# Car Driving without Cameras

By

JONES MABEA AGWATA

Department of Computer Science
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of MASTER OF SCIENCE in the Faculty of Engineering.

SEPTEMBER 2018

Word count: ten thousand and four

# Abstract

The need for robust object detection in 3D point clouds has greatly increased with the ongoing push for autonomous vehicles(AVs). Most of these systems use Light Detection and Ranging(LiDAR), cameras or a combination of both in order to perform object detection [6]. LiDAR presents objects as point clouds in a 3D space thus offering critical shape information of objects in view. However, this representation is sparse and suffers from drawbacks such as occlusion. As a result, LiDAR-based detection performs poorly as compared to multimodal methods that have helped overcome this.[12]. However, multimodal methods are often complex to set up and synchronise [6] with the cost of components running into thousands of pounds. Furthermore, if one of the systems fails, the whole system would be rendered redundant. This can be catastrophic for self driving cars and therefore it is important to address this form of dependence.

# Dedication and acknowledgements

Here goes the dedication.

## Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .................................................... DATE: ..........................................

# Table of Contents

# List of Tables

| Table | Page |
|---|---|

# List of Figures

# Chapter 1

# Introduction

In the coming decade there is expected to be proliferation of new technologies that have been aided by recent developments in Artificial Intelligence(AI) and Machine Learning(ML). AI and ML have become topics of increasing interest in academic fields and industries globally. As a result, collaborations between these two fields has become increasingly common. With increase in computational power and the development of newer AI algorithms robotic technologies have greatly advanced and are now able to perform complex tasks that were once too difficult, dangerous or even impossible for humans to perform. Consequently, industries have benefited from the accuracy and efficiency provided by these robots.

An area that has developed a wide range of interest currently is the topic of autonomous vehicles. The idea of autonomous vehicles is not new and as early as 2005, DARPA had invested heavily in the creation of unmanned trucks and organised for the Urban Challenge to allow for different teams to showcase their unmanned vehicles. However, due to the challenges such as low computational power and underdeveloped AI and ML systems, the resulting implementations were not practical and had a high fault rate of 1 fault in 100 miles compared to the human fault rate of around 1 in 100 million miles. Nonetheless, from this challenge, it was clear that the prospect of autonomous vehicles was plausible and indeed possible.

## 1.1 Why Autonomous Cars

### 1.1.1 Safety

A major argument for self driving cars is the improved safety that they will provide on the road. According to a report released by the United States Department of Transport, National Highway Traffic Safety Administration(NHTSA), around 6,296,000 crashes occured in the year 2015 with

35,092 people losing their lives in these crashes. Of these crashes, around 94% of them were as a result of human error. Worldwide, it is estimated that there were 1.2 million deaths in 2013 due to road crashes. In light of this, self driving cars could greatly reduce the number of road crashes. According to the Eno Centre of Transportation, if 90% of the cars on the road were autonomous, there would be a reduction of 4,220,000 road crashes, this would save 21,700 lives. This is based on the fact that a large number of road crashes are a as a result of human error and therefore autonomous vehicles will be able to significantly reduce this number. However, this estimate is dependent on how well the AV system is designed to be able to handle complex, dynamic driving situations.

### 1.1.2 Environmental Impact

According to the Environmental Protection Agency, more than a quarter of greenhouse gases are from the transportation sector. A major contributor of this is traffic congestion due to various factors such as traffic destabilizing shockwave propagation and road accidents. Autonomous vehicles are able to mitigate this as they are able to gauge and calculate the motion vectors of different objects around them. By using traffic smoothing algorithms and smarter implementations such as the slot mechanism developed by MIT, traffic congestion can be greatly reduced and as a result lower fuel consumption. Furthermore, most companies working with AVs are moving towards electrical vehicles and thus reducing the impact of fossil fuels on the environment.

This project seeks to develop further s

## 1.2 Aims and Objectives

The objectives of the project are as follows:

1. Detailed analysis of state-of-the-art Lidar-Based object detection deep neural networks.

2. Implement voxel feature encoding layer for grouping point clouds

3. Implement region proposal network for object detection from voxels.

4. Test and evaluate the implemented neural network against results of state-of-the-art point cloud object detection methods.

## 1.3 Deliverables

The deliverables are split into technical and analytical.

- **End to end point cloud object detection RPN.** This will be a software implementation of the system that will be publicly available through a Github repository.

- **Evaluation report.** In this report, the following topics will be discussed.

    1. A review of related research and implementations tackling object detection using LiDAR cloud points.
    2. Performance analysis of system and analysis criteria.
    3. A comparison between the implemented system and other state-of-the-art detection systems, potentially through a public benchmark.
    4. The ethical and safety implications of the system and its viability in a real world setting.
    5. Economic analysis of the LiDAR-based system and its potential impact on the development of AVs.
    6. Validation of DNN performance with other public datasets containing data from AVs.

## 1.4 Added Value

This project will implement and open source a sophisticated cloud point detection technique for use in autonomous vehicles.

In doing so, the project will democratize access to proprietary technology by Apple [12] to be used and developed further by future researchers working with object detection in point clouds. Following the detailed review, I intend to propose ways through which this project can be implemented to reduce the cost of AVs in order to make them more viable. If successful, this project will provide a possible framework for the main stream adoption of AVs.

## 1.5 Research scope

The focus of this project is mainly with regard to computer vision, deep learning and robotics. Computer vision is the task of obtaining, processing, analysing and contextualising visual information to produce numerical information that can be understood and manipulated by computers. This is necessary in order to process and analyse the LiDAR data. Deep learning is a broad term used to describe methods that utilise the use of deep neural networks that have a large number of layers. DNNs have become a heavily researched an invested area due to their ability to capture complex underlying models from data. This will be crucial for detecting objects from point clouds. This project will combine existing research using computer vision and deep learning such as [8][12] to develop a Region Proposal Network capable of accurately detecting objects in point clouds.

## 1.6 Report structure

## 1.7 History and Current Setups

In 2009, Google was the first major technology company to announce its self driving car project and within 18 months, it had developed a highly robust system that could handle some difficult roads. In the following years, there was an avalanche of companies that also announced their interest in developing self driving cars ranging from car manufacturers to other technology companies such as Apple, Samsung and NVIDIA. Startups and other tertiary technology company began investing in developing systems that can be used in these autonomous vehicles.

Currently, autonomous vehicles are grouped into 5 different categories by the NHTSA:

- **Level 0** - No autonomy.
- **Level 1** - Basic driver assistance built into vehicle design.
- **Level 2** - Partially autonomous but driver expected to monitor environment at all times.
- **Level 3** - Conditionally autonomous with the driver not required to monitor the environment but is required to take back control if need be.
- **Level 4** - Highly autonomous with the vehicle capable of handling most conditions but the driver has the option to take control.
- **level 5** - Completely autonomous with the vehicle capable of handling all conditions.

### 1.7.1 Componenents of a Self Driving Car

Most self-driving cars consist of 4 main components:

- **LiDAR** - LiDAR provides highly detailed 3D information about the evnironment around the vehicle and objects in it. LiDAR operates by sending out pulses of lasers and recording the reflections of the pulses from objects. By comparing this with the time taken for the lasers to be reflected(time of flight) and their direction, the distance of these objects can be calculated and mapped in a point cloud.

  $$distance = \frac{time \times \text{speed of light}}{2}$$

  To achieve a high level of accuracy, the LiDAR has to send out a large enough number of lasers in different directions fast enough to create an accurate point cloud representation of the environment around it. As such, LiDAR systems have multiple channels(emitter/receiver pairs) angled vertically that emit hundreds of thousands of lasers per second.

#### 1.7.1.1 Cost

LiDAR systems require complex optical systems that are expensive to build. As such they are the most expensive sensor in AVs. Consequentially, the cost of production of LiDARs increase greatly as the number of channels increase. More channels allow for more accurate representations of the surrounding environment which is necesary for safer navigation of AVs, however this would not be economically feasible. As such, different companies have developed different types of LiDAR in order to still produce accurate point clouds at a reasonable price.

  - **Mechanical Mirror**
  - **Solid State**
  - **Optical Phase Array**
  - **Microelectromechanical systems (MEMS)**
  - **3D Flash**

- **Cameras** - Cameras mounted on the vehicle are used for classification and identification of various objects on the road. This is important for recognising traffic rules from traffic signs or road markings as well as determining the nature of objects on the road. Cameras can also be used to create 3D maps of the surrounding environment.By combining two cameras, a stereo image can be captured that provides depth information. Alternatively, by combining a camera and IR Laser sensor for depth estimation, RGB-D images are obtained and mapped in a point cloud.

#### 1.7.1.2 Cost

- **Position Estimators** - Position estimators are a group of sensors used for navigation of the vehicle. These include GPS systems, odometers and gryometers.
- **Distance Sensors** - Distance sensors such as radars and sonars are important for gauging the distance of objects on the road. Radars are the most commonly used distance sensors and they work by transmitting radio waves and recording the reflected radio waves from objects. As compared to cameras and LiDARs, radars work well in a variety of low visibility scenarios such as poor weather. However, the reflectivity of these radio waves depends on the nature of objects, their size, absorbtion characteristics and the transmitting power. As such, it is may not be effective for detecting objects with low absorbtion characteristics such as pedestrians and animals.
- **Processing Unit** - In order to process all the data from the sensors in the vehicle, AVs require powerful processing units in order to be able to process all this data in real time.

Most of the ML/AI algorithms used for detecting and identifying objects from LiDAR and camera data demand large amounts of processing power. This is achieved through the use of CPUs, GPUs, FPGA or combinations with each other.

#### 1.7.1.3 Cost

Due to

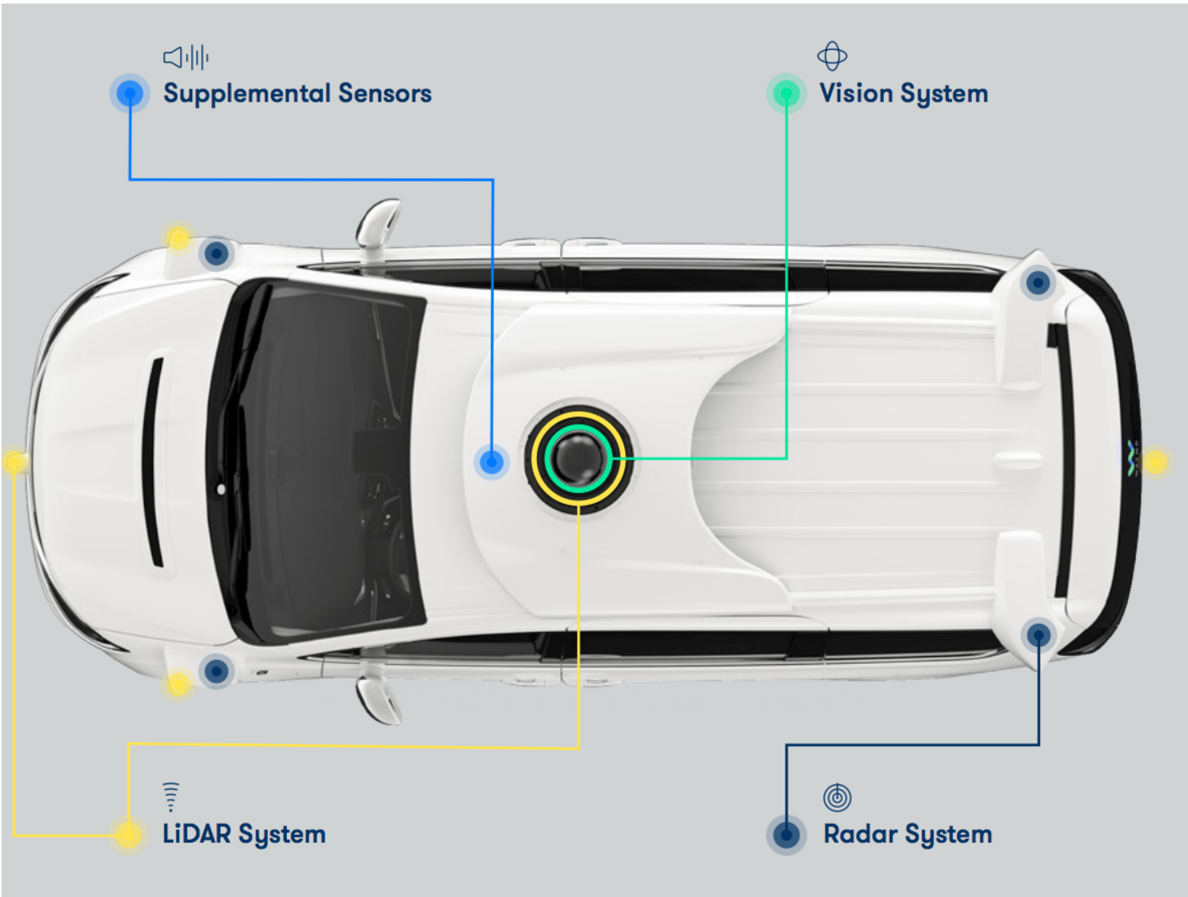



Figure 1.1: Components of Waymo's Self Driving Car (Waymo)

As seen from the table above, a wide range of sensors are used in AVs with each . This is essential for accurate navigation of the AV and therefore multiple sensors are fused together in order to provide enough data to achieve this. Given the large number of sensors, a major inhibiting factor in the production of AVs is the cost of sensors. Table **??** highlights the cost of the various sensors.

- Cost analysis - Attempts to minimise cost - Tradeoffs

## 1.8 Conclusion

As highlighted from the previous section, each of the components serve a crucial purpose in AVs. It is evident that these components complement each other in order to adapt to their shortcomings.

From this a tradeoff between safety, cost, autonomy, power and viability cleary emerges with regard to producing autonomous vehicles. In this case, the following terms can be explained as below

- **Safety**
- **Cost**
- **Autonomy**
- **Power**
- **Viability**

Cost increase viability decrease, safety increase, autonomy increase. power increase, vice versa is true Autonomy increase, cost increase, power increase, safety increase, viability decrease. Power increase, viability decrease, cost increase

# Chapter 2

## Literature Review

AVs are grouped into 5 different categories by the NHTSA:

- **Level 0** - No autonomy.
- **Level 1** - Basic driver assistance built into vehicle design.
- **Level 2** - Partially autonomous but driver expected to monitor environment at all times.
- **Level 3** - Conditionally autonomous with the driver not required to monitor the environment but is required to take back control if need be.
- **Level 4** - Highly autonomous with the vehicle capable of handling most conditions but the driver has the option to take control.
- **level 5** - Completely autonomous with the vehicle capable of handling all conditions.

### 2.0.1 Componenents of a Self Driving Car

Most self-driving cars consist of 4 main components:

- **LiDAR** - LiDAR provides highly detailed 3D information about the evnironment around the vehicle and objects in it. LiDAR operates by sending out pulses of lasers and recording the reflections of the pulses from objects. By comparing this with the time taken for the lasers to be reflected(time of flight) and their direction, the distance of these objects can be calculated and mapped in a point cloud.

$$distance = \frac{time \times \text{speed of light}}{2}$$

To achieve a high level of accuracy, the LiDAR has to send out a large enough number of lasers in different directions fast enough to create an accurate point cloud representation of the environment around it. As such, LiDAR systems have multiple channels(emitter/receiver pairs) angled vertically that emit hundreds of thousands of lasers per second.

#### 2.0.1.1 Cost

LiDAR systems require complex optical systems that are expensive to build. As such they are the most expensive sensor in AVs. Consequentially, the cost of production of LiDARs increase greatly as the number of channels increase. More channels allow for more accurate representations of the surrounding environment which is necesary for safer navigation of AVs, however this would not be economically feasible. As such, different companies have developed different types of LiDAR in order to still produce accurate point clouds at a reasonable price.

- **Mechanical Mirror**
- **Solid State**
- **Optical Phase Array**
- **Microelectromechanical systems (MEMS)**
- **3D Flash**

- **Cameras** - Cameras mounted on the vehicle are used for classification and identification of various objects on the road. This is important for recognising traffic rules from traffic signs or road markings as well as determining the nature of objects on the road. Cameras can also be used to create 3D maps of the surrounding environment.By combining two cameras, a stereo image can be captured that provides depth information. Alternatively, by combining a camera and IR Laser sensor for depth estimation, RGB-D images are obtained and mapped in a point cloud.

#### 2.0.1.2 Cost

- **Position Estimators** - Position estimators are a group of sensors used for navigation of the vehicle. These include GPS systems, odometers and gryometers.

- **Distance Sensors** - Distance sensors such as radars and sonars are important for gauging the distance of objects on the road. Radars are the most commonly used distance sensors and they work by transmitting radio waves and recording the reflected radio waves from objects. As compared to cameras and LiDARs, radars work well in a variety of low visibility scenarios such as poor weather. However, the reflectivity of these radio waves depends on the nature of objects, their size, absorbtion characteristics and the transmitting power. As such, it is may not be effective for detecting objects with low absorbtion characteristics such as pedestrians and animals.

- **Processing Unit** - In order to process all the data from the sensors in the vehicle, AVs require powerful processing units in order to be able to process all this data in real time. Most of the ML/AI algorithms used for detecting and identifying objects from LiDAR and camera data demand large amounts of processing power. This is achieved through the use of CPUs, GPUs, FPGA or combinations with each other.

#### 2.0.1.3 Cost

Due to

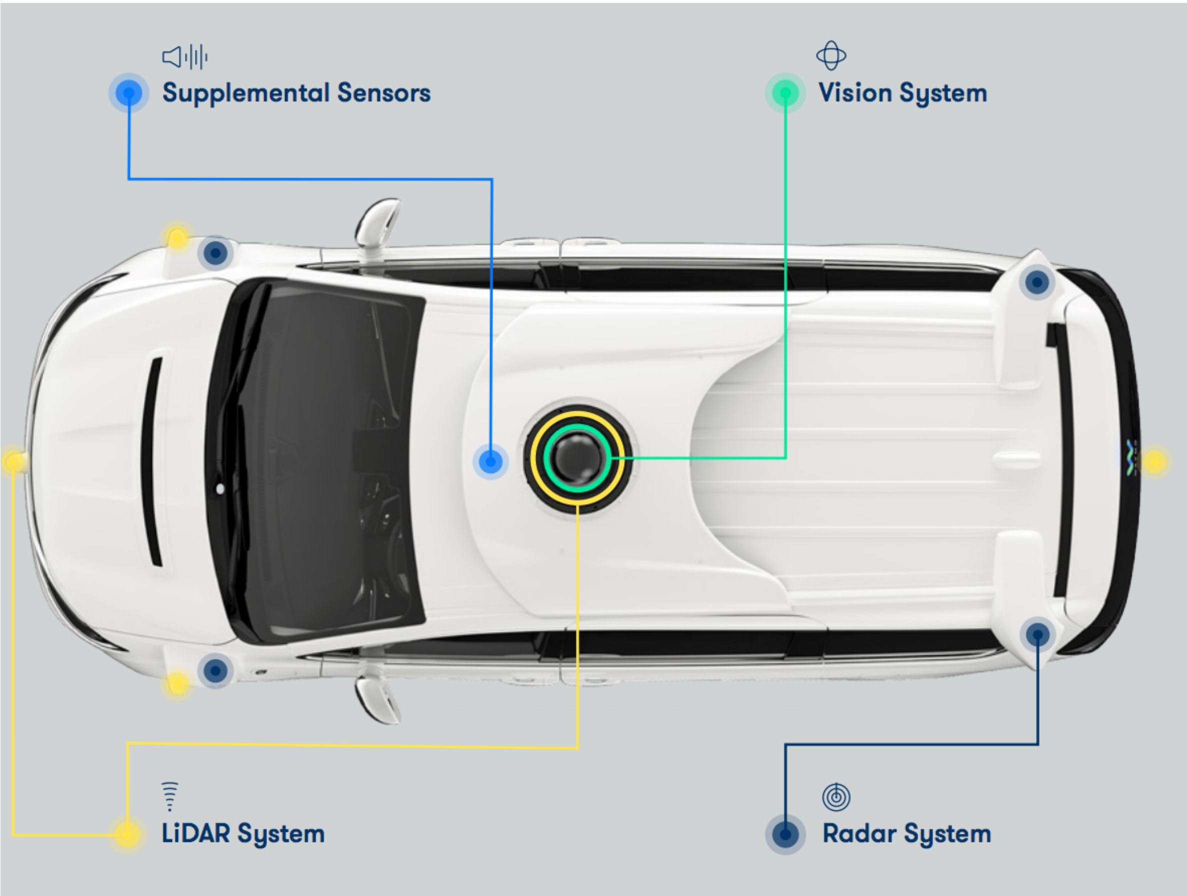



Figure 2.1: Components of Waymo's Self Driving Car (Waymo)

As seen from the table above, a wide range of sensors are used in AVs with each . This is essential for accurate navigation of the AV and therefore multiple sensors are fused together in order to provide enough data to achieve this. Given the large number of sensors, a major inhibiting factor in the production of AVs is the cost of sensors. Table **??** highlights the cost of the various sensors.

### 2.0.2 Legal, Ethical and Economic Considerations

The classic ethical dilemma for self driving cars poses a scenario whereby AVs are presented with a situation whereby a fatal accident is inevitable. For example, the AV either has to crash into a group of people in order to save the life of a passenger or to crash itself and sacrifice the life of the passenger. This dilemma highlights important legal, ethical and economic considerations to be considered by companies involved in the production of AVs and their corresponding systems.

According to [2], there are no public laws that cover the use of independent autonomous vehicles in public spaces. In his article, he cites the fundamental issue as "understanding and accepting the effect of AVs as an independent action by a machine". Due to the lack of public laws on their use, he proposes extending fundamental rights such as the right to life into the framework for creating laws that cover emerging technologies that have an effect on the public, that are otherwise not accounted for in traditional laws.

Bearing this in mind, it is important to note that despite improvements in traffic safety over the years, 94% are still caused by human beings with a large majority of them being fatal with the current human error rate being 1 in 100 million miles. This creates a realistic baseline that can then be extended in evaluating the performance of AV systems. As such, if the risk of automation is lesser than the risk of human vehicle control then the AV would be beneficial. Consequently, the AVs are not to be considered as perfect systems.

With regard to economic considerations, car manufucturers involved in the production of AVs have to ensure that their AVs are able to handle numerous scenarios even if they are rare or considered statistically impossible. They should also be required to take legal liability in case any of their system components are defective and result in failures or accidents. This is important for customer trust without which they will not be able to convince customers to buy AVs. In addition, for the adoption of AVs to be widespread, they have to be reasonably priced. With the current prices of the the various components, the adoption of AVs at the moment is not viable.

In light of the tradeoffs discussed in the previous chapter, different companies working on AVs have come up with different implementations to achieve realisitic AV systems. It is therefore important to highlight the main modes of operation required in such systems which can be grouped in to three main operations.

- **Perception** - This is the first step which involves processing the input from the sensors. In this mode tasks such as object detection and tracking, lane detection, traffic sign detection and recognition are performed.

- **Planning** - After the detection and recognition tasks are performed in the perception stage, route and trajectory planning algorithms are performed. These algorithms are required to handle complex situations to ensure safety of the passengers and other road users.

- **Control** - This stage involves the execution of the plans created in the previous stage. This stage is crucial as the actuators involved in steering and movement have to be able to be able to accurately follow the plans. This involves calculation of energy and forces. At this stage the trajectories and movement of other road users and objects have to be calculated in order to anticipate and avoid any accidents.

## 2.1 Object Detection

A recurring theme that is central to the operation of AVs is object detection. Object detection is crucial for safe operation of AVs as it forms the first step before any planning and control. As such, various companies have been working to come up with accurate object detection systems through different combination of sensors in order to achieve this. To achieve real-time results, deep learning techniques are being developed for this task.

Following the discussion in the previous section, a few requirements have to be fulfilled for mainstream AV adoption. These requirements are as below:

- **Robust**

- **Reproducible**

- **Validatable**

- **Viable**

- **VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection** [12] In this paper, they were able to create a point cloud based 3D object detection network by adding a Voxel Feature Extraction layer before a RPN that was able to divide the point cloud into voxels that were used as input for the RPN. This research forms the main basis of my project and will be replicated with the aim of open sourcing it.

- **Complex-YOLO: Real-time 3D Object Detection on Point Clouds** [11] In this paper, they were able to use point clouds as direct input into Recurrent Neural Networks in order to classify and segment 3D object. They were able to create a global point cloud signature that could be used to achieve this.

- **3D Fully Convolutional Network for Vehicle Detection in Point Cloud** [7] In this paper, the use of a 3D Fully Convolutional Network(FCN) to detect and localise objects in a point cloud as 3D bounding boxes is discussed. They were able to achieve this by discretising the point clouds into square grids represented by a 4D array containing the dimensions and a channel indicating if there was a point at that space in the grid. This paper will be useful in understanding how the point clouds can be represented in a discretised space.

- **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks** [10] In this paper, the idea of real-time RPNs is introduced which combines region-based detectors[4] and CNNs. This is a key paper as the implementation will utilise a RPN that has to be real-time due to the time-sensitive nature of autonomous navigation.
- **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation** [8] In this paper, they were able to use point clouds as direct input into Recurrent Neural Networks in order to classify and segment 3D object. They were able to create a global point cloud signature that could be used to achieve this.
- **PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space** [9] This paper built up on the PointNet implementation by recursively applying a Hierarchical Neural Network on nested partitions of the point clouds. In doing so they were able to capture the local structures of the point clouds as a result of their metric nature and thus achieve better performance than PointNet.

## 2.2 Conclusion

# Chapter 3

# Project Outline

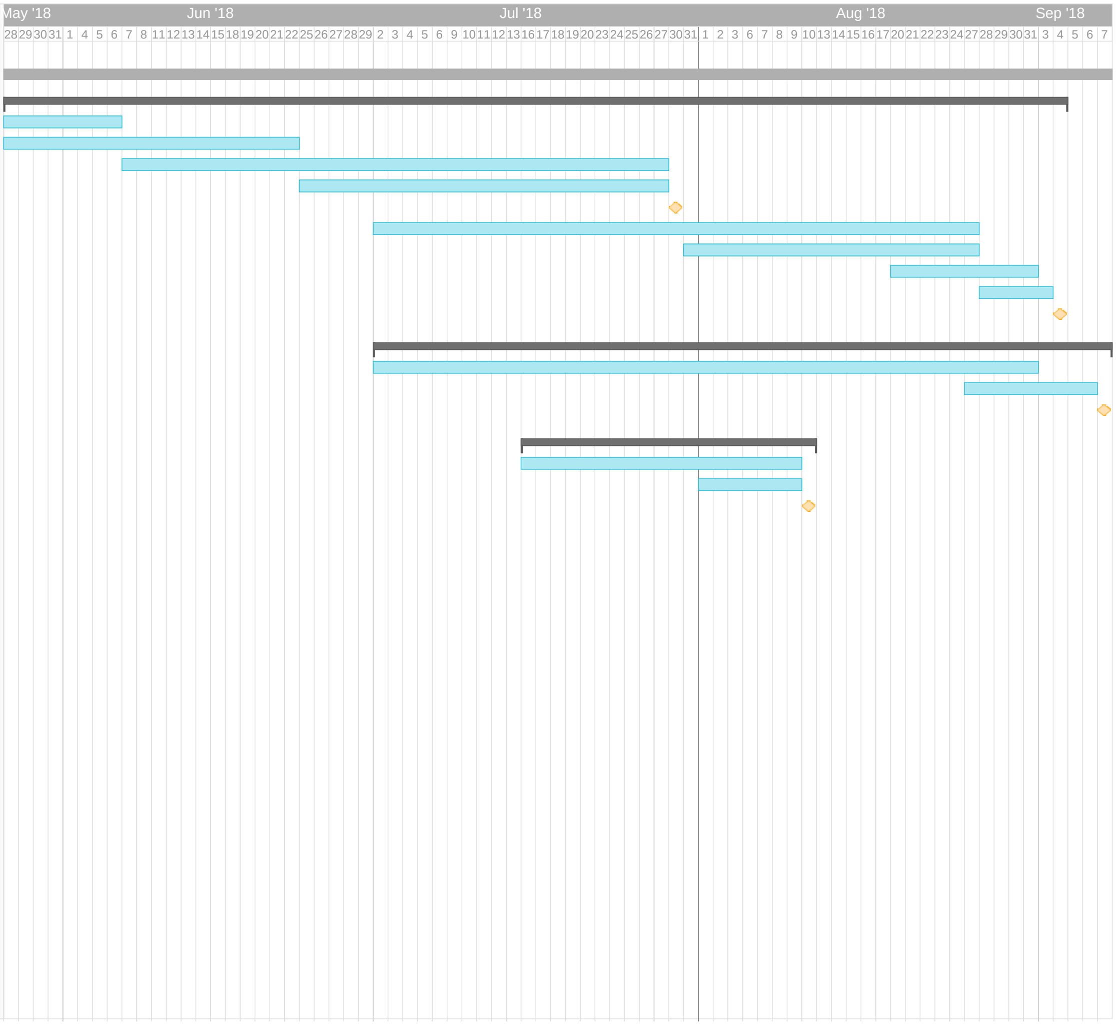
## 3.1 Description of tasks

- **Project environment setup**
- **Review of related research**
- **CNN implementation**
- **Testing and debugging**
- **Training of CNN**
- **Parameter tuning**
- **Evaluation**
- **Validation with different datasets**

teamgantt
Created with Free Edition

Printed: 04/19/2018

| Research Outline | start | end |
|---|---|---|
| **Technical** | **28/05/18** | **04/09/18** |
| Environment setup | 28/05 | 06/06 |
| Related research review | 28/05 | 22/06 |
| CNN implementation | 07/06 | 27/07 |
| Testing and Debugging | 25/06 | 27/07 |
| Working CNN Model | 30/07 | 30/07 |
| Training of CNN | 02/07 | 27/08 |
| Parameter tuning | 31/07 | 27/08 |
| Evaluation | 20/08 | 31/08 |
| Validation | 28/08 | 03/09 |
| Publish source code | 04/09 | 04/09 |
| **Report** | **02/07/18** | **07/09/18** |
| Write up | 02/07 | 31/08 |
| Editing and formatting | 27/08 | 06/09 |
| Submission | 07/09 | 07/09 |
| **Poster** | **16/07/18** | **10/08/18** |
| Design | 16/07 | 09/08 |
| Editing and formatting | 01/08 | 09/08 |
| Submission | 10/08 | 10/08 |



## 3.2 Risk Analysis

#### Technology incompatibility

#### Unrealistic timeline

#### Poor evaluation results

#### Altered requirements

#### Overestimated capabilities

#### Illness

#### Data loss

| Risk Factor | Risk Rank | Contributors | Mitigation and Contingencies |
|---|---|---|---|
| Technology incompatibility | High | Cutting edge software | |
| Unrealistic timeline | High | | |
| Poor evaluation results | Medium | | |
| Altered requirements | Medium | | |
| Overestimated capabilities | High | | |
| Illness | Medium | | |
| Data loss | Low | | Frequent pushing to Github repository |

CHAPTER 4

# CONCLUSION

APPENDIX A

# APPENDIX A

Begins an appendix

# Bibliography


[1] J.-F. Bonnefon, A. Shariff, and I. Rahwan, *Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars?*, arXiv preprint arXiv:1510.03346, (2015).

[2] T. M. Gasser, *Fundamental and special legal questions for autonomous vehicles*, in Autonomous Driving, Springer, 2016, pp. 523–551.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, *Vision meets robotics: The KITTI dataset*, International Journal of Robotics Research (IJRR), (2013).

[4] R. Girshick, *Fast r-cnn*, in International Conference on Computer Vision (ICCV), 2015.

[5] J. Ku, A. Harakeh, and S. L. Waslander, *In defense of classical image processing: Fast depth completion on the cpu*, arXiv preprint arXiv:1802.00036, (2018).

[6] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. Waslander, *Joint 3d proposal generation and object detection from view aggregation*, arXiv preprint arXiv:1712.02294, (2017).

[7] B. Li, *3d fully convolutional network for vehicle detection in point cloud*, arXiv preprint arXiv:1611.08069, (2016).

[8] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, *Pointnet: Deep learning on point sets for 3d classification and segmentation*, Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 1 (2017), p. 4.

[9] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, *Pointnet++: Deep hierarchical feature learning on point sets in a metric space*, in Advances in Neural Information Processing Systems, 2017, pp. 5105–5114.

[10] S. Ren, K. He, R. Girshick, and J. Sun, *Faster r-cnn: Towards real-time object detection with region proposal networks*, in Advances in neural information processing systems, 2015, pp. 91–99.

[11] M. Simon, S. Milz, K. Amende, and H.-M. Gross, *Complex-YOLO: Real-time 3D Object Detection on Point Clouds*, ArXiv e-prints, (2018).

[12] Y. ZHOU AND O. TUZEL, *Voxelnet: End-to-end learning for point cloud based 3d object detection*, arXiv preprint arXiv:1711.06396, (2017).