Master Thesis - Project Plan

Project Goals

The overall goal of the project is to detect and classify golf clubs in images. Images are taken from a portable **Trackman IV** unit and the algorithm should be geared towards running on embedded devices with limited compute capabilities.

(a) The portable **TrackMan IV** unit

(b) Different types of golf clubs

The method consists of using deep neural networks to do detection and classification on images.

The project has a set of steps as defined below:

- **Data Gathering**: Data needs to be collected and annotated. Efficient methods are needed to be able to have a large enough dataset in time.
- **Detection Step**: Make a detection model to crop the club head from the image using a very small NN, *mobileNET* or standard CV techniques
- **Preprocess Step**: Use preprocessing to assist in making the *classification step* easier
- **Classification Step:** Using the cropped images, try to identify club type. Use majority vote over many frames.
- **Embed Device**: Adapt the network to make it run on an embedded device

Metrics

We need to decide on desired and measurable metrics for the overall system. The users will have little use of the system if it has poor performance.

- **Detection**: We desire an acceptable crop but the detection is mostly to help with classification. Thus we can expand the bounding box given by the detection step to make sure, that the whole club are contained within the cropped image. For this step to work in an acceptable manner, a desired IoU (intersection over union) of **50%** is specified as a minimum.

- **Basic Classification**: For 4 main categories of clubs (woods/drivers, hybrids (utilities), irons/wedges and putters) we specify an accuracy of **95%** as being acceptable
- **Extended Classification**: For the distinguishing the different types of irons (5-iron, 6-iron, 7-iron, etc), we specify an accuracy of **90%** as being acceptable
- **Embedded device**: Need an FPS of 5 (it's ok to not process every single frame) and a **85%** accuracy.

Learning Objectives

Upon completion of this thesis, the student should have gained knowledge and skills that will enable him to:

- Give an overview of state of the art data annotation tools for videos.
- Explain methods for generating artificial data, and ideas for making the process of collection and annotation easier.
- Explain different CNN object detection architectures.
- Use CNN object detection architectures and benchmark their performance.
- Explain different CNN classification architectures.
- Use CNN classification architectures and benchmark their performance.
- Run the chosen architectures on an embedded device.
- Construct methods for image preprocessing and data augmentation.

Overall Project Plan

Below is the overview of the general project plan. The report will be written in parallel with the work being done.

Week	Activity	Risk
1	Project plan, data analysis and brain storm	3
2+3	Literature study and enhanced project plan	1
2+3	Data annotation (for detection step)	2
3+4	Basic model (detection) | **IoU 50%**	2
5	Artificial data generation	3
6	Club category classification (classification problem) - 4 categories of clubs | **95% Acc**	3
7	Iron type classification (extended classification problem) | **90% Acc**	4
8	Combination of models	2
9-10	Improvement of model with more elaborate preprocessing and multiple training runs	3
11-12	Run on embedded device (maybe shrink and try smaller networks)	4
12-16	Finishing touches and finishing report (time buffer)	2

Risk Analysis

In the plan the risk is classified using a scale from 1 (no risk) to 5 (high risk). The risk is described as the chance of the activity being delayed.

- High risk is data gathering, annotation and analysis. The is a very good chance that

there is too little data to really do much besides simple category classification of 4 types of clubs. Furthermore there is a good chance that the data can be messy and annotation / labeling is very time consuming.

- Classification step is where we will really see if we have enough data. Detection 4 different types of clubs at 95% accuracy should be doable. Expanding the number of outputs on the neural network or creating more NN's for further detection (not desired in hardware) is very data demanding. If the bottleneck is the data, there is little I can do from a model perspective.
- Improving model can be time consuming and setting up environment for multiple runs with spaced hyper param search is not something I have much experience in. It also requires significant computational resources if it should be fast.
- Embedded device can be very hard depending on the configuration at TrackMan. If I'm given a device that is ready to go it is easy. If I have to start from scratch installing driver etc. for webcam it can take a surprising amount of time.

Status

As of week 4, we currently have around 200 videos. 145 have been annotated and a basic detection model has been constructed. There is not quite enough data, so I have had to rethink the project in terms of data collection and annotation.