Object Recognition Assigment 3 Report: Bird Species Categorization

Arnaud Massenet
ENS Paris-Saclay, 61 Avenue du President Wilson

Abstract

Categorizing pictures of objects has been at the heart of computer vision work in the last decades. We are now able to identify the location and category of objects on pictures by using datasets with millions of labeled images. However, in cases where the data is scarce, our current models struggle to learn on the data. To remedy the learning problem, one can use transfer learning to use the knowledge gained on previous similar categorization tasks to our new problem. In this report, we will discuss our approach for categorizing birds from the CUB-200-2011 dataset.

1. Introduction

In this report, we are trying to classify 20 species of birds taken from the CUB-200-2011 dataset. We are essentially trying to perform fine-grained image categorization. In other words, our classification task deals with birds that look alike and are hard to distinguish even to the human eye. For this task, we tried three approaches. The first one involved the building and training of a CNN from scratch and the second approach was based on transfer learning using pre-trained models such as ResNet and ResNext. In the last one, we used image embeddings from [1] and applied Linear Regression.

2. Dataset and Preprocessing

The subset of CUB-200-2011 which we used for training was comprised of 20 classes with around 50 samples for each class. The validation set had around 5 samples per class.

2.1. Images Transformations

Given the small amount of data we have, both for training and validation, we started by applying transformations to the datasets to avoid overfitting. To do so, at training time, we randomly applied vertical and horizontal flips to the images and then resized all images to 256*256. This step allows adding a small amount of noise that will help the model to generalize better. On top of these transformations, we have also created cropped pictures based on object detection models that were pre-trained to detect birds (taken from [2]).

2.2. Image Embedding

Another preprocessing technique we used was to use the feature extraction provided in the code from the *Large Scale fine-grained categorization and Domain-specific transfer learning* [1]. This feature extraction is based on the Inceptionv3 architecture presented in [3]. Essentially, we reduce all images to a 2048 1D vector which we will use to train a shallow Dense Neural network.

3. Models and Training

We trained all models for 100 epochs.

3.1. Homebaked Convolutional Neural Network

As a first step, we designed a convolution neural network (HCNN). For the sake of brevity, we invite the interested user to read the code for $Net2$ in the $model.py$ file. We have essentially 11 convolutional layers, 2 dense layers with dropout for the training and ReLu to introduce non-linearity.

3.2. Transfer Learning with Resnet and ResNext

We used both ResNet-152 and ResNext50 with pre-trained weights (on ImageNet) to train on our model. We initialized the last layer with a dense layer and tried to train the model on only a portion of the layers, but training all layers gave the best results. For more information, we invite the curious reader to refer to the code in $transfer\_model.py$.

3.3. Homebaked Dense Network on Image Embeddings

Using the Image Embeddings we trained a shallow 2-layers Dense Neural Net with dropout (HDN). We invite the interested user to read the code for Net in the $model.py$ file.

4. Results

	HCNN	ResNet	ResNext	HDN
Train Accuracy	60.3%	93.7%	100.0%	95.6%
Val Accuracy	54.0%	80.0%	91.0%	95.0%
Test Accuracy	**41.3%**	**68.4%**	**81.3%**	**90.1%**

Table 1. Table 1: Train, Validation and partial test accuracy for 4 models that we used to classify our dataset.

5. Conclusion

Through various models and strategies to represent the data, we discovered that image embedding allowed us to get good results even with a simple model such as ours and little data. We believe that further exploration in embedding models could help to train simple models with small datasets.

References

[1] Yin Cui, Yang Song, Chen Sun, Andrew Howard, and Serge J. Belongie. Large scale fine-grained categorization and domain-specific transfer learning. *CoRR*, abs/1806.06193, 2018.

[2] Yusuke Niitani, Toru Ogawa, Shunta Saito, and Masaki Saito. Chainercv: a library for deep learning in computer vision. In *ACM Multimedia*, 2017.

[3] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.