EWI3615TU - Computer Science Project

Evaluation Document

Group 1

Gilbert Hardeman	4357337	Daniek Dieben	4448138
Lisa-Marie Plag	4969669	Britt Rooijakkers	4551451
Sean van der Meer	4445546		

Delft University of Technology

January 25, 2019

TU Delft

Delft
University of
Technology

Challenge the future

Contents

1 Initial Motivation

When we first came together, it was clear that we all wanted to do something with computer vision. We all really liked the Computational Intelligence course and wanted to use our newly learned skills during this project as well. Furthermore, this project was called "CS Project Big Data and AI", so we figured that we wanted to do something in combination of Big Data and couple both parts together as to have a straight forward data pipeline. We did not had the Big Data Processing course yet, but it would be parallel to the project so we thought this would be a nice opportunity to apply the course content to this project.

We settled on using YouTube as big data source as we thought we could do something with recognizing objects in YouTube videos. Eventually, we settled on recognizing foods in images and forward this data to the YouTube data set to get recipe videos about this recognized foods returned to the user. We really saw the use of this as it happens very frequently to students that there are 4 leftover ingredients left some night and you don't know what to cook with it. It would be really nice if you could let a program recognize your ingredients and return a recipe video including all leftover ingredients. We all agreed this was an application we would definitely use ourselves if it existed and therefore we were really motivated to accomplish to program it.

However, we had to take into account our (limited) program capabilities and the duration of the project, so we settled on trying to recognize one ingredient in one picture first before trying to recognize more.

2 Objective and Goals

Our main objectives of the project were to correctly implement a convolutional neural network (CNN) for recognizing food objects in images and retrieving and filtering useful data from the YouTube database. Then we needed to couple these two together and make a pipe line where data flows from one part to another, an image as input and video links as output.

The goal for the CNN was to train the neural network on three food objects, bananas, broccoli and apples and make sure that it recognized these images from outside the training set. The goal for the big data part was first to be able to work with the YouTube 8M database and retrieve videos from there. However, this appeared to be really hard and we switched to web scraping instead.

The new goal was to be able to automatically, given a search term from the CNN, use the YouTube search engine for receiving recipe videos. So the returned videos had to be filtered for recipe videos by default. We also wanted to let the user of our program choose how the videos should be filtered, so our program should have other filtering options available as well.

3 Implementation Plan

We planned on splitting the project in two parts: the implementation of the CNN and the implementation of the big data part. The big data part would be retrieving and filtering useful data from the YouTube database given a search term.

In the beginning, we planned that three people were going to work on the CNN and two people on the big data part. In the middle of the project, this switched to three people working on the big data part, because there was still some progress to be made there.

The planning was to first do a lot of research during the first couple of weeks and then start implementing right a way. In this way we also had the Christmas Break in case we had to catch up. Then the two weeks after the Christmas Break were meant for testing and implementation wrap-up and preparing our final presentation.

We also planned to start writing the report immediately, so we could evaluate week by week. So every week we had to write a section about our planning (at the beginning of the week) and what we had accomplished and which obstacles we faced (at the end of the week). In this way we

could also look back every week to have a nice overview of what was done and what needed to be done still. Also, we could evaluate how quick the implementation process went and use this to plan what things we were going to implement next.

4   Final Result

In the end, we built an application where you can make a picture with your web cam or you can upload a picture and the application returns a search for recipe videos for the object that is recognized in the picture.

The CNN is trained on recognizing bananas and broccoli only, so pictures of other images will not return anything and instead give an error message on the screen. This will also be returned if the CNN is not able to recognize the banana or broccoli. As soon as the picture is fed into the program, the CNN gets it as an input and returns one word to the big data part of our program, which on turn scrapes the web for watchable links to recipe videos for the recognized food object.

This is all done inside the scope of a graphical user interface where it is also possible to filter the video search on relevance, upload time, view count, rating, upload date, duration and if the video has subtitles, is 4K or high definition or if the video is live. The search is by default filtered on recipe videos and only videos that have an existing thumbnail and do not contain any disturbing content are returned.

5   Result vs Objective

Our main objectives were to implement a convolutional neural network for recognizing bananas, apples and broccoli in images and to efficiently retrieve and filter YouTube videos using the YouTube 8M database. Then our third objective was to couple these two parts together as to create a controlled data flow or data pipeline.

In the end, we did meet all of our objectives but achieved some of them in a different way. Firstly, the current CNN does only recognize bananas and broccoli, no apples. This was because there were too little labeled images of apples available on the COCO database and we did not want to switch databases. Secondly, we did not use the YouTube 8M database for retrieving YouTube videos. There was really little information available on how to work with this database and how to retrieve data from it. We used web scraping instead. So in the end, we did not use the skills we learned during our Big Data Processing course, but instead used the knowledge we got from the lecture about web scraping in the middle of the project.

Also, we thought it was nice to make a graphical user interface and make a real life application for our program. We added some extra features on this GUI as well. We planned to let our program recognize bananas and broccoli and apples in pictures that the user took on his or her webcam, but we also added that the user can upload pictures from his or her computer or can do manual search. Manual search is an extra feature where the user can provide his or her own search term, in the case the user does not like bananas or broccoli, but wants recipe videos of other products.

Thus, in the end we have built a program which suits all our goals at the beginning of the project (apart from our CNN not recognizing apples), but we achieved some in a different way than we planned. We agreed however that this was not a problem to us as we really liked the end result of our project. Also, we think that the way we had to adapt during the project taught us how to be flexible.

6 Process Evaluation

6.1 Planning

In the beginning of the project, we had heard that the first couple of weeks were meant for research. We did a lot of research, however, this leaded us to start a bit too late with the implementation of our program and we had to catch up during the Christmas break. After this, we had made a lot of progress but still did not couple the two parts of our project together and this lead to a bit of stress for the group in the last two weeks, because the final presentation was approaching quickly. In this last week, we made sure we focused on the important things that were left to do, such as the report, finish the data pipeline and testing.

6.2 Implementation

The implementation of our idea was a bit harder than we thought it would be at the beginning of the project. Building the CNN was particularly hard at first because we did not know much about the subject and where to start. Using examples we first build a draft version with all the parts we needed. A network with all the layers connected to each other, code to feed in the data and to evaluate the performance. After that we could focus on trying out different architectures. Inspired by the current trend research to use a lot of layers we tried that first. We found however that since our task is simpler and our amount of data is limited a shallow network performs better.

The implementation of the web scraping was a bit late, as in the beginning of the project we encountered that it was really difficult to work with the YouTube 8M database. There was only limited information available on the internet on how to retrieve data from it. Also, there were only two people working on this part, where one of them did not have a laptop so he could not really work that efficient on the universities computers and the other one just had her laptop crashed. This was the point where we decided to let one extra person work on this part.

We then settled for web scraping, as this had the exact same (wanted) results as we would have had with the YouTube 8M database. The implementation of this went very fast, also because of the lecture on web scraping we had a couple of weeks earlier. From then on, we worked on a graphical user interface and did a lot of refactoring as well.

The two parts of the project were operational around the same time. Because the plan on how to couple these parts was already planned prior, it made it clear how to connect the two. The coupling started a bit rough due to an interpreter version difference. This however, was resolved quite quickly and the way the graphical user interface was designed made it modular enough to add all functionalities together.

6.3 Overall

Looking back at the end, we all agreed that we had a lot of fun during the project. The main things that we learned were how to structure such a large project and how to evenly divide work load over the five group members. We also learned how to be flexible when we encountered something that we did not expect to happen. For all the problems we faced, we eventually found some solutions.

Something we can improve in in next projects is to start implementing sooner. We did not start too late, but we did had to do some catch up work during the Christmas Break, which we did not plan for initially.

Something we did really well during this project was evaluation. We communicated a lot about our different opinions on how to handle and tackle the problems we faced and we also mentioned what we thought went well. This resulted in a really relaxed atmosphere in the group, especially at the end of the project.