Assignment 1 - Initiate

Inés Broto Clemente

October 2021

1 Project summary

- **Topic:** Classification of images
- **Type:** Bring your own method and bring "part" of the data.

2 Project Description

As it is my first time in Vienna and also my first time living in a German speaking country, I've started to study German. I'm still a beginner and I have to learn a lot of vocabulary in order to be able to start practicing in a more realistic way, here it comes the idea of my project: I will develop an Image Classification tool that given a picture (probably taken in the street) answers with the German word for the object in the picture. There's a lot of literature taking about this topic, most of it using convolutional neural networks to face the challenge and that's also what I'm planning to do. As a baseline, I will build up a VGG-based model [SZ15], a network that combines convolutional layers to extract insights with max pooling layers to reduce dimensions and finally, some fully connected layers are the ones in charge of making the final classification decision. However, more recent papers have introduced the idea of an Inception Structure which has the potential to analyse one image in different resolution (using different convolution sizes) in order to get better insights. This was the base idea for GoogLeNet [SLJ+14], which also proposes an improved utilization of the computing resources inside the network. My idea is to start with the VGG structure and try to optimize it as well and introduce this Inception Structure to try to make it perform as good as possible.

3 Data Description

Looking once again at the literature, the most popular dataset for image classification is definetely ImageNet [RDS+15] which is a large data set of annotated photographs intended for computer vision research. It is organized according to the WordNet (a large lexical database of English) hierarchy. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". There are more than 100,000 synsets in WordNet; the majority of them are nouns (80,000+). In ImageNet, on average 1000 images to illustrate each synset are provided and human-annotated. ImageNet holds 1,281,167 images for training and 50,000 images for validation, organised in 1,000 categories. However, taking into acount the dimensions of this huge data set, I will be working with the tiny version of it, Tiny ImageNet, which contains 100000 images of 200 classes (500 for each class) downsized to 64×64 colored images. Each class has 500 training images, 50 validation images and 50 test images. In the end, then, If there's time left I'll try to work with the hole ImageNet data set but it wouldn't be worth it to work with it while choosing the best architecture and hyper-parameters.
Moreover, since I'm looking forward a German Image Classifier I will translate the labels to German using a static translation map in the beginning.

4 Work break-down structure

It follows a work break-down structure in order to have a proper planning for the work I'll develop during this semester. I've tried to be as realistic s possible since is my first "on-my-own" and I'm not really sure about how much time I'll invest.

Image Classification App

1. Data set collection - 15h
   (a) Downloading or accessing the data - 5h
   (b) Understanding, translating, cleaning and organizing the data - 10h
2. Network designing and building - 10h
3. Network training and fine-tuning - 40h
4. Final application - 30h
5. Final report - 12h
6. Final presentation - 6h

Total amount of hours: 113h

References

[RDS+15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge, 2015.

[SLJ+14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.

[SZ15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.