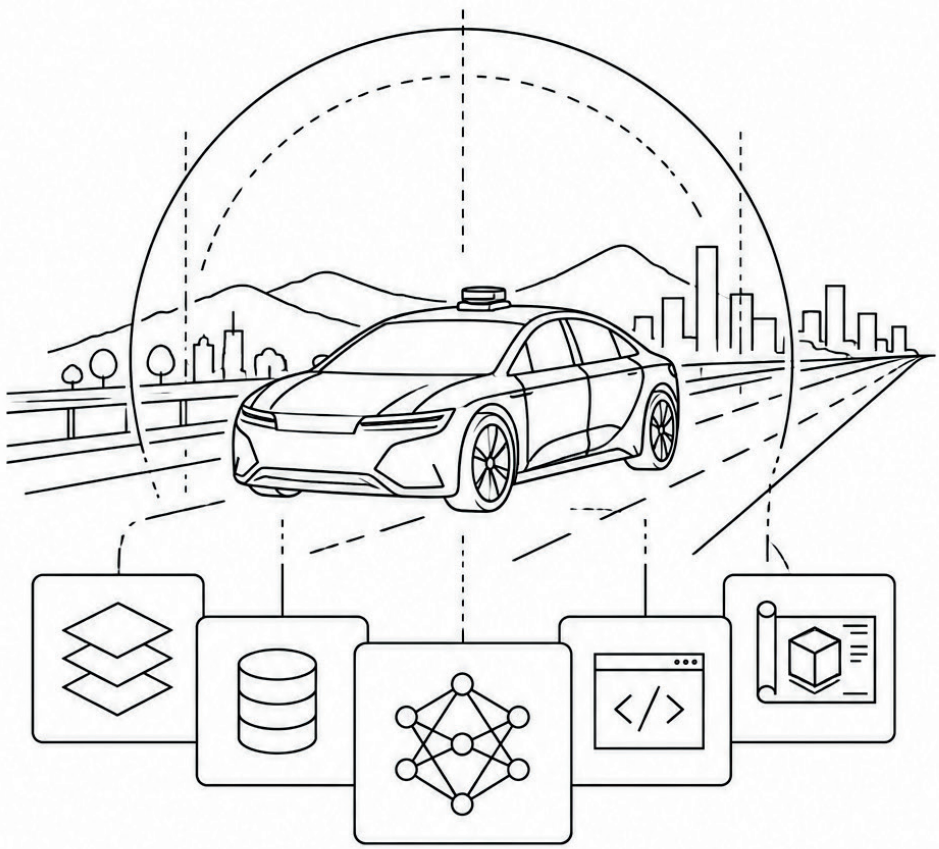
Mälardalen University Doctoral Dissertation 476

EFFICIENT MODEL DESIGN AND SELF-SUPERVISED LEARNING FOR REAL-TIME AUTONOMOUS VEHICLE PERCEPTION

Ali Zoljodi



Mälardalen University

Mälardalen University Press Dissertations
No. 476

EFFICIENT MODEL DESIGN AND SELF-SUPERVISED LEARNING FOR REAL-TIME AUTONOMOUS VEHICLE PERCEPTION

Ali Zoljodi

2026

Department of Computer Science & Engineering

Copyright © Ali Zoljodi, 2026
ISBN 978-91-7485-771-9
ISSN 1651-4238
Printed by E-Print AB, Stockholm, Sweden

Mälardalen University Press Dissertations
No. 476

EFFICIENT MODEL DESIGN AND SELF-SUPERVISED LEARNING FOR REAL-TIME AUTONOMOUS VEHICLE PERCEPTION

Ali Zoljodi

Akademisk avhandling

som för avläggande av teknologie doktorsexamen i datavetenskap vid Institutionen för datavetenskap och datateknik kommer att offentligen försvaras fredagen den 6 november 2026, 13.30 i Kappa, Mälardalens universitet och digitalt, Västerås.

Fakultetsopponent: Professor Enzo Tartaglione, Télécom Paris, France.

Institutionen för datavetenskap och datateknik

Abstract

Developing vehicles capable of operating without human drivers has long been a dream. Autonomous driving systems are the culmination of this dream, potentially becoming a reality if we can overcome challenges related to accuracy, response time, safety, and hardware efficiency.

Deep Neural Networks (DNNs) represent a significant advancement in data-driven perception models, revolutionizing computer vision technologies. If developed and trained properly, these models can provide accurate results within an acceptable inference time and become reliable components of autonomous driving systems.

This thesis presents our contributions to the development of accurate yet fast and hardware-efficient DNN models for safety-critical autonomous driving applications such as lane detection, stereo vision, and trajectory prediction. It particularly highlights our neural architecture search (NAS) methods to address the aforementioned challenges. Additionally, we introduce our distinct self-supervised learning method for lane detection, which is critical for training accurate models without relying on human-labeled data. This method significantly enhances the accuracy and generalization of lane detection methods across a wide range of images featuring different weather conditions, textures, and geographies.

The research presents several key contributions. Firstly, the thesis introduces a novel NAS method for 3D lane detection, named 3DLaneNAS, which provides accurate lane capturing on mountain roads and outperforms existing models in both accuracy and latency. Secondly, we propose a fast and accurate NAS method for stereo vision, named FastStereoNet, which is crucial for computing the distance of surrounding obstacles using cameras, and demonstrates superior accuracy and latency across GPUs, CPUs, and FPGAs. Thirdly, we develop a self-supervised lane detection method named CLLD that leverages strong lane priors and a cross-similarity operation to detect lanes robustly under diverse conditions. Fourthly, we introduce a NAS method for trajectory prediction using Lidar data, named TrajectoryNAS, which achieves state-of-the-art results in trajectory prediction with the lowest inference time. Finally, the thesis contributes to model compression by introducing advanced post-training quantization (PTQ)techniques, including cluster-affine transformations and output reconstruction, which significantly improve the accuracy of ultra-low-bit models while ensuring efficiency on resource-constrained hardware.

In conclusion, this thesis has successfully introduced several innovative methods tailored for autonomous driving applications, making substantial advancements in 3D lane detection, stereo vision, lane detection, trajectory prediction, and quantization. Each proposed approach demonstrates significant improvements over existing techniques in terms of accuracy, latency, generalization, and hardware efficiency across different driving conditions and platforms.

For future work, we aim to further enhance robustness and generalization by extending our methods with vision transformers and exploring hybrid NAS–quantization pipelines. Such approaches are expected to leverage deeper contextual information from visual data while further reducing computational demands, thereby improving the reliability of autonomous driving technologies in diverse real-world environments.

ISBN 978-91-7485-771-9
ISSN 1651-4238

Sammanfattning

Att utveckla fordon som kan framföras utan mänskliga förare har länge varit en vision. System för autonom körning utgör förverkligandet av denna vision och kan bli verklighet om utmaningar relaterade till noggrannhet, responstid, säkerhet och hårdvarueffektivitet kan övervinnas.

Djupa neurala nätverk (DNN) utgör ett betydande framsteg inom datadrivna perceptionsmodeller och har revolutionerat tekniker inom datorseende. Om dessa modeller utvecklas och tränas på ett lämpligt sätt kan de ge noggranna resultat inom en acceptabel inferenstid och bli tillförlitliga komponenter i system för autonom körning.

Denna avhandling presenterar våra bidrag till utvecklingen av noggranna, snabba och hårdvarueffektiva DNN-modeller för säkerhetskritiska tillämpningar inom autonom körning, såsom körfältsdetektering, stereoseende och trajektorieförutsägelse. Särskild tonvikt läggs på våra metoder för neural arkitektursökning (NAS), vilka syftar till att hantera de ovannämnda utmaningarna. Dessutom introducerar vi en egen metod för självövervakad inlärning för körfältsdetektering, vilken är viktig för att träna noggranna modeller utan att vara beroende av data som annoterats av människor. Metoden förbättrar avsevärt noggrannheten och generaliseringsförmågan hos metoder för körfältsdetektering över ett brett spektrum av bilder med varierande väderförhållanden, texturer och geografiska miljöer.

Forskningen presenterar flera centrala bidrag. För det första introducerar avhandlingen en ny NAS-metod för detektering av körfält i 3D, benämnd **3DLaneNAS**, som möjliggör noggrann detektering av körfält på bergsvägar och överträffar befintliga modeller vad gäller både noggrannhet och latens. För det andra föreslår vi en snabb och noggrann NAS-metod för stereoseende, benämnd **FastStereoNet**, vilken är viktig för att beräkna avståndet till omgivande hinder med hjälp av kameror och uppvisar hög noggrannhet och låg latens på GPU:er, CPU:er och FPGA:er. För det tredje utvecklar vi en självövervakad metod för körfältsdetektering, benämnd **CLLD**, som utnyttjar stark förhandskunskap om körfält samt en korslikhetsoperation för robust detektering av körfält under varierande förhållanden. För det fjärde introducerar vi en NAS-metod för trajektorieförutsägelse baserad på LiDAR-data, benämnd **TrajectoryNAS**, som uppnår resultat i nivå med den senaste forskningen inom trajektorieförutsägelse samtidigt som den har mycket låg inferenslatens. Slutligen bidrar avhandlingen till modellkomprimering genom att introducera avancerade tekniker för **kvantisering efter träning (PTQ)**, inklusive klustervisa affina transformationer och utgångsrekonstruktion. Dessa tekniker förbättrar avsevärt noggrannheten hos modeller med extremt låg bitbredd samtidigt som effektiviteten bibehålls på hårdvara med begränsade resurser.

Sammanfattningsvis introducerar denna avhandling flera innovativa metoder anpassade för tillämpningar inom autonom körning och bidrar därigenom till

framsteg inom 3D-körfältsdetektering, stereoseende, körfältsdetektering, trajektorieförutsägelse och kvantisering. Varje föreslagen metod uppvisar betydande förbättringar jämfört med befintliga tekniker vad gäller noggrannhet, latens, generaliseringsförmåga och hårdvaruеffektivitet under olika körförhållanden och på olika beräkningsplattformar.

I framtida arbete avser vi att ytterligare förbättra robustheten och generaliseringsförmågan genom att utvidga våra metoder med **Vision Transformers** och undersöka hybrida lösningar som kombinerar NAS och kvantisering. Sådana metoder förväntas kunna utnyttja djupare kontextuell information från visuella data samtidigt som beräkningskraven minskas ytterligare, vilket kan förbättra tillförlitligheten hos tekniker för autonom körning i varierande verkliga miljöer.

Abstract

Developing vehicles capable of operating without human drivers has long been a dream. Autonomous driving systems are the culmination of this dream, potentially becoming a reality if we can overcome challenges related to accuracy, response time, safety, and hardware efficiency.

Deep Neural Networks (DNNs) represent a significant advancement in data-driven perception models, revolutionizing computer vision technologies. If developed and trained properly, these models can provide accurate results within an acceptable inference time and become reliable components of autonomous driving systems.

This thesis presents our contributions to the development of accurate yet fast and hardware-efficient DNN models for safety-critical autonomous driving applications such as lane detection, stereo vision, and trajectory prediction. It particularly highlights our neural architecture search (NAS) methods to address the aforementioned challenges. Additionally, we introduce our distinct self-supervised learning method for lane detection, which is critical for training accurate models without relying on human-labeled data. This method significantly enhances the accuracy and generalization of lane detection methods across a wide range of images featuring different weather conditions, textures, and geographies.

The research presents several key contributions. Firstly, the thesis introduces a novel NAS method for 3D lane detection, named **3DLaneNAS**, which provides accurate lane capturing on mountain roads and outperforms existing models in both accuracy and latency. Secondly, we propose a fast and accurate NAS method for stereo vision, named **FastStereoNet**, which is crucial for computing the distance of surrounding obstacles using cameras, and demonstrates superior accuracy and latency across GPUs, CPUs, and FPGAs. Thirdly, we develop a self-supervised lane detection method named **CLLD** that leverages strong lane priors and a cross-similarity operation to detect lanes robustly under diverse conditions. Fourthly, we introduce a NAS method for trajectory prediction using LiDAR data, named **TrajectoryNAS**, which achieves state-of-the-art results in trajectory prediction with the lowest inference time. Finally, the thesis contributes to model compression by introducing advanced **post-training quantization (PTQ)** techniques, including cluster-affine transformations and output reconstruction, which significantly improve the accuracy of ultra-low-bit models while ensuring efficiency on resource-constrained hardware.

In conclusion, this thesis has successfully introduced several innovative methods tailored for autonomous driving applications, making substantial advancements in 3D lane detection, stereo vision, lane detection, trajectory prediction, and quantization. Each proposed approach demonstrates significant improvements over existing techniques in terms of accuracy, latency,

generalization, and hardware efficiency across different driving conditions and platforms.

For future work, we aim to further enhance robustness and generalization by extending our methods with **vision transformers** and exploring hybrid NAS–quantization pipelines. Such approaches are expected to leverage deeper contextual information from visual data while further reducing computational demands, thereby improving the reliability of autonomous driving technologies in diverse real-world environments.

Acknowledgments I would like to express my sincere gratitude to all those who have supported me throughout this doctoral journey.

First and foremost, I would like to thank my supervisor, **Professor Masoud Daneshtalab**, for his guidance, encouragement, and continuous support during my PhD studies. His insight, trust, and academic mentorship have played a central role in shaping this research and in my development as a researcher.

I am also deeply grateful to **Mälardalen University** for providing me with the opportunity, academic environment, and resources to pursue my doctoral studies. I sincerely appreciate the inspiring atmosphere and the support of colleagues, collaborators, and staff members who contributed to a productive and rewarding research experience.

I am profoundly thankful to my mother, **Narges**, whose sacrifices, love, and unconditional belief in me have guided me throughout my life. Whatever I have achieved is deeply rooted in her dedication, strength, and unwavering support.

I would also like to express my deepest gratitude to my father, **Iraj**, for his constant support, encouragement, and everything he has done for me throughout my life.

I am sincerely grateful to my brother, **Pouya**, and my girlfriend, **Setayesh**, for their love, kindness, and unwavering support. I am deeply thankful to my family and friends for their kindness, encouragement, and support during both joyful and challenging times.

Finally, I would like to thank everyone who, in one way or another, contributed to this journey and helped make this achievement possible.

Ali Zoljodi
Västerås, Sweden
September 17, 2026

List of Publications

Papers included in the thesis[1]

Paper A: **Zoljodi, A.**, Loni, M., Abadijou, S., Alibeigi, M., Daneshtalab, M. (2022) 3DLaneNAS: Neural Architecture Search for Accurate and Light-Weight 3D Lane Detection. *31st International Conference on Artificial Neural Networks (ICANN)*, United Kingdom, September 7, 2022.

Paper B: Loni, M., **Zoljodi, A.**, Majd, A., Ahn, B. H., Daneshtalab, M., Sjödin, M., Esmaeilzadeh, H. (2022) FastStereoNet: A Fast Neural Architecture Search for Improving the Inference of Disparity Estimation on Resource-Limited Platforms. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(8), August 2022.

Paper C: **Zoljodi, A.**, Abadijou, S., Alibeigi, M., Daneshtalab, M. (2024) Contrastive Learning for Lane Detection via Cross-Similarity. *Pattern Recognition Letters*, 185:175–183, September 2024.

Paper D: Sharifi, A. A.*, **Zoljodi, A.***, Daneshtalab, M. (2024) TrajectoryNAS: A Neural Architecture Search for Trajectory Prediction. *MDPI Sensors*. *Equal contribution.

Paper E: **Zoljodi, A.**, Tanveer, H., Brenig, J., Timofte, R., Daneshtalab, M. (2026) CAT: Post-Training Quantization Error Reduction via Cluster-based Affine Transformation. *In preparation.*

Additional Publications, not included in the thesis

Paper 1*: Loni, M., Sinaei, S., **Zoljodi, A.**, Daneshtalab, M., Sjödin, M. (2020) DeepMaker: A Multi-Objective Optimization Framework for Deep Neural Networks in Embedded Systems. *Microprocessors and Microsystems*, 73:102989.

Paper 2*: Loni, M., **Zoljodi, A.**, Sinaei, S., Daneshtalab, M., Sjödin, M. (2019) NeuroPower: Designing Energy Efficient Convolutional Neural Network Architecture for Embedded Systems. *International Conference on Artificial Neural Networks (ICANN)*, pp. 208–222.

Paper 3*: Loni, M., **Zoljodi, A.**, Maier, D., Majd, A., Daneshtalab, M., Sjödin, M., Juurlink, B., Esmaeilzadeh, H. (2020) DenseDisp: Resource-Aware Disparity Map Estimation by Compressing Siamese Neural Architecture. *IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8.

Paper 4*: Sharifi, A. A., **Zoljodi, A.**, Daneshtalab, M. (2024) DAT: Deep Learning-Based Acceleration-Aware Trajectory Forecasting. *Journal of Imaging*, 10(12):321.

[1]The included papers have been reformatted to comply with the thesis layout.

Paper 5*: Mousavi, H., **Zoljodi, A.**, Daneshtalab, M. (2023) Analysing Robustness of Tiny Deep Neural Networks. *European Conference on Advances in Databases and Information Systems (ADBIS)*, pp. 150–159.

Contents

I Thesis 1

1 Introduction 2
1.1 What is an Autonomous Vehicle? 2
1.2 Autonomous Vehicles as an Artificial Intelligence Challenge 4
1.2.1 Perception in Autonomous Driving 4
1.2.2 Deep Learning for Autonomous-Driving Perception 5
1.2.3 Why Autonomous Driving Matters? 6
1.2.4 Human Behavior, Determinism, and Driving 7
1.2.5 Learning from the Human Brain 8
1.2.6 From Biological Inspiration to Efficient Autonomous Vehicles 10
1.2.7 Adaptive and Efficient Learning 10
1.2.8 Representation Learning and Generalization 11
1.2.9 Dissertation Overview 12

2 Research 14
2.1 Problem Statement 14
2.2 Research Goals 15
2.3 Research Methodology 16

3 Background and Related Work 18
3.1 Motivation for Autonomous Vehicles 18
3.1.1 Human Cognitive Capabilities Relevant to Autonomous Vehicles 19
3.2 Autonomous Vehicles 20
3.2.1 Sensors and Sensor Characteristics 20
3.2.2 Modular and End-to-End Approaches 21
3.3 Perception for Autonomous Driving 21
3.3.1 Object Detection and Three-Dimensional Scene Understanding 21
3.3.2 Lane Detection 22
3.3.3 Tracking and Motion Forecasting 22
3.3.4 Perception Challenges 23
3.4 Neural Network Architectures 23
3.4.1 Convolutional Neural Networks 23
3.4.2 Temporal and Recurrent Models 24

3.4.3 Transformers and Attention ... 24
3.4.4 Computational Cost ... 25
3.5 Neural Architecture Search ... 25
3.5.1 Search Spaces ... 26
3.5.2 Search Strategies ... 26
3.5.3 Performance Estimation and Weight Sharing ... 26
3.5.4 Hardware-Aware and Joint Search ... 27
3.6 Neural Network Quantization ... 27
3.6.1 Uniform Quantization ... 27
3.6.2 Quantization-Aware Training and Post-Training Quantization ... 27
3.6.3 Sources of Quantization Error ... 28
3.7 Representation Learning ... 28
3.7.1 Supervised and Self-Supervised Learning ... 29
3.7.2 Contrastive Learning ... 29
3.7.3 Global, Local, and Dense Representations ... 29
3.7.4 Transferability and Generalization ... 30
3.8 Chapter Summary ... 30

4 Thesis Contributions ... 31
4.1 Contributions and Research Subgoal Relationship ... 33

5 Discussion, Conclusion and Future Work ... 35
5.1 Discussion ... 35
5.2 Limitations ... 36
5.3 Conclusion ... 37
5.4 Future Work ... 37

II Included Papers ... 56

Paper A: 3DLaneNAS: Neural Architecture Search for Accurate and Light-Weight 3D Lane Detection ... 56
6.1 Introduction ... 57
6.2 Related Work ... 58
6.2.1 Lane Detection ... 58
6.2.2 Neural Architecture Search ... 59
6.3 3DLaneNAS ... 60
6.3.1 Search Space ... 60
6.3.2 Search Algorithm ... 61
6.3.3 Training Procedure ... 62
6.4 Experimental Setup ... 63
6.4.1 Dataset ... 63
6.4.2 Configuration Setup ... 63
6.4.3 Evaluation Metrics ... 63
6.5 Experimental Results ... 63

6.5.1 3D Lane Detection Performance Metrics ... 63
6.5.2 Analyzing Search Methods ... 64
6.5.3 Statement of Reproducibility ... 65
6.6 Conclusion ... 66

Bibliography ... 66

Paper B: FastStereoNet: A Fast Neural Architecture Search for Improving the Inference of Disparity Estimation on Resource-Limited Platforms ... 71
7.1 Introduction ... 72
7.2 Related Work ... 74
7.2.1 Disparity Estimation on Resource-limited devices ... 74
7.2.2 Neural Architecture Search (NAS) ... 74
7.3 Search Strategy ... 76
7.3.1 Proposed Method ... 76
7.3.2 FastStereoNet Practical Proof of Convergence ... 80
7.3.3 FastStereoNet Complexity Analysis ... 81
7.4 Experimental Setup ... 81
7.4.1 Evaluation Dataset ... 81
7.4.2 Hardware Specifications ... 81
7.4.3 Experimental Configuration ... 82
7.5 Results ... 83
7.5.1 Disparity Estimation Performance ... 83
7.5.2 Estimation Error Across Different Metrics ... 85
7.5.3 Smoothing Comparison ... 85
7.5.4 Search Convergence Analysis ... 88
7.5.5 Analyzing Search Methods ... 88
7.5.6 Reproducibility of the Results ... 88
7.5.7 FPGA Implementation Results ... 89
7.6 Architecture Design Space ... 90
7.6.1 Siamese Network Architecture ... 90
7.6.2 Representation of CNN Architectures ... 91
7.6.3 Efficient NAS by Transferred Weights Mechanism ... 92
7.7 Conclusion ... 93

Bibliography ... 93

Paper C: Contrastive Learning for Lane Detection via Cross-Similarity ... 103
8.1 Introduction ... 104
8.2 Related work ... 106
8.2.1 lane detection ... 106
8.2.2 contrastive learning ... 107
8.3 Methodology ... 108
8.3.1 Contrastive Learning for Lane Detection ... 109
8.3.2 Masking ... 109

8.3.3 cross-similarity ... 110
8.3.4 Loss function ... 112
8.4 Experimental setup ... 113
8.4.1 Benchmarks ... 113
8.5 Results ... 114
8.5.1 Comparison with prior works ... 114
8.5.2 Comparison with supervised learning ... 116
8.5.3 Visual demonstration ... 117
8.6 Ablation study ... 118
8.6.1 Similarity loss impact ... 118
8.6.2 Impact of masking as the augmentation ... 118
8.7 Discussion ... 119
8.8 Conclusion ... 119

Bibliography ... 120

Paper D: TrajectoryNAS: A Neural Architecture Search for Trajectory Prediction ... 126
9.1 Introduction ... 127
9.2 Related works ... 128
9.2.1 Trajectory Prediction ... 128
9.2.2 Neural Architecture Search ... 132
9.3 TrajectoryNAS ... 132
9.3.1 Search Space ... 134
9.3.2 Search Algorithm ... 136
9.4 Experimental setup ... 138
9.4.1 Dataset ... 138
9.4.2 Evaluation Metrics ... 138
9.4.3 Configuration Setup ... 139
9.5 Results ... 140
9.5.1 Trajectory Prediction Performance ... 140
9.5.2 Analysing Search Methods ... 140
9.5.3 Visual Demonstration ... 141
9.6 Conclusion ... 142

Bibliography ... 142

Paper E: CAT: Post-Training Quantization Error Reduction via Cluster-based Affine Transformation ... 149
10.1 Introduction ... 150
10.2 Related Work ... 151
10.3 Method ... 153
10.3.1 Quantization Notation ... 153
10.3.2 Cluster-based Affine Transformation for Error Reduction ... 154

10.4 Experiments 156
10.5 Ablation Study 159
10.6 CAT improvement on ViT quantization via different PTQ methods 163
10.7 Inference-Time Overhead 164
10.8 Discussion 164
10.9 Conclusion 165
10.10 Enhance PTQ methods with CAT 168
10.11 Ablation on CAT fitting sample size 172

Bibliography 179

I

Thesis

1. Introduction

"The problem of understanding intelligence is the problem of understanding how experience changes the brain."

. Donald Hebb

1.1 What is an Autonomous Vehicle?

Autonomous vehicles are among the most ambitious real-world applications of artificial intelligence. However, before discussing their technological foundations, it is important to note that there is no universally accepted definition of autonomous vehicle [92]. Unlike many artificial intelligence systems that operate in constrained digital environments, an autonomous vehicle must perceive, interpret, and interact with an open and continuously changing physical world. It must recognize relevant objects and road structures, estimate their spatial relationships, predict the future behavior of other traffic participants, plan a safe trajectory, and execute control actions within strict real-time constraints [150].

A *Driving Automation System (DAS)* performs part or all of the *Dynamic Driving Task (DDT)* on a sustained basis. An *Autonomous Vehicle (AV)* is capable of performing the complete DDT within its Operational Design Domain (ODD), where the operational conditions under which the system is designed to function are explicitly defined [11]. Driving automation is categorized into six levels, ranging from Level 0 (no driving automation) to Level 5 (full driving automation under all driving conditions), as illustrated in Fig. 1.1.

A typical autonomous vehicle's software consists of five major functional components: perception [122, 155, 96, 8, 60], localization [17, 57, 132, 15], prediction [54, 80], planning [126, 43], and control [133]. Perception refers to the ability to perceive the surroundings of the autonomous vehicle, detect vehicles and pedestrians, and understand road scenes [159]. Localization estimates the precise position of the ego vehicle within its environment in a global coordinate system [70]. Prediction and planning leverage the generated environmental model to anticipate the future trajectories of surrounding traffic participants and generate a safe and efficient driving strategy for the ego vehicle [44]. Finally, the control system converts the planned trajectory into steering, braking, and acceleration commands that guide the vehicle safely through the environment [161, 151, 6].

Autonomous vehicles require the aforementioned components to operate continuously and cooperatively under highly dynamic and uncertain conditions. Modern autonomous vehicles employ multiple complementary sensing modalities, including cameras [18, 62], LiDAR [73] , radar [147] , Global Navigation Satellite Systems (GNSS) [63] , and inertial measurement units (IMUs) [16] , to obtain a comprehensive understanding of their surroundings. Sensor fusion [140] approaches integrate these heterogeneous observations to improve perception accuracy, robustness, and redundancy.

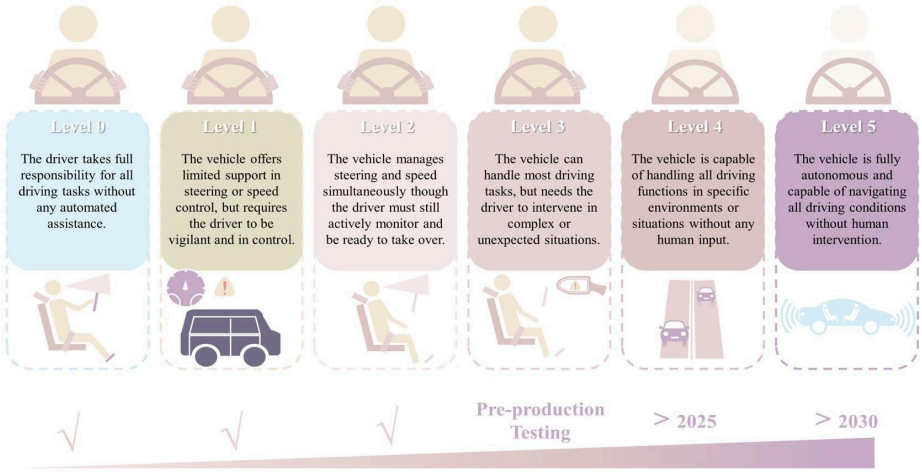


Figure 1.1. The six levels of driving automation defined by the SAE J3016 standard, ranging from Level 0 (No Driving Automation) to Level 5 (Full Driving Automation). The figure also illustrates the gradual shift of responsibility for the Dynamic Driving Task (DDT) from the human driver to the automated driving system.

Recent advances in deep learning have fundamentally transformed autonomous vehicles [160]. Deep neural networks [38] have become the dominant approach for solving perception tasks such as object detection [105] , semantic segmentation [36] , lane detection [8] , stereo matching [128] , depth estimation [112] , and trajectory forecasting [80] . Their ability to learn hierarchical representations directly from large-scale data has significantly improved the capability of autonomous vehicles to understand complex traffic scenes. However, these improvements have also introduced new challenges, including computational complexity, latency, memory consumption, robustness, uncertainty estimation, and deployment on resource-constrained embedded platforms [65].

Consequently, autonomous vehicles have become a multidisciplinary research field that combines artificial intelligence, computer vision, robotics, sensor fusion, embedded systems, optimization, and control theory. Beyond improving prediction accuracy, future autonomous vehicles must achieve high levels of robustness, adaptability, computational efficiency, and functional safety to operate reliably under diverse environmental conditions and previously unseen scenarios [160]. These challenges motivate the research presented throughout this dissertation, which focuses on developing accurate, efficient, and hardware-aware deep learning methods for safety-critical autonomous vehicles.

1.2 Autonomous Vehicles as an Artificial Intelligence Challenge

As an Artificial Intelligence (AI) system, an Autonomous Vehicle (AV) must continuously perceive its environment, process sensory information, make driving decisions, and execute appropriate actions. While these stages are common to most AI systems, autonomous driving imposes far stricter requirements on each of them, particularly under stringent real-time constraints.

Consider a general-purpose chatbot based on a Large Language Model (LLM). Such a system is expected to generate responses quickly and remain relevant to the user's query. However, minor inaccuracies or slight deviations from the user's intent are often tolerable because the interaction is iterative—the user can ask follow-up questions, allowing the system to refine or correct its responses over time.

Now consider an AI system designed for medical diagnosis, where blood test results or medical imaging are analyzed to detect diseases. In this scenario, diagnostic accuracy is of paramount importance. Nevertheless, the system typically has the luxury of additional processing time, ranging from minutes to days, and the analysis can be repeated or verified before a final diagnosis is made.

Autonomous driving, however, combines the strictest requirements of both domains. An Autonomous Vehicle cannot tolerate either inaccurate decisions or excessive processing delays. Even a decision that is highly accurate but produced a few seconds too late may result in a catastrophic accident. Likewise, a fast but incorrect decision can have equally severe consequences.

Furthermore, Autonomous Vehicles must operate in complex, dynamic, and largely unstructured environments without human supervision. Their perception systems must cope with imperfect sensor measurements, changing weather and lighting conditions, occlusions, and both adversarial and non-adversarial perturbations, all while maintaining reliable performance.

Therefore, autonomous driving represents one of the most demanding AI applications, requiring an exceptional balance of non-functional qualities, including low latency, high accuracy, strong generalization, and robustness. Achieving these properties simultaneously is fundamental to ensuring the safety and reliability of Autonomous Vehicles.

1.2.1 Perception in Autonomous Driving

Perception forms the interface between an autonomous vehicle and its environment. Its purpose is not merely to record sensor observations, but to transform them into structured information that supports prediction, planning, and control. A reliable perception system must identify relevant traffic participants, estimate their positions and motions, recognize road boundaries and lane markings,

interpret traffic signs and signals, estimate traversable space, and represent uncertainty in its observations [152].

Autonomous vehicles commonly employ complementary sensing modalities, including cameras, LiDAR, radar, ultrasonic sensors, Global Navigation Satellite Systems, and inertial measurement units. Cameras provide rich appearance and semantic information but can be sensitive to illumination and weather conditions.LiDAR provides accurate geometric measurements and three-dimensional structure, although its measurements may become sparse at long ranges and can be degraded by adverse weather. Radar offers reliable range and velocity measurements and is comparatively robust under poor visibility, but generally provides lower spatial resolution. The complementary characteristics of these sensors motivate multi-sensor fusion, through which information from several modalities is combined to improve environmental understanding and system redundancy [151, 6, 55].

Autonomous-driving perception includes several closely related tasks. Object detection identifies and localizes vehicles, pedestrians, cyclists, and other obstacles. Semantic and instance segmentation assign semantic meaning to image pixels or three-dimensional points. Lane detection and road-boundary estimation provide representations of the drivable structure. Stereo matching and monocular depth estimation recover the geometry of the scene. Object tracking associates detections over time, while occupancy estimation represents free, occupied, and potentially occluded regions. These outputs form an intermediate representation upon which motion forecasting, behavior prediction, and trajectory planning can operate.

The safety of the complete autonomous system therefore depends strongly on the quality of perception. An object that is not detected cannot be considered by the planner, and an inaccurate estimate of its location or velocity may lead to an unsafe prediction. Perception errors may also propagate through the downstream pipeline. Consequently, autonomous-driving perception must be evaluated not only in terms of average accuracy, but also in terms of temporal consistency, uncertainty, robustness to environmental variation, computational latency, and performance in safety-critical corner cases.

1.2.2 Deep Learning for Autonomous-Driving Perception

Convolutional neural networks (As one of the most influential deep learning architectures in computer vision) substantially improved image classification, object detection, semantic segmentation, lane detection, and depth estimation. Recurrent models and temporal convolutional architectures enabled the integration of sequential information for tracking and motion prediction. More recently, attention-based and transformer architectures have provided mechanisms for modeling long-range spatial relationships, temporal interactions, and cross-modal dependencies. Deep learning is now used throughout both modular

autonomous-driving pipelines and end-to-end systems that learn mappings from sensor inputs to driving decisions [39, 55].

The principal advantage of deep neural networks is their ability to learn task-relevant representations from data. Rather than explicitly programming every visual pattern associated with a pedestrian, lane marking, or vehicle, a network learns statistical structures that support recognition across many observations. This representation-learning capability is particularly important in autonomous driving, where the number of possible combinations of road geometry, traffic behavior, weather, illumination, and object appearance is effectively unbounded.

Nevertheless, high benchmark accuracy does not by itself guarantee safe real-world performance. Deep neural networks commonly assume that training and deployment data follow similar distributions. In practice, autonomous vehicles may encounter new cities, sensor configurations, road markings, weather conditions, traffic conventions, and previously unseen objects. Such domain shifts can significantly reduce model performance. Generalization, domain adaptation, uncertainty estimation, continual learning, and robustness to sensor degradation therefore remain central research problems in autonomous-driving perception [83, 2].

Deep models must also operate under the strict computational constraints of embedded automotive platforms. Larger networks may improve predictive accuracy, but they also increase latency, memory consumption, communication overhead, and energy demand. In a safety-critical system, late information can be nearly as harmful as incorrect information. The design of autonomous-driving perception consequently requires a balance among accuracy, robustness, computational efficiency, and response time. This balance motivates methods such as neural architecture search, network pruning, quantization, knowledge distillation, efficient representation learning, and hardware-aware optimization.

1.2.3 Why Autonomous Driving Matters?

The motivation for autonomous driving begins with the human and societal cost of road transportation. The World Health Organization estimates that ≈ 1.19 million people die in road traffic crashes each year. Road traffic injuries are also a leading cause of death among children and young adults between 5 and 29 years of age [138].

Driving is a complex task performed by a biological agent whose behavior and performance vary over time. Human drivers may be affected by fatigue, distraction, stress, intoxication, emotional state, limited visibility, divided attention, and differences in experience. Crash investigations have frequently attributed the immediate critical reason preceding a crash to driver-related factors. However, such findings should not be interpreted as showing that humans are the sole cause of crashes, because road design, vehicle condition, envi-

ronmental factors, and interactions among these variables may also contribute [120].

An automated driving system is not affected by fatigue, intoxication, impatience, or emotional distraction in the human sense. It can continuously monitor multiple sensing modalities and execute the same programmed safety policy without becoming tired. It can also potentially react faster than a human driver, communicate with other vehicles, and benefit from improvements learned across a fleet. These characteristics provide a strong motivation for developing driving automation.

However, automated systems possess their own failure modes. They may misinterpret unfamiliar objects, fail under distribution shifts, become overconfident in ambiguous situations, or be affected by sensor occlusion, calibration errors, weather, software faults, and adversarial interference. The safety argument for autonomous vehicles should therefore not be that machines are inherently infallible or already superior to humans in every condition. Rather, autonomous vehicles offer the possibility of replacing variable human performance with systems that can be systematically tested, replicated, monitored, updated, and improved.

Beyond road safety, autonomous vehicles may improve mobility for people who cannot drive, including some elderly and disabled individuals. It may increase the efficiency of freight transportation, enable more consistent vehicle utilization, reduce some forms of congestion, and support new mobility services. At the same time, its deployment raises significant questions concerning employment, legal liability, cybersecurity, privacy, public acceptance, and the distribution of technological benefits.

1.2.4 Human Behavior, Determinism, and Driving

Sapolsky presents human behavior as the outcome of interacting causes operating over multiple timescales. An action is influenced by neural activity occurring immediately before it, hormonal and physiological states operating over longer periods, prior experiences, childhood development, social environment, culture, genetics, and evolutionary history [114, 115]. This framework challenges explanations that treat a driver's decision as the product of an isolated and unconstrained moment of choice.

Applied to driving, this perspective suggests that a human response to a dangerous situation is affected by far more than the visible road scene. A decision may also depend on fatigue accumulated during the preceding hours, stress experienced earlier in the day, learned habits, driving experience, attentional state, age, cultural norms, and biological variation. Two drivers observing the same event may therefore perceive its significance differently and produce different actions.

This variability provides one conceptual motivation for driving automation. An autonomous vehicle can be designed to respond consistently to specified sensory conditions without being directly affected by anger, sleep deprivation, social pressure, or impulsivity. Nevertheless, Sapolsky's framework should not be used to claim that present autonomous vehicles are universally more robust than human drivers. Human biological variability creates vulnerabilities, but the human brain also provides remarkable abilities for contextual reasoning, adaptation, transfer, and response to novelty. Current autonomous vehicles often perform well in familiar distributions while struggling with scenarios that experienced humans can interpret using broad contextual knowledge.

The deeper implication is therefore not that biological intelligence should simply be replaced, but that it should also be studied as a source of computational principles. Autonomous vehicles can potentially become more capable by incorporating mechanisms inspired by human perception and learning, while avoiding some of the biological limitations that affect human driving.

1.2.5 Learning from the Human Brain

The human brain operates under severe physical constraints. It receives noisy and incomplete sensory input, processes information with limited energy, and must act in real time. Despite these limitations, humans can learn new environments, transfer prior knowledge to unfamiliar situations, infer hidden intentions, focus attention on relevant stimuli, and adapt behavior following relatively few experiences. These capabilities are directly relevant to autonomous vehicles.

Representation Learning

Human perception does not produce an exact reconstruction of the physical world. Instead, the brain develops representations that preserve information relevant to behavior. For example, the human brain can develop a robust understanding of driving lanes regardless of the road surface or environment. Whether the vehicle is traveling on an asphalt highway, a gravel rural road, or a snow-covered street, humans are generally able to recognize lane boundaries and adapt their driving behavior accordingly. A driver need not remember every pixel in a traffic scene. The driver must instead recognize that an object is a pedestrian, infer that the pedestrian may enter the road, and understand that the situation requires increased caution.

Deep neural networks similarly learn internal representations that transform raw sensor observations into task-relevant features. However, learned representations can become overly dependent on superficial correlations in the training data. Brain-inspired research may help motivate representations that better encode objects, relationships, causality, temporal continuity, and uncertainty.

Adaptability and Plasticity

The brain changes through experience. Synaptic strengths, functional connectivity, and task-relevant representations are modified through learning and plasticity. This adaptability allows humans to become familiar with new vehicles, roads, traffic cultures, and environmental conditions.

Most deployed deep networks, by contrast, remain fixed after training. A model optimized for one sensor arrangement or geographic region may lose accuracy when deployed elsewhere. Continual learning, online adaptation, meta-learning, and domain adaptation seek to provide artificial systems with greater flexibility. For autonomous vehicles, such mechanisms could allow perception models to adjust to gradual changes in sensors, seasons, road infrastructure, or traffic behavior without forgetting previously acquired knowledge.

Generalization and Transfer

Humans frequently apply knowledge beyond the exact conditions in which it was learned. A driver who has never seen a particular construction vehicle can still infer that it is large, moving, and potentially hazardous. This capacity reflects the use of compositional representations, prior knowledge, and causal understanding.

Deep learning systems often generalize effectively within the statistical neighborhood of their training data but may fail under substantial distribution shifts. Improving generalization requires diverse training data, appropriate inductive biases, simulation, self-supervised learning, domain adaptation, multimodal learning, and representations that capture more stable properties of the world rather than dataset-specific patterns.

Attention and Selective Processing

The sensory environment contains more information than either a human or an artificial system can process with equal priority. Human attention dynamically allocates resources toward objects and events that are relevant to current goals. In driving, attention may rapidly shift from lane keeping to a child approaching the road, an occluded intersection, or the sudden deceleration of a preceding vehicle.

Attention mechanisms in deep neural networks provide a computational analogue by weighting relevant spatial locations, temporal observations, objects, or sensing modalities. For autonomous vehicles, adaptive attention can improve both accuracy and efficiency by allocating computation according to scene complexity and potential risk.

Temporal Reasoning and Prediction

Human drivers interpret traffic as a continuous process rather than a sequence of independent images. The motion of a pedestrian, orientation of a cyclist, and interaction between vehicles provide evidence about what may happen

next. The brain integrates past observations with contextual knowledge to form predictions.

Autonomous vehicles similarly require temporal representations. A perception system that processes each frame independently may produce unstable detections or overlook motion cues. Recurrent models, temporal convolution, transformers, tracking systems, and world models provide mechanisms for integrating information over time and anticipating future events.

1.2.6 From Biological Inspiration to Efficient Autonomous Vehicles

The brain should not be treated as a complete engineering approach. Biological intelligence evolved under constraints and objectives that differ from those of an autonomous vehicle. Human perception contains biases, illusions, limited attention, and inconsistent decision-making. Nevertheless, the brain demonstrates valuable principles: efficient representation, structural specialization, adaptation, multimodal integration, temporal prediction, learning from limited supervision, and resilience to incomplete information.

The relationship between neuroscience and artificial intelligence can therefore be bidirectional. Neuroscience can suggest computational mechanisms for artificial systems, while artificial models can provide formal tools for investigating biological intelligence. Research at this intersection has argued that the brain may be understood as a heterogeneous learning system in which different structures, learning objectives, and timescales interact to produce efficient behavior [91].

For autonomous vehicles, these ideas motivate several technical directions. Neural architecture search can identify task-specific structures rather than relying on uniformly designed networks. Self-supervised learning can exploit the large amount of unlabeled sensory data generated by vehicles. Quantization and model compression can reduce computational and energy costs.

The objective is not to reproduce the human brain in silicon. It is to develop artificial driving systems that combine the consistency, speed, and reproducibility of computation with selected capabilities associated with biological intelligence, including adaptability, generalization, contextual representation, and efficient learning. This perspective provides the central motivation for the research presented in this dissertation.

1.2.7 Adaptive and Efficient Learning

One remarkable characteristic of the human brain is its ability to continuously reorganize itself according to environmental demands. Neural pathways are strengthened through repeated experience, unnecessary connections are weakened, and computational resources are allocated according to task requirements.

During adolescence, the human frontal cortex undergoes a process of synaptic pruning, reorganizing neural networks by removing redundant connections and strengthening important ones. This refinement increases the efficiency and specialization of the brain for the cognitive functions and skills acquired through experience [9]. This adaptive capability enables biological intelligence to solve complex problems while operating under extremely limited energy consumption.

Inspired by these observations, the first part of this dissertation investigates adaptive optimization techniques for deep neural networks used in autonomous vehicles' perception. Rather than designing network architectures manually, the presented research explores methods that automatically discover efficient model structures while maintaining high predictive performance. Particular emphasis is placed on Neural Architecture Search (NAS), which searches for architectures optimized jointly for accuracy and computational efficiency. Furthermore, model compression techniques such as neural network quantization are investigated to reduce latency, memory footprint, and energy consumption, enabling deployment on embedded automotive hardware.

Collectively, these approaches seek to improve the efficiency of perception systems without sacrificing the reliability required by safety-critical autonomous vehicles.

1.2.8 Representation Learning and Generalization

Human perception extends far beyond memorizing visual appearances. Instead, the brain constructs internal representations that capture stable concepts, relationships, and regularities of the surrounding world. Such representations enable humans to recognize unfamiliar objects, adapt to new environments, and transfer previously acquired knowledge to unseen situations.

Inspired by these capabilities, the second research direction of this dissertation investigates representation learning methods for autonomous vehicles' perception. In particular, it focuses on self-supervised contrastive learning techniques that exploit the vast amount of unlabeled driving data to learn informative visual representations without relying exclusively on manual annotation.

Learning robust representations improves the ability of perception systems to generalize across different weather conditions, road geometries, camera configurations, geographical regions, and traffic scenarios. This research therefore addresses one of the fundamental challenges of autonomous vehicles: enabling perception models to remain reliable when operating beyond the specific data distribution on which they were originally trained.

1.2.9 Dissertation Overview

Taken together, the research presented in this dissertation investigates autonomous vehicle perception from two complementary perspectives. The first perspective focuses on improving computational efficiency through adaptive optimization techniques, including neural architecture search, and quantization. The second perspective focuses on improving robustness and generalization through representation learning and self-supervised learning. These two perspectives address a central challenge in autonomous driving: perception models must provide reliable representations of complex driving environments while satisfying the computational and latency constraints required for real-time deployment.

The first part of the dissertation investigates computationally efficient model design using neural architecture search (NAS). **Paper A** introduces 3DLaneNAS, which applies NAS to monocular 3D lane detection with the objective of identifying accurate yet lightweight architectures. This work demonstrates how automated architecture design can be used to balance perception accuracy and computational complexity for a safety-critical perception task. Extending this direction to depth perception, **Paper B** presents FastStereoNet, a NAS-based approach for disparity estimation that searches for efficient stereo-matching architectures suitable for resource-constrained platforms. Together, these studies establish NAS as an effective mechanism for adapting perception architectures to both task-specific requirements and computational constraints.

The dissertation subsequently extends architecture optimization from perception toward motion understanding. **Paper D** introduces TrajectoryNAS, where NAS is applied to trajectory prediction to explore architectures that balance prediction performance and computational efficiency. While Papers A and B focus primarily on extracting spatial information from the surrounding environment, Paper D addresses the temporal dimension of autonomous driving by considering the future motion of traffic participants. Collectively, these works demonstrate that adaptive architecture optimization can be applied across different stages of an autonomous driving perception pipeline, from geometric scene understanding to motion prediction.

Complementary to efficient architecture design, the second research direction investigates how models can learn more effective representations from data. **Paper C** introduces a self-supervised contrastive learning framework for lane detection based on cross-similarity. Instead of relying exclusively on supervised learning from annotated samples, this work investigates how representations learned through self-supervision can improve downstream lane detection. This direction is particularly relevant to autonomous driving, where collecting large amounts of sensor data is relatively inexpensive, whereas obtaining accurate task-specific annotations can be costly and time-consuming. The study therefore complements the architecture-search contributions by ad-

dressing efficiency from a data and representation perspective rather than solely from a computational perspective.

Finally, the dissertation considers the efficiency of already-trained models at deployment time. **Paper E** introduces CAT, a post-training quantization error reduction method based on cluster-wise affine transformations. Rather than modifying the underlying network architecture or requiring computationally expensive retraining, CAT aims to compensate for errors introduced by low-precision quantization after training. This contribution extends the dissertation's efficiency perspective from architecture-level optimization to model-level compression and deployment, demonstrating how the computational and memory requirements of neural networks can be reduced while mitigating the associated degradation in predictive performance.

Overall, the included papers form a progression across different levels of efficient and reliable autonomous driving intelligence. Papers A, B, and D investigate *how computationally efficient architectures can be automatically designed*; Paper C investigates *how stronger representations can be learned from largely unlabeled data*; and Paper E investigates *how trained models can be efficiently compressed for deployment*. Although these research directions address different technical problems, they share a common objective: developing learning-based systems that are not only accurate, but also computationally efficient, adaptable, and capable of operating reliably under the diverse and uncertain conditions encountered in real-world autonomous driving. The following chapters review the theoretical foundations and relevant literature underlying these research directions before presenting the individual contributions of the dissertation in greater detail.

2. Research

If you recognize that self-driving cars are going to prevent car accidents, AI will be responsible for reducing one of the leading causes of death in the world."

Mark Zuckerberg

2.1 Problem Statement

In recent years, deep neural networks have revolutionized computer vision, driving significant progress in areas such as image classification, object detection, semantic segmentation, and specialized tasks like lane detection for autonomous vehicles. This progress has largely been fueled by increasingly large and complex models that demand immense computational resources and vast amounts of labeled data. While the performance of such models has steadily improved, these trends also present substantial challenges, particularly in terms of data availability, accessibility, and the sustainability of model development.

Model scaling in image perception has been dramatic, beginning with breakthroughs like AlexNet, which used deep convolutional networks to outperform traditional methods on the ImageNet challenge [68]. This set the stage for a rapid evolution of different architectures, such as VGG [119], ResNet [47], and EfficientNet [125] with increasing parameter counts and depth. More recently, Vision Transformers (ViTs) [4] have demonstrated strong performance, though they require even more data and computational power to train effectively. For example, Vision Transformers trained on JFT-300M and ImageNet-21k illustrate the necessity of large-scale data to achieve competitive results.

However, providing enough high-quality labeled data remains a bottleneck, particularly for downstream tasks like lane detection. This task presents unique challenges due to the dynamic and often unpredictable conditions in real-world driving. Lane markings are highly sensitive to adverse lighting, weather conditions, occlusion by vehicles or pedestrians, and deterioration over time. Their appearance can vary significantly due to differences in road types, construction, or even cultural standards across countries. These complexities demand large and diverse datasets that are carefully annotated, which is a resource-intensive process.

The computational intensity of training large models is another hurdle. Vision Transformer models like ViT-H/14 contain ≈ 600 million parameters and require enormous hardware infrastructure, often available only to major tech companies. This raises concerns about the environmental impact of deep learning and the accessibility of state-of-the-art models to smaller institutions. While some models have been optimized for efficiency like MobileNet [51] or EfficientNet. There is still a noticeable trade-off between model complexity and deployability.

To reduce reliance on labeled data, the community has embraced methods such as self-supervised learning, domain adaptation, and synthetic data generation. Techniques like SimCLR [19] and MoCo [46] have shown that high-quality representations can be learned from unlabeled data.

While the trend toward larger, more powerful models in computer vision has yielded impressive results, it also highlights the growing tension between performance and feasibility. Tasks like lane detection exemplify how data collection remains a critical and difficult component of model development. Addressing these challenges will require continued innovation in model design, training strategies, and data-efficient learning approaches to ensure progress in computer vision remains both inclusive and sustainable. Based on these requirements, we defined the following research questions:

1. **What are the bottlenecks of deep learning-based perception systems in autonomous vehicles with respect to latency, memory footprint, and accuracy?** This question aims to investigate the sub-modules of perception models in autonomous vehicles, such as feature extraction and feature fusion, and analyze their impact on the final performance, including potential drawbacks.
2. **How can we leverage neural architecture search (NAS) to address hardware challenges in autonomous vehicle perception?** This question aims to investigate perception accuracy and latency bottlenecks and how to design a novel NAS method to mitigate these bottlenecks.
3. **How can Neural Architecture Search (NAS) be optimized to reduce response time and computational cost?** This question explores optimization methods for Neural Architecture Search (NAS), such as one-shot training and transfer learning, and how they can be effectively leveraged in perception systems for autonomous vehicles.
4. **How can self-supervised learning reduce the dependence of autonomous vehicle perception models on labeled training data?** This question investigates methods to enhance model training with minimal labeled datasets, focusing on self-supervised techniques that utilize available unlabeled data effectively.
5. **How can quantization reduce runtime and memory footprint in DNNs?** This question explores quantization methods, such as Quantization-Aware Training (QAT) and Post-Training Quantization (PTQ), and how they can effectively reduce DNN runtime and memory footprint.

2.2 Research Goals

To address our research question, we conducted a study on Neural Architecture Search (NAS) and its deployment strategies for autonomous vehicle perception systems. The objectives of this research are to design and develop a NAS methodology that enables the creation of highly accurate perception algorithms

for autonomous vehicles, targeting a wide range of hardware platforms, including GPUs, CPUs, and FPGA boards.

Additionally, we explored the training and generalization of perception models in autonomous vehicles using the minimal amount of labeled training data. The outcomes of these studies are presented in the five papers included in this thesis:

Overall goal: Develop methods for designing, training, and deploying generalizable and computationally efficient autonomous vehicle perception models, while considering hardware constraints, response-time requirements, data efficiency, and memory footprint.

For further clarification, the overall goal is divided into the following four subgoals. These subgoals are directly derived from the research questions defined in the previous section. In particular, **Subgoal 1** addresses **RQ1** by identifying and analyzing the main bottlenecks of autonomous vehicle perception systems in terms of accuracy, inference latency, and memory footprint. Based on these identified bottlenecks, **Subgoal 2** addresses **RQ2 and RQ3** by investigating NAS-based solutions and efficient search strategies for designing resource-efficient architectures. **Subgoal 3** corresponds to **RQ4** by addressing the limited availability of labeled training data through self-supervised representation learning, while **Subgoal 4** addresses **RQ5** by investigating quantization techniques for reducing runtime and memory requirements during deployment.

The following subgoals are defined clearly.

- **Subgoal 1:** Analyze the key bottlenecks of autonomous vehicle perception with respect to accuracy, inference latency, and memory footprint.
- **Subgoal 2:** Design and implement low-cost and fast-converging Neural Architecture Search (NAS) methods to address the identified accuracy, inference latency, and memory-footprint bottlenecks under resource-constrained deployment conditions.
- **Subgoal 3:** Design and implement low-cost self-supervised learning algorithms for autonomous vehicle perception based on their characteristics and limitations.
- **Subgoal 4:** Investigate post-training quantization and error-reduction techniques to compress perception models, enabling efficient deployment under strict hardware and real-time constraints.

2.3 Research Methodology

This thesis follows a structured research methodology consisting of four main stages: (1) problem formulation, (2) solution proposition, (3) implementation, and (4) evaluation, following the methodological framework described in [64].

Problem formulation. The research begins by identifying key limitations of learning-based autonomous vehicle perception and prediction systems. In particular, the thesis focuses on challenges related to predictive performance,

inference latency, memory requirements, architecture-search cost, and dependence on large amounts of labeled training data. These challenges form the basis of the research questions and subgoals defined in the previous sections.

Solution proposition. Different methodological approaches are investigated according to the identified challenges. Neural Architecture Search (NAS) is employed to address architecture-level trade-offs between predictive performance and computational efficiency in autonomous vehicle perception. In particular, the thesis investigates NAS-based solutions for 3D lane detection, disparity estimation, and trajectory prediction. Self-supervised representation learning is investigated as a complementary approach for reducing dependence on manually labeled data and improving the generalization of lane-detection models. Finally, post-training quantization and output-level error-reduction techniques are investigated to improve the efficiency of trained models during deployment.

Implementation. The proposed methods are implemented using established deep-learning frameworks, including PyTorch, TensorFlow, and JAX. Each method is developed and evaluated using application-specific datasets, architectures, and experimental configurations appropriate to the corresponding autonomous vehicle perception task.

Evaluation. The proposed solutions are evaluated empirically using several complementary criteria:

- **Predictive performance:** Task-specific evaluation metrics are used to measure the accuracy and effectiveness of the proposed methods, including accuracy, mean intersection over union (mIoU), disparity and trajectory prediction metrics, and other application-specific measures.
- **Computational efficiency:** Inference latency, model size, memory requirements, and architecture-search cost are evaluated where relevant to assess the computational efficiency of the proposed approaches.
- **Hardware evaluation:** Where applicable, inference performance is measured on different hardware platforms to investigate the hardware-dependent behavior and deployment suitability of the resulting models.
- **Comparative evaluation:** The proposed methods are compared with relevant baseline and state-of-the-art approaches using established benchmark datasets and evaluation protocols.
- **Ablation studies:** Where appropriate, ablation experiments are conducted to analyze the contribution of individual components and design choices within the proposed methods.

The results obtained from these stages provide the empirical basis for answering the research questions and evaluating the extent to which the research subgoals are achieved. The outcomes of the studies have been disseminated through the publications included in this thesis and related scientific presentations.

3. Background and Related Work

"We can only see a short distance ahead, but we can see plenty there that needs to be done."

Alan Turing

This chapter introduces the foundations required to position the contributions of this thesis. It begins by motivating the development of Autonomous Vehicles (AVs), highlighting the societal and technological need for autonomous driving. It first discusses the limitations and costs associated with human driving, as reported in the literature, and then explains how autonomous vehicles have the potential to address and mitigate these challenges. It then reviews the functional organization of autonomous vehicles and focuses on perception, the component most directly related to the contributions of this thesis. The chapter subsequently introduces the principal neural-network architectures used in visual perception and motion forecasting, followed by neural architecture search, neural-network quantization, and representation learning. Together, these topics establish the technical background for developing autonomous vehicles that are accurate, transferable, computationally efficient, and suitable for real-time deployment.

3.1 Motivation for Autonomous Vehicles

A large-scale comparative study published in 2024 by researchers from Waymo and Swiss Re [77] provides compelling evidence for the potential safety benefits of autonomous vehicles. The study established a human-driving benchmark using insurance data comprising more than 600,000 liability claims and over 125 billion miles of driving exposure. Compared with this benchmark, trained Waymo autonomous specialists manually operating the vehicles achieved a 45% reduction in bodily-injury claim frequency. More remarkably, during approximately 3.8 million miles of fully autonomous rider-only operation, the Waymo Driver incurred no bodily-injury liability claims, corresponding to a 100% reduction relative to comparable human driving. These findings naturally raise the question of why human driving remains considerably more prone to safety-critical incidents.

Although the human brain possesses extraordinary cognitive capabilities, its ability to sustain high levels of cognitive control is inherently limited. Extended periods of demanding cognitive work induce cognitive fatigue, which elevates the perceived cost of cognitive control and biases behavior toward low-effort, automatic responses. In safety-critical tasks such as driving, these limitations can impair attention, decision making, and hazard response, thereby increasing the risk of severe incidents [108].

The efficiency of human cognitive processing varies according to both internal and external factors. Prolonged mental effort [25], circadian rhythms [10], nutritional state [37], hormonal changes [45], emotional responses [23], and aversive sensory stimuli [34] can all degrade attention and cognitive control, reducing performance in demanding tasks.

Collectively, these limitations often lead to unsafe decisions even before a journey begins, such as driving under the influence of alcohol or drugs, operating a vehicle despite severe fatigue, or driving while experiencing impaired cognitive or physiological conditions. Such examples demonstrate that human decision-making alone is not always sufficient for ensuring safety in safety-critical environments. Consequently, there is an increasing need to automate safety-critical systems that traditionally rely on human judgment, with road transportation representing one of the most important application domains. This necessity constitutes one of the primary motivations behind the development of Autonomous Vehicles.

3.1.1 Human Cognitive Capabilities Relevant to Autonomous Vehicles

Automated systems and human drivers possess complementary strengths and weaknesses. Humans have broad semantic understanding, social intuition, and the capacity to reason about unfamiliar events using prior knowledge [26]. They can often infer the intentions of pedestrians and other drivers from subtle contextual cues. However, humans have limited attention, variable reaction time, biological fatigue, and susceptibility to distraction and impairment.

Automated systems can continuously process multiple sensors, maintain repeatable response times, and avoid biological fatigue or intoxication. Cameras, radar, LiDAR, global navigation satellite systems, and inertial sensors can provide complementary measurements beyond unaided human vision. Nevertheless, automated systems remain constrained by their sensors, training data, model assumptions, operational design domain, and validation coverage. They may fail systematically under rare weather conditions, unusual object appearances, sensor contamination, mapping errors, adversarial configurations, or interactions that were poorly represented during development.

Table 3.1 summarizes the comparison.

Mueller examined which human-like errors automated vehicles would need to avoid in order to realize their theoretical safety potential [95]. The study is important because replacing the human driver does not automatically eliminate crashes. An automated system must correctly sense relevant objects, interpret the scene, predict future motion, select an appropriate manoeuvre, and execute

Table 3.1. *Conceptual comparison between human drivers and autonomous vehicles. The comparison is conditional on the operational design domain and should not be interpreted as evidence that current automated systems universally outperform humans.*

Capability	Human driver	Autonomous Vehicles
Attention	Limited and vulnerable to distraction	Persistent sensor monitoring, subject to sensing and software failures
Fatigue and impairment	Performance varies with sleep, alcohol, drugs, stress, and health	No biological fatigue or intoxication, but hardware may degrade or fail
Reaction	Variable and affected by expectation and workload	Potentially fast and repeatable within the designed pipeline
Semantic understanding	Strong contextual and commonsense reasoning	Limited by learned representations and training coverage
Rare situations	Can sometimes generalize from broad experience	Often vulnerable to distribution shift and long-tail events
Social interaction	Interprets gestures, negotiation, and informal road conventions	Implicit communication remains difficult to model
Sensing	Primarily vision, hearing, proprioception, and memory	Can combine camera, radar, LiDAR, GNSS, IMU, maps, and connectivity
Consistency	Varies across drivers and internal states	Repeatable under equivalent inputs, but may reproduce systematic errors
Failure mode	Perceptual, cognitive, behavioural, or physiological	Sensor, data, model, software, hardware, or integration failure

it safely. A failure in any stage may reproduce an error analogous to a human perception, prediction, or decision failure.

3.2 Autonomous Vehicles

A typical autonomous vehicle is organized as a sequence of sensing, localization, perception, prediction, planning, and control modules [158]. Sensors measure the vehicle and its surroundings. Localization estimates the vehicle pose relative to a map or global frame. Perception transforms raw sensor measurements into structured scene representations. Prediction estimates the future states of surrounding agents. Planning selects a safe and efficient trajectory, and control converts the trajectory into steering, braking, and acceleration commands.

3.2.1 Sensors and Sensor Characteristics

Cameras provide dense appearance information, colour, texture, and semantic cues at relatively low cost, but their output is sensitive to illumination [33], weather [59], motion blur [99], and occlusion [162]. LiDAR [118] directly measures three-dimensional geometry and provides accurate range observations, although its point density decreases with distance and its performance may

be affected by adverse weather [28] and sensor contamination [116]. Radar [31] is robust to lighting and provides direct radial-velocity measurements, but conventional automotive radar has lower angular resolution and noisier spatial structure. GNSS [61] and inertial sensors support localization, while high-definition maps provide prior information about road geometry and traffic structure.

No single sensor is reliable under all conditions. Multi-modal fusion [140] therefore combines complementary measurements to improve coverage and robustness. Fusion [7] may occur at the raw-data level, feature level, object level, or within a shared bird's-eye-view representation. However, fusion introduces calibration, synchronization, missing-modality, and computational challenges.

3.2.2 Modular and End-to-End Approaches

Modular systems decompose driving into interpretable components with explicit interfaces. Their advantages include easier testing, modular replacement, and direct diagnosis of component failures. However, errors may propagate between modules, and hand-designed interfaces may discard information useful to downstream tasks.

End-to-end systems learn a direct mapping from sensor observations to trajectories or control commands. They may optimize the complete driving objective and reduce manually specified intermediate representations, but generally require extensive data and are more difficult to interpret and validate. Hybrid systems preserve intermediate perception or planning representations while allowing joint learning across multiple stages [22].

3.3 Perception for Autonomous Driving

Perception converts sensor data into a representation of the environment suitable for prediction and planning. Its outputs may include object classes [164, 107, 113] and poses [129], lane boundaries [88, 143, 18, 8, 89, 8, 127], drivable space[144], semantic maps[157], depth [56], occupancy [21], object tracks [3], and motion estimates [124]. Perception is particularly important because downstream decisions can only be as reliable as the scene representation on which they operate.

3.3.1 Object Detection and Three-Dimensional Scene Understanding

Object detection estimates the class and spatial extent of relevant road users and obstacles. Camera-based methods typically predict two-dimensional bounding

boxes or infer three-dimensional properties from monocular or multi-view imagery [35, 20, 134, 52, 146]. LiDAR-based methods [5] operate on points, voxels, pillars, or range images. Multi-modal methods combine appearance and geometry.

Modern detectors are commonly divided into two-stage and one-stage architectures [94]. Two-stage methods [50] first generate candidate regions and then refine their classification and localization. One-stage methods [58, 97] predict objects directly from dense feature maps or queries. Transformer-based detectors [1, 84, 163] formulate detection as set prediction, reducing dependence on hand-designed components such as anchor boxes and non-maximum suppression.

Bird's-eye-view [75, 156, 145, 90, 72] representations have become central to multi-camera and multi-sensor perception because they provide a common metric space for detection, mapping, occupancy, and motion reasoning. Recent surveys distinguish geometry-based view transformation, depth-based lifting, attention-based transformation, and multi-modal fusion strategies.

Bird's-eye-view representations aggregate spatially distributed scene information into a common ground-plane coordinate system, which facilitates metric reasoning for vehicles, pedestrians, lanes, and other road elements.

3.3.2 Lane Detection

Lane detection estimates lane markings, lane boundaries, or lane center-lines from images [106, 162, 101, 85, 102, 110] or three-dimensional [35, 100, 42] observations. Early deep methods formulated the task as semantic segmentation [69]. More recent approaches use row-wise classification [82, 76, 104], anchors [123, 53, 141], keypoints [67, 131, 111], parametric curves [32, 109], graph representations [86, 103], or transformer queries [87]. Lane detection is difficult under shadows, worn markings, strong illumination changes, occlusion, merging and splitting lanes, and complex road topology [149].

Unlike generic object recognition, lane detection depends strongly on local geometry and elongated spatial structure [162]. A representation useful for lane detection must distinguish neighboring lane instances while preserving continuity across partially occluded regions (fig. 3.1).

This observation motivates local and structure-aware representation-learning approaches discussed in Section 3.7.

3.3.3 Tracking and Motion Forecasting

Multi-object tracking associates detections over time and estimates persistent agent states [29]. Tracking-by-detection systems commonly combine a motion model, an appearance or geometric affinity function, and a data-association

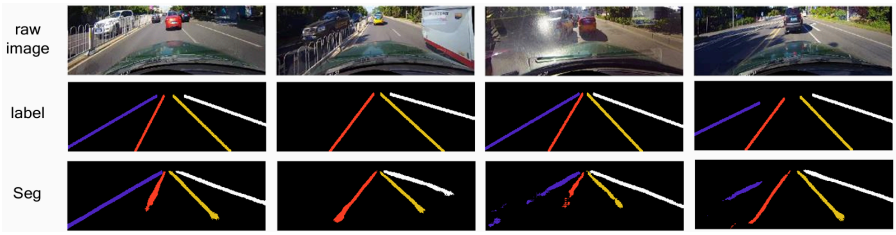


Figure 3.1. Examples illustrating the impact of vehicle occlusion on lane detection. The first row shows the input images, the second row the ground-truth lane annotations, and the third row the predicted segmentation results. Occluded lane markings often lead to fragmented or incomplete predictions, highlighting the importance of exploiting spatial context to infer lane continuity. Figure adapted from [106].

algorithm. Errors originate from missed detections, false positives, occlusions, and identity switches [24].

Motion forecasting [54] estimates multiple plausible future trajectories conditioned on agent history, road structure, and interactions. Classical approaches use kinematic models [117], while learning-based methods employ recurrent networks [66, 14], graph neural networks [71], attention and transformers [154]. Because future behavior is inherently uncertain, forecasting methods often predict multi-modal distributions rather than a single deterministic trajectory. Map context, social interaction, and temporal consistency are essential for forecasting manoeuvres such as yielding, merging, turning, and lane changing.

3.3.4 Perception Challenges

Autonomous-driving perception must operate under adverse weather, darkness, glare, occlusion, sensor failure, domain shift, and long-tail events [121]. It must also satisfy real-time latency and onboard-resource constraints. Increasing model capacity can improve benchmark accuracy but may violate deployment limits. Conversely, aggressive compression can damage the representations required for small objects, distant lanes, or rare classes. These competing objectives motivate efficient architecture design, architecture search, quantization, and transferable representation learning.

3.4 Neural Network Architectures

3.4.1 Convolutional Neural Networks

Convolutional neural networks [139] (fig. 3.2) exploit local connectivity and shared weights to learn translation-sensitive feature hierarchies. Early layers capture edges and texture, while deeper layers encode increasingly abstract patterns. Residual connections enable very deep networks by learning residual mappings and improving gradient propagation [48]. Feature pyramids [78]

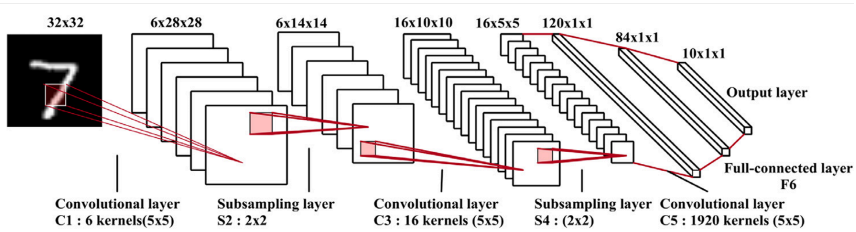


Figure 3.2. Architecture of LeNet-5, one of the earliest and most influential convolutional neural network (CNN) models [153]. The network consists of alternating convolutional and subsampling (pooling) layers followed by fully connected layers for classification. Figure adapted from [41].

combine multiple spatial resolutions and are widely used for detecting objects of different sizes. Encoder-decoder [148] architectures recover spatial resolution for segmentation and lane detection.

For an input tensor $\mathbf{X}$ and convolutional kernel $\mathbf{W}$, a convolutional feature at location (i, j) can be expressed as

$$\mathbf{Y}_{i,j} = \sum_{u,v} \mathbf{W}_{u,v} \mathbf{X}_{i+u,j+v}. \tag{3.1}$$

Although convolution provides a strong inductive bias for vision, its local receptive field may require many layers to model long-range interactions.

3.4.2 Temporal and Recurrent Models

Recurrent neural networks [93] maintain a hidden state that summarizes previous observations. Long short-term memory and gated recurrent units alleviate vanishing gradients and have been widely used for trajectories and temporal sensor features. Temporal convolution provides a parallel alternative, while explicit state-space and filtering models incorporate motion assumptions. Recurrent computation can become difficult to parallelize over long sequences, motivating attention-based temporal models.

3.4.3 Transformers and Attention

Transformers replace recurrent processing with attention over a set or sequence [130] (fig. 3.3). Given queries $\mathbf{Q}$, keys $\mathbf{K}$, and values $\mathbf{V}$, scaled dot-product attention is

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}. \tag{3.2}$$

Self-attention captures interactions within one representation, whereas cross-attention transfers information between representations, such as image features and object queries. Vision Transformers apply attention to image patches, and

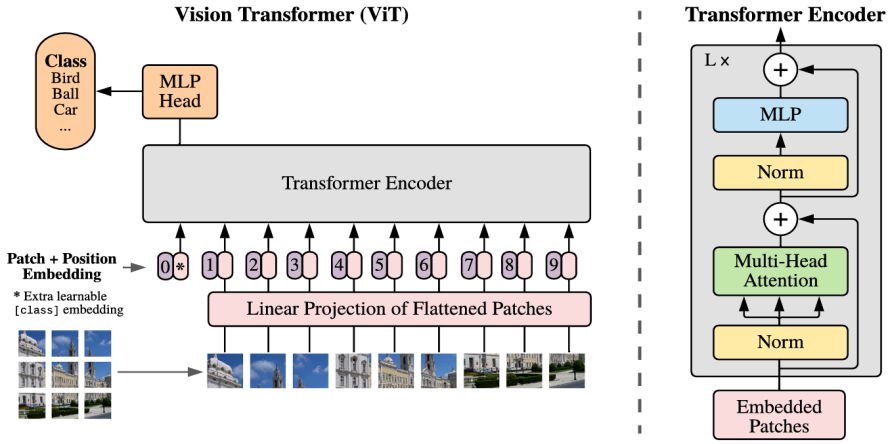


Figure 3.3. Architecture of the Vision Transformer (ViT), proposed by Dosovitskiy *et al.* [27]. Unlike convolutional neural networks, ViT partitions an input image into a sequence of fixed-size patches, linearly projects each patch into an embedding space, and augments them with positional embeddings before processing them through a stack of Transformer encoder blocks. Each encoder consists of multi-head self-attention (MHSA), multi-layer perceptrons (MLPs), layer normalization, and residual connections. A learnable classification token aggregates the global image representation for image classification. ViT demonstrated that transformer-based architectures can achieve state-of-the-art performance in computer vision when trained on sufficiently large datasets. Figure adapted from Dosovitskiy *et al.* [27].

detection transformers use learned queries to predict object sets. In autonomous driving, transformers are used for multi-view fusion, bird's-eye-view construction, temporal aggregation, tracking, mapping, and trajectory prediction.

3.4.4 Computational Cost

Model complexity is commonly described using parameter count and floating-point operations. These measures are useful but do not fully predict deployment performance. Runtime also depends on memory access, operator implementation, parallelism, tensor dimensions, hardware utilization, numerical precision, and batch size. A model with fewer operations can be slower than a larger model if its operators are poorly supported by the target accelerator. Architecture design for autonomous vehicles should therefore consider measured latency, memory consumption, throughput, and energy in addition to predictive accuracy.

3.5 Neural Architecture Search

Neural architecture search automates the design of network architectures. A NAS method is commonly characterized by its search space, search strategy,

and performance-estimation strategy [30]. The search space determines which architectures can be represented; the search strategy selects candidate architectures; and the performance-estimation method approximates the quality of each candidate without fully training every model.

3.5.1 Search Spaces

A search space may vary network depth, width, operator type, kernel size, expansion ratio, resolution, skip connections, feature-fusion topology, or task-specific heads. Cell-based spaces search for a repeated computational motif, whereas macro-level spaces optimize the complete network. For autonomous-driving tasks, search spaces may also include temporal fusion, multi-modal branches, detection and forecasting heads, or precision assignments.

3.5.2 Search Strategies

Early NAS methods used reinforcement learning or evolutionary algorithms to optimize discrete architectures. Random search remains an important baseline. Bayesian optimization uses a surrogate model to guide evaluation toward promising regions. Differentiable NAS relaxes discrete operator choices into continuous architecture parameters and optimizes them using gradient descent, substantially reducing search cost [79]. However, weight sharing and continuous relaxation can introduce ranking inconsistency, optimization bias, and instability.

Multi-objective NAS optimizes conflicting criteria such as accuracy, latency, energy, and memory. The problem may be expressed as

$$\max_{\alpha \in \mathscr{A}} \quad [\mathrm{Acc}(\alpha), -\mathrm{Lat}(\alpha), -\mathrm{Mem}(\alpha)], \tag{3.3}$$

or as constrained optimization:

$$\max_{\alpha \in \mathscr{A}} \mathrm{Acc}(\alpha) \quad \text{s.t.} \quad \mathrm{Lat}(\alpha) \leq L_{\max}, \quad \mathrm{Mem}(\alpha) \leq M_{\max}. \tag{3.4}$$

The result is often a Pareto set rather than a single architecture.

3.5.3 Performance Estimation and Weight Sharing

Training every candidate to convergence is usually prohibitive. Low-fidelity evaluation uses fewer epochs, reduced data, lower input resolution, or proxy tasks. Predictor-based methods learn a surrogate for architecture accuracy. One-shot methods train a supernetwork whose weights are shared by many subnetworks. Once-for-all networks further train an elastic network that can be specialized for different devices without retraining the entire search space [13].

Weight sharing reduces cost but may make candidate rankings dependent on interference within the supernetwork. Consequently, final candidates should be retrained or carefully re-evaluated under the target deployment conditions.

3.5.4 Hardware-Aware and Joint Search

Hardware-aware NAS directly includes measured or predicted device latency rather than relying solely on operation count. ProxylessNAS searches architectures directly on target tasks and hardware [12]. Joint optimization may additionally search pruning ratios and numerical precision. APQ, for example, combines architecture, pruning, and quantization-policy search within a hardware-aware framework [135]. This is particularly relevant to autonomous vehicles, where perception and forecasting must satisfy hard latency and resource limits.

3.6 Neural Network Quantization

Quantization maps high-precision weights and activations to a smaller set of discrete values. It reduces storage, memory traffic, and potentially inference latency and energy when the target hardware provides efficient low-precision arithmetic.

3.6.1 Uniform Quantization

For a real-valued tensor x, affine uniform quantization can be written as

$$q(x) = \text{clip}\left(\left\lfloor \frac{x}{s} \right\rceil + z, q_{\min}, q_{\max}\right), \tag{3.5}$$

where s is the scale, z is the zero point, and $[q_{\min}, q_{\max}]$ is determined by the bit width. Dequantization gives

$$\hat{x} = s\,(q(x) - z)\,. \tag{3.6}$$

Symmetric quantization typically fixes $z = 0$, while asymmetric quantization allows a non-zero offset. Quantization may be applied per tensor, per channel, per group, or with mixed precision across layers.

3.6.2 Quantization-Aware Training and Post-Training Quantization

Quantization-aware training simulates low-precision operations during optimization and adapts the weights to quantization noise. The non-differentiable

rounding operation is commonly approximated by a straight-through estimator. Quantization-aware training can achieve high accuracy but requires access to training data and substantial retraining.

Post-training quantization converts a pretrained model using a small calibration set and little or no task-level retraining. It is attractive when original training is expensive or data are unavailable. Calibration estimates activation ranges and quantization parameters. At very low precision, nearest rounding and independent layer calibration become insufficient because errors accumulate and change downstream feature distributions.

AdaRound formulates weight rounding as a task-aware optimization rather than always rounding to the nearest grid point [98]. BRECQ improves post-training quantization through block reconstruction [74], while QDrop improves low-bit calibration by randomly dropping activation quantization during reconstruction [137]. PD-Quant introduces a prediction-difference metric to better align quantized and full-precision outputs [81]. These methods show the progression from local parameter matching toward output- and task-aware reconstruction.

3.6.3 Sources of Quantization Error

Quantization error depends on tensor range, outliers, bit width, distribution shape, layer sensitivity, nonlinearities, and interactions between layers. Weight quantization perturbs learned filters, while activation quantization changes sample-dependent intermediate representations. Low-bit errors may alter class rankings even when average feature reconstruction error is small. This distinction is important for post-training correction methods: reducing global mean-squared error does not necessarily repair structured, class-dependent, or cluster-dependent prediction errors.

Mixed-precision quantization assigns higher precision to sensitive layers and lower precision to robust layers. Precision selection can be formulated as an optimization or architecture-search problem. Quantization should ultimately be evaluated using actual hardware because theoretical reductions in bit operations do not guarantee latency improvements when low-precision kernels are unavailable or conversion overhead is substantial.

3.7 Representation Learning

Representation learning aims to transform raw observations into features that preserve task-relevant information and suppress irrelevant variation. In autonomous driving, transferable representations are required to cope with limited labels, changes in weather and illumination, geographical shift, novel objects, and differences between sensor configurations.

3.7.1 Supervised and Self-Supervised Learning

Supervised learning optimizes representations using task labels. Although effective, dense annotation for lanes, objects, depth, tracks, and trajectories is expensive. Learned features may also overfit dataset-specific biases.

Self-supervised learning obtains supervisory signals from the structure of unlabelled data. Major families include reconstruction, masked prediction, temporal prediction, contrastive learning, clustering, and teacher–student consistency. These approaches enable pretraining on large unlabelled driving datasets before fine-tuning for a downstream task.

3.7.2 Contrastive Learning

Contrastive learning encourages representations of related views to be similar and unrelated samples to be distinguishable. For a positive pair (i, j), the normalized temperature-scaled cross-entropy loss used by SimCLR can be written as [19]

$$\ell_{i,j} = -\log \frac{\exp\left(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau\right)}{\sum_{k \neq i} \exp\left(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau\right)}, \tag{3.7}$$

where τ is a temperature parameter and sim is commonly cosine similarity. Momentum Contrast maintains a queue of negative keys and a momentum-updated encoder, enabling a large and consistent dictionary [49]. BYOL shows that strong representations can also be learned without explicit negative pairs using online and target networks [40].

3.7.3 Global, Local, and Dense Representations

Image-level self-supervision is effective for classification but may not preserve the spatial detail required for detection, segmentation, and lane estimation. Dense prediction requires correspondence at the pixel, patch, region, or structural level. DenseCL extends contrastive learning to local features by matching dense correspondences [136], while PixPro learns pixel-level representations through propagation between neighbouring features [142].

Autonomous-driving scenes contain repeated local patterns, elongated lane structures, small distant objects, and strong geometric relations. Therefore, positive-pair construction and similarity modelling should reflect the downstream structure rather than relying exclusively on global image identity. For lane detection, useful pretraining should preserve lane continuity, distinguish neighbouring lanes, and remain robust to shadows, occlusion, and appearance changes.

3.7.4 Transferability and Generalization

Representation quality is evaluated through fine-tuning, linear probing, few-shot learning, cross-domain transfer, and robustness tests. A representation that performs well only on the pretraining distribution may be unsuitable for autonomous driving. Cross-dataset evaluation and difficult-condition subsets are particularly important because benchmark averages can hide failures under night-time, shadow, rain, or rare road geometry.

Representation learning is also related to efficiency. Strong pretrained features can reduce labelled-data requirements and may allow smaller downstream models to achieve competitive performance. However, compression and quantization can distort these representations. Efficient autonomous-driving models therefore require joint consideration of representation quality, architecture, and numerical precision.

3.8 Chapter Summary

Human driving is affected by distraction, fatigue, impairment, and heterogeneous decision-making performance. Autonomous driving can reduce some of these failure mechanisms, but current evidence remains conditional on the operational domain, automation level, and comparison methodology. The safety of an autonomous vehicle ultimately depends on a complete system that accurately perceives the environment, predicts dynamic agents, and executes decisions under real-time constraints.

Perception models increasingly rely on convolutional, graph-based, and transformer architectures. Their growing computational demands motivate neural architecture search and quantization, while the scarcity and diversity of driving data motivate self-supervised and contrastive representation learning. These foundations are closely connected: architecture search determines how information is processed, quantization determines how accurately it is represented numerically, and representation learning determines whether the extracted features generalize across tasks and environments. The following chapters build on these foundations to present the contributions of this thesis.

4. Thesis Contributions

This thesis presents five main contributions toward the development of efficient and generalizable learning-based methods for autonomous driving applications. The contributions are directly connected to the research questions and subgoals defined in chapter 2. In particular, **Subgoal 1** concerns the identification and analysis of accuracy, inference-latency, and memory-footprint bottlenecks; **Subgoal 2** addresses these bottlenecks through efficient and fast-converging Neural Architecture Search (NAS); **Subgoal 3** addresses the limited availability of labeled data through self-supervised learning; and **Subgoal 4** investigates quantization and model compression for efficient deployment. The following contributions address these subgoals at different stages of model design, training, and deployment.

Contribution 1: 3DLaneNAS for Accurate and Lightweight 3D Lane Detection

We proposed 3DLaneNAS, a multi-objective NAS approach for monocular 3D lane detection that searches for architectures considering both prediction performance and computational efficiency. The contribution investigates the trade-off between accuracy and inference latency in 3D lane detection and uses this analysis to guide architecture optimization for resource-constrained autonomous driving systems.

- **Reference Paper:** [Paper A – *3DLaneNAS: Neural Architecture Search for Accurate and Light-Weight 3D Lane Detection*]
- **Subgoals Addressed: Subgoal 1 and Subgoal 2.**
- **Relation to Subgoals:** 3DLaneNAS contributes to **Subgoal 1** by investigating the accuracy and inference-latency trade-offs associated with different architectural choices in monocular 3D lane detection. It addresses **Subgoal 2** by applying multi-objective NAS to automatically identify architectures that improve this trade-off rather than relying on manually designed network structures.
- **Achievements:** Yielded a minimum of 5.2% higher accuracy and approximately 1.33$\times$ lower latency on the evaluated benchmarks.

Contribution 2: CLLD for Self-Supervised Lane Detection

We developed Contrastive Learning for Lane Detection via Cross-Similarity (CLLD), a self-supervised representation-learning approach designed specifically for lane detection. CLLD exploits unlabeled data to learn representations

that can subsequently be transferred to the downstream lane-detection task, reducing reliance on large amounts of manually annotated training data.

- **Reference Paper:** [Paper C – *Contrastive Learning for Lane Detection via Cross-Similarity*]
- **Subgoal Addressed: Subgoal 3.**
- **Relation to Subgoal:** CLLD directly addresses **Subgoal 3** by developing a self-supervised learning strategy adapted to the characteristics of lane detection. It investigates how unlabeled driving data can be used to learn useful representations before supervised fine-tuning, thereby reducing dependency on task-specific labeled data while improving generalization under challenging driving conditions.
- **Achievements:** Demonstrated improved lane-detection performance, particularly under challenging visibility conditions, while reducing dependency on labeled training data.

Contribution 3: FastStereoNet for Efficient Disparity Estimation

We introduced FastStereoNet, an efficient NAS framework for disparity estimation on resource-limited computing platforms. The contribution considers the computational characteristics of stereo-matching architectures and develops an accelerated architecture-search strategy for identifying compact models suitable for deployment on constrained hardware.

- **Reference Paper:** [Paper B – *FastStereoNet: A Fast Neural Architecture Search for Improving the Inference of Disparity Estimation on Resource-Limited Platforms*]
- **Subgoals Addressed: Subgoal 1 and Subgoal 2.**
- **Relation to Subgoals:** FastStereoNet contributes to **Subgoal 1** by investigating computational and model-size limitations of disparity-estimation networks when deployed on resource-constrained hardware. It addresses **Subgoal 2** by introducing an efficient NAS strategy that reduces architecture-search cost while identifying compact architectures with competitive predictive performance. This contribution therefore addresses both the efficiency of the resulting architecture and the computational cost of the NAS process itself.
- **Achievements:** Provided a $5.25\times$ reduction in search time and a $44.4\times$ reduction in model size while maintaining competitive accuracy.

Contribution 4: TrajectoryNAS for Efficient Trajectory Prediction

We introduced TrajectoryNAS, a NAS approach for trajectory prediction in autonomous driving systems. Using LiDAR point-cloud data, TrajectoryNAS searches for architectures that jointly consider prediction accuracy and in-

ference latency, targeting the requirements of real-time autonomous driving applications.

- **Reference Paper:** [Paper D – *TrajectoryNAS: A Neural Architecture Search for Trajectory Prediction*]
- **Subgoals Addressed: Subgoal 1 and Subgoal 2.**
- **Relation to Subgoals:** TrajectoryNAS contributes to **Subgoal 1** by examining the trade-off between prediction accuracy and inference latency in trajectory-prediction architectures. It addresses **Subgoal 2** by incorporating these performance criteria into the NAS process and automatically searching for architectures that provide a more favorable balance between predictive performance and computational requirements.
- **Achievements:** Achieved a minimum of 4.8% higher accuracy and approximately $1.1\times$ lower latency compared with the evaluated approaches.

Contribution 5: Post-Training Quantization Error reduction

We proposed CAT, a post-training quantization error-reduction approach based on cluster-wise affine transformations. CAT targets the accuracy degradation introduced by low-bit quantization while preserving the memory and computational advantages of quantized neural networks. Unlike architecture-level optimization, this contribution operates at the deployment stage and improves the efficiency of already-trained models without requiring full retraining.

- **Reference Paper:** [Paper E – *CAT: Post-Training Quantization Error reduction via Cluster-based Affine Transformation*]
- **Subgoal Addressed: Subgoal 4.**
- **Relation to Subgoal:** CAT directly addresses **Subgoal 4** by investigating post-training quantization as a mechanism for reducing the memory footprint and computational requirements of neural networks. In particular, CAT addresses the accuracy degradation associated with aggressive low-bit quantization by restoring the relationship between quantized and full-precision outputs through cluster-wise affine transformations. It therefore contributes to the deployment of compressed models under memory, computation, and real-time constraints.
- **Achievements:** Demonstrated that quantization-induced accuracy degradation can be mitigated with a lightweight post-training correction mechanism while retaining the efficiency benefits of low-bit models.

4.1 Contributions and Research Subgoal Relationship

The five contributions collectively address the research subgoals at different stages of the model-development and deployment pipeline. Contributions 1, 3, and 4 address **Subgoal 1** by investigating accuracy, latency, and/or memory-related limitations across different autonomous driving tasks, including 3D

lane detection, disparity estimation, and trajectory prediction. These analyses provide the basis for addressing **Subgoal 2**, where the same contributions employ NAS to automatically identify architectures with improved trade-offs between predictive performance and computational requirements. Contribution 2 addresses **Subgoal 3** from a complementary perspective by improving data efficiency through self-supervised representation learning. Finally, Contribution 5 addresses **Subgoal 4** at the deployment stage through post-training quantization and error reduction.

Consequently, the contributions follow the overall research direction established by the research questions and subgoals: first, identifying the principal *accuracy, latency, and memory-footprint bottlenecks*; second, addressing architecture-related bottlenecks through *NAS*; third, addressing the labeled-data limitation through *self-supervised learning*; and finally, reducing deployment-related computational and memory requirements through *quantization*. Table 4.1 summarizes this relationship.

Table 4.1. *Relationship between thesis contributions, papers, and research subgoals.*

	C1	**C2**	**C3**	**C4**	**C5**
Subgoal 1: Bottleneck Analysis	✓		✓	✓	
Subgoal 2: Efficient NAS	✓		✓	✓	
Subgoal 3: Self-Supervised Learning		✓			
Subgoal 4: Quantization and Compression					✓
Paper	A	C	B	D	E

The mapping illustrates that the contributions are complementary rather than independent. Contributions 1, 3, and 4 address architectural efficiency across different autonomous driving tasks; Contribution 2 addresses representation and data efficiency during training; and Contribution 5 addresses model efficiency during deployment. Together, they address the thesis's overall objective of developing accurate, computationally efficient, and generalizable learning-based methods suitable for real-time autonomous driving systems.

5. Discussion, Conclusion and Future Work

"It is through science that we prove, but through intuition that we discover."

Henri Poincare

5.1 Discussion

This thesis investigated the development of efficient and generalizable learning-based methods for autonomous driving, with particular emphasis on the trade-offs between accuracy, inference latency, memory footprint, search cost, and dependence on labeled data. Rather than addressing these challenges through a single optimization technique, the thesis considered them at different stages of the model-development pipeline, including architecture design, representation learning, and model deployment.

The first research direction focused on computationally efficient architecture design through Neural Architecture Search (NAS). The studies on 3D lane detection, stereo disparity estimation, and trajectory prediction demonstrated that architecture design can be formulated as a multi-objective optimization problem in which predictive performance must be considered together with computational requirements. The proposed approaches showed that automatically searched architectures can provide improved trade-offs between accuracy and efficiency compared with manually designed alternatives. These findings directly address the thesis subgoals related to identifying computational bottlenecks and designing efficient NAS strategies for autonomous driving applications.

An important observation across these studies is that computational bottlenecks are application-dependent. In 3D lane detection, the central challenge is to maintain accurate geometric estimation while limiting inference latency and architectural complexity. In stereo disparity estimation, model size and search cost become particularly important for deployment on resource-constrained hardware. In trajectory prediction, the latency–accuracy trade-off becomes critical because prediction is part of a time-sensitive decision-making pipeline. Therefore, a single architecture or optimization objective is unlikely to be optimal for all autonomous driving tasks. Instead, hardware constraints and application-specific performance requirements should be incorporated directly into the design process.

The second research direction addressed another fundamental limitation of modern perception systems: their dependence on large amounts of labeled training data. The proposed Contrastive Learning for Lane Detection via Cross-Similarity (CLLD) method demonstrated how self-supervised representation learning can exploit unlabeled driving data before downstream supervised training. This is particularly relevant for autonomous driving, where sensor

data can be collected at large scale, while obtaining high-quality annotations for tasks such as lane detection is considerably more expensive and time-consuming. The results suggest that self-supervised learning can complement architecture optimization by improving the quality and generalization of learned representations rather than focusing solely on reducing computational cost.

The third research direction considered efficiency after model training through quantization. The proposed CAT method investigates post-training reduction of quantization-induced errors through cluster-based affine transformations. This extends the efficiency considerations of the thesis from architecture-level optimization to deployment-level optimization. While NAS determines which architecture should be executed, quantization determines how efficiently the parameters and activations of the selected model can be represented and processed. These two directions are therefore complementary and can potentially be combined within a unified hardware-aware optimization framework.

Taken together, the contributions show that efficiency in autonomous driving cannot be reduced to a single metric such as latency or model size. A practical autonomous driving model must simultaneously satisfy several constraints, including sufficient predictive accuracy, low response time, manageable memory consumption, efficient training and architecture search, and robustness to variations in the available data. Improving one of these dimensions in isolation can lead to degradation in another. The broader challenge is therefore to identify suitable trade-offs between these objectives according to the requirements of the target application and hardware platform.

5.2 Limitations

The contributions of this thesis should be interpreted in the context of several limitations. First, the proposed methods are evaluated primarily at the level of individual perception and prediction tasks rather than within a complete autonomous-driving stack. Consequently, improvements in task-level accuracy or inference efficiency do not directly establish system-level safety improvements. Second, latency and computational efficiency depend on the target hardware and software implementation, and measurements obtained on the evaluated platforms may not transfer directly to other deployment environments. Third, NAS, self-supervised learning, and quantization are investigated as complementary but largely separate research directions; their joint optimization is not experimentally evaluated in this thesis. Finally, the reported results depend on the datasets and evaluation protocols used in the individual studies, and further cross-domain and system-level validation would be required to establish broader generalizability.

5.3 Conclusion

The overall objective of this thesis was to investigate methods for developing accurate, computationally efficient, and generalizable learning-based models for autonomous driving applications. This objective was addressed through five contributions covering architecture search, self-supervised representation learning, and post-training quantization.

The NAS-based contributions demonstrated how automated architecture design can be applied across multiple autonomous driving tasks. 3DLaneNAS investigated lightweight and accurate architectures for monocular 3D lane detection, FastStereoNet addressed efficient disparity estimation and reduced architecture-search cost for resource-constrained platforms, and TrajectoryNAS extended multi-objective architecture optimization to trajectory prediction. Collectively, these studies addressed the thesis subgoals concerning the analysis of accuracy, latency, and memory-related bottlenecks and the development of efficient NAS methods to mitigate them.

CLLD addressed the data-efficiency objective of the thesis by introducing a self-supervised learning strategy tailored to lane detection. The study demonstrated that task-specific representation learning can exploit unlabeled data to improve downstream perception and reduce the dependence on extensive manual annotation.

Finally, CAT extended the thesis toward deployment-oriented model compression by addressing the accuracy degradation introduced by aggressive post-training quantization. This contribution complements the architecture-search studies by showing that efficiency can also be improved after architecture design and training.

Overall, the findings of this thesis support the view that efficient autonomous driving perception should be approached as a joint optimization problem involving architecture, representation, and numerical precision. The contributions provide different mechanisms for addressing these dimensions and together establish a foundation for the design of perception systems that better balance predictive performance with computational and data constraints.

5.4 Future Work

Although the contributions of this thesis address several important aspects of efficient autonomous driving models, they also reveal a number of opportunities for further research.

One particularly important direction is the efficient deployment of Vision Transformers and related attention-based architectures. Transformer-based models have demonstrated strong representation capabilities because of their ability to model long-range and global dependencies. However, their computational complexity, memory consumption, and inference latency remain

important limitations for real-time autonomous driving, particularly on embedded and edge hardware.

A natural extension of the NAS contributions in this thesis is therefore to develop hardware-aware search spaces specifically for transformer-based perception models. Instead of searching only over conventional architectural components, future NAS frameworks could jointly optimize transformer depth, embedding dimensions, attention mechanisms, token reduction strategies, and numerical precision. Such approaches could search directly for architectures that satisfy explicit latency and memory constraints on target automotive hardware.

Quantization represents another promising direction for improving transformer efficiency. The CAT framework presented in this thesis can motivate further investigation into post-training correction mechanisms for transformer architectures, where quantization errors may affect attention operations, activation distributions, and layer normalization differently from conventional convolutional networks. Extending cluster-based correction strategies to mixed-precision and hardware-aware quantization could enable more aggressive compression while preserving predictive performance.

Another promising research direction is the joint optimization of architecture search and quantization. In this thesis, NAS and quantization were studied as complementary but largely separate optimization stages. Future work could integrate them into a unified search process in which architectural choices and numerical precision are optimized simultaneously. Such a framework could search not only for which operations and connections should be used, but also for the appropriate bit-width of individual layers or components according to their sensitivity and computational cost.

The interaction between self-supervised learning and efficient architecture design also deserves further investigation. Self-supervised pretraining is commonly performed using large models and substantial computational resources. Future research could explore whether NAS can identify architectures that are specifically efficient for self-supervised pretraining and downstream transfer. Similarly, the representations learned through self-supervision could potentially guide architecture search by identifying which components are most important for preserving transferable information.

A further extension would be to evaluate the proposed optimization strategies within a more complete autonomous driving pipeline. The contributions in this thesis focus on individual perception and prediction tasks, including lane detection, disparity estimation, and trajectory prediction. In practical autonomous driving systems, however, these tasks operate as interconnected components. Future work could investigate system-level optimization in which the computational budget is distributed dynamically across perception, prediction, and planning modules according to scene complexity and driving conditions.

Finally, future research should increasingly consider adaptability at runtime. Instead of deploying a single static model, an autonomous driving system

could dynamically modify its computational behavior depending on available hardware resources, environmental complexity, and uncertainty. For example, the system could activate a larger architecture or higher numerical precision in difficult scenes while using a more efficient configuration under simpler conditions. Combining adaptive architectures, self-supervised representations, and mixed-precision inference could ultimately lead to autonomous driving systems that are not only efficient by design, but also capable of adapting their computational resources to changing real-world conditions.

In summary, the work presented in this thesis provides a foundation for efficient model design across architecture, representation learning, and deployment. The next step is to integrate these dimensions more closely, moving from individually optimized models toward adaptive and hardware-aware learning systems that jointly optimize accuracy, latency, memory usage, data efficiency, and computational cost for real-time autonomous driving.

References

[1] Nicolas "Carion et al. “"End-to-End Object Detection with Transformers"”. In: *"Computer Vision – ECCV 2020"*. "Cham": "Springer International Publishing", "2020", "213–229". ISBN: "978-3-030-58452-8".

[2] Stefan Abrecht et al. “Deep Learning Safety Concerns in Automated Driving Perception”. In: *arXiv preprint arXiv:2309.03774* (2023).

[3] Momir Adžemović. “Deep Learning-Based Multi-Object Tracking: A Comprehensive Survey from Foundations to State-of-the-Art”. In: *The Visual Computer* 42.9 (2026), p. 368. ISSN: 1432-2315. DOI: 10.1007/s00371-026-04607-7. URL: https://doi.org/10.1007/s00371-026-04607-7.

[4] Dosovitskiy Alexey. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv: 2010.11929* (2020).

[5] Nang Htet Htet Aung et al. “A Review of LiDAR-Based 3D Object Detection via Deep Learning Approaches Towards Robust Connected and Autonomous Vehicles”. In: *IEEE Transactions on Intelligent Vehicles* 10.1 (2025), pp. 526–547. DOI: 10.1109/TIV.2024.3415771.

[6] Claudine Badue, Rânik Guidolini, Raphael Vivacqua Carneiro, et al. “Self-Driving Cars: A Survey”. In: *Expert Systems with Applications* 165 (2021), p. 113816. DOI: 10.1016/j.eswa.2020.113816.

[7] Hichem Belgacem et al. “Multi-Sensor Data Fusion: An Overview of Concepts, Methods, and Classifications”. In: *IEEE Access* 14 (2026), pp. 41202–41229. DOI: 10.1109/ACCESS.2026.3673744.

[8] Jiping Bi et al. “Lane Detection for Autonomous Driving: Comprehensive Reviews, Current Challenges, and Future Predictions”. In: *IEEE Transactions on Intelligent Transportation Systems* 26.5 (2025), pp. 5710–5746. DOI: 10.1109/TITS.2024.3524603.

[9] Sarah-Jayne Blakemore and Suparna Choudhury. “Development of the adolescent brain: implications for executive function and social cognition”. In: *Journal of Child Psychology and Psychiatry* 47.3-4 (2006), pp. 296–312. DOI: https://doi.org/10.1111/j.1469-7610.2006.01611.x. eprint: https://acamh.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-7610.2006.01611.x.

[10] Katharina Blatter and Christian Cajochen. “Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings”. In: *Physiology and Behavior* 90.2 (2007). Includes a Special Section on Chronobiology Aspects of the Sleep–Wake Cycle and Thermoreregulation, pp. 196–208. ISSN: 0031-9384. DOI: https://doi.org/10.1016/j.physbeh.2006.09.009.

[11] Daniel Bogdoll et al. “Taxonomy and Survey on Remote Human Input Systems for Driving Automation Systems”. In: *Advances in Information and Communication*. Ed. by Kohei Arai. Springer International Publishing, 2022, pp. 94–108.

[12] Han Cai, Ligeng Zhu, and Song Han. “ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware”. In: *International Conference on Learning Representations*. 2019.

[13] Han Cai et al. “Once-for-All: Train One Network and Specialize It for Efficient Deployment”. In: *International Conference on Learning Representations*. 2020.

[14] Samuele Capobianco et al. “Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks”. In: *IEEE Transactions on Aerospace and Electronic Systems* (2021).

[15] Athanasios Chalvatzaras, Ioannis Pratikakis, and Angelos A. Amanatiadis. “A Survey on Map-Based Localization Techniques for Autonomous Vehicles”. In: *IEEE Transactions on Intelligent Vehicles* 8.2 (2023), pp. 1574–1596. DOI: 10.1109/TIV.2022.3192102.

[16] Changhao Chen and Xianfei Pan. “Deep Learning for Inertial Positioning: A Survey”. In: *IEEE Transactions on Intelligent Transportation Systems* 25.9 (2024), pp. 10506–10523. DOI: 10.1109/TITS.2024.3381161.

[17] Changhao Chen et al. “Deep Learning for Visual Localization and Mapping: A Survey”. In: *IEEE Transactions on Neural Networks and Learning Systems* 35.12 (2024), pp. 17000–17020. DOI: 10.1109/TNNLS.2023.3309809.

[18] Li Chen et al. “PersFormer: 3D Lane Detection via Perspective Transformer and the OpenLane Benchmark”. In: *Computer Vision – ECCV 2022*. Ed. by Shai Avidan et al. Springer Nature Switzerland, 2022, pp. 550–567. DOI: 10.1007/978-3-031-19839-7_32.

[19] Ting Chen et al. “A Simple Framework for Contrastive Learning of Visual Representations”. In: *Proceedings of the 37th International Conference on Machine Learning*. Ed. by Hal Daumé III and Aarti Singh. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 1597–1607. URL: https://proceedings.mlr.press/v119/chen20j.html.

[20] Xiaozhi Chen et al. “Monocular 3D Object Detection for Autonomous Driving”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[21] Xuesong Chen et al. “M3Net: Multimodal Multi-task Learning for 3D Detection, Segmentation, and Occupancy Prediction in Autonomous Driving”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 39. 2. Association for the Advancement of Artificial Intelligence (AAAI), 2025, pp. 2275–2283. DOI: 10.1609/aaai.v39i2.32227. URL: https://doi.org/10.1609/aaai.v39i2.32227.

[22] Pranav Singh Chib and Pravendra Singh. “Recent Advancements in End-to-End Autonomous Driving Using Deep Learning: A Survey”. In: *IEEE Transactions on Intelligent Vehicles* 9.1 (2024), pp. 103–118. DOI: 10.1109/TIV.2023.3318070.

[23] Michelle E. Costanzo et al. “Neural Efficiency in Expert Cognitive-Motor Performers During Affective Challenge”. In: *Journal of Motor Behavior* 48.6 (2016). PMID: 27715496, pp. 573–588. DOI: 10.1080/00222895.2016.1161591.

[24] “Data association in multiple object tracking: A survey of recent techniques”. In: *Expert Systems with Applications* 192 (2022), p. 116300.

[25] Frédéric Dehais et al. “A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance”. In: *Frontiers in Neuroscience* Volume 14 - 2020 (2020). ISSN: 1662-453X. DOI: 10.3389/fnins.2020.00268. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2020.00268.

[26] Veronica Diveica. “How the brain constructs and uses meaning: Semantic cognition”. In: *Nature Reviews Psychology* 4.11 (2025), pp. 682–682.

[27] Alexey Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020).

[28] Mariella Dreissig et al. “Survey on LiDAR Perception in Adverse Weather Conditions”. In: *2023 IEEE Intelligent Vehicles Symposium (IV)*. 2023, pp. 1–8. DOI: 10.1109/IV55152.2023.10186539.

[29] Chenjie Du et al. “Exploring the State-of-the-Art in Multi-Object Tracking: A Comprehensive Survey, Evaluation, Challenges, and Future Directions”. In: *Multimedia Tools and Applications* 83.29 (2024), pp. 73151–73189. DOI: 10.1007/s11042-023-17983-2. URL: https://doi.org/10.1007/s11042-023-17983-2.

[30] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. “Neural Architecture Search: A Survey”. In: *Journal of Machine Learning Research* 20.55 (2019), pp. 1–21.

[31] Tai Fei, Andreas Becker, and Markus Gardill. “Interference Mitigation in Automotive Radar Systems: A Current State Survey and Future Trends”. In: *IEEE Transactions on Intelligent Transportation Systems* 27.1 (2026), pp. 53–80. DOI: 10.1109/TITS.2025.3629041.

[32] Zhengyang Feng et al. “Rethinking Efficient Lane Detection via Curve Modeling”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 17062–17070.

[33] Graham D. Finlayson. “Colour and Illumination in Computer Vision”. In: *Interface Focus* 8.4 (2018), p. 20180008. ISSN: 2042-8898. DOI: 10.1098/rsfs.2018.0008. URL: https://doi.org/10.1098/rsfs.2018.0008.

[34] N. Gaoua et al. “Sensory displeasure reduces complex cognitive performance in the heat”. In: *Journal of Environmental Psychology* 32.2 (2012), pp. 158–163. ISSN: 0272-4944. DOI: https://doi.org/10.1016/j.jenvp.2012.01.002.

[35] Nicholas Garnett et al. “3D-LaneNet: End-to-end 3D multiple lane detection”. In: *ICCV*. 2019.

[36] Swarnendu Ghosh et al. “Understanding Deep Learning Techniques for Image Segmentation”. In: *ACM Comput. Surv.* 52.4 (Aug. 2019). ISSN: 0360-0300. DOI: 10.1145/3329784. URL: https://doi.org/10.1145/3329784.

[37] E. Leigh Gibson and M. W. Green. “Nutritional Influences on Cognitive Function: Mechanisms of Susceptibility”. In: *Nutrition Research Reviews* 15.1 (2002), pp. 169–206. DOI: 10.1079/NRR200131.

[38] Ian Goodfellow. *Deep learning*. 2016.

[39] Sorin Grigorescu et al. “A Survey of Deep Learning Techniques for Autonomous Driving”. In: *Journal of Field Robotics* 37.3 (2020), pp. 362–386. DOI: 10.1002/rob.21918.

[40] Jean-Bastien Grill et al. “Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 21271–21284.

[41] Jiuxiang Gu et al. “Recent advances in convolutional neural networks”. In: *Pattern recognition* 77 (2018), pp. 354–377.

[42] Yuliang Guo et al. “Gen-LaneNet: A Generalized and Scalable Approach for 3D Lane Detection”. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer, 2020, pp. 666–681.

[43] Yuqing Guo et al. “A Survey of Trajectory Planning Methods for Autonomous Driving—Part I: Unstructured Scenarios”. In: *IEEE Transactions on Intelligent Vehicles* 9.9 (2024), pp. 5407–5434. DOI: 10.1109/TIV.2023.3337318.

[44] Steffen Hagedorn et al. “The Integration of Prediction and Planning in Deep Learning Automated Driving Systems: A Review”. In: *IEEE Transactions on Intelligent Vehicles* 10.5 (2025), pp. 3626–3643. DOI: 10.1109/TIV.2024.3459071.

[45] Elizabeth Hampson and Doreen Kimura. “Reciprocal Effects of Hormonal Fluctuations on Human Motor and Perceptual-Spatial Skills”. In: *Behavioral Neuroscience* 102.3 (1988), pp. 456–459. DOI: 10.1037/0735-7044.102.3.456.

[46] Kaiming He et al. “Momentum Contrast for Unsupervised Visual Representation Learning”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 9729–9738.

[47] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[48] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.

[49] Kaiming He et al. “Momentum Contrast for Unsupervised Visual Representation Learning”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 9729–9738.

[50] O. Hmidani and E. M. Ismaili Alaoui. “A comprehensive survey of the R-CNN family for object detection”. In: *2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*. 2022, pp. 1–6. DOI: 10.1109/CommNet56067.2022.9993862.

[51] Andrew G. Howard et al. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.” In: *CoRR* abs/1704.04861 (2017).

[52] Kuan-Chih Huang et al. “MonoDTR: Monocular 3D Object Detection With Depth-Aware Transformer”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 4012–4021.

[53] Shaofei Huang et al. “Anchor3DLane: Learning To Regress 3D Anchors for Monocular 3D Lane Detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2023, pp. 17451–17460.

[54] Yanjun Huang et al. “A Survey on Trajectory-Prediction Methods for Autonomous Driving”. In: *IEEE Transactions on Intelligent Vehicles* 7.3 (2022), pp. 652–674. DOI: 10.1109/TIV.2022.3167103.

[55] Yu Huang and Yue Chen. “Autonomous Driving with Deep Learning: A Survey of State-of-the-Art Technologies”. In: *arXiv preprint arXiv:2006.06091* (2020).

[56] Zhiwei Huang et al. “A systematic review of monocular depth estimation for autonomous driving: Methods and dataset benchmarking”. In: *Results in Engineering* 26 (2025), p. 105359. ISSN: 2590-1230. DOI: https://doi.org/10.1016/j.rineng.2025.105359.

[57] Omar Y. Al-Jarrah et al. “Exploring Deep Learning-Based Visual Localization Techniques for UAVs in GPS-Denied Environments”. In: *IEEE Access* 12 (2024), pp. 113049–113071. DOI: 10.1109/ACCESS.2024.3440064.

[58] Peiyuan Jiang et al. “A Review of Yolo Algorithm Developments”. In: *Procedia Computer Science* 199 (2022). The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 and 2021): Developing Global Digital Economy after COVID-19, pp. 1066–1073. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2022.01.135.

[59] Wei Jiang et al. “Research on Performance Limitations of Visual-based Perception System for Autonomous Vehicle under Severe Weather Conditions”. In: *2022 IEEE Intelligent Vehicles Symposium (IV)*. 2022, pp. 586–593. DOI: 10.1109/IV51971.2022.9827169.

[60] Yuzhu Jiang et al. “Pedestrian Group Activity Recognition for Autonomous Vehicles and Robots: A Survey and Perspectives”. In: *IEEE Transactions on Cybernetics* 56.3 (2026), pp. 1515–1528. DOI: 10.1109/TCYB.2025.3629343.

[61] Ronghe Jin et al. “A Survey on Cooperative Positioning Using GNSS Measurements”. In: *IEEE Transactions on Intelligent Vehicles* 9.11 (2024), pp. 7402–7420. DOI: 10.1109/TIV.2024.3397879.

[62] Xin Jin. “A Review of Cityscape Research Based on Dynamic Visual Perception”. In: *Land* 12.6 (2023). ISSN: 2073-445X. DOI: 10.3390/land12061229. URL: https://www.mdpi.com/2073-445X/12/6/1229.

[63] Niels Joubert, Tyler G. R. Reid, and Fergus Noble. “Developments in Modern GNSS and Its Impact on Autonomous Vehicle Architectures”. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020, pp. 2029–2036. DOI: 10.1109/IV47402.2020.9304840.

[64] Ivan Jureta. “Requirements Engineering Methods: A Classification Framework and Research Challenges”. In: (Mar. 2012).

[65] Tala Talaei Khoei, Hadjar Ould Slimane, and Naima Kaabouch. “Deep Learning: Systematic Review, Models, Challenges, and Research Directions”. In: *Neural Computing and Applications* 35.31 (2023), pp. 23103–23124. ISSN: 1433-3058. DOI: 10.1007/s00521-023-08957-4. URL: https://doi.org/10.1007/s00521-023-08957-4.

[66] ByeoungDo Kim et al. “Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network”. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. 2017.

[67] Yeongmin Ko et al. “Key Points Estimation and Point Instance Segmentation Approach for Lane Detection”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.7 (2022), pp. 8949–8958. DOI: 10.1109/TITS.2021.3088488.

[68] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012.

[69] Sunil Kumar, Ankur Pandey, and Sudeep Varshney. “Exploring the Impact of Deep Learning Models on Lane Detection Through Semantic Segmentation”. In: *SN Computer Science* 5.1 (2024), p. 139. DOI: 10.1007/s42979-023-02328-5. URL: https://doi.org/10.1007/s42979-023-02328-5.

[70] Sampo Kuutti et al. “A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications”. In: *IEEE Internet of Things Journal* 5.2 (2018), pp. 829–846. DOI: 10.1109/JIOT.2018.2812300.

[71] Feng-Jie Li, Chun-Yang Zhang, and C. L. Philip Chen. “STS-DGNN: Vehicle Trajectory Prediction via Dynamic Graph Neural Network With Spatial–Temporal Synchronization”. In: *IEEE Transactions on Instrumentation and Measurement* (2023).

[72] Hongyang Li et al. “Delving into the Devils of Bird’s-Eye-View Perception: A Review, Evaluation and Recipe”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023). DOI: 10.1109/TPAMI.2023.3333838.

[73] You Li and Javier Ibanez-Guzman. “Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems”. In: *IEEE Signal Processing Magazine* 37.4 (2020), pp. 50–61. DOI: 10.1109/MSP.2020.2973615.

[74] Yuhang Li et al. “BRECQ: Pushing the Limit of Post-Training Quantization by Block Reconstruction”. In: *International Conference on Learning Representations*. 2021.

[75] Zhiqi Li et al. “BEVFormer: Learning Bird’s-Eye-View Representation From LiDAR-Camera Via Spatiotemporal Transformers”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP.99 (2024), pp. 1–18. DOI: 10.1109/TPAMI.2024.3515454. URL: https://doi.org/10.1109/TPAMI.2024.3515454.

[76] Zhuoling Li et al. “GroupLane: End-to-End 3D Lane Detection With Channel-Wise Grouping”. In: *IEEE Robotics and Automation Letters* 9.11 (2024), pp. 10487–10494. DOI: 10.1109/LRA.2024.3475881.

[77] Luigi Di Lillo et al. “Comparative Safety Performance of Autonomous- and Human Drivers: A Real-World Case Study of the Waymo Driver”. In: *Heliyon* 10.14 (2024), e34379. ISSN: 2405-8440. DOI: 10.1016/j.heliyon.2024.e34379. URL: https://doi.org/10.1016/j.heliyon.2024.e34379.

[78] Tsung-Yi Lin et al. “Feature Pyramid Networks for Object Detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.

[79] Hanxiao Liu, Karen Simonyan, and Yiming Yang. “DARTS: Differentiable Architecture Search”. In: *International Conference on Learning Representations*. 2019.

[80] Jianbang Liu et al. “A Survey on Deep-Learning Approaches for Vehicle Trajectory Prediction in Autonomous Driving”. In: *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. 2021, pp. 978–985. DOI: 10.1109/ROBIO54168.2021.9739407.

[81] Jiawei Liu et al. “PD-Quant: Post-Training Quantization Based on Prediction Difference Metric”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 24427–24437.

[82] Lizhe Liu et al. “CondLaneNet: A Top-To-Down Lane Detection Framework Based on Conditional Convolution”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2021, pp. 3773–3782.

[83] Xinyu Liu et al. “Deep Transfer Learning for Intelligent Vehicle Perception: A Survey”. In: *arXiv preprint arXiv:2306.15110* (2023).

[84] Yingfei Liu et al. “PETR: Position Embedding Transformation for Multi-view 3D Object Detection”. In: *Computer Vision – ECCV 2022*. Ed. by Shai Avidan et al. Cham: Springer Nature Switzerland, 2022, pp. 531–548. ISBN: 978-3-031-19812-0. DOI: 10.1007/978-3-031-19812-0_31.

[85] Ze Liu et al. “CondLaneNet: a top-to-down lane detection framework based on conditional convolution”. In: *CVPR*. 2021.

[86] Pingping Lu, Shaobing Xu, and Huei Peng. “Graph-Embedded Lane Detection”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 2977–2988. DOI: 10.1109/TIP.2021.3057287.

[87] Yueru Luo et al. “LATR: 3D Lane Detection from Monocular Images with Transformer”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2023, pp. 7941–7952.

[88] Fulong Ma et al. “Monocular 3D Lane Detection for Autonomous Driving: Recent Achievements, Challenges, and Outlooks”. In: *IEEE Transactions on Intelligent Transportation Systems* 26.6 (2025), pp. 7380–7400. DOI: 10.1109/TITS.2025.3550745.

[89] Nan Ma, Guilin Pang, and Yiu-Ming Cheung. “Exploiting the Complementarity of Bilateral Domains for Fast Lane Detection”. In: *IEEE Transactions on Intelligent Transportation Systems* 26.11 (2025), pp. 20109–20121. DOI: 10.1109/TITS.2025.3597352.

[90] Yuexin Ma et al. “Vision-Centric BEV Perception: A Survey”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024). DOI: 10.1109/TPAMI.2024.3449912.

[91] Adam H. Marblestone, Greg Wayne, and Konrad P. Kording. “Toward an Integration of Deep Learning and Neuroscience”. In: *Frontiers in Computational Neuroscience* 10 (2016), p. 94. DOI: 10.3389/fncom.2016.00094.

[92] Markus Maurer et al. *Autonomous driving: technical, legal and social aspects*. Springer, 2016.

[93] Larry R Medsker, Lakhmi Jain, et al. “Recurrent neural networks”. In: *Design and applications* 5.64-67 (2001), p. 2.

[94] Samiyaa Yaseen Mohammed. “Architecture review: Two-stage and one-stage object detection”. In: *Franklin Open* 12 (2025), p. 100322. ISSN: 2773-1863. DOI: https://doi.org/10.1016/j.fraope.2025.100322. URL: https://www.sciencedirect.com/science/article/pii/S2773186325001100.

[95] Alexandra S. Mueller, Jessica B. Cicchino, and David S. Zuby. “What Humanlike Errors Do Autonomous Vehicles Need to Avoid to Maximize Safety?” In: *Journal of Safety Research* 75 (2020), pp. 310–318. DOI: 10.1016/j.jsr.2020.10.005.

[96] Khan Muhammad et al. “Vision-Based Semantic Segmentation in Scene Understanding for Autonomous Driving: Recent Achievements, Challenges, and Outlooks”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.12 (2022), pp. 22694–22715. DOI: 10.1109/TITS.2022.3207665.

[97] Ruta Mulajkar and Sanjay Yede. “YOLO Version v1 to v8 Comprehensive Review”. In: *2024 International Conference on Inventive Computation Technologies (ICICT)*. 2024, pp. 472–478. DOI: 10.1109/ICICT60155.2024.10544452.

[98] Markus Nagel et al. “Up or Down? Adaptive Rounding for Post-Training Quantization”. In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*. Ed. by Hal Daumé III and Aarti Singh. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 7197–7206. URL: https://proceedings.mlr.press/v119/nagel20a.html.

[99] Fernando Navarro, Francisco J. Serón, and Diego Gutierrez. “Motion Blur Rendering: State of the Art”. In: *Computer Graphics Forum* 30.1 (2011), pp. 3–26. DOI: https://doi.org/10.1111/j.1467-8659.2010.01840.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2010.01840.x.

[100] Sergiu Nedevschi et al. “3D lane detection system based on stereovision”. In: *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*. IEEE. 2004, pp. 161–166.

[101] Davy Neven et al. “Towards end-to-end lane detection: an instance segmentation approach”. In: *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 286–291.

[102] Erkan Oğuz et al. “A deep learning based fast lane detection approach”. In: *Chaos, Solitons and Fractals* 155 (2022), p. 111722. ISSN: 0960-0779. DOI: https://doi.org/10.1016/j.chaos.2021.111722. URL: https://www.sciencedirect.com/science/article/pii/S0960077921010766.

[103] Halil İbrahim Öztürk, Muhammet Esat Kalfaoğlu, and Ozsel Kilinc. “GLane3D: Detecting Lanes with Graph of 3D Keypoints”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2025, pp. 27508–27518.

[104] Dong-Hee Paek, Kevin Tirta Wijaya, and Seung-Hyun Kong. “Row-wise LiDAR Lane Detection Network with Lane Correlation Refinement”. In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. 2022, pp. 4328–4334. DOI: 10.1109/ITSC55140.2022.9922341.

[105] Vrushali Pagire, Murthy Chavali, and Ashish Kale. “A comprehensive review of object detection with traditional and deep learning methods”. In: *Signal Processing* 237 (2025), p. 110075. ISSN: 0165-1684. DOI: https://doi.org/10.1016/j.sigpro.2025.110075.

[106] Xingang Pan et al. “Spatial As Deep: Spatial CNN for Traffic Lane Detection”. In: *AAAI*. 2018.

[107] Ying Peng et al. “Survey on Image and Point-Cloud Fusion-Based Object Detection in Autonomous Vehicles”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.12 (2022), pp. 22772–22789. DOI: 10.1109/TITS.2022.3206235.

[108] Mathias Pessiglione et al. “Origins and Consequences of Cognitive Fatigue”. In: *Trends in Cognitive Sciences* 29.8 (Aug. 2025), pp. 730–749. ISSN: 1364-6613. DOI: 10.1016/j.tics.2025.02.005. URL: https://doi.org/10.1016/j.tics.2025.02.005.

[109] Maximilian Pittner, Alexandru Condurache, and Joel Janai. “3D-SpLineNet: 3D Traffic Line Detection Using Parametric Spline Representations”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2023, pp. 602–611.

[110] Zequn Qin, Huanyu Wang, and Xi Li. “Ultra Fast Structure-Aware Deep Lane Detection”. In: *Computer Vision – ECCV 2020*. Vol. 12346. Lecture Notes in Computer Science. Springer, 2020, pp. 276–291. DOI: 10.1007/978-3-030-58586-0_17. URL: https://doi.org/10.1007/978-3-030-58586-0_17.

[111] Zhan Qu et al. “Focus on Local: Detecting Lane Marker from Bottom Up via Key Point”. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 14117–14125. DOI: 10.1109/CVPR46437.2021.01390.

[112] Uchitha Rajapaksha et al. “Deep Learning-based Depth Estimation Methods from Monocular Image and Videos: A Comprehensive Survey”. In: *ACM Comput. Surv.* 56.12 (Oct. 2024). ISSN: 0360-0300. DOI: 10.1145/3677327. URL: https://doi.org/10.1145/3677327.

[113] Waseem Rawat and Zenghui Wang. “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review”. In: *Neural Computation* 29.9 (2017), pp. 2352–2449. DOI: 10.1162/neco_a_00990.

[114] Robert M. Sapolsky. *Behave: The Biology of Humans at Our Best and Worst*. New York: Penguin Press, 2017.

[115] Robert M. Sapolsky. *Determined: A Science of Life Without Free Will*. New York: Penguin Press, 2023.

[116] Birgit Schlager et al. “Contaminations on Lidar Sensor Covers: Performance Degradation Including Fault Detection and Modeling as Potential Applications”. In: *IEEE Open Journal of Intelligent Transportation Systems* 3 (2022), pp. 738–747. DOI: 10.1109/OJITS.2022.3214094.

[117] Matthias Schreier, Volker Willert, and Jürgen Adamy. "An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments". In: *IEEE Transactions on Intelligent Transportation Systems* (2016).

[118] Brent Schwarz. "Mapping the World in 3D". In: *Nature Photonics* 4.7 (2010), pp. 429–430. ISSN: 1749-4893. DOI: 10.1038/nphoton.2010.148. URL: https://doi.org/10.1038/nphoton.2010.148.

[119] K Simonyan and A Zisserman. "Very deep convolutional networks for large-scale image recognition". In: Computational and Biological Learning Society, 2015, pp. 1–14.

[120] Santokh Singh. *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*. Tech. rep. DOT HS 812 115. National Highway Traffic Safety Administration, 2015.

[121] Chen Sun et al. "Toward Ensuring Safety for Autonomous Driving Perception: Standardization Progress, Research Advances, and Perspectives". In: *IEEE Transactions on Intelligent Transportation Systems* (2024).

[122] Pei Sun et al. "Scalability in Perception for Autonomous Driving: Waymo Open Dataset". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020.

[123] Lucas Tabelini et al. "Keep your Eyes on the Lane: Real-time Attention-guided Lane Detection". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 294–302. DOI: 10.1109/CVPR46437.2021.00036.

[124] William Talbot et al. "Continuous-Time State Estimation Methods in Robotics: A Survey". In: *IEEE Transactions on Robotics* 41 (2025), pp. 4975–4999. DOI: 10.1109/TRO.2025.3593079.

[125] Mingxing Tan and Quoc Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 6105–6114. URL: https://proceedings.mlr.press/v97/tan19a.html.

[126] Siyu Teng et al. "Motion Planning for Autonomous Driving: The State of the Art and Future Perspectives". In: *IEEE Transactions on Intelligent Vehicles* 8.6 (2023), pp. 3692–3711. DOI: 10.1109/TIV.2023.3274536.

[127] Xiaoqiang Teng et al. "Deep Learning for 3-D Lane Detection in Autonomous Driving: A Survey". In: *IEEE Internet of Things Journal* 13.6 (2026), pp. 10432–10453. DOI: 10.1109/JIOT.2026.3654005.

[128] Fabio Tosi, Luca Bartolomei, and Matteo Poggi. "A Survey on Deep Stereo Matching in the Twenties". In: *International Journal of Computer Vision* 133.7 (2025), pp. 4245–4276. ISSN: 1573-1405. DOI: 10.1007/s11263-024-02331-0. URL: https://doi.org/10.1007/s11263-024-02331-0.

[129] Ritam Upadhyay et al. "Real-Time Deep Learning–Based Image Processing for Pose Estimation and Object Localization in Autonomous Robot Applications". In: *The International Journal of Advanced Manufacturing Technology* 127.3 (2023), pp. 1905–1919. ISSN: 1433-3015. DOI: 10.1007/s00170-022-09994-4. URL: https://doi.org/10.1007/s00170-022-09994-4.

[130] Ashish Vaswani et al. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[131] Jinsheng Wang et al. "A Keypoint-Based Global Association Network for Lane Detection". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 1392–1401.

[132] Shoude Wang and Nur Syazreen Ahmad. "A Comprehensive Review on Sensor Fusion Techniques for Localization of a Dynamic Target in GPS-Denied Environments". In: *IEEE Access* 13 (2025), pp. 2252–2285. DOI: 10.1109/ACCESS.2024.3519874.

[133] Song Wang and Zhixia Li. "Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches". In: *PloS one* 14.3 (2019), e0214550.

[134] Tai Wang et al. "FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*. Oct. 2021, pp. 913–922.

[135] Tianzhe Wang et al. "APQ: Joint Search for Network Architecture, Pruning and Quantization Policy". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 2078–2087.

[136] Xinlong Wang et al. "Dense Contrastive Learning for Self-Supervised Visual Pre-Training". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 3024–3033.

[137] Xiuying Wei et al. "QDrop: Randomly Dropping Quantization for Extremely Low-Bit Post-Training Quantization". In: *International Conference on Learning Representations*. 2022.

[138] World Health Organization. *Global Status Report on Road Safety 2023*. Geneva: World Health Organization, 2023.

[139] Jianxin Wu. “Introduction to convolutional neural networks”. In: *National Key Lab for Novel Software Technology. Nanjing University. China* 5.23 (2017), p. 495.

[140] Chao Xiang et al. “Multi-Sensor Fusion and Cooperative Perception for Autonomous Driving: A Review”. In: *IEEE Intelligent Transportation Systems Magazine* 15.5 (2023), pp. 36–58. DOI: 10.1109/MITS.2023.3283864.

[141] Lingyu Xiao et al. “ADNet: Lane Shape Prediction via Anchor Decomposition”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2023, pp. 6404–6413.

[142] Zhenda Xie et al. “Propagate Yourself: Exploring Pixel-Level Consistency for Unsupervised Visual Representation Learning”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 16684–16693.

[143] Hang Xu et al. “CurveLane-NAS: Unifying Lane-Sensitive Architecture Search and Adaptive Point Blending”. In: *Computer Vision – ECCV 2020*. Cham: Springer International Publishing, 2020, pp. 689–704. ISBN: 978-3-030-58555-6. DOI: 10.1007/978-3-030-58555-6_41.

[144] Pengyu Xue et al. “Drivable Area Detection Method in Dark Unstructured-Roads Based on CNN Data Fusion With Surface Normal Estimation”. In: *IEEE Transactions on Intelligent Transportation Systems* 26.6 (2025), pp. 8694–8706. DOI: 10.1109/TITS.2025.3557013.

[145] Chenyu Yang et al. “BEVFormer v2: Adapting Modern Image Backbones to Bird’s-Eye-View Recognition via Perspective Supervision”. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 17830–17839. DOI: 10.1109/CVPR52729.2023.01710.

[146] Sunghun Yang et al. “MonoCLUE: Object-Aware Clustering Enhances Monocular 3D Object Detection”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 40. 14. 2026, pp. 11721–11729. DOI: 10.1609/aaai.v40i14.38157. URL: https://doi.org/10.1609/aaai.v40i14.38157.

[147] Shanliang Yao et al. “Exploring Radar Data Representations in Autonomous Driving: A Comprehensive Review”. In: *IEEE Transactions on Intelligent Transportation Systems* 26.6 (2025), pp. 7401–7425. DOI: 10.1109/TITS.2025.3554781.

[148] Jong Chul Ye and Woon Kyoung Sung. “Understanding Geometry of Encoder-Decoder CNNs”. In: *Proceedings of the 36th International Conference on Machine Learning*. 2019, pp. 7064–7073.

[149] Sibel Yenikaya, Gökhan Yenikaya, and Ekrem Düven. “Keeping the vehicle on the road: A survey on on-road lane detection systems”. In: (2013).

[150] Ekim Yurtsever et al. “A Survey of Autonomous Driving: Common Practices and Emerging Technologies”. In: *IEEE Access* 8 (2020), pp. 58443–58469. DOI: 10.1109/ACCESS.2020.2983149.

[151] Ekim Yurtsever et al. “A Survey of Autonomous Driving: Common Practices and Emerging Technologies”. In: *IEEE Access* 8 (2020), pp. 58443–58469. DOI: 10.1109/ACCESS.2020.2983149.

[152] Jiaxing Zhang et al. “Roadside lidar-based scene understanding toward intelligent traffic perception: A comprehensive review”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 233 (2026), pp. 69–88. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2026.01.012.

[153] Jingsi Zhang et al. “A novel deep LeNet-5 convolutional neural network model for image recognition”. In: *Computer Science and Information Systems* 19.3 (2022), pp. 1463–1480.

[154] Kunpeng Zhang et al. “Trajectory Prediction for Autonomous Driving Using Spatial-Temporal Graph Attention Transformer”. In: *IEEE Transactions on Intelligent Transportation Systems* (2022).

[155] Peng Zhang et al. “A New Literature Review of 3D Object Detection on Autonomous Driving”. In: *J. Artif. Int. Res.* 82 (June 2025). DOI: 10.1613/jair.1.15961. URL: https://doi.org/10.1613/jair.1.15961.

[156] Yi Zhang et al. “Accurate and Real-Time Object Detection Based on Bird’s Eye View on 3D Point Clouds”. In: *2019 International Conference on 3D Vision (3DV)*. 2019, pp. 214–221.

[157] Ying Zhang et al. “Demand-Oriented Autonomous Navigation for Mobile Robots With Vector-Semantic Mapping in Urban Environments”. In: *IEEE Transactions on Vehicular Technology* 75.6 (2026), pp. 9510–9521. DOI: 10.1109/TVT.2025.3649304.

[158] Jianfeng Zhao et al. “Autonomous Driving System: A Comprehensive Survey”. In: *Expert Systems with Applications* 242 (2024), p. 122836. DOI: 10.1016/j.eswa.2023.122836.

[159] Jingyuan Zhao et al. “A Survey of Autonomous Driving from a Deep Learning Perspective”. In: *ACM Computing Surveys* 57.10 (Oct. 2025), pp. 1–60. DOI: 10.1145/3729420. URL: https://doi.org/10.1145/3729420.

[160] Jingyuan Zhao et al. “A Survey of Autonomous Driving from a Deep Learning Perspective”. In: *ACM Comput. Surv.* 57.10 (May 2025). ISSN: 0360-0300. DOI: 10.1145/3729420. URL: https://doi.org/10.1145/3729420.

[161] Jingyuan Zhao et al. “Autonomous Driving System: A Comprehensive Survey”. In: *Expert Systems with Applications* 242 (2024), p. 122836. DOI: 10.1016/j.eswa.2023.122836. URL: https://doi.org/10.1016/j.eswa.2023.122836.

[162] Tu Zheng et al. “RESA: Recurrent Feature-Shift Aggregator for Lane Detection”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 4. 2021, pp. 3547–3554. DOI: 10.1609/aaai.v35i4.16469. URL: https://doi.org/10.1609/aaai.v35i4.16469.

[163] Xizhou Zhu et al. “Deformable {DETR}: Deformable Transformers for End-to-End Object Detection”. In: *International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=gZ9hCDWe6ke.

[164] Zhengxia Zou et al. “Object Detection in 20 Years: A Survey”. In: *Proceedings of the IEEE* 111.3 (2023), pp. 257–276. DOI: 10.1109/JPROC.2023.3238524.