Object Recognition and Computer Vision
Assignement 3: Bird classification

Loubna Ben Allal
ENS Paris Saclay
loubna.ben_allal@ens-paris-saclay.fr

Abstract

In this project, we tried to predict bird breeds using Computer Vision techniques. The images were taken from the Caltech-UCSD Birds-200-2011 bird [3]. Given 1187 images and 20 target categories, the goal is to build a model with the highest accracy on a test dataset containing the same classes. The challenge is that the size of the dataset is small, and the birds can sometimes be hard to find and to differentiate from other breeds. This report summarizes our approach for solving this problem.

1. Introduction

We want to classify birds images into 20 classes. This problem can be challenging given the small size of our dataset. To avoid overfitting and make use of the performant models that were trained on large datasets such as ImageNet, we used the concept of transfer learning [2]. We also did some preprocessing, before applying the model, using birds cropping and data augmentation techniques. This enabled us to detect bird breeds better and to enrich our dataset.

2. Approach

2.1. Birds detection

We noticed that the background doesn't always have an added value to the breed identification and can rather be confusing in some cases. So we decided to crop the images around the birds using an object detection model. We used YOLO v3 [1] for this task, which is a fast and accurate model. We were able to retrive more than 80% of the birds with at least 0.9 confidence score. If a bird isn't succesfully detected by the model, we just keep the original image.

2.2. Data Augmentation and preprocessing

The total number of images in our training and validation datasets is only 1187, and they are unevenly ditributed through the 20 categories. So to prevent overfitting, we used data augmentation techniques such as horizontal and vertical flips, as well as random crops. We also resized the images to a square of dimension 320x320 to have a good resolution that saves all the details of the images.

2.3. Model selection

We tried different models from those pretrained on ImageNet such as VGG, Inception and ResNet. We got the best performance using fine-tuned ResNet101, where we froze all the layers except the last one and we added a dropout layer and another fully connected layer, that were trained on our training set. We tried different optimizers, the best one was SGD with 0.9 momentum, and we adapted the learning rate during the training using a scheduler with the one cycle learning rate policy. We used 16 as batch size.

This model gave us 93% validation accuracy and 80.6% accuracy on the test set.

3. conclusion

Our approach reaches good results on the birds dataset and makes use of different techniques. Some improvement perspectives would be to analyze the classes with less images and try to balance the data, or use self-supervised learning to include unlabeled images.

References

[1] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 1

[2] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *International conference on artificial neural networks*, pages 270–279. Springer, 2018. 1

[3] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical report, 2011. 1