***Self-assessment – Comănac Dragoș-Mihail***

***Applied Computational Intelligence***

***Advanced Methods in Data Analysis (2022-2023)***

[***dragos.comanac@stud.ubbcluj.ro***](mailto:dragos.comanac@stud.ubbcluj.ro)

***2Y109464***

1. **Introduction**

The research activity I have performed in order to complete the assignments has been strongly focused on the computer vision task of object detection which consists of recognizing various objects in an image by both locating and classifying them. Following the recent trends of the research into this domain, I choose to investigate how a method based on deep neural networks could be used to solve this problem.

Also, the motivation behind my work is that I wanted to investigate how an intelligent system can be integrated in an assistive technology used by visually impaired persons to better understand their surroundings such that they can perform their tasks more efficiently and with less risks. Therefore, to achieve this goal, an accurate, but low computational machine model is needed. Computer vision is a good fit for this kind of solution, because it aims to replicate human vision. Nowadays, cameras even surpass the human eye in terms of quality of the image and computer vision aims to replicate the functions of the visual cortex by understanding the complex scenarios found in images.

1. **Challenges**

Having this motivation in mind, several challenges quickly arrived, mostly related to the development of the machine learning system.

First of all, finding the model architecture fit for the problem was a crucial challenge to overcome. This kind of architecture had to meet several requirements such as being designed for object detection and being based on neural networks, but another important aspect was size. The machine learning model had to be fast and accurate, which means that another challenge was to also optimize the size of the model, in terms of learnable parameters.

The next challenge that occurred was related to the used dataset. Here, various problems were involved regarding class imbalance.

Another challenge was training the model. The optimization process can be quite tricky in the context of a complex task such as object detection, because both classification and localization need to be optimized, and sometimes they might not get along too well, simply because they require different parts of the object.

Lastly, fine-tuning and hyper-parameter tuning were two challenges that once overcome, the object detection model has better perception abilities, essentially for free, because these two processes only affect the training time, not the inference time.

1. **Strengths**

I feel that I have learned a lot during this research and that it helped me improve my skills in this domain substantially.

Firstly, in order to begin the actual research activity, a lot of reading and documenting was needed. I had to extensively study the recent literature on object detection to better understand the task and be able to choose from the vast variety of architectures. As such, I also learned a lot about the history of object detection models and why some are better than others in some regards.

After researching the purpose of various design choices behind object detection architectures, I chose to go forward with the You Only Look Once neural network architecture. This was best suited because it involved a single deep neural network that can efficiently detect objects directly from the image. This is notably faster, but slightly less accurate, than its counterparts, the Region-based convolutional neural networks which are composed of a complicated and hard to train two-stage pipeline.

This process implied the acquisition of several other useful machine learning related skills. For instance, I improved my coding abilities by better understanding how a machine learning framework such as Keras works, including other needed auxiliary libraries.

Also, probably more important is that I acquired a stronger grasp on the whole process of designing, training, and deploying in a useful context a deep convolutional neural network model.

1. **Weaknesses**

If during this work there were a lot of strong positive aspects, as it is natural, there were also several difficulties and limitations involved.

One of the first things that I learned is that it’s better to start from simpler models which is not necessarily that intuitive at the first glance. For a beginner it might make sense to add more and more complexity such as extra layers, but this is actually the wrong move. Ideally, at first a simple model should be used in order to verify the whole pipeline and add extra complexity later. Although, after I started with a simple model, and added gradually more complexity, still it was hard to optimize a very large model, due to overfitting. This might be a limitation of the used dataset.

Another limitation of my solution is that the neural network does not behave as expected when deployed on a mobile device, meaning that the performance is slightly worse than on the original dataset. This is probably mostly due to the calibration of the camera which gives images from a slightly different distribution than that of the dataset.

1. **Opportunities**

This work is far from being perfect, therefore there are many possibilities for improvement. The first obvious improvement would be to investigate some more complex architectures for object detections, or other aspects related to the optimization process. For example, focal loss would address the issue of class imbalance, which should help as the used dataset is a bit unbalanced.

Also, the dataset itself could be improved. In this era of deep learning the quality of the data is usually more important than the architecture itself, therefore improvements here might have a stronger impact. The same can be said about the processing of the images on the mobile device. The performance should be improved by investigating the distribution of the images acquired from the camera.

1. **Threats**

From my point of view, the main risk when developing a machine learning system is the data. The whole point is to learn the complex patterns hidden in the data, therefore if there are inconsistencies in the data, they will be translated into the performance of the algorithm. In my case, I encountered the problem of class imbalance which I overcame by removing some instances of the dominant class and adding new instances of the most underrepresented class by adding the predictions of a dedicated open-source object detector.

Another problem for me was the data processing. Even though at first it seemed that the model learns partially, it still felt that it should be doing more. It turned out that I inversed the location of the bounding boxes in the encoding phase. Technically speaking, the model can learn in this way, but it is an extra unnecessary layer of complexity. This shows that there can be subtle problems in parts of the system which don’t seem that complex.

1. **Impact**

In conclusion, I believe that the impact of this work is two-fold. First of all, from a research point of view, I think it helps in understanding of a widely used architecture and how it behaves in a real-world scenario. Secondly, there is the prospect of using this work in a wider context, such as a mobile cheap and accessible assistive technology that could help visually impaired persons in performing their daily tasks better by acquiring more information about their surroundings. Also, this work could be used not only in an assistive technology, but also to other problems which require fast machine learning models to be run on a mobile device with scarce resources.