

# A distribution balance-based data augmentation method for light-trap pest detection

Yue Teng<sup>#a,b</sup>, Rujing Wang<sup>a,b</sup>, Ziliang Huang<sup>a,b</sup>, Shijian Zheng<sup>c</sup>, Qiong Zhou<sup>a,b</sup>, Jie Zhang<sup>\*a</sup>

<sup>a</sup>Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Science, Hefei 230031, China; <sup>b</sup>Science Island Branch of Graduate School, University of Science and Technology of China, Hefei 230026, China; <sup>c</sup>College of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China

## ABSTRACT

Agriculture pest disaster is one of the most important reasons that reduce grain yield. Accurate recognition and detection are the core of Integrated Pest Management (IPM). Existing deep learning-based methods improve the capacity of feature extraction, but ignore the imbalance of object number and size distribution result in insufficient performance. Therefore, we design a Joint Balance-Distribution Oriented Composition (JBDOC) to detect multi-class pests with the synthetic dataset. Object bounding boxes and white background boards are used to construct the balanced synthetic dataset for training the convolutional neural network (CNN). Our JBDOC solves the distribution imbalance without methods restriction and improves the test performance without extra time consumption. We combine the JBDOC with current popular detection methods to verify the validity. Experimental results show that the JBDOC greatly improves the performance of deep learning-based detectors in the pest field.

**Keywords:** Agriculture pest detection, convolutional neural network, data augmentation, data balance

## 1. INTRODUCTION

Crop pest disaster is one of the important reasons that affect grain yield. The foundation of Integrated Pest Management (IPM) is accurately detecting pests<sup>1</sup>. Manual pest detection is time-consuming and laborious work. In recent years, with the rapid development of computer vision (CV), researchers tend to automatically recognize and locate pests with images. The traditional methods based on machine learning extract features manually for specific pest detection. However, the hand-designed feature extraction method limits the generalization ability of pest detection. In addition, different environments (background, illumination, and pest attitude) present challenges for pest detection.

Since the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)<sup>2</sup>, deep learning-based methods have achieved state-of-the-art performance in the field of general object detection. Ren et al. use region proposal network (RPN) instead of selective search<sup>3</sup> to design a two-stage baseline network Faster R-CNN<sup>4</sup>. Cai et al. use coarse-to-fine intersection over Union (IoU) threshold to improve the performance of Faster R-CNN<sup>5</sup>. Tian et al. design a point regression-based one-stage detector and use the center-ness branch to filtrate low-quality prediction bounding boxes<sup>6</sup>. Lin et al. use the focal loss to balance the positive and negative samples in the one-stage paradigm<sup>7</sup>.

Due to the rapid development of general object detection, researchers attempt to detect pests with deep learning-based methods. Liu et al. designed Channel-Spatial Attention to improve the feature extraction capability of CNN for small-size pest identification<sup>8</sup>. Jiao et al. use the feature fusion method to solve the problem of multiscale pest detection in the light-trap dataset Pest-21<sup>9</sup>. Wang et al. fuse context-aware information (longitude, latitude, temperature, humidity, etc) to improve the performance of in-field pest detection<sup>10</sup>. Eayan et al. use three CNNs for pest identification simultaneously and the genetic algorithm is employed to select the network weight<sup>11</sup>. Liu et al. design a SAFS network to detect pests by selecting features using a region-guided and scale-aware method<sup>12</sup>.

Existing pest detections improve the feature extraction capability from deep learning-based object detections. However, the balance of class number and size distribution are ignored result in insufficient practical performance. Firstly, due to the emergence of pests being random and irregular, the category number is uneven. The CNNs tend to recognize objects

#yueteng@mail.ustc.edu.cn; \*zhangjie@iim.ac.cn

with high frequency. Secondly, due to the different feature extraction levels of multi-scale objects, the performance of small-size pests is inferior to large-size pests. To solve the above-mentioned problems (category imbalance and size distribution imbalance), we propose a Joint Balance-Distribution Oriented Composition (JBDOC) with the synthetic dataset for light-trap pest detection.

## 2. MATERIALS AND METHODS

### 2.1 Original data collection

The pest images are collected automatically with the device from Jiaduo Company. The data collection device is distributed in the field environment of Anhui Province and consists of the pest light-trap device, camera, white background board, and pest cleaning device, as shown in Figure 1. Crop pests are attracted and stunned by the light-trap device. Then, pests fall to the white background board with grid lines. The pest camera collects images and the cleaning device cleans the white background board every 30 seconds. We invite agriculture experts to annotate data using Lableme software. After data cleaning and annotation, we constructed our light-trap pest dataset (LPD) includes 18585 JPEG images with 26 classes. The resolution of our LPD is 2592×1944 and the annotation files format is XML like PASCAL VOC. We divide the dataset into a trainset with 14868 images and a test set with 3717 images.



Figure 1. light-trap pest collection device. (1) Pest light-trap device; (2) interior structure; (3) camera; (4) pest cleaning device; (5) white background board; (6) original light-trap dataset.

### 2.2 Joint balance-distribution oriented composition

The deep learning-based detector shows poor ability in recognizing pests with small-size and few-shot characters. We use the Faster R-CNN to explore the relation of Average Precision (AP) with object size and number in our LPD, as shown in Figure 2. Due to the fact that the AP has a positive correlation with the object size and number, we design a Joint Balance-Distribution Oriented Composition (JBDOC) with the synthetic dataset to solve the imbalance of category and size distribution.

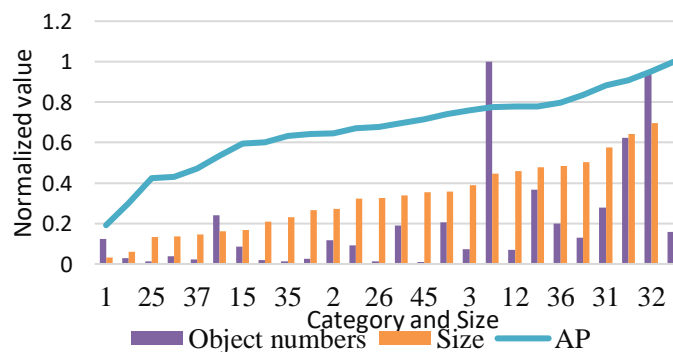


Figure 2. The performance of category-size distribution.

Since the light-trap device cleans the pests on the background board every 30 seconds, we can easily select a background board image without pests from the original data. We use the white background board and the bounding box in the original train set to expand the data. Specifically, we hope to expand the data as much as possible for the small-size and few-shot categories. Therefore, we use equation (1) to determine the object number of the synthetic dataset for each category:

$$E_C = \partial \cdot (Max(N_{DC}) - N_C) + (1 - \partial) \frac{mSize_D}{mSize_C} \cdot N_C \quad (1)$$

where  $E_C$  represents the number of bounding boxes to be expanded for category  $C$ ,  $\lfloor \cdot \rfloor$  represents the rounding down operation,  $N_C$  represents the number of existing bounding boxes of class  $C$ ,  $Max(N_{DC})$  represents the category with the most number in the train set,  $mSize_D$  represents the average size of class  $C$ ,  $mSize_C$  represents the mean object size of LPD,  $\partial$  is a scale factor, and in this paper  $\partial = 0.5$ .

Although the white background board of our dataset is conducive to reconstruct data, the extracted object bounding boxes exist black grid lines. We use the Grabcut algorithm<sup>13</sup> to remove excess grid lines. To calculate the mask, the Grabcut algorithm needs to artificially present the area  $B$  with the foreground object and the whole image  $B_{in}$ . To solve the defects of the Grabcut, we designed an Automatic Grabcut Algorithm (AGA). Specifically,  $\forall B(x, y, w, h) \in T$ , we have  $B_{in} = B(x, y, 2w, 2h)$ . Where  $B(x, y, w, h)$  indicates the bounding box coordinates that have been labeled,  $T$  represents the train set from the original dataset,  $B_{in}$  represents the input image of the Grabcut algorithm. For  $B_{in}$ , we choose only the bounding box without other objects in the range of  $\forall B(x, y, w, h) - B(x, y, 2w, 2h)$  to avoid overlapped objects sneaking into AGA. We take  $B_{in}$  as the whole image and  $B$  as the object area parameter, then using the Grabcut algorithm to calculate the mask of the pest object.

In addition, overlap may occur when the pest bounding box images are inserted one by one into the blank background board images, so we design a 9-point method to avoid overlapping insertion, as shown in figure 3. For the 9 points in the pest image to be inserted (right side of Figure 3), if one of them falls into the existing bounding box, it will be judged to be overlapping and not inserted.

Due to the irregular appearance of the pest, some nodes of the 9-point method are not on the pest pixel, as shown in the dotted box on the right of Figure 3. To solve the problem that the 9-point method is too strict in judgment, we propose a mask-9-point method to refine the object boundary points, which combines the pest mask and the 9-point method. Firstly, we default the bounding box center point to be inside the pest object. Secondly, we use the midpoint from the center point to the boundary point as the judgment point. Thirdly, if the judgment points fall in the pest mask, the boundary point is added to the 9 points, otherwise, the midpoint of the line is added to 9 points, as shown in Figure 4.

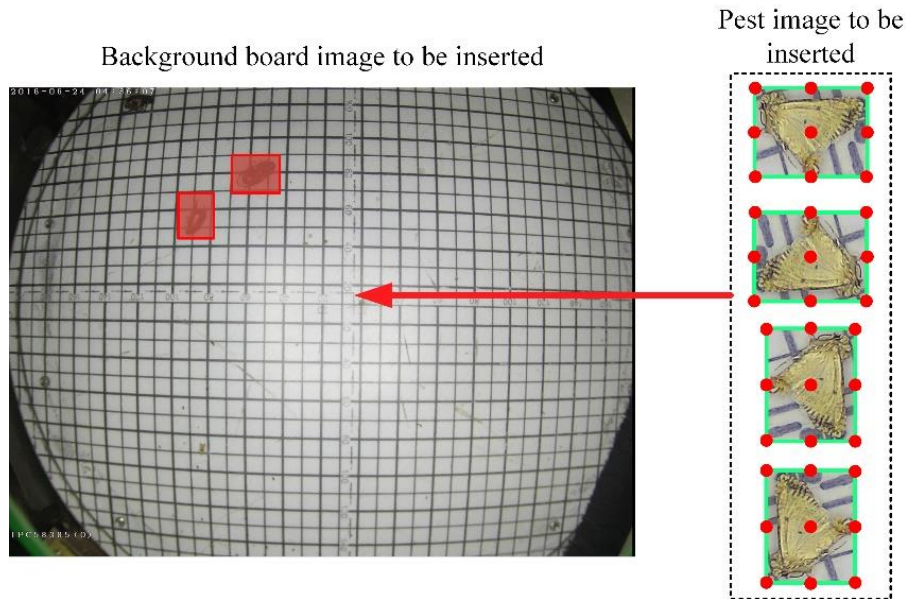


Figure 3. The 9-point method for overlap judge.

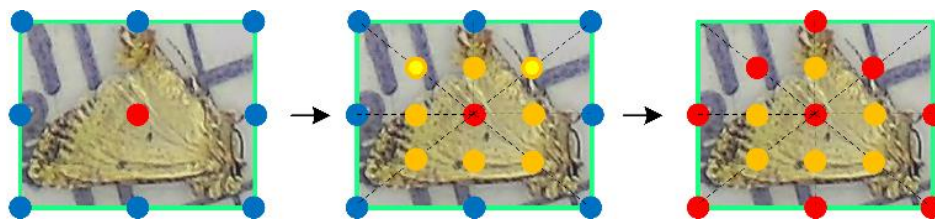


Figure 4. Mask-9-point overlap judgment.

In addition, Figure 5 shows the overall structure of the JBDOC and we list the specific steps below:

Step 1: we select 1309 background board images without boxes as expanded data.

Step 2: we construct the expanded bounding box set from the training set using equation (1) and our AGA to get the pest mask.

Step 3: flipping the images in the expanded bounding box set horizontally and vertically to get more generality.

Step 4: the pest image in Step 3 is randomly selected and inserted into the image in Step 1 using the Poisson Fusion. Determine whether there is an overlapping bounding box. If there is, discard the insertion and re-execute Step 4. Finish Step 4 until all the pests bounding box images are inserted into the expanded dataset.

### 3. EXPERIMENTAL ANALYSIS AND DISCUSSION

#### 3.1 Experiment settings

We combine our method with Faster R-CNN, Cascade R-CNN, FCOS, and Retinanet respectively to show performance improvements on LPD datasets. Firstly, we use JBDOC to construct a synthetic dataset. Secondly, we use the synthetic dataset to preliminary train the detector. Thirdly, we use the original trainset to train the network and use the test set to obtain the result. All the processes of training use backpropagation and Stochastic Gradient Descent (SGD) to optimize network parameters. For Faster R-CNN and Cascade R-CNN, the RPN is requested to generate 1000 candidate boxes. We use the same training parameter for preliminary training with 12 epochs and the Momentum is set to 0.9. The learning rate follows the principle of linear scaling, in which the first 8 epochs are 0.0025 and the last 4 epochs are 0.00025. The device of servers using an NVIDIA Titan RTX GPU and Mmdetection2.0.0 framework in the environment of Python 3.8.

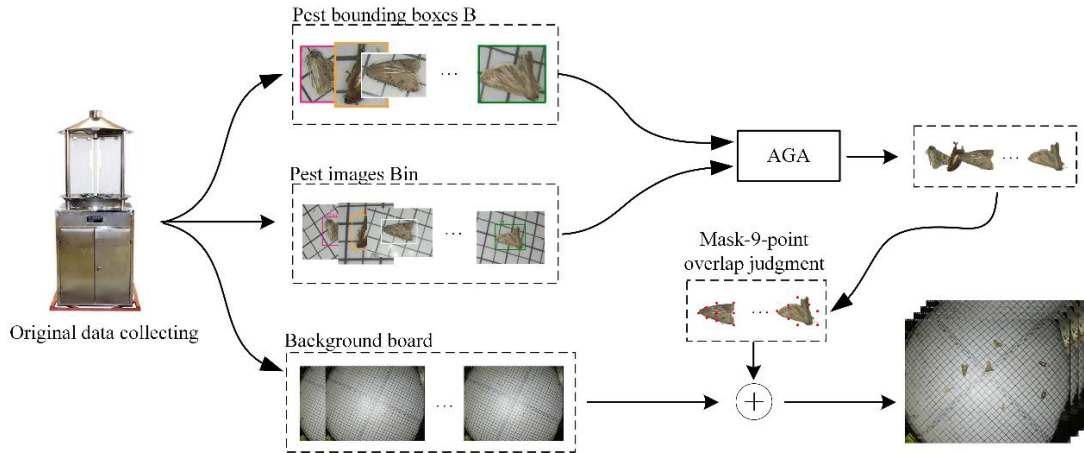


Figure 5. The overall framework of JBDOC.

### 3.2 Experiment results

The experiment results show in Table 1. Where the JBDOC represents the method of Joint Balance-Distribution Oriented Composition,  $AP_{50}$  represents the AP (average accuracy) with the IoU threshold of 0.5, mRecall and mAP represents the mean AP and mean Recall under the IoU threshold of 0.5, 0.75, 0.95. To be fair, we use ResNet50 as the backbone network for all methods. Our JBDOC achieves state-of-the-art performance with Cascade R-CNN. In addition, with the help of our JBDOC, the performance of all methods is improved.

Table 1. The performance of our JBDOC.

Method	JBDOC	mAP	$AP_{50}$	$AP_{75}$	mRecall
Faster R-CNN		35.4	62.3	37.3	50.5
	√	<b>36.4</b>	<b>63.7</b>	<b>38.3</b>	<b>51.0</b>
Cascade R-CNN		36.0	62.6	38.5	50.2
	√	<b>36.8</b>	<b>64.1</b>	<b>39.2</b>	<b>53.2</b>
Fcos		33.1	57.4	35.5	55.3
	√	<b>33.9</b>	<b>58.6</b>	<b>36.5</b>	<b>55.2</b>
Retinanet		30.6	52.6	33.0	53.6
	√	<b>31.0</b>	<b>53.5</b>	<b>33.0</b>	<b>53.4</b>

Further, the class-by-class performance compares as shown in Figure 6. The pest number and pest average size are displayed using normalized values (the absolute values divided by the sum values). The  $AP_{50}$  is improved obviously by our JBDOC at few-shot small-size categories (24, 25, 26, 28, and 45). This shows that our JBDOC effectively solves the imbalance of class number and size distribution.

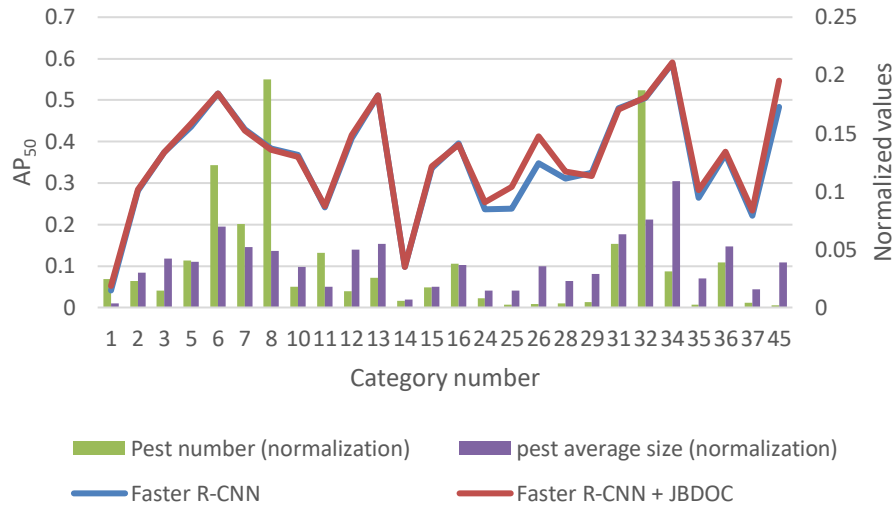


Figure 6. The performance improvement of JBDOC compared with Faster R-CNN.

### 3.3 Qualitative results

To visually observe the accuracy, we visualized the detection results of Faster R-CNN, Retinanet, Faster R-CNN+JBDOC (ours), Retinanet+JBDOC (ours), as shown in Figure 7. Among them, the first row shows the performance of the minimum-sized category, the third row shows the test result with the interference of other categories. Our JBDOC contributes to detecting the small-size pest as shown in the first row of Figure 7. With our JBDOC, the detection bounding boxes become more precise as shown in the second line of Figure 7. Even if the images are mixed with other categories and noise, our JBDOC still serves the purpose of networks to discover pests, as shown in the third row in Figure 7.

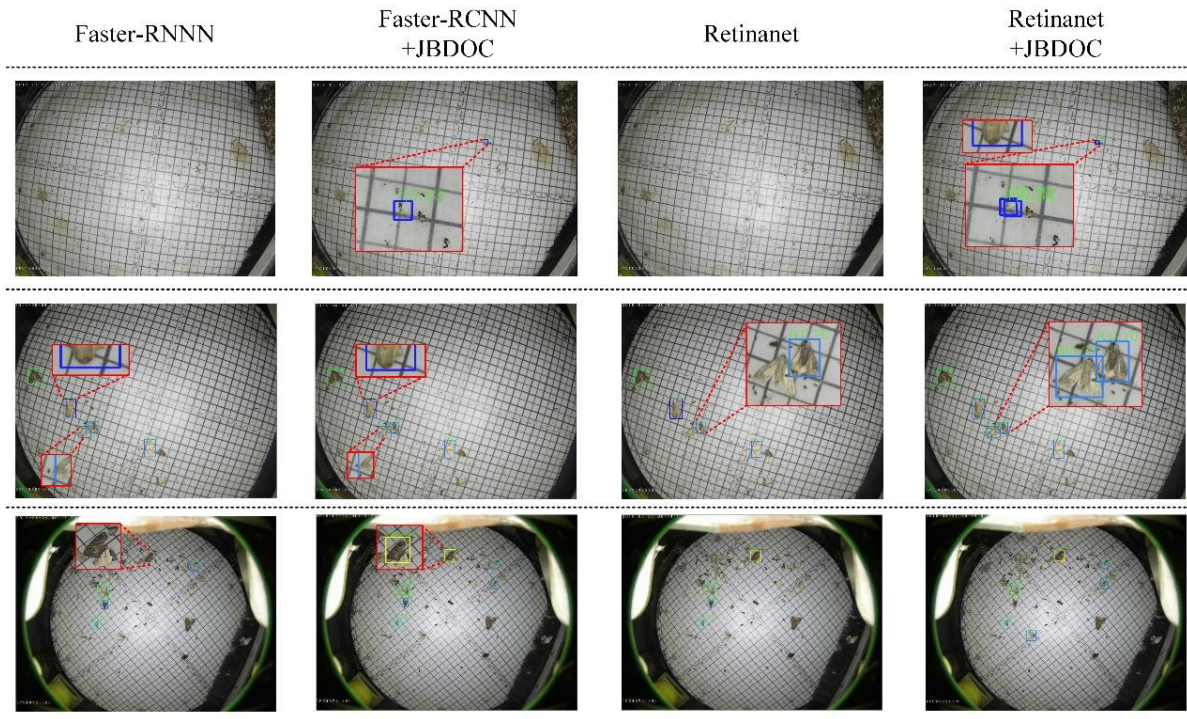


Figure 7. Visualization results.

## 4. CONCLUSION

Due to the imbalance of pest categories and size distribution, the performance of pest detection is insufficient to the requirements of Integrated Pest Management (IPM). Therefore, we propose a Joint Balance-Distribution Oriented Composition (JBDOC) to improve the performance of pest detection. Firstly, we construct a light-trap pest dataset (LPD) including 18585 images and 26 pest classes. Secondly, we design an Automatic Grabcut Algorithm (AGA) to automatically obtain the mask of pest bounding boxes in the LPD. Thirdly, we design a data augmentation method to expand the data with few-shot and small-size categories. With the synthetic dataset, the experimental results show that the proposed method can effectively solve the problem of small-size detection difficulty and category imbalance. Future work will focus on reducing manual labeling costs with synthetic datasets and few-shot learning.

## REFERENCES

- [1] Singh, S. P., (ed.) [Common Bean Improvement in the Twenty-First Century] (vol. 7), Springer Science & Business Media, (2013).
- [2] Krizhevsky, A., Sutskever, I. and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 25, 1097-1105 (2012).
- [3] Uijlings, J. R., Van De Sande, K. E., Gevers, T. and Smeulders, A. W., "Selective search for object recognition," *International Journal of Computer Vision*, 104(2), 154-171 (2013).
- [4] Ren, S., He, K., Girshick, R. and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, 28, 91-99 (2015).
- [5] Cai, Z. and Vasconcelos, N., "Cascade R-CNN: Delving into high quality object detection," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 6154-6162 (2018).
- [6] Tian, Z., Shen, C., Chen, H. and He, T., "FCOS: Fully convolutional one-stage object detection," *Proc. of the IEEE/CVF Inter. Conf. on Computer Vision*, 9627-9636 (2019).
- [7] Lin, T. Y., Goyal, P., Girshick, R., He, K. and Dollár, P., "Focal loss for dense object detection," *Proc. of the IEEE Inter. Conf. on Computer Vision*, 2980-2988 (2017).
- [8] Liu, L., Wang, R., Xie, C., Yang, P., Wang, F., Sudirman, S. and Liu, W., "PestNet: An end-to-end deep learning approach for large-scale multi-class pest detection and classification," *IEEE Access*, 7, 45301-45312 (2019).
- [9] Jiao, L., Dong, S., Zhang, S., Xie, C. and Wang, H., "AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection," *Computers and Electronics in Agriculture*, 174, 105522 (2020).
- [10] Wang, F., Wang, R., Xie, C., Yang, P. and Liu, L., "Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition," *Computers and Electronics in Agriculture*, 169, 105222 (2020).
- [11] Ayan, E., Erbay, H. and Varçın, F., "Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks," *Computers and Electronics in Agriculture*, 179, 105809 (2020).
- [12] Liu, L., Wang, R., Xie, C., Li, R., Wang, F., Zhou, M. and Teng, Y., "Learning region-guided scale-aware feature selection for object detection," *Neural Computing and Applications*, 33(11), 6389-6403 (2021).
- [13] Rother, C., Kolmogorov, V. and Blake, A., "Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, 23, 3(2012).