End-to-end learning for autonomous vehicles

a narrow approach

Adauton Heringer

Thesis presented to the
Institute of Mathematics and Statistics
of the University of São Paulo
in partial fulfillment
of the requirements
for the degree of
Master of Science

Program: Computer Science
Advisor: Prof. Dr. Flávio Soares Corrêa da Silva

São Paulo
June, 2023

End-to-end learning for autonomous vehicles

a narrow approach

Adauton Heringer

This version of the thesis includes the corrections and modifications suggested by the Examining Committee during the defense of the original version of the work, which took place on June 23, 2023.

A copy of the original version is available at the Institute of Mathematics and Statistics of the University of São Paulo.

Examining Committee:

Prof. Dr. Flávio Soares Corrêa da Silva (advisor) – IME-USP

Prof. Dr. Cláudio Barbieri – Poli-USP

Dr. Andrea Gorrini – Transform Transport

To my lovely wife, Liège Heringer, to whom I share every single word of this dissertation.

To my daughters, Martina and Celeste Heringer, for endless inspiration.

Acknowledgments

My profound gratitude to Professor Flávio Corrêa da Silva, who in the process of being my supervisor, became a wise friend.

Resumo

Adauton Heringer. **Aprendizado end-to-end para veículos autônomos: *uma abordagem restrita***. Dissertação (Mestrado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

Veículos autônomos tem prometido revolucionar nossa civilização. Contudo, nas últimas duas décadas as expectativas tem sido consistentemente frustradas. Baseado na diferença entre inteligência artificial geral e restrita, e equipado com o arcabouço teórico dos imaginários sociotécnicos, criticamos a autonomia geral: o estudo dos veículos autonomos como previsto por seu artificialmente fabricado imaginário sociotécnico utópico. Por outro lado, conceitualizamos autonomia restrita como o estudo de veículos autônomos em contextos limitados. Desta forma, propomos uma abordagem restrita: em vez de treinar um veículo em um contexto ilimitado, delimitamos precisamente o caminho em que ele deve dirigir. Acompanhando os últimos avanços no aprendizado profundo end-to-end, treinamos uma rede neural convolucional para mapear imagens e comandos de alto nível diretamente para o controle do veículo como esterçamento, aceleração e frenagem em um ambiente simulado. Embora este seja um trabalho conceitual e multidisciplinar, nossos resultados indicam que podemos melhorar significativamente o desempenho do veículo ao delimitar seu caminho, e dessa forma contribuir com o avanço da tecnologia autônoma.

Palavras-chave: Aprendizado end-to-end. Veículos Autônomos. Imaginário Sociotécnico. Inteligência Artificial Geral. Autonomia Restrita. Redes Neurais Convolucionais.

Abstract

Adauton Heringer. **End-to-end learning for autonomous vehicles: *a narrow approach***. Thesis (Master's). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

Autonomous vehicles are long promised to revolutionize our civilization. Nevertheless, it has consistently failed to meet expectations in the past two decades. Based on the fundamental difference between narrow and general artificial intelligence and equipped with the theoretical approach of sociotechnical imaginaries, we criticize general autonomy: the study of autonomous vehicles as envisaged by its artificially fabricated sociotechnical imaginary utopia. By contrast, we conceptualize narrow autonomy as the study of context-limited autonomous vehicles. Accordingly, we propose a narrow approach: instead of training a vehicle in a context-free environment, we set clear boundaries for the path the vehicle is supposed to drive. Using the latest advancements in end-to-end deep learning, we trained a convolutional neural network to map images and high-level commands straight to vehicle control, such as steering angle, throttle, and brake, in a simulated environment. Although this is a multidisciplinary conceptual work, our results indicate that by delimiting its path we can significantly improve performance and contribute to the advancements of autonomous technology.

Keywords: End-to-end Learning. Autonomous Vehicles. Sociotechnical Imaginary. Artificial General Intelligence. Narrow Autonomy. Convolutional Neural Networks.

Contents

Chapter 1

Introduction

Autonomous vehicles is one of the most important technologies of our time. It is expected to revolutionize how humans travel and completely change urban commuting scenarios. Besides the convenience of relying upon robots instead of human beings for traveling and commuting, autonomous vehicles are believed to reduce to virtually zero the number of injuries and deaths caused by automobile accidents. It becomes even more important when considering that 1,250,000 people are killed, and 50,000,000 are seriously injured every year in road traffic accidents (World Health Organization, 2015).

Vehicle automation was first envisioned as early as 1918 (Faisal *et al.*, 2019, p. 47), and the first successful demonstrations of autonomous technology date from the 1980s. Currently, there are two major autonomous driving systems. First is the traditional autonomous driving system, which combines several smaller modules such as object detection, semantic segmentation, motion estimation, and others. Second, the end-to-end autonomous driving system is a deep learning self-contained system that maps sensory input such as raw pixel images directly to vehicle control such as steering, throttling, and braking. The latter has gained popularity since Bojarski *et al.* (2016) demonstrated its use for steering a vehicle in an urban area 98% of the time without human interference.

According to several Silicon Valley leaders, autonomous vehicles would have long been in full operation by now, 2022, taking account all promises made through last decade (Topham, 2021; Davies, 2016; Stewart, 2016; Topham, 2020). However, expectations have been consistently frustrated, and we are not even close to seeing autonomous vehicles on streets or highway roads: "*The perspectives have changed since 2015, when it was probably peak hype. Reality is setting in about the challenges and complexity*" says professor Nick Reed to The Guardian (Topham, 2021). These frustrations happened along with significant advances in technologies that surround autonomous vehicles, e.g., deep learning, convolutional neural networks, adaptive cruise control, and lane following, to name some. Likewise, we are still waiting for IBM's bold promise to revolutionize medicine with its AI superdoctor, announced two days after its powerful *Watson* shocked the world with the Jeopardy! victory in 2011 (Strickland, 2019). Delusional expectations, not the technology itself, seem to be the problem.

Jasanoff and Kim (2009) introduces the analytical concept of Sociotechnical Imaginary.

The vision of the desirable future of autonomous vehicles is an exemplary instance of a sociotechnical imaginary. They are fabricated through a complex network of institutional power shared by scientific and non-scientific actors such as automotive manufacturers, government-based research institutes, academia, marketing agencies, and mass media. It creates the image of an inevitable future that has yet to come true, leading people to believe that "it's just a matter of time". In other words, it produces an over-optimistic view of a future that will hardly become true. In this dissertation, we use the theoretical framework of the Sociotechnical Imaginary to break free from this utopia and propose a new approach to autonomous vehicles.

In Artificial Intelligence (AI), there is an important distinction between narrow and general AI. Artificial Narrow Intelligence (ANI) aims to make machines perform specific tasks that "would require intelligence if done by men" as said by Marvin Minsky. *Deep Blue*, a chess player program by IBM that defeated the world champion Garri Kasparov in 1997, is a good example of narrow AI. It requires intelligence to complete a task but has a specific, limited purpose. Despite performing brilliantly at its target task, it has nothing to do with human-like intelligence.

On the other hand, Artificial General Intelligence (AGI) aims for human-like generalizations. Its scope is sufficiently broad to encompass any unforeseen situation. It is a synthetic intelligence whose goal is to deal with any circumstance at human-level performance. When IBM desired to transform *Watson* from a *Jeopardy!* player to a superdoctor, it was migrating from ANI to AGI.

We differentiate narrow from general autonomy using the same terminology as Artificial Intelligence. We conceptualize general autonomy as the area that pursues the goal of having a general autonomous vehicle as envisioned by its sociotechnical imaginary. This futuristic machine is expected to be able to surpass human-level performance in any circumstance and be prepared to handle any unforeseen situation. It is a system designed to replace a human driver in context-free environments. By contrast, we define narrow autonomy as the study of context-limited autonomous vehicles. It aims to achieve smaller, delimited tasks such as traveling on specific paths. As examples of narrow autonomy, driverless trains are fully autonomous since they do not require continuous human supervision, yet they are narrow, with precise context delimitation.

Private cars serve us in individual, specific needs. The vast majority of a personal car lifespan is used for repetitive tasks on the same path, for example, going from home to work, bringing children from school to home, and visiting parents. Those are repetitive tasks since those addresses rarely change. Now, imagine an autonomous vehicle that can drive a specific person from home to work and nothing else. It is not prepared to deal with every intersection it may encounter, yet it is well prepared to drive safely through those specific intersections on the route from home to work. Despite very far from its sociotechnical imaginary, this narrow autonomous vehicle would still be a significant advancement from the current state of the art. Aiming for narrow autonomous vehicles that serve individual needs is much more realistic than aiming for the fabricated utopic visions desired by its sociotechnical imaginary.

We want to contribute to autonomous vehicle studies by proposing a new approach. Breaking free from the fantasized sociotechnical imaginary, we imagine a future where individuals could teach their own cars to drive specific routes autonomously. Haavaldsen *et al.* (2019) has achieved great results using an end-to-end learning approach based on convolutional neural networks. Starting from Haavaldsen *et al.* (2019), we teach our model to drive safely over specific paths so that it can mimic trajectories such as from home to work. Further, we measure how the scope amplitude affects the model's performance. In other words, as we narrow the scope, we expect our autonomous vehicle to operate better.

To make the vehicle learn to recover from perturbations, Haavaldsen *et al.* (2019) randomly injects noise into the vehicle trajectory. Although trivial to do in a simulated environment, noise injection is costly and impractical in the physical world. Since our goal is to contribute to the physical world, instead of noise injection, we use more cameras so that we increase our agent's sensory interaction with the environment. We measure the impact of a higher degree of interaction with the environment on the model's performance.

Although we attempt to quantify our results as much as possible, this is, by all means, and by its very nature, a qualitative multidisciplinary work. By using the sociotechnical analytical concept, we have equipped ourselves with a strong theoretical foundation to criticize how autonomous vehicles are currently viewed. From that, we propose a new approach based on what we conceptualized as narrow autonomy. Finally, we used state-of-the-art knowledge in deep learning and CNN to experiment and measure our proposal. As seen in the section 4, our results indicate that the narrow approach can be more productive for autonomous vehicle development.

The next chapter lays down the theoretical background necessary to support our work. Also, we review the main works on the literature important for this dissertation. In chapter 3 we present our methodology as well as all the setup used for our experiments. Results are then presented in chapter 4, followed by a discussion on our findings in chapter 5. Chapter 6 is dedicated to a general overview of our work and a conclusion of the main insights.

Chapter 2

Background

2.1 Sociotechnical Imaginaries

Defined for the first time by Jasanoff and Kim (2009), the term "sociotechnical imaginaries" has increasingly gained more popularity in recent years. Its theoretical basis has provided a solid foundation in which many works in science and technology studies (STS) are being produced (Sismondo, 2020; Schiølin, 2020; Levidow and Raman, 2020; Lawrence, 2020; Smallman, 2020; Mladenović *et al.*, 2020). Despite the common idea that imagination resides primarily in the individual minds of creative scientists and engineers, STS have shown that visions and expectations of the future are actually embedded in social organization, therefore influencing the practices of science and technology (Jasanoff and Kim, 2009, p. 122). With that, sociotechnical imaginaries cannot only influence, but also shape trajectories of scientific research, development, and innovation.

In a comparative work, Jasanoff and Kim (2009) shows how two highly distinct nation-states, the United States and South Korea, adopted political agendas on nuclear power that guided them through completely different paths over the next 50 years following the end of World War II. Based on how each country's collectively held imaginary on nuclear technology, both scientific and non-scientific actors joined together to move nuclear power development toward what was already imagined by each nation-state: "In the US, the explicit, central move was to create a newly manageable entity, the 'atom for peace', which converts nuclear energy from terrifying to benign form. As custodian of *this* atom, the state represents itself as a responsible regulator of a potential runaway technology that demands effective *containment*. At the same time, the state has largely delegated the task of development and promotion to the private sector. South Korea, by contrast, retains responsibility not just for regulation but also for the development of nuclear power through a logic of self-reliance. In the national imaginary, the implicit source of nuclear power is the *atom for development*" (Jasanoff and Kim, 2009, p.121).

In 1953, American president Dwight Eisenhower gave a speech to the United Nations General Assembly entitled "Atoms for peace". The speech was designed to calm international and domestic anxiety as well as to contain fear: in a 20-minute speech, the words "fear" or "fearful" were used 16 times. One fragment of the speech, in particular, exemplifies the containment necessity president Eisenhower was trying to address: "The United States would seek more than the mere reduction or elimination of atomic materials for military purposes. It is not enough to take this weapon out of the hands of the soldiers. It must be put into the hands of those who will know how to strip its military casing and adapt it to the arts of peace" (Jasanoff and Kim, 2009, p. 126). After World War II, Americans were terrified of the destructive potential of nuclear power. An apocalyptic catastrophe was a real threat if atomic energy was not contained and controlled.

South Korea, however, had a very different view on nuclear power after World War II. Despite the more than 40,000 Koreans estimated to have died in Hiroshima and Nagasaki, the atomic bombings of 1945 liberated Korea from 36 years of harsh colonial rule under Japan. Instead of seeing nuclear power as a dangerous, fearful, and potentially runaway technology that needed to be contained, the Korean people saw it as critical to achieving freedom and sovereignty for their nation. It signified that the role of science and technology in developing a powerful, strong nation-state was indispensable. The following passage in the journal *Donga-ilbo* from 1947 captures well the general thinking at the time regarding nuclear power:

> What did the atomic bomb that defeated Japan in this war teach us? [...] The only thing that would make our fatherland wealthy and strong is the power of science. No matter how independent we are politically, without scientific independence, we will be enslaved again. The total mobilization of science and technology and the scientification of production ... are the only ways to place our beloved nation on a stable foundation of self-reliance and independence.
>
> (Jasanoff and Kim, 2009, p. 131)

Based on this general sociotechnical imaginary, South Korea took responsibility for not only regulating but developing and expanding nuclear technology. Unlike the US, containment was never a concern to be included on any political agenda in South Korea.

These two vastly different sociotechnical imaginaries mentioned above led those two nations through totally divergent paths. By the end of 20$^{\text{th}}$ century, levels of development of nuclear technology in these countries were entirely unlike. While the American government instituted a virtual ban on the construction of any new nuclear power plants from the 1970s onward, South Korea developed model plants, trained two generations of nuclear scientists, physicists, and engineers, and went from external dependent to one of the leaders in the exportation of nuclear technology to the world: "The US 'contained' its nuclear power know-how to the point of virtual paralysis [...]. South Korea, by contrast, has so thoroughly incorporated nuclear know-how into its scientific and political practices that it now offers resources for the region, and possibly beyond" (Jasanoff and Kim, 2009, p. 140).

The comparison between the US and South Korea serves as a brief introduction to the concept of sociotechnical imaginary. However, to fully comprehend this idea, it is necessary to discuss the concept of co-production, upon which sociotechnical imaginary is built. STS has long stated that scientific knowledge does not shape our lives and values unidirectionally. Instead, our norms, values, and sense of how we should organize and govern ourselves symmetrically influence and shape what we make of our nature and society.

The norms and values of a determined society indicate the path science and technology should advance and develop. For that, a technological apparatus produced by a society's scientific advancement is co-produced by its own norms, values, and interpretations of how the future ought to be.

This dynamic can be more profoundly understood on the following passage:

> Briefly stated, co-production is shorthand for the proposition that the ways in which we know and represent the world (both nature and society) are inseparable from the ways in which we choose to live in it. Knowledge and its material embodiments are at once products of social work and constructive of forms of social life; *society cannot function without knowledge any more than knowledge can exist without appropriate social supports.* Scientific knowledge, in particular, is not a transcendent mirror of reality. It both embeds and is embedded in social practices, identities, norms, conventions, discourses, instruments and institutions - in short, in all the building blocks of what we term the social. The same can be said even more forcefully of technology.
>
> (Jasanoff, 2015, p. 2-3, italics ours)

The co-production concept is essential to the STS, especially by its theoretical sophistication when it confronts the idea of "pure science": a science that does not allow itself to be influenced by noises of social and political daily lives. It incorporates the subjectivity that was missing in the previous theoretical approaches to STS and helps to understand scientific progress paths better. However, co-production itself lacks specificity to explain persistent problems in the modern world. Questions such as why attempts to remake the world fail in some cultures and prosper in others remain unanswered: "it is important to understand in a time of globalization why different moral valences attach to new scientific ideas and technological inventions throughout the world, and why differences persist in what we might call the constitutional position of science and technology in political order" (Jasanoff, 2015, p. 5).

In the answers to these questions, the theoretical framework of sociotechnical imaginaries gives its most significant contribution. Initially limited to explaining disparate technological development between nation-states, Jasanoff (2015) expanded the sociotechnical imaginary to embrace other forms of power organizations such as corporations, social movements, and others. In a highly quoted passage, sociotechnical imaginary is defined as:

> collectively held, institutionally stabilized and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology
>
> (Jasanoff, 2015, p. 6)

Unlike mere ideas or fashions, sociotechnical imaginaries are collective, durable, and capable of being performed. Captured by the term "sociotechnical", these imaginaries are at once products and instruments of the co-production of science, technology, and society in the modern world. Also, they are temporally situated and culturally particular

(Jasanoff, 2015, p. 28).

2.1.1 Autonomous Vehicle Sociotechnical Imaginary

Visions of self-driving cars or autonomous vehicles are a perfect example of a sociotechnical imaginary as described above (Braun and Randell, 2020, p. 1). Autonomous vehicles are a future technology that has yet to be developed or is in the process of being developed. It has been co-produced by the efforts of both scientific and non-scientific actors in society: marketing and advertisement agencies; universities and research institutes; nation-states; governments; transport experts; social and mass media; and automobile manufacturers. All these actors are engaged in the process of interpreting, defining, and predicting the contours of an imagined autonomous automobility future (Braun and Randell, 2020, p. 1).

The autonomous vehicle sociotechnical imaginary is the most recent of a series of automobility sociotechnical imaginaries, each of which predicted a utopic automobile future with no success in becoming as envisioned and promised (Gundler, 2013; Curts, 2015) . As stated by Braun and Randell (2020, p. 1): "automobility, to borrow the metaphor used to name the Norman Bel Geddes exhibit at the General Motors pavilion at the 1939/1940 New York World's Fair, is a continuous 'Futurama'."

The definition of sociotechnical imaginary of Jasanoff (2015) in section 2.1 is a good starting point to understand the autonomous vehicles sociotechnical imaginary (Braun and Randell, 2020, p. 3). It is composed of visions of a desirable future (M *et al.*, 2019, p. 60-67); it is spread and institutionally stabilized by private and public actors (Mckinsey & Company, 2013); it is frequently discussed in the mass media and several other public spaces by scientific and non-scientific actors - celebrities, politicians, etc. (Topham, 2021); and "it is represented as a significant scientific and technological development, thus appealing to shared positive evaluations of science and technology" (Braun and Randell, 2020, p. 3).

Several institutions, from small ventures to big corporations and governmental organizations, are investing a large amount of capital in the development of autonomous vehicles. These same institutions not only develop the technology itself but also fabricate and promote the visions of the desirable future of autonomous vehicles (Braun and Randell, 2020, p. 3). As observed by (Urry, 2016, p. 9): "Powerful actors seeking to realize the future they envisaged often deploy complex rhetorical imaginaries and visions of a future 'heaven'." These actors create a complex of positive visions that are promoted and disseminated in the mass media as the "inevitable" future.

This "future they envisage" is a true paradise. The currently spread autonomous vehicles sociotechnical imaginary promises to reduce car accidents to virtually zero by removing the human factor in decision making (Braun and Randell, 2020; Singh, 2018); to increase efficiency and consequently reduce traffic jams (Litman, 2015; Faisal *et al.*, 2019); to encourage longer commute distances since the "driver" never gets tired (Martínez-Díaz and Soriguera, 2018). When we eliminate "human errors", there is nothing left but perfect machines (Sparrow and Howard, 2017). Also, sophisticated rhetorical aspects are undeniable. Exhibitions such as 2019 North American International Auto Show, Mercedes

Benz pavilion at the 2019 Frankfurt motor show, or the one entitled "Cars: Accelerating the Modern World", present autonomous vehicles with moving escalators, lighting, visual and sound effects that "more than anything else resemble a Disneyland exhibit" (Braun and Randell, 2020, p. 3). Even the current terminology of "self-driving", "driverless" or "autonomous" imply a generalization able to completely replace the human in the driving task: frequently used as synonym for "automated", the word "autonomous" actually means "independence" or "self-sufficient" (Shladover, 2018).

This heavenly future is clearly an over-optimistic utopia. While the autonomous vehicle sociotechnical imaginary has led people to believe that "it's just a matter of time", reality has consistently failed to meet expectations: "The perspectives [of autonomous vehicles] have changed since 2015, when it was probably peak hype. Reality is setting in about the challenges and complexity" says professor Nick Reed (Topham, 2021). A truly autonomous vehicle, as envisaged by its sociotechnical imaginary, will hardly ever cease being a futuristic fantasy.

2.2 General vs. Narrow Autonomy

There is an important distinction in terminology when it comes to the field of artificial intelligence (AI). The so-called narrow AI, or Artificial Narrow Intelligence (ANI), is the term used to refer to systems that have specific "intelligent" behaviors in specific contexts (Kurzweil, 2005; Goertzel, 2014; Fjelland, 2020). When Marvin Minsky, one of the pioneers in the field, defines AI as "...the science of making machines do things that would require intelligence if done by men" (quoted from Fjelland, 2020), he is referring to ANI. Exemplary instances of ANI are IBM's chess player program called *Deep Blue* that was able to defeat the world champion Garri Kasparov in 1997; the also IBM's *Watson*, able to defeat two champions in the game *Jeopardi!* in 2011; and the Alphabet's *AlphaGo*, that defeated the world champion Le Sedol in 2016 in the ancient board game *Go*, regarded as being more complex than chess (Fjelland, 2020, p. 3-4).

Emerged as an antonym to "narrow AI", the term general AI or Artificial General Intelligence (AGI) refers to systems that have broader generalization capabilities (Goertzel, 2014, p. 1). While ANI has specific goals, such as winning a chess game, AGI aims for human-like generalizations. As stated by Ben Goertzel, the core AGI hypothesis is:

> The creation and study of synthetic intelligences with sufficiently broad (e.g human-level) scope and strong generalization capability, is at bottom qualitatively different from the creation and study of synthetic intelligences with significantly narrower scope and weaker generalization capability.
>
> (Goertzel, 2014, p. 3)

Borrowing the terminology from AI, we will distinguish between narrow autonomy and general autonomy. General autonomy is the area that pursues the goal of having a general autonomous vehicle as envisioned by its sociotechnical imaginary. As with AGI, this futuristic autonomous vehicle is expected to be able to make human-like general decisions in context-free traffic situations. Even further, since "more people are killed in road crashes than any other form of violent death, war included" (Braun and Randell, 2020, p. 7), general autonomous vehicles are expected to surpass human-level decision-making performance by reducing deaths and injuries on road accidents.

Narrow autonomy, in contrast, is conceptualized as the study of context-limited autonomous vehicles. Breaking free from the current autonomous vehicle sociotechnical imaginary, Narrow Autonomy aims to achieve smaller, context-limited tasks, such as traveling specific paths that serve determined people. This context delimitation is what differs narrow from general autonomy. One instance of narrow autonomy is driverless trains. They are autonomous since they do not require continuous human supervision, yet they are context limited: their path is confined to the rail network (Lunawat, 2022).

The reason for us to make this distinction is that, in this dissertation, narrow autonomy is our main focus. Instead of pursuing the ambitious goal of creating a general autonomous vehicle that would work safely in any city, we narrowed our scope to delimit specific paths in which our vehicle would work well. Anything outside that path is out of scope. In other words, we are using the end-to-end learning approach to teach a vehicle to operate autonomously on a specific path. Then we compare the results with the general approach where data is collected in one town and the trained model is then tested in another town.

2.3 A Brief History of Autonomous Vehicles

Automated vehicles were envisioned for the first time as early as 1918. The first concept was exhibited two decades later by General Motors in the 1939 iconic New York World's Fair (Faisal *et al.*, 2019, p. 47). Research on completely autonomous vehicles has been done since at least the 1980s when several private and public institutions started projects to explore intelligent transportation systems, reaching the first promising results in the area Zhang, 2019; Janai *et al.*, 2020.

In 1986 PROMETHEUS project was launched in Europe, involving more than 13 automotive manufacturers, several government research units, and universities from 19 European countries. In 1995 this team executed the first long-distance travel, from Munich, Germany, to Odense, Denmark. They achieved about 95% of autonomous driving, reaching velocities up to 175 km/h. This same year in the US, Carnegie Mellon University reached an important milestone by completing the first autonomous drive from Pittsburgh, PA to San Diego, CA (Janai *et al.*, 2020, p. 2).

The following year, in the "No Hands Across America" tour, D. Pomerleau and Jochem (1996) completed the "No Hands Across America" tour from Washington DC to San Diego with 98% of autonomy in steering, although manual throttling and braking. A trained neural network was used to steer the vehicle while humans controlled longitudinal direction. It was one of the first works to use the end-to-end approach for lane keeping (Zhang, 2019; Janai *et al.*, 2020).

After "No Hands Across America", the American Defence Advanced Research Projects Agency (DARPA) started the DARPA Grand Challenges in 2003. The objective was that vehicles drive autonomously in an off-road environment without lane markings Zhang (2019, p. 3). DARPA offered $ 1 million prize for the team that first finished the 240 km drive between California and Nevada. None of the teams was able to complete the route. In the following year, the contest's second edition was launched and had five autonomous vehicles completing the route (Janai *et al.*, 2020).

At around this time, DARPA Autonomous Vehicles (DAVE) was introduced using an end-to-end learning system to train a convolutional neural network (CNN) using as input raw images collected by two attached cameras: one at the right and one at the left of the vehicle (Zhang, 2019; Bojarski *et al.*, 2016). Project DAVE was strongly influenced by the pioneering work of D. A. Pomerleau (1988) called ALVINN: Autonomous Landing Vehicle in a Neural Network. This work demonstrated that a trained neural network with an end-to-end approach could steer vehicles on public roads. Although inspiring, project DAVE was not sufficient for the end-to-end approach to become a viable alternative to traditional modular systems: the average distance between collisions was 20 meters in complex locations (Bojarski *et al.*, 2016).

Recently, the end-to-end system created by Bojarski *et al.* (2016) was able to drive 98% autonomously on an urban route from the Nvidia office in Holmdel to Atlantic Highlands, Monmouth county, NJ. Further, this trained CNN traveled 16km of Garden State Parkway without human intervention. This work has inspired several recent studies on the end-to-end approach.

2.4 Autonomous Driving Systems

We divide autonomous driving systems in two large groups: (1) traditional systems and (2) end-to-end systems (Zhang, 2019; Chen *et al.*, 2015)

Traditional Autonomous Driving Systems

Traditional autonomous driving systems are made by a combination of several smaller modules, such as object detection (traffic lights, traffic signs, cars, pedestrians, and others), semantic segmentation (lane keeping), motion estimation, tracking of traffic participants, and reconstruction. Each of these elements represents one different self-contained system, and all these systems combined result in a rule-based control system (Janai *et al.*, 2020, p. 57).

Figure 2.1: *Traditional Autonomous Driving System*

Zhang (2019, p. 4) divide traditional driving systems into five main components as seeing in figure 2.1: (1) perception, (2) localization and mapping, (3) path planning, (4) decision making, and (5) vehicle control. The perception component uses sensors to scan and analyze the environment, similar to the human eyes. It is responsible for lane marker detection, object and obstacle detection, traffic sign recognition, and others. The Localization and mapping component maps the environment and calculates the vehicle's global and local locations. Path planning uses perception and localization outputs to infer possible routes for driving. The Decision-making component uses possible routes as input and outputs the optimal route based on the vehicle's state and environmental information. Vehicle control then is responsible for the driving commands, such as steering, throttling, and braking, that guide the vehicle along the optimal route (Zhang, 2019, p. 4-5).

One of the problems of this approach is that it requires a robust solution to many still open challenges in scientific research, for instance, semantic segmentation (Janai *et al.*, 2020; Zhang, 2019). Another challenge is the amount of necessary manually-labeled data for the perception component to recognize human-design features such as traffic signs, traffic lights, and lane markers. Besides, traditional driving systems are highly dependent on pre-built high-definition maps. Since pre-built, these maps do not respond quickly enough to environmental changes, e.g., construction zones. Many end-to-end autonomous driving systems have been developed and studied as an alternative to these drawbacks.

End-to-end Autonomous Driving Systems

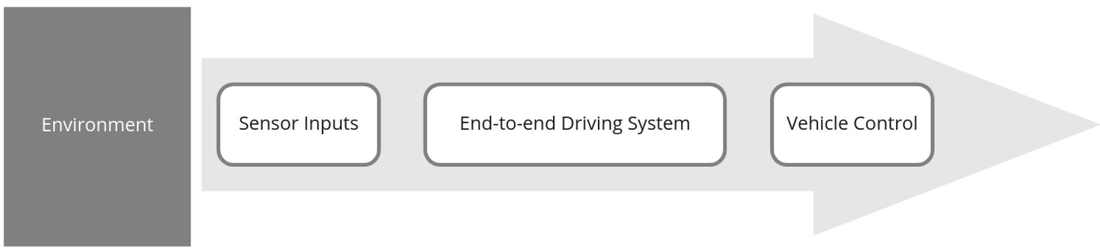


Figure 2.2: *End-to-End Autonomous Driving System*

End-to-end autonomous driving is to use a self-contained system that maps from sensory input such as raw images collected by front-facing vehicle cameras directly to driving actions such as throttle, brake, and steering (Zhang, 2019; Janai *et al.*, 2020; Russell, 2019). These systems are projected to learn directly from expert demonstrations, which means that it does not rely on specific sub-components. For example, the system learns through trained networks that a red sign means it should break, but there is no red sign recognition system for this particular task. End-to-end systems have gained popularity since Alphabet's *AlphaGo* defeated the world champion in the game "Go" using this approach.

Bojarski *et al.* (2016) proposed a deep convolutional neural network trained with raw images from three attached cameras to a vehicle to learn lane keeping. Images were collected and labeled with the steering angle defined by the expert, i.e., the driver. After hours of data collection, the network presented promising results keeping the car in the center of the lane driving through the urban environment of New Jersey (Janai *et al.*, 2020, p. 57).

One of the main advantages of this approach is the simplification of the system compared to traditional, multi-modular methods. It is leaner, with less space for bugs and software engineering problems. Besides, traditional driving systems require their developers to determine what is relevant to the system to observe. For instance, a bird and a horse can appear to be the same size in a pixel image, but a bird could be ignored while the horse could not. The end-to-end approach envisages learning all these policies at once by observing the actions of an expert.

2.5 Literature Review

Since the study and development of autonomous driving systems have become more popular over the last decades, many recent works have attempted to review autonomous vehicle literature and state-of-the-art (Bachute and Subhedar, 2021; Cheng *et al.*, 2022; Ruslan *et al.*, 2023). Other studies aim to assess the future impact of new technologies such as 5G regarding the adoption of autonomous vehicles in our society (Hakak *et al.*, 2023; Orfanou *et al.*, 2022). Finally, several works aim to improve autonomous vehicles' performance by combining different technologies or approaches (Hu *et al.*, 2022; P *et al.*, 2022; Unal *et al.*, 2023).

Some recent works that seem related to this dissertation were judiciously analyzed. One example is Xiao *et al.* (2022), which uses Light Detection and Ranging (Lidar) along with RGB images to create what they called a multimodal approach to improve autonomous vehicle performance. Another example is Lee and Ha (2020), which adds temporal information by applying Long short-term memory (LSTM) to the model, achieving exciting results. However, our work is not about improving autonomous vehicles' performance. Instead, we aim to analyze how reducing the scope of an intelligent agent trained with the end-to-end deep learning approach affects its performance. Because of that, many works on the literature were left out of this literature review.

In his Master's dissertation, Russell (2019) concisely describes the recent advancements in end-to-end deep learning for autonomous vehicles, but not chronologically. Zhang (2019) presents a brief historical overview of the field in his Ph.D. thesis. Finally, Dammen (2019) presents an excellent literature review chronologically so the reader can glance at the critical milestones in this area.

Inspired by Dammen (2019), this literature review is presented chronologically so the reader can understand the historical evolutionary aspects of end-to-end learning for autonomous vehicles. Based on the main studies on the field, we present below the ones we carefully judge as more relevant to the area and have most contributed to this present work. Some of these works are broadly cited in the literature, and their importance is consensual as a technical reference.

1988 - ALVINN: An Autonomous Land Vehicle In a Neural Network

One of the pioneering works on end-to-end learning for autonomous vehicles using neural networks is *ALVINN: An autonomous land vehicle in a neural network*, by D. A. Pomerleau (1988). They proposed a 3-layered neural network to learn the task of road following, i.e., to steer the vehicle so that it keeps itself in the center of the road (in this case, an off-road drive). This autonomous system contained two input sources: a front-facing attached camera that captured 30x32 pixel images and a *laser range finder* - distance sensor - that captured 8x32 pixel images. The output was a 45-length one-hot encoded vector with each unit representing one of the 45 possible steering angles.

ALVINN was trained with simulated images. The real test was done by NAVLAB, the Carnegie Mellon University Navigation Laboratory. The test with simulated images showed that the trained neural network could accurately predict the right steering angle 90% of the time within two units. In the real test, a vehicle could drive at half a meter per second over a 400-meter road in a wooded area. What shocked the scientific community was that a simple network using backpropagation could learn in half an hour what other groups at Carnegie Mellon University would take months. Also, for the first time, it was demonstrated that the end-to-end approach could effectively guide an autonomous vehicle in very specific conditions. The work of D. A. Pomerleau (1988) was of crucial importance for the end-to-end learning field, influencing and inspiring important scientific studies up to date, as is the case of the work of Bojarski *et al.* (2016) 27 years later (Dammen, 2019).

2005 - Off-road obstacle avoidance through end-to-end learning

About fifteen years after ALVINN, LeCun *et al.* (2005) demonstrated the application of an end-to-end learning approach for obstacle avoidance. As in the model of D. A. Pomerleau (1988), the system maps raw pixels directly to the steering angle the robot should apply to avoid obstacles. Input images were collected by two front-facing colored cameras attached to the robot. Data was collected by the demonstration of the experts: humans remotely controlled the robots under different conditions of lighting, weather, and obstacles. The final result showed, as stated by LeCun *et al.* (2005, p. 1), that "The robot exhibits an excellent ability to detect obstacles and navigate around them in real-time at speeds of 2 m/s".

Figure 2.3: *50 cm-long truck robot used by LeCun et al., 2005*

We can point out three main differences between models of LeCun *et al.* (2005) and D. A. Pomerleau (1988): (1) the use of stereo cameras by LeCun *et al.* (2005); (2) the application on obstacle avoidance instead of road following; and (3) the use of a convolutional neural network instead of a fully connected neural network (Dammen, 2019).

One of the problems found by LeCun *et al.* (2005) is multiple viable choices when the robot approach an obstacle. Avoiding the obstacle by either going right or left was equally acceptable. For their measurements, only one way would be the right way, resulting in an accuracy metric fairly high of 35,8%. This problem of multiple viable choices would be later addressed by Codevilla *et al.* (2017) with the approach called conditional imitation learning, further explained later in this section (Dammen, 2019).

2016 - End-to-end Learning for Self-Driving Cars

Called DAVE-2, the work of Bojarski *et al.* (2016) demonstrated that a vehicle could learn to maintain itself in the center of the lane without the external help of what to look in an input image. With a convolutional neural network that maps raw RGB images directly to lateral vehicle control, i.e., steering, the system automatically detects relevant characteristics for that operation, such as lanes and end of roads. As put by Bojarski *et al.* (2016, p. 1): “We never explicitly trained it to detect, for example, the outline of roads”. The data collection was done in some cities and roads in New Jersey for more than 72 hours with three attached cameras: a central one and the others on each side. The CNN contained nine layers: a normalization layers, five convolutional layers, and three fully-connected dense layers.

The results were remarkable, achieving 98% of autonomous driving, i.e., with no human intervention for steering, on the route from Holmdel to Atlantic Highlands in Monmouth County, New Jersey. Also, in Monmouth County, the vehicle drove 10 miles on the Garden State Parkway, a multi-lane road with on and off ramps, with zero intervention.

The name DAVE-2 refers to the 10-year earlier project DAVE - DARPA for Autonomous Vehicle. However, DAVE-2 was highly inspired by D. A. Pomerleau (1988):

> In many ways, DAVE-2 was inspired by the pioneering work of Pomerleau who in 1989 built the Autonomous Land Vehicle in a Neural Network (ALVINN) system. It demonstrated that an end-to-end trained neural network can indeed steer a car on public roads. Our work differs in that 25 years of advances let us apply far more data and computational power to the task. In addition, our experience with CNNs lets us make use of this powerful technology. (ALVINN used a fully-connected network which is tiny by today’s standard.)
>
> (Bojarski *et al.*, 2016, p. 1)

This work is a reference for many recent studies on end-to-end learning. It showed that a CNN could learn the task of lane following with no need for manual decomposition of the problem as needed in the traditional approach. This implies promising advances in autonomous vehicle technology. However, this work only predicted lateral controlling of the vehicle, i.e., steering angles, with longitudinal controlling (throttle and brake) left aside for humans.

2017 - End-to-end Driving via Conditional Imitation Learning

A problem with end-to-end learning is that when facing an intersection, the autonomous agent has two possible ways to go (left or right). Since, from the point of view of the CNN, both are equally correct paths to go, the trained network cannot infer the right way to go only with the image at hand. Codevilla *et al.* (2017) addressed this problem with what they called *conditional imitation learning*. Besides raw pixel images, they used high-level commands as input to simulate the driver's intention so that the CNN could now know which way to go. With this, the intelligent agent acts as a chauffeur by responding to navigational commands given by the passenger.

> At test time, the learned driving policy functions as a chauffeur that handles sensorimotor coordination but continues to respond to navigational commands [...] this resolves some of the ambiguity in the perceptuomotor mapping and creates a communication channel that can be used to guide the autonomous car as one would guide a chauffeur.
>
> (Codevilla *et al.*, 2017, p. 1-2)

Using a CNN inspired by the work of Bojarski *et al.* (2016) and adding the driver's intention as input, Codevilla *et al.* (2017) was able to successfully finish 88% of the time a predefined circuit in the simulated environment of CARLA, in a town that the data was collected (Town 01). In Town 02, where the CNN had no data input, the route was finished 64% of the time. By comparison, the same circuit was completed respectively 20% and 24% of the time when no high-level commands (intentions) were used as input.

The work of Codevilla *et al.* (2017) was the first to use CARLA Simulator (Dosovitskiy *et al.*, 2017), the same we choose to use in this dissertation. Also, whereas all the above-mentioned works have dealt with only lateral control, i. e., steering the vehicle, Codevilla *et al.* (2017) was the first to use an end-to-end approach to aim autonomous vehicle to learn longitudinal controls such as throttling and braking, highly contributing for the advancements in the area (Dammen, 2019).

2019 - Autonomous Vehicle Control: End-to-End Learning in Simulated Urban Environments

The study of Haavaldsen *et al.* (2019) examines the end-to-end learning approach in a simulated urban environment using CARLA Simulator (Dosovitskiy *et al.*, 2017) using Long Short Term Memory to capture temporal dynamics. Starting from the work of Bojarski *et al.* (2016) and adding the high-level commands introduced by Codevilla *et al.* (2017), Haavaldsen *et al.* (2019) were able to improve the performance of the autonomous vehicle by adding temporal dependencies to the model. When the vehicle approaches another vehicle or a red traffic light, using only fixed images is confusing for the CNN to know precisely when to stop and by what intensity to use the brakes. To overcome this, Haavaldsen *et al.* (2019) added the LSTM approach to the CNN to capture the temporal dynamics of the policy to be learned.

The results were quite remarkable. The LSTM-CNN model braked more smoothly when approximating a car or a red traffic light. Out of ten turns of a preset route, the LSTM-CNN model ended successfully five times, while the CNN approach (without LSTM) only finished twice. The average distance traveled was significantly better, 56% and 81%, respectively.

Chapter 3

Methodology

Based on the study of the autonomous vehicles' sociotechnical imaginary, we stated two hypotheses. This methodology is designed to test these hypotheses with some empirical results. This study, however, is a conceptual work, meaning that the central idea is to criticize the current situation of autonomous vehicles' development and all the expectations that reality refuses to meet. This chapter states our hypotheses and describes our plan to test them. For each hypothesis, we elaborated one experiment to test it. Overall, we collect data from a simulated environment, use this data to train a convolutional neural network, and test the trained CNN in the simulator to compare results.

3.1 Hypotheses

3.1.1 Hypothesis 1

Grounded on the theoretical approach of the sociotechnical imaginary proposed by Jasanoff and Kim (2009) and the recent work on end-to-end learning by Haavaldsen *et al.* (2019), we hypothesize that we can improve the performance of autonomous vehicles when using a narrow approach to path learning, everything else constant. Instead of aiming for a generalized vehicle that would work in any circumstances, we envisioned a vehicle that could learn individual needs, such as a specific path that mimics going from home to work. We designed an experiment to test the power of a convolutional neural network to learn a specific path in a simulated environment, as opposed to when the goal is to make an autonomous vehicle that can drive in a never seen path. More, we test how reducing the action's scope of the autonomous vehicle affects its performance. Based on this, we state our first hypothesis:

Hypothesis 1 - The scope amplitude is negatively correlated with the performance of the autonomous vehicle

In other words, as we narrow the scope, we improve the performance of our intelligent agent on its specific task.

3.1.2 Hypothesis 2

Haavaldsen *et al.* (2019) have used noise injection in his work to make the agent recover from perturbations. In their words:

> a model trained only using expert data in ideal environments may not learn how to recover from perturbations[...] To capture more volatile data, a randomly generated noise value was added to the autopilot's outgoing control signal. This resulted in sudden shifts in the vehicle's trajectory and speed, which the autopilot subsequently tried to correct. To eliminate undesirable behavior in the training set, only the autopilot's response to the noise was collected, not the noisy control signal.
>
> (Haavaldsen *et al.*, 2019, p. 4)

In this dissertation, however, we increment the number of cameras used to make the agent learn how to recover from perturbations. Instead of using noise injection to capture data of the vehicle recovering from perturbations, we proposed a multi-camera approach in which we vary the yaw and position of multiple cameras coupled to the vehicle. With this, we were able to collect more data in less time, and by compensating the camera angles with steering offset, our vehicle is able to learn lane following with a high degree of precision, recovering from perturbations and maintaining itself in the center of the lane most of the time. As far as we are concerned, we are the first work to test how the increment in the number of cameras affects autonomous vehicles' performance. With this, we state our second hypothesis:

Hypothesis 2 - The degree to which the agent interacts with the environment is positively correlated with the performance of the autonomous vehicle

As we increase the number of sensors the agent uses, we increase the degree of interaction with the environment and thus improve performance.

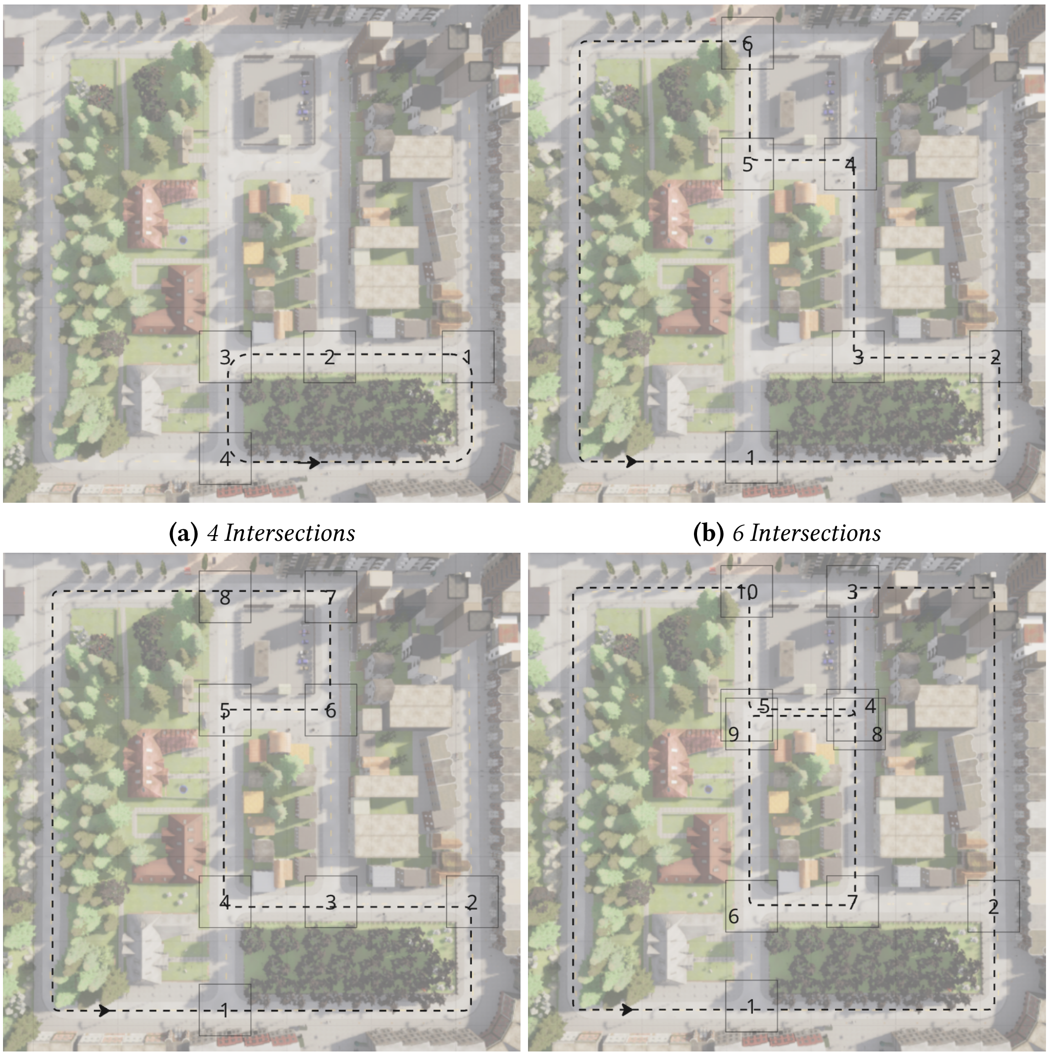
3.2 Metrics

Although this is, by all means, a qualitative work, we want to measure two aspects of autonomous vehicles so that we can test our hypotheses: (1) the scope amplitude; and (2) the degree of interaction with the external environment. Once with these metrics, we can then validate or invalidate our hypotheses. Below we explain how we measured each of our metrics.

3.2.1 Scope Amplitude

When working with general autonomous vehicles, the scope amplitude tends to be infinite. As explained, general autonomy is based on a utopic sociotechnical imaginary with close to no boundaries. The autonomous vehicle is supposed to operate well in any circumstance. On the other hand, when we narrow the scope by delimiting the path subject to our agent, we reduce the amplitude of the scope. In this work, we will use the number of intersections in a path to measure the scope amplitude. It is a proxy metric that has the potential to make our point and test our first hypothesis.

We define four paths with different number of intersections, namely four, six, eight and ten. Each path is shown in figure 3.1. We expect that as we decrease the number of intersections, our autonomous vehicle's performance will increase. As we narrow the scope amplitude, we have clearer objectives and limits for our agent as opposed to a broad, sparse objective such as a general autonomous vehicle. With that, we can validate or invalidate our hypothesis 1.



(a) *4 Intersections*

(b) *6 Intersections*

(c) *8 Intersections*

(d) *10 Intersections*

Figure 3.1: *Four different paths to test hypothesis 1.*

3.2.2 Degree of Interaction with the Environment

The degree of interaction with the environment measures our agent's sensitivity to its environment. We use the number of cameras attached to the vehicle as a proxy variable for the degree of interaction with the environment. As we increase the number of sensors that our agent uses to collect data, we increase its interaction with the environment it is involved.

We expect that the more cameras attached to our vehicle, the better it will perform. To measure that, we run four experiments and compare then. Each autonomous vehicle has a number of cameras attached, namely three, seven, eleven and fifteen. With that, we will have a better understanding of how the degree of interaction relates to the agent's performance and consequently validate or invalidate our hypothesis 2. For more details of each camera position, see section 3.4.1. Table 3.1 summarizes our plan.

Hypothesis	Variable	Metric	Plan
Hypothesis 1	Scope Amplitude	Number of intersections in a path	Compare the performance of models trained over paths with 4, 6, 8 and 10 intersections
Hypothesis 2	Degree of Interaction	Number of cameras attached to the vehicle	Compare the performance of models with 3, 7, 11, and 15 cameras attached

Table 3.1: *Summary of our methodological plan.*

3.3 Environment

3.3.1 Simulator

A simulator can have many advantages over the physical world for academic studies, especially in engineering and computer science. Firstly, it is cost-efficient. Real-life experiments demand expensive tools such as cars, strong GPUs, human resources, large areas, and safety equipment. Secondly, since our main subject is automotive vehicles, testing in the physical world can be dangerous, whereas in simulated environments, we are safe. Finally, by working in a simulated environment, we can quickly test hypotheses and have reasonable indications of whether or not it is worth experimenting with these ideas in the physical world.

One option that has become popular is using real databases made available by several large companies and institutions, such as Nvidia, Kitti, Landmarks, Berkley Deep Drive, and Waymo Open Dataset, to name a few (Niranjan *et al.*, 2021). This work could not benefit from such open, real databases because it lacks an essential part of our data, which are high-level commands. With this, for this work, using a real dataset without the proper labeling - throttle, brake, steering, current speed, speed limit, and high-level commands - would not be possible. As explained in the literature review, Codevilla *et al.* (2017) and Haavaldsen *et al.* (2019) uses high-level commands to overcome the problem of more than one correct path to take when approaching an intersection. As far as we are concerned, we could not find any real, open dataset with high-level commands.

Therefore, this dissertation uses the CARLA (Car Learning to Act) Simulator, an open-source urban environment explicitly created to develop, train, and validate autonomous vehicle systems (Dosovitskiy *et al.*, 2017). There are several driving simulators, such as Udacity, TORCS (Wang. *et al.*, 2019) and AirSim (Dammen, 2019), but CARLA gathers some characteristics that make it the best choice for this dissertation: it is open source, it works on Linux, it has state of the art realism, and it has a server-client architecture.

CARLA's main resources and limitations

CARLA Simulator was created in 2017 with the specific purpose of supporting the development of autonomous driving studies. Its source code is open and free for non-commercial use, as well as its protocols and digital assets resources such as urban layouts, buildings, and vehicles. Also, CARLA is built on top of Unreal Engine 4, thus providing state-of-the-art rendering quality, realistic physics, and basic NPC (non-player character) logic.

Recent studies have deeply analyzed several driving simulators to be used in autonomous vehicle research. According to (Niranjan *et al.*, 2021, p. 7), "CARLA simulator has an advantage over other simulators based on its open source nature, robustness, compatibility, and real-world environment modeling.". Further, Malik *et al.* (2022, p. 1) concludes that CARLA is a "currently state-of-the-art open-source simulator for end-to-end testing of autonomous driving solutions". With that, CARLA Simulator is an excellent choice to meet our expectations in this present work.

Malik *et al.* (2022, p. 1) points out that CARLA's main limitation is its lack of support for off-road environments. Another limitation is its hardware demands, which makes it impossible to use with low-end devices, most commonly encountered in developing countries. To overcome this, CARLA has a synchronous mode with a fixed time step, which means that the client controls the time of the simulated world. This feature is handy because we can reproduce some experiments without depending too much on the processing capacity of the equipment since the simulator time will always be the same. Also, if one has enough processing power, one can run experiments faster than in real life. In this work, we used synchronous mode with a fixed time step of 0.05 seconds, which gave us 20 frames per second (fps).

CARLA's sensors

Among many resources, CARLA has several types of sensors, such as RGB Cameras, RADAR, and LIDAR. Recent works have used LIDAR information and deep neural networks to improve autonomous vehicle performance, such as (Terapaptommakol *et al.*, 2022). In this dissertation, we restrict ourselves to the use of RGB Cameras, following the studies of Bojarski *et al.* (2016), Codevilla *et al.* (2017), Haavaldsen *et al.* (2019), and Dammen (2019). Instead of using new sensors to improve autonomous vehicles' performance, we stick to what has been done in the specialized literature to test our hypotheses. The ultimate goal of this work is to test how scope delimitation and the degree of interaction with the environment impact an autonomous vehicle's performance.

Figure 3.2: *CARLA Simulator.*

3.3.2 Hardware and Software Setup

Two Dell notebooks equipped with Nvidia GPUs were used in this work: one for CARLA server and one for CARLA client. The server is where the CARLA world is simulated, and the client runs scripts for collecting data. These notebooks were also used to build the model and train the CNN. Also, the official Tensorflow-GPU Docker image was used to facilitate Nvidia CUDA setup. A brief summary of the system setup can be seen in table 3.2.

Hardware	
CPU	Intel i7 9750H
GPU	GeForce RTX 2060
RAM	16gb
Software	
Python Dist.	3.8.10
CARLA Version	0.9.13
Tensorflow-GPU	2.9.1
Keras	2.9.0
Numpy	1.22.4
Docker	20.10.17

Table 3.2: *System setup.*

3.4 Data Collection

Data was gathered at a rate of 10 (ten) shots per second, following Haavaldsen *et al.* (2019) proposed work, and the total time for each model was 1 hour at simulation time. For each recorded frame, photos of size 300x180 pixels were taken in different positions. We labeled each recorded image with the control signal applied to the vehicle at that frame, namely steering angle, throttle, and brake. Values of steering angle are scalar values to control lateral movement of the vehicle [-1.0, 1.0], -1.0 being the max angle of that car to the left side and 1.0 the same to the right side. Throttle and brake are scalar values ranging in the interval [0.0, 1.0].

Also, we recorded the vehicle's current speed, the road speed limit, and the high-level command (HLC) the car would take in the next step. HLC is the intention of the user for the autonomous vehicle coded in four types: (1) *turn left at next intersection*, (2) *turn right at next intersection*, (3) *go straight at next intersection*, and (4) *follow lane*, used whenever the vehicle is not close to any intersection and should just keep itself in the center of the lane.

Data	**Description**
Images	300x180 pixel images
Steering	Control steering applied at that frame [-1.0, 1.0]
Throttle	Control throttle applied at that frame [0.0, 1.0]
Brake	Control brake applied at that frame [0.0, 1.0]
Current Speed	Snapshot of the speed of the vehicle at that frame
Speed Limit	Speed Limit of the road
High-Level-Command (HLC)	One-hot-encoding vector meaning (1) turn left, (2) turn right, (3) go straight and (4) follow lane

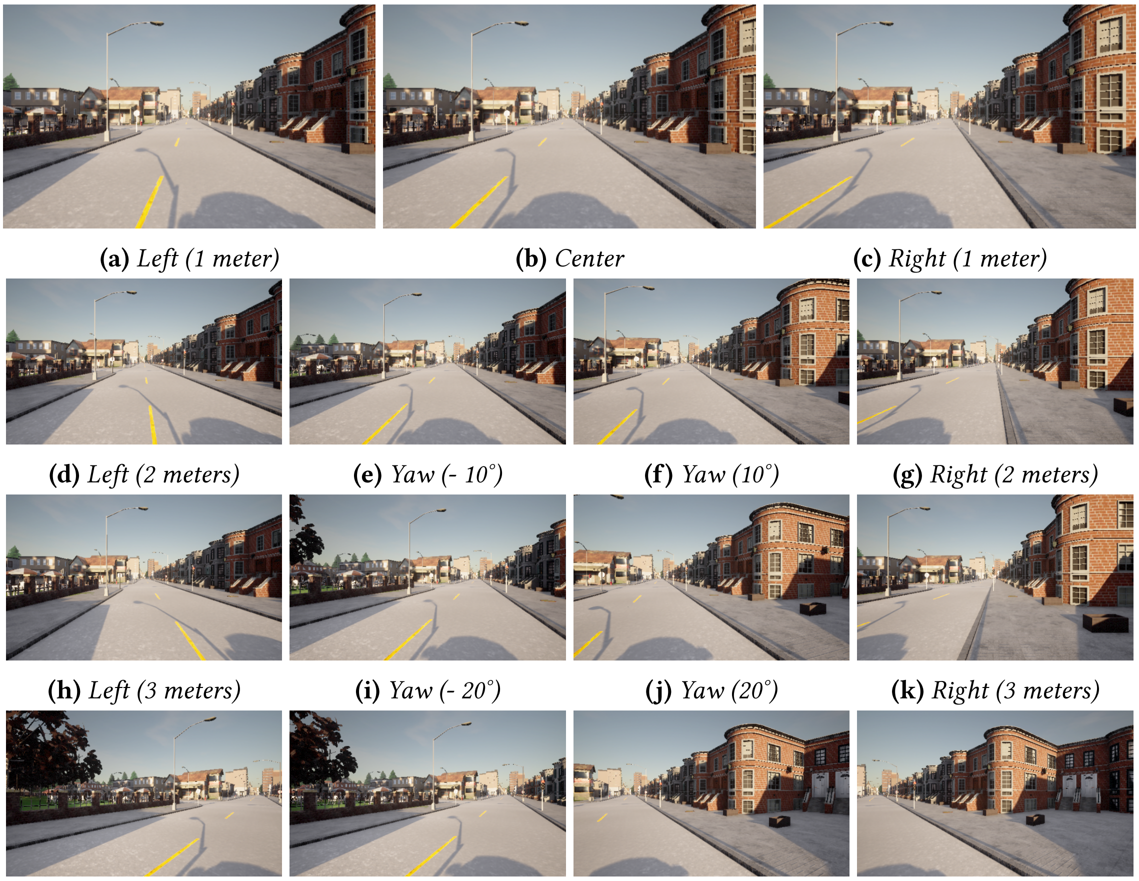
Table 3.3: *Data collected in each observation and its description.*

3.4.1 Camera positioning

To collect data, both Bojarski *et al.* (2016) and Haavaldsen *et al.* (2019) used three forward-facing cameras positioned at the vehicles' left, center, and right sides. Haavaldsen *et al.* (2019) added an offset value to the steering angle to compensate for the camera position: "To counteract the left and right image's positional offset, the associated steering angle was shifted by +0.1 and -0.1 respectively." Our work uses this same rationale, but different offset values were used, and more cameras were added to measure the impact of more cameras on the agents' performance. The offset values were extracted with the CARLA Simulator autopilot as a reference. For instance, when shifting the vehicle one meter to the left, we observe an increment of 0.2 in the steering angle applied by the autopilot to correct the car's trajectory. Table 3.4 summarizes all offset values used.

With that, to test hypothesis 2 we laid out four different models. The first model uses three cameras, (a), (b), and (c) of figure 3.3. The second model has seven attached cameras: the previous three cameras, plus (d), (e), (f), and (g). The third model increments four cameras, (h), (i), (j), and (k). Finally, the last model adds (l), (m), (n), and (o), totaling 15 cameras. Apart from the one positioned at the center, each camera has an associated offset value to compensate for its positional or angle shift.

To test hypothesis 1, all four models uses seven cameras: (a), (b), (c), (d), (e), (f), and (g). The difference between each model is the path where we collect data, seen in figure 3.1.



(a) *Left (1 meter)* **(b)** *Center* **(c)** *Right (1 meter)*

(d) *Left (2 meters)* **(e)** *Yaw (- 10°)* **(f)** *Yaw (10°)* **(g)** *Right (2 meters)*

(h) *Left (3 meters)* **(i)** *Yaw (- 20°)* **(j)** *Yaw (20°)* **(k)** *Right (3 meters)*

(l) *Yaw (- 40°)* **(m)** *Yaw (- 30°)* **(n)** *Yaw (30°)* **(o)** *Yaw (40°)*

Figure 3.3: *One shot at the same frame for all camera positions. The first model uses three cameras, (a), (b), and (c) of figure 3.3. The second model has seven attached cameras: the previous three cameras, plus (d), (e), (f), and (g). The third model increments four cameras, (h), (i), (j), and (k). Finally, the last model adds (l), (m), (n), and (o), totaling 15 cameras.*

Lateral Shift	**Steering Offset**
1 meter	±0.20
2 meters	±0.35
3 meters	±0.60
Angular Shift	**Steering Offset**
10 degrees	±0.20
20 degrees	±0.40
30 degrees	±0.60
40 degrees	±0.80

Table 3.4: *Steering offsets for all camera positions. Negative shifts mean positive offsets, whereas positive shifts mean negative offsets. Also, negative means steering left, and positive means steering right.*

3.4.2 Weather Conditions

CARLA Simulator contains 14 preset weather conditions ranging from sun positioning to climate conditions, such as clear, wet, cloudy, and rainy. From the 14 possibilities, we used a subset of eight presets, excluding night and raining modes. The reason for this is that our CNN cannot generalize for these types of conditions within the short amount of data we are aiming for only one hour of data collection. For comparison, Haavaldsen *et al.* (2019, p. 4) uses only four different weather conditions: Clear noon, cloudy noon, clear sunset, and cloudy sunset. Dammen (2019, p. 17) uses up to eight weather conditions chosen randomly but with a higher probability of getting cloudy weather: "*The reason to amplify episodes with good weather, was that important features as lane markings were not visible during weather like heavy rain.*"

Therefore, to keep our work in line with the specialized literature, we completely excluded night and rainy modes and chose randomly between 8 preset weather conditions. The probability was the same for all preset weather, and every 25 seconds, we changed the simulator's current weather. We kept the sun positioned at noon and sunset, and weather conditioning as clear, cloudy, wet, and wet-cloudy. Table 3.5 shows all eight possible weather combinations submitted when collecting data.

Weather Conditions
ClearNoon
CloudyNoon
WetNoon
WetCloudyNoon
ClearSunset
CloudySunset
WetSunset
WetCloudySunset

Table 3.5: *All possible weather conditions to be chosen randomly every 25 seconds when collecting data.*

3.4.3 Autopilot and routes

Carla has a built-in autopilot that we used to collect data. As shown in figure 3.1, we predefined four paths to collect data, each one with a different number of intersections. Since this town has eight intersections, the path with 10 intersections uses two intersections repeatedly. However, each time the vehicle uses this intersection, it comes from different directions and have different values of High-level command. For this reason we counted the same intersection twice.

To test our hypothesis 2, we used the path with 6 intersections for all models, varying the number of camera sensors used for each model. Also, for all models, we collect data for one hour in simulator time.

(a) *Clear Noon* **(b)** *Clear Sunset*

(c) *Cloudy Noon* **(d)** *Wet Cloudy Sunset*

Figure 3.4: *Examples of different weather conditions in CARLA Simulator.*

3.4.4 Other conditions and assumptions

In this section, we would like to highlight other main conditions we defined in the simulator environment and the reasons behind our choices. We sought to balance the realism provided by CARLA Simulator with some justifiable simplifications. These assumptions do not lead us to achieve a better autonomous vehicle, and this is by any means our goal. We aim to test our hypotheses, and all our decisions go toward this objective. With this, we tried to simplify as much as possible while keeping sight of our primary goal.

Pedestrians

Following the work of Haavaldsen *et al.* (2019), we disabled pedestrians on the simulator. Having an autonomous vehicle unprepared to deal with pedestrians is not useful. However, this work is about using a convolutional neural network for path learning. Further, we want to test if, by limiting our scope, we can improve autonomous vehicles' performance. Therefore, having pedestrians in the data collected would not help us test our hypotheses.

Another reason to disable pedestrians is that there are other technologies to detect pedestrians. Convolutional neural networks dedicated to pedestrian detection and other sensors, such as distance and crash avoidance, should be used for this work. The usage we are applying our CNN is in path learning, so removing pedestrians has no impact on our results.

Other vehicles

Both Dammen (2019) and Haavaldsen *et al.* (2019) included other vehicles when using CARLA Simulator in their works. We opted not to include other vehicles in this dissertation. Having other vehicles would not help us to test our hypotheses. Moreover, other techniques could be used for vehicle detection. Haavaldsen *et al.* (2019) uses Long Short Term Memory (LSTM) to make the agent approximation to the other car's tail smoother, and the leading car manufacturers already use several anti-collision systems. Therefore, although we know that autonomous vehicles should be prepared to deal with other cars in real life, we decided to remove other vehicles from our data.

Traffic lights

We disabled traffic lights in our tests. With the same rationale for removing pedestrians and other cars, having traffic lights would not help us to test our hypotheses, which is to test if reducing the scope of action of a CNN would improve the autonomous vehicles' performance. Also, having traffic lights is unnecessary since we have neither pedestrians nor other vehicles. Moreover, several works are dedicated exclusively to traffic light detection and recognition (Wang *et al.*, 2022). Thus, we leave this problem for other studies and focus on end-to-end path learning.

3.5 CNN Architecture

Our CNN architecture is primarily inspired by Haavaldsen *et al.* (2019), which in turn was inspired by Bojarski *et al.* (2016). It has two main components. The first is a feature extractor where inputs are the 300X180 pixel RGB images and outputs a vector size 100 through a dense layer. Its first and second layers are cropping and normalization layers. The cropping layer reduces the top part of the image by 70 pixels to exclude the sky part. The second scale the pixel values from [0.0, 255] to [-0.5, 0.5]. Following, there are six convolutional layers, with the first three layers using a kernel size of 5x5 and the last three with a kernel size of 3x3. Strides for the first four convolutional layers are size two, while the last 2 are size 1. After that, the output is then flattened to a 768 size vector, and lastly, a fully connected layer reduces the one-dimension vector to 100 nodes.

The second is a predictor, where we concatenate the output of the feature extractor with a vector size 6 containing information on current speed, speed limit, and the high-level command (HLC), meaning turn left, turn right, go straight and follow the lane. The HLC is encoded in a one-hot vector size 4, which along with the current speed and speed limit, gives us a vector size 6. Then this 106-size vector goes through 4 fully connected layers with outputs of 100, 50, 10, and 3, respectively. The final output contains the information we need to control the vehicle: steer, throttle, and brake. A ReLU activation functions follow all fully connected and convolutional layers. The architecture is plot in the figure 3.5.

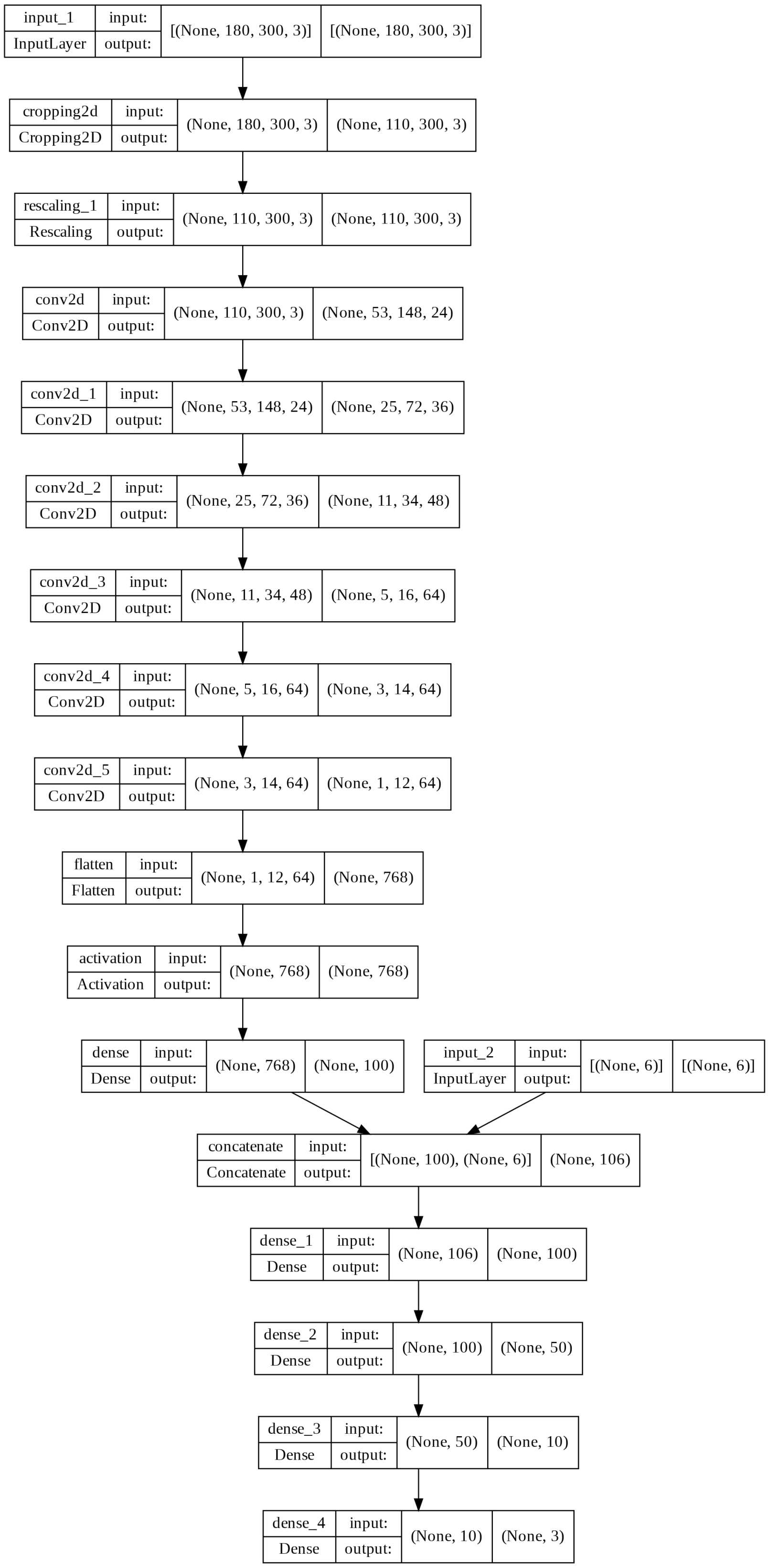


Figure 3.5: *CNN Architecture.*

3.6 Training setup

All datasets were split by a ratio of 0.8: 80% for the training set and 20% validation set. Adam optimizer was used for training along with Mean Squared Error (MSE) loss function. All four models were trained for 100 epochs. A callback was used to save checkpoints whenever the model reached a better loss value to prevent overfitting. Then, the checkpoint with the best loss value was the one we used for testing. Also, if after 30 epochs we did not see any better value for validation loss, we interrupted the training with a callback called "early stop". The mini-batch was size 32, following the recommendation of Masters and Luschi (2018).

3.7 Testing setup

The trained models were tested in the CARLA Simulator, mapping raw pixels and high-level commands directly to vehicle control. At each frame, an RGB image was collected from a camera sensor in front of the vehicle along with the vehicle's current speed, the road speed limit, and the high-level command to be applied in order to follow the predefined route. This information was then submitted to the CNN model to predict steering, throttle, and brake. We ran this experiment 100 times for each model to compare them. A random preset weather was chosen in each turn.

The information to compare is as follow: average number of lane invasions, average distance traveled, number of collisions, number of missed turn, and the number of successfully completed turns. Whenever we detected a collision, we finished that round and considered it unsuccessful. The same criteria were used if the vehicle missed a turn at an intersection. We keep running if it touches the lane since it is a minor failure.

Data gathered when testing each trained CNN model
average number of lane invasions
average distance traveled
total number of collisions
total number of missed turns
total successful rounds

Table 3.6: *Metrics to be used for analysis.*

Chapter 4

Results

Below are shown the results we have with our experiments. In order to present the results in a structured way for hypotheses 1 and 2, this chapter is divided into two sections: experiment 1 and experiment 2. Then, for each experiment, we exhibit both training and testing results. For training results, we show the validation and training loss values achieved for each model and a chart with the historical values for each epoch. Afterward, we present the results of letting the trained network control the vehicle on CARLA Simulator.

4.1 Experiment 1

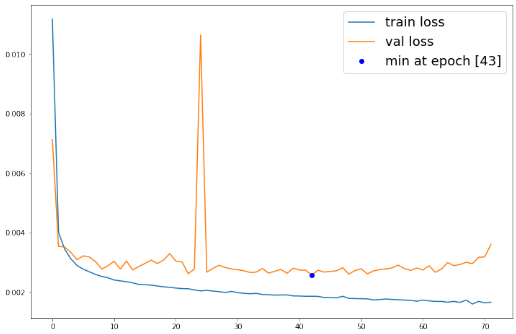
To test our hypothesis 1 - the scope amplitude is negatively correlated with the performance of the autonomous vehicle - we trained our CNN using data from four different paths with an increasing number of intersections, namely 4, 6, 8, and 10. The number of intersections is our metric for the amplitude of scope since, with more intersections, the path is more difficult for the CNN to learn when all else is kept constant. Below we present both training and testing results.

4.1.1 Training

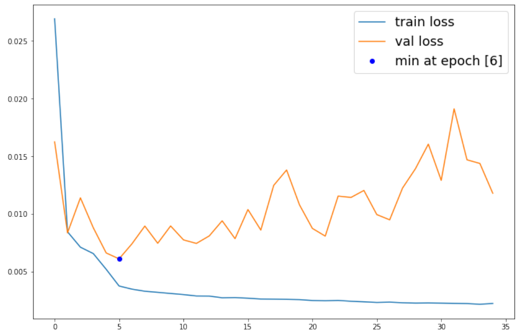
For all models, 100 epochs were enough for training and validation loss to stabilize. All models could reach low training and validation loss values, indicating that the CNN could learn effectively (table 4.1). The history of loss values can be seen in figure 4.1.

Intersections	Training Loss	Validation Loss
4	0.0016	0.0026
6	0.0022	0.0061
8	0.0021	0.0039
10	0.0026	0.0055

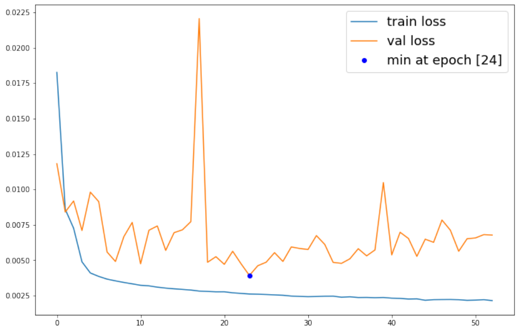
Table 4.1: *Minimum training and validations losses.*



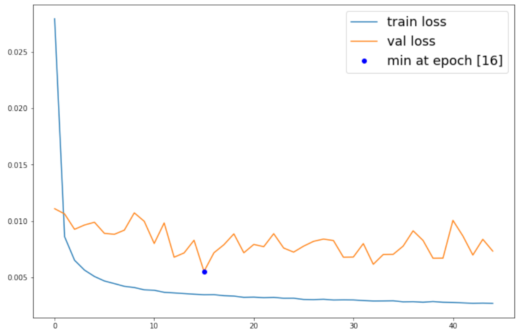
(a) *4 Intersections*



(b) *6 Intersections*



(c) *8 Intersections*



(d) *10 Intersections*

Figure 4.1: *Training and validation losses history.*

4.1.2 Testing

We ran all four models in a real-life test inside the CARLA Simulator. Each time the world ticks, we collect one 300X180 RGB picture from a camera in front of the vehicle - the exact center position of the data-gathering phase. Along with the picture, we collect the current speed, speed limit, and the high-level command for the next intersection. We input these data into the trained model to predict the steer, throttle, and brake values. Then, we apply the predicted values to the vehicle control.

Each model was submitted to test the same path of the data collection phase. Lane invasion was considered a minor failure, so it did not interrupt the round test but was accounted for later comparison. Collisions and missed turns, however, were considered severe failures. On the occurrence of any of these events, we stopped that round and saved the event's location for later analysis. A round is considered successful only if it reaches the endpoint without missing any turn at the predefined intersections and without any collision. The experiment was run 100 times for each model, and metrics for comparison are shown in table 4.2.

As stated in our hypothesis, we can see a downward trend of the successful rounds as the scope, measured by the number of intersections in a path, increases. Measured in terms of Pearson's Correlation Coefficient, we reached a value of -0.90, statistically significant at $\alpha = 0.05$ (see figure 4.2). With that, we can reject the null hypothesis of zero correlation. Thus, the overall result goes toward validating our hypothesis 1.

It is important to remember that although this statistical value serves us to have an

Intersections	**4**	**6**	**8**	**10**
Successful Rounds	98	85	79	21
Collisions	1	5	20	17
Missed Turns	1	10	1	62
Avg. Lane Invasions (σ)	0.03(0.30)	0.6(1.61)	1.75(2.13)	0.65(1.68)
Avg. Distance Traveled (σ)	343(10)	750(143)	811(200)	856(246)
Total Distance	345	812	910	1230
Distance Traveled / Total Distance	99.4%	92.4%	89.1%	69.8%
Lane Invasions per 100 meters	0.01	0.08	0.22	0.09

Table 4.2: *Test metrics for all models.*

accurate measure of our work, this work is not, by any means, a quantitative work. Our effort is to have a conceptual work that criticizes the current situation of autonomous vehicle development and proposes a more productive way to use the technology at our disposal. In other words, our goal is to empirically investigate what we conceptually concluded based on the sociotechnical imaginary of autonomous vehicles.

The average lane invasions grow as we increase the number of intersections, except for the last model, with ten intersections. This model could learn better to keep the agent between the lanes than its predecessor with eight intersections. One possible explanation is that the diversity of images benefited this CNN in preventing overfitting. Somehow the model with eight intersections could not drive as stably as the other models. However, for all models, an average of 1 or 2 lane invasions per each round is impressively low, considering the total distance of the path was at least 812 meters for models with 6, 8, and 10 intersections. In conclusion, all models were able to learn lane following reasonably well. The number of lane invasions on every 100 meters follows the same trend as the average lane invasion.

Finally, when we look at the percentage of distance traveled, we can see a downward trend as the number of intersections increases. The simplest path, with four intersections, achieved 99.4% of the distance traveled and 98 successful rounds out of 100, a remarkable achievement considering that the data collection phase lasted only 1 hour. In other words, we could train a CNN to learn a predefined path with a precision of about 99% using only 1 hour of data collection as input. Of course, the path is quite simple and, by any means, similar to what we expect in real life, but it shows that the end-to-end approach can have significant results as we narrow the scope of our experiments.

Crash and missed turn locations

Figure 4.3 shows maps of Town 02 with collision and missed turn locations for each one of our models. Collisions are marked with red circles, whereas missed turns are marked with blue circles. We numbered each intersection to reference the precise location in our analysis.

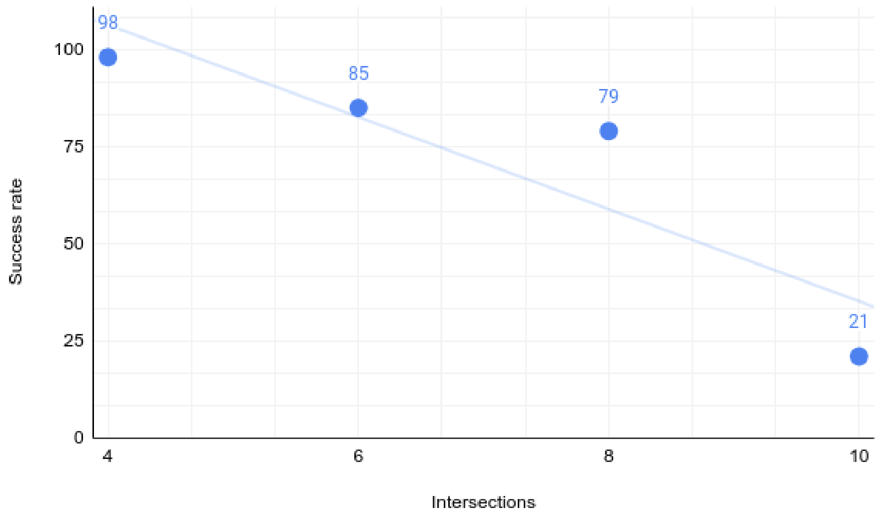


Figure 4.2: *Successful Rounds vs. Number of Intersections.* $r(2) = -0.90$, *significant at* $\alpha = 0.05$.

(a) *4 Intersections* **(b)** *6 Intersections*

(c) *8 Intersections* **(d)** *10 Intersections*

Figure 4.3: *Crash and missed turn locations*

4.2 Experiment 2

Our second hypothesis states that the degree to which the agent interacts with the environment is positively correlated with the performance of the autonomous vehicle. The metric we use for “degree of interaction with the environment” is the number of cameras we use in each vehicle. With that, we trained four models, all going through the same path but with different quantities of cameras collecting data: 3, 7, 11, and 15. As stated in our hypothesis 2, we expect the agent’s performance to improve as we increase the number of cameras in the agent.

4.2.1 Training

One hundred epochs were enough for all models to stabilize. We can see in image 4.4 that the model converges more smoothly as we increase the number of cameras collecting data. It happens because there is more data to train. Also, we could reach a validation loss of 0.0032 in the model with 15 cameras, quite lower than the value of 0.0104 of the model with three cameras. Table 4.3 shows training and validation losses for all models, and figure 4.4 shows the history of training data.

Cameras	Training Loss	Validation Loss
3	0.0018	0.0104
7	0.0022	0.0061
11	0.0044	0.0088
15	0.0029	0.0032

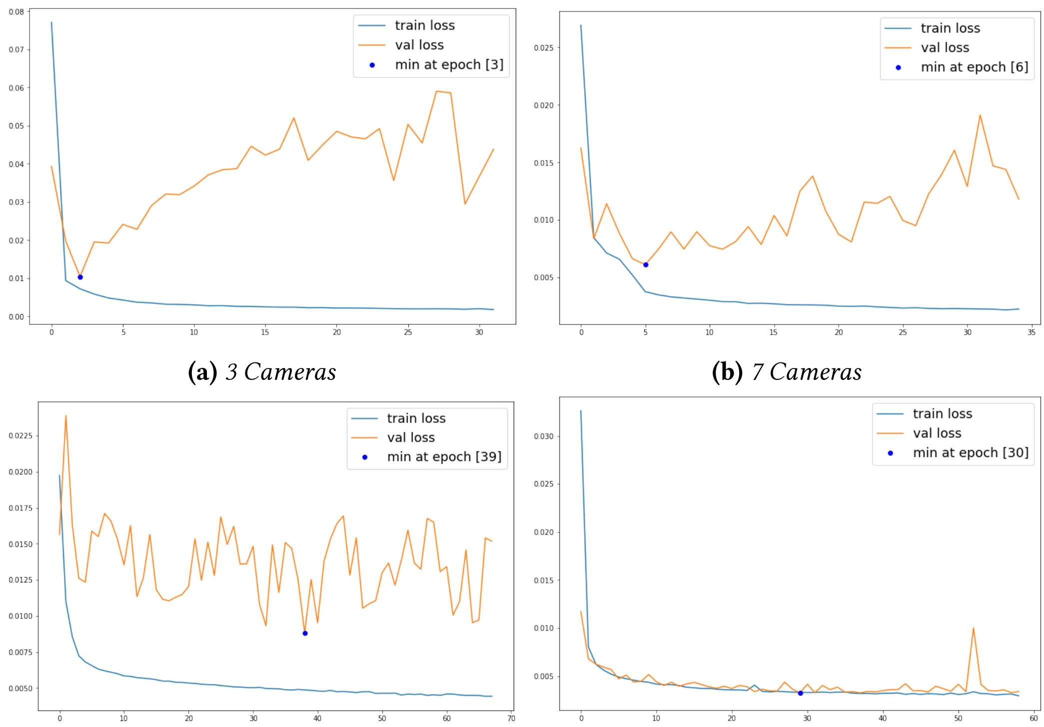
Table 4.3: *Minimum training and validations losses.*

4.2.2 Testing

The scatter plot of figure 4.5 shows an upward trend in successful rounds as we increase the number of cameras. Precisely, Pearson’s Correlation Coefficient is 0.87, statistically significant at $\alpha = 0.10$. As we expected, it confirms that as we increase the degree of interaction with the environment, in this case, the number of camera sensors we coupled in the car, the autonomous vehicle’s performance increases significantly.

Although finding a trend in the average lane invasions is hard, the average distance traveled increases as we increase the number of cameras. It goes in line with our hypothesis since the distance traveled is an indicator of the autonomous vehicle’s performance. When looking at lane invasions per 100 meters, models with 3, 11, and 15 cameras follow a downward trend, except with the case with the model with seven cameras. The black-box nature of convolutional neural networks makes this hard to explain, but it is expected to have this type of disparity. If we were to rerun the experiment with the same data, we could achieve very different results. This characteristic of neural networks is still considered an open problem in the literature.

Our second experiment shows that all models could learn lane following reasonably well. Moreover, it shows that increasing the number of cameras in the vehicle can significantly improve the agent’s performance, which validates our second hypothesis. Lastly,



(a) *3 Cameras* **(b)** *7 Cameras*

(c) *11 Cameras* **(d)** *15 Cameras*

Figure 4.4: *Training and validation losses history.*

it is worth mentioning that we could achieve 99% of successful rounds when using 15 cameras, an impressive result considering the complexity of the path - an urban area with six different intersections and more than 800 meters - and the amount of data collection of only 60 minutes.

Cameras	**3**	**7**	**11**	**15**
Successful Rounds	28	85	89	99
Collisions	41	5	10	1
Missed Turns	31	10	1	0
Avg. Lane Invasions (σ)	5.12(4.37)	0.6(1.61)	3.98(2.90)	0.82(1.95)
Avg. Distance Traveled (σ)	597(214)	750(143)	781(122)	811(6)
Total Distance	812	812	812	812
Distance Traveled / Total Distance	73.5%	92.4%	96.2%	99.9%
Lane Invasions per 100 meters	0.86	0.08	0.51	0.10

Table 4.4: *Test metrics for all models.*

Crash and missed turn locations

Figure 4.6 shows the locations where collisions and missed turns happened when testing in the simulator. We can see that only models with 11 and 15 cameras could learn to pass through intersection three safely. For models with 3 and 7 cameras, there were 31 and 10 missed turns in this intersection, respectively. It is also worth noticing that intersection six was hard to get through without colliding with a pole on the sidewalk, as

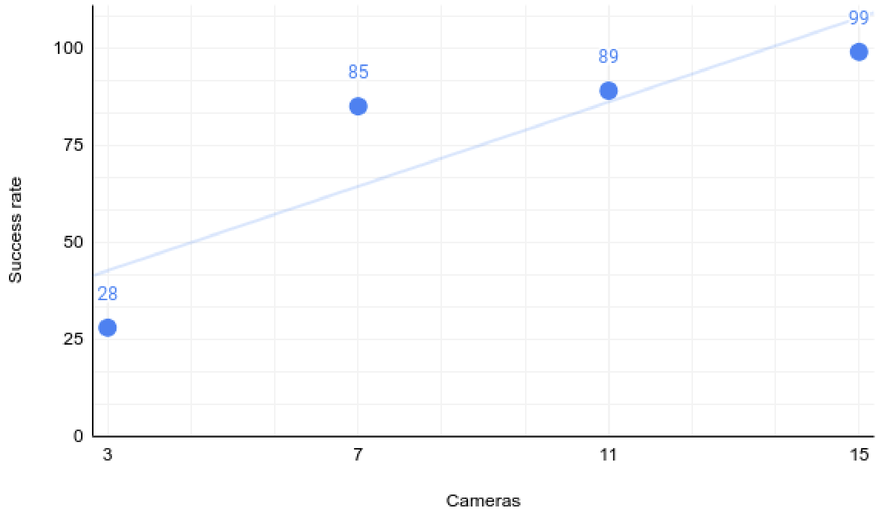
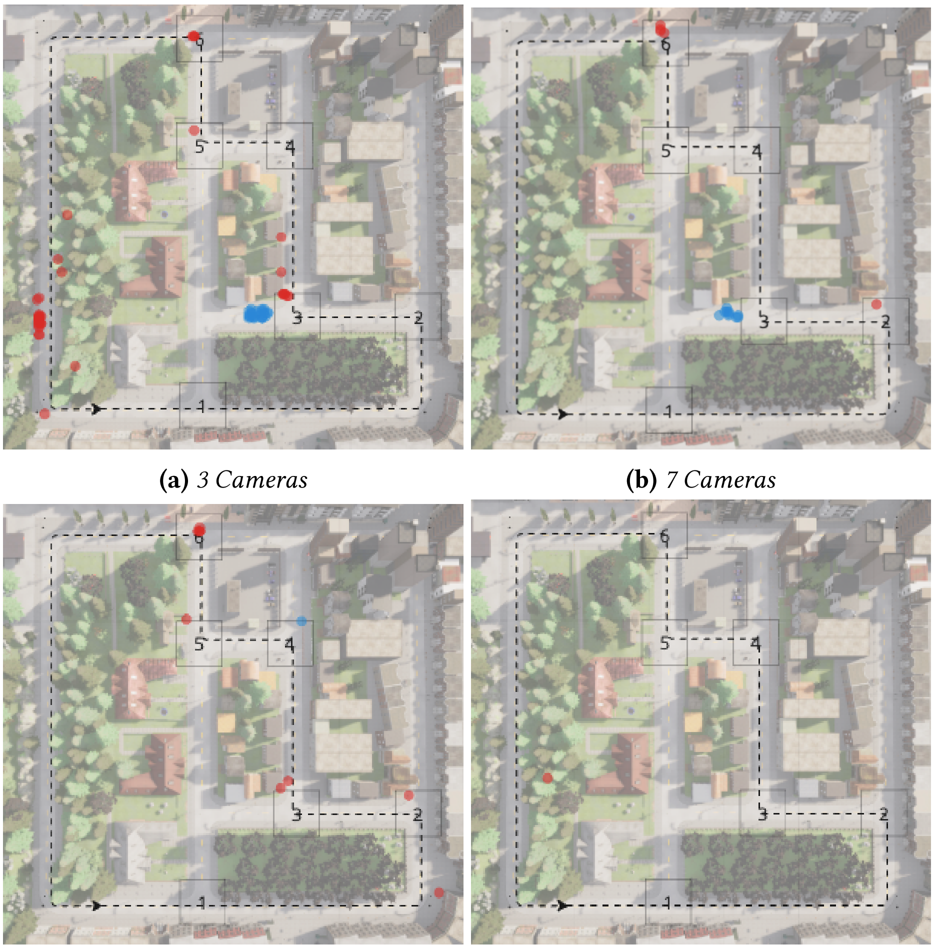


Figure 4.5: *Successful Rounds vs. Number of Cameras.* $r(2) = 0.87$*, significant at* $\alpha = 0.10$*.*

seen in figure 5.1. Lastly, the model with three cameras had difficulty passing by a guard rail close to the final point. One of the reasons is that at this point, the maximum speed of the road is 90 km/h, more than any other stretch on the path. At this speed, the vehicle had some instability because of the time between the prediction of the vehicle control and applying the control to the vehicle. However, for models with 7, 11, and 15 cameras, this was not a significant problem.



(a) *3 Cameras*

(b) *7 Cameras*

(c) *11 Cameras*

(d) *15 Cameras*

Figure 4.6: *Crash and missed turn locations*

Chapter 5

Discussion

The result of experiment 1 validates our hypothesis 1: "The scope amplitude is negatively correlated with the performance of the autonomous vehicle". As we narrowed the scope by reducing the number of intersections in a path, our intelligent agent performed better in our experiment. In the most straightforward route, with only four intersections, we could achieve a 98% successful rate, an impressive result considering only one hour of data collection and the simplicity of our end-to-end system.

Based on our findings, it would be more effective for producers of autonomous technology to focus on solving more minor aspects of the problem rather than aiming for a fully general autonomous vehicle. By achieving acceptable milestones in controlled environments, we can gradually expand the scope to larger areas until we are confident enough to introduce autonomous vehicles in more challenging settings, like densely populated cities. Unfortunately, autonomous vehicles are treated differently, with private companies and academic researchers spending vast capital to pursue the utopia dictated by its sociotechnical imaginary rather than real progress.

For our second experiment, the results show a strong positive correlation between the degree of interaction with the environment and the autonomous vehicle's performance. The trained network achieved 28 successful rounds out of 100 using three cameras. When we increased the number of cameras to 15, the number of successful rounds went to 99. It validates our hypothesis 2: "the degree to which the agent interacts with the environment is positively correlated with the performance of the autonomous vehicle".

This result suggests that one way to increase performance is to capture more data as input for the neural network. It is essential to mention that this data must be correctly labeled. In our experiment, we added different offset values for each camera. It helped the neural network to converge better to a minimal loss value. Therefore, collecting more data and correctly labeling it is key to improving performance for end-to-end deep learning.

5.1 Common pitfalls

There were places where crashes were more frequent, such as poles and hydrants. The trained network confused these places and could not distinguish them from the road. Figure 5.1 shows images of common pitfalls.

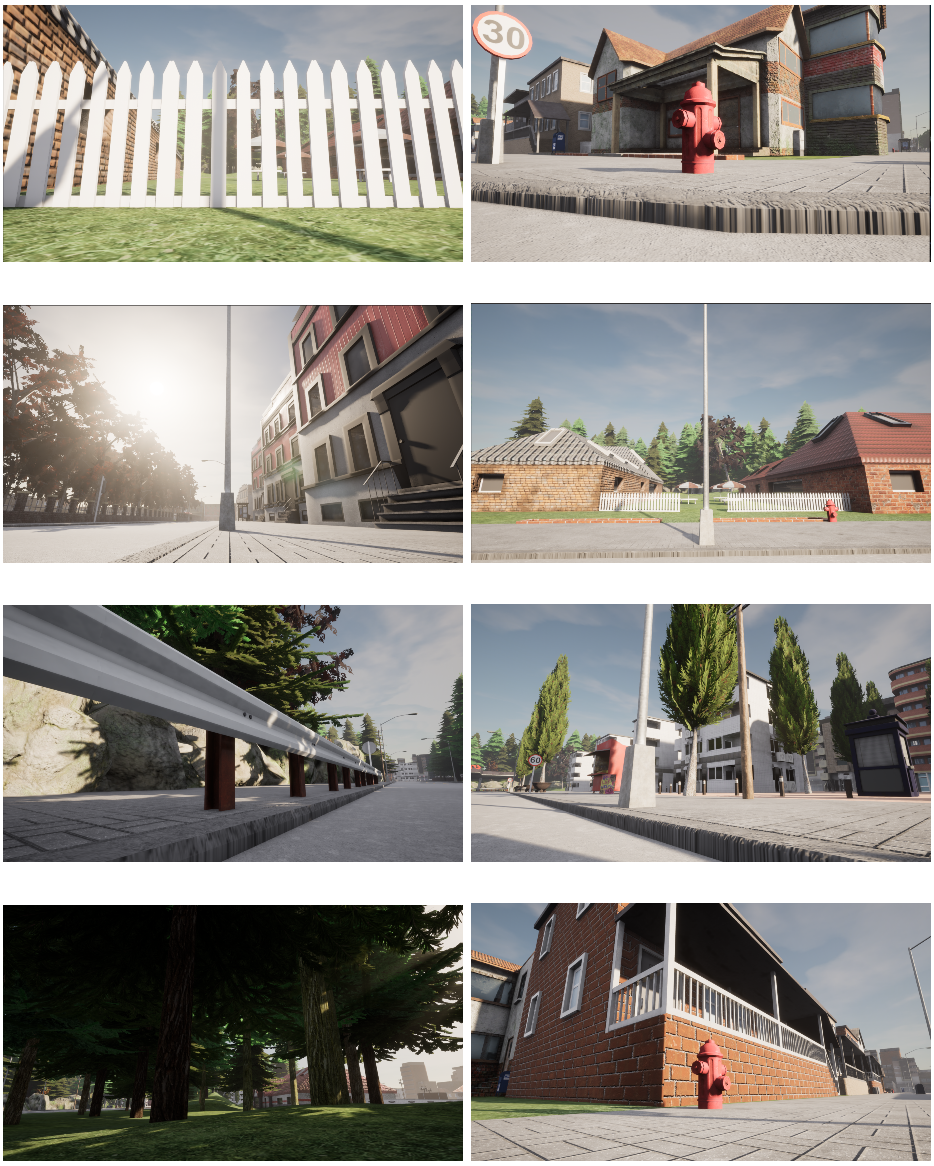


Figure 5.1: *Common pitfalls*

These hot spots suggest that the neural network can have trouble interpreting some places. Identifying points where errors are persistent may indicate that specialized signaling in roads to guide autonomous vehicles can be an additional important topic to study.

5.2 Future work

A simulated environment's advantage is its ease of application and reduced costs. A suggestion for future work is to apply these techniques in the physical world. Significant differences, such as infinite lighting possibilities, might help the convolutional neural network for better generalization. Also, our experiment did not consider other vehicles, pedestrians, and traffic lights. Future work should consider that via end-to-end learning or other sensors, thus using a mix of technologies to create a safe autonomous vehicle.

Advanced technologies such as Lidar (Light Detection and Ranging) can help improve autonomous vehicle advancements. It was outside the scope of this dissertation to explore this technology, but it is important to consider it in future work. Regarding neural networks, using Long short-term memory (LSTM) has presented remarkable results (Haavaldsen *et al.*, 2019). Future work could focus on improving the CNN architecture and applying more advanced techniques.

An additional benefit of narrow autonomy is it could reduce the discrepancy in driving skills by helping in common specific tasks. One example is autonomous emergency braking, which is already present in some commercial vehicles. This assistance is one instance of a narrow approach that can increase safety on public roads by minimizing human errors. As such, many other satellite tasks could be automated. The potential of increasing safety by reducing variance in driving skills is an interesting topic that should be further investigated by future research.

5.3 Open problems of deep learning

Neural network results are hard to explain or reproduce. Although we can see a pattern for successful rounds, this is different for average lane invasions. Our trained model performed better with eleven cameras compared to the model with seven cameras, 89 against 85 successful rounds, which was expected. However, regarding lane invasions, the model with seven cameras had much fewer invasions per 100 meters, 0.08 against 0.51.

Training the network with the same dataset does not guarantee that the network will converge, and results vary greatly in each case. Problems such as overfitting are common and can significantly impact the results. These are still open problems in the literature of deep learning.

In recent work, Berner *et al.* (2022) discuss the several open problems of deep learning and propose a new field of "mathematical analysis of deep learning" to deal with these open problems. With that, the black box characteristic of neural networks will remain open for the following years as we still do not have the answers to questions such as, in the words of Berner *et al.* (2022): "the outstanding generalization power of overparametrized neural networks, the role of depth in deep architectures, the apparent absence of the curse of dimensionality, the surprisingly successful optimization performance despite the non-convexity of the problem, understanding what features are learned, why deep architectures perform exceptionally well in physical problems, and which fine aspects of an architecture affect the behavior of a learning task in which way". Therefore, despite the impressive results we achieve when using deep learning, it is crucial to consider these open problems when applying these techniques in the real world.

5.4 Autonomous vehicles revisited

Autonomous vehicle development has failed to meet expectations in the last decade. Instigated by this frustration, we studied autonomous vehicles in light of the sociotechnical imaginary theoretical framework. We concluded that the autonomous vehicle field resembles general artificial intelligence. While artificial intelligence development went from narrow to general approaches, autonomous vehicles pursue general autonomy without passing through essential milestones in narrow autonomy.

This work is a manifesto against the current situation of autonomous vehicle development. We proposed a narrow approach, and to prove our point, we laid out a set of experiments to test the performance of our autonomous vehicle with different scope amplitudes. The challenge here is not to reduce the scope but to break free from the utopia of the autonomous vehicle sociotechnical imaginary. The results of our experiments indicate that we can achieve better performance when we narrow down the scope of our intelligent agent, maintaining everything else the same.

Our findings challenge the current situation of autonomous vehicles by acknowledging the rhetorical utopia promoted by, in the words of Urry (2016, p. 9) "powerful actors seeking to realize the future they envisaged". Instead of assuming general autonomy as an inevitable future, we advocate for actionable strategies to achieve results that benefit society by proposing a narrow approach. In that sense, autonomous vehicles' development should be revisited.

Chapter 6

Conclusion

We started our studies with the concept of sociotechnical imaginaries, which can be described as visions of desirable futures. They are fabricated by a complex network of institutional power shared by scientific and non-scientific actors. These visions are at once products and instruments of the co-production of science and technology. In the case of autonomous vehicles, these fabricated visions are extremely over-optimistic; the "inevitable future" refused to meet the expectations over the last two decades.

Inspired by the distinction between Narrow and General Artificial Intelligence, we propose the same distinction in the field of autonomous technology. Therefore, we conceptualize narrow autonomy as the study of context-limited autonomous vehicles. In contrast, general autonomy is the area that pursues the goal of achieving an autonomous vehicle as envisaged by its sociotechnical imaginary.

Grounded on this theoretical background, we propose that narrow autonomy is a more productive way to develop autonomous vehicles. In fact, important milestones were achieved in the artificial intelligence field before the proposition of such a general artificial intelligence. Autonomous vehicles would benefit from such a mindset. To pursue this goal, however, we must first break free from its institutionalized sociotechnical imaginary.

Therefore, we stated two testable hypotheses. The first says that "the scope amplitude is negatively correlated with the performance of the autonomous vehicle". In other words, as we narrow the scope for our intelligent agent, its performance tends to be better. The second hypothesis states that "the degree to which the agent interacts with the environment is positively correlated with the performance of the autonomous vehicle". This means that as we have more sensory capability to read the environment, we improve our agent's performance.

Using the CARLA Simulator, we set up two experiments to test our hypotheses based on the latest advancement in end-to-end deep learning. In the first experiment, we trained a convolutional neural network (CNN) with images of four pre-defined paths of different levels of complexity, measured by the number of intersections in each path. For the second experiment, differentiate the autonomous vehicles by the number of cameras it uses to collect data. Then, after training the respective neural networks, we submitted the trained networks to 100 rounds of testing in the simulator and compared the results.

Our results validate our hypotheses, which suggest that by narrowing down the scope of our autonomous vehicle, we increase its performance, everything else constant. Also, another way to improve the agent's performance is to increase the degree of interaction with the environment, in our case, measured by the number of cameras used in the data collection phase.

These findings suggest that there could be a more productive way to approach autonomous vehicle development. The first step is to break free from the fabricated fantasy that autonomous vehicle is "only a matter of time". Autonomous vehicles' sociotechnical imaginary promote visions that intelligent agents will not only act along humans on traffic roads but also surpass human-level performance, virtually eliminating traffic accidents caused by human errors. However, the frustration of the last decade indicates that without actionable strategies, this will hardly ever cease being a futuristic delusion.

References

[Bachute and Subhedar 2021] Mrinal R. Bachute and Javed M. Subhedar. "Autonomous driving architectures: insights of machine learning and deep learning algorithms". In: *Machine Learning with Applications* 6 (2021), p. 100164. issn: 2666-8270. doi: https://doi.org/10.1016/j.mlwa.2021.100164. url: https://www.sciencedirect.com/science/article/pii/S2666827021000827 (cit. on p. 13).

[Berner *et al.* 2022] Julius Berner, Philipp Grohs, Gitta Kutyniok, and Philipp Petersen. "The modern mathematics of deep learning". In: *Mathematical Aspects of Deep Learning*. Cambridge University Press, Dec. 2022, pp. 1–111. doi: 10.1017/9781009025096.002. url: https://doi.org/10.1017%5C%2F9781009025096.002 (cit. on p. 43).

[Bojarski *et al.* 2016] Mariusz Bojarski *et al.* *End to End Learning for Self-Driving Cars*. 2016. doi: 10.48550/ARXIV.1604.07316. url: https://arxiv.org/abs/1604.07316 (cit. on pp. 1, 11, 12, 14–16, 23, 25, 29).

[Braun and Randell 2020] Robert Braun and Richard Randell. "Futuramas of the present: the "driver problem" in the autonomous vehicle sociotechnical imaginary". In: *Humanities and Social Sciences Communications* 7 (Dec. 2020). url: https://www.nature.com/articles/s41599-020-00655-z (visited on 08/11/2021) (cit. on pp. 8, 9).

[Chen *et al.* 2015] Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao. "Deepdriving: learning affordance for direct perception in autonomous driving". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 2722–2730. doi: 10.1109/ICCV.2015.312 (cit. on p. 11).

[Cheng *et al.* 2022] Jun Cheng, Liyan Zhang, Qihong Chen, Xinrong Hu, and Jingcao Cai. "A review of visual slam methods for autonomous driving vehicles". In: *Engineering Applications of Artificial Intelligence* 114 (2022), p. 104992. issn: 0952-1976. doi: https://doi.org/10.1016/j.engappai.2022.104992. url: https://www.sciencedirect.com/science/article/pii/S0952197622001853 (cit. on p. 13).

[Codevilla *et al.* 2017] Felipe Codevilla, Matthias Müller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. *End-to-end Driving via Conditional Imitation Learning*. 2017. doi: 10.48550/ARXIV.1710.02410. url: https://arxiv.org/abs/1710.02410 (cit. on pp. 15, 16, 22, 23).

[Curts 2015] Kati Curts. “Temples and turnpikes in "the world of tomorrow": religious assemblage and automobility at the 1939 new york world’s fair”. In: *Journal of the American Academy of Religion* 83.3 (2015), pp. 722–749. issn: 00027189, 14774585. url: http://www.jstor.org/stable/24488182 (visited on 09/15/2022) (cit. on p. 8).

[Dammen 2019] Jonas Dammen. “End-to-end deep learning for Autonomous Driving”. PhD thesis. Trondheim, Norway: Department of Computer Science, Norwegian University of Science and Technology, July 2019 (cit. on pp. 13–16, 23, 27, 29).

[Davies 2016] Rob Davies. *Lyft to trial driverless cars on US roads*. May 2016. url: https://www.theguardian.com/business/2016/may/06/lyft-driverless-cars-uber-taxis-us-roads-chevrolet-bolt (visited on 09/21/2022) (cit. on p. 1).

[Dosovitskiy *et al.* 2017] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. “CARLA: An open urban driving simulator”. In: *Proceedings of the 1st Annual Conference on Robot Learning*. 2017, pp. 1–16 (cit. on pp. 16, 23).

[Faisal *et al.* 2019] Asif Faisal, Md Kamruzzaman, Tan Yigitcanlar, and Graham Currie. “Understanding autonomous vehicles: a systematic literature review on capability, impact, planning and policy”. In: *Journal of Transport and Land Use* 12.1 (2019), pp. 45–72. issn: 19387849. url: https://www.jstor.org/stable/26911258 (visited on 09/15/2022) (cit. on pp. 1, 8, 10).

[Fjelland 2020] Ragnar Fjelland. “Why general artificial intelligence will not be realized”. In: *Humanities and Social Sciences Communications* 7 (June 2020). url: https://doi.org/10.1057/s41599-020-0494-4 (visited on 08/04/2021) (cit. on p. 9).

[Goertzel 2014] Ben Goertzel. “Artificial general intelligence: concept, state of the art, and future prospects”. In: *Journal of Artificial General Intelligence* 5.1 (2014), pp. 1–48. doi: doi:10.2478/jagi-2014-0001. url: https://doi.org/10.2478/jagi-2014-0001 (cit. on p. 9).

[Gundler 2013] Bettina Gundler. “Promoting german automobile technology and the automobile industry: the motor hall at the deutsches museum, 1933–1945”. In: *The Journal of Transport History* 34.2 (2013), pp. 117–139. doi: 10.7227/TJTH.34.2.3. eprint: https://doi.org/10.7227/TJTH.34.2.3. url: https://doi.org/10.7227/TJTH.34.2.3 (cit. on p. 8).

[Haavaldsen *et al.* 2019] Hege Haavaldsen, Max Aasbø, and Frank Lindseth. “Autonomous vehicle control: end-to-end learning in simulated urban environments”. In: *Nordic Artificial Intelligence Research and Development*. Ed. by Kerstin Bach and Massimiliano Ruocco. Cham: Springer International Publishing, 2019, pp. 40–51. isbn: 978-3-030-35664-4 (cit. on pp. 3, 16, 19, 20, 22–25, 27–29, 43).

[HAKAK *et al.* 2023] Saqib HAKAK *et al.* "Autonomous vehicles in 5g and beyond: a survey". In: *Vehicular Communications* 39 (2023), p. 100551. ISSN: 2214-2096. DOI: https://doi.org/10.1016/j.vehcom.2022.100551. URL: https://www.sciencedirect.com/science/article/pii/S2214209622000985 (cit. on p. 13).

[HU *et al.* 2022] Jie HU, Huifang KONG, Qian ZHANG, and Runwu LIU. "Enhancing scene understanding based on deep learning for end-to-end autonomous driving". In: *Engineering Applications of Artificial Intelligence* 116 (2022), p. 105474. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2022.105474. URL: https://www.sciencedirect.com/science/article/pii/S095219762200464X (cit. on p. 13).

[JANAI *et al.* 2020] Joel JANAI, Fatma GÜNEY, Aseem BEHL, and Andreas GEIGER. "Computer vision for autonomous vehicles: problems, datasets and state of the art". In: 12.1-3 (2020). DOI: 10.1561/0600000079 (cit. on pp. 10–12).

[JASANOFF 2015] Sheila JASANOFF. "Dreamscapes of modernity: sociotechnical imaginaries and the fabrication of power". In: 1st ed. Chicago, IL: University of Chicago Press, 2015. Chap. Future Imperfect: Science, technology and the imagination of modernity (cit. on pp. 7, 8).

[JASANOFF and KIM 2009] Sheila JASANOFF and Sang-Hyun KIM. "Containing the atom: sociotechnical imaginaries and nuclear power in the united states and south korea". In: (2009). DOI: 10.1007/s11024-009-9124-4 (cit. on pp. 1, 5, 6, 19).

[KURZWEIL 2005] Ray KURZWEIL. *The Singularity is near: When Humans Transcend Biology*. New York: Viking, 2005. ISBN: 978-0-670-03384-3 (cit. on p. 9).

[LAWRENCE 2020] Christopher LAWRENCE. "Heralds of global transparency: remote sensing, nuclear fuel-cycle facilities, and the modularity of imagination". In: *Social Studies of Science* 50.4 (2020). PMID: 31603373, pp. 508–541. DOI: 10.1177/0306312719879769. eprint: https://doi.org/10.1177/0306312719879769. URL: https://doi.org/10.1177/0306312719879769 (cit. on p. 5).

[LECUN *et al.* 2005] Yann LECUN, Urs MULLER, Jan BEN, Eric COSATTO, and Beat FLEPP. "Off-road obstacle avoidance through end-to-end learning". English (US). In: *Advances in Neural Information Processing Systems 18 - Proceedings of the 2005 Conference*. Advances in Neural Information Processing Systems. 2005 Annual Conference on Neural Information Processing Systems, NIPS 2005 ; Conference date: 05-12-2005 Through 08-12-2005. 2005, pp. 739–746. ISBN: 9780262232531 (cit. on pp. 14, 15).

[LEE and HA 2020] Myoung-jae LEE and Young-guk HA. "Autonomous driving control using end-to-end deep learning". In: *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*. Feb. 2020, pp. 470–473. DOI: 10.1109/BigComp48618.2020.00-23 (cit. on p. 13).

[Levidow and Raman 2020] Les Levidow and Sujatha Raman. "Sociotechnical imaginaries of low-carbon waste-energy futures: uk techno-market fixes displacing public accountability". In: *Social Studies of Science* 50.4 (2020). PMID: 32077372, pp. 609–641. doi: 10.1177/0306312720905084. eprint: https://doi.org/10.1177/0306312720905084. url: https://doi.org/10.1177/0306312720905084 (cit. on p. 5).

[Litman 2015] Todd Litman. "Autonomous vehicle implementation predictions: implications for transport planning". In: 2015 (cit. on p. 8).

[Lunawat 2022] Dev Lunawat. *How do Driverless Trains Work?* July 2022. url: https://www.scienceabc.com/innovation/driverless-trains-work.html (visited on 09/16/2022) (cit. on p. 10).

[M *et al.* 2019] Alonso Raposo M *et al.* *The future of road transport.* Scientific analysis or review KJ-1A-29748-EN-N (online),KJ-1A-29748-EN-C (print). Luxembourg (Luxembourg), 2019. doi: 10.2760/668964(online),10.2760/524662(print) (cit. on p. 8).

[Malik *et al.* 2022] Sumbal Malik, Manzoor Khan, and Hesham El-Sayed. "Carla: car learning to act — an inside out". In: *Procedia Computer Science* 198 (Jan. 2022), pp. 742–749. doi: 10.1016/j.procs.2021.12.316 (cit. on p. 23).

[Martínez-Díaz and Soriguera 2018] Margarita Martínez-Díaz and Francesc Soriguera. "Autonomous vehicles: theoretical and practical challenges". In: *Transportation Research Procedia* 33 (2018). XIII Conference on Transport Engineering, CIT2018, pp. 275–282. issn: 2352-1465. doi: https://doi.org/10.1016/j.trpro.2018.10.103. url: https://www.sciencedirect.com/science/article/pii/S2352146518302606 (cit. on p. 8).

[Masters and Luschi 2018] Dominic Masters and Carlo Luschi. "Revisiting small batch training for deep neural networks". In: *arXiv preprint arXiv:1804.07612* (2018) (cit. on p. 31).

[Mckinsey & Company 2013] Mckinsey & Company. *The road to 2020 and beyond: what's driving the global automotive industry.* 2013. url: https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/automotive%5C%20and%5C%20assembly/pdfs/mck_the_road_to_2020_and_beyond.ashx (visited on 05/11/2022) (cit. on p. 8).

[Mladenović *et al.* 2020] Miloš N. Mladenović, Dominic Stead, Dimitris Milakis, Kate Pangbourne, and Moshe Givoni. "Chapter ten - governance cultures and sociotechnical imaginaries of self-driving vehicle technology: comparative analysis of finland, uk and germany". In: *Policy Implications of Autonomous Vehicles.* Ed. by Dimitris Milakis, Nikolas Thomopoulos, and Bert van Wee. Vol. 5. Advances in Transport Policy and Planning. Academic Press, 2020, pp. 235–262. doi: https://doi.org/10.1016/bs.atpp.2020.01.001. url: https://www.sciencedirect.com/science/article/pii/S2543000920300019 (cit. on p. 5).

[Niranjan *et al.* 2021] D.R. Niranjan, B C VinayKarthik, and Mohana. "Deep learning based object detection model for autonomous driving research using carla simulator". In: *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. Oct. 2021, pp. 1251–1258. doi: 10.1109/ICOSEC51865.2021.9591747 (cit. on pp. 22, 23).

[Orfanou *et al.* 2022] Foteini P. Orfanou, Eleni I. Vlahogianni, George Yannis, and Evangelos Mitsakis. "Humanizing autonomous vehicle driving: understanding, modeling and impact assessment". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 87 (2022), pp. 477–504. issn: 1369-8478. doi: https://doi.org/10.1016/j.trf.2022.04.008. url: https://www.sciencedirect.com/science/article/pii/S1369847822000766 (cit. on p. 13).

[P *et al.* 2022] Ramkumar M. P *et al.* "Autonomous navigation system based on a dynamic access control architecture for the internet of vehicles". In: *Computers and Electrical Engineering* 101 (2022), p. 108037. issn: 0045-7906. doi: https://doi.org/10.1016/j.compeleceng.2022.108037. url: https://www.sciencedirect.com/science/article/pii/S0045790622003007 (cit. on p. 13).

[D. Pomerleau and Jochem 1996] D. Pomerleau and T. Jochem. "Rapidly adapting machine vision for automated vehicle steering". In: *IEEE Expert* 11.2 (1996), pp. 19–27. doi: 10.1109/64.491277 (cit. on p. 10).

[D. A. Pomerleau 1988] Dean A. Pomerleau. "Alvinn: an autonomous land vehicle in a neural network". In: *Advances in Neural Information Processing Systems*. Ed. by D. Touretzky. Vol. 1. Morgan-Kaufmann, 1988. url: https://proceedings.neurips.cc/paper/1988/file/812b4ba287f5ee0bc9d43bbf5bbe87fb-Paper.pdf (cit. on pp. 11, 13–15).

[Ruslan *et al.* 2023] Noor Amira Ilyanie Ruslan *et al.* "Modelling and control strategies in path tracking control for autonomous tracked vehicles: a review of state of the art and challenges". In: *Journal of Terramechanics* 105 (2023), pp. 67–79. issn: 0022-4898. doi: https://doi.org/10.1016/j.jterra.2022.10.003. url: https://www.sciencedirect.com/science/article/pii/S0022489822000714 (cit. on p. 13).

[Russell 2019] Keith Russell. "End-to-end learning: using neural networks for vehicle control and obstacle avoidance". In: *University Honors Program Theses*. Vol. 417. 2019. url: https://digitalcommons.georgiasouthern.edu/honors-theses/417 (cit. on pp. 12, 13).

[Schiølin 2020] Kasper Schiølin. "Revolutionary dreams: future essentialism and the sociotechnical imaginary of the fourth industrial revolution in denmark". In: *Social Studies of Science* 50.4 (2020). doi: 10.1177/0306312719867768 (cit. on p. 5).

[Shladover 2018] Steven E. Shladover. "Connected and automated vehicle systems: introduction and overview". In: *Journal of Intelligent Transportation Systems* 22.3 (2018), pp. 190–200. doi: 10.1080/15472450.2017.1336053. eprint: https://doi.org/10.1080/15472450.2017.1336053. url: https://doi.org/10.1080/15472450.2017.1336053 (cit. on p. 9).

[Singh 2018] Santokh Singh. "Critical reasons for crashes investigated in the national motor vehicle crash causation survey". In: 2018 (cit. on p. 8).

[Sismondo 2020] Sergio Sismondo. "Sociotechnical imaginaries: an accidental themed issue". In: *Social Studies of Science* 50.4 (2020). PMID: 32721275, pp. 505–507. doi: 10.1177/0306312720944753. eprint: https://doi.org/10.1177/0306312720944753. url: https://doi.org/10.1177/0306312720944753 (cit. on p. 5).

[Smallman 2020] Melanie Smallman. ""nothing to do with the science": how an elite sociotechnical imaginary cements policy resistance to public perspectives on science and technology through the machinery of government". In: *Social Studies of Science* 50.4 (2020). PMID: 31603380, pp. 589–608. doi: 10.1177/0306312719879768. eprint: https://doi.org/10.1177/0306312719879768. url: https://doi.org/10.1177/0306312719879768 (cit. on p. 5).

[Sparrow and Howard 2017] Robert Sparrow and Mark Howard. "When human beings are like drunk robots: driverless vehicles, ethics, and the future of transport". In: *Transportation Research Part C: Emerging Technologies* 80 (2017), pp. 206–215. issn: 0968-090X. doi: https://doi.org/10.1016/j.trc.2017.04.014. url: https://www.sciencedirect.com/science/article/pii/S0968090X17301262 (cit. on p. 8).

[Stewart 2016] Heather Stewart. *George Osborne to back driverless car trials on UK motorways.* Mar. 2016. url: https://www.theguardian.com/uk-news/2016/mar/12/george-osborne-driverless-car-trials-budget (visited on 09/21/2022) (cit. on p. 1).

[Strickland 2019] Eliza Strickland. *How IBM Watson overpromissed and underdelivered on AI Health Care.* Apr. 2019. url: https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care (visited on 08/02/2021) (cit. on p. 1).

[Terapaptommakol *et al.* 2022] Wasinee Terapaptommakol, Danai Phaoharuhansa, Pramote Koowattanasuchat, and Jartuwat Rajruangrabin. "Design of obstacle avoidance for autonomous vehicle using deep q-network and carla simulator". In: *World Electric Vehicle Journal* 13.12 (2022). issn: 2032-6653. doi: 10.3390/wevj13120239. url: https://www.mdpi.com/2032-6653/13/12/239 (cit. on p. 23).

[Topham 2020] Gwyn Topham. *Self-driving cars could be allowed on UK motorways next year.* Aug. 2020. url: https://www.theguardian.com/technology/2020/aug/18/self-driving-cars-allowed-motorways-industry-risk (visited on 09/21/2022) (cit. on p. 1).

[Topham 2021] Gwyn Topham. *'Peak hype': why the driverless car revolution has stalled.* Jan. 2021. url: https://www.theguardian.com/technology/2021/jan/03/peak-hype-driverless-car-revolution-uber-robotaxis-autonomous-vehicle (visited on 07/29/2021) (cit. on pp. 1, 8, 9).

[Unal *et al.* 2023] Devrim Unal, Ferhat Ozgur Catak, Mohammad Talal Houkan, Mohammed Mudassir, and Mohammad Hammoudeh. "Towards robust autonomous driving systems through adversarial test set generation". In: *ISA Transactions* 132 (2023), pp. 69–79. issn: 0019-0578. doi: https://doi.org/10.1016/j.isatra.2022.11.007. url: https://www.sciencedirect.com/science/article/pii/S0019057822005997 (cit. on p. 13).

[Urry 2016] John Urry. *What is the Future?* Cambridge: Polity, 2016 (cit. on pp. 8, 44).

[Wang *et al.* 2022] Qingyan Wang *et al.* "Traffic lights detection and recognition method based on the improved yolov4 algorithm". In: *Sensors* 22.1 (2022). issn: 1424-8220. doi: 10.3390/s22010200. url: https://www.mdpi.com/1424-8220/22/1/200 (cit. on p. 29).

[Wang. *et al.* 2019] Yaqin Wang., Dongfang Liu., Hyewon Jeon., Zhiwei Chu., and Eric Matson. "End-to-end learning approach for autonomous driving: a convolutional neural network model". In: *Proceedings of the 11th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,* INSTICC. SciTePress, 2019, pp. 833–839. isbn: 978-989-758-350-6. doi: 10.5220/0007575908330839 (cit. on p. 23).

[World Health Organization 2015] World Health Organization. *Global status report on road safety 2015.* World Health Organization, 2015, 323 p. (Cit. on p. 1).

[Xiao *et al.* 2022] Yi Xiao, Felipe Codevilla, Akhil Gurram, Onay Urfalioglu, and Antonio M. López. "Multimodal end-to-end autonomous driving". In: *IEEE Transactions on Intelligent Transportation Systems* 23.1 (Jan. 2022), pp. 537–547. issn: 1558-0016. doi: 10.1109/TITS.2020.3013234 (cit. on p. 13).

[Zhang 2019] Jiakai Zhang. "End-to-end learning for Autonomous Driving". PhD thesis. New York, US: Department of Computer Science, New York University, May 2019 (cit. on pp. 10–13).