Undergraduate Project Plan – 2017/2018

Project: Image to Speech – An Everyday Assistant for Blind People

Student: Ovidiu-Horatiu Ilie, 3rd year student, Dept. of Computer Science, UCL

Supervisor: Prof. Niloy J. Mitra, Professor of Geometry Processing, Dept. of Computer Science, UCL

Aims and Objectives:

- **Aim:**
  To learn more about Image Segmentation with Convolutional Neural Networks and to create an object detection model capable of identifying common objects in indoor (and even outdoor) scenes.
  To understand how to use text-to-speech and natural language processing tools to create an application capable of describing scenes by identifying objects in images / videos.
- **Objectives:**
  - **1.** Create a good CNN model for object detection, train and evaluate it.
  - **2.** Create and Android app that can use the phone camera and the CNN model to identify objects in images.
  - **3.** Extend app to work with video too – contiguous sequence of frames.
  - **4.** Use text-to-speech to deliver audio information to user – describing scenes.
  - **5.** Try implementing natural language processing to better describe scenes in a human fashion.
  - **6.** Implementing intelligent assistant functionalities – voice control of the app (users being able to interact with the app and to ask for further information).
  - **7.** Create a good documentation for the application, alongside what tools were used.
  - **8.** Describe the model used, how it was trained and how it compares to current methods.
  - **9.** Deploy project for testing and gather feedback from possible users.

Expected Outcomes / Deliverables:

- Trained CNN model for object detection, with high accuracy.
- Full comparison of the model with existent state of the art algorithms.
- Android application with features described in objectives.
- Documentation for how the application was created, how it can be used and design decisions.
- Strategy for testing and evaluating the solution.
- Interim Report and Final Report clearly describing the project and submitted on time.
- Getting feedback from people on how good and useful the solution is.

Work Plan:

- Project start to end of October (4 weeks) – Literature review and CNN study. Defining project details.

- End of October to late November (4 weeks) – Research the implementation of similar methods. Identify datasets for training the model. Find and use pre-trained models for semantic segmentation.
- End of November to end of December (5 weeks) – Start implementation of Android application. Re-iterate requirements. Test and evaluate semantic segmentation so far.
- End of December to late January (5 weeks) – Produce a first working version. Interim report submission. Research ways of improving current segmentation model used.
- Beginning of February to beginning of March (5 weeks) – Improve features of the project. Test with users, get feedback and re-iterate.
- Beginning of March to mid-April (7 weeks) – Create documentation for all components. Polishing the solution. Final report submission.

Milestones:

- Preliminary knowledge for the project acquired and research finished by end of November.
- First working version of the project available by late January.
- Interim report finished by mid-January.
- Final version of the project by mid-March, including testing and polishing the application.
- Final report finished by April.