



PhD. Program in Electronics: Advanced Electronic
Systems. Intelligent Systems

**Predictive Techniques for Scene
Understanding by using Deep
Learning in Autonomous Driving
applications**

PhD. Thesis Presented by
Carlos Gómez Huélamo

2023



PhD. Program in Electronics: Advanced Electronic
Systems. Intelligent Systems

Predictive Techniques for Scene Understanding by using Deep Learning in Autonomous Driving applications

PhD. Thesis Presented by
Carlos Gómez Huélamo

Advisors

Dr. Luis Miguel Bergasa Pascual
Dr. Rafael Barea Navarro

Alcalá de Henares, TBD

A mi Madre, allá donde esté

*“En este vasto mundo
navegáis en pos de un sueño,
surcando el ancho mar
que se extiende frente a vosotros.
El puerto de destino es el mañana
cada día más incierto.
Encontrad el camino,
cumplid vuestros sueños,
estáis todos en el mismo barco
y vuestra bandera es la libertad”
Espero que te guste Mamá, allá donde estés ...”*
Opening 3 de One Piece.
Autor: The Babystars

Acknowledgements

Esta Tesis Doctoral supone el culmen a cuatro años (Abril 2019 - Abril 2023) realmente duros, cargado de emociones, triunfos, pandemias, estafas y tropiezos, todo a partes iguales. Este es probablemente (aunque como diría Sean Connery interpretando a James Bond en 1983, *Never Say Never Again*) mi último gran documento individual, académicamente hablando. Durante mi etapa universitaria (2013 hasta el momento, 2023) he tenido ciertos momentos puntuales en los que he sentido un salto cualitativo como profesional: El primero fue en el segundo cuatrimestre de segundo de carrera, cuando las cosas se pusieron tensas con Control II e Informática Industrial. Vaya sudores. El segundo probablemente fue con el fallecimiento de mi madre durante mi ERASMUS+ en Irlanda. Duros y oscuros momentos, alejado de mis seres queridos. El tercer momento llega en segundo de máster, durante mi querido ERASMUS+ en Finlandia, donde compagino una estancia preciosa en Tampere con el máster y un pre-inicio de doctorado. Me equivoqué al empezar tan pronto con la beca, "queriendo cobrar" cuanto antes, en vez de terminar tranquilamente el TFM y plantear tranquilamente la tesis, pero eso no lo sabría hasta tiempo después. Por último, toda la tesis han sido sube y bajas, con mala planificación por mi parte, momentos puntuales donde me equivoqué rotundamente al no estudiar PyTorch tras el WAF 2018 tras la sugerencia de mi tutor, el no enfocarme en técnica individual hasta bien entrado el doctorado, no querer hacer nada hasta que no tuviese la teoría perfectamente asimilada, tener demasiado respeto a la Inteligencia Artificial y escurrir el bulto de mi tesis en un compañero mientras yo me dedicaba a integrar y corregir los bugs del grupo que para mí era lo fácil. Mal. Todo mal. Pero todo cambió tras mi segunda estancia, en Estados Unidos, cuando tras llorar por no entender el camino a seguir, decidí crear mi propio camino, con paciencia, fé y práctica y error compaginado con lectura de artículos, para mejorar mi confianza y autoestima, y finalmente logré empezar a entender lo que era el Deep Learning. Gracias a todos mis errores, desventuras y discusiones, a día de hoy, excepto momentos inevitables, me encuentro con muchísima capacidad en atacar prácticamente cualquier problema.

Cada año, desde hace ya varios, mi primera publicación en Instagram viene seguida de la frase "Trabaja duro en silencio y deja que tu éxito haga todo el ruido". Filosofía Kaizen, de mejora y aprendizaje continuo, para así cada día entender el mundo un

poquito mejor. Si toda la dedicación y estudio que he depositado en este trabajo sirven para algo en mi futuro, sé que todo el esfuerzo habrá merecido la pena.

Después de este particular monólogo, a lo cual soy muy propenso y de lo cual mis amigos y compañeros no cesan en su empeño de recordádmelo, debo, como no puede ser de otra manera, dar paso a los agradecimientos.

En primer lugar, me gustaría agradecer a mis profesores del grupo RobeSafe, especialmente a mis tutores Luis Miguel Bergasa Pascual y Rafael Barea Navarro, por ofrecerme estar en el grupo (así como aguantarme) durante todos estos años e intentar que tuviésemos el mejor *experimental setup* y *roadmap* en el laboratorio, aunque no fuese siempre sencillo. Sin dudas considero realmente interesante la temática propuesta en esta tesis doctoral, predicción de agentes en el contexto de vehículo inteligente, ya que entra en el plano filosófico sobre cómo razonar el futuro de los objetos y cómo podría afectar a la capacidad ejecutiva del agente que deba tomar una decisión. Mi mente hace tiempo que cambió y me fijo siempre que conduzco de todo lo que intento reproducir con mis estudios. Habrá que seguir esta tendencia muy de cerca en los próximos años, porque personalmente considero que sus aplicaciones son fantásticas.

A mis tutores en las estancias de doctorado, Christoph Stiller y Eduardo Molinos en el Karlsruhe Institute of Technology (KIT, Alemania) y Wei Zhan y Masayoshi Tomizuka en la University of California, Berkeley (UCB, Estados Unidos). Se suele decir que unas veces se gana y otras se aprende, y yo en estas estancias quizás aprendí demasiado ... No obstante, me guardo grandísimos momentos (admirar las secuoyas gigantes o hacer mi primera escalada en roca entre ellas) y amigos, como Su Shaoshu o Frank Bieder, con los que aún guardo un cierto contacto.

A mis compañeros, mejor dicho, amigos, de laboratorio: Javier Araluce, Rodrigo, Felipe (chavalín), Santiago, Miguel Antunes, Miguel Eduardo, antiguos compañeros como Javier del Egidio, Óscar, Alejandro, Eduardo, Roberto y Pablo, y a nuevos becarios como Fabio, Navil y Pablo. Gracias de corazón por estar ahí, en nuestras charlas sobre tecnología, empleos, el camino correcto a seguir y la vida en general.

Especial mención vuelvo a hacer a Miguel Eduardo y mi compañero Marcos Conde, cuya apoyo intenso ayudó a centrar mi camino en técnica y escritura. Muchísimas gracias por todo lo que aprendí a vuestro lado.

A mis amigos de la universidad, especialmente a Rocío, Juan Carlos, Esther, Sergio, Pablo, Rubén y Adrián Rocandio. Aún me acuerdo de cuando empezamos con la carrera

y como el paso inexorable del tiempo nos moldea a conveniencia. Os deseo lo mejor en vuestro futuro.

A mis buenos amigos Samuel y Adrián, con quien gran parte de mi vida he compartido. Con especial cariño guardo las interminables charlas sobre la vida y el futuro después de Karate, de comer, de cenar, en el coche, siempre quejándonos de la hora que marcaba el reloj al final de tan interminables conversaciones.

A mi familia, uno de los pilares de mi vida. A mi padre Juan Antonio y a mi madre Petra, que en paz descanse, les debo todo lo que soy y es por ello por lo que les estaré siempre agradecido. Querido padre, gracias por ser tan Genaro, arisco y pesado. Siempre has sido mi ejemplo a seguir, aunque cuando tenga 42 años me sigas regañando por subir con las zapatillas puestas. Querida madre, no se muere quien se va, sólo se muere el que se olvida, y tú nunca caerás en el olvido. A mi querida hermana-calili-chessmaster Silvia, con quien tantas regañinas he tenido, pero el cariño que nos tenemos las supera a todas. A mi perrita Nuka (a.k.a. dragón o Nuki-Nuki), cuyos paseos matutinos son probablemente el ingrediente secreto para la elaboración de esta tesis, dando rienda suelta a mi cabeza para imaginar nuevas propuestas mientras miraba el cielo azul.

Al resto de la familia, amigos compañeros, entrenadores y profesores, gracias por todo.

Y, por último, la persona más importante de mi vida ahora mismo. Mi querida Marta, la persona más maravillosa y buena que conozco. Hemos compartido risas, lloros, besos y abrazos. Nunca me cansaré de repetirte lo suave que tienes la piel tras darte un beso en la mejilla y después hacerte de rabiar. Espero que esta situación esté dentro de un while cuya condición sea True.

"Te quiero más que ayer, pero menos que mañana. Hoy, y siempre"

Mi querido lector, disculpa mi monólogo de agradecimientos, es mi forma de ser y la cual tengo por bandera, aunque creo que ha quedado bonito. Podría decir mil anécdotas más de mi doctorado, pero como diría Aragorn, legítimo Rey de Gondor, enfrente de la mismísima Puerta Negra: *Hoy no es ese día*.

Vamos a la lectura importante, que empiece el *Rock and Roll* !!.

Resumen

Este documento ha sido generado con una plantilla para memorias de trabajos fin de carrera, fin de máster, fin de grado y tesis doctorales. Está especialmente pensado para su uso en la Universidad de Alcalá, pero debería ser fácilmente extensible y adaptable a otros casos de uso. En su contenido se incluyen las instrucciones generales para usarlo, así como algunos ejemplos de elementos que pueden ser de utilidad. Si tenéis problemas, sugerencias o comentarios sobre el mismo, dirigidlas por favor a Carlos Gómez Huélamo <carlos.gomezh@uah.es>.

Palabras clave: Autonomous Driving, Deep Learning, Motion Prediction, Scene Understanding.

Abstract

This document has been generated with a template for Bsc and Msc Thesis (trabajos fin de carrera, fin de máster, fin de grado) and PhD. Thesis, specially thought for its use in Universidad de Alcalá, although it should be easily extended and adapted for other use cases. In its content we include general instructions of use, and some example of elements than can be useful. If you have problemas, suggestions or comments on the template, please forward them to Carlos Gómez Huélamo <carlos.gomezh@uah.es>.

Keywords: Autonomous Driving, Deep Learning, Motion Prediction, Scene Understanding.

Contents

Acknowledgements	vii
Resumen	xi
Abstract	xiii
Contents	xv
List of Figures	xvii
List of Tables	xix
List of Algorithms	xxi
List of Acronyms	xxi
1 Introduction	1
1.1 Motivation	1
1.2 Historical Context	2
1.3 Problem Statement	2
1.4 Objectives and Structure of this work	2
2 Related Works	5
2.1 Introduction	5
2.2 Physic-based Motion Prediction	5
2.3 Deep Learning based Motion Prediction	5
2.4 Vehicle Motion Prediction	5

3	Theoretical Background	7
3.1	Kalman Filtering	7
3.2	Convolutional Neural Networks	7
3.3	Recurrent Neural Networks	7
3.4	Generative Adversarial Networks	7
3.5	Attention Mechanisms	7
3.6	Graph Neural Networks	7
3.7	Training losses	7
4	Predictive Techniques for Scene Understanding	9
4.1	SmartMOT	9
4.2	GAN based Vehicle Motion Prediction	9
4.3	Exploring Map Features	9
4.4	Leveraging traffic context via GNN	9
4.5	Improving efficiency of Vehicle Motion Prediction	9
5	Applications in Autonomous Driving	11
5.1	Motion Prediction Datasets	11
5.2	Multi-Object Tracking	11
5.3	Decision-Making	11
5.4	Holistic Simulation	11
6	Conclusions and Future Works	13
6.1	Conclusions	13
6.2	Future Works	13
	Bibliography	15

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

*Aaay, el oro, la fama, el poder.
Todo lo tuvo el hombre que en su día se autoproclamó
el rey de los piratas, ¡GOLD ROGER!
Mas sus últimas palabras no fueron muy afortunadas:
"¿¡MI TESORO!? Lo dejé todo allí, buscadlo si queréis,
ojalá se le atragante al rufián que lo encuentre.*

Opening 1 de One Piece: "We are"

Autor original: Hiroshi Kitadani

1.1 Motivation

Autonomous Vehicles (AVs) have held the attention of technology enthusiasts and futurists for some time as evidenced by the continuous development and research in Autonomous Vehicle Technologies (AVT) over the past two decades, being one of the emerging technologies of the Fourth Industrial Revolution, and particularly of the Industry 4.0. The phrase Fourth Industrial Revolution was first introduced by Klaus Schwab, CEO (Chief Executive Officer) of the World Economic Forum, in a 2015 article in Foreign Affairs (American magazine of international relations and United States foreign policy). A technological revolution is defined as a period in which one or more technologies are replaced by other kinds of technologies in a short amount of time. Hence, it is an era of accelerated technological progress featured by Researching, Development and Innovation whose rapid application and diffusion cause an abrupt change in society. In particular, the Fourth Industrial Revolution is expected to be marked by breakthroughs in emerging technologies in fields such as Artificial Intelligence (AI), Computer Vision, Internet of Things (IoT), fifth-generation wireless technologies (5G), Robotics, 3D printing and the scope of this master thesis, fully autonomous vehicles. The sum of all these advances are resulting in machines that can potentially see, hear and what is more important, think, moving more deftly than humans. Moreover, Industry 4.0 is the subset of the Fourth Industrial Revolution that concerns industry, that is, this concept focuses the existence of factories

in which machines are enhanced with sensors and wireless connectivity, connected to a system that can visualize the whole production line and make decision on its own. In fact, if it is substituted “the whole production” by the environment, the concept refers to a self-driving car. A self-driving car (also known as driverless car or autonomous car) is a vehicle that can sense its environment and moving safely with little or even no human input. They combine a variety of sensors to recognize their environments, such as GPS, camera, Inertial Measurements Units (IMUs), radar, sonar or LiDAR (Light Detection and Ranging). Then, advanced control systems process this sensory information in order to calculate in a proper way navigation paths, traffic signs or detect and track the road obstacles (which is the main purpose of this thesis) to ensure a safe driving. Furthermore, statistics show that 69 including associated states, lives in urban areas. According to the World Health Organization, nearly one third of the world population will live in cities by 2030, leading to an overpopulation in most of them. Aware of this problem, the Transport White Paper published by the European Commission in 2011 indicated that new forms of mobility ought to be proposed so as to provide sustainable solutions for people and goods safely. For example, regarding safety, it sets the ambitious goal of halving the overall number of road deaths in the EU between 2010 and 2020. Nevertheless, this goal does not seem to be easy since only in 2014 more than 25,700 people died on the roads in the EU, many of them caused by an improper behaviour of the driver on the road. Autonomous driving is considered as one of the solutions to the before mentioned problems and one of the greatest challenges of the automotive industry today. The existence of reliable and economically affordable autonomous vehicles will create a huge impact on society affecting social, demographic, environmental and economic aspects. Besides this, it is estimated to cause a reduction in road deaths, reduce fuel consumption and harmful emission associated and improve traffic flow, as well as an improvement in the overall driver comfort and mobility in groups with impaired faculties, such as disable or elderly people. Other industrial applications of autonomous vehicles are agriculture, retail, manufacturing, commercial and freight transport or mining.

1.2 Historical Context

1.3 Problem Statement

[1]

1.4 Objectives and Structure of this work

Most of the current tracking systems are based on traditional techniques. In that sense, the main scope of this work is to study the state-of-the-art of Deep Learning based Multi-

Object Tracking approaches and implement and validate an optimal architecture both in simulation and real world, mainly focused on the Autonomous Vehicles paradigm. It is a hard issue due to the lack of literature in this specific branch of deep-learning based object tracking applications, as shown in [2] [3] [4]. Moreover, in order to achieve the main scope, the following objectives will be met: 1. Researching of current Deep-Learning based Multi-Object Tracking approaches. 2. Study of state-of-the-art software technologies and sensors to perform the MOT problem both in simulation and real-world. 3. Explanation of a real-world project named SmartElderlyCar, including its hardware and software architecture. 4. Propose an architecture for Deep-Learning based Multi-Object Tracking. 5. Validate the proposed architecture for MOT both in CARLA simulator and real world

The organization of this document has been done as follows: • Chapter 2 presents a technical background about current object tracking approaches, including Visual Object Tracking, LiDAR based and sensor fusion. Then, as this master thesis focuses of 2D tracking, challenges in Visual Object Tracking are shown. Then, Deep Learning in MOT is studied in addition to some state-of-the-art approaches. • Chapter 3 focuses on the software technologies used in this master thesis, that is, ROS for sensor communication, PCL as point cloud processing, Docker as a tool to increase the portability and testability of the project and CARLA as simulator environment. • Chapter 4 presents the SmartElderlyCar project, an autonomous electric car able to drive in the University of Alcalá campus, as the reference to develop the architecture proposal of this work. • Chapter 5 shows the Deep Learning based Multi-Object Tracking architecture proposal. This is the main chapter of this master thesis. • Chapter 6 shows the validation of the architecture proposal in the KITTI benchmark as well as quantitative and qualitative results both in CARLA simulator, KITTI benchmark and in the real prototype of the SmartElderlyCar. • Chapter 7 illustrates the conclusions and future works of this project. • Appendix A details how Kalman filter works in the object tracking context. • Appendix B shows the Artificial Intelligence paradigm, including Machine Learning and Deep Learning concepts. In addition, a brief explanation of how CNNs and RNNs work due to its tight relation with object detection and tracking. • Appendix C shows some interesting parts of the code created and developed in order to perform most of exposed tasks throughout this master thesis. • Appendix D illustrates the user's manual so as to install the system requirements and reproduce the obtained results. • Appendix E represents the main hardware and software specifications used in this project. • Appendix F illustrated an estimation of the required budget to develop this thesis.

Chapter 2

Related Works

*Llegaré a ser el mejor, El mejor que habrá jamás
Mi causa es ser su entrenador, Tras poderlos capturar.
Viajaré a cualquier lugar, Llegaré a cualquier rincón
Y al fin podré desentrañar, El poder de su interior.
¡Pokémon! Hazte con todos (solos tú y yo),
Es mi destino, mi misión
¡Pokémon! Tú eres mi amigo fiel,
Nos debemos defender.*

Opening 1 de Pokémon: "Gotta catch 'em all!"

Autor original: Jason Paige

2.1 Introduction

2.2 Physic-based Motion Prediction

2.3 Deep Learning based Motion Prediction

2.4 Vehicle Motion Prediction

Chapter 3

Theoretical Background

*Desde que el mundo cambió, estamos mucho más unidos
con los Digimon, luchamos juntos contra el mal.*

*Algo extraño pasaba, Digievolucionaban,
en tamaño y color, ellos son los Digimon.*

Opening 1 de Digimon: "Butterfly"

Autor original: Kōji Wada

3.1 Kalman Filtering

3.2 Convolutional Neural Networks

3.3 Recurrent Neural Networks

3.4 Generative Adversarial Networks

3.5 Attention Mechanisms

3.6 Graph Neural Networks

3.7 Training losses

Chapter 4

Predictive Techniques for Scene Understanding

Avanzad, sin temor a la oscuridad.

Luchad jinetes de Theoden.

Caerán las lanzas, se quebrarán los escudos.

Aún restará la espada.

Rojos será el día, hasta el nacer del sol.

*Cabalgad, cabalgad, cabalgad hacia la desolación
y el fin del mundo. Muerte, muerte, muerte.*

Discurso de Theoden, Rey de Rohan

El Señor de los Anillos: El Retorno del Rey

4.1 SmartMOT

4.2 GAN based Vehicle Motion Prediction

4.3 Exploring Map Features

4.4 Leveraging traffic context via GNN

4.5 Improving efficiency of Vehicle Motion Prediction

Chapter 5

Applications in Autonomous Driving

*La fuerza de tus convicciones determina tu éxito,
no el número de tus seguidores.*

Reamus Lupin

Harry Potter y Las Reliquias de la Muerte, Parte 2

5.1 Motion Prediction Datasets

5.2 Multi-Object Tracking

5.3 Decision-Making

5.4 Holistic Simulation

Chapter 6

Conclusions and Future Works

*El mundo no es todo alegría y color, es un lugar terrible
y por muy duro que seas es capaz de arrodillarte a
golpes y tenerte sometido a golpes permanente si no se
lo impides; Ni tú ni yo ni nadie golpea mas fuerte que la
vida. Pero no importa lo fuerte que golpeas, sino lo
fuerte que pueden golpearte y los aguantas mientras
avanzas, hay que soportar sin dejar de avanzar.*

¡Así es como se gana!

*Si tú sabes lo que vales, vé y consigue lo que mereces
pero tendrás que soportar los golpes y no puedes estar
diciendo que no estás donde querías llegar por culpa de
él o de ella, eso lo hacen los cobardes y tú no lo eres.*

TÚ ERES CAPAZ DE TODO.

Discurso de Rocky a su hijo

Rocky Balboa

6.1 Conclusions

6.2 Future Works

Bibliography

- [1] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, “A survey on trajectory-prediction methods for autonomous driving”, *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 652–674, 2022.

