Università degli studi di Milano-Bicocca

Advanced Machine Learning

Final Project

Flowers recognition

Author:

David Bertoldi - 735213 - d.bertoldi@campus.unimib.it

January 14, 2023

Abstract

The ABSTRACT is not a part of the body of the report itself. Rather, the abstract is a brief summary of the report contents that is often separately circulated so potential readers can decide whether to read the report. The abstract should very concisely summarize the whole report: why it was written, what was discovered or developed, and what is claimed to be the significance of the effort. The abstract does not include figures or tables, and only the most significant numerical values or results should be given.

1 Introduction

The aim of this work is to build a machine learning model able to learn from a small knowledge base and to classify similar images but belonging to different classes. The main issue to overcome was the high chance that the model could overfit and fail to classify unseen samples. The strategy adopted was to find the model exploiting transfer learning the best and tryed to freeze the model such that it maintained similar performances.

This document describes the research on trained models, hyperparameters and generalization techniques that allowed the model to operate on a large variety of images.

2 Datasets

The dataset used is the *102 Category Flower Dataset* [1] created by the researchers of the *Visual Geometry Group* of *Oxford*. The dataset is composed of 8 189 RGB images of variable size, each image contains one or more flowers on a neutral background and is labeled with a single category extracted from a set of 102 possible categories. The original dataset also contains the flowers segmented from the background (Figure 1); these images can be used for example as further input for the neural network. In this work they were not used in order to force the model to be more elastic with respect to the background of the images.

The subdivision of the dataset defined in the original publication has been maintained, in particular there are 1 020 images in the training set, 1 020 images in the validation set and 6 149 images in the test set.

Figure 1: Training images and their segmentation

Each category is represented by 10 images in the training set and validation set, while the proportion of images for each category varies in the test set. The difficulty of operating on a dataset of this type is evident: the number of images for training is limited while the test set is larger.

Another peculiarity of the dataset is the presence of similar images belonging to different categories.

3 The Methodological Approach

In order to classify images correctly two types of experiment were taken in accont: the first one aimed to find the best architecture from the one found in litterature that could fit the available hardware (see Section 3.1) and the second tried to find a good trade-off between number of trainable layers and accuracy by freezing the layers' weights during training.

Before proceeding with the description of the experiments, there was a preliminar implicit objective: how to overcome the scarcity of training data.

3.1 Technology

The hardware used for this work was composed by a GPU NVIDIA 2070 SUPER with 8GB of VRAM, a CPU Intel i7-9700K 3.6GHz and 32GB of RAM.

The libraries used were *Keras*, based on *Tensorflow*, for learning and preprocessing, *sklearn* for metrics and *numpy* for generic calculation

3.2 Preprocessing and Data Augmentation

In order to raise the number of samples it was used `ImageDataGenerator` from *Keras*. This allowed to virtually loop infinitely on the images during the training and most importantly to transform each of these images by applying none, one or more of the following transformations:

- Horizontal flipping
- Rotation of angle $\alpha \in [-20°, 20°]$
- Shift of brightness $\gamma \in [0.7, 1.3]$
- Zoom $\zeta \in [0.8, 1.2]$

Additionally every image is transformed with the original preprocessing function used for the training of the original architecture on *ImageNet*.

3.3 Transfer Learning

In order to facilitate the learning process

4 Results and Evaluation

The Results section is dedicated to presenting the actual results (i.e. measured and calculated quantities), not to discussing their meaning or interpretation. The results should be summarized using appropriate Tables and Figures (graphs or schematics). Every Figure and Table should have a legend that describes concisely what is contained or shown. Figure legends go below the figure, table legends above the table. Throughout the report, but especially in this section, pay attention to reporting numbers with an appropriate number of significant figures.

5 Discussion

The discussion section aims at interpreting the results in light of the project's objectives. The most important goal of this section is to interpret the results so that the reader is informed of the insight or answers that the results provide. This section should also present an evaluation of the particular approach taken by the group. For example: Based on the results, how could the experimental procedure be improved? What additional, future work may be warranted? What recommendations can be drawn?

6 Conclusions

Conclusions should summarize the central points made in the Discussion section, reinforcing for the reader the value and implications of the work. If the results were not definitive, specific future work that may be needed can be (briefly) described. The conclusions should never contain "surprises". Therefore, any conclusions should be based on observations and data already discussed. It is considered extremely bad form to introduce new data in the conclusions.

References

[1] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2008.