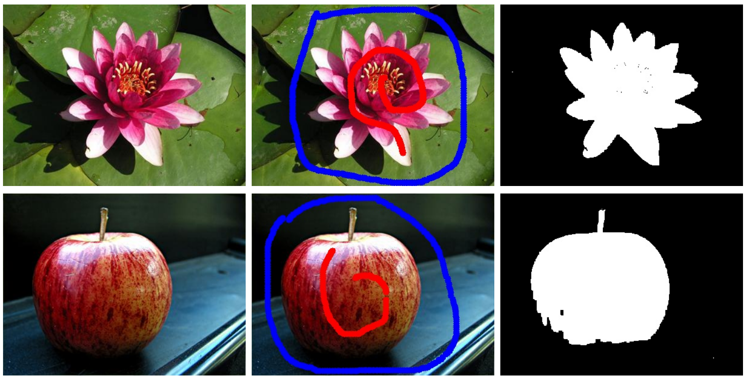
$$n_k^s = \sum_{p\in P} \delta(A_p - s). \tag{30}$$

In order to improve the optimization speed of GrabCut, Tang et al. proposed a fast global optimal binary segmentation technique, OneCut, by modifying the energy function of GrabCut [30]. OneCut uses L1-norm to measure the appearance overlap penalty, which replaces the original region of the GrabCut algorithm with L1-norm and solves the NP-hard problem. The calculation of L1-norm is given by Equation (31), where θ^S is the appearance model with label s. $E_{L_1}(\theta^s, \theta^{\bar{s}})$ is incorporated and optimized using one graph cut with Equation (32), n_k^s is the number of pixels falling into bin k and having label s, and Ω is a collection of all pixels. OneCut can completely separate the foreground and background for simple images. For an image with a complex background, although the outline of the image can be drawn more accurately locally, it is difficult for a complete image outline. Often, where the color changes suddenly, the OneCut algorithm works generally.

$$E_{L_1}(\theta^s, \theta^{\bar{s}}) = -\|\theta^s - \theta^{\bar{s}}\|_{L_1}, \tag{31}$$

$$E_{L_1}(\theta^s, \theta^{\bar{s}}) = \sum_{k=1}^{K} \min(n_k^s, n_k^{\bar{s}}) - \frac{1}{2}|\Omega|. \tag{32}$$

Figure 6 shows two examples of segmentation using OneCut. Figure 6a shows two input images. Figure 6b shows the result of artificial labelling. Figure 6c shows two mask images of the segmentation result.



(**a**) (**b**) (**c**)

Figure 6. Results of OneCut. (**a**) Original images, (**b**) Labelled images, (**c**) Segmentation mask.

Inspired by co-segmentation, Gao et al. proposed mutual GrabCut [31]. Co-segmentation refers to the segmentation of a common foreground from a similar set of images. In this model, a rectangular box with a fixed distance from the edge of the image is automatically set for the similar image. Because the foreground model with similar images is used as a constraint, the selection of the foreground does not need to be too compact. The foreground similarity constraint for similar images is added to the regional term of the energy function. A new energy function is designed by considering the similarities of the foreground and background of this group of co-segmentation images, respectively. The new regional term is given by Equation (33), where D^1 evaluates the fit of the label A_p^l to p^l and D^2 evaluates the similarity of the foreground.

$$U(A^l, k^l, \theta^{1-l}, P^l) = \begin{cases} \sum\limits_{p \in P} \Big(\lambda D^1(A_p^l, k_p^l, \theta^l, p^l) + (1-\lambda) D^2(A_p^l, k_p^{1-l}, \theta^{1-l}, p^l) \Big), A_p^l = 1 \\ D^1(A_p^l, k_p^l, \theta^l, p^l), otherwise. \end{cases} \tag{33}$$

At the same time, Zhou et al. proposed four technical components to improve the algorithm [32]. First, the algorithm combines the texture information of the input image and uses the result of the texture detection to enhance the original image with Equation (34). I_{ij} is the original image and t_{ij} is the texture image, which are added by the coefficient α. Second, the boundary term of the energy function is modified in combination with the structural tensor, where the structural tensor is given by Equation (35) and $Q(x, y)$ is defined as a 3×3 window from V in the paper. The new boundary term is given by Equation (38), where λ_+ and λ_- represent the maximum and minimum characteristics of S, respectively, and κ is a global parameter. Third, the foreground and background input from the user and the segmentation results are delivered to the adapted active contour to refine the initial segmentation. Fourth, it can refine the segmentation results again using local boundary editing. The algorithm can precisely segment various images containing textures through multiple components and produce smooth contours aligned with real boundaries.

$$v_{ij} = (\alpha I_{ij}, (1-\alpha) t_{ij}), \tag{34}$$

$$S(x, y) = \begin{pmatrix} (\frac{\partial Q(x,y)}{\partial x})^2 & \frac{\partial Q(x,y)}{\partial x} \frac{\partial Q(x,y)}{\partial y} \\ \frac{\partial Q(x,y)}{\partial x} \frac{\partial Q(x,y)}{\partial y} & (\frac{\partial Q(x,y)}{\partial y})^2 \end{pmatrix}, \tag{35}$$

$$w_s(v_{ij}, v_{kl}) = \lambda_+(i,j) \frac{\lambda_+(i,j)}{\lambda_-(i,j) + \varepsilon(\lambda_+(i,j) - \lambda_-(i,j))} (k-i, l-j) S(i,j) \begin{pmatrix} k-i \\ l-j \end{pmatrix}, \tag{36}$$

$$st(v_{ij}, v_{kl}) = \frac{\kappa}{2} (w_s(v_{ij}, v_{kl}) + w_s(v_{kl}, v_{ij})), \tag{37}$$

$$E_b(A) = \sum_{(ij,kl)\in N} (1 - \delta(A_{ij}, A_{kl})) \frac{1}{dist(v_{ij}, v_{kl})} \exp(-st(v_{ij}, v_{kl})). \quad (38)$$

In 2015, Cheng et al. proposed DenseCut [33], which uses a densely connected conditional random field (CRF) to replace the time-consuming iterative refinement of the global color model in traditional GrabCut. First, in order to achieve efficient GMM estimation, color histograms are employed to select the most frequent color portions for GMM training data samples. Then the efficient CRF inference is used to perform effective label consistency modeling and the energy function is modified with Equation (39) so that the boundary complexity of the new energy function is linear with the number of pixels. $1/Z_p$ is a normalization factor that constrains $Q(A_p)$, and l is a binary label with $l, l' \in \{0,1\}$. w is the weighting factor and θ_α, θ_β, θ_γ and θ_μ control the degree of nearness, similarity and smoothness, respectively. In the paper, the value of $w_1 = 6$, $w_2 = 10$, $w_3 = 2$, $\theta_\alpha = 20$, $\theta_\beta = 33$, $\theta_\gamma = 3$ and $\theta_\mu = 43$ according to the experience. The last term of Equation (39) is rewritten by adding and then subtracting $Q_p(l')$ to obtain Equation (45), where $\sum_{q\in P} g(p,q)Q_q(l')$ is essentially a Gaussian filter. The algorithm achieves a large increase in speed in the case of obtaining a more accurate segmentation result.

$$Q_p(A_p = l) = \frac{1}{Z_p} \exp\left(\sum_{p\neq q} g(p,q)Q_q(l') - \psi_p(A_p) \right), \quad (39)$$

$$\psi_p(A_p) = -\log P_{A_p}, \quad (40)$$

$$g(p,q) = w_1 g_1(p,q) + w_2 g_2(p,q) + w_3 g_3(p,q), \quad (41)$$

$$g_1(p,q) = \exp\left(-\frac{|p-q|^2}{\theta_\alpha^2} - \frac{|I_p - I_q|^2}{\theta_\beta^2} \right), \quad (42)$$

$$g_2(p,q) = \exp\left(-\frac{|p-q|^2}{\theta_\gamma^2} \right), \quad (43)$$

$$g_3(p,q) = \exp\left(-\frac{|I_p - I_q|^2}{\theta_\mu^2} \right), \quad (44)$$

$$\sum_{p\neq q} g(p,q)Q_q(l') = \sum_{q\in P} g(p,q)Q_q(l') - Q_p(l'). \quad (45)$$

Figure 7 shows two examples of segmentation with DenseCut. Figure 7a,c are input images. Figure 7b,d are two mask images of the segmentation result.

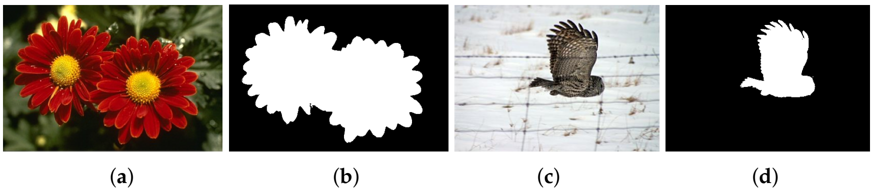


Figure 7. Results of DenseCut. (**a**) Original image, (**b**) Segmentation mask, (**c**) Original image, (**d**) Segmentation mask.

Similarly, Guan et al. proposed an improved GrabCut algorithm by changing the input cues and modifying the energy functions [34]. First, a rectangular bounding box is drawn, as in ref. [35]. Inside the bounding box is the content that needs to be processed, and the outside of the bounding box is the content that does not need to be processed. This is to reduce the image and speed up the calculation. Then the rectangle is used to select a part of the object while trying to include all the colors of the object. Then, by modifying the energy function, the regional term is modified to Equation (46), where Φ refers to the probability distribution, and the paper takes the normal distribution for calculation. The calculations of the mean μ and variance σ of the normal distribution are given by Equations (47) and (48), respectively. N_F is the number of foreground pixels and N_{PF} is the number of background pixels. Finally, the boundary of the segmentation result is drawn using edge detection and the excess boundary is deleted to obtain the final segmentation result.

$$R_p(A_p) = -\ln(\|A_p - \mu\|_\Phi), \tag{46}$$

$$\mu = \frac{1}{N_F + N_{PF}}\left(\sum_{p\in F\bigcup PF} A_p\right), \tag{47}$$

$$\sigma = \frac{1}{N_F + N_{PF}}\left(\sum_{p\in F\bigcup PF} (A_p-\mu)(A_p-\mu)^T\right). \tag{48}$$

Yong et al. used pixel values to construct a compact structure tensor that improved GrabCut [36]. The algorithm extracts texture information using nonlinear compacted structural tensor (NCST) and extracts color information using pixel values. In order to improve the simplicity and efficiency of the calculation, the mixed Gaussian model constructed using GrabCut is extended to the tensor space, and the common Riemann metric is replaced by the Kullback–Leible (KL) divergence (that is, the measurement method of GMM has changed, a new energy function is obtained, and the convergence criteria have also changed). The NCST is given by Equation (49), and the KL distance between point m and point n in the NCST is given by Equation (50). In the NCST space, the original GrabCut energy construction is updated to Equation (54) with K_T GMM components, each component having a mean M_T, a variance σ_T^2 and a weight ς_T. (A,j) represents the jth GMM component of label A, τ is a constant, β_T is an adaptive value and $|O|$ is the number of pairs of pixels. The energy function iteratively segments to satisfy the convergence of Equation (56), stopping the iteration. L_Λ represents the tensor obtained at the Λth iteration, and $N_{fg,\kappa}$ and $N_{bg,\phi}$ represent the κth component of the foreground GMM and the ϕth component of the background GMM, respectively. The algorithm realizes the non-parametric fusion of texture information and color information.

$$T_C = \begin{bmatrix} \hat{D}_{xx} & \hat{D}_{xy} \\ \hat{D}_{xy} & \hat{D}_{yy} \end{bmatrix}, \tag{49}$$

$$Z(T_C, T_M) = \sqrt{\frac{1}{4}(tr(T_P^{-1}T_M + T_M^{-1}T_P) - 4)}, \tag{50}$$

$$\sigma_T^2 = \frac{1}{|\Omega_T|}\sum_{i=1}^{|\Omega_T|} Z^2(T_i, \bar{M}_T), \tag{51}$$

$$\varphi = \begin{cases} 1, \alpha_m \neq \alpha_n \\ 0, \alpha_m = \alpha_n \end{cases}, \tag{52}$$

$$\beta_T = \left(\frac{2}{|O|}\sum_{1\le m,n\le N} Z^2(T_m, T_n)\right)^{-1}, \tag{53}$$

$$\begin{aligned} E(A) = \sum_{u\in U}\left(-lb\sum_{j=1}^{K_T}\left(\frac{\varsigma_T(A,j)}{2\pi\sigma_T^2(A,j)}\exp\left(-\frac{Z^2(T_u,M_T(A,j))}{2\sigma_T^2(A,j)}\right)\right)\right) \\ + \sum_{1\le(m,n)\le N}\varphi\left(\varsigma_T Z^{-1}(T_m,T_n)\exp\left(-\beta_T Z^2(T_m,T_n)+\tau\right)\right) \end{aligned}, \tag{54}$$

$$L_T(N_{fg,\kappa}\|N_{bg,\phi}) = \frac{1}{2}\left(lb\frac{(\sigma_{bg,\phi})^2}{(\sigma_{fg,\kappa})^2} + lb\frac{(\sigma_{fg,\kappa})^2}{(\sigma_{bg,\phi})^2} - 1\right), \tag{55}$$

$$\|L_\Lambda - L_{\Lambda-1}\|^2 \le \sigma\|L_1 - L_0\|^2. \tag{56}$$

In general, GrabCut interactions cannot select objects tightly. Yu et al. proposed an algorithm for dealing with bounding boxes loosely covering objects called LooseCut [37]. The algorithm allows the energy function to include an additional energy term to encourage consistent labelling of similar pixels with Equations (57) and (58). Among them, E_{GC} is the original energy function, and E_{LC} is the new label consistency term. For the iterative optimization of the GMM, the global similarity constraint is added with the Equations (59) and (60). Among them, μ_f^i is the mean of the ith Gaussian component of the foreground GMM, $\mu_b^{j(i)}$ is the mean of the jth Gaussian component of the background GMM and $Sim(M_f, M_b)$ is the global similarity constraint. The $Sim(M_f, M_b)$ must satisfy the constraints of $Sim(M_f, M_b) \le \delta$ during the iteration.

$$E(A,\theta) = E_{GC}(A,\theta) + \beta E_{LC}(X), \tag{57}$$

$$E_{LC}(A) = \sum_k \sum_{p\in C_k} \phi(A_p \ne A_{C_k}), \tag{58}$$

$$S(M_f^i, M_b) = \frac{1}{|\mu_f^i - \mu_b^{j(i)}|}, \tag{59}$$

$$Sim(M_f, M_b) = \sum_{i=1}^{K_f} S(M_f^i, M_b). \tag{60}$$

In order not to rely on more prior information, Long et al. proposed a method of pre-segmenting image with MeanShift and modifying the energy function [17]. Each of the resulting superpixel regions is represented by a color histogram, replacing the previous method using only the color mean. The boundary term in GrabCut replaces the Euclidean distance by using the Bhattacharyya coefficient. Finally, the results obtained are edge optimized. Among them, Equation (61) is a new energy function and Equation (63) is a boundary term. The Bhattacharyya coefficient is obtained using the color histogram of every two superpixels. H represents the color histogram and Z represents the size of the color histogram.

$$E(A) = \sum_{p\in P} R(A_p) + \sum_{(p,q)\in N} |A_p - A_q| B(A_p, A_q), \tag{61}$$

$$\begin{cases} R(A_p = 1) = \Upsilon, R(A_p = 0) = 0, \forall p \in\ F \\ R(A_p = 1) = 0, R(A_p = 0) = \Upsilon, \forall p \in\ B \\ R(A_p = 1) = \rho(p, O), R(A_p = 0) = \rho(p, B), \forall p \in U, \end{cases} \tag{62}$$

$$B(A_p, A_q) = \lambda\rho(p, q), \tag{63}$$

$$\Upsilon = 1 + \max_{p \in P} \sum_{j:\{p,q\} \in N} \rho(p,q), \tag{64}$$

$$\begin{cases} \rho(p,F) = \sum\limits_{k=1}^{Z} \sqrt{H_p(k)H^F(k)} \\ \rho(p,B) = \sum\limits_{k=1}^{Z} \sqrt{H_p(k)H^B(k)} \\ \rho(p,q) = \sum\limits_{k=1}^{N} \sqrt{H_p(k)H_q(k)}, \end{cases} \tag{65}$$

In addition, He et al. proposed a unified GrabCut model that combines feature extraction with optimized segmentation and multi-scale decomposition [38,39]. The model consists of two parts, smoothing and segmentation, which complement each other. Segmentation relies on a smooth multi-scale appearance. It uses the total variation (the image is iteratively smoothed, but the edge information remains) to maintain the geometry of the foreground and achieves a smooth effect for segmentation. Combining multi-scale edges and appearance, a new Gibbs energy function (seen Equation (66)) is proposed for segmentation, where p' is the pixel of the smoothed image.

$$E(u) = \frac{\lambda}{2} \int_P (I_{p'} - I_p)^2 dP + \int_P |\nabla I_{p'}| dP. \tag{66}$$

Most improved GrabCut algorithms improve its performance using optimizing functions. To some extent, the accuracy is improved, but it makes improved GrabCut more complex and requires more computation. Some other algorithms do not effectively change the Gibbs energy function. In addition, some algorithms easily fall into a local optimum.

3.4. Non-Interactive GrabCut

Although GrabCut has excellent segmentation ability, the application of this algorithm is narrowed because of the requirement of artificial interaction. Therefore, there is no way to meet the requirements of some fully automatic applications. Therefore, the improvement of GrabCut and the introduction of high-quality non-interactive GrabCut have become a key issue in the research on this algorithm.

Among them, Fu et al. first proposed the application of saliency detection to GrabCut, which realized the automatic segmentation of non-manually assigned tags [20]. Similarly, studies [22,23] also achieved the automatic segmentation effect of GrabCut through the saliency-detection algorithm.

Fu et al. implemented an improved algorithm using a pre-trained Deep Convolutional Neural Network (DCNN) combined with GrabCut [40]. Through a lot of training, the object type on the image is automatically recognized, and the recognition result of DCNN is used as the object of GrabCut. The trimap is initialized using the selective search method and the DCNN, and finally, GrabCut segmentation is performed. The algorithm has difficulty achieving ideal segmentation for images of multiple objects, but deep learning may solve this problem. Halil et al. used the advanced Yolov3 model [41] instead of DCNN [42]. The advantage is that the deep model has a strong learning ability and can recognize multiple objects on the image. The author applies this method to dermoscopic images.

In addition, Zhang et al. proposed an improved GrabCut algorithm based on a probabilistic neural network (PNN) [43]. The algorithm replaces the Gaussian mixture model in the GrabCut algorithm with a PNN model to calculate the weight of t-links to improve the calculation efficiency of the algorithm. The results show that the segmentation accuracy of the PNN GrabCut algorithm can improve the problems of under-segmentation and over-segmentation.

Kim et al. implemented an adaptive region of interest selection algorithm using the depth image extraction GrabCut mask [44]. The algorithm uses GrabCut to segment the depth image to obtain the depth segmentation mask. It then performs morphological operations on the depth segmentation mask to zoom in and out. The area outside the zoom in is used as the background, and the area inside the zoom out is used as the foreground. Finally, the three-channel GMM is converted to a four-channel GMM, that is, the conversion from RGB channels to RGB-D channels with depth information. The algorithm adaptively selects the region of interest, effectively suppressing the error detection of the foreground. However, to process the depth image, manual interaction process is actually needed, and real non-interaction is not realized.

Sanguesa et al. implemented an improved algorithm for the initial segmentation of the foreground using four different colorimetric methods [45]. In the literature, four types of color differences are used for segmentation. These color differences are the intensity difference, Euclidean difference, color distortion and CIEDE2000 (a uniform measure of color difference). The foreground is then initially extracted using threshold segmentation, and finally GrabCut segmentation is performed based on the initial extraction. All four methods have obtained a good automatic segmentation effect, but have not been experimentally analyzed on images affected by natural light.

The calculation of the intensity difference is given by Equation (67), the images are converted into grayscale, and then, one image is subtracted from the other using an absolute value. The calculation of the Euclidean difference is given by Equation (68), which is computed in the RGB colorspace like a normal difference between vectors. The calculation of the color distortion is given by Equations (69)–(73). The calculation of CIEDE2000 is given by Equations (74)–(76), where $\Delta L'$, $\Delta C'$ and $\Delta H'$ are the differences between pixels in their corresponding channels; S_L, S_C and S_H are compensation terms; k_L, k_C and k_H are weighting factors that depend on the application; and R_T is the hue-rotation term.

$$I_{dif} = |I_{fg}^{gray} - I_{bg}^{gray}|, \tag{67}$$

$$I_{euclidian} = \sqrt{(I_{bgR} - I_{fgR})^2 + (I_{bgG} - I_{fgG})^2 + (I_{bgB} - I_{fgB})^2}, \tag{68}$$

$$\|I_{x_t}\|^2 = I_{fgR}^2 + I_{fgG}^2 + I_{fgB}^2, \tag{69}$$

$$\|I_{v_t}\|^2 = I_{bgR}^2 + I_{bgG}^2 + I_{bgB}^2, \tag{70}$$

$$\langle I_{x_t}, I_{v_t} \rangle^2 = (I_{bgR} I_{fgR} + I_{bgG} I_{fgG} + I_{bgB} I_{fgB})^2, \tag{71}$$

$$I_{p_2}^2 = \frac{\langle I_{x_t}, I_{v_t} \rangle}{\|I_{v_t}\|}, \tag{72}$$

$$I_{colorDist} = \sqrt{\|I_{x_t}\|^2 - I_{p_2}^2}, \tag{73}$$

$$C_{ab}^* = \sqrt{a^{*2} + b^{*2}}, \tag{74}$$

$$h_{ab}^* = \arctan \frac{b^*}{a^*}, \tag{75}$$

$$I_{\Delta E_{00}^*} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H}}. \tag{76}$$

In order to segment clothing images, Deng et al. used a combination of face detection and edge detection to realize an automatic segmentation algorithm for clothing [46]. First, the position of the face is obtained using face detection, and then the position of the clothing is roughly positioned and a rectangular frame of the clothing area is roughly obtained. Then the edge-detected canny operator further refines the position of the image foreground area to make the four sides closer to the edge of the garment. After finalizing the exact boundaries, GrabCut is used to segment the garment.

Khattab et al. use Orchard–Bouman [47] clustering technology to initialize GrabCut [48]. Orchard–Bouman is a clustering technique using color quantization that uses the eigenvectors of the color covariance matrix to determine good clustering. The algorithm uses the unsupervised Orchard–Bouman clustering technique to initialize the cluster foreground and background, generate the GMM with the clustered foreground and background pixels and finally execute non-interactive GrabCut. Robust and accurate segmentation is provided by Orchard–Bouman clustering, so the effect of initial clustering on segmentation is very important.

Similarly, they also use SOFM [49] clustering technology to initialize and improve GrabCut [50]. First, the foreground and background pixels are clustered using the SOFM clustering technique, the GMM is generated using the foreground and background pixels obtained by clustering and finally GrabCut segmentation is performed.

They also improved GrabCut using K-means and Fuzzy C-means (FCM) [51] as new clustering techniques [52]. Similarly, the clustered foreground and background pixels are generated into a GMM, and finally the GrabCut segmentation is performed. They compared these different clustering techniques and found that using K-means for clustering and then performing GrabCut has the best accuracy.

Ye et al. replaced the interaction of GrabCut by combining saliency detection and preset meshing [53]. The algorithm first reduces the input image to increase the speed of the operation. The input image is then divided into 14 × 14 grids, it is determined by saliency detection whether each grid contains the foreground area, and the grid is pre-labeled as foreground and background. Then the image resolution is adjusted to reduce the amount of data and GrabCut segmentation is performed. Finally, the result is converted into a binary image, and the mathematical morphology method is used instead of matting to further smooth the boundary, reduce noise and solve the problem of the target boundary roughness.

Sun et al. designed a GrabCut model of the visual attention mechanism for apple images [54]. The model uses the graphic-based visual saliency GBVS algorithm to obtain the automatic input of the adaptive rectangle. The author also uses Ncut segmentation to solve the problem of identifying overlapping fruits and realizes instance segmentation. However, it is particularly important for the algorithm to design an appropriate initial recognition model, and it is impossible to avoid manually adjusting some parameters.

Non-interactive GrabCut is an image-segmentation algorithm that automatically separates foreground objects from the background in an image without requiring any user input. It is an extension of the original GrabCut algorithm that relies on user-defined scribbles to initialize the segmentation process. The non-interactive version of the algorithm works by first generating an initial segmentation based on color and texture cues, which are computed using a Gaussian mixture model. The initial segmentation is then refined iteratively using Graph Cuts to minimize an energy function that considers both appearance and spatial information. The energy function is based on a Markov random field model that captures the spatial relationships between pixels in the image. It considers the likelihood of each pixel belonging to the foreground or background, as well as the smoothness of the boundary between the two regions. The non-interactive GrabCut algorithm is particularly useful for segmenting large datasets where manual annotation would be impractical. It has been successfully applied to a wide range of applications, including medical imaging, video surveillance and image editing.

However, the accuracy of the segmentation results can be limited by the quality of the initial segmentation, which is based on color and texture cues. The algorithm may not perform well on images with ambiguous object boundaries or poor contrast between the object and background. The segmentation results may be sensitive to the choice of parameters, which can require some tuning for optimal performance.

3.5. Others

In addition to the above classifications, there are many researchers who have improved GrabCut from different perspectives. Deshpande et al. proposed an image-segmentation technique using MRI images [55]. This technique use two algorithms, including random walks and GrabCut in One Cut to deal with complexity in texture, indistinct and/or noisy object boundaries, lower contrast, etc. Jiang et al. proposed a fully automatic segmentation method [56] accomplished using an objective object-weight detection and modified GrabCut segmentation. This method is developed only based on the inherent image features and can be applied to different scenarios. Hua et al. proposed a GrabCut color-image-segmentation algorithm based on ROI [35]. The user selects the ROI by dragging the rectangle, and the GrabCut algorithm is only used for the ROI. That is, the ROI area is initially selected with a rectangular input box, the pixels outside the ROI area are discarded pixels, and the pixels within the ROI continue to select the object through a rectangular input box for GrabCut segmentation. This algorithm reduces the complexity of the image because it greatly reduces the pixels in the image. For complex background images, the new algorithm is less expensive and more accurate than GrabCut. However, for the case where the object accounts for a large proportion of the image, the segmentation effect is not satisfactory.

Sallem et al. improved GrabCut with RGB-D images based on appearance and geometric criteria [57]. For the depth information of the RGB-D image, the normal is changed where the plane changes. The parallelism of the normal direction is a good regional mandatory criterion, and the strong change of the edge is a clear indication of the boundary. Then, by modifying the Orchard–Bouman clustering technique, it is used to account for changes in color and normal.

Wu et al. expressed the interactive segmentation problem as a Multiple Instance Learning (MIL) [58] task and proposed MILCut [59]. The algorithm uses SLIC to obtain a superpixel map of the input image, using MIL on a superpixel basis. The bounding box of the user input is reduced to a certain range to make the bounding box and the foreground more compact, and the slices in the bounding box are taken as the positive bags. The bounding box is extended to a certain range so that it does not contain the foreground at all, and the slices outside the bounding box are taken as the negative bags. After obtaining the positive and negative bags, the regional term of the energy function is obtained according to the probability map of MIL.

Lee et al. proposed an improved GrabCut algorithm that uses clustering techniques to reduce the image noise [60]. The algorithm uses a median filter to filter the original image to reduce the noise, then uses the K-means algorithm to cluster the quantized image, and the quantized image is used for the conventional GrabCut of foreground segmentation.

In order to solve the redundancy problem of n-links construction, Niu et al. proposed an improved algorithm for dynamically constructing n-links [61]. When the image is initialized, the algorithm only constructs the t-links of the regional term without constructing n-links of any adjacent pixels. The traditional algorithm searches for the pixels belonging to the source node and calculates all the paths of the maximum flow as the starting node when the maximum flow is calculated. The algorithm first searches for all pixel nodes located at the foreground and background boundaries as the search start node. Once there is a start node to search, n-links are built between the explored and undetected pixel nodes. Each n-link is constructed only once, and the algorithm will determine if n-links have been constructed before constructing new n-links between two pixel nodes.

Lu et al. proposed an improved algorithm based on GrabCut and GMM for medical images to obtain simplified interaction and a better segmentation accuracy [62]. The algorithm obtains the parameters of the foreground and background GMM in advance through the collated training set (the image size, window width, window level and biopsy position are adjusted to be consistent on the same type of medical image). A mask of the same size as the image is then created where each point α on the mask corresponds to each pixel. $\alpha = (01)_2$ represents the foreground, $\alpha = (11)_2$ represents the possible foreground, $\alpha = (00)_2$ represents the background and $\alpha = (10)_2$ represents the possible background.

When testing, the mask is used, the parameters of the GMM are obtained using the training set and the user's brush interaction (without the rectangular input box), the GMM is updated and finally the segmentation result is obtained. The improved algorithm simplifies the interactive operation of GrabCut and is suitable for medical images.

Rajchl et al. proposed training a neural network classifier to improve GrabCut given an image dataset labelled with weak annotations, named DeepCut [63]. The labelled weak annotation of DeepCut is a bounding box, replacing the GMM with a convolutional neural network (CNN) model and solving it on a densely connected CRF. Compared to GrabCut, the algorithm uses the transfer learning and reinitializes the CNN using the parameters of the last iteration, rather than recalculating the model. The algorithm is easy to apply to medical images.

Lee et al. used depth sensors to improve GrabCut for human segmentation [64]. First, a depth sensor is used to obtain depth image and skeleton information is obtained from the depth image. The resulting skeleton is then projected onto the color image as a shape prior. The skeleton information is used to obtain the prior probability that the pixels around the skeleton belong to the background or belong to the foreground (pixels close to the skeleton belong to the foreground with a large probability, and pixels far from the skeleton belong to the background with a large probability).

Xu et al. considered the inconsistency of the bounding box, combined GrabCut's interactive mode with deep learning and proposed a new segmentation method [65]. This method uses a rectangle as a soft constraint and transforms it into a Euclidean distance map. By concatenating the image and the distance map as the input and predicting the mask as the output, the convolutional codec network is trained end-to-end. This method can have a correct output even when the rectangle is not accurate. At the same time, the author develops the network to a curve-based input and applies the network to instance-level semantic segmentation.

Figure 8 shows two examples of segmentation using Deep GrabCut. Figure 8a shows two input images. Figure 8b shows the result of artificial labelling. Figure 8c shows two mask images of the segmentation result.

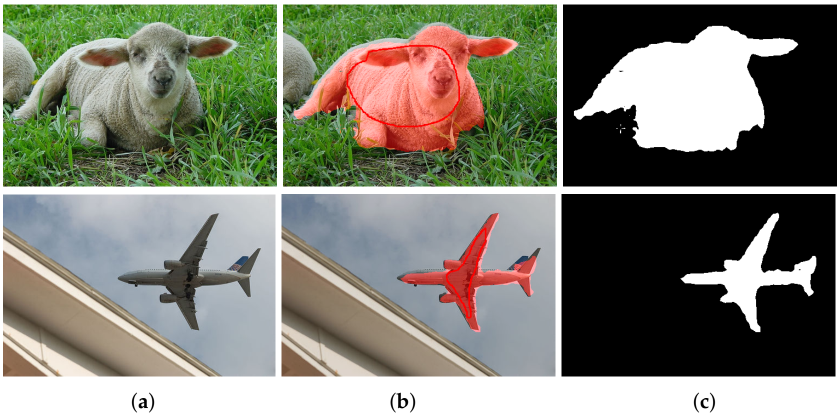


Figure 8. Results of Deep GrabCut. (**a**) Original images, (**b**) Labelled images, (**c**) Segmentation mask.

4. GrabCut Applications

With the development of computer technology and the widespread application of computer vision principles, computer image-processing technology has occupied an indispensable position in many fields. Image-segmentation technology is the basis of many image-processing technologies, so it is very important to choose image-segmentation technology that meets the needs and has a superior performance.

Interactive GrabCut can effectively extract objects from complex background images. The algorithm has a high segmentation precision and high execution efficiency, and the amount of interaction is very small. The operation is simple and can be applied to images in various fields.

4.1. Medical Images

In medical images, information extraction is a key step, and it is a research hotspot in disease diagnosis, surgical planning and the evaluation of treatment effects. The foundation and basic task of information extraction on medical images is image segmentation. However, due to various noise interferences and artifacts in medical imaging processing, as well as the diversity of pathologies, there are many difficulties in many aspects.

The formation of medical images is greatly affected by medical imaging devices and external environmental noise, resulting in blurred boundaries of medical images, making it difficult to identify subtle structures and achieve a medical diagnosis. In addition, small changes in the same target at different modalities and imaging angles can vary greatly in the imaging results. Moreover, some organs or cells in medical images may overlap with each other. The spatial complexity is higher than that of ordinary natural images. Objects also have small changes in different individuals. For the above problems, the analysis framework of medical images needs to be established for different image content and analysis tasks, so as to achieve the role of assisting doctors in observation.

Segmentation is the primary task of analyzing medical images, but, as already mentioned, there are many interfering factors in medical images. It is difficult to obtain perfect results using the existing image-segmentation methods. However, the advantages of GrabCut combined with various image-processing methods are applied to medical images, which can obtain ideal effects and solve the problems of cumbersome manual operation and low efficiency. It is one of the most popular applications of GrabCut.

It is difficult to segment the regions of interest from magnetic resonance imaging (MRI) and X-ray computed tomography (X-ray CT) images because of their high complexity, blurred boundaries and rich noise. At the same time, manual segmentation is not feasible in terms of time and cost. Ref. [66] proposes an interactive segmentation framework called MIST (Medical Image Segmentation Tool) and develops software for experimentation with the existing segmentation method. The framework automatically generates a binary mark image of the region of interest using mathematical morphology and then inputs the generated mark image as a mask into GrabCut to generate an output image. At the same time, users can use matting to further refine the region of interest, which provides accurate results for most medical images. This method is suitable for medical image segmentation with low real-time performance, and there is still room for improvement in efficiency.

Similarly, for the segmentation puzzle of medical images, ref. [62] pre-obtains the parameters of the foreground and background GMM through the collated training set. Subsequently, GrabCut is initialized using the GMM and the segmentation result is obtained through the user's brush interaction. The proposed algorithm simplifies the interaction and accelerates the convergence speed of the model. However, this requires a medical image dataset for training as a premise. As we all know, the acquisition of medical image data is not easy and may be expensive.

Cardiovascular disease is one of the prime causes of death worldwide, and cardiac fibrosis is key in the development of heart disease. The degree of white fibrosis in the appearance of the heart plays an important role in the diagnosis of and research on myocardial infarction. There are many existing segmentation technologies to segment heart images and extract relevant information, but each has its own advantages or disadvantages. In ref. [67], GrabCut is used to segment heart images, then the segmentation result is combined with the equalization of the heart map using the fuzzy clustering algorithm FCM, and finally, the threshold processing and the morphological operation are used. A segmentation map with clear myocardial ischemia is obtained. The algorithm demonstrates a good segmentation ability in the area of myocardial ischemia/myocardial infarction, which is easy to identify by visually inspection but is difficult to recognize, while being robust to low-quality photographs produced by cardiac motion. This is important for helping future clinical research or assessing the risk of heart attack.

The automation of white blood cell (WBC) detection and counting brings convenience to doctors. However, due to the different shapes and sizes of WBCs, which are prone to deformation, and some external factors, the segmentation of WBC images has a certain challenge. In ref. [68], an algorithm for automatically selecting white blood cell regions and using GrabCut segmentation is proposed. In the selection of the white blood cell area, the Canny operator is used for edge detection, and the edge density of the rectangular regions in different size ranges is scored. Similarly, the scores are based on color differences, and the combination of the two is more likely to be white blood cells. In the segmentation of white blood cells, the selected region is iteratively segmented as a mask of GrabCut, and the selected small region (nuclear region) is replaced with cytoplasmic pixels. The result of each iteration of the segmentation is expanded to iteratively split again. Compared to segmentation based directly on the original input image, the proposed framework can effectively avoid adverse effects from background factors such as red blood cells (RBC) and platelets. However, some parameters of the method proposed in the literature are set experimentally, without considering the existence of the optimal value of the parameter and the adaptability to different datasets. In addition, the experiment did not consider the overlapping of WBC.

Figure 9 shows two examples of WBC segmentation using GrabCut. Figure 9a,c are two input WBC images. Figure 9b,d are images segmented using GrabCut (images from CellaVision dataset).

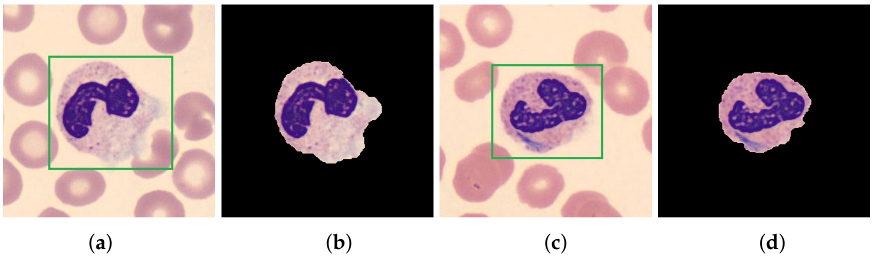

(**a**) (**b**) (**c**) (**d**)

Figure 9. Results of WBC segmentation. (**a**) Original image, (**b**) Result of GrabCut, (**c**) Original image, (**d**) Result of GrabCut.

Ultrasound tomography (UST) images for risk assessment after breast segmentation are one of the primary means of breast cancer screening and also play a key role in cancer treatment. Several segmentation algorithms for UST images have been developed today, but usually require a lot of time and excessive manual interaction and are not suitable for large-scale research. To overcome these problems, ref. [69] proposes a method called AUGC for automatically segmenting UST images. First, the input UST image is enhanced in contrast, edge detection is used to obtain the edge of the breast and then the convex hull searching algorithm is used to obtain the point where the polygon protrudes. A closed breast edge is obtained using curve fitting, and finally the closed edge is segmented as a mask of GrabCut. The algorithm shows a good performance in UST image segmentation, which greatly reduces the segmentation time.

Similarly, for the effective segmentation of breasts in UST images, ref. [70] proposes a three-dimensional GrabCut (GC3D) algorithm for breast segmentation. The algorithm needs to artificially place several points between the circular transducer and the breast boundary, then uniformly generate nine points between adjacent pairs of points using the Hermite cubic curve interpolation and finally connect all the points in order and generate a mask. The mask is supplied to GrabCut. At the same time, the mask is also provided to other slices of the UST image to save time and energy. GC3D achieves a good performance in an acceptable amount of time, saving the time for doctors to perform manual breast segmentation. This method has the potential to be fully automated when the position of the circular transducer is fixed.

For dermoscopic images, it is difficult to detect and classify skin lesions because of the variety of morphological features and the complexity of histopathological changes. In addition, skin lesion images have various colors with abrupt boundaries and pseudo-features, which make image segmentation more difficult. In ref. [71], filtering techniques are used to remove noise from this type of image, and then skin lesions are segmented using GrabCut. Finally, the skin lesion area is obtained using K-means clustering and post-processing. The difference between this algorithm and most segmentation algorithms is the use of GrabCut, which uses edge and region information to locate global lesions. Then K-means fine-tunes the localized area to the segment lesion area, effectively extracting the skin-lesion area. At the same time, there is a certain problem that when the lesion exceeds the boundary, the detection result sometimes misses part of the lesion area.

In processing dermoscopic images, Halil et al. also proposed a new method [42]. They first used the DullRazor algorithm to remove the hair's effect on the lesion, then detected the lesion area through the Yolov3 depth model and segmented it using GrabCut, and finally used morphological operators for post-processing to obtain the skin-lesion area. This method cannot detect lesions when they are low-contrast or the lesion area occupies the entire image surface, because the Yolov3 model does not learn these types of data. Therefore, a large and diverse dataset is the key to the success of the algorithm.

Due to the corrosion resistance, high melting point and high hardness of teeth, dental biometric technology plays an important role in modern forensic science. One of the main steps in personal tooth identification is the complete segmentation of dental images. Ref. [72] uses a morphological opening operation on dental X-ray images and GrabCut to obtain contour images and crown images of the teeth, respectively. The contour image of the teeth and the crown image are combined to obtain a complete image of the tooth segmentation. The algorithm can segment the complete teeth image in the case of uneven gray-scale distribution and teeth connected with other parts. The limitation is also obvious. When the gray value of the image does not change much, the final result easily appears incomplete.

The synapse is the structural basis of the functional activities of the nervous system. The information transmission between functional neurons must have a mature synaptic structure. Therefore, it is of great significance to learn about the relevant factors of synapse formation and its mechanism of action. The verification of synapses in electron microscopy (EM) requires much heavy and repeated manual work, so automatic synaptic reconstruction pipelines are essential for analyzing large amounts of brain tissue. In ref. [73], in order to avoid an incorrect distinction between the synaptic gap and membrane, the presynaptic membrane is considered as background information, while the postsynaptic membrane and synaptic gap are considered as a whole, and then the famous deep network Faster R-CNN is used to locate the synapse. A z-continuity screening method is used for the output of the deep network in order to improve the detection accuracy. In fine segmentation, the Dijkstra algorithm is used to obtain the optimal path of the synaptic gap, and then fed to GrabCut for fine segmentation. Finally, ImageJ is used to display the three-dimensional structure of the synaptic crack. The algorithm improves the detection accuracy, ensures the accuracy of segmentation, improves the efficiency of synaptic verification and facilitates the analysis of connectomics and synaptic plasticity.

Frants and Agaian proposed an extended GrabCut image-segmentation algorithm for foreground/background dermoscopic image-segmentation applications [74]. The algorithm integrates octree color quantization and a modified GrabCut method with a new energy function. This method effectively solves the automatic skin-lesion segmentation problem and has great significance for the precise diagnosis of skin cancer. To address the problems associated with detecting low-grade tumors and CSF fluid leaks in the initial phase of brain cancer, Saeed et al. proposed a new framework of the hybrid k-nearest neighbors model that is a combination of the hybridization of Graph Cut and support vector machines and a hidden Markov model of the K-means clustering algorithm [75]. They used a GrabCut segmentation method, which is the application of the Graph Cut algorithm, and extracted the data with a scale-invariant features transform. In conclusion, this model gives better results than existing models.

MRI images play an important role in the diagnosis of childhood chronic kidney disease (CKD), providing a more comprehensive kidney anatomy and function assessment, which is usually necessary for the diagnosis of CKD. Ref. [76] proposes a fully automated kidney segmentation technique for the assessment of glomerular filtration rate (GFR) in children. This method uses GrabCut for time-resolved 3D DCE-MRI data sets. A random forest classifier further divides the kidney tissue into the cortex, medulla and collection systems. The automatic segmentation method has a similar effect to the manual segmentation on the GFR estimation. However, when the medulla clusters are more than one cortical thickness apart, the algorithm may fail to automatically segment due to labelling the kidneys incorrectly.

4.2. Non-Medical Images

In addition to its wide application in medical images, GrabCut has become an application hotspot in many other fields because of its excellent graphics-segmentation capabilities.

4.2.1. Applications in Object Detection and Recognition

It is an important task for the intelligent monitoring system to detect abandoned objects in public. In the field of target detection, Liang et al. proposed a simple and effective calibration guidance scheme to solve the problem of target detection in the aviation field [77]. To detect camouflaged objects, Hongwei Zhu et al. present a novel boundary-guided separated attention network (call BSA-Net) [78]. BSA-Net utilizes two-stream separated attention modules to highlight the separator between an image's background and foreground. The results show that BSA-Net has an obvious detection effect on camouflaged objects. To solve the problem of complex background and poor imaging quality in target detection, Bin Kang et al. proposed a simple yet effective calibrated-guidance scheme to enhance channel communications in a feature-transformer fashion that can adaptively determine the calibration weights for each channel based on the global feature-affinity correlations [79]. The results show that this method has a strong performance trend in directional target detection and horizontal target detection. There are huge hidden dangers in abandoned objects, such as luggage and backpacks, which may be dangerous objects placed by terrorist attackers deliberately. In ref. [80], a new framework for the automatic detection of abandoned objects is proposed. In order to obtain an accurate target in the frame, the author uses GrabCut to obtain the precise detection result after obtaining the boundary box of the abandoned object. The detection system does not need to add a tracking mechanism and can obtain robust and accurate results in a complex and real environment.

Vision-guided remote robots are typically used to perform tasks such as crawling and categorization in various environments that contain unfamiliar objects in addition to matching libraries. Therefore, breaking the traditional algorithm to create templates for unfamiliar objects is an important research direction. Ref. [81] proposes the use of a superpixel algorithm to obtain the superpixel map of an image captured by a robot and then remotely artificially divide the target with GrabCut to create the target and a template for later matching. This method can replace a large number of tasks, such as grabbing and sorting, and release the operator's workload. At the same time, GrabCut at the superpixel level separates objects from texture-rich backgrounds, reducing iterations and time consumption.

The automatic identification of seabirds using machines is helpful to monitor the activity areas and rules of seabirds in the wild. However, automatic recognition is difficult because of different lighting, complex backgrounds or different directions and postures of birds. For the automatic identification method of seabirds, ref. [82] introduces GrabCut to segment seabird units from complex backgrounds. Then, by combining the global features, such as the shape, texture and color, and local features such as SIFT, it overcomes the difficulty of identification caused by various postures and directions. Finally, based on their integrated features, the seabirds are identified using a combined classifier. This method does not require annotations of bird body and attribute vocabulary and also achieves a better performance.

Figure 10 shows three examples of seabird segmentation using GrabCut. Figure 10a–c are seabird images in flight, standing and swimming positions, respectively and the corresponding GrabCut segmented images (images from MSRA10K dataset [23]).

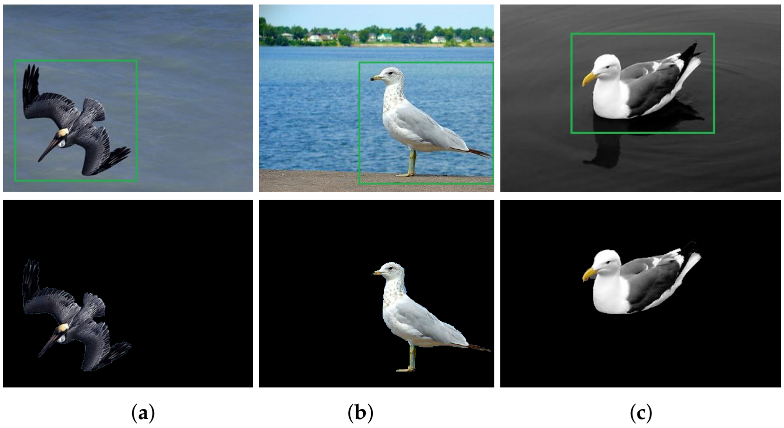


Figure 10. Results of seabirds segmentation. (**a**–**c**) Original images and results of GrabCut.

Icing on the transmission line may cause ice flashing of the insulator, collapse of the tower, tripping of the transmission line, and other accidents. Therefore, serious ice formation on transmission lines will have serious consequences. However, there is no suitable way to represent and evaluate the icing conditions between insulator sheds. Ref. [83] studied image-processing method to detect natural icing on glass insulators. By identifying the convex defects of the contour of the icing insulator string based on GrabCut, the calculation method of the pattern spacing and the pattern cantilever is proposed to estimate the icing degree of the entire insulator string. This method, due to the superior segmentation performance of GrabCut, easily identifies icing conditions from significant changes in the pattern spacing and pattern overhang.

In a natural scene, the text or character area is the area with the most significant semantics, which conveys important information about the image. However, text or character detection is still a challenging research problem. Ref. [24] uses GrabCut based on salient regions to segment regions containing text content, then uses the maximally stable extremal region (MSER) feature detector for text detection and finally the Canny edge detector generates salient text. This method using GrabCut can only segment a large fuzzy area containing text and cannot obtain a pure text result. However, after introducing the MSER feature detector, it can separate the text area and the non-text area well and obtain accurate results.

With the rapid development of UAV tracking technology in agriculture, aviation, navigation, transportation and public security, Li et al. proposed and evaluated the residue-aware correlation filters and the method of refining scale estimates with GrabCut [84]. The accuracy and precision of a UAV tracker can be greatly improved using Grabcut technology. Salau et al. proposed a modified GrabCut algorithm for localizing vehicle plate numbers [85]. It extends the use of the traditional GrabCut algorithm with the addition of a feature-extraction method that uses geometric information to give accurate foreground extraction. The experimental result shows that this algorithm has high accuracy and plays an important role in traffic control and surveillance systems. In order to solve the problem of difficult and inefficient license-plate localization in complex environments, Shi et al. proposed an improved GrabCut Chinese license plate location-detection algorithm [86]. They replace the candidate frame by introducing the aspect ratio of the license plate as the foreground-extraction feature to automate the detection of the license plate using the GrabCut algorithm. The results show that the improved GrabCut algorithm has a better accuracy and real-time performance.

4.2.2. Applications in Video Processing

In video processing, since video surveillance is ubiquitous and indispensable to improve personal safety, it is very important to solve the privacy problem in video surveillance. Ref. [87] proposes an automatic de-identification technique in surveillance videos based on computer vision. This method uses background subtraction to detect pedestrians and background areas. It then reduces the contrast of the background area while preserving the contrast of the pedestrian area and uses the pedestrian area as a mask for GrabCut segmentation. The segmented pedestrians use a neural art algorithm, that is, the deep features of an image are replaced with other features. After such processing, automatic de-identification of video surveillance is realized. The resulting de-identified image has many appearance features that are different from the original image (e.g., hair and clothing colors). At the same time, it keeps the naturalness of the contours of the de-identified humans and scenes. This method takes the shape of a person as an important feature in a small dataset, which makes it easy to distinguish pedestrians by mistake. In large datasets, because many people are similar in shape, this problem is alleviated to some extent.

Video object segmentation is based on automatically segmenting unmarked objects in the video. The application of this technology often requires a good segmentation quality and time efficiency. Ref. [88] uses a non-iterative version of GrabCut to develop a new video-object-segmentation framework. The framework only needs to perform interactive processing on the first frame of the video, and the subsequent frame segmentation can achieve efficient non-interactive and non-iterative processing using the previous GMM. However, the limitation is also obvious. It can be applied to videos with simple backgrounds, but it inevitably produces deviations in complex backgrounds and cannot provide accurate outlines.

4.2.3. Applications in Agriculture and Animal Husbandry

With the application of intelligent and refined technologies in agriculture, the requirements for the quality inspection of crops are becoming higher and higher. Taking cucumber as a representative, its growth status and appearance quality directly affect the yield and farmers' income. Therefore, it is important to evaluate the appearance, quality and growth of cucumber. In order to improve the quality-detection accuracy and processing efficiency of cucumber images, ref. [53] used pre-processed cucumber images to extract cucumbers with GrabCut. Pre-processing reduces the number of iterations and operation time of GrabCut. Finally, the image noise and jagged borders are removed using morphological operations and the complete contour of the cucumber is segmented. This method shortens the average running time; the effect is better than SLIC and traditional GrabCut. It realizes the non-destructive extraction of a cucumber on a complex background, which can meet the evaluation of basic growth conditions.

In order to improve the fruit quality and optimize orchard management, Sun et al. used the GrabCut and Ncut algorithms to identify apples in orchard images [54]. They designed a GrabCut model based on the visual attention mechanism, used Ncut to segment and identify overlapping fruits and finally used the three-point circle-fitting method to reconstruct the apple. The apple-identification method has the potential to realize early growth monitoring and yield estimation, but it is difficult for non-professionals to improve the identification accuracy because it requires manual parameter adjustment.

In the digital and intelligent pig industry, reducing the cost and labor intensity of enterprises and increasing production and income are the fundamental goals. The precise segmentation of pigs is the basic work of artificial intelligence object tracking and behavior recognition. It is one of the important techniques for identifying piglet movement or rest and judging whether the piglet has been squeezed for a long time. Ref. [89] first performed a series of pre-processing operations on piglet images to reduce the influence of the lighting and surrounding environment on object segmentation. Then they obtained the input of GrabCut through a morphological operation, achieved fine segmentation and finally performed feature recognition to judge the status of the piglet. This method has high accuracy, and the average processing time meets the real-time requirements of the agricultural video-surveillance system.

In order to prevent a decrease in plant yield caused by diseases and pests, Qi et al. proposed a lightweight convolutional neural network [90]. First, they use the GrabCut algorithm to unify the background of the experimental data and the real data to black. Second, they propose a new coordinate attention block to improve the classification accuracy of convolutional neural networks. Finally, to make the trained network more available for agricultural platforms with limited resources, model compression is applied to the trained network. As shown in the study, this model can be well applied in agriculture to identify plant disease categories and improve the yield and quality of crops.

At present, weeding in China mainly relies on chemical herbicide spraying on a large area. To improve the efficiency and reduce environmental pollution, Zhang et al. proposed a modified Grabcut algorithm [91]. They first used filtering technology to enhance and suppress the noise in the original weed image. In the segmentation stage, they used an improved GrabCut algorithm to roughly segment each weed image and used adaptive fuzzy dynamic K-means to segment the original weed image. Finally, the weed species is recognized using SRC. The results validate that the proposed method is effective for weed-species recognition, which can be used as a preliminary step for precision-applying pesticide.

4.2.4. Applications in Human Body Images

The segmentation of the human body area is essential in many applications, such as human activity recognition, virtual reality games and video surveillance. Due to the complex shape and structure of the human body and irregular movements, human body segmentation is still a challenging problem. At present, human body segmentation with a single background is widely used, but it is still a research hotspot in complex backgrounds. Ref. [64] uses depth sensors to obtain depth images and human skeletons and project human skeleton information onto color images. The energy function is established based on the ideas of Graph Cuts and GrabCut, the prior probability provided by the human skeleton is increased and finally the energy function is optimized to obtain the segmented human body region. Compared with traditional human segmentation, this method still provides high-quality segmentation results in a complex background and is universal.

When taking passport images, a high-precision automatic human-body-extraction algorithm is very important, because passport images must meet high-demand ICAO standards. Therefore, designing a high-precision algorithm is helpful to the machine that automatically takes passport images. Ref. [45] used four different methods to automatically extract the human body region from passport images. The extraction results are segmented using GrabCut to realize non-interactive human body extraction of the passport images. The method in the paper obtains high-quality results in simple scenes, but it will be affected by the background and lighting in more challenging scenes, which will greatly affect the segmentation accuracy.

4.2.5. Other Applications

In remote-sensing images, clouds that appear above a ground object due to weather factors have become a research hotspot for better interpretation of image information, and cloud extraction is an important process. Ref. [92] first splits satellite images into superpixels through SLIC and extracts the unique features from remote sensing images. The probabilistic latent semantic analysis (PLSA) model is used to extract the deep information in the superpixels, and the descriptor of each superpixel is calculated to obtain the feature vector. Finally, the feature vector is output using the support vector machine (SVM) and the cloud mask is fed to GrabCut with threshold processing to achieve accurate cloud segmentation. This method uses GrabCut to further refine the cloud-detection results at the pixel level, effectively improving the cloud-detection accuracy and achieving robust results.

Peng et al. used an improved GrabCut based on the visual attention model to identify rare-earth ore-mining areas in remote-sensing images [25]. They used the ITTI visual attention model to generate a saliency map to initialize GrabCut and added a normalized difference vegetation index (NDVI) term to the energy function of GrabCut to constrain the segmentation results. This identification method will cause errors in part of the impervious surface and part of the reclaimed areas in the abandoned rare-earth ore-mining region.

With the continuous innovation of existing technology, a large number of existing clothing images are the main research objects in clothing sales systems. In the face of clothing image segmentation, ref. [46] uses GrabCut to propose a fully automatic clothing-image-segmentation framework. The framework can be divided into two situations: with or without models. It uses face-detection and edge-detection algorithms to provide a basis for clothing positioning. Finally, according to the positioning results, automatic GrabCut is used to obtain the clothing segmentation results. The accuracy of the framework lags behind the classic algorithm, which improves the efficiency. In the case of low-accuracy requirements, it can be applied to the retrieval system of massive images during online clothing shopping.

Yamasaki et al. proposed a support system that uses ICT and also investigated a method of extracting stone-contour information [93]. They set restricted regions for background-likely characteristics using a convex hull of a pre-extraction result using GrabCut close to the original iteratively. The results show that the method improves the problem of over-segmentation or insufficient segmentation. Zhang et al. proposed a new building-extraction method from high-resolution remote-sensing images based on GrabCut that can automatically select foreground and background samples under the constraints of building elevation contour lines [94]. GrabCut and geometric features are used to carry out image segmentation and extract buildings. The results show that image segmentation with GrabCut can better preserve the entire building boundary.

5. Discussion

After the previous description, we know that GrabCut is a powerful image-processing tool with superior performance. In this section, we will employ the typical GrabCut methods to verify their performance and also compare the specific performance of different improved GrabCut models.

5.1. Experimental Results

For the improved GrabCut algorithm reviewed above, we have selected some classic and high-frequency algorithms (GrabCut, LazySnapping, OneCut, Saliency Cuts, method of [11], DenseCut and Deep GrabCut) in the experiments in this section. The results are shown in Figure 11. The segmented image was randomly selected from the GrabCut dataset [2] containing 50 images and the corresponding binary segmentation masks and the MSRA-B dataset [95] containing 5000 images and the corresponding binary segmentation masks. Moreover, in order to compare the experimental results objectively, we used five evaluation indicators to evaluate the segmentation results [96], which are the recall, precision, F-measure (FMS_η), Jaccard index (JAC) and time [97]. The results are shown in Table 1 and Figures 12–14. The recall, precision, FMS_η and JAC are all widely used metrics in image-segmentation evaluation. The recall is a measure of coverage, measuring how many actual positive cases are divided into positive, and the precision is a measure of how much is divided into positive cases that are actually positive. FMS_η is the weighted harmonic average of the recall and precision, combining the results of recall and precision. JAC is used to compare similarities and differences between the classifications and actual categories. Their formulas are Equations (77)–(80).

$$Recall = TPR = \frac{TP}{TP + FN}, \tag{77}$$

$$Precision = PPV = \frac{TP}{TP + FP}, \tag{78}$$

$$FMS_{\beta} = \frac{(\beta^2 + 1) \cdot PPV \cdot TPR}{\beta^2 \cdot PPV + TPR}, \tag{79}$$

$$JAC = \frac{TP}{TP + FP + FN}, \tag{80}$$

where TP is the true positives, FP is the false positives, TN is the true negatives and FN is the false negatives. $\eta = 1$ in FMS_{η} is the most common (evaluation index $\eta = 1$ in this paper).

From the results of the segmentation, it is easy to segment errors because some of the pixels in the foreground of IMG1 are similar to the background, and these pixels are located in the center of the object. The six traditional algorithms have obvious under-segmentation phenomena in this region, but a good segmentation can be achieved for the overall segmentation of the object. Although Deep GrabCut has no under-segmentation of the object center, it has the problem of fuzzy edge segmentation inherent in deep learning methods. The most ideal is GrabCut, which has the best performance value on recall, FMS_1 and JAC.

Compared with IMG1, the segmentation of IMG2–IMG6 is more difficult. Among them, IMG2 is likely to cause segmentation errors because the color of kangaroo fur is similar to the color of an area in the grass. The kangaroo in the picture occupies most of the area, and the color distribution range of the image is small. This is unfavorable for partial segmentation algorithms to establish foreground and background color models. For example, the bounding box required by GrabCut occupies almost the entire image. This makes the background GMM unable to obtain most of the background color range initially and finally leads to over-segmentation. The edges of LazySnapping and the method of [11] are very rough, and the segmentation effect can be improved only by investing more human interaction to paint the boundaries. The OneCut effect is the worst, and each item in the evaluation index of segmentation quality is the lowest because the segmentation failed (this kind of failure case will be explained below). The most ideal segmentation is Saliency Cuts, which has the highest scores for FMS_1 and JAC, but the kangaroo's hand is considered as the background because it is similar to part of the ground area. DenseCut is the opposite of Saliency Cuts, which treats similar areas on the ground as the foreground. The approximate area of Deep GrabCut can be segmented, but the boundary segmentation is not ideal.

The object of IMG3 is significant, but the complex background brings difficulties to segmentation. LazySnapping and OneCut divide the background person into the foreground. The method of [11] has rough edges. The Deep GrabCut segmentation is incomplete and the under-segmentation area is large. GrabCut scored the highest in recall, FMS_1 and JAC.

IMG4 has a high background complexity, the foreground boundary is complex, and some foreground areas are very similar to the background area pixels, so the division is difficult. The segmentation result of GrabCut has a large area of over-segmentation, and the precision, FMS_1 and JAC are relatively low in the evaluation of indicators. The result of Deep GrabCut is not ideal, but its robustness is strong. Despite the high complexity of the image, the approximate object area can be segmented. The other algorithms have a small area of error segmentation basically, but the overall effect is ideal. In terms of index evaluation, LazySnapping has the best performance value in precision, FMS_1 and JAC.

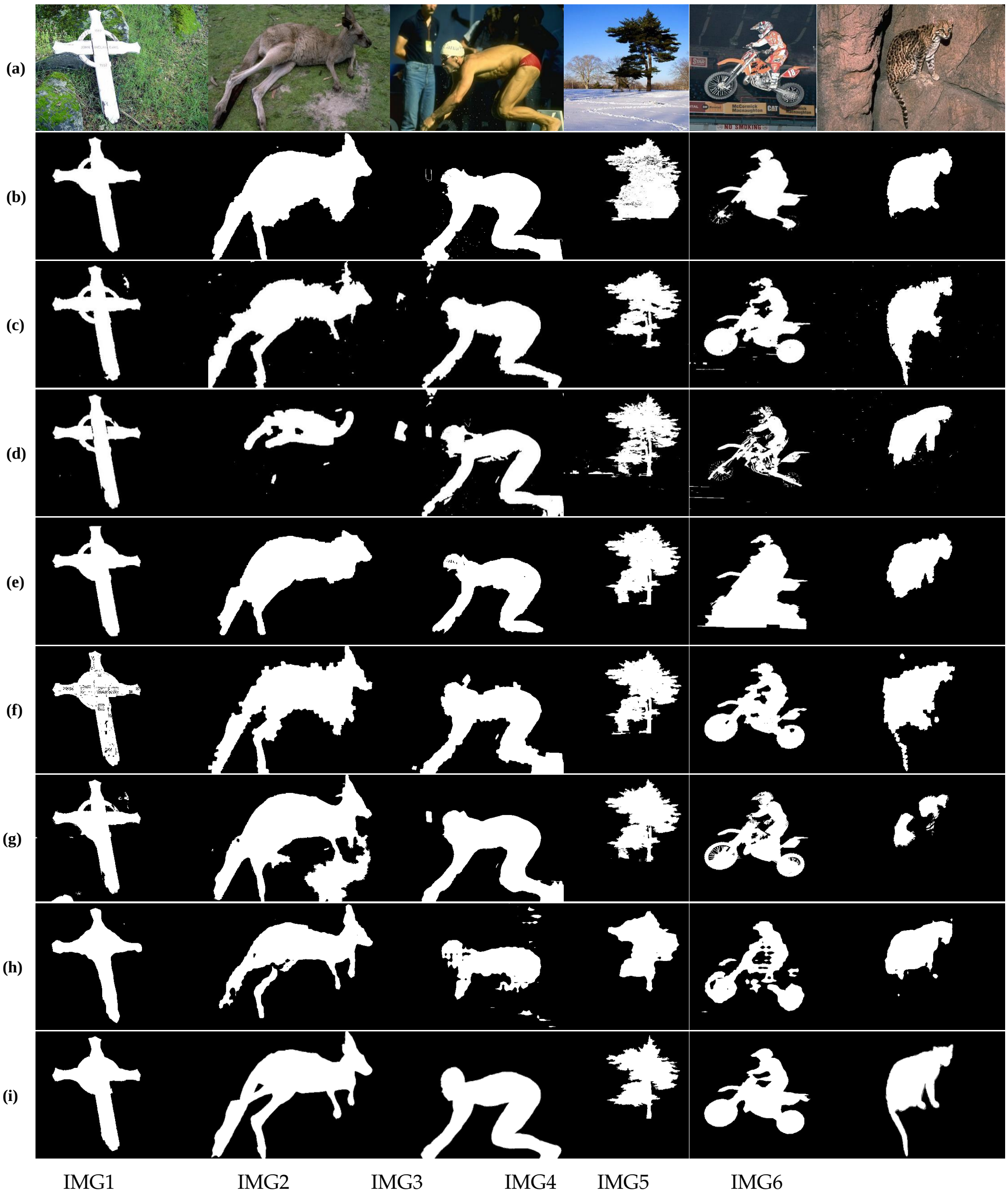


Figure 11. Comparison of experimental results. (**a**) Original images, (**b**) GrabCut, (**c**) LazySnapping, (**d**) OneCut, (**e**) Saliency Cuts, (**f**) Method of [11], (**g**) DenseCut, (**h**) Deep GrabCut, (**i**) Ground truth.

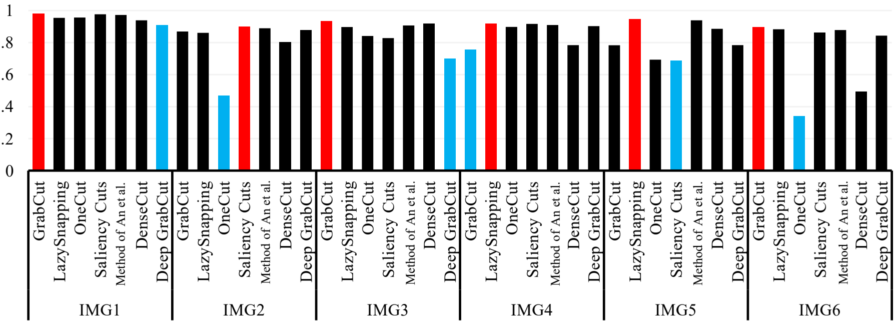


Figure 12. GrabCut, LazySnapping, OneCut, Saliency Cuts, method of [11], DenseCut, Deep GrabCut FMS_1 on 6 images.

Table 1. Performance comparison of 6 images. The best and worst performance values are shown in red and blue, respectively.

Image	Method	Recall	Precision	FMS_1	JAC	Time (secs)
IMG1	GrabCut	0.9752	0.9889	0.9820	0.9647	8.8695
	LazySnapping	0.9494	0.9597	0.9545	0.9130	0.2570
	OneCut	0.9201	0.9969	0.9569	0.9175	0.5093
	Saliency Cuts	0.9641	0.9892	0.9765	0.9541	1.7063
	Method of [11]	0.9244	0.9876	0.9550	0.9139	3.8284
	DenseCut	0.9445	0.9342	0.9393	0.8856	3.7621
	Deep GrabCut	0.8752	0.9447	0.9085	0.8324	5.8396
IMG2	GrabCut	0.9911	0.7739	0.8691	0.7685	10.9130
	LazySnapping	0.8737	0.8481	0.8607	0.7554	0.9667
	OneCut	0.7026	0.3530	0.4700	0.3072	7.5650
	Saliency Cuts	0.9075	0.8901	0.8987	0.8160	0.9005
	Method of [11]	0.9749	0.8177	0.8894	0.8008	2.9662
	DenseCut	0.9806	0.6820	0.8045	0.6729	1.3253
	Deep GrabCut	0.8018	0.9713	0.8785	0.7833	7.4800
IMG3	GrabCut	0.9748	0.8983	0.9350	0.8779	7.7970
	LazySnapping	0.8763	0.9186	0.8969	0.8131	1.0038
	OneCut	0.8213	0.8624	0.8413	0.7261	2.5656
	Saliency Cuts	0.7304	0.9561	0.8281	0.7067	0.9423
	Method of [11]	0.9422	0.8743	0.9070	0.8298	2.8496
	DenseCut	0.9373	0.9006	0.9186	0.8495	1.2150
	Deep GrabCut	0.5559	0.9507	0.7016	0.5403	7.5170
IMG4	GrabCut	0.9355	0.6369	0.7577	0.6099	8.1819
	LazySnapping	0.9439	0.8941	0.9183	0.8490	0.2536
	OneCut	0.9379	0.8617	0.8982	0.8152	0.6877
	Saliency Cuts	0.9904	0.8529	0.9166	0.8460	1.7416
	Method of [11]	0.9802	0.8487	0.9098	0.8345	5.4663
	DenseCut	0.9980	0.6455	0.7839	0.6447	3.0919
	Deep GrabCut	0.9147	0.8915	0.9029	0.8231	6.3885
IMG5	GrabCut	0.6700	0.9425	0.7833	0.6437	20.2160
	LazySnapping	0.9183	0.9740	0.9453	0.8964	0.2394
	OneCut	0.5358	0.9824	0.6934	0.5307	1.6940
	Saliency Cuts	0.8033	0.6021	0.6883	0.5247	3.7740
	Method of [11]	0.9077	0.9723	0.9389	0.8848	2.9830
	DenseCut	0.7989	0.9937	0.8857	0.7949	2.6913
	Deep GrabCut	0.7058	0.8824	0.7843	0.6451	6.8088
IMG6	GrabCut	0.8858	0.9068	0.8962	0.8119	4.1540
	LazySnapping	0.9332	0.8383	0.8832	0.7908	1.0684
	OneCut	0.6809	0.2262	0.3397	0.2046	2.6872
	Saliency Cuts	0.7904	0.9497	0.8628	0.7587	4.0841
	Method of [11]	0.9406	0.8240	0.8785	0.7833	2.8732
	DenseCut	0.3270	1.0000	0.4929	0.3270	1.2402
	Deep GrabCut	0.7619	0.9445	0.8434	0.7292	3.4945

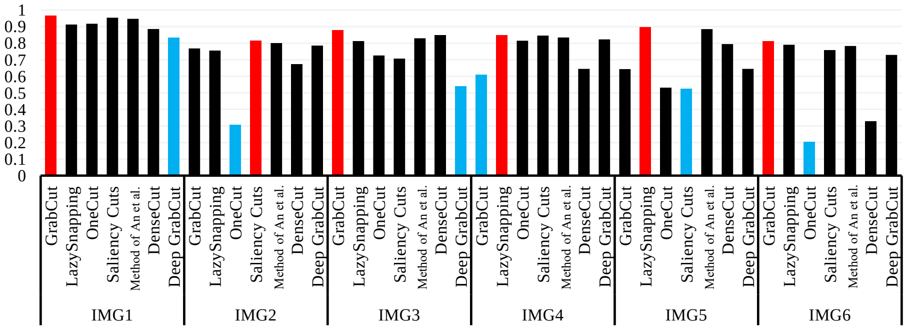


Figure 13. GrabCut, LazySnapping, OneCut, Saliency Cuts, method of [11], DenseCut, Deep GrabCut JAC on 6 images.

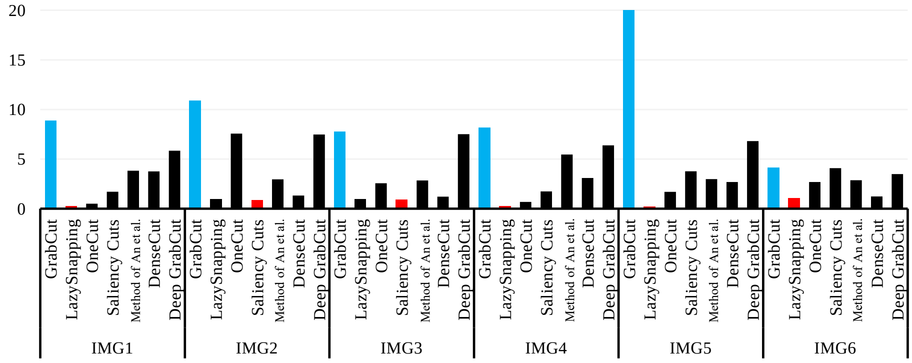


Figure 14. GrabCut, LazySnapping, OneCut, Saliency Cuts, method of [11], DenseCut, Deep GrabCut Time (secs) on 6 images.

For IMG5, it is an image with a high background complexity, and a large area on the wheel is similar to the background pixel. A large area of under-segmentation occurred in the segmentation result of OneCut, and both the recall and JAC were lower in the evaluation index. A large area of over-segmentation occurred in the segmentation result of Saliency Cuts, and both the precision and FMS_1 were low in the evaluation index. The ideal algorithms for segmentation are LazySnapping, DenseCut and the method of [11], especially LazySnapping, for which the evaluation index is the most ideal.

For IMG6, not only is the color range of the image concentrated, but the contrast between the object and the background is low, which makes segmentation difficult. The segmentation problem of each algorithm is obvious. GrabCut scored the highest in the evaluation index, but segmentation errors appeared in details such as tails. Saliency Cuts and Deep GrabCut also have this problem. LazySnapping and the method of [11] are very rough. OneCut and DenseCut have a large area of under-segmentation.

For the above experiment, comparing the running time, GrabCut has the longest segmentation time. Because the optimization of the energy function is NP-hard, the amount of calculation is large and there is no improvement. The second is Deep GrabCut. The algorithm runs slowly and the edge segmentation accuracy is low. However, because of the introduction of an interactive mode, the classification of the dataset is weakened and the generalization is improved. LazySnapping shows the fastest segmentation speed in most images. It has the highest efficiency and the best real-time performance, and the segmentation of LazySnapping is stable.

In addition to a detailed analysis of this section, we have conducted more experiments. We evaluated extensively on the GrabCut dataset and MSRA-B dataset, and the statistical results are shown in Tables 2 and 3, respectively. Part of the segmentation results are shown in Figures 15 and 16.

Table 2. Performance comparison on the GrabCut dataset. The best and worst performance values are shown in red and blue, respectively.

Method	Recall	Precision	FMS_1	JAC	Time (secs)
GrabCut	0.9668	0.9213	0.9407	0.8927	11.0076
LazySnapping	0.9681	0.9104	0.9357	0.8842	1.3669
OneCut	0.8585	0.7926	0.7899	0.6974	6.1393
Saliency Cuts	0.8371	0.8892	0.8255	0.7458	0.6803
Method of [11]	0.9614	0.8878	0.9212	0.8597	3.5718
DenseCut	0.8427	0.9418	0.8561	0.7927	1.3851
Deep GrabCut	0.8854	0.8774	0.8701	0.7849	10.3698

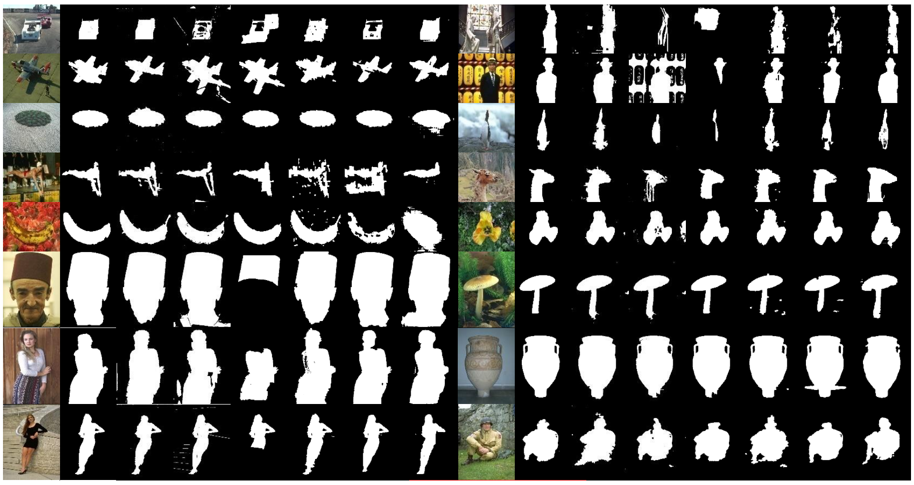


Figure 15. Results of different methods on the GrabCut dataset. For each image block, the original images are shown in the first column. The results of GrabCut, LazySnapping, OneCut, Saliency Cut, the method of [11], DenseCut and Deep GrabCut are shown in the second-to-last column, respectively.

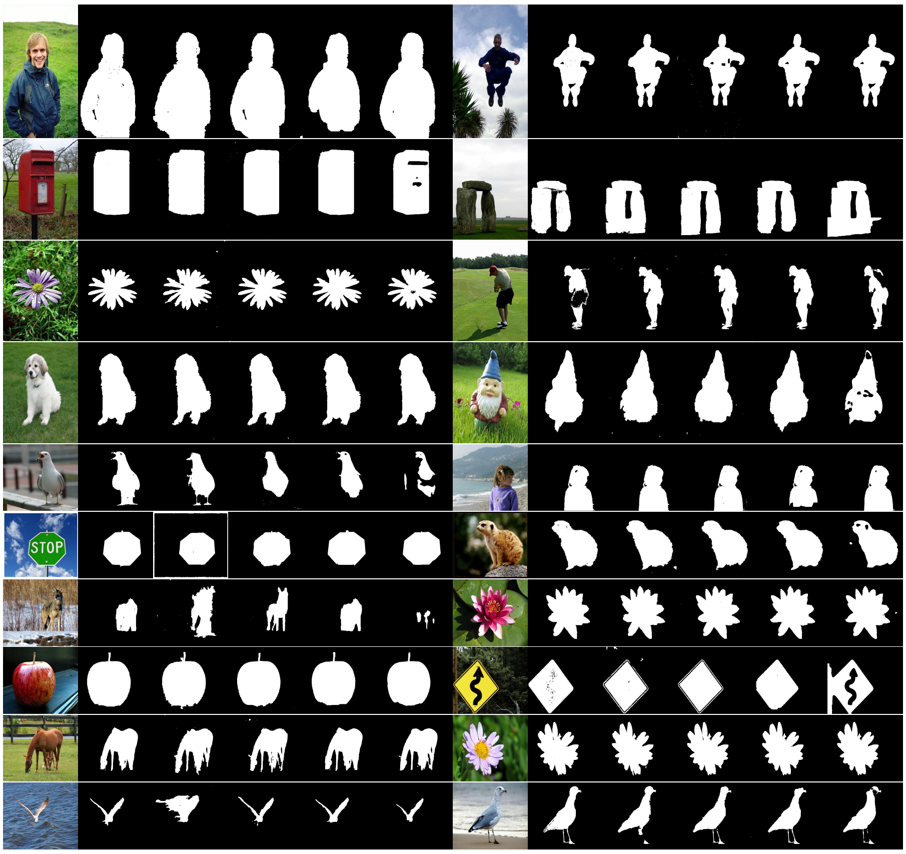


Figure 16. Results of different methods on the MSRA-B dataset. For each image block, the original images are shown in the first column. The results of GrabCut, LazySnapping, OneCut, Saliency Cuts, the method of [11], DenseCut and Deep GrabCut are shown in the second-to-last column, respectively.

Table 3. Performance comparison on the MSRA-B dataset. The best and worst performance values are shown in red and blue, respectively.

Method	Recall	Precision	FMS_1	JAC	Time (secs)
GrabCut	0.9429	0.9251	0.9301	0.8772	7.2807
LazySnapping	0.9548	0.8680	0.9008	0.8348	0.6805
OneCut	0.8609	0.8531	0.8363	0.7539	1.6462
Saliency Cuts	0.8704	0.8764	0.8614	0.7933	0.7436
Method of [11]	0.9463	0.8905	0.9141	0.8507	2.5903
DenseCut	0.8323	0.9419	0.8676	0.7951	0.9125
Deep GrabCut	0.8702	0.8765	0.8641	0.7833	5.9086

From the evaluation of the two datasets, it can be seen that GrabCut has the best effect and the highest stability, but the disadvantage is that it takes a long time. Because LazySnapping uses the watershed algorithm and replaces the complex GMM coefficient iteration with simple statistical methods, the speed is significantly faster and can meet most real-time requirements. However, it is known from the experimental process and results that LazySnapping easily classifies the background as the foreground, regardless of whether it is connected to the foreground area. This situation occurs if the pixels in part of the color range in the background are not marked by the background brush. Similarly, the foreground area has this problem.

OneCut has failed completely during the experiment. When a large area of the foreground is similar to the background, the segmentation result is only the pixels marked by the foreground brush. Because OneCut is constrained by unary potentials, it will roughly classify pixels similar to the foreground brush as the foreground, and the background is the same. If the foreground and background brush input are concentrated in part of the color range, the more pixels the brush inputs, the greater the binding force. At this time, the algorithm is easy to trap in a local optimum. Here are two failure cases, one is a complete failure and the other is a local failure, as shown in Figure 17. OneCut is suitable for situations where the background is complex but the foreground and background areas are similar with few pixels.

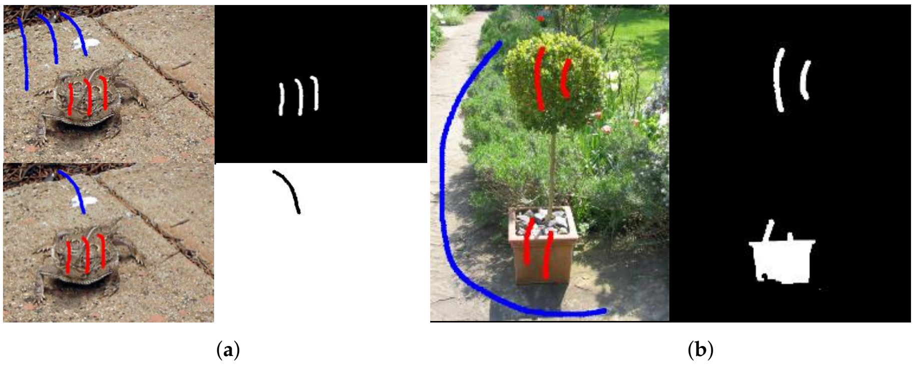

(**a**) (**b**)

Figure 17. Two OneCut failure cases. (**a**) Complete failure, (**b**) Local failure.

Saliency Cut is an automatic segmentation algorithm that is fast and can achieve real-time performance. However, the mask image of part of the data in the GrabCut dataset is not made entirely dependent on saliency. It is subjective and is suitable for measuring the performance of interactive segmentation algorithms. Therefore, Saliency Cut is more difficult to segment on the GrabCut dataset. During the experiment, Saliency Cut needed to adjust the saliency threshold. The best threshold for each image was not consistent. In order to conduct batch experiments, we use a fixed threshold (the threshold is 70).

The method of [11] can efficiently segment the approximate foreground area. However, it has a weak segmentation performance for small foreground areas and poor edge segmentation. It needs to be further segmented and optimized at the pixel level. DenseCut, which is also based on the bounding box, can segment most images with good results, and the speed also has advantages. Because DenseCut uses edge detection and connected domain filling, it is sensitive to the background connected domain in the bounding box. If the contrast between the background and the foreground is low, segmentation errors easily occur.

The Deep GrabCut algorithm enhances the fault tolerance of user interaction. It allows interactive brushes to be drawn inside the foreground area and outside. However, it was observed during experiments that the effect is more stable when the area drawn by the interactive brush only contains the foreground area. If the interactive brush draws outside the foreground area, it is easy to segment the background area contained in the brush as the object. At this time, the segmentation effect of the foreground and background in the interactive boundary is unstable.

From the review of the improved algorithms of GrabCut and their application in various fields, it can be seen that GrabCut is still a research hotspot in the field of image segmentation. Although there are many improved GrabCut algorithms, some disadvantages of GrabCut have not been overcome at the same time.

Since the result is an iterative solution to the energy function to obtain the optimal solution, and the form of the energy function is NP-hard, the solution of the algorithm is time-consuming. Therefore, how to reduce the amount of computation and find optimization methods is an important task for GrabCut. One possible solution is to modify the segmentation model or energy function, for example, modifying the GMM with large computational complexity or using a simpler model to replace a complex function. Preprocessing can also be added to reduce the amount of calculation. For example, the image is transferred from the segmentation at the pixel level to the segmentation at the superpixel level.

In achieving high-quality segmentation, it should use fewer complex formulations, which often lead to slow techniques, to avoid obstructing the actual use. For example, morphological operations can be introduced to optimize the segmentation edges. It is a good method to obtain better segmentation results by modifying the energy function. In addition, the main focus of the GrabCut model is on color features, so it is possible to explore the introduction of geometric features, texture features or some local features for segmentation.

More efficient automatic segmentation methods should be explored because manual operations have a greater impact on segmentation results. Incorrect foreground and background area selection can lead to unsatisfactory segmentation. Therefore, how to avoid the influence of human factors will receive more attention in the future. This aspect of exploration will make the algorithm more practical and more applicable. For example, with the development of deep learning, the introduction of depth models has greater possibilities for improvement.

5.2. Influence of Deep Learning

Image segmentation is a very difficult process in the field of image processing. Dense semantic segmentation has been rapidly developed and applied with the development and popularization of deep learning [98–101]. At present, there are many semantic segmentation methods based on deep learning that can successfully achieve fully automatic segmentation and have good segmentation results in a short time, for example, a series of excellent models, from Fully Convolutional Neural Network (FCN) [98] to DeepLabV3+ [101]. Not only can they realize automatic segmentation, but the segmentation effect becomes more and more significant with the development. This undoubtedly has a huge impact on traditional image-segmentation algorithms. However, deep learning algorithms require a large amount of high-quality pixel-level label information for semantic segmentation, such as the famous PASCAL VOC dataset [102] and the Cityscapes dataset [103] for urban driving scenes. They are applied to the training of deep models and can only achieve good segmentation effects for the classifications that exist in the dataset. The classification outside the dataset will be regarded as the background, which limits the applicability of the segmentation model to different scenarios. Models such as Deep GrabCut use interactive mode to improve generalization, but the segmentation accuracy is not as good as traditional segmentation methods.

Therefore, if the deep learning model wants to have an ideal segmentation effect for various salient objects, the dataset needs rich classification and a large number of images. However, labeling image data is a time-consuming and laborious process. Taggers need to spend a lot of time viewing each image, then manually adding labels or drawing annotation information such as bounding boxes. This manual marking process often takes hours or weeks to complete. In addition, for some complex tasks, such as target detection and image segmentation, the labeling process requires more time and effort, which is undoubtedly difficult. For GrabCut, it is an interactive segmentation algorithm. No matter what kind of salient objects need to be segmented, as long as the user continues to perform the correct manual interaction, it can always achieve good results. This is the advantage of GrabCut. In the improved GrabCut algorithms mentioned above, some algorithms simplify the user interaction or improve fault tolerance, and even achieve non-interactive effects in some applications. This is why GrabCut can continue to be used in the field of image segmentation. However, it is undeniable that deep learning is gradually replacing traditional image-segmentation algorithms because of the emergence and rapid expansion of a large number of datasets. It is difficult for GrabCut to be completely replaced by deep learning, and it has its own applicability in some aspects. At the same time, GrabCut can closely combine the spatial relationship of pixels and spread the interaction between pixels, so that the relationship between pixels can be well-described in image segmentation. This is generally lacking in semantic segmentation based on deep learning. Therefore, GrabCut is likely to continue to develop, and the concept of the GrabCut model can also provide fine segmentation ideas for deep learning models. In Table 4, we provide a comparison of the current characteristics between deep learning and GrabCut.

Table 4. Comparision between deep learning and GrabCut.

Characteristic	**Deep Learning**	**GrabCut**
accuracy	Higher	Secondary
problems during learning	It may require a lot of tag data and a lot of computing resources to train the model	For complex images, we need to manually specify the foreground and background
training effort	Relatively time-consuming, requiring a lot of tag data	Need to label foreground and background, but not too much tag data
applicable scenario	Process complex image tasks	Process foreground extraction in still images or videos

6. Future Work and Challenges

The GrabCut algorithm is a commonly used semi-automatic algorithm for image segmentation that combines image segmentation and interactive editing to quickly and accurately segment images. However, the GrabCut algorithm still faces challenges and issues that need to be addressed in practical applications. The future work and challenges mainly include the following aspects:

Improving the robustness and generalization ability of the algorithm: The GrabCut algorithm is sensitive to factors such as image quality, background noise, and lighting changes, and it is necessary to improve the robustness and generalization ability of the algorithm to increase the reliability and stability of the algorithm in practical applications.

Enhancing the segmentation accuracy and efficiency of the algorithm: The current GrabCut algorithm requires users to manually label the foreground and background, which consumes a lot of time and effort. Future work should focus on improving the segmentation accuracy and efficiency of the algorithm, reducing user interaction costs and further increasing the practicality of the algorithm.

Improving the application scenarios and adaptability of the algorithm: The current GrabCut algorithm is mainly used in the field of image segmentation. Future work should explore more application scenarios and fields, such as video segmentation, 3D image segmentation, medical image segmentation, etc., to increase the adaptability and practical value of the algorithm.

Developing more flexible and scalable algorithms: The success of the GrabCut algorithm is inseparable from its optimization model based on Graph Cuts. Future work should explore more flexible and scalable optimization models to meet different requirements in different scenarios. At the same time, more universal algorithm frameworks should be developed to make the algorithm more easily applied to different fields and scenarios.

In conclusion, future work should focus on improving the robustness and generalization ability of the algorithm, enhancing the segmentation accuracy and efficiency of the algorithm, improving the application scenarios and adaptability of the algorithm and developing more flexible and scalable algorithms to further improve the performance and practicality of the GrabCut algorithm in practical applications.

7. Conclusions

We provide a comprehensive review of GrabCut, an important image-segmentation method. It was first proposed by Carsten Rother, Vladimir Kolmogorov and Andrew Blake in 2004. The algorithm uses an iterative process to refine an initial foreground and background labelling of pixels based on a combination of color and texture features. In this paper, GrabCut and its improved models are explained in detail, e.g., the energy function of the weighted undirected graph is introduced and then its regional and boundary terms are analyzed. The GMM is also analyzed in detail and then the working process of GrabCut is given.

The main advantage of the GrabCut algorithm is its ability to accurately segment an object in an image with minimal user input. Unlike other segmentation methods that rely on hand-drawn masks, the GrabCut algorithm requires only a rough initial estimate of the object's location in the image. Additionally, the algorithm can handle complex object boundaries and occlusions.Therefore, GrabCut has a good segmentation performance with less resource consumption and high segmentation accuracy. However, there are also some limitations to the GrabCut algorithm. The algorithm's performance is highly dependent on the quality of the initial estimate and may require multiple iterations to achieve accurate segmentation. Additionally, the algorithm is sensitive to changes in lighting and color, which can affect the accuracy of the segmentation. Finally, the algorithm may be computationally intensive and may require significant processing power to run in real time on large images.

Author Contributions: Conceptualization, Z.W.; resources, Y.Z.; writing—original draft preparation, R.W.; writing—review and editing, Y.L.; supervision, Z.W.; project administration, Y.Z.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 61201421).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

References

1. Wang, Z.; Ma, B.; Zhu, Y. Review of level set in image segmentation. *Arch. Comput. Methods Eng.* **2021**, *28*, 2429–2446. [CrossRef]
2. Rother, C.; Kolmogorov, V.; Blake, A. "GrabCut"—Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [CrossRef]
3. Blake, A.; Rother, C.; Brown, M.; Perez, P.; Torr, P. Interactive image segmentation using an adaptive GMMRF model. In *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004*; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3021, pp. 428–441.
4. Rother, C.; Minka, T.; Blake, A.; Kolmogorov, V. Cosegmentation of Image Pairs by Histogram Matching—Incorporating a Global Constraint into MRFs. In Proceedings of the Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006.
5. Geman, S. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell* **1984**, *6*, 721–741. [CrossRef] [PubMed]
6. Selim, S.Z.; Ismail, M.A. K-Means-Type Algorithms—A Generalized Convergence Theorem And Characterization Of Local Optimality. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 81–87. [CrossRef] [PubMed]
7. Vincent, L.; Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 583–598. [CrossRef]
8. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Susstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2281. [CrossRef]
9. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
10. Li, Y.; Sun, J.; Tang, C.K.; Shum, I.Y. Lazy snapping. *Acm Trans. Graph.* **2004**, *23*, 303–308. [CrossRef]
11. An, N.; Pun, C. Iterated Graph Cut Integrating Texture Characterization for Interactive Image Segmentation. In Proceedings of the 2013 10th International Conference Computer Graphics, Imaging and Visualization, Los Alamitos, CA, USA, 6–8 August 2013; pp. 79–83.
12. Ren, D.Y.; Jia, Z.H.; Yang, J.; Kasabov, N.K. A Practical GrabCut Color Image Segmentation Based on Bayes Classification and Simple Linear Iterative Clustering. *IEEE Access* **2017**, *5*, 18480–18487. [CrossRef]
13. Li, X.L.; Liu, K.; Dong, Y.S. Superpixel-Based Foreground Extraction With Fast Adaptive Trimaps. *IEEE Trans. Cybern.* **2018**, *48*, 2609–2619. [CrossRef]
14. e Silva, R.H.L.; Machado, A.M.C. Automatic measurement of pressure ulcers using Support Vector Machines and GrabCut. *Comput. Methods Programs Biomed.* **2021**, *200*, 105867. [CrossRef]
15. Wu, S.Q.; Nakao, M.; Matsuda, T. SuperCut: Superpixel Based Foreground Extraction With Loose Bounding Boxes in One Cutting. *IEEE Signal Process. Lett.* **2017**, *24*, 1803–1807. [CrossRef]
16. Van den Bergh, M.; Boix, X.; Roig, G.; Van Gool, L. SEEDS: Superpixels Extracted Via Energy-Driven Sampling. *Int. J. Comput. Vis.* **2015**, *111*, 298–314. [CrossRef]
17. Long, J.W.; Feng, X.; Zhu, X.F.; Zhang, J.X.; Gou, G.L. Efficient Superpixel-Guided Interactive Image Segmentation Based on Graph Theory. *Symmetry* **2018**, *10*, 169. [CrossRef]
18. Zhou, X.N.; Wang, Y.N.; Zhu, Q.; Xiao, C.Y.; Lu, X. SSG: Superpixel segmentation and GrabCut-based salient object segmentation. *Vis. Comput.* **2019**, *35*, 385–398. [CrossRef]
19. Borji, A.; Cheng, M.M.; Hou, Q.; Jiang, H.; Li, J. Salient object detection: A survey. *Comput. Vis. Media* **2019**, *5*, 117–150. [CrossRef]
20. Fu, Y.; Cheng, J.; Li, Z.L.; Lu, H.Q. Saliency Cuts: An Automatic Approach to Object Segmentation. In Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; Volumes 1–6, pp. 696–699.
21. Kim, K.S.; Yoon, Y.J.; Kang, M.C.; Sun, J.Y.; Ko, S.J. An Improved GrabCut Using a Saliency Map. In Proceedings of the 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), Tokyo, Japan, 7–10 October 2014; pp. 317–318.
22. Li, S.Z.; Ju, R.; Ren, T.W.; Wu, G.S. Saliency Cuts Based on Adaptive Triple Thresholding. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 4609–4613.
23. Cheng, M.M.; Mitra, N.J.; Huang, X.L.; Torr, P.H.S.; Hu, S.M. Global Contrast Based Salient Region Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 569–582. [CrossRef]
24. Gupta, N.; Jalal, A.S. A robust model for salient text detection in natural scene images using MSER feature detector and Grabcut. *Multimed. Tools Appl.* **2019**, *78*, 10821–10835. [CrossRef]
25. Peng, Y.; Zhang, Z.; He, G.; Wei, M. An Improved GrabCut Method Based on a Visual Attention Model for Rare-Earth Ore Mining Area Recognition with High-Resolution Remote Sensing Images. *Remote Sens.* **2019**, *11*, 987. [CrossRef]
26. Niu, Y.Z.; Su, C.R.; Guo, W.Z. Salient Object Segmentation Based on Superpixel and Background Connectivity Prior. *IEEE Access* **2018**, *6*, 56170–56183. [CrossRef]
27. Wang, Y.; Ren, T.; Zhong, S.H.; Liu, Y.; Wu, G. Adaptive saliency cuts. *Multimed. Tools Appl.* **2018**, *77*, 22213–22230. [CrossRef]
28. Vicente, S.; Kolmogorov, V.; Rother, C. Joint optimization of segmentation and appearance models. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision (Iccv), Kyoto, Japan, 29 September–2 October 2009; pp. 755–762.
29. Komodakis, N.; Paragios, N.; Tziritas, G. MRF Energy Minimization and Beyond via Dual Decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 531–552. [CrossRef] [PubMed]
30. Tang, M.; Gorelick, L.; Veksler, O.; Boykov, Y. GrabCut in One Cut. In Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 1769–1776.

31. Gao, Z.S.; Shi, P.; Karimi, H.R.; Pei, Z. A mutual GrabCut method to solve co-segmentation. *EURASIP J. Image Video Process.* **2013**, *2013*, 1–11. [CrossRef]
32. Zhou, H.L.; Zheng, J.M.; Wei, L. Texture aware image segmentation using graph cuts and active contours. *Pattern Recognit.* **2013**, *46*, 1719–1733. [CrossRef]
33. Cheng, M.M.; Prisacariu, V.A.; Zheng, S.; Torr, P.H.S.; Rother, C. DenseCut: Densely Connected CRFs for Realtime GrabCut. *Comput. Graph. Forum* **2015**, *34*, 193–201. [CrossRef]
34. Guan, Q.; Hua, M.; Hu, H.G. A Modified Grabcut Approach for Image Segmentation Based on Local Prior Distribution. In Proceedings of the 2017 International Conference on Wavelet Analysis And Pattern Recognition (ICWAPR), Ningbo, China, 9–12 July 2017; pp. 122–126.
35. Hua, S.Y.; Shi, P. GrabCut Color Image Segmentation Based on Region of Interest. In Proceedings of the 2014 7th International Congress on Image And Signal Processing (CISP 2014), Dalian, China, 14–16 October 2014; pp. 392–396.
36. Yong, Z.; Jiazheng, Y.; Hongzhe, L.; Qing, L. GrabCut image segmentation algorithm based on structure tensor. *J. China Univ. Posts Telecommun.* **2017**, *24*, 38–47. [CrossRef]
37. Yu, H.K.; Zhou, Y.J.; Qian, H.; Xian, M.; Wang, S. Loosecut: Interactive Image Segmentation with Loosely Bounded Boxes. In Proceedings of the 2017 24th IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3335–3339.
38. He, K.; Wang, D.; Tong, M.; Zhang, X. Interactive Image Segmentation on Multiscale Appearances. *IEEE Access* **2018**, *6*, 67732–67741. [CrossRef]
39. He, K.; Wang, D.; Tong, M.; Zhu, Z. An Improved GrabCut on Multiscale Features. *Pattern Recognit.* **2020**, *103*, 107292. [CrossRef]
40. Fu, R.G.; Li, B.; Gao, Y.H.; Wang, P. Fully automatic figure-ground segmentation algorithm based on deep convolutional neural network and GrabCut. *IET Image Process.* **2016**, *10*, 937–942. [CrossRef]
41. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
42. Nver, H.M.; Ayan, E. Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. *Diagnostics* **2019**, *9*, 72.
43. Zhang, N.C. Improved GrabCut Algorithm Based on Probabilistic Neural Network. *Adv. Laser Optoelectron.* **2021**, *58*, 0210024. [CrossRef]
44. Kim, G.; Sim, J.Y. Depth Guided Selection of Adaptive Region of Interest for Grabcut-Based Image Segmentation. In Proceedings of the 2016 Asia-Pacific Signal And Information Processing Association Annual Summit And Conference (APSIPA), Jeju, Republic of Korea, 13–16 December 2016.
45. Sanguesa, A.A.; Jorgensen, N.K.; Larsen, C.A.; Nasrollahi, K.; Moeslund, T.B. Initiating GrabCut by Color Difference for Automatic Foreground Extraction of Passport Imagery. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016.
46. Deng, L.L. Pre-detection Technology of Clothing Image Segmentation Based on GrabCut Algorithm. *Wirel. Pers. Commun.* **2018**, *102*, 599–610. [CrossRef]
47. Orchard, M.T.; Bouman, C.A. Color Quantization Of Images. *IEEE Trans. Signal Process.* **1991**, *39*, 2677–2690. [CrossRef]
48. Khattab, D.; Ebied, H.M.; Hussein, A.S.; Tolba, M.F. Color Image Segmentation Based on Different Color Space Models Using Automatic GrabCut. *Sci. World J.* **2014**, *2014*, 126025. [CrossRef]
49. Kohonen, T.; Oja, E.; Simula, O.; Visa, A.; Kangas, J. Engineering applications of the self-organizing map. *Proc. IEEE* **1996**, *84*, 1358–1384. [CrossRef]
50. Khattab, D.; Ebied, H.M.; Hussein, A.S.; Tolba, M.F. Automatic GrabCut for Bi-label Image Segmentation Using SOFM. *Intell. Syst. Vol 2 Tools, Archit. Syst. Appl.* **2015**, *323*, 579–592.
51. Wang, P.H. Pattern-Recognition with Fuzzy Objective Function Algorithms-Bezdek, Jc. *Siam Rev.* **1983**, *25*, 442.
52. Khattab, D.; Ebeid, H.M.; Tolba, M.F.; Hussein, A.S. Clustering-based Image Segmentation using Automatic GrabCut. In Proceedings of the International Conference on Informatics and Systems (INFOS 2016), Cairo, Egypt, 9–11 May 2016; pp. 95–100.
53. Ye, H.J.; Liu, C.Q.; Niu, P.Y. Cucumber appearance quality detection under complex background based on image processing. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 193–199.
54. Sun, S.; Jiang, M.; He, D.; Long, Y.; Song, H. Recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm. *Biosyst. Eng.* **2019**, *187*, 201–213. [CrossRef]
55. Deshpande, A.; Dahikar, P.; Agrawal, P. An Experiment with Random Walks and GrabCut in One Cut Interactive Image Segmentation Techniques on MRI Images. In *Proceedings of the International Conference on Computational Vision and Bio Inspired Computing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 993–1008.
56. Jiang, F.; Pang, Y.; Lee, T.N.; Liu, C. Automatic object segmentation based on grabcut. In *Proceedings of the Science and Information Conference*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 350–360.
57. Sallem, N.K.; Devy, M. Extended GrabCut for 3D and RGB-D Point Clouds. *Adv. Concepts Intell. Vis. Syst. Acivs* **2013**, *8192*, 354–365.
58. Dietterich, T.G.; Lathrop, R.H.; LozanoPerez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.* **1997**, *89*, 31–71. [CrossRef]

59. Wu, J.J.; Zhao, Y.B.; Zhu, J.Y.; Luo, S.W.; Tu, Z.W. MILCut: A Sweeping Line Multiple Instance Learning Paradigm for Interactive Image Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 256–263.
60. Lee, G.; Lee, S.; Kim, G.; Park, J.; Park, Y. A Modified GrabCut Using a Clustering Technique to Reduce Image Noise. *Symmetry* **2016**, *8*, 64. [CrossRef]
61. Niu, S.X.; Chen, G.S. The Improvement of the Processes of a Class of Graph-Cut-Based Image Segmentation Algorithms. *Ieice Trans. Inf. Syst.* **2016**, *E99d*, 3053–3059. [CrossRef]
62. Lu, Y.W.; Jiang, J.G.; Qi, M.B.; Zhan, S.; Yang, J. Segmentation method for medical image based on improved GrabCut. *Int. J. Imaging Syst. Technol.* **2017**, *27*, 383–390. [CrossRef]
63. Rajchl, M.; Lee, M.C.H.; Oktay, O.; Kamnitsas, K.; Passerat-Palmbach, J.; Bai, W.; Damodaram, M.; Rutherford, M.A.; Hajnal, J.V.; Kainz, B.; et al. DeepCut: Object Segmentation From Bounding Box Annotations Using Convolutional Neural Networks. *IEEE Trans. Med. Imaging* **2017**, *36*, 674–683. [CrossRef]
64. Lee, J.; Kim, D.W.; Won, C.S.; Jung, S.W. Graph Cut-Based Human Body Segmentation in Color Images Using Skeleton Information from the Depth Sensor. *Sensors* **2019**, *19*, 393. [CrossRef]
65. Xu, N.; Price, B.L.; Cohen, S.; Yang, J.; Huang, T.S. Deep GrabCut for Object Selection. *arXiv* **2017**, arXiv:1707.00243.
66. Kalshetti, P.; Bundele, M.; Rahangdale, P.; Jangra, D.; Chattopadhyay, C.; Harit, G.; Elhence, A. An interactive medical image segmentation framework using iterative refinement. *Comput. Biol. Med.* **2017**, *83*, 22–33. [CrossRef]
67. Baracho, S.F.; Pinheiro, D.J.L.L.; de Godoy, C.M.G.; Coelho, R.C. A segmentation method for myocardial ischemia/infarction applicable in heart photos. *Comput. Biol. Med.* **2017**, *87*, 285–301. [CrossRef]
68. Liu, Y.H.; Cao, F.L.; Zhao, J.W.; Chu, J.J. Segmentation of White Blood Cells Image Using Adaptive Location and Iteration. *IEEE J. Biomed. Health Inf.* **2017**, *21*, 1644–1655. [CrossRef]
69. Wu, S.B.; Yu, S.D.; Zhuang, L.; Wei, X.H.; Sak, M.; Duric, N.; Hu, J.N.; Xie, Y.Q. Automatic Segmentation of Ultrasound Tomography Image. *Biomed Res. Int.* **2017**, *2017*, 2059036. [CrossRef]
70. Yu, S.D.; Wu, S.B.; Zhuang, L.; Wei, X.H.; Sak, M.; Neb, D.; Hu, J.N.; Xie, Y.Q. Efficient Segmentation of a Breast in B-Mode Ultrasound Tomography Using Three-Dimensional GrabCut(GC3D). *Sensors* **2017**, *17*, 1827. [CrossRef]
71. Jaisakthi, S.M.; Mirunalini, P.; Aravindan, C. Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. *IET Comput. Vis.* **2018**, *12*, 1088–1095. [CrossRef]
72. Mao, J.F.; Wang, K.H.; Hu, Y.H.; Sheng, W.G.; Feng, Q.X. GrabCut algorithm for dental X-ray images based on full threshold segmentation. *IET Image Process.* **2018**, *12*, 2330–2335. [CrossRef]
73. Xiao, C.F.; Li, W.F.; Deng, H.; Chen, X.; Yang, Y.; Xie, Q.W.; Han, H. Effective automated pipeline for 3D reconstruction of synapses based on deep learning. *BMC Bioinform.* **2018**, *19*, 263. [CrossRef] [PubMed]
74. Frants, V.; Agaian, S. Dermoscopic image segmentation based on modified GrabCut with octree color quantization. In *Proceedings of the Mobile Multimedia/Image Processing, Security, and Applications 2020*; SPIE: Bellingham, DC, USA, 2020; Volume 11399, pp. 119–130.
75. Saeed, S.; Abdullah, A.; Jhanjhi, N.; Naqvi, M.; Masud, M.; AlZain, M.A. Hybrid GrabCut Hidden Markov Model for Segmentation. *Comput. Mater. Contin.* **2022**, *72*, 851–869. [CrossRef]
76. Yoruk, U.; Hargreaves, B.A.; Vasanawala, S.S. Automatic renal segmentation for MR urography using 3D-GrabCut and random forests. *Magn. Reson. Med.* **2018**, *79*, 1696–1707. [CrossRef]
77. Wei, Z.; Liang, D.; Zhang, D.; Zhang, L.; Geng, Q.; Wei, M.; Zhou, H. Learning calibrated-guidance for object detection in aerial images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 2721–2733. [CrossRef]
78. Zhu, H.; Li, P.; Xie, H.; Yan, X.; Liang, D.; Chen, D.; Wei, M.; Qin, J. *I Can Find You! Boundary-Guided Separated Attention Network for Camouflaged Object Detection*; AAAI: Washington, DC, USA, 2022.
79. Kang, B.; Liang, D.; Mei, J.; Tan, X.; Zhou, Q.; Zhang, D. Robust RGB-T Tracking via Graph Attention-Based Bilinear Pooling. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**. [CrossRef]
80. Lin, C.Y.; Muchtar, K.; Yeh, C.H. Robust Techniques for Abandoned and Removed Object Detection Based on Markov Random Field. *J. Vis. Commun. Image Represent.* **2016**, *39*, 181–195. [CrossRef]
81. Zhang, H.; Song, A.G. A map-based normalized cross correlation algorithm using dynamic template for vision-guided telerobot. *Adv. Mech. Eng.* **2017**, *9*, 1687814017728839. [CrossRef]
82. Xu, S.X.; Zhu, Q.Y. Seabird image identification in natural scenes using Grabcut and combined features. *Ecol. Inf.* **2016**, *33*, 24–31. [CrossRef]
83. Hao, Y.P.; Wei, J.; Jiang, X.L.; Yang, L.; Li, L.C.; Wang, J.K.; Li, H.; Li, R.H. Icing Condition Assessment of In-Service Glass Insulators Based on Graphical Shed Spacing and Graphical Shed Overhang. *Energies* **2018**, *11*, 318. [CrossRef]
84. Li, S.; Liu, Y.; Zhao, Q.; Feng, Z. Learning residue-aware correlation filters and refining scale estimates with the grabcut for real-time UAV tracking. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; pp. 1238–1248.
85. Salau, A.O.; Yesufu, T.K.; Ogundare, B.S. Vehicle plate number localization using a modified GrabCut algorithm. *J. King Saud Univ.-Comput. Inf. Sci.* **2021**, *33*, 399–407. [CrossRef]
86. Shi, H.; Zhao, D. License Plate Localization in Complex Environments Based on Improved GrabCut Algorithm. *IEEE Access* **2022**, *10*, 88495–88503. [CrossRef]

87. Brkic, K.; Hrkac, T.; Kalafatic, Z. Protecting the privacy of humans in video sequences using a computer vision-based de-identification pipeline. *Expert Syst. Appl.* **2017**, *87*, 41–55. [CrossRef]
88. Dong, L.; Feng, N.; Mao, M.; He, L.; Wang, J. E-GrabCut: An economic method of iterative video object extraction. *Front. Comput. Sci.* **2017**, *11*, 649–660. [CrossRef]
89. Kang, F.; Wang, C.; Li, J.; Zong, Z. A Multiobjective Piglet Image Segmentation Method Based on an Improved Noninteractive GrabCut Algorithm. *Adv. Multimed.* **2018**, *2018*, 1083876. [CrossRef]
90. Qi, F.; Wang, Y.; Tang, Z. Lightweight Plant Disease Classification Combining GrabCut Algorithm, New Coordinate Attention, and Channel Pruning. *Neural Process. Lett.* **2022**, *54*, 5317–5331. [CrossRef]
91. Zhang, S.; Huang, W.; Wang, Z. Combing modified Grabcut, K-means clustering and sparse representation classification for weed recognition in wheat field. *Neurocomputing* **2021**, *452*, 665–674. [CrossRef]
92. Tan, K.; Zhang, Y.; Tong, X. Cloud Extraction from Chinese High Resolution Satellite Imagery by Probabilistic Latent Semantic Analysis and Object-Based Machine Learning. *Remote Sens.* **2016**, *8*, 963. [CrossRef]
93. Yamasaki, Y.; Migita, M.; Koutaki, G.; Toda, M.; Kishigami, T. ISHIGAKI Region Extraction Using Grabcut Algorithm for Support of Kumamoto Castle Reconstruction. In Proceedings of the International Workshop on Frontiers of Computer Vision, Daegu, Republic of Korea, 22–23 February 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 106–116.
94. Zhang, K.; Chen, H.; Xiao, W.; Sheng, Y.; Su, D.; Wang, P. Building Extraction from High-Resolution Remote Sensing Images Based on GrabCut with Automatic Selection of Foreground and Background Samples. *Photogramm. Eng. Remote Sens.* **2020**, *86*, 235–245. [CrossRef]
95. Liu, T.; Yuan, Z.J.; Sun, J.A.; Wang, J.D.; Zheng, N.N.; Tang, X.O.; Shum, H.Y. Learning to Detect a Salient Object. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 353–367. [PubMed]
96. Wang, Z.; Wang, E.; Zhu, Y. Image segmentation evaluation: A survey of methods. *Artif. Intell. Rev.* **2020**, *53*, 5637–5674. [CrossRef]
97. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef]
98. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
99. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
100. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
101. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
102. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]
103. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NE, USA, 26 June–1 July 2016.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.