HW3 - Bird Image Classification

Ghassen BETTAIEB

Abstract

Bird species classification is a delicate task that requires fine-grained feature learning. We will try to solve this problem using Transfer Learning on models pretrained on ImageNet. The best model achieves 80% *test accuracy on Caltech-UCSD Birds-200-2011 dataset in this Challenge.*

1. Introduction

In this challenge, we provide different transfer-learning based approaches to classify bird species. The main challenge of this classification problem is related to the big variation in the provided data images. We had to shuffle and resplit the train and validation images to the same distribution as before, because the initialvalidation images were not very representative. To reduce the noise introduced by the image background we use an object detection method to localize and segment Birds. The ImageNet pretrained models used for transfer learning are ResNet50 and ResNext101.

2. Proposed Approach

2.1. Data augmentation

To artificially increase the number of images at training time, relevant data augmentation are used: Horizontal Flip, color jitter and adding white noise. This will create variations of the images to help the model generalize better and reduce overfitting.

2.2. Bird detection

To reduce the noise in the image background and learn features that only describe the body of the birds we perform bird image detection using Faster-RCNN, pretrained on 80 object classes (including bird class at class label=16).

2.3. Transfer Learning

ImageNet Pretrained models are deep nets trained on 1.2 million labeled images with 1000 classes. Hence using pretrained models for transfer learning helps to learn highly generalizable deep features for this task. The last convolutional layer usually outputs the global level features that are passed after to the task specific fully connected layers to output the 20 class probability distribution. A softmax function is finally used to determine the bird specie label. In a deep net the first convolutional layers learn simple features like lines and edges whereas the last convolutional layers learn fine-grained task specific features; Thus it is useful to retrain the last few layers on bird images to add granularity to the learned features. This is done by unfreezing few layers before the Fully-Connected layers to update its weights when training.

3. Experiments

We use pretrained resnet50 and resnext101 with a fully-connected layer of output size 20. We train the fully-connected head and the few last unfreezed layers. We use the SGD optimizer with an initial small learning rate ($lr = 0.01$) that decreases with training time to avoid altering the learned conv filters.

4. Results

Table 1: Accuracy of finetuning ResNet50 and ResNext101 on original and cropped images

Model	images	train accuracy	validation accuracy	test accuracy
ResNet50	original	0.92	0.92	0.73
	cropped	0.90	0.83	0.68
ResNext101 32x8d	original	0.94	0.95	0.80
	cropped	0.85	0.84	0.73

5. Conclusion

After trying different model architectures, the best results were obtained by finetuning ResNext101 that surprisingly yielded better test accuracy on non-cropped images. This may be explained by the fact that background scenes are correlated to bird species like water for the Brant-Cormora or tree trunc for Brown-Creeper.
A potential approach to improve results is to finetune the model on original images to learn the global spatial features and after retrain the same model on cropped images to learn finegrained features.