A Method for Object Extraction from Crop Image Based on Visual Saliency

Yuting Wang¹, Zhouyu Liao¹*
¹School of Big Data and Computer Science, Hechi University, Yizhou, Hechi, Guangxi, China
*email: 07040@hcnu.edu.cn

Abstract. In order to solve the problem of accurate extraction of crop targets in natural background, an image target extraction method based on visual saliency is proposed. Firstly, an image is divided into multiple super pixels. Secondly, the significance detection method of ‘background first’ is used to calculate the significance of the whole image. Thirdly, the optimal data source is generated according to the characteristics of the high saliency region. Finally, the optimal data source is divided into target region and background region, and the crop target is extracted from the original image. Experimental results on a large number of crop images show that this method can extract targets accurately and completely.

1. Introduction
With the development of agricultural science and technology, automatic machines are more and more widely used in agricultural production. For example: automatic irrigation of crops, automatic picking of crops, automatic monitoring of diseases and pests, etc. Much of them are inseparable from the support of digital image processing technology. Just as, the automatic picking technology of crops completely depends on the accurate extraction of object.

At present, many researchers have adopted different target extraction algorithms. Yue Y J et al [1] used the improved cascade RCNN to detect tomato fruits in the greenhouse. Fan Z Z et al [2] combined with deep learning and traditional methods, proposed a fruit detection algorithm integrating color and fruit diameter features for apple fruit detection. Yin H et al [3] proposed a flower image classification method based on selective convolution feature fusion. This method uses an unsupervised way to locate the significant area in the flower image, removes the interference of the background and noise to the flower target. Zhang H Q et al [4] introduced the region information near the threshold point into the segmentation algorithm according to the characteristics of tomato fruit image.

Most of the existing target extraction methods are mainly aimed at a specific flower or fruit target, which does not have universal applicability; Secondly, the complex background brings serious interference to the accurate segmentation of targets; Thirdly, the strong reflection on the target surface also seriously affects the recognition of the target area. Therefore, the research on how to segment different kinds and colors of flowers and fruits from various natural background is of great significance to improve the practicability and universal usability of the algorithm.

2. Method and steps

2.1. SSLIC super pixel segmentation algorithm
The algorithm in this paper is based on super-pixel when calculating saliency map.
The super-pixel segmentation algorithm adopts SSLIC algorithm. It is called SSLIC because it is very similar to SLIC algorithm.

The division of super pixels in SSLIC algorithm depends on the color similarity and proximity between pixels. The color similarity is calculated in RGB color space, as shown in formula (1); The proximity is calculated by formula (2); The comprehensive measurement formula is shown in formula (3).

\[
d_{rgb} = \sqrt{(R_i - R_k)^2 + (G_i - G_k)^2 + (B_i - B_k)^2}
\]

\[
d_{xy} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}
\]

\[
dist = d_{rgb} + k \cdot d_{xy}
\]

In formula (1), \(d_{rgb}\) represents the color similarity between pixel \(i\) and the center of super pixel \(k\); In formula (2), \(d_{xy}\) represents the proximity degree from pixel \(i\) to the center of super-pixel \(k\); \(dist\) represents the comprehensive similarity from pixel \(i\) to the center of super-pixel \(k\).

2.2. Generate saliency map

In this paper, the detection method of ‘background first’ is used to calculate the significance. The steps of it are as follows:

Use SSLIC algorithm to segment an image into multiple super pixels.

Get the real initial background area. For an image, the focus of vision is usually in its center, the target area is close to the center and the background area is close to the edge of the image. So, preset these super pixels on the four sides of the super pixel map as the background area, as shown in Figure 2 (Figure 1 shows the super pixel map). However, the preset background area may contain the target area, as shown in the red area in Figure 3. Therefore, we need to exclude the suspected target area from the preset background area, so as to obtain the real initial background area, as shown in Figure 4.

![Fig. 1 Super pixel map](image1.png) ![Fig. 2 Preset initial background area](image2.png) ![Fig. 3 Incorrectly divided area](image3.png) ![Fig. 4 Real initial background area](image4.png)

It is necessary to calculate the average distance from each super pixel to all other super pixels in the preset background area, then calculate the total average value of these average distances. Conditions for excluding the suspected target area are as follows formula (4) (5). If both are satisfied, the super pixel is excluded, otherwise the super pixel still belongs to the background.

Formula (4) indicates that when the average distance from one super pixel \(n\) to other super pixels in the background is greater than the total average distance \(sum_{avg}\).

Formula (5) indicates that the R or B component value of the super-pixel \(n\) is greater than its G component value. The purpose of this condition is to avoid mistaking the white sky or other rapidly changing background factors as the target area.

\[
avg_n > sum_{avg}
\]
Generate saliency map. Calculate the shortest distance (Euclidean distance) from each super pixel in the graph to the initial background area, and use this distance to represent the saliency of each super pixel. The greater the distance, the higher the significance value. The saliency map of the image is obtained according to the saliency of each super-pixel.

2.3. Optimal data source
According to the conclusion of reference [5], if we can find a data source suitable for target segmentation (optimal data source), we can easily segment the target accurately and completely from an image.

The steps to generate the optimal data source are as follows:

1) Calculate the average value of the saliency of all super pixels according to the saliency map. Super pixels greater than this average belong to the target, and those less than this average belong to the background.

2) Calculate the average R, G and B values of the target area and the background area respectively, and find out the component to which the maximum average value belongs, that is the optimal component.

3) The optimal data source is generated according to the component to which the maximum average value of the target and background belongs. In order to obtain the optimal data source, it is necessary to improve the significance of the target region and suppress the significance of the background region. The method to obtain the optimal data source is shown in formula (6, 7):

\[
W_{rgb_{r,c}} = (1 + w_{XZ} * X_{r,c})(f_g * r_{gb_{r,c,best}} - b_g * R_{gb_{r,c,best}}) * w_R
\]  

In formula (6), \(W_{rgb_{r,c}}\) represents the optimal value of the pixel point in column \(c\) of row \(r\) in the image, \(w_{XZ}\) is the saliency coefficient, \(X_{r,c}\) is the saliency value of the pixel in column \(c\) of row \(r\), \(r_{gb_{r,c,best}}\) represents the component of the point to which the maximum average value is taken (for example, if the average value of the target is the largest under the R component, the value of the R component is taken), \(f_g\) is the foreground coefficient, \(b_g\) is the background coefficient, and \(r_{gb_{r,c,best}}\) represents the value of the point under the G component, \(w_R\) represents a coefficient used to amplify the value of R at a special time.

\[
\begin{align*}
\text{if (best == G)} & \quad f_g = 1, b_g = 0, w_R = 1, w_{XZ} = 0.5 \\
\quad \text{if (avg \_fg_{best} >= avg \_bg_{best} &\& avg \_fg_{best} > 110 &\& avg \_bg_{best} > 110)} & \quad \begin{cases} 
  f_g = \frac{\text{avg \_fg}_{best}}{\text{avg \_bg}_{best}} \\
  f_b = \frac{\text{avg \_bg}_{best}}{\text{avg \_bg}_{best}} \\
  w_R = 1, w_{XZ} = 0
\end{cases} \\
\text{if (best != G)} & \quad \text{if (avg \_fg_{best} < avg \_bg_{best})} \\
\quad & \quad \begin{cases} 
  f_g = \frac{\text{avg \_fg}_{best}}{\text{avg \_bg}_{best}} \\
  f_b = \frac{\text{avg \_bg}_{best}}{\text{avg \_bg}_{best}} \\
  w_R = \frac{255}{r_{gb_{r,c,best}}}, w_{XZ} = 0
\end{cases} \\
\quad \text{else} & \quad f_g = 1, b_g = 1, w_R = 1, w_{XZ} = 0
\end{align*}
\]  

In formula (7), the values of each coefficient in formula (6) are given when \(\text{best}\) takes different values. \(\text{avg \_fg}_{best}\) represents the average significance of the target area under the optimal
component, \( \text{avg}_{bG_{best}} \) represents the average significance of the background area under the optimal component, \( \text{avg}_{fG} \) represents the average significance of the target area under the G component.

2.4. Target extraction method

To extract the target, we must first get the accurate region where the target is located. The target area is obtained by segmenting the optimal data source through FCM algorithm. That is, FCM algorithm is used to divide the optimal data sources into two categories, one belongs to the target and the other belongs to the background. Then the corresponding target region in the original image is extracted according to the segmentation result. Since FCM algorithm is introduced in many articles, it will not be repeated here.

2.5. Overall algorithm flow

According to the previous sections, the overall flow of the algorithm is shown in Figure 5.

![Fig. 5 Overall algorithm flow chart](image)

3. Experimental results and analysis

In order to verify the performance of this algorithm, a large number of experiments are carried out, and 10 images with different colors and different background factors are selected as experimental representatives. In these images, there are many factors that will affect the accurate extraction of targets, such as: branches and leaves, soil, fallen petals, cement buildings, dried branches, plastic film on the ground, water droplets, mud spots, reflections from leaf edges, water droplets, red leaf buds, mud spots, falling petals, dry white branches, etc. The programming platform is Visual C++ and MATLAB. The experimental results are compared with the effects of SUN [6] and FT [7], and the results are shown in the Table 1.

| Tab. 1 Comparison of experimental results on flower and fruit images |
|--------------------------------------------------|
| 
| No.1 | No.2 | No.3 | No.4 | No.5 | No.6 | No.7 | No.8 | No.9 | No.10 |
|------|------|------|------|------|------|------|------|------|------|
| Original image | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) |
| FT | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) |
| SUN | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) |
| Optimal data source | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) |
| Extracted target | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) | ![Image](image) |

It can be seen from table 1, the algorithm in this paper can highlight the target area more completely and accurately, and suppress the significance of the background area at the same time. The last row in the table is the target extracted by using the saliency map in this paper, and its effect is good.
In order to objectively evaluate the effect of this algorithm, we use precision-recall curve to quantitatively analyze and compare this algorithm with other algorithms. Firstly, adjust the saliency map obtained by various algorithms to the range of 0 to 255. Then, select the threshold from 0 to 255 to binarize the saliency map of each algorithm, and compare the results with the ground truth, so as to calculate the precision and recall, and draw the precision-recall curve. The results are shown in Figure 6. It can be seen from Figure 6 that the algorithm in this paper has better processing effect on crop images than SUN and FT algorithms.

4. Conclusions
In this paper, a method of using saliency map to obtain image prior knowledge is proposed, and the prior knowledge is applied to the extraction of flowers and mature fruits, and good experimental results are obtained. Due to the variety of target colors and the ever-changing background factors, it is even more difficult to make an algorithm adapt to all environments, so the method in this paper needs to be further improved. How to design a more adaptive target extraction method is the next direction work. I would like to express my sincere thanks to the author who provided the source code of the comparison algorithm.

Acknowledgments
This paper is one of the phased achievements of the 2020 basic ability improvement project Research on Human Abnormal Behavior Detection in Intelligent Video Surveillance System (2020ky15015) for middle-aged and young teachers in colleges and universities in Guangxi.

References
[1] Yue Y J, Sun B Y, Wang H J and Zhao H 2021 Object Detection of Tomato Fruit Based on Cascade RCNN J. Science Technology and Engineering 21 2387
[2] Fan Z Z, Liu Q, Chai J W, Xiao F and Li H F 2020 Apple detection and grading based on color and fruit-diameter J. Computer Engineering & Science 42 1599
[3] Yin H, Fu X, Zeng J X, Duan B and Chen Y 2019 Flower image classification with selective convolutional descriptor aggregation J. Image and Graphics 24 762
[4] Zhang H Q, Liu Y and He M 2015 The Research of Tomato Image Segmentation Based on Machine Vision J. Agricultural Mechanization Research 37 58
[5] Wang Y T and Liao Z Y 2016 Influence of different color components on image segmentation results J. Software Guide 15 221
[6] Zhang L, Tong M, Marks H, Tim K, Shan H and Cottrell G 2008 SUN: a bayesian framework for saliency using natural statistics J. Vision 8(1)
[7] Achanta R, Hemami S, Estrada F and Süsstrunk S 2009 Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (Miami) vol 4 (USA: IEEE Press) p 1597