ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0043

Detection of Agricultural Pests based by YOLO

Yingjie Luo, Lei Ni

Chengdu Technological University, Chengdu 610000, China

Abstract

Focus on the problem of the species pests are difficult to identify in agricultural production an application based on the YOLO algorithm is presented. The algorithm extracts features fastly and complete detection task simultaneously. Through collecting agricultural pests, the algorithm establishes the a dataset build target object detection model and train them, the problem of difficulties in identifying agricultural pest species and locating them has been solved. The result shows that the mAP'value of YOLO algorithm is 92.42%, the prediction precision is 96.8%. The algorithm has a high recognition rate, and identify the species of pests fastly and accurately, it can meet the detection and identification requirements.

Keywords

Object Detection; Pests in Agriculture; YOLO.

1. Introduction

Agricultural pests rely largely on manual identification leading to misclassification, and research has been carried out on the use of convolutional neural networks to identify pests to ensure the effective and rational application of pesticides against crop diseases. Zhang Hongtao[1] et al. used the Ant Colony Optimization (ACO) algorithm for stored grain pests and confirmed the possibility of ACO-based grain insect feature extraction. Chen Juen[2] based on an improved residual network to extract more pest image features for image recognition of garden pests by adding convolutional layers and enlarging the number of channels.

Traditional two-stage object detection techniques mainly use algorithms such as selective search or edge box. Regions in the image potentially containing detection objects are extracted [3-4], while the candidate regions are categorized and their positions are regressed to get detection results. They are based on feature extraction, in which most of the candidate regions are generated by independent network branches, followed by classification and regression [5]. Not only does it take time, but also it has weak generalization power that does not fit well to complicated almost. YOLO (You Only Look Once) [6], a one-stage object detection, regard object detection as a regression problem, slicing the bounding box and associated probabilities of classes resulting in faster processing of images. To distribute pests is uneven usual. Object detection model will be trained by a convolutional neural network with one stage object detection algorithm YOLO in this paper. Following the model was used as the target detection model, the model prediction accuracy was 96.8% with an average mAP (mean Average Precision) of 0.944.

2. YOLO Algorithm

2.1 Design Approach

2.1.1Detection Principle

Using the whole image as the network input, and directly regressing the position of the bounding box and the class to which the bounding box belongs at the output layer, the YOLO algorithm. Images are divided into S*S grid cells, and S is the number of grids, in this paper S is 7. Each grid cell is predicted with two grid cells. One grid is in charge of predicting an object if its center belongs to this grid. That means the whole image with pixel resolution of 448*448 is input to the network, and the features of insects are extracted by convolution to get the feature vector. The final output feature map is a tensor of S*S*(B*5+C) after filtering out the smaller pixels with the activation. B is 2 bounding boxes, C is 20 pests, and the number 5 represents 4 coordinate values (x, y, w, h) and 1 confidence level. As shown in equation (1) the confidence level is defined as the product of the presence or absence of the object and the IOU value, The output is 7*7*30, as shown in Figure 1.

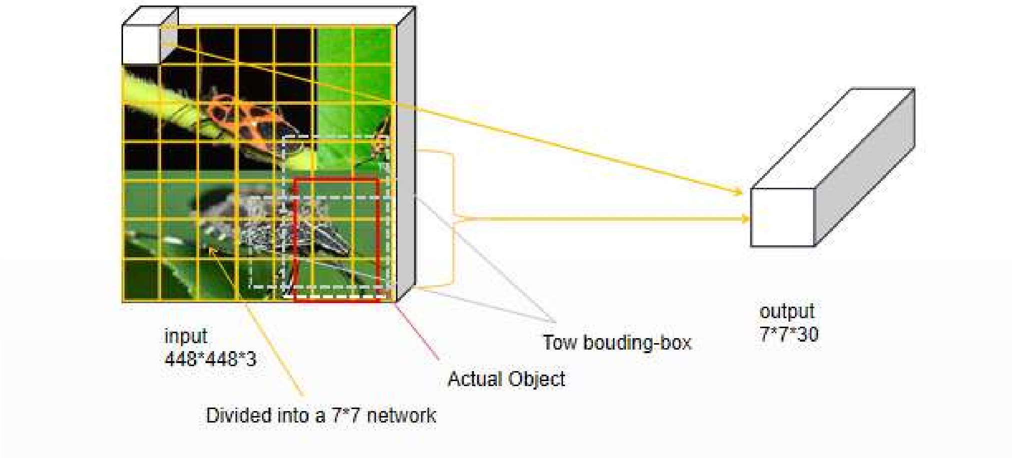

Figure 1. Detection diagram

$$Confidence\ =\ Pr(Object) * IOU_{pred}^{trut} \tag{1}$$

In this way, each bounding box is labelled with the highest ranked class label according to the confidence ranking of the 20 classes. The confidence is then calculated for each of the 20 categories in a turn, in a category (i.e. a row of the matrix), the score less than the threshold is set to 0. The confidence is calculated as the IOU (intersection over union, i.e. the ratio of the intersection and union of two regions) of the predicted bounding box and ground truth. The probability that a grid cell has an object is set to a value of 1 for ground truth in the presence of an object and 0.5 for ground truth without an object. Finally the NMS algorithm (NonMaximum Suppression) is used to remove the probability of a bounding box with a large repetition rate, and the bounding box is calculated by equation (2).

$$\Pr(\mathrm{Class}_i|\mathrm{Object}) * \Pr(\mathrm{Object}) * \mathrm{IOU}_{\mathrm{pred}}^{\mathrm{truth}} = \Pr(\mathrm{Class}_i) * \mathrm{IOU}_{\mathrm{pred}}^{\mathrm{truth}} \tag{2}$$

Pr(Class|Object) in equation (2) is the class probability mentioned above. The eventual product is the final confidence score of the current bounding box (note that and confidence obviously do not mean the same). IOU is multiplied in the equation to show the larger the IOU the larger the confidence score, so the final confidence score combines the coordinate optimum and the category optimum.

2.1.2 Training Principles

Object detection problem is considered as a regression problem by the YOLO algorithm [7] which both uses the sum of squared errors for the loss function. An essential in training the model is to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0043

reduce the losses by gradient descent and back propagation. The loss function is intended to find a balance between coordinates (x,y,w,h), confidence level and classification ranking in a good way. Loss Function Equation (3) can be divided into three parts: ① position error, i.e. coordinate loss ② confidence error, composed of the confidence loss when objects are included plus the confidence loss when objects are not included ③ classification error, i.e. the probability the grid cell contains a class of objects (20 pests in this paper).

$$\begin{aligned}&\lambda_{\text{coord}} \sum_{i=0}^{S2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \left[(x_i-\hat{x}_i)^2+(y_i-\hat{y})^2\right] + \lambda_{\text{coord}} \sum_{i=0}^{S2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \left[\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right]+\\&\sum_{i=0}^{S2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \left(C_i-\hat{C}_i\right)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S2} \sum_{j=0}^{B} I_{ij}^{\text{noobj}} \left(C_i-\hat{C}_i\right)^2 + \sum_{i=0}^{S^2} I_i^{\text{objj}} \sum_{c\in \text{ classes}} \left(p_i(c)-\hat{p}_i(c)\right)^2\end{aligned} \quad (3)$$

Equation 3, Ii indicates whether the center of the object falls into grid i, if so Ii = 1, the opposite is 0. Iij indicates the j-th detection bounding box in grid i is responsible for the object, if so it is 1, the opposite is 0 [8]. The image contains many grids with no object centres falling in, and such grids predict B detection bounding boxes with a confidence 0. Typically, the gradient of such grid cell during training is far greater than the gradient of the grid containing the centre of the object, resulting in unstable or even scattered training. To address these issues, YOLO assigns a larger weight (λcoord = 5) to the localization error. A smaller weight (λnoobj = 0.5) is assigned to the confidence error of the grid that does not contain object centers. And the classification error and confidence error of the grid containing object centers remains unchanged.

2.2 Structural Improvements

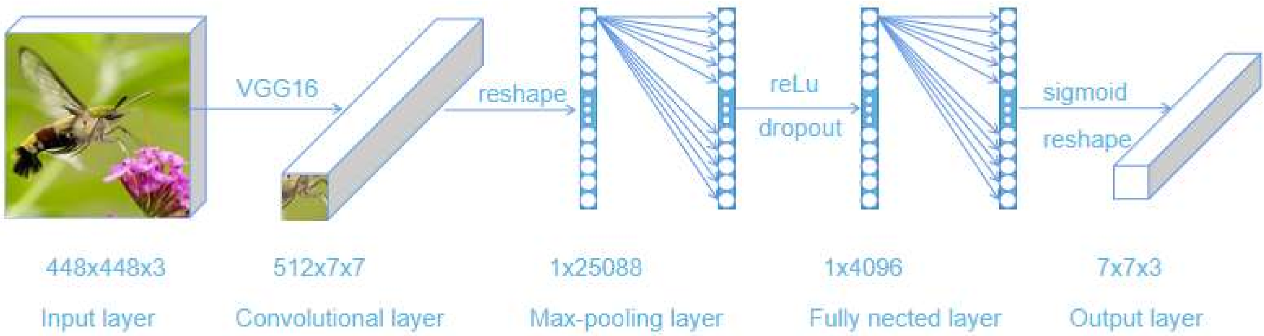

Figure 2. Network structure with VGG16 backbone

The VGG16 is used as the backbone network in this paper, as shown in Figure 2. VGG16 has a 16-layer network containing 13 convolutional layers and 3 fully connected layers[9]. The convolution layer is for extracting features. The reshape is an one-dimensional fully-connected layer mainly designed to predict category probabilities and coordinates, and the output layer is 7*7*30, with 7*7 being the number of grid cells. During the experiments, the parameters were updated using the mini-batch gradient descent scheme and Back-Propagation, reducing the time spent on training by reducing the number of fully connected and pooled layers. According to the confidence score of the 20 bug species, the bounding boxes are ranked and labelled with their own top ranking. Taking out all the bounding boxes of the same kind and their confidence score. This allows only one prediction per bounding box, which is more accurate if NMS algorithm is used again.

Volume 8 Issue 7, 2022
DOI: 10.6919/ICJE.202207_8(7).0043

3. Concrete Implementation

3.1 Source of Data

About 1200 different images of pests were collected from the Internet in total, and the samples were selected as pests of Hemiptera, Lepidoptera, and Sphingidae. Included categories such as Cicadidae, Cnidaria, Ruler Moths, Elephant Wax Cicadidae, Lampyridae, and Fruit Moths, it covers 20 categories of common agricultural pests. The resolutions of all images were adjusted to 480*480, and we needed to strengthen the data of the original images, including flip, rotation, scale, crop, and translation to increase data for training. It also improves the generalization power and robustness of the model to avoid overfitting of some of the images earned during the training as shown in Figure 3.

Figure 3. Pest organism Samples

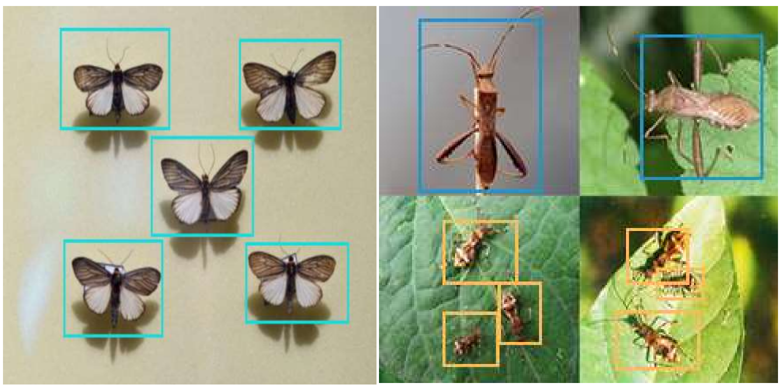

Figure 4. Labeling Pests

3.2 Process of Training

During training, the PASCAL VOC dataset format was used and the results are shown in Figure 4. The objective of the YOLO algorithm is the predictions the information values of these coordinates from the input images. Following note of all pest samples, an arrangement file was created to read the training data path for the project, which was automatically split into a training set and a test set using a data processing script file. The training is shown in Figure 5.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202207_8(7).0043

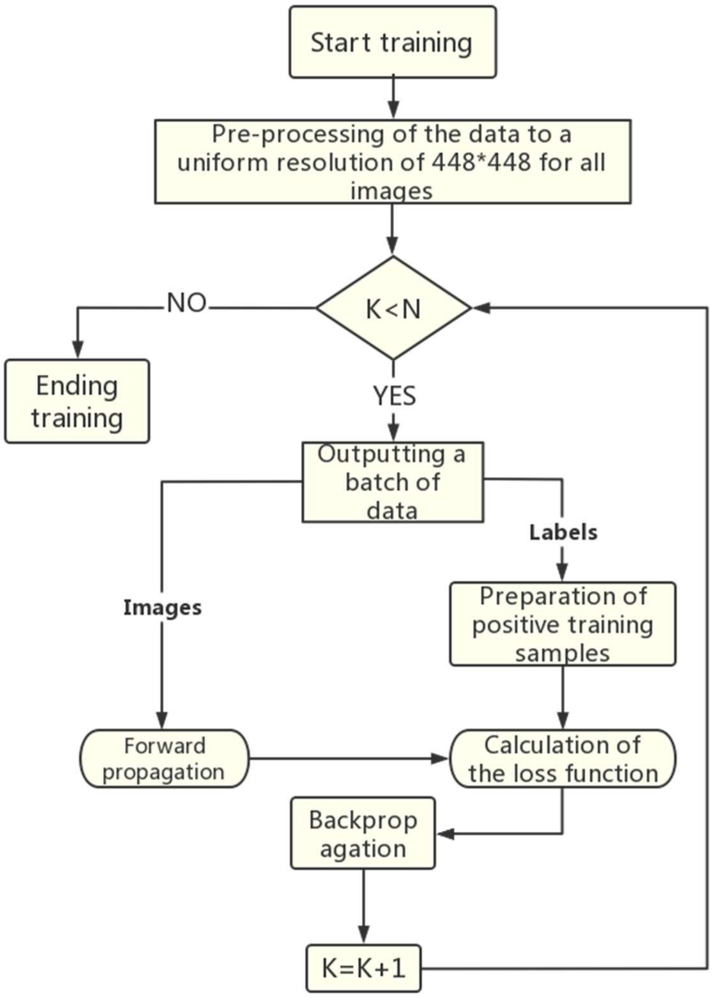

Figure 5. Training graphs

3.3 Results of Analysis

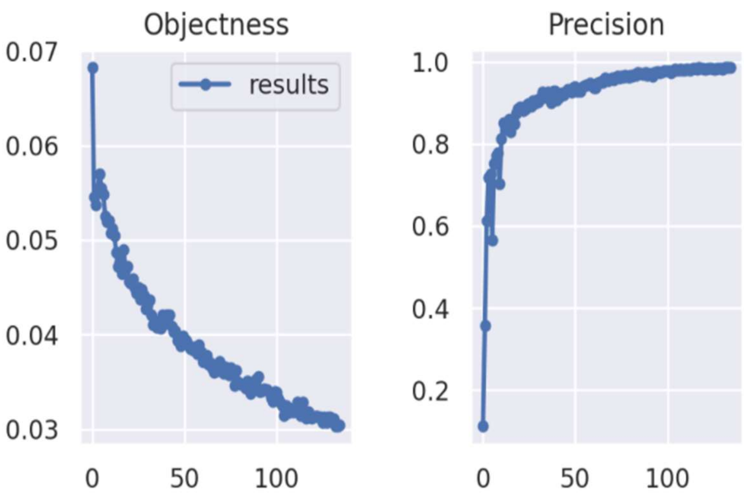

Figure 6. Loss values and precision of the trained set

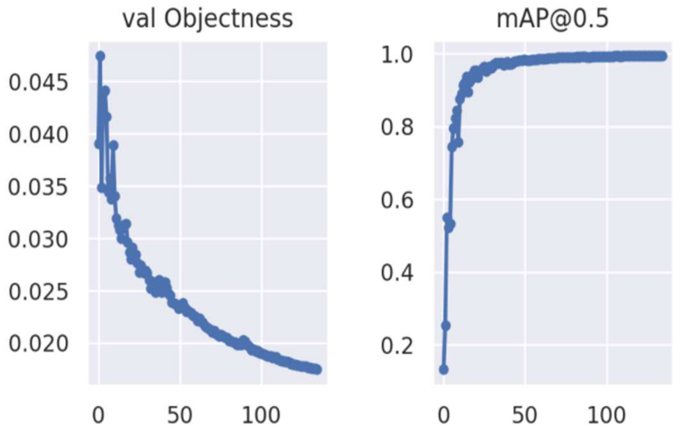

Figure 7. Loss values and mAP plots of confirmation sets

According to the results of this training, the model prediction accuracy was 96.8% (Figure 6.), and the average accuracy of mAP (Figure 7.) was 0.9442, with a high overall recognition rate. Overall set training 200 epochs, the loss of the training of the first epoch of rapid decrease, the final loss value (as in Figure 6) converged to 0.0124, after the loss convergence, the training ended. Once the training was finished, separate testing sets were imported into the trained network model for verification testing to obtain the detection results of the object detection algorithm on the test sets, as shown in Figure 8. As can be seen from Figure 8, the model identified the pests in the different band positions in the image accurately, with confidence levels of 87% and 92% respectively, and recognition speed of 1.965s.

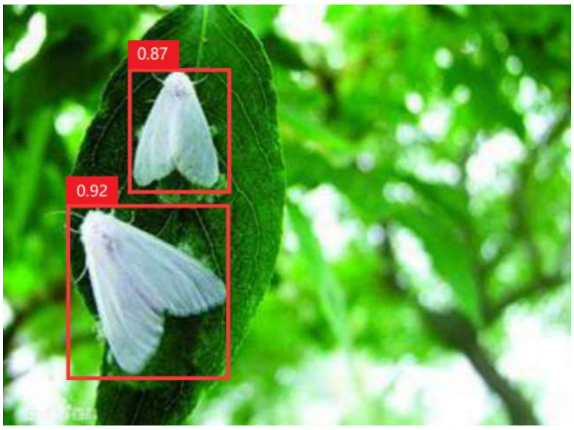

Figure 8. Pest prediction results graph

4. Conclusion

To achieve agricultural pest identification and improve the overall identification accuracy and efficiency, this paper uses a single-stage object detection algorithm to identify pests. About 1200 images of bug dataset and its labeled files are used in this paper. The model was trained for 200 epochs, saving the weights every 10 epochs, eventually the best model was chosen among all the weights, with an mAP of 95.23%. Experimentally, the proposed YOLO-based target detection algorithm has a high recognition rate for common agricultural pests. The method can identify the species and location of bugs quickly and accurately, which meets the needs of pest detection and identification.

Acknowledgments

Natural Science Foundation.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0043

Fund Source: College Students' Innovative Entrepreneurial Training Plan Program and Project No.S202111116050.

References

[1] H.T. Zhang, H.P. Mao, D.Y. Qiu. Feature extraction in image recognition of grain storage pests[J]. Journal of Agricultural Engineering,2009,25(02): p.126-130.

[2] J. Chen, J. Chen, L. Y. Chen, S. Wang, H. Zhao, H. Y. Zhao, and C. J. Wen. Image recognition of garden pests based on improved residual networks[J]. Journal of Agricultural Machinery,2019,50(05):p.187-195.

[3] ZITNICK C L, DOLLAR P.Edge boxes: locating object proposals from edges[C]//European Conference on Computer Vision.Cham: Springer, 2014: p.391-405.

[4] HU Q, ZHAI L.RGB- D image multi-target detection method based on 3D DSF.

[5] R-CNN[J].International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(8): 1954026.

[6] B. J. Xiao, R. J. Wan, and J. K. Chen. Research on mask wearing recognition using YOLOV5 model[J]. Fujian Computer,2021,37(03):p.35-37.

[7] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: ;IEEE, 2016. 779 −788.

[8] Y.N. Wang, Z.H. Pang, D.M. Yuan. Real-time vehicle detection based on YOLO algorithm[J]. Journal of Wuhan University of Technology,2016,38(10):p.41-46.

[9] Y. Mei, Y. Yin, Y. L. Shi, Z. H. Liu, L. Y. Chang. Research on intelligent recognition algorithm of leafy vegetable downy mildew based on improved VGG convolutional neural network[J]. Shanghai Vegetable, 2021(06):p.76-84.

[10] https://zhuanlan.zhihu.com/p/166254616.