Deep Learning – Case Study

Title: -Bird Species Classification using VGG16

Name: Kunal Zaveri

Enrollment Number: 18012021094

Batch: DL2

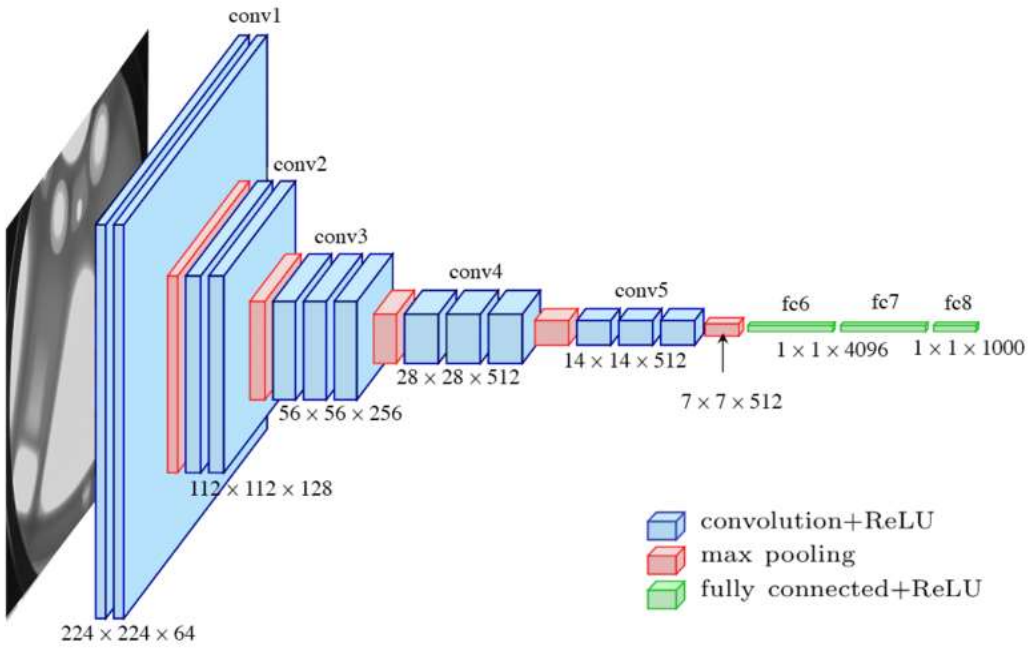
1. Introduction

Birds are an integral part of any environment and they are of the utmost importance to nature. The presence of birds of different species in an ecosystem is also important for many environmental reasons. This is another area where our classification program can be of use. With the help of our classification method, the authorities can also keep track of bird hunting in an area by monitoring the population rise and fall of each species of bird. Nowadays, image classification has become one of the major research fields of machine learning and deep learning.

2. Architecture

We are using the VGG16 architecture. VGG16is a convolutional neural network architecture that seeks to perform. The 16 in VGG16 refers to it having 16 layers that have weights. This network is a pretty large network and it has about 138 million (approx.) parameters.

- I am creating a sequential model. A sequential model means that all the layers of the model will be arranged in sequence. Here I have imported ImageDataGenerator from Keras. pre-processing. The objective of ImageDataGenerator is to import data with labels easily into the model
- I had created an object of ImageDataGenerator for both training and testing data
- I have used RELU activation for both the dense layer of 4096 units so that I stop forwarding negative values through the network.



3. Working

- The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (Conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, centre). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of Conv. layer input is such that the spatial resolution is preserved after convolution, i.e., the padding is 1-pixel for 3×3 Conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the Conv. layers (not all the Conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2-pixel window.
- For our feature extraction, we used a VGG-16 network that consists of 16 layers. During training, the input images are converted to images of size 224 x 224 as is required by VGG-16. A full stack of convolutional layers has three fully connected (FC) layers. The first two layers have 4096 channels each and the third layer contains 1000 channels. The final layer is a soft-max layer.

4. Code

The code is open source and publish on GitHub

Link: - https://github.com/kz2511/Bird-Species-Classification

5. Output