Assignment 3: Bird image classification

Hugo Blanc
École normale supérieure Paris-Saclay
4 Av. des Sciences, 91190 Gif-sur-Yvette
hugo.blanc.2022@ensta-paris.fr

Abstract

Object recognition is an easy task for human vision, but its algorithmic implementation is a complex challenge. Moreover, recent advances in computer vision allow, thanks to deep learning, to perform this task with a better accuracy. This is why, in this work, we will focus on image classification. We will deal here with the classification of birds from 20 different species, the dataset being derived from Caltech-UCSD Birds-200-2011 bird dataset. Also, the proposed method aims to provide a high classification accuracy, for its conception we studied various fields: preprocessing and data augmentation on the initial dataset, neural network design and transfer learning, adjustment of the network hyperparameters. Finally, our method proved to have good results including on the test set (>80%), which allowed us to obtain scores close to state-of-the-art algorithms.

1. Introduction

In this work, we are faced with the problem of classifying birds from a small data set. In a first step, we will artificially increase this dataset thanks to data augmentation techniques. Then, a transfer learning method will be used, consisting in retrieving a pre-trained model on a large dataset and readjusting it to our task. Here, the main pre-trained model used will be the ViT model, pre-trained on ImageNet. Finally, the hyperparameters of our model will be adjusted to obtain better results.

2. Image preprocessing and Data Augmentation

The images are initially of different sizes with different brightness and colorimetry. For these reasons, we resize the set of images to a standard size (224x224) and normalize the images to obtain a uniform dataset. Then, data augmentation allows to add more data to train and evaluate our model, this reduces data overfitting during the training. It also helps to solve class imbalance problems. In the context of image processing, we augment the dataset by performing random transformations on our image set. The selected transformations here are: horizontal flipping, cropping, rotation, they allow to improve the accuracy of our method.

3. Design of the model

3.1. Traditionnal CNN approch

A convolutional neural network was first created to provide a working basis.This network consists of 5 convolutional layers combined with max-pooling, Batch-Normalization and dropout operations with ReLu activation function and a final fully connected classifier. This method leads to low performances (<60% accuracy on test set).

3.2. Transfer Learning and ViT

The objective of transfer learning here is to retrieve a pre-trained model on a large dataset, then adapt the last layers of our model to allow it to classify our data. Finally, we train the model starting from the pre-trained weights to obtain a classifier for our task. The ViT model pre-trained on ImageNet was used here for the good performances it demonstrated in [1] [2]. Other networks such as ResNet, EfficientNet and ResNeXt were tested but gave slightly inferior results. Finally, a new network is created by paralleling the two previous best networks (ensemble learning).

4. Results

Different hyperparameters of the network have been adjusted (batch-size, learning rate, number of epoch) in order to improve the accuracy. The best results obtained are summarized below:

	Validation Accuracy	Test accuracy
ViT	93%	81.3%
ResNeXt-101	90%	80%
ViT+ResNeXt-101	95%	83.9%

Table 1. Transfer Learning method results

References

[1] Neil Houlsby Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *Int. Conf. Learn. Represent.* 1

[2] Xiaohua Zhai Ross Wightman Jakob Uszkoreit Lucas Beyer Andreas Steiner, Alexander Kolesnikov. How to train your vit? data, augmentation, and regularization in vision transformers. 2021. 1