



بسم الله الرحمن الرحيم

∞∞∞∞

تم رفع هذه الرسالة بواسطة / سامية زكى يوسف

بقسم التوثيق الإلكتروني بمركز الشبكات وتكنولوجيا المعلومات دون أدنى

مسئولية عن محتوى هذه الرسالة.

ملاحظات: لا يوجد





AIN SHAMS UNIVERSITY

FACULTY OF ENGINEERING

Computer and Systems Engineering

## 3D Object Detection for Autonomous Driving using Deep Learning

A Thesis submitted in partial fulfilment of the requirements of the degree of

Master of Science in Electrical Engineering

(Computer and Systems Engineering)

by

**Mohamed Khaled Gamil Ismail Hussein**

Bachelor of Science in Electrical Engineering

(Computer and Systems Engineering)

Faculty of Engineering, Ain Shams University, 2017

Supervised By

**Prof. Dr. Mahmoud Ibrahim Khalil**

Professor at Computer and Systems Engineering Department

Faculty of Engineering, Ain Shams University

**Dr. Bassem Amin Hamed Abdullah**

Associate Professor at Computer and Systems Engineering Department

Faculty of Engineering, Ain Shams University

Cairo - (2022)





AIN SHAMS UNIVERSITY

FACULTY OF ENGINEERING

Computer and Systems Engineering

# 3D Object Detection for Autonomous Driving using Deep Learning

by

**Mohamed Khaled Gamil Ismail Hussein**

Bachelor of Science in Electrical Engineering

(Computer and Systems Engineering)

Faculty of Engineering, Ain Shams University, 2017

## Examiners' Committee

### Name and Affiliation

**Prof. Dr. Hassen Taher Hassen Dorrah**

Prof. at Electrical Engineering,  
Faculty of Engineering, Cairo University

### Signature

.....

**Prof. Dr. Mohamed Watheq Ali Kamel El-Kharashi**

Prof. at Computer and Systems Engineering,  
Faculty of Engineering, Ain Shams University

.....

**Prof. Dr. Mahmoud Ibrahim Khalil**

Prof. at Computer and Systems Engineering,  
Faculty of Engineering, Ain Shams University

.....

**Dr. Bassem Amin Hamed Abdullallah**

Associate Prof. at Computer and Systems Engineering,  
Faculty of Engineering, Ain Shams University

.....

Date: 8 May 2022

# Statement

This thesis is submitted as a partial fulfilment of Master of Science in Electrical Engineering, Faculty of Engineering, Ain shams University.

The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

**Mohamed Khaled Gamil Ismail Hussein**

Signature

.....

Date: 8 May 2022



# Researcher Data

Name : Mohamed Khaled Gamil Ismail Hussein  
Date of birth : 1/12/1994  
Place of birth : Cairo, Egypt  
Last academic degree : Bachelor of Science in Electrical Engineering  
Field of specialization : Computer and Systems Engineering  
University issued the degree : Ain Shams University  
Date of issued degree : 19/09/2017  
Current job : Senior Embedded Software Engineer at eJad





# Abstract

3D Object detection is one of the most important perception tasks needed by autonomous vehicles to detect different road agents like other vehicles, cyclists, and pedestrians which is required for driving tasks like collision avoidance and path planning. State of the art 3D Object detection deep learning models take as input point clouds from very expensive LIDAR sensors because it provides accurate depth information. Another alternative is using cameras which are far cheaper but lack accurate depth information which causes a huge gap in average precision from LIDAR models vs camera models. In this thesis, our work is focused on 3D Object Detection for car class from stereo images without LIDAR supervision neither during training nor during inference and the challenging task of running 3D Object Detection on an embedded target Nvidia Jetson TX2 by modifying Stereo R-CNN model and reducing the model size to approximately one third the size of the original model to be more suitable for embedded targets. Experiments on KITTI dataset showed that our model's inference time is 1.8 seconds and its' average precision for moderate car class is 17% on the test set. Our model decreases training and inference time by approximately 60% with a 13% drop on the test set which is an expected trade-off when decreasing the number of parameters inside the model.

**Keywords: 3D Object Detection, Stereo Vision, Autonomous Driving, Embedded Systems**

# Thesis Summary

In this thesis, we explore 3D object detection using deep learning using stereo images as the only input during both training and inference. Our main contribution is also running 3D object detection on an embedded target Nvidia Jetson TX2 to test the performance of existing model Stereo R-CNN on it and proposing a new model Mobile Stereo R-CNN which is more suited to embedded targets that has faster inference time and less memory footprint with the expected sacrifice in accuracy due to decreasing the model size.

The thesis is divided into 5 chapters, along with a list of figures, a list of tables, a list of abbreviations, and a bibliography.

- In Chapter 1, we introduce the topic of 3D Object Detection and its' importance in the context of Autonomous Driving
- In Chapter 2, we review deep learning and object detection theoretical concepts and we do a literature review for different methods and models used for 3D Object Detection using deep learning
- In Chapter 3, we describe the new proposed model and the modifications made to our base model to make the new model suitable for running on an embedded system device
- In Chapter 4, we describe the experiments made and their results and compare our proposed model with our base model
- In Chapter 5, we give a summary of the work done in this thesis and proposals for future work

# Acknowledgment

I would like to thank my supervisors Dr Bassem and Dr Mahmoud for their guidance and support during my work on this thesis.

I would also like to thank eJad (<http://www.ejad.com.eg>) for their technical and logistics support for this work.

Finally, I would like to thank my family and friends for their support to finish this work.

**Mohamed Khaled, 2022**

# Table of Contents

List of Figures .....	xiv
List of Tables .....	xvi
List of Abbreviations .....	xvii
1 Introduction .....	1
1.1 Autonomous Driving .....	1
1.2 3D Object Detection .....	2
1.3 Main Contribution .....	5
1.4 Thesis Outline .....	5
2 Background .....	6
2.1 Sensors .....	6
2.1.1 LIDAR .....	6
2.1.2 Camera .....	7
2.2 Deep Learning Theoretical Review .....	8
2.2.1 Perceptron .....	8
2.2.2 Activation Functions .....	9
2.2.3 Multilayer Perceptron .....	13
2.2.4 Convolutional Neural Networks .....	14
2.3 2D Object Detection Literature Review .....	16
2.3.1 Metrics .....	17
2.3.2 Two-Stage Detectors .....	19
2.4 3D Object Detection Literature Review .....	25
2.4.1 KITTI Dataset .....	25
2.4.2 Stereo Camera Based Methods .....	26
2.4.3 LIDAR Based Methods .....	28
2.4.4 Fusion Based Methods .....	28
2.4.5 Mono Images Based Methods .....	29
2.5 3D Object Detection Research Challenges .....	29
3 Proposed Method: Mobile Stereo R-CNN .....	30

3.1 Base Model Selection .....	30
3.2 Stereo R-CNN.....	31
3.2.1 Backbone Network.....	32
3.2.2 Stereo RPN.....	32
3.2.3 Keypoint Prediction .....	32
3.2.4 Stereo Regression.....	32
3.2.5 3D Box Estimation.....	32
3.2.6 Loss Function.....	33
3.3 Optimization Approaches .....	33
3.3.1 MobileNetV2 .....	33
3.3.2 Backbone Replacement.....	36
3.4 Software Implementation.....	37
4 Experiments and Results.....	45
4.1 Hardware Setup.....	45
4.1.1 Nvidia GeForce GTX 1080Ti.....	45
4.1.2 Nvidia Jetson TX2 .....	45
4.2 Training.....	46
4.2 Inference.....	46
4.3 Results.....	46
4.4 Experiments and Analysis.....	49
4.4.1 The Original Stereo R-CNN .....	49
4.4.2 Mobile Stereo R-CNN with Base Model Output Depth 256.....	52
4.4.3 Mobile Stereo R-CNN with Base Model Output Depth 128.....	56
4.4.4 Effect of Changing Number of Frozen Blocks in MobileNetV2 .....	59
4.4.5 Experiments Summary.....	59
5 Conclusion and Future Work .....	61
5.1 Conclusion .....	61
5.2 Future Work .....	61
Bibliography.....	64



# List of Figures

Figure 1.1: Levels of Autonomy .....	2
Figure 1.2: 2D Object Detection Example.....	3
Figure 1.3: 2D Object Detection (top) and 3D Object Detection (bottom).....	4
Figure 2.1: LIDAR Pointcloud .....	7
Figure 2.2: Depth Estimation .....	8
Figure 2.3: Perceptron Example .....	9
Figure 2.4: Binary Step Function.....	10
Figure 2.5: Linear Activation Function .....	10
Figure 2.6: Sigmoid/Logistic Activation Function.....	11
Figure 2.7: Tanh Function (Hyperbolic Tangent).....	12
Figure 2.8: ReLU Activation Function .....	13
Figure 2.9: Multilayer Perceptron.....	14
Figure 2.10: Convolution Layer.....	15
Figure 2.11: Max Pooling .....	16
Figure 2.12: Intersection over Union .....	17
Figure 2.13: IoU Calculation .....	18
Figure 2.14: Precision-recall curve [50] .....	19
Figure 2.15: R-CNN architecture [47] .....	20
Figure 2.16: Transformation between Predicted and Ground Truth Bounding Boxes [49] .....	21
Figure 2.17: Fast R-CNN architecture [48] .....	21
Figure 2.18: Faster R-CNN architecture [2] .....	22
Figure 2.19: Pyramid Images vs Pyramid Features [53] .....	23
Figure 2.20: FPN [53] .....	24
Figure 2.21: Faster R-CNN without FPN [52] .....	24
Figure 2.22: Faster R-CNN with FPN [52].....	25
Figure 2.23: KITTI sensor setup.....	26
Figure 3.1: Mobile Stereo RCNN Network Architecture .....	30
Figure 3.2: Depthwise Separable Convolution Block .....	34
Figure 3.3: Bottleneck Residual Block .....	35
Figure 3.4: Mobile Stereo R-CNN PyTorch Model .....	42
Figure 3.5: Software Flowchart .....	43
Figure 3.6: Example of an Input Image .....	43
Figure 3.7: Example of an Input Image After Subtracting Pixel Means .....	43
Figure 3.8: Left Image Feedforward in Backbone.....	44
Figure 4.1: Nvidia Jetson TX2.....	45