Object recognition and computer vision : Assignment 3
MVA Fall 2021

David Soto
École Normale Supérieure Paris-Saclay
david.soto@ens-paris-saclay.fr

Abstract

This assignment consists in participating in a Kaggle competition with the rest of the class on the Caltech-UCSD Birds-200-2011 bird dataset. The goal of this assignment is to produce a model that gives the highest possible accuracy on the test set. The dataset used for this Kaggle competition is a subset of the Caltech bird dataset, containing only 20 different species of birds. In this paper, we present a method for the classification of birds on the new dataset.

1. Dataset

The bird dataset given for the competition provides many challenging photos of birds. For instance, in some photos birds are occluded by objects such as leaves, tree branches or bird cages. The bird dataset contains 1702 images (1185 for training and 517 for testing) and 20 different species of birds is considered.

2. Approach

2.1. Ideas

The idea behind this method is to build a model using pre-trained neural networks (trained on ImageNet dataset of course) that already have well initialised filters, so that there is no need to learn them from scratch.

2.2. Data Augmentation

In order to improve the ability of the models to learn bird representation, I used data augmentation. It is well known that deep neural networks require a lot of training samples. Using data augmentation allowed me to generate new labeled images for the training of the models. For data augmentation, I mainly used the following : random horizontal flip, random rotation (45 degrees) and random resized crop.

2.3. The bird classification model

Using pre-trained models (trained on ImageNet dataset) is a good way to start building a bird classification model. Doing so will surely help us to produce a good model since pre-trained models have already learned some useful features, such as vertical lines, that are likely to appear in the photos of the bird dataset. Another reason for using pre-trained models is that we do not train them from scratch which saves us time and allows us to obtain good performances faster. Hence, using a pre-trained model seems a good way to produce a good model for the bird classification task. PyTorch provides pre-trained models in torch.models. Following this approach, I tested several pre-trained neural networks such as ResNet-50 or VGG-16 (which are all pre-trained on the ImageNet dataset) but the models that I chose to work with were ResNet-152 and ResNet-101. I just fine-tuned them for the bird classification task. PyTorch provides the ResNet-152 and ResNet-101 models on torchvision.models.

For the fine-tuning of ResNet-152 and ResNet-101, I started by freezing the first layers of the networks. This is because the first layers extract universal features (such as horizontal lines or edges) that are present in all photos. Therefore these layers do not need to be re-trained. I only trained the last layers. Moreover, I modified the network's architecture by adding two fully connected layers at the end (with 20 output units).

3. Implementation

After the images are loaded, I use data augmentation. Every image is resized to 256x256, we apply random horizontal flip, random rotation of 45 degrees and random resized crop. I use SGD optimizer with a learning rate of 0.0005 and a momentum of 0.8. Due to stockage issues that I encountered while training the models, I set the batch size to 4. The construction of a pre-trained ResNet-152 and ResNet-101 was done by passing pretrained=True into constructor.

Finally, the model was trained up to 37 epochs. Afterwards, I generated a csv file containing the predictions of the trained model on the images of the mystery categories dataset.

4. Results

Using this method, I obtained an average accuracy of 90% on the validation dataset, and a score of 0.81290 on the public leaderboard. This score was attained by training the model up to 37 epochs.

5. References

1. Wah, Catherine, et al. *The Caltech-UCSD Birds 200-2011 dataset.* Computation & Neural Systems Technical Report, CNS-TR-2011-001. (2011).
2. Ekin D. Cubuk, Barret Zoph , Jonathan Shlens,. Google Research, Brain Team. *RandAugment: Practical automated data augmentation with a reduced search space*. arXiv:1909.13719v2, nov. 2019.