Object Recognition and Computer Vision 2022/2023 - A3

Théo Di Piazza (theo.dipiazza@gmail.com)
Ecole normale supérieure Paris-Saclay (MVA)
November 2022

Abstract

The Caltech-UCSD Birds-200-2011 dataset which contains pictures of different types of birds. The objective is to develop a model capable of identifying the type of bird from its image, with the most accurate model possible. We will present the dataset, the pre-processing, the training and then we will present the model with the best score for an accuracy of 76% on the test sample.

1 Introduction

The dataset contains 1702 images: 1082 for the training sample, 103 for the validation sample and 517 for the test sample.

Figure 1: Rusty Blackbird (left) and Yellow Breasted (right)

Since the dataset contains few images, a first step can be devoted to pre-processing and data augmentation, in order to help the model perform better on the classification task.

2 Pre-Processing & Data Augmentation

If the bird is difficult to identify in the image due to its distance or position, this can complicate its classification. The first step proposed is, if possible, to extract the bird from the image with a pre-trained segmentation model: Mask R-CNN. If the bird is detected by the model, a new image will be created from the boundary of the area in which it is located.

Then, the original dataset is enriched via classical augmentation methods: rotation, glaussian blur, flip. These methods increase the accuracy score.

3 Models

The architecture of the chosen model is as follows: a ResNet152 followed by linear layers with activation and dropout.

The interest of using ResNet is to benefit from a complex architecture, which has already been trained on a larger dataset to segment and classify different types of objects, including birds. Thus, the first layers of the proposed architecture allow to identify edges the bird in a way, before classifying it according to its race with final layers. The added final linear layers allow the architecture to be fine-tuned to the bird classification. Using dropout increase the accuracy on the test sample, by preventing the network from over-training on the training dataset.

For training, the learning rate used for the ResNet layers is 10^{-5}, and 10^{-4} for the linear output layers. This allows to go from an accuracy of 31% (initial architecture of the project, cf github) to an accuracy of 91% on the validation sample.

4 Conclusion

The advantage of the proposed method is that it benefits from a pre-trained architecture via transfer learning : here ResNet152 was used. The final accuracy obtained on the test sample is 76%.

To improve the proposed solution, it is possible to study the accuracy of the model for each class of bird to estimate eventuel biases, and possibly use ensemble learning.

References

[1] *Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, Mask R-CNN, https://arxiv.org/abs/1703.06870 (2017)*
[2] *Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition, https://arxiv.org/abs/1512.03385 (2015)*