Object Recognition and Computer Vision : Assignment 3

Rafaël Brutti
rafael.brutti@telecom-paris.fr

Abstract

This assignment focuses on image classification on a subset of the Caltech-UCSD Birds-200-2011 dataset. The goal is to produce a model that gives the highest possible accuracy on a test dataset containing the same categories.

1. Introduction

The dataset used during the competition is composed of 1082 images for the training set, 103 images for the validation set and 517 images for the test set. One of the first big problems for the classification of these images is that the shooting is not standardized because the lighting is not always the same, the bird is not always in the center of the image, the head of the bird is not always in the same direction and sometimes the bird is even placed behind an object. To limit these effects, one of the solutions could be the preprocessing of the data such as data augmentation or the use of bird shape detection algorithms. After doing the preprocessing, the challenge is to find the best neural network architecture that leads to the best classification of bird species.

2. Architecture

Firstly, adding new layers or non-linear activation functions to the baseline model didn't increase the accuracy dramatically. This is why, I preferred to use models as Inception or ResNet which gave better results. However, without preprocessing the data, these neural networks tend to overfit due to their big structures and the few number of images for training and for validation. Finally, I decided to work with a ResNet152 pre-trained on ImageNet which increased the accuracy drastically. One other way to overpass the overfitting issue could have been to increase the number of images for each of the three sets.

3. Data Preprocessing

3.1. Data Augmentation

Amongst all my tests, the most successful changes were random horizontal flipping and randomly changing the orientation of the image.

3.2. Faster R-CNN

I used a Faster R-CNN [1], model with a ResNet50-FPN backbone and pre-trained on COCO train2017. This model allows deeper neural networks to improve accuracy and then is well suited for ResNet152. Then, I modified this object detection model to keep only the bounding boxes where birds are detected in order to crop the images by not considering insignificant objects.

4. Learning Rate Scheduler

I ran several tests in order to find a convenient value for the learning rate[1] which seems to be around 1e-3. Then, I set the learning rates of the model using discriminative fine-tuning [2] to avoid overfitting by forcing the model to learn more from the later layers.

5. Results

Model (in 15 epochs)	Validation Acc.	Test Acc.
ResNet152 pre-trained only	62%	54%
- with preprocessing	80%	72%
- with preprocessing & LR Scheduler	89%	82%

References

[1] S. Hong Y.Han A. Lee and G. Kim. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors (Basel)*, 2019. 1

[2] Leslie N. Smith. Cyclical learning rates for training neural networks. *IEEE*, 2017. 1

[1] https://www.kaggle.com/darthmanav/explaining-resnet-model-fine-tuning-pca-t-sne/notebook