Object recognition and computer vision 2019/2020

Marouen Chafrouda
Marouen.chafrouda@student.ecp.fr

1. Introduction

In this project we'll work on Fine-Grained Image Classification of 20 Bird species Using a combination of techniques that involve using Deep Neural Networks.

2. Model Architecture

2.1. Feature Extraction

In our Classification we'll use the feature Extraction module used in Resnet 101 model and already available, pre-trained, in pytorch's Torchvision library. This network extracts as many as 1000 features which are then used to classify the birds into our 20 species using fully connected layers. The main difference between Resnet and other networks is the usage of Residual Blocks which help solve the performance degradation problem usually occurring in very deep Neural Networks : with the network depth increasing, accuracy gets saturated and then degrades rapidly, of course this is totally empirical, but it occurs in networks when experimenting with them.

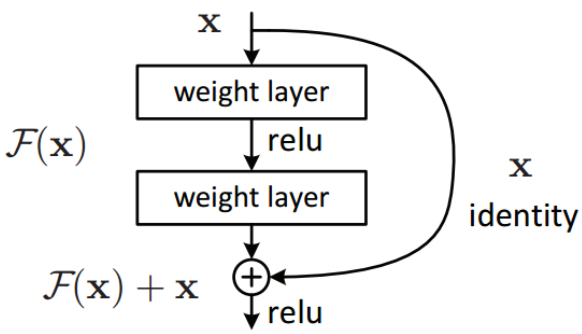

Figure 1. Residual Block Architecture

2.2. Data Preprocessing

For pre-processing we first started off using Faster-RCNN to extract bounding boxes that contain the bird object in our photos.This reduces the work required from our network and helps it extract relevant features only from the regions that are needed as we don't expect the picture background to provide any information that might help with the classification. Faster R-CNN is composed from 3 parts, typical Convolution Layers,a Region Proposal Network (RPN) which is the one responsible for the better performance of this network, and finally typical fully connected layers for classes and bounding boxes prediction. We also do normalization and resizing on images later on as part of the pre-processsing.

2.3. Data Augmentation

For data augmentation we use Imgaug, a very Handy library that provides a wide range of possible image augmentation techniques. Our augmentation pipeline includes Gaussian Blur, Dropouts and Coarse Dropouts, sharpening embossing Horizontal Flipping and rotations. These transformations are applied randomly on parts of the data-set on each epoch.

3. Results

The training is done first only on the fully connected layers that use the 1000 extracted features to predict the classes and then the whole network is trained at a lower learning rate so the weights can be more adapted to the problem(i.e bird species classificiation). The model was trained on a subset of the "Caltech-UCSD Birds-200-2011 bird dataset" which was divided to training and validation (80%/20%). We get around 90% accuracy on Validation

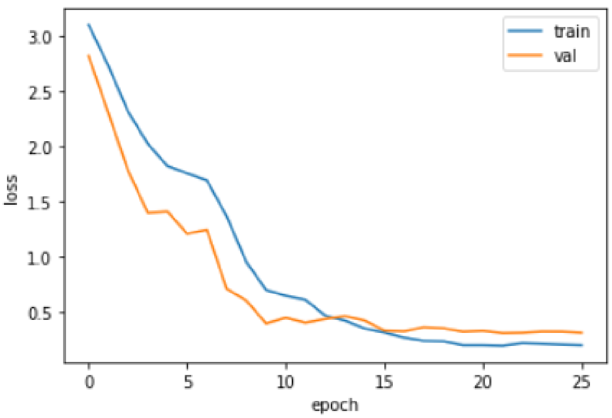

Figure 2. Training and Validation Loss

set

.	Training	Validation
Accuracy	93.3544	89.8734

4. References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 [cs.CV], 2015.

[2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497v3 [cs.CV] 6 Jan 2016