Butterfly Recognition Based on Faster R-CNN

Ruoyan Zhao¹, Cuixia Li¹,*, Shuai Ye¹ and Xinru Fang¹

¹Software and Applied Science and Technology Institute of Zhengzhou University, Zhengzhou, China

*Corresponding author e-mail: qyliying@126.com

Abstract. There are more than 18,000 butterflies in the world. Butterflies are essential cultural insects and have great significance in the field of insect research. However, due to its high similarity, various type, and the characteristics of the distinction is not apparent, the recognition and classification of butterflies at this stage has the problems of low accuracy and slow recognition speed, so it is of great significance to make research on the automatic identification of butterflies and improve the efficiency of automatic identification of butterflies. This paper uses the Faster R-CNN algorithm for butterfly recognition and describes the whole process from butterfly dataset selection, butterfly dataset processing, to butterfly classification. The experimental results show that the butterfly automatic recognition system based on the Faster R-CNN deep learning framework can realize automatic detection and species identification of butterfly photos in the ecological environment, and the average classification accuracy can reach 70.4%.

1. Introduction

In recent years, the automatic recognition of butterflies has attracted the attention of more and more researchers, and there have been many kinds of research on butterfly recognition applications [1]. In the classification problem of the butterfly, the butterfly's shape, color, texture, and other features are extracted according to the input butterfly image, and then the corresponding classification result is returned according to the similarity. Over the years, the research on automatic recognition of butterflies has emerged in an endless stream, with automatic recognition of butterfly species based on Extreme Learning Machines (ELM) [2], using Local Binary Patterns (LBP) [3] and Grey-Level Co-occurrence Matrix (GLCM) [4]. Extract features in butterfly images, then use the extreme learning machine [5] for classification, and butterfly recognition based on single hidden layer neural network [6], use binary length similarity (BLS) [7] to extract butterfly shape features and then use Single hidden layer neural networks are classified. There are two problems in the existing butterfly species identification research. Firstly, the butterfly dataset is difficult to collect, and the number of butterflies included in the butterfly data set is not comprehensive. Secondly, the butterfly photos used for training are all pattern photos with obvious morphological features, lacking the ecological photos of butterflies in the natural ecological environment, and the differences between the two photos are obvious, which makes the combination of production and research difficult and the recognition accuracy is low [8]. To solve the above two problems, this paper uses the up and down flip, left and right rotation, increase noise and other methods [9] to expand the data set, and in the training process, ensure that the training data set contains pattern photos and ecological photos to ensure training accurately. The dataset used in this article is the official data released by the 3rd China Data Mining Competition in the 2018 China
International Butterfly Recognition Competition, which includes standard photos of all existing butterfly specimens in China, and a large number of butterfly ecology environment photos. Part of the butterfly dataset was filmed in the wild and offered by other butterfly lovers. Since the position of the butterfly in the ecological photo is not fixed, it is necessary to first detect the position of the object and then process the multi-category.

The butterfly dataset contains a total of 5,695 butterfly photos, of which there are 1,425 eco-photographs, 111 species, of which 17 species of butterfly have only one ecological photograph. Due to the lack of ecological photos of various butterflies, the need for data sets to be expanded is obvious. The ecological photos and pattern photos of the butterflies are very different. The number of butterflies in each photo is uncertain, and because of the butterfly's self-protection function, it is always in the environment conducive to hiding itself, making the butterfly distinguish from each other in the surrounding environment is difficult. The standard photo of the butterfly specimen has 4270 sheets and 1176 kinds. The standard photo contains all kinds of ecological photos. Which is characterized by a picture only with a butterfly, and the picture is clear, the features are obvious, and the picture is easy to distinguish. Samples of the butterfly dataset photos are shown in Figure 1.

Deep learning is an important technology in the field of machine learning. Through many studies and experiments, deep learning has shown great potential in the field of image processing. Therefore, this paper attempts to use deep learning technology to apply Faster R-CNN to butterfly recognition. The Faster R-CNN algorithm can simultaneously locate and classify the target, and complete the butterfly detection and classification in the butterfly ecological environment photo.

In the second part, this paper will introduce the Faster R-CNN algorithm in detail. In the third part, the whole experimental process of butterfly recognition will be described, and the experimental results will be analyzed. In the fourth part, we will summarize and analyze this experiment and look forward to the application prospects of this technology.

![Figure 1. Samples of the butterfly dataset photos.](image)

### 2. Faster R-CNN

The task of butterfly detection is to find the location of butterflies accurately in a given picture and to mark the category of butterflies. R-CNN detection algorithm is based on the traditional method to find out the area that may be an object firstly, then normalize the size of the area into the format of the convolution network input, and finally determine whether the area is an object, which kind of object, as well as further regression and micro-adjustment of the area that is an object, so as to make the area framed more accurate [10]. Fast R-CNN proposes a special layer, named ROI layer. After the deposition of R-CNN [11] and Fast R-CNN [12], Faster R-CNN has been proposed. It mainly solves the following two problems: firstly, it proposes a regional recommendation network RPN to quickly generate candidate regions; secondly, it makes RPN and Fast R-CNN network share parameters through alternate training; In its structure, Faster R-CNN integrates feature extraction, regression, and
classification into a network, which greatly improves the comprehensive performance, especially in classification speed. The basic structure of Faster R-CNN is shown in the following figure 2.

![Figure 2. The structure diagram of Faster R-CNN.](image)

The convolutional layer is a target detection method in the CNN network. Faster R-CNN network firstly uses a set of Convolutional, Relu and Pooling layers to extract the feature values of images. Then the extracted eigenvalues are used in subsequent RPN layer and full connection layer. Region Proposal Networks is also known as RPN network; RPN Network is used to generate area target values. The anchor is the core of RPN network. In RPN networks, it is necessary to determine whether the target exists in the corresponding receptive field of each sliding window center. Because the target size and the ratio of length to length are different, it needs sliding windows of several sizes, the function of the anchor is to give the size of a reference window and get different sizes of windows according to the multiple and the ratio of length to width. Training in RPN network involves the problems of ground truth and a loss function. Ground truth is the criterion to judge whether anchor is the target, expressed as 0 or 1. In this paper, the decision rule is that if the Intersection-over-Union (IoU) value of an anchor of any target area is greater than 0.7, the target area is determined. If the IoU of an anchor of any target area is less than 0.3, the background area is determined. The IoU is the coverage of the predicted box and the real box. Whose value is equal to the intersection of two boxes divided by the union of two boxes. Other anchors do not participate in the calculation. After determining the region as an object, bounding box regression is used to modify anchors so that to obtain accurate proposals. ROI Pooling layer collects feature maps and proposals of input, then extracts proposal feature maps after synthesizing these features, then sends them to the subsequent full connection layer to determine the target category. Next, the classification process is carried out, proposal feature maps are used to calculate the category of a proposal, and the bounding box regression is used to obtain the final exact location of the detection box.

In the VGG16 network model, for an image of arbitrary size P*Q, the image is first scaled to a fixed size M*N, and then the image of M*N is sent to the network. Convolutional layers contain 13 Convolutional layers, 13 relu layers, and 4 pooling layers. RPN network first passes through the
convolution of 3*3, then generates foreground anchors and bounding box regression offsets respectively, and finally calculates proposals. The ROI Pooling layer uses proposals to extract proposal features from feature maps and send them to subsequent full-connection and softmax networks for classification so that to determine which category of objects a proposal belongs to. The process of testing a picture is shown in Figure 3 below.

3. The expression of the experimental process and results

The process of dealing with data sets is very important for training and testing. Our process of official given data sets can be divided into the following steps: filtering, naming and transcoding, divide into sets, marking, refinement and expansion. First of all, filtering out the butterfly species which only have one photo from the official images and remove them from the data set. Secondly, naming and transcoding. Name all photos as a unified format, such as IMG_003501. There are two formats of photos in the data set as JPG and png. All PNG formats are converted into jpg format. Thirdly, divide into sets, the whole dataset contains ecological photos and model photos. The feature of model photos is that one photo contains only one butterfly, while the feature of ecological photos is that there are different kinds of butterflies in one photo. According to the papers provided by the official organization of competition, the test set contains only the butterfly of the ecological photos. According to the training set and the test set, the ecological photos are divided into 50% and 50% respectively. The second method to generate the training set is that only add the butterfly model photos corresponding to the butterfly species in the training set. Forth, marketing, labeling tool is used to tag each picture of the training set and generate the XML file. The whole process is manually labeled. The marking process is divided into four steps: the first step is to set the path, storing the XML file generated after marking in that location; The second step is to frame the butterfly area in the photo; The third step is to specify the species, after the area is framed, fill in the name number of the butterfly, such as aaaa0007014; the fourth step is saving, make the butterfly area, store the XML file in the specified path. In the process of marking, several problems have been noticed. No part of the butterfly in the photo can be missing. By this principle, the marking area can be minimized as far as possible. The path of XML file storage after tagging needs to be set accurately, and the types of each photo should not be confused. Fifthly, refinement. Considered that training needs four documents, namely training set document, validation set document, training validation set document and test set document, it is decided to refine the data set first and then expand it. The document contains the number of the photo, such as 003501. After completed these four documents, the total training set is refined into the training set and verification set, each accounting for about 50%. Training verification set is the union of the training set and verification set. Fill in the number of the corresponding pictures and separate the corresponding pictures for expansion. After separation, the photo folders are named train2.0 and val2.0. Sixthly, expend the data set. Filter, increase noise and rotation are used to expand the training set. Two different filters were added to the train 2.0 and val 2.0 photos, namely, heating

Figure 3. The process of testing a picture.
filter (85) and cooling filter (LBB). Seventhly, modify, because the file name and path in the XML file generated after marking the photo with the tool are the same as that in the XML file before copying, so we need to change it to the name and path of the corresponding picture. Eighthly, check. Check the XML file in batches to see if there is one XML file containing more than one object or the name under the object is None. The corresponding solution to these two problems is as the following, for the first problem, only one object is reserved in an XML file, and the redundant object is deleted directly. For the second, change None in the name to a label corresponding to the butterfly species, which is done manually. Ninthly, merging. All of the above data set processing will ultimately be attributed to four documents. We need to add the names of all the extended photos to the corresponding text documents. In the process of processing, we only need to change the training set documents and verification set documents because the test set documents are not expanded and need not be changed, while the training verification set document content only needs to merge the training set document content with the verification set document content.

During the experiment, the learning rate is set to 0.001, the momentum value is set to 0.9, the training step is set to 5000, the results are displayed once each 10 training sessions, and the size of each training batch is 128. The evaluation and recognition process can be divided into two parts. The first is whether the selection of candidate boxes is correct or not: using IoU as the identification index of regional detection. This experiment stipulates that if the IoU value is greater than 0.7, the selection of candidate boxes is correct, and if the IoU value is less than 0.3, the selection of candidate boxes is wrong. The second is whether the target detection classification is correct or not, multi-class image classification evaluation index mAP is used as the target detection evaluation index. In the first classification test, it was found that the mAP could reach 73.3%, but after the second replacement test set, the mAP dropped to 68.2%. In order to ensure that the accuracy of the test is always stable at a high level, we replaced the data set several serval times, and finally, the mAP is stable at about 70.4%.

4. Conclusion

In this paper, the whole process from butterfly dataset selection, butterfly data set processing to butterfly recognition and classification using Faster R-CNN is described, and finally, the average classification accuracy is stable at 70.4%. Summarize the whole experiment process and draw the following conclusions. First, during the training process, we must ensure that the sample is sufficient. We can use many ways to expend the data set, such as upside-down, left-right rotation, cooling, noise adding and so on, to ensure the training is effective; Second, when testing the accuracy, we should ensure the randomness of samples, which requires constant change of test set and constant adjustment of parameters to achieve a stable classification and recognition accuracy at a high level. The research on butterfly recognition is not only important for the progress of image recognition, but also for the protection of butterfly culture.

Acknowledgments

This research was financially supported by the Key Technology R&D Program projects of Henan Province (172102210003) and College Students Innovation and Entrepreneurship Training Program of Zhengzhou University (2018cxy443).

References

[1] Fischer K, Karl I, Heuskin S, et al. Kin recognition and inbreeding avoidance in a butterfly[J]. Ethology, 2015, 121(10): 977-984.
[2] Kaya Y, Kayci L, Tekin R, et al. Evaluation of texture features for automatic detecting butterfly species using extreme learning machine[J]. Journal of Experimental & Theoretical Artificial Intelligence, 2014, 26(2): 267-281.
[3] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(7): 971-987.
[4] Marceau D J, Howarth P J, Dubois J M M, et al. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 1990, 28(4): 513-519.

[5] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1-3): 489-501.

[6] Li C, Wu H, Yang Z, et al. A novel SHLNN based robust control and tracking method for hypersonic vehicle under parameter uncertainty[J]. Complexity, 2017, 2017.

[7] Ji J, Li J, Tian Q, et al. Angular-similarity-preserving binary signatures for linear subspaces[J]. IEEE Transactions on Image Processing, 2015, 24(11): 4372-4380.

[8] Pinzari M, Santonico M, Pennazza G, et al. Chemically mediated species recognition in two sympatric Grayling butterflies: Hipparchia fagi and Hipparchia hermione (Lepidoptera: Nymphalidae, Satyrinae)[J]. PloS one, 2018, 13(6): e0199997.

[9] Sharma V, Eckels J, Schilling B, et al. Panorama Public: A public repository for quantitative data sets processed in Skyline[J]. Molecular & Cellular Proteomics, 2018, 17(6): 1239-1244.

[10] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[11] Li W, Li J, Sarma K V, et al. Path R-CNN for Prostate Cancer Diagnosis and Gleason Grading of Histological Images[J]. IEEE transactions on medical imaging, 2018.

[12] Faster R. Towards Real-Time Object Detection with Region Proposal Networks Shaoqing Ren[J]. Kaiming He, Ross Girshick, and Jian Sun.