Classification of bird species using the data set UCSD Birds-200-2011

OSSELIN Pierre
École Normale Supérieure Paris-Saclay
pierre.osselin@gmail.com

Abstract

Fine-grained recognition consists in discriminating categories with only small subtle visual differences. Here we study a particular instance of that problem where we aim at classifying birds with regards to their species, with the Caltech-UCSD Birds-200-2011 data set [3]. First, our designed model should be able to detect birds in an image, which is a difficult task in some cases. Second, the model should detect very small differences, sometimes regarding solely the body shape or the beak of the birds. Finally, we have a very small training set, consisting of roughly 50 *images per class. To solve this task, we pre-processed the data by detecting birds in the images, intersecting the results given by the YOLOv3 [2] and Mask-RCNN [1] algorithms, and then fine-tuned a ResNext-*101*. This pipeline allows us to attain a test score of* 82.6% *on Kaggle.*

1. Introduction

Fine-grained recognition requires a good detection and recognition model. To do this, we use the python wrapper [1] [2] of YOLOv3 [2] and Mask-RCNN [1]. After preprocessing the data, we perform data augmentation, and fine-tune a ResNeXt-101 architecture on our data.

2. Pipeline

2.1. Bird detection

The quality of the data set is not homogeneous, some pictures zoom on birds, in others birds are hidden. This observation prompts us to use bird detection to focus on the relevant information. We use python wrappers of YOLOv3 and Mask-RCNN. We processed the data by intersecting the results of both algorithms, compared by their output probabilities that the object detected is a bird. We then used the bounding box given to crop the images, and created two data sets. The first one consisted of the cropped images fully resized in the resolution 224×224, the resolution used for our model. In the second one we also resized to 224×224, but kept the aspect-ratio and filled the missing pixels with black backgrounds. The rationale behind this second data set is that some bird species differ only by their body shape, a feature that can be altered by a classic resizing.

Type of Cropping	Validation Error	Test Error
Normal	0.91262135922	0.82565
Constant Aspect Ratio	0.90291262135	0.80645

Table 1. Results

2.2. Data Augmentation

We introduced data augmentation in our pipeline. In particular, images are randomly horizontally flipped, rotated to up to 30%, and their color is slightly perturbed before being fed to our model.

2.3. Classification

To perform classification, we fine-tuned a pre-trained model of ResNeXt-101 with our data set. The weights are initially fitted to classify the ImageNet data set. We used an SGD optimizer with parameters $lr = 0.0001$ and $m = 0.9$ on 100 epochs. The results are displayed on table 1.

3. Conclusion

We could improve our performances by using other object detection algorithms, better architectures for classification, and applying transfer learning techniques.

References

[1] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[2] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

[3] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

[1] https://github.com/qqwweee/keras-yolo3.git

[2] https://github.com/matterport/Mask$_{R}CNN.git$