Research on Flower Image Classification Method Based on YOLOv5

Ming Tian1*, Zhihao Liao2
1Department of Automation, Faculty of Artificial Intelligence and Automation, Hust, Wuhan, China
2Department of Artificial Intelligence, Faculty of Information and Engineering, Nanchang University, Nanchang, China
*tianming0811@163.com

Abstract—The rapid development of deep learning has accelerated the progress of related technologies in the computer vision field and it has broad application prospects. Due to flower inter-class similarity and intra-class differences, flower image classification has essential research value. To achieve flower image classification, this paper proposes a deep learning method using the current powerful object detection algorithm YOLOv5 to achieve fine-grained image classification of flowers. Overlap and occluded objects often appear in the images of the flowers, so the DIoU_NMS algorithm is used to select the target box to enhance the detection of the blocked objects. The experimental dataset comes from the Kaggle platform, and experimental results show that the proposed model in this paper can effectively identify five types of flowers contained in the dataset. Precision reaching 0.942, Recall reaching 0.933, and mAP reaching 0.959. Compared with YOLOv3 and Faster-RCNN, this model has high recognition accuracy, real-time performance, and good robustness. The mAP of this model is 0.051 higher than the mAP of YOLOv3 and 0.102 higher than the mAP of Raster-RCNN.

1. INTRODUCTION
The study of flowers has many aspects, such as the study of ornamental value, nutritional, and medicinal value of flowers[1]. Accurate classification of flowers is the primary work of related research. Researchers can realise flower classification accurately under knowledge accumulation and data access, but it takes much time. With the development of society and the improvement of living standards, people have an increasing demand for flower appreciation, but ordinary people are often challenging to accurately identify the types of flowers. Therefore, the automatic classification of flower images can assist researchers in flower research and provide popular science and convenience for people.

Flower image classification belongs to fine-grained image classification[2], which classifies different flower subcategories. There are three difficulties in the image classification of flowers. (1) There are similarities between different categories of flowers, as shown in Figure 1; (2) Flowers of the same category have different characteristics, as shown in Figure 2; (3) Plants usually have more than one flower, so there are overlapping flowers in the image, as shown in Figure 3.
Traditional methods extract image features mainly by feature extraction operators and combine the extracted features with traditional machine learning algorithms to realise image classification. Many researchers have used complex algorithms to divide flower images and perform feature extraction in recent years, achieving good results [3-6]. Nilsback et al. [3] proposed to fuse the SIFT and HOG features and classify them using SVM. Wang L et al. [4] proposed transforming original colour flower images from RGB space to Lab space and performing segmentation using the OTSU algorithm. Angelova et al. [5] proposed a target segmentation method and extracted 4-scale HOG features using locality-constrained linear coding (LLC) for feature coding. Xie X et al. [6] proposed the significance-based flower image segmentation method Saliency Based flower Image Segmentation. However, the algorithm for extracting the image characteristics in the traditional methods is closely related to the current image's characteristics, and the algorithm's versatility is not good.

Deep learning can automatically extract features from images, avoiding the subjectivity of manually extracting features, and is widely used in object detection and classification problems. Xia et al. [7] use Inception v3 network and migration learning techniques to improve classification accuracy, but the model is complex and flawed in real-time. Yin H et al. [8] proposed an unsupervised flower image classification method based on selective convolution descriptor aggregation. For fine-grained image classification it refers to the literature for bird classification [9]. Literature [9] Use bounding box and part annotation to train the R-CNN model to obtain the optimal candidate regions and classify them with SVM.

The YOLO algorithm is a one-stage target detection algorithm in deep learning. The YOLO algorithm was proposed in 2016 by Redmon et al. [8], which implements the classification and localisation of object detection with one network, and YOLO has been widely used in object detection. Currently, YOLO has
experienced development from v1 to v5 [10-13]. YOLOv5, launched in 2020, has the advantages of small volume, fast speed, high precision, and implementation in ecologically mature PyTorch. R-CNN belongs to the Two-stage object detection algorithm, which can effectively improve the problem of the target to be tested in the image, but the Two-stage model is more complex and computational than the One-stage model. From all of the above, the latest YOLOv5 target detection algorithm is selected to classify the flower dataset of the Kaggle platform.

2. YOLOv5 MODEL
YOLOv5 is the latest object detection algorithm in the YOLO algorithm, which has high detection accuracy and is fast and good in real-time. YOLOv5 includes four models of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, where the YOLOv5s has the smallest volume. This paper selects the YOLOv5s model, consisting of four parts, Input, Backbone, Neck, Prediction, respectively. The network structure is shown in Figure 4.

Figure 4. YOLOv5 network
2.1 Input
On the input side, YOLOv5 draws on the CutMix method and uses Mosaic data enhancement to improve the recognition of small targets effectively. Add adaptive scaling processing, the image unified scaling to a unified size and then sent to the network learning, to enhance the ability of network data processing.

2.2 Backbone
Backbone includes CSP networks and Focus structures, etc. The Focus structure contains four-slice operations and one convolution of 32 convolutional cores, transforming the original 608 * 608 * 3 picture into 304 * 304 * 32 feature maps. CSPNet performs local cross-layer fusion, utilising the feature information of different layers to obtain richer feature maps.

2.3 Neck
The Neck section contains both PANet and SPP. PANet (PathAggregation), to fully integrate the image features of different layers, aggregate the top feature information and the output features of different CSP networks in top-down order, and then aggregate the shallow features from the bottom-up. SPP (space pyramid pooling) uses four different sized nuclei for maximum pooling and then performs tensor splicing.

2.4 Prediction

2.4.1 loss function: The loss function of YOLOv5s uses GIOU_Loss, which alleviates the situation that IOU_Loss cannot handle the two boxes. As shown in Figure 5, we assume that the minimum external rectangle of the prediction box and the real box is C, The union set of the forecast box and the real box is N. The intersection of the prediction box and the real box is M, IOU is the ratio of intersection to the union, as shown in (1). D is the difference set of C and N, as shown in (2). GIOU is IOU minus the ratio of D TO C, as shown in (3), then the formula for GIOU_Loss is shown in (4).

\[
\text{IOU} = \frac{M}{N} \quad (1)
\]

\[
D = |C - N| \quad (2)
\]

\[
\text{GIOU} = \text{IOU} - \frac{D}{C} \quad (3)
\]

\[
\text{GIOU}_\text{LOSS} = 1 - \text{GIOU} = 1 - \left(1 - \frac{M}{N} - \frac{D}{C}\right) \quad (4)
\]

Figure 5. Geometric relationship
2.4.2 Non-Maximum Suppression: locally removes redundant detection boxes, retaining the best one. YOLOv5 uses NMS to select the detection box, and this article uses DIOU_NMS. DIOU_NMS can improve the detection accuracy of the overlapping and occluded targets.

3. RESULTS AND ANALYSIS OF THE EXPERIMENTS

3.1 Flower detection framework
The framework of this model is shown in Figure 6, divided into two parts: the training phase and flower classification phase. During the training phase, the training set is entered into the model for training. During the flower classification phase, the model completes the flower detection and classification of the input image. The model processes the prediction scores in the five categories of daisies, sunflowers, dandelion, roses, and tulips, and then plotted the predicted flower species and scores on the output image.

![Figure 6. Flower detection framework](image)

3.2 Dataset
The dataset of this paper adopts the flower dataset disclosed on the Kaggle platform, which includes flower images of daisy, dandelion, rose, sunflower and tulip. See Fig. 7 for the schematic diagram of various flower images. The flowers were labelled with LabelMe, and the data set was divided into the training set, verification set and test set according to the ratio of 8:1:1. The dataset consists of 400 images, each of different sizes.

![Figure 7. Five types of flower images](image)

3.3 Results of the training
The Pytorch framework is adopted in this experiment, and the model is trained on GeForce GTX 1080ti (video memory 12G) GPU, CUDA 10.0 experimental environment. The training parameters are set as follows: the input image pixel is 640*640, the Batch size is 16, and the initial learning rate is 0.01. The
momentum factor is set to 0.95, the weight decays to 0.001, and the epoch is 300. Precision, Recall and mAP after model training are shown in Figure 8.

![Figure 8. Training results](image)

It can be seen from Figure 8 that the accuracy, recall, and mAP of the model after training is high and stable, which shows that the training effect of the model is good.

3.4 The evaluation index

This experiment uses Precision (P), Recall (R), and mean Average Precision (mAP) as evaluation indicators. The results of the detection contained four cases, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The definition is shown in Table 1.

| Abbreviation       | Define                              |
|--------------------|-------------------------------------|
| True Positive      | TP                                  |
| False Positive     | FP                                  |
| True Negative      | TN                                  |
| False Negative     | FN                                  |

A positive sample with a positive forecast
A negative sample with a positive forecast
A negative sample with a negative forecast
A positive sample with a negative forecast

For category C, the accuracy is the ratio of the number of correct samples detected to the number of samples detected, and the calculation formula is shown in (5). The Recall of category C is the ratio of the correct number of samples detected to the total number of samples of this class, and the calculation formula is shown in (6). AP is the area of the curve enclosed by P and R, and mAP is the average of the AP for all categories, and the calculation formula is shown in (7).

\[
\text{Precision} = \frac{TP}{TP+FP} \quad (5)
\]

\[
\text{Recall} = \frac{TP}{TP+FN} \quad (6)
\]

\[
\text{Mean Average Precision} = \frac{\sum \text{AveragePrecision}}{C} \quad (7)
\]
3.5 Test results and analysis
The experimental results are shown in Figure 9. The proposed model can correctly detect and classify flower images with insect interference, overlapping ones, and fuzzy flower images from the detection results.

![General flower image](image1)
(a) General flower image

![Insect interference image](image2)
(b) Insect interference image

![Overlapping flower image](image3)
(c) Overlapping flower image

![Fuzzy flower image](image4)
(d) Fuzzy flower image

Figure 9. Test results

Furthermore, we detect and classify the test sets on YOLOv3 and Faster-RCNN models, and the indicators of the different models are shown in Figure 10.

![Precision, Recall and mAP in different models](image5)

From Figure 10, the algorithm in this paper, which is YOLOv5s, has higher Precision, Recall and mean Average Precision than other algorithms and performs better than other algorithms, which is suitable and excellent as an algorithm for implementing the automatic classification of flower images.

4. CONCLUSION
Based on the powerful YOLOv5 algorithm in target detection, this paper realised flower image detection and fine-grained classification. The experimental results show that the model can detect and recognise the flower images with insect interference, overlapping flower images and fuzzy flower images. Compared with the YOLOv3 and Faster-RCNN algorithms, the flower classification model proposed in this paper has good detection and classification results and obvious performance advantages. In addition, this experiment mainly focuses on the detection and classification of large flower targets in the image. The next step is to optimise the model further to classify and detect small flower targets in the image.

REFERENCES
[1] Ou J, YANG C H. Quantitative evaluation of the ornamental value of wild herbal flowers [J].
Guizhou Agricultural Science, 2009, 37(06): 166-170.

[2] Weng Y C, Tian Y, Lu D M, et al. Fine-grained bird classification based on deep region networks [J]. Journal of Image and Graphics, 2017, 22(11): 1521-1531.

[3] Nilsback M E, Zisserman A. Automated flower classification over a large number of classes [C] / / Proceedings of the 6th Indian Conference on Computer Vision, Graphics & Image Processing. Bhubaneswar, India: IEEE, 2008: 722-729. [DOI: 10.1109 /ICVGIP. 2008. 47]

[4] Wang L, Fang L, Chen E, et al. Flower image segmentation algorithm based on Lab color space [J]. Journal of Zhejiang Wanli University, 2018112(3): 67-73.

[5] Angelova A, Zhu S H. Efficient object detection and segmentation for fine-grained recognition [C] / / Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013: 811-818. [DOI: 10.1109 / CVPR. 2013. 110]

[6] Xie X D, Lyu Y P, Cao D L. Saliency detection based flower image foreground segmentation [EB / OL]. 2014-09-18 [2018-06-20]

[7] Xia X L, Xu C, Nan B. Inception-v3 for flower classification [C] / /Proceedings of the 2nd International Conference on Image, Vision and Computing. Chengdu: IEEE, 2017: 783-787 [DOI: 10.1109 /ICIVC. 2017. 7984661]

[8] Yin H, Fu X, Zeng J X, Duan B, Chen Y. Flower image classification with selective convolutional descriptor aggregation [J]. Journal of Image and Graphics, 2019, 24(05): 0762-0772.

[9] Zhang N, Donahue J, Girshick R, et al. Part-Based R-CNNs for fine-grained category detection [C] / / Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 834-849. [DOI: 10.1007 /978-3-319-10590-154]

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, Real-time Object Detection [C]. Las Vegas: IEEE Conference on Computer Vision 56 and Pattern Recognition, 2016.

[11] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]. Honolulu, USA.: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[12] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2021-01-15].

[13] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-01-15].