Assignment 1: Multi-Object Detection with CNNs

Course held by: Prof. Giuseppe Serra, Dr. Beatrice Portelli, Dr. Simone Scaboro

November 23, 2022

1 Instructions

DEADLINE
Wednesday December 7th, 2022, 23:59

1.1 Submission instructions

- you can do this assignment either alone or in group (at most 3 students)
- please submit no more than *one submission* per group
- alongside your code, please submit a *short report* (in .pdf), containing a written description of what you have done, including used methods, results, discussion, conclusions and eventual references
- indicatively, the report should not exceed 3 pages of text (images and tables don't count towards this limit)
- in case of multiple source code files please summarize the content of each .py/.ipynb file in the report
- put both the source code and the report in a .zip file named after your student id: "idstudent.zip" for individual works and "idstudent1_idstudent2_idstudent3.zip" for group works (e.g. "012345_678901.zip" or "012345_678901_124311.zip")
- the .zip file must be uploaded to the appropriate folder of the e-Learning page of this course no later than the specified deadline
- any submission that will be submitted after the deadline won't be evaluated.

1.2 Source code instructions

- be sure to write your code in Python 3
- you can use the libraries seen in class (pandas, sklearn, numpy, matplotlib...). If you choose to use additional libraries, please mention them in the report and explain why you chose them. The only **constraint** is to use PyTorch (not TensorFlow, Keras, or other alternatives) for the construction and training of the model.
- please comment your code. The code can be presented as either a .py (python source code) or a .ipynb (Jupyter Notebook) file
- if you have more than one source code file please name them so that it is easy to understand what's inside each file and which file contains the code to run your work
- alongside your code, please submit a *requirements.txt* file which specifies the versions of the used libraries (you can use pipreqs `https://github.com/bndr/pipreqs` to automatically create a minimal clean requirements file). Alternatively, make sure that the code can be run on Google Colaboratory.

2 Assignment

In this assignment, you will get familiarized with CNNs, and you will develop a solution for a **multi-object detection problem**.

Given a dataset of images with people, you will need to implement a model based on CNNs, to predict the **position** and the **type** of their clothes.

You will analyze the role and effect of different architectures and parameters of your CNN model (e.g. number of layers, filters, pooling, normalization, etc.), and report the correct evaluation metrics and observations in a report.

In a nutshell, you have to:

- download the dataset on which you will perform your experiments from the following link: https://drive.google.com/file/d/1LNKILhCM6hvC1unvxdt3mLa4jUEQvuIN/view?usp=sharing (password: *dluniud2022*). The dataset consists of two parts: train and test. Use the training samples to train your model (part of it can be used as a validation set), and the test set to test it and reports its performance

- you have to read, parse and load the data. The JSON files with the annotations are full of information but what you need for this assignment are only two features: `bounding_box` and `category_name`. Remember that transformations can be applied to your images for improving their representation (e.g. cropping, contrast, saturation changes)

- you have to design the architecture of your model. Building your CNN model is one of the most important issues for this task, and you should try different architectures and configurations, including a different number of layers (e.g. experiment with max. 10 layers), filters (e.g. check the Conv2D function from `torch.nn`, which provides many options for configuring CNN parameters). You can use a Sequential model, for which you can specify the types of desired operations after each convolutional layer, including batch normalization and the activation function (e.g. ReLU)

- you need to select a loss function, and an optimization module, along with a learning rate. For the task of object detection, you can find some basic losses online[1]. Don't stress yourself by spending too much time looking for fancy loss functions, it isn't the main goal of the assignment

- train and test your CNN model using the usual pipeline we have seen during the practical lessons. Please try different architectures and parameters and then, based on the performed experiments, select a final model. **The decisions made and the rationale behind them will be explained in the report** (e.g. why did you use specific filter sizes and how did it help at improving the accuracy?). Please add diagrams of the architectures you designed for this task (we suggest using tools like draw.io for this)

[1] https://ambolt.io/how-object-detectors-learn/

- to measure the performance of the model on the test data, please use methods based on Intersection-over-Union to calculate the precision of the bounding boxes (see `https://blog.mturk.com/tutorial-measuring-the-accuracy-of-bounding-box-image-annotations-from-mturk-ad3dfcdf8aa0`). Use accuracy, precision, and recall to measure the performance of the task of cloth type classification

- in any case, display some relevant results of correct and/or incorrect predictions (the image with the bounding boxes, like in the figure above)

- **formulate your critical reflections about the performed experiment and the lessons learnt** (e.g. what is important when constructing a CNN model and how to optimize it?).

Remember that you can optionally compare your model with a pre-trained one like VGG. However, the goal is *not* to achieve state-of-the-art results but to learn how to configure a good CNN model.

Good luck!

2.1 Contacts

For any problem or doubt about the assignment, you can contact us. Our emails are:

- scaboro.simone@spes.uniud.it

- portelli.beatrice@spes.uniud.it