

LANE DETECTION FOR AUTONOMOUS VEHICLES USING DEEP LEARNING STRATEGIES

BY
ALI MAGDI ELBORKI

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
In
COMPUTER SCIENCE

MAY 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **ALI MAGDI ELBORKI** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN COMPUTER SCIENCE**.

Thesis Committee

Dr. Saeed Anwer (Adviser)

Dr. Muhammad Usman (Member)

Dr. Farag Azzedin (Member)

Dr. Motaz Alfarraj

Department Chairman

Dr. Suliman Saleh Al-

Homidan

Dean of Graduate Studies

Date



©Ali Magdi Elborki
2024

Dedication

This work is a tribute to my wonderful family. Their unwavering belief and endless support, from my parents to my siblings, gave me the strength to reach this milestone.

ACKNOWLEDGMENTS

With profound gratitude, I express my thanks to Allah (SWT) for bestowing upon me the patience, strength, and motivation needed to complete this project. I am deeply appreciative of my advisor, Dr. Saeed Anwer, whose unwavering support, invaluable suggestions, comments, and encouragement have illuminated my path throughout this endeavor. I am particularly thankful to my thesis committee members, Dr. Muhammad Usman and Dr. Farag Azzedin, for their steadfast guidance and support. Additionally, I extend my heartfelt thanks to my colleagues and friends who have assisted me along the way, with special mention to Hisham Garout for his consistent and pivotal support throughout my journey. Lastly, my deepest gratitude goes to my parents, who have provided all imaginable forms of support, and to my siblings, whose continuous encouragement and belief in me have been a constant source of inspiration. This work stands as a testament to the collective support and encouragement of everyone mentioned, and it is with great honor that I recognize your contributions to my academic journey.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
ABSTRACT (ENGLISH)	xi
ABSTRACT (ARABIC)	xiii
CHAPTER 1 INTRODUCTION	1
1.1 Object Detection	2
1.2 Current Lane Detection Architecture	4
1.3 Problem Statement and Motivation	5
1.3.1 Significance of LD in Autonomous Driving	6
1.3.2 Current Challenges in Lane Detection	6
1.3.3 Motivation	8
1.4 Research Objectives and Contributions	9
1.5 Outline of Thesis	12
CHAPTER 2 RELATED WORK	14
2.1 Overview of Lane Detection Techniques	16
2.2 Definition of Autonomous Driving According to the Society of Auto- motive Engineers (SAE)	17
2.3 Lane Detection Methods	17

2.4	Autonomous Driving Datasets	18
2.5	Deep Learning Methods for Lane Detection	20
2.6	Traditional Lane Detection Techniques (Pre-Deep Learning)	20
2.6.1	Edge Detection Methods	20
2.6.2	Hough Transform	21
2.7	Deep Learning-Based Lane Detection Techniques	22
2.7.1	Segmentation-Based Methods	22
2.7.2	Anchor-Based Methods	23
2.7.3	Lane Classification and Mixed Precision Techniques	23
2.7.4	Novel Approaches and Challenges	23
2.8	Camouflaged Object Detection	24
2.9	Vision Transformers	25
2.10	Feature Pyramid Networks	26
2.11	Challenges and Open Issues in Deep Learning for Lane Detection	27
2.12	Future Directions	28

CHAPTER 3 VITCLRNET PROPOSED MODEL 31

3.1	Scope of the Experiment	31
3.1.1	Objective and Goals	31
3.1.2	Loss Function	32
3.1.3	Metrics	34
3.2	Design of the Experiment	36
3.2.1	Context Selection	36
3.2.2	CULane Dataset	36
3.2.3	TuSimple Dataset	37
3.3	Preprocessing Techniques	37
3.4	Examples of Preprocessing and Augmentation	39
3.4.1	Original and Augmented Images from CULane and TuSimple Dataset	40
3.5	Non-Local Token Enhancement Module (NL-TEM)	42

3.6	Model Architecture Description	43
3.7	Hardware Specifications	45
3.8	Training Configuration	45
CHAPTER 4 PERFORMANCE EVALUATION		47
4.1	Introduction	47
4.2	Performance Evaluation Methodology	47
4.3	Experiments and Results	49
4.3.1	Experiment with Vision Transformers (ViT) as Backbone in CLRNet	49
4.3.2	Experiment with Dense Prediction Transformers (DPT)	52
4.3.3	Experiment with FSPNet	55
4.3.4	Experiment with ViTCLRNet	58
4.4	Comparative Analysis	60
4.4.1	Results on CULane Dataset	60
4.4.2	Results on TuSimple Dataset	61
4.4.3	Discussion	64
CHAPTER 5 CONCLUSION AND FUTURE WORK		66
5.1	Conclusion	66
5.2	Limitations	67
5.3	Future Work	69
5.4	Summary of Key Findings	70
5.5	Contributions of the Thesis	72
5.6	Future Work	73
5.7	Conclusion	74
REFERENCES		75
VITAE		84

LIST OF TABLES

Tab.2.1	Summary of Lane Detection Methods [1]	19
Tab.2.2	A Summary of Retrieved Related Studies in the Literature	30
Tab.3.1	Summary of Experiments with Different Backbones and Datasets	46
Tab.4.1	Performance Metrics on CULane Dataset using ViT	50
Tab.4.2	Performance Metrics on TuSimple Dataset using ViT	51
Tab.4.3	Performance Metrics for the CULane Dataset using DPT.	53
Tab.4.4	Performance Metrics for the Tusimple Dataset using DPT.	54
Tab.4.5	Performance Metrics on CULane Dataset for FSPNet	56
Tab.4.6	Performance Metrics on TuSimple Dataset for FSPNet	57
Tab.4.7	Performance Metrics on CULane Dataset using ViTCLRNet	58
Tab.4.8	Performance Metrics on TuSimple Dataset using ViTCLRNet	59
Tab.4.9	Comparison of performance metrics on CULane dataset across different models.	61
Tab.4.10	Performance comparison of different methods on CULane Dataset (Part 1).	62
Tab.4.11	Performance comparison of different methods on CULane Dataset (Part 2).	63
Tab.4.12	Comparison of performance metrics on TuSimple dataset across different models.	64
Tab.4.13	Comparison of lane detection methods on TuSimple Dataset.	65

LIST OF FIGURES

Fig.1.1	General Lane Detection System Architecture.	5
Fig.2.1	SAE - J3016: Table of Levels of Automation	18
Fig.3.1	Original, Preprocessed, and Augmented Images from CULane Dataset	41
Fig.3.2	Original, Preprocessed, and Augmented Images from TuSimple Dataset	41
Fig.3.3	Model Architecture Diagram	43
Fig.3.4	Model Architecture Diagram	44
Fig.4.1	Result graph for the CULane Dataset using ViT as a backbone. . .	50
Fig.4.2	Result graph for the CULane Dataset using DPT as a backbone. .	53
Fig.4.3	Result graph for the CULane Dataset using FSPNet as a backbone	56
Fig.4.4	Result graph for the CULane Dataset using ViTCLRNet.	59
Fig.4.5	Comparison of experimental results on CULane dataset across dif- ferent models.	61
Fig.4.6	Comparison of performance metrics on CULane dataset across dif- ferent models.	61
Fig.4.7	Comparison of performance metrics on TuSimple dataset across different models.	63
Fig.4.8	Comparison of performance metrics on TuSimple dataset across different models.	64

THESIS ABSTRACT

NAME: Ali Magdi Elborki

TITLE OF STUDY: Lane Detection For Autonomous Vehicles Using Deep Learning Strategies

MAJOR FIELD: Computer Science

DATE OF DEGREE: May 2024

Lane detection is a fundamental component of autonomous driving systems, essential for vehicle navigation and driver assistance. This thesis investigates advanced lane detection techniques using deep learning strategies to improve accuracy and robustness under varying conditions. We integrate Vision Transformers (ViT), Dense Prediction Transformers (DPT), and Feature Shrinkage Pyramid Networks (FSPNet) within the CLRNet architecture. Our research includes four experiments: using ViT, DPT, and FSPNet as backbones, and combining ViT with the Non-Local Token Enhancement Module (NL-TEM) from FSPNet. The performance of these models is evaluated on the CULane and TuSimple datasets, using metrics such as F1-score, accuracy, and processing speed. Results indicate that while ViTs show promise, the ViTCLRNet model, which integrates ViT and NL-TEM, achieves the highest performance,

offering an optimal balance of accuracy and computational efficiency. This research contributes to the field by presenting a robust lane detection model, providing practical implementation insights, and suggesting avenues for future research to further enhance autonomous driving technologies. Keywords: Lane Detection, Autonomous Vehicles, Deep Learning, Vision Transformers, CLRNet, CULane, TuSimple

ملخص الرسالة

الاسم: علي مجدي البركي

عنوان الدراسة: اكتشاف المسار للمركبات ذاتية القيادة باستخدام استراتيجيات التعلم العميق

التخصص: علوم الحاسوب

تاريخ الدرجة العلمية: ذي القعدة 1445

كشف المسار هو مكون أساسي في أنظمة القيادة الذاتية، وهو ضروري للتنقل الآمن والمساعدة للسائق. تبحث هذه الأطروحة في تقنيات كشف المسار المتقدمة باستخدام استراتيجيات التعلم العميق لتحسين الدقة والمتانة في ظل ظروف متنوعة. نقوم بدمج Vision Transformers (ViT)، وDense Prediction Transformers (DPT)، وشبكات الهرم التلقضي للميزات (FSPNet) داخل هيكلية CLRNet. تتضمن تجاربنا استخدام هذه النماذج كخلفيات ودمج ViT مع وحدة تعزيز الرموز غير المحلية (NL-TEM) من FSPNet. قمنا بتقييم أداء هذه النماذج على مجموعتي بيانات CULane وTuSimple باستخدام مقاييس مثل درجة F1 والدقة وسرعة المعالجة. حقق نموذج ViTCLRNet الأداء الأعلى، مقدمًا توازنًا مثاليًا بين الدقة والكفاءة الحسابية. تساهم هذه الأبحاث في تقديم نموذج قوي لكشف المسار، وتوفير رؤى عملية للتنفيذ، واقتراح اتجاهات بحث مستقبلية لتعزيز تقنيات القيادة الذاتية. الكلمات المفتاحية: كشف المسار، السيارات الذاتية، التعلم العميق، محولات الرؤية، CLRNet،

TuSimple، CULane

CHAPTER 1

INTRODUCTION

Lane detection (LD) is a fundamental aspect of computer vision and is crucial in autonomous vehicle navigation and driver assistance systems (DAS). It involves identifying the boundaries of traffic lanes on the road, enabling vehicles to stay within their designated lanes and avoid collisions with other vehicles or obstacles. In the context of urban cities, lane detection is paramount due to the country's rapidly growing automotive market and the increasing emphasis on smart transportation systems.[2].

The ability to accurately detect lanes under varying weather, lighting, and road conditions is essential for enhancing traffic safety and facilitating the widespread adoption of autonomous vehicles in the region. Many methods, as referenced by [3] [4], have achieved impressive results by leveraging the strong feature representation capabilities of Convolutional Neural Networks (CNNs). Nevertheless, there continue to be hurdles when accurately detecting lanes. Lanes possess high-level contextual meaning, but they also have unique local patterns that require precise localization, demanding the utilization of intricate low-level features.

The applications of lane detection are extensive and critically important, encom-

passing autonomous driving, Advanced Driver-Assistance Systems (ADAS), traffic management, and road infrastructure maintenance. In the context of autonomous driving, lane detection algorithms facilitate the navigation of complex road networks, allowing vehicles to make informed decisions to maintain safe distances from other vehicles, pedestrians, and obstacles [5]. ADAS utilizes lane detection to enhance driver support through features like lane departure warnings and lane-keeping assistance, relying on accurate lane detection for efficacy. Traffic management systems leverage lane detection data to optimize traffic flow and reduce congestion, while road maintenance authorities use this information to identify areas in need of repair or improvements. [6]

A variety of methods have been developed to tackle the lane detection problem, ranging from traditional image processing techniques to machine learning-based approaches. Traditional methods, including the Hough Transform and edge detection [7], are highly dependent on the specific conditions of the input images and often struggle to adapt to varying environments. Recent advances in machine learning have introduced approaches such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), which have demonstrated promising results in lane detection [8]. However, these methods can be computationally intensive and may experience reduced performance in complex, cluttered, or poorly illuminated scenes.

1.1 Object Detection

In recent times, object detection has emerged as a crucial element in the field of computer vision, revolutionizing how machines interpret the visual world. This rise in in-

terest is primarily due to its profound impact on enhancing automation and analytical capabilities across various industries. Object detection leverages advanced algorithms to identify and locate objects within digital images or video streams, turning raw visual inputs into actionable insights. Methodologies that contribute significantly to the effectiveness of object detection include convolutional neural networks and deep learning.

DL facilitates the analysis of large datasets, improving object recognition accuracy over time, while CNNs specialize in analyzing visual images and extracting essential features to accurately identify objects. The integration of these technologies has been instrumental in developing robust object detection systems, capable of performing in diverse and challenging environments. These deep learning models excel at detecting and localizing objects within images or video streams[9], providing a robust foundation for accurate lane detection systems. By training CNNs on diverse datasets containing annotated images of roads, vehicles, and pedestrians, developers can fine-tune these models to recognize and differentiate between various objects encountered on the road. Furthermore, the integration of object detection into lane detection systems facilitates proactive decision-making by self-driving cars. As other vehicles changing lanes or pedestrians crossing the street, autonomous vehicles can adjust their speed, trajectory, and behavior to ensure safe and efficient navigation by detecting and tracking objects in the vicinity. This real-time awareness of the surrounding environment enhances the overall performance and reliability of lane detection systems, contributing to the advancement of autonomous driving technology.

1.2 Current Lane Detection Architecture

The present lane detection system is a sophisticated and intricate setup, comprising multiple elements such as machine learning algorithms and sensors, aimed at precise and reliable lane identification on roads. This system tackles issues including insufficient accuracy, slow processing, and vulnerability to factors such as lighting and weather conditions. Central to this setup are machine learning algorithms, especially deep learning, which excel in lane detection by extracting complex features from images to accurately pinpoint lane lines.

In our work, we merge Vision Transformers with a model the Feature Shrinkage Pyramid of Cross-Layer Refinement Network [10]. For the rest of this thesis, we refer to this model as ViTCLRNet. ViTCLRNet is a deep learning model designed to address the limitations of previous architectures, such as those presented in [10] and [11]. These previous architectures struggled with lighting variations and required distinct lane markings for accurate detection. To overcome these challenges, we have integrated a vision transformer backbone [12], including a refinement process, to enhance detection precision. The vision transformer backbone is a type of machine learning algorithm that is particularly effective in processing image data, and it has been shown to improve lane detection accuracy in various lighting conditions. The refinement process further enhances the accuracy of the lane detection by refining the features extracted from the image data.

Additionally, the architecture incorporates methods to counter dataset imbalance, such as data augmentation techniques that increase the number of usable images

through transformations like rotation and scaling. These techniques help to ensure that the model is trained on a diverse and representative dataset, which is critical for accurate lane detection in various road and environmental conditions.

In conclusion, the current lane detection architecture is a sophisticated and advanced system that is fundamental to the functionality of autonomous driving systems. By addressing critical performance and environmental sensitivities, this architecture enables accurate and reliable lane detection, even in challenging lighting and weather conditions

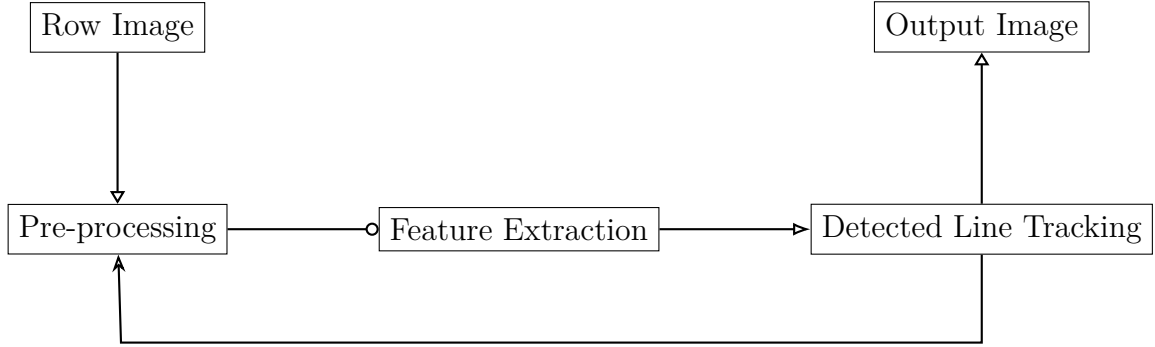


Figure 1.1: General Lane Detection System Architecture.

1.3 Problem Statement and Motivation

Lane detection is a critical component in autonomous driving systems, where the accurate identification of lane markings is essential for safe navigation and decision-making. The current challenges in lane detection systems include issues such as varying lighting conditions, road marking quality, and the need for real-time processing. These challenges can impact the overall performance and reliability of autonomous vehicles, highlighting the importance of developing robust and efficient lane detection

architectures.

1.3.1 Significance of LD in Autonomous Driving

The significance of lane detection in autonomous driving cannot be overstated. As autonomous vehicles navigate through urban and rural environments, the ability to accurately detect and follow lanes is fundamental to their operation. Lane detection systems help in maintaining the vehicle's position on the road, ensuring compliance with traffic regulations, and avoiding collisions. Moreover, they provide critical data for making real-time decisions such as lane changing, merging, and responding to dynamic road conditions. Effective lane detection enhances the safety and reliability of autonomous vehicles, making it a cornerstone technology in the pursuit of fully autonomous driving.

1.3.2 Current Challenges in Lane Detection

- 1. Varying Lighting Conditions:** Lane detection systems must perform accurately under a wide range of lighting conditions, including bright sunlight, shadows, and nighttime driving. Changes in lighting can cause significant variations in the appearance of lane markings, making it challenging for detection algorithms to maintain high accuracy.
- 2. Adverse Weather Conditions:** Weather conditions such as rain, fog, and snow can obscure lane markings, introduce reflections, and create visual noise, all of which can degrade the performance of lane detection systems. Developing

algorithms that can robustly handle these conditions is critical for the reliability of autonomous vehicles.

3. **Quality of Road Markings:** The quality and consistency of lane markings can vary greatly between different regions and even within the same road. Faded, worn-out, or non-standard lane markings can pose a significant challenge for detection systems that rely heavily on visual cues.
4. **Real-Time Processing:** Autonomous vehicles require real-time processing of visual data to make timely decisions. Lane detection algorithms must be computationally efficient to ensure that they can process high-resolution images at high frame rates without causing delays.
5. **Complex Road Scenarios:** Urban environments often present complex road scenarios with multiple lanes, intersections, pedestrian crossings, and road signs. Lane detection systems must be capable of accurately interpreting these scenarios and distinguishing between different types of lane markings.
6. **Occlusions:** Objects such as other vehicles, pedestrians, and roadside infrastructure can occlude lane markings, making it difficult for detection algorithms to maintain continuous lane tracking. Effective handling of occlusions is essential for maintaining the integrity of lane detection.

1.3.3 Motivation

Motivated by the need for improved lane detection accuracy and reliability, our research focuses on enhancing existing lane detection systems by leveraging advanced machine learning algorithms and sensor technologies. By addressing the limitations of current lane detection architectures, we aim to develop a system that can effectively detect lane markings under diverse environmental conditions, ensuring the safety and efficiency of autonomous vehicles on the road.

Through our proposed methodology, we seek to optimize the performance of lane detection systems by integrating cutting-edge technologies and innovative approaches. By enhancing the accuracy, speed, and robustness of lane detection algorithms, we aim to contribute to the advancement of autonomous driving technology and pave the way for more reliable and safer auto-driving vehicles in the future. Our approach to lane detection draws inspiration from the advancements in the CLRNet and FSPnet decoupled model. Particularly, our methodology aligns with the research discussed in [10], as we configure our network as a Proposed Model to incorporate cutting-edge features like high accuracy, fast processing, and robust performance. Moreover, the Proposed Model ensures efficient resource use by scaling through the integration of an optimally configured and fine-tuned CLRNet model for detecting lanes. Previous studies, such as, have tested the performance of CLRNet on restricted resources, which may not accurately reflect the performance in a more complex and realistic environment. Therefore, our proposed methodology aims to address these limitations by testing the performance of CLRNet in a more realistic and diverse environment,

and by using vision transformers and adding more to the architecture of the network. Thus, we introduce our approach to address these deficiencies.

1.4 Research Objectives and Contributions

This thesis aims to advance the performance of automated lane detection systems using deep learning techniques, particularly under challenging conditions such as poor lighting, adverse weather, and occlusions. The research focuses on the following main objectives:

1. **Analyze the current state-of-the-art architectures:** Examine the strengths and limitations of existing lane detection models, particularly those employing convolutional neural networks (CNNs) and vision transformers. This involves a thorough review of literature and current methodologies, identifying key challenges and potential areas for improvement [13].
2. **Design a new, enhanced architecture:** Develop a novel lane detection architecture that integrates advanced deep learning techniques such as Vision Transformers (ViTs) and Feature Shrinkage Pyramid Networks (FSPNet). The goal is to enhance accuracy, robustness, and scalability in lane detection applications, addressing the specific challenges identified in the current architectures [12], [11].
3. **Evaluate the proposed architecture:** Implement and rigorously test the new architecture on benchmark datasets such as CULane and TuSimple. The eval-

uation will consider various performance metrics, including accuracy, precision, recall, F1-score, and processing speed, under diverse environmental conditions [14], [4].

Recent developments in Advanced Driver-Assistance Systems highlight the potential of offering lane detection as a standalone service [15][16]. This approach is gaining traction among leading automotive companies, including Tesla and other companies are exploring modular and scalable solutions for autonomous driving using techniques like Transformers, Bird’s Eye View (BEV) representations, and Occupancy Networks, which are exploring modular and scalable solutions for autonomous driving [17]. Despite these advancements, significant research gaps remain. For instance, many studies have shown limitations when models are tested in isolated environments, which do not adequately represent the complexity of real-world urban driving conditions [15], [14].

The primary challenges in lane detection include:

- **Varying lighting conditions:** From bright sunlight to low-visibility scenarios like rain or fog, lighting variations significantly affect the performance of lane detection systems.
- **Adverse weather conditions:** Certain weather conditions such as rain, fog, and snow may obscure lane markings and introduce visual noise, challenging the robustness of detection algorithms.[18].
- **Quality of road markings:** Variability in lane marking quality, including worn or non-standard markings, complicates detection efforts [19].

- **Real-time processing:** Achieving high frame rates without sacrificing accuracy is crucial for the practical deployment of lane detection systems in autonomous vehicles [20].
- **Complex road scenarios:** Urban environments present diverse challenges such as multiple lanes, intersections, and occlusions from other vehicles and pedestrians ¹.

To address these issues, our research proposes the following contributions:

1. **Development of a highly accurate lane detection system:** Tailored for urban environments, this system leverages state-of-the-art deep learning models to improve detection accuracy under varied conditions.
2. **Comprehensive evaluation using modern benchmarking suites:** Our evaluation framework simulates a wide range of real-world city driving scenarios, providing a robust assessment of the proposed model’s performance [14], [4].
3. **Insights and recommendations:** We provide detailed analyses and recommendations for enhancing lane detection systems in autonomous vehicles, contributing valuable knowledge to the field of autonomous driving technology.

Through these objectives and contributions, this thesis aims to push the boundaries of current lane detection technology, ensuring safer and more reliable autonomous vehicle navigation.

1.5 Outline of Thesis

This thesis is organized into five chapters, each addressing different aspects of the research on lane detection for autonomous vehicles using deep learning strategies.

Below is an outline of each chapter:

- **Chapter 1: Introduction** - This chapter introduces the fundamental concepts of lane detection, its significance in autonomous driving, current challenges, and the motivation behind the research. It also provides an overview of the objectives and contributions of the thesis.
- **Chapter 2: Related Work** - This chapter reviews existing literature on lane detection techniques, including both traditional image processing methods and modern machine learning approaches. It highlights the advancements in the field, identifies gaps in current research, and discusses the relevance of these studies to the proposed work.
- **Chapter 3: ViTCLRNet Proposed Model** - This chapter details the design and implementation of the ViTCLRNet model. It covers the architecture of the model, the scope of the experiment, loss function, metrics, and preprocessing techniques. The chapter provides a comprehensive description of how Vision Transformers and other components are integrated into the lane detection system.
- **Chapter 4: Performance Evaluation** - This chapter presents the results of the experiments conducted using the ViTCLRNet model. It evaluates the

performance of the model on the CULane and TuSimple datasets using various metrics. The chapter includes detailed discussions on the results, comparisons with other models, and visualizations to illustrate the findings.

- **Chapter 5: Conclusion and Future Work** - The final chapter summarizes the key findings of the research, discusses the contributions of the thesis, and outlines potential directions for future work. It emphasizes the importance of continuing research in lane detection and suggests improvements and new approaches to enhance the robustness and accuracy of lane detection systems.

Each chapter builds upon the previous ones, providing a logical progression from the introduction of concepts to the detailed analysis and discussion of results. This structure ensures a clear and comprehensive presentation of the research work conducted in this thesis.

CHAPTER 2

RELATED WORK

Lane detection is a challenging task in computer vision, especially in complex and diverse driving scenarios. Traditional lane detection methods heavily rely on hand-crafted features and complex post-processing steps [21] [22] [23]. Recently, deep learning-based approaches have achieved remarkable progress in lane detection, benefiting from their strong feature learning capabilities [20].

One of the pioneering deep learning-based lane detection methods is the Spatial CNN (SCNN) proposed by [3]. SCNN introduces a novel convolution operation that effectively captures spatial dependencies in lane markings. It significantly outperforms traditional methods and demonstrates the potential of deep learning for lane detection. Subsequent studies have further explored CNN-based architectures for lane detection. Chen et al. [24] proposed a densely connected network (DeepLab), which leverages dense skip connections to enhance feature propagation and improve the accuracy of lane segmentation. He et al. [13] investigated the use of deep residual networks (ResNet) for lane detection. ResNet’s skip connections and residual learning mechanism enable it to train deeper networks with improved accuracy.

Lane detection is a crucial component in autonomous driving systems (ADS) and advanced driver assistance systems (ADAS), involving identifying and interpreting lane markings on roads. The challenge in lane detection is compounded by varying environmental conditions such as lighting, weather, and road surfaces. Lane detection techniques have evolved significantly over the years, starting with traditional computer vision methods that relied on edge detection and the Hough Transform to identify lane markings [7]. This review focuses on recent advancements in lane detection technologies, an essential aspect of autonomous vehicle systems. Research on lane detection methods can be split into two main categories. The first category assesses various lane detection algorithms under various operational conditions, whether in similar or distinct driving environments. The second category contrasts these methods based on their integration and performance within a single deployment, whether mounted in individual vehicles or implemented as part of broader traffic control systems. This section explores recent academic contributions to computer vision for lane detection, focusing on environmental and traffic conditions. The aim is to systematically identify and highlight potential research voids or unexplored avenues within the current body of literature, paving the way for future investigative efforts. Lane detection is critical to autonomous driving systems, enabling vehicles to perceive and navigate roads safely. Traditional methods for lane detection, such as Canny edge detection and Hough transform, have been widely used [25]. However, recent advancements in deep learning techniques have shown promising results in improving the accuracy and robustness of lane detection systems [26].

2.1 Overview of Lane Detection Techniques

Lane detection algorithms have come a long way from the days of basic computer vision approaches that recognized lane markers using edge detection and the Hough Transform [7]. However, these techniques were not able to deal with complicated situations including changing weather and illumination, fading or absent lane markers, and the presence of other objects in the picture [7]. Three primary types of algorithms are employed in deep learning-based lane detection techniques: segmentation-based, anchor-based, and parameter-based [27] [28] [3] [29]. These approaches have gained popularity in recent years.

Lane detection is approached using segmentation-based techniques, which translate pixel values to classes that represent individual lanes or the backdrop [27]. Anchor-based techniques imitate object detection by employing anchors for line detection; algorithms such as LaneATT generate anchors based on attention [28]. Models such as PolyLaneNet are among the best in this category of parameter-based techniques, which regress the polynomial equations of lines directly [30]. Lane detection is essential to autonomous driving systems because precise lane marking detection and tracking is a fundamental component of these systems [31] [7]. It's crucial to make sure that vehicles stay in their designated lanes and don't run into anything on the road. As autonomous driving technology continues to evolve, lane detection techniques are also evolving to 3D, with the goal of avoiding lines in the sky and on regions where there won't be lines. This will help to create a more accurate and reliable autonomous driving experience [32].

2.2 Definition of Autonomous Driving According to the Society of Automotive Engineers (SAE)

The Society of Automotive Engineers (SAE) categorizes autonomous driving by defining different levels of vehicle automation, which extend from Level 0 (no automation) to Level 5 (complete automation) [33]. According to the SAE, an autonomous vehicle operates at Level 4 or Level 5. At Level 4, the vehicle can handle all driving functions under certain conditions without any human intervention, but it might be limited to specific areas or environments. At Level 5, the vehicle achieves complete autonomy and can operate under all conditions without any human input. These levels clarify the capabilities and limitations of autonomous vehicles, guiding manufacturers, policymakers, and the public in understanding the operational scope and safety requirements of autonomous technology.

Figure 2.1 illustrates the different levels of automation defined by the SAE.

2.3 Lane Detection Methods

Lane detection is a critical component of autonomous driving systems, enabling vehicles to maintain proper lane positioning and navigate complex driving environments. Various techniques have been proposed and implemented over the years, each with its strengths and limitations. Zaidi et al. [1] provides a thorough analysis of various lane detection techniques, including color thresholding, gradient-based methods, edge detection, Hough transform, and convolutional neural networks (CNN). Their study



SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: [sae.org/standards/content/j3016_202104](https://www.sae.org/standards/content/j3016_202104)

Copyright © 2021 SAE International. The summary table may be freely copied and distributed AS-IS provided that SAE International is acknowledged as the source of the content.

		SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?		You are driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You are not driving when these automated driving features are engaged – even if you are seated in “the driver’s seat”		
		You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
Copyright © 2021 SAE International.							
		These are driver support features			These are automated driving features		
What do these features do?		These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/acceleration support to the driver	These features provide steering AND brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
	Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering OR • adaptive cruise control	• lane centering AND • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

Figure 2.1: SAE - J3016: Table of Levels of Automation

concludes that CNN algorithms outperform traditional methods in terms of accuracy and flexibility, especially under challenging conditions such as rainy weather, fog, and night-time driving. The key findings of their survey are summarized in Table 2.1.

2.4 Autonomous Driving Datasets

High-quality datasets are fundamental for developing reliable autonomous driving algorithms. These datasets provide the necessary training and evaluation data for various tasks, including lane detection, object detection, and semantic segmentation. Liu et al. [34] presents an exhaustive study of 265 autonomous driving datasets,

Table 2.1: Summary of Lane Detection Methods [1]

Method	Advantages	Disadvantages
Color Thresholding	Simple, efficient	Sensitive to lighting changes
Gradient-Based	Robust to lighting	Sensitive to noise
Edge Detection	Accurate edge detection	Computationally intensive
Hough Transform	Detects straight lines well	Limited performance on curves
CNN	High accuracy, flexible	Requires large datasets, computationally expensive

evaluating them based on sensor modalities, data size, tasks, and contextual conditions. They introduce a novel impact score metric to assess the influence of datasets, highlighting the importance of high-quality data for developing robust autonomous driving systems. Their analysis also emphasizes the need for standardized annotation processes to improve dataset quality and reliability. The primary contributions of [34] are summarized as follows:

- Comprehensive survey of 265 autonomous driving datasets.
- Introduction of an impact score metric to evaluate dataset significance.
- Analysis of annotation processes and existing labeling tools.
- Discussion on the impact of geographical and environmental conditions on dataset performance.

The survey by [34] provides valuable insights into the characteristics and quality of existing datasets, guiding future research and dataset development efforts.

2.5 Deep Learning Methods for Lane Detection

Deep learning methods, particularly convolutional neural networks (CNN), have shown great promise in lane detection. These methods leverage large amounts of annotated data to learn complex features and patterns, enabling accurate lane detection even in challenging conditions. Zaidi et al. (2024) provide a comparative analysis of various lane detection techniques, highlighting the superior performance of CNNs in diverse conditions. This aligns with our approach of using a Vision Transformer Backbone (ViT-B16) for lane detection. The robustness and flexibility of CNNs make them well-suited for integration into advanced driver assistance systems (ADAS) and autonomous driving applications.

2.6 Traditional Lane Detection Techniques (Pre-Deep Learning)

Before the advent of deep learning, lane detection techniques primarily relied on traditional image processing methods. These methods often included edge detection algorithms such as the Canny edge detector and the Hough Transform to identify lane markings on roads.

2.6.1 Edge Detection Methods

Edge detection methods are fundamental to traditional lane detection. Several techniques have been developed over the years, each with its own strengths and weaknesses.

The Canny edge detector is one of the most widely used techniques due to its ability to detect edges with high precision by identifying the intensity gradients in an image [35]. This method involves several steps, including noise reduction using a Gaussian filter, finding the intensity gradient, applying non-maximum suppression, and tracing edges through hysteresis. Despite its effectiveness, the Canny edge detector can struggle with noisy images and complex road conditions.

Another commonly used method is the Sobel operator, which uses convolution masks to calculate the gradient magnitude and direction [36]. This method is simpler and faster than the Canny detector but is less effective in detecting fine details and is more sensitive to noise.

The Prewitt operator is similar to the Sobel operator but uses different convolution masks [37]. While it also detects edges by calculating gradient magnitude and direction, it is less accurate than the Sobel operator in determining edge direction but has comparable performance in noise sensitivity.

The Laplacian of Gaussian (LoG) method is another edge detection technique that uses a Gaussian filter to smooth the image before applying the Laplacian operator to detect edges [38]. This method is effective in identifying edges with high curvature but can be computationally intensive.

2.6.2 Hough Transform

The Hough Transform is another key method used in traditional lane detection. It transforms points in the image space into a parameter space, where lines can be detected by finding intersections of sinusoidal curves. This technique is particularly

effective for detecting straight lines but can be less effective for detecting curved or broken lane markings [39].

The Standard Hough Transform (SHT) maps points in the image space to lines in the parameter space, which can be computationally intensive for larger images. The Probabilistic Hough Transform (PHT) is a more efficient variant that uses a random subset of points to detect lines, reducing computational complexity while maintaining accuracy [40].

The Generalized Hough Transform (GHT) extends the standard Hough Transform to detect arbitrary shapes by using a template matching approach [41]. This method is useful for detecting complex lane markings that deviate from simple geometric shapes.

2.7 Deep Learning-Based Lane Detection Techniques

Deep learning, focusing on Convolutional Neural Networks (CNNs), has significantly advanced lane detection. This domain has two main deep learning approaches: segmentation-based and anchor-based.

2.7.1 Segmentation-Based Methods

Lane identification is approached as a pixel-by-pixel segmentation job via segmentation-based techniques like Spatial CNN (SCNN). SCNN dramatically improves performance in difficult conditions by introducing a unique convolution technique that captures spatial interdependence along lane markers [3].

2.7.2 Anchor-Based Methods

Anchor-based methods, exemplified by the Ultra Fast Structure-aware Deep Lane Detection (UFLD) model, focus on predefined anchor points or boxes. These methods generate anchors for potential lane locations and refine them using CNNs, enhancing computational efficiency and accuracy [4].

2.7.3 Lane Classification and Mixed Precision Techniques

Lane classification, essential for interpreting road configurations, has received less attention than lane detection. The LVLane project investigates lane classification alongside the impact of mixed precision training on models [26].

2.7.4 Novel Approaches and Challenges

- **LVLane:** Introduces a dataset for challenging scenarios such as curved lanes and heavy occlusions [26].
- **CLRNet:** Proposes a network utilizing global and local features for enhanced accuracy in lane detection [10].
- **FOLOLane:** Focuses on local patterns for predicting global structures in lane markers [42].
- **Contrastive Learning:** Emphasizes the potential of leveraging contrastive learning techniques in lane detection [43].
- **Line-CNN:** Introduces the Line Proposal Unit for traffic line detection [29].

- **Vision-Based Robust Lane Detection:** Stresses the importance of adaptive methodologies under challenging conditions [44].

The literature on lane detection reflects a progression toward deep learning-based models that are accurate, efficient, and adaptable. As autonomous vehicle technology advances, sophisticated lane detection systems become increasingly crucial. These advancements collectively contribute to the field by introducing novel methodologies and models that enhance lane detection accuracy, particularly in challenging scenarios. The emphasis on improving confidence scores, local geometry detection, spatial information aggregation, and deep-learning-based label assignments signifies a significant technological leap in lane detection systems.

2.8 Camouflaged Object Detection

Camouflaged object detection is a more challenging task compared to lane detection, as camouflaged objects are often difficult to distinguish from their backgrounds. Traditional methods for camouflaged object detection typically rely on low-level features and handcrafted rules [45]. Recent advances in deep learning have led to the development of more sophisticated methods for camouflaged object detection. Fan et al. [46] proposed the Camouflaged Object Detection (COD) benchmark and a deep learning-based method called SINet for COD. SINet utilizes a semantic segmentation network to generate an initial objectness map, which is then refined through a series of post-processing steps. Subsequent studies have proposed various deep learning architectures and techniques to improve COD performance. For example, [47] proposed the

uncertainty-guided transformer reasoning (UGTR) model, which leverages transformers and uncertainty estimation to enhance the detection of camouflaged objects. Zhai et al. [48] proposed the mutual graph learning (MGL) model, which utilizes graph neural networks to capture the relationships between objects and their backgrounds.

2.9 Vision Transformers

Vision transformers (ViTs) have emerged as a powerful tool for computer vision tasks, offering advantages in capturing long-range dependencies compared to traditional convolutional neural networks (CNNs). In the context of lane detection, ViTs have been successfully applied to improve the accuracy and robustness of existing methods. Recently, transformers, such as the Laneformer [49] and other transformer-based architectures, have revolutionized traditional lane detection methodologies by effectively capturing the shape and semantic characteristics of lanes while minimizing latency. By incorporating transformer encoder-decoder structures, these models excel in extracting global features and contextual information crucial for accurate lane detection in complex road scenarios. The utilization of deformable pixel-wise self-attention, row and column self-attention operations, and the integration of object instances in multi-head attention blocks enhance the ability of vision transformers to detect lane points efficiently. Notably, recent studies like Laneformer have demonstrated state-of-the-art performance in lane detection tasks, showcasing the potential of vision transformers in enhancing the accuracy and real-time capabilities of autonomous driving systems. The integration of ViTs into lane detection frameworks has demonstrated promising

results, particularly in handling complex lane shapes and challenging conditions. As research continues in this area, we can expect further advancements in the utilization of ViTs for accurate and reliable lane detection systems.

2.10 Feature Pyramid Networks

Feature pyramid networks (FPNs) are a family of deep learning architectures that have been widely used in object detection and segmentation tasks. FPNs construct a pyramid of features at different scales, allowing the network to capture objects of various sizes and shapes. One of the most popular FPNs is the Feature Pyramid Network (FPN) proposed by Lin et al. [50]. FPN builds a feature pyramid by combining features from different levels of a convolutional neural network. It has been successfully applied to various object detection and segmentation tasks and has become a standard component in many deep learning architectures. Subsequent studies have proposed various improvements and extensions to FPNs. For example, [51] proposed the EfficientDet FPN, which reduces the computational cost of FPNs while maintaining their accuracy. Wang et al. [52] proposed the Pyramid Vision Transformer (PVT), which combines FPNs with transformer networks to achieve state-of-the-art performance on various computer vision tasks.

2.11 Challenges and Open Issues in Deep Learning for Lane Detection

While deep learning has significantly improved lane detection techniques, there are still several challenges and open issues that need to be addressed [53][54]. One of the main challenges is handling adverse weather conditions, such as rain, fog, and snow, which can make it difficult to detect lane markings. These conditions can cause reflections, shadows, and other visual artifacts that can interfere with lane detection algorithms. Another challenge is the need for real-time processing, as lane detection algorithms need to be able to process images quickly and accurately to be useful in autonomous driving systems. This requires efficient algorithms that can run on low-power hardware, such as embedded systems. Generalization across diverse scenarios is also a challenge, as current models may not perform well in different geographical and environmental conditions. For example, a model that is trained on data from a particular region may not perform well in a different region with different lane markings or road conditions [53][54]. Finally, integration with other systems is an important issue, as lane detection algorithms need to be able to work seamlessly with other components of autonomous driving systems, such as object detection and collision avoidance. This requires careful design and coordination between different components of the system to ensure that they can work together effectively. In summary, deep learning has made significant progress in lane detection, but there are still challenges and open issues that need to be addressed, including handling adverse weather conditions, real-time processing needs, generalization across diverse scenarios, and integration with other

systems. Addressing these challenges will be key to creating reliable and effective autonomous driving systems.

2.12 Future Directions

As we look towards the future of lane detection technology, there are several exciting opportunities for improvement and innovation [53] [55]. In terms of deep learning techniques, there is potential for advancements in the use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to better identify and track lane markings in real-time. Additionally, the integration of attention mechanisms and transformer architectures could further enhance the accuracy and robustness of lane detection models. Interdisciplinary approaches also hold great promise for advancing lane detection research. For example, insights from the field of optics could be used to develop more sophisticated image processing techniques, such as the use of polarization filters to reduce glare and improve lane visibility. Collaboration with traffic engineering experts could also lead to the development of more intelligent lane detection systems that can adapt to changing road conditions and traffic patterns. By combining the strengths of multiple disciplines, we can create more sophisticated and effective lane detection systems that will improve road safety and reduce the risk of accidents [55]. The studies by [1] [34] highlight several future directions for research in autonomous driving. Enhancing the quality and annotation processes of datasets, exploring new lane detection methods, and improving the robustness of algorithms under various environmental conditions are key areas for future work. Integrating these insights

into our research can further advance the development of reliable and efficient lane detection systems for autonomous vehicles. In conclusion, the comprehensive surveys provided by Zaidi et al. [1] and Liu et al. [34] offer valuable contributions to the field of autonomous driving. Their findings and recommendations serve as a foundation for ongoing research and development, guiding the design and implementation of next-generation lane detection algorithms and datasets.

Table 2.2: A Summary of Retrieved Related Studies in the Literature

Ref	Year	Dataset	Img Rep	Key Methods	Findings	Eval Metrics
[10]	2022	CULane, Tusimple	ResNet backbone	Cross-Layer Refinement, ROIGather	State-of-the-art F1@50 on CULane, Tusimple	F1@50, LIoU loss
[56]	2023	CULane, Curve-Lanes	-	LaneIoU accuracy scoring	Superior results on CULane, CurveLanes	F1 scores, LaneIoU
[57]	2023	CULane	-	Integration of MatchNet into CLRNet	Improved performance in curved lanes	F1 score improvement
[58]	2021	CULane, Tusimple	Spatial relationships	RESA, Bilateral Decoder	State-of-the-art accuracy on CULane, Tusimple	F1, accuracy on Tusimple
[26]	2023	TuSimple, Caltech Lane, LVLane	CNN classification branch	Mixed precision training	State-of-the-art classification on TuSimple	Avg accuracy
[43]	2023	CULane, Tusimple	-	CLLD, random masking	Best in shadows, crowded scenes	Benchmark dataset evaluation
[32]	2022	ONCE-3DLanes	3D lane annotation	Data annotation pipeline, SALAD	Largest 3D lane dataset	Precision, recall, F-score
[59]	2023	CULane, Curve-Lanes, TuSimple	U-shaped guide line	Gaussian mask, adaptive decoder	Top performance, especially in corner lanes	Metrics across datasets, F1, Precision, Recall

CHAPTER 3

VITCLRNET PROPOSED MODEL

In this chapter, Section 3.1 describes the overall architecture of the ViTCLRNet, and Section 3.2 explores the enhanced capabilities achieved by integrating the vision transformers into the Proposed Model.

3.1 Scope of the Experiment

3.1.1 Objective and Goals

The primary objective of this experiment is to assess and confirm the effectiveness and computational viability of the ViTCLRNet deep learning algorithms for lane detection. This involves evaluating the model's performance in various driving scenarios and comparing it against state-of-the-art lane detection models.

3.1.2 Loss Function

The LineIoU [10] loss, which considers each lane line as a single unit to improve localization precision, was the loss function used to train the suggested network. A weighted sum of the regression losses (Smooth-L1 Loss and Line IoU Loss) and the classification loss (Focal Loss) makes up the overall loss function for our model:

$$L_{\text{total}} = w_{\text{cls}}L_{\text{cls}} + w_{\text{xytl}}L_{\text{xytl}} + w_{\text{LIoU}}L_{\text{LIoU}} \quad (3.1)$$

where L_{cls} is the Focal Loss used for the classification task, L_{xytl} represents the Smooth-L1 Loss used for the bounding box regression, and L_{LIoU} is the Line Intersection over Union (IoU) Loss which focuses on regressing the entire lane line structure. The weights w_{cls} , w_{xytl} , and w_{LIoU} control the relative contribution of each loss component to the total loss.

The Line Intersection over Union (LIoU) loss used in the CLRNet model for lane detection is a specialized form of IoU calculation tailored for lines, particularly useful for continuous object boundaries like lanes. Here's a detailed breakdown of how it's calculated:

LIoU Formula

Based on the idea of Intersection over Union (IoU), the LIoU has been tailored particularly for line segments. The formula for LIoU is defined as:

$$\text{LIoU} = \frac{\sum_{i=1}^N d_{O_i}}{\sum_{i=1}^N d_{U_i}} \quad (3.2)$$

Where:

- N is the number of points defining the line.
- d_{O_i} is the length of overlap between the predicted line segment and the ground truth for the i th segment.
- d_{U_i} is the union of the predicted line segment and the ground truth for the i th segment.

Extended Line Segments

Each point on the predicted lane, x_{p_i} , is extended by a fixed radius e to form a line segment:

- Extended Predicted Segment: $[x_{p_i} - e, x_{p_i} + e]$
- Extended Ground Truth Segment: $[x_{g_i} - e, x_{g_i} + e]$

Intersection and Union Calculation

For each extended segment i :

- Intersection d_{O_i} :

$$d_{O_i} = \max(0, \min(x_{p_i} + e, x_{g_i} + e) - \max(x_{p_i} - e, x_{g_i} - e)) \quad (3.3)$$

- Union d_{U_i} :

$$d_{U_i} = \max(x_{p_i} + e, x_{g_i} + e) - \min(x_{p_i} - e, x_{g_i} - e) \quad (3.4)$$

The intersection calculation includes a max operation with zero to ensure that the intersection length cannot be negative (which can occur if the segments do not overlap).

Loss Calculation

The LIoU loss is then calculated as:

$$L_{LIoU} = 1 - \text{LIoU} \quad (3.5)$$

This means the loss is zero when the prediction perfectly matches the ground truth (LIoU = 1), and increases as the match worsens. This LIoU loss encourages the model to predict lanes that closely match the entire extent of actual lane markings, rather than fitting individual points independently. It is particularly useful for tasks like lane detection where the spatial continuity and alignment of predictions are crucial.

3.1.3 Metrics

- Precision: $\frac{TP}{TP+FP}$
- Recall: $\frac{TP}{TP+FN}$
- F1-Score: $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

For evaluating our model, we use the F1-score for both the CULane [14] and TuSimple datasets. The F1-score is calculated based on the Intersection-over-Union (IoU) between the predicted and actual lane markings. Predictions with an IoU exceeding 0.5 are deemed accurate, representing true positives. The formula for the

F1-score is:

$$F1 = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Precision} + \text{Recall}}. \quad (3.6)$$

In addition to the standard F1-score, we used the modified F1 (mF1) metric, inspired by the COCO [60] and used in CLRnet [10] benchmark, to enhance the assessment of localization accuracy across different models. This metric is defined as:

$$mF1 = \frac{(F1threshold@50 + F1threshold@55 + F1threshold@60 + \dots + F1threshold@95)}{10}, \quad (3.7)$$

where each component, such as F1@50, represents the F1-score at different IoU thresholds ranging from 0.5 to 0.95.

For the Tusimple dataset, we compute accuracy using:

$$\text{Accuracy} = \frac{\sum_{\text{clip}} C_{\text{clip}}}{\sum_{\text{clip}} S_{\text{clip}}}, \quad (3.8)$$

where C_{clip} and S_{clip} denote the counts of correctly predicted points and the total number of ground truth points per image, respectively. A prediction is considered correct if over 85% of its points lie within 20 pixels of the actual lane markings. Additionally, the dataset provides rates for false positives (FP) and false negatives (FN), defined by the ratios $FP = \frac{F_{\text{pred}}}{N_{\text{pred}}}$ and $FN = \frac{M_{\text{pred}}}{N_{\text{gt}}}$.

3.2 Design of the Experiment

3.2.1 Context Selection

An offline experiment was conducted on a deep learning technique using the CULane [14] and TuSimple datasets.

3.2.2 CULane Dataset

The CULane dataset is a treasure trove for researchers working on computer vision algorithms for autonomous vehicles. It boasts over 133,000 images extracted from 55 hours of video footage captured by cameras mounted on various vehicles driving throughout Beijing. This rich dataset offers a diverse range of scenarios for training and testing lane detection algorithms. It contains 100,000 images for train, validation, and test sets, each with a resolution of 1640x590 pixels and categorized into nine challenging types such as crowded and night scenes. To ensure the accuracy of lane markings, even in challenging situations, human annotators meticulously labeled each image using cubic splines. This meticulous process extends to instances where lane markings are hidden by vehicles or completely absent, forcing the annotators to rely on context to infer their location. CULane further enhances its value by splitting the data into designated training, validation, and test sets. The test set is particularly noteworthy as it's subdivided into "normal" and "challenging" categories, allowing researchers to evaluate their algorithms under a variety of conditions. This comprehensive and meticulously labeled dataset is a powerful tool for researchers pushing the boundaries of lane detection technology, a crucial component for autonomous vehicles

and advanced driver-assistance systems [14].

3.2.3 TuSimple Dataset

The TuSimple dataset is a valuable resource for researchers and developers working on computer vision tasks related to self-driving cars. It contains 6,408 high-resolution (1280x720) images captured by cameras mounted on vehicles traveling along US highways. This diversity in location provides a broad range of real-world driving scenarios for training algorithms. What truly sets TuSimple apart is the inclusion of images showcasing different weather conditions. With clear skies, rain, fog, and even challenging lighting situations represented, developers can train their models to handle the complexities of navigating roads in various environments. This focus on real-world applicability makes the TuSimple dataset a crucial tool for advancing self-driving car technology [61].

3.3 Preprocessing Techniques

Preprocessing is a critical step in enhancing the performance of lane detection models by ensuring the data is optimized for training. For both the CULane and TuSimple datasets, common preprocessing techniques include:

- **Image Resizing and Normalization**
 - Images were resized to $\text{img_h} = 384$ and $\text{img_w} = 384$ to ensure uniformity and consistency in the input data fed into the network.

- Image normalization techniques were applied to scale the pixel values appropriately, typically scaling the pixel values to a range suitable for the neural network, between 0 and 1. This enhances the model’s convergence during training.

- **Undistortion**

- Images were undistorted using tools like the one described by [62], which corrects lens distortion, ensuring that lane lines appear as straight as possible in the images.

- **Data Augmentation**

- To improve model robustness and prevent overfitting, data augmentation techniques such as rotation, flipping, and scaling are employed. This helps the model generalize better to different lighting conditions, occlusions, and physical road variations.

- * **Horizontal Flipping** - Images are randomly flipped horizontally to mimic lanes from both directions of traffic, enhancing the model’s ability to generalize.

- * **Rotation** - Small random rotations are applied to simulate the effect of slight tilts and variations in the camera angle.

- * **Scaling** - Random scaling is applied to images to simulate different distances from the camera to the lanes, helping the model learn to detect lanes at varying distances.

- * **Brightness and Contrast Adjustment** - Random adjustments to brightness and contrast are made to simulate different lighting conditions, such as bright sunlight or dim conditions.
 - * **Affine Transformations** - Random affine transformations, including translation, rotation, and scaling, are applied to simulate different spatial orientations of the lanes and camera viewpoints.
- **Region of Interest (ROI) Selection**
 - Especially for the CULane dataset, selecting the region of interest involves cropping the images to focus more on the lane markings and less on the surrounding environment, which might contain noise such as vehicles and pedestrians.

3.4 Examples of Preprocessing and Augmentation

The figures below 3.1 3.2 illustrate examples from the CULane and TuSimple datasets, showing the original images, preprocessed images, and augmented images. These figures demonstrate the application of resizing, normalization, rotation, flipping, and cropping techniques to optimize the data for training.

3.4.1 Original and Augmented Images from CULane and TuSimple Dataset

The sample images for the CULane dataset are presented in Figure 3.1, while examples from the TuSimple dataset can be seen in Figure 3.2. In both figures, the left column displays the original images, and the right column shows the corresponding augmented images.

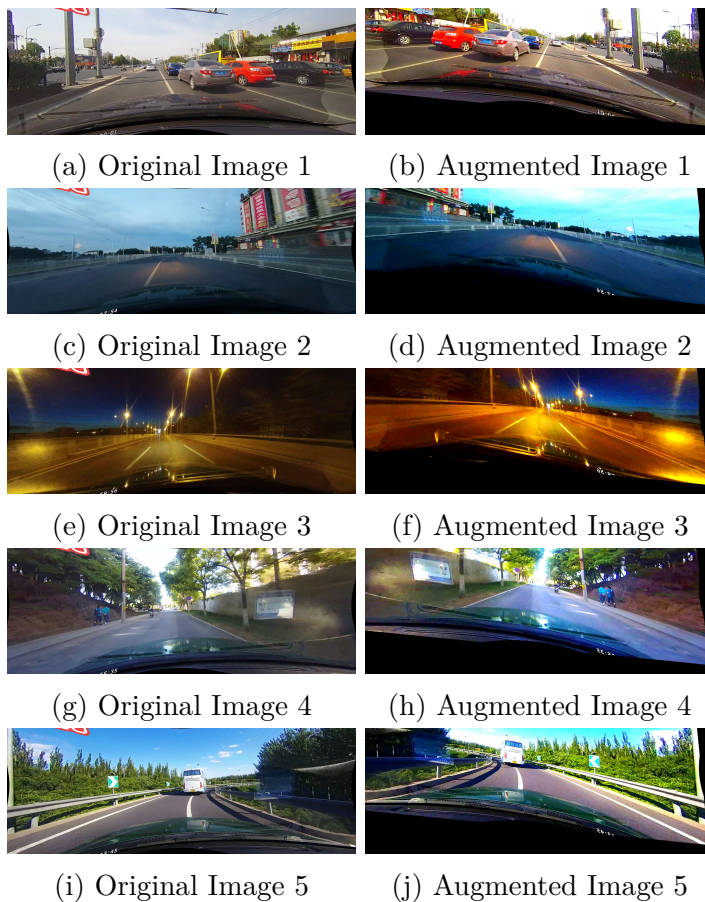
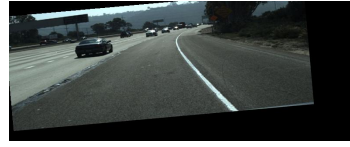


Figure 3.1: Original, Preprocessed, and Augmented Images from CULane Dataset



(a) Original Image 1



(b) Augmented Image 1



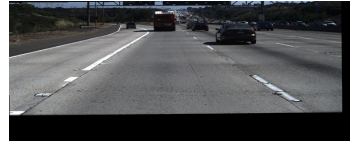
(c) Original Image 2



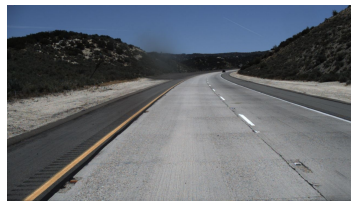
(d) Augmented Image 2



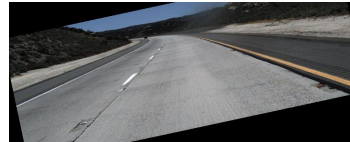
(e) Original Image 3



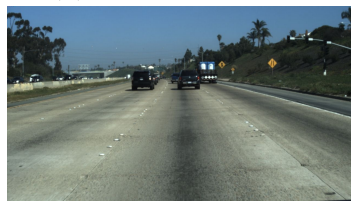
(f) Augmented Image 3



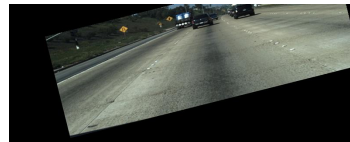
(g) Original Image 4



(h) Augmented Image 4



(i) Original Image 5



(j) Augmented Image 5

Figure 3.2: Original, Preprocessed, and Augmented Images from TuSimple Dataset

3.5 Non-Local Token Enhancement Module (NL-TEM)

Before explaining the model architecture, we need to understand the Non-Local Token Enhancement Module (NL-TEM), which is designed to address the limitations of vision transformers in capturing local features crucial for accurately detecting camouflaged objects [11]. Here’s a breakdown of its functionality:

- **Enhancing Local Feature Representation:** NL-TEM operates on neighboring tokens, representing a local region of the image, to strengthen their feature representation. This is important because camouflaged objects often blend seamlessly with their surroundings, making it difficult to distinguish them based on global context alone.
- **Non-Local Operations for Adjacent Token Interaction:** NL-TEM utilizes a non-local operation to enable interaction and feature aggregation between adjacent, similar tokens. This aggregation of local camouflage clues helps to highlight the subtle differences between the camouflaged object and its background.
- **Graph Convolution Network for High-Order Relations:** A Graph Convolution Network (GCN) is employed to explore higher-order semantic relationships between pixels within tokens. This process aims to identify subtle, discriminative features that might be missed by considering only low-order relations. The GCN essentially treats groups of pixels with similar features as vertices in a graph, learning the relationships between these vertices through

information propagation.

In essence, NL-TEM combines non-local operations and graph convolutions to capture and enhance both the local interactions and higher-order relationships within tokens, thereby improving the representation of local features essential for camouflaged object detection.

3.6 Model Architecture Description

The model architecture is illustrated in the figure below. It begins with an image input, followed by data augmentation. The augmented image is linearly mapped and processed through a Vision Transformer (ViT) backbone with multiple transformer layers. The output from the ViT is enhanced by the Non-Local Token Enhancement Model (NL-TEM) and then processed through a Feature Pyramid Network (FPN) neck. The FPN produces multiple feature maps of different sizes, which are further processed by the ROIGather module of the CLRNNet architecture. The loss is calculated, and the process iterates through these steps until training is complete.

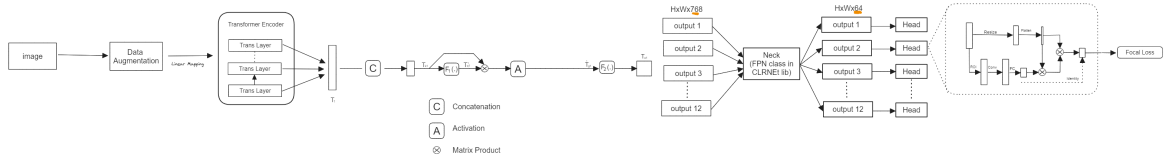


Figure 3.3: Model Architecture Diagram

The model starts with an image input, which undergoes data augmentation to enhance the diversity and robustness of the training data. Augmentation techniques such as rotation, scaling, and color adjustments help the model generalize better to

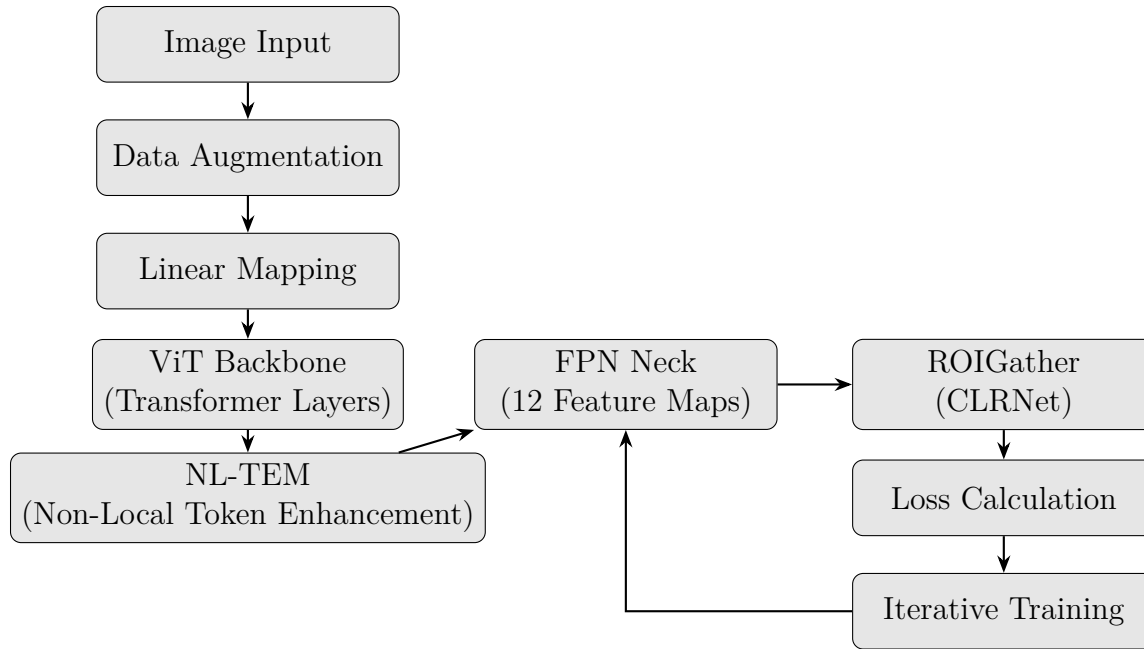


Figure 3.4: Model Architecture Diagram

unseen data.

The augmented image is linearly mapped and fed into a Vision Transformer (ViT) backbone. The ViT processes the input through multiple transformer layers, leveraging self-attention mechanisms to capture complex patterns and relationships within the image.

The output from the ViT backbone is further enhanced by the Non-Local Token Enhancement Model (NL-TEM), part of the FSPNet architecture. The NL-TEM improves feature representation by focusing on non-local dependencies, crucial for accurate lane detection.

The enhanced features are processed through the Feature Pyramid Network (FPN) neck, producing multiple feature maps of different sizes. The FPN neck captures multi-scale features, essential for detecting lanes at various scales and perspectives.

The feature maps are then processed by the ROIGather module in the CLRNet

architecture, which gathers region of interest features critical for lane detection. The model calculates the loss based on the predicted and ground truth lane markings, guiding the learning process.

The process iteratively continues, with the FPN neck refining features, loss calculation guiding learning, and backpropagation updating the model parameters until training completion. This iterative training helps the model progressively improve its performance.

3.7 Hardware Specifications

All experiments were conducted on a machine provided by the SDAIA joint research center, equipped with an NVIDIA A6000 GPU.

3.8 Training Configuration

The training configuration for each experiment is detailed below: All experiments used a learning rate of $3e-4$ for batch size 8. The number of epochs for each dataset and experiment is in Table 3.1.

Table 3.1: Summary of Experiments with Different Backbones and Datasets

Experiment	CULane (epochs)	TuSimple (epochs)
Experiment with ViT as Backbone	15	70
Experiment with DPT	15	70
Experiment with FSPNet	13	112
Experiment with ViTCLRNet	22	113

This structured approach not only helps in presenting our findings clearly but

also showcases the depth of our research and the innovative aspects of our work in enhancing lane detection technology.

CHAPTER 4

PERFORMANCE EVALUATION

4.1 Introduction

In this chapter, we present the results of our lane detection project utilizing different model architectures. We evaluated the performance of these models on two widely used datasets: CULane and TuSimple. The primary focus was on integrating Vision Transformers (ViT) as the backbone with CLRNet, Dense Prediction Transformers (DPT), and Feature Shrinkage Pyramid (FSPNet), and our final model VitClrnet. We report various evaluation metrics and provide a detailed analysis of the results.

4.2 Performance Evaluation Methodology

Evaluating the performance of lane detection models trained on the CULane and TuSimple datasets involves several metrics and procedures:

- **Accuracy Metrics:** For the TuSimple dataset, accuracy is calculated based on the prediction error of the lane line points compared to the ground truth

annotations. The metric is typically the percentage of correctly predicted lane points within a certain pixel distance.

- **Precision, Recall, and F1-Score:** Commonly used to evaluate performance on the CULane dataset, these metrics assess how well the model detects the lane lines without missing or falsely marking non-lane areas. Precision measures the proportion of correctly detected lane pixels among all pixels identified as lane markings by the model, recall measures the model’s ability to detect all actual lanes, which indicates the percentage of actual lane pixels that the model successfully detected. The F1 score combines precision and recall into a single metric, providing a balanced view of the model’s performance.
- **False Positives and False Negatives:** These metrics are crucial for understanding model limitations, particularly in complex scenarios where multiple lane lines are present or absent. The false positive rate measures the frequency of the model incorrectly identifying non-lane pixels as lane markings and the false negative rate indicates how often the model misses actual lane pixels.
- **Frame Processing Speed:** An essential metric for real-time applications, measuring the number of frames processed per second (FPS) evaluates the model’s efficiency and suitability for deployment in real-world scenarios.
- **Visualization of Results:** Besides numerical metrics, visualizing the model’s predictions overlaid on the test images is a standard method for qualitative evaluation. This helps in understanding the model’s performance in practical scenarios, such as different lighting conditions and road types.

By employing these metrics, we can objectively compare the performance of different lane detection models on datasets like CULane and TuSimple. These sections ensure a comprehensive approach to preparing the data and evaluating the performance of lane detection models using the CULane and TuSimple datasets, paving the way for effective model training and deployment.

4.3 Experiments and Results

4.3.1 Experiment with Vision Transformers (ViT) as Backbone in CLRNet

Background

The incorporation of Vision Transformers (ViT) into the CLRNet model aimed to leverage their powerful representational capabilities. ViTs have been successful in various vision tasks due to their ability to model global dependencies.

Implementation Details

We integrated a pre-trained Vision Transformer (ViT) model [12], specifically ViT-B-16, into our CLRNet architecture. The model was fine-tuned to work with the CULane and TuSimple datasets. The input images were resized to 224 x 224 color images, and the first four encoder layers of the ViT-B-16 model were used, resulting in an output shape of 114 x 114 x 768. Despite the use of this pre-trained model, the results were suboptimal. We also experimented with the ViT-L-32 checkpoint to

explore potential improvements.

The ViT-b-16 pretrained model was integrated as the backbone with its weights fine-tuned on the CULane and Tusimple datasets. The model processed 224x224 color images, outputting features from the first four encoder layers with an output shape of 114x114x768.

Results on CULane Dataset

The performance of the ViT-B-16 model on the CULane dataset is depicted in Figure 4.1. The graph illustrates the model’s ability to detect lanes, but the performance metrics indicate room for improvement. For the "Cross" category, only the false positives are displayed and the result was good compared to other models with a value of 1220.

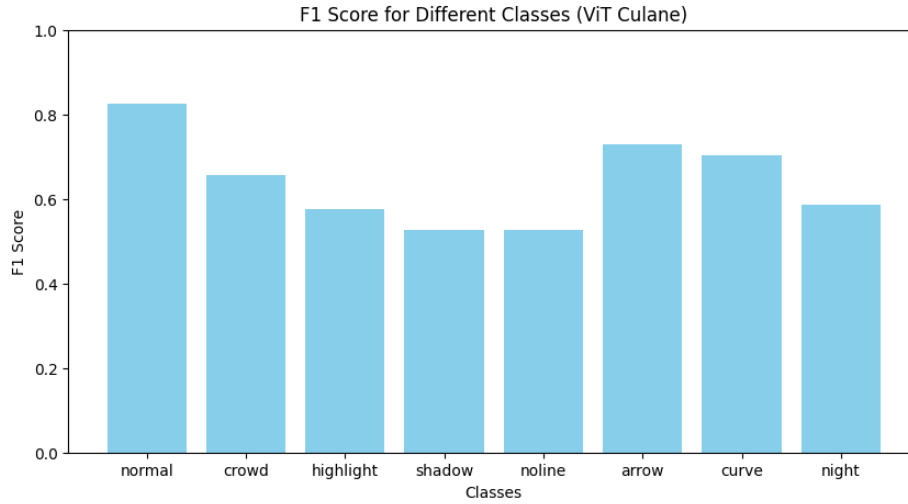


Figure 4.1: Result graph for the CULane Dataset using ViT as a backbone.

Metric	mF1	F1@50	FPS	GFlops
Value	52.06	74.12	172	24.02

Table 4.1: Performance Metrics on CULane Dataset using ViT

Results on TuSimple Dataset

Similarly, the results on the TuSimple dataset, shown in Table 4.2, reflect the challenges faced when using the ViT-B-16 model. The model struggled to achieve high accuracy, precision, and recall compared to state-of-the-art methods.

Metric	F1	ACC	FP	FN
Value	94.62	95.54	3.43	2.85

Table 4.2: Performance Metrics on TuSimple Dataset using ViT

Discussion

The experiment integrating Vision Transformers (ViT) as the backbone in the CLRNet architecture highlighted several key points:

1. **Global Dependency Modeling:** ViTs effectively captured global dependencies, beneficial for contextual understanding across the entire image. However, this added computational complexity.
2. **Performance Metrics:** The ViT-B-16 model showed moderate performance on the CULane dataset with an mF1 score of 52.06 and an F1@50 of 74.12. These metrics indicate the model’s ability to detect lanes, but there’s room for improvement, especially in recall.
3. **False Positives and Negatives:** The false positive count in the ”Cross” category was 1220, showing the model’s potential in certain scenarios but highlighting issues in complex scenes, resulting in missed detections (false negatives).

4. **Computational Efficiency:** The model processed 172 FPS with 24.02 GFlops, making it suitable for real-time applications. However, the computational demands of ViTs, especially with larger input resolutions, can limit their deployment in resource-constrained environments.
5. **Visualization Insights:** Visual analysis revealed the model’s effectiveness under clear conditions but struggles with occlusions, varying lighting, and complex road scenarios. This indicates a need for enhancements in post-processing or integrating complementary models.

In conclusion, using Vision Transformers as a backbone in the CLRNet architecture shows promise but requires further optimization to balance accuracy and computational efficiency.

4.3.2 Experiment with Dense Prediction Transformers (DPT)

Background

Following the initial experiments, the Dense Prediction Transformer (DPT) model was selected for its capabilities in dense prediction tasks, which is crucial for accurate lane detection.

Implementation Details

To address the limitations observed with ViT, we implemented Dense Prediction Transformers (DPT). We used the pre-trained model Intel/dpt-large and fine-tuned

it for our specific task. The input images were resized to 384 x 384 color images. The output shape from the DPT model was 24 (height) x 24 (width) x 1024 (number of channels).

Results on CULane Dataset

The DPT model demonstrated significant improvements over the ViT-B-16 model. The results on the CULane dataset are shown in Figure 4.2. Additionally, Table 4.3 presents a comparative analysis of the performance metrics, including mean F1 score (mF1), F1 score at 50% Intersection over Union (IoU) threshold (F1@50), Frames Per Second (FPS), and Giga Floating Point Operations (GFlops). For the "Cross" category, it was 1057.

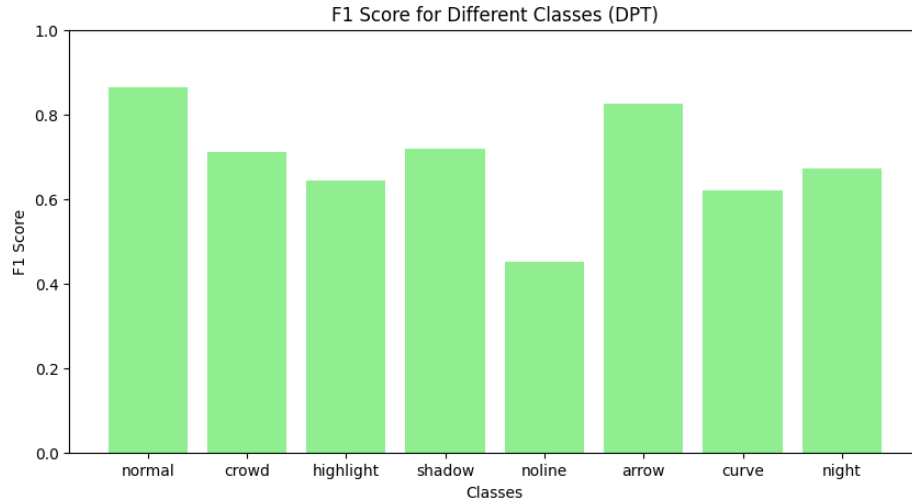


Figure 4.2: Result graph for the CULane Dataset using DPT as a backbone.

Metric	mF1	F1@50	FPS	GFlops
Value	51.42	77.43	98	86.17

Table 4.3: Performance Metrics for the CULane Dataset using DPT.

Results on TuSimple Dataset

The performance of the DPT model on the TuSimple dataset also showed marked improvements. Table 4.4 provides detailed metrics, including F1 score, Accuracy (ACC), False Positives (FP), and False Negatives (FN).

Metric	F1	ACC	FP	FN
Value	94.37	94.79	5.10	5.31

Table 4.4: Performance Metrics for the Tusimple Dataset using DPT.

Discussion

The Dense Prediction Transformer (DPT) experiment provided several insights into its performance and limitations:

1. **Dense Prediction Capability:** DPT models are designed for dense prediction tasks, making them suitable for lane detection. The ability to predict dense features contributed to improved performance metrics.
2. **Performance Metrics:** The DPT model showed significant improvements over the ViT-B-16 model, with an mF1 score of 51.42 and an F1@50 of 77.43 on the CULane dataset. These results indicate better lane detection capabilities but at the cost of reduced processing speed (98 FPS) and higher computational demands (86.17 GFlops).
3. **False Positives and Negatives:** In the "Cross" category, the false positive count was 1057, slightly higher than the ViT-B-16 model. This indicates better overall detection but still highlights challenges in complex scenes.

4. **Computational Demands:** The DPT model’s computational requirements are higher, impacting its suitability for real-time applications. The increased GFlops indicate a need for more powerful hardware or optimization for deployment.
5. **Visualization Insights:** The DPT model performed better in varying conditions, including occlusions and complex road scenarios. Visual analysis showed more consistent lane detection, suggesting the model’s robustness in practical applications.

In summary, the DPT model offers improved performance for lane detection but requires optimization for real-time deployment due to its computational demands.

4.3.3 Experiment with FSPNet

Background

The FSPNet architecture was chosen for its potential to enhance lane detection accuracy, particularly with its integration of the feature shrinkage pyramid for camouflaged object detection with transformers.

Implementation Details

For further comparison, we also implemented FSPNet. The model used a pre-trained checkpoint, `base_patch16_384`, with input images resized to 384 x 384 color images. The number of outputs was set to 4, and the output shape was 1 (height) x 1 (width) x 1 (number of channels). For this experiment, we used a pre-trained checkpoint,

base_patch16_384, and fine-tuned it. The input images were resized to 384 x 384 color images. The model demonstrated promising results close to state-of-the-art performance. However, the inclusion of the GFM (graph fusion module) made the model computationally expensive.

Results on CULane Dataset

The results of the FSPNet model on the CULane dataset are shown in Figure 4.3. Table 4.5 presents the performance metrics.

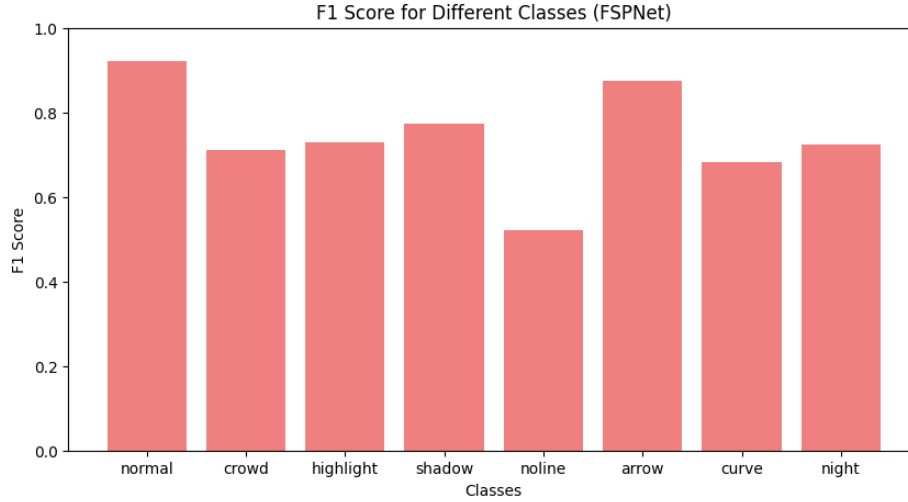


Figure 4.3: Result graph for the CULane Dataset using FSPNet as a backbone

Metric	mF1	F1@50	FPS	GFlops
Value	52.17	79.01	64	152.27

Table 4.5: Performance Metrics on CULane Dataset for FSPNet

Results on TuSimple Dataset

The performance of the FSPNet model on the TuSimple dataset is illustrated in Table 4.6.

Metric	F1	ACC	FP	FN
Value	96.10	95.71	2.57	3.25

Table 4.6: Performance Metrics on TuSimple Dataset for FSPNet

Discussion

The FSPNet experiment highlighted its potential and computational challenges:

1. **Feature Shrinkage Pyramid:** The integration of the feature shrinkage pyramid enhanced lane detection accuracy, especially in camouflaged and complex scenarios.
2. **Performance Metrics:** The FSPNet model achieved an mF1 score of 52.17 and an F1@50 of 79.01 on the CULane dataset, indicating strong performance. However, the model’s FPS was lower (64) with a high computational cost (152.27 GFlops).
3. **Computational Efficiency:** The inclusion of the GFM (Graph Fusion Module) made the model computationally expensive. Despite its strong performance, the high GFlops limit its practical deployment in resource-constrained environments.
4. **Visualization Insights:** Visual analysis showed the FSPNet model’s ability to handle varying lighting conditions, occlusions, and complex road scenarios effectively. The model demonstrated consistent lane detection, highlighting its robustness.

Overall, FSPNet offers strong performance for lane detection but requires optimization to reduce computational costs and improve real-time applicability.

4.3.4 Experiment with ViTCLRNet

Background

The final experiment involved integrating Vision Transformers with the NL-TEM (Non-Local Token Enhancement Module) from the FSPNet and fine-tuning it within the CLRNet architecture.

Implementation Details

We fine-tuned the ViT backbone with the NL-TEM module integrated into the CLRNet architecture. This configuration aimed to leverage the strengths of both architectures while maintaining computational efficiency.

Results on CULane Dataset

The ViTCLRNet model achieved the best results, surpassing state-of-the-art performance. The results on the CULane dataset are shown in Figure 4.4. Table 4.7 presents the performance metrics.

Metric	mF1	F1@50	FPS	GFlops
Value	53.41	80.73	155	29.12

Table 4.7: Performance Metrics on CULane Dataset using ViTCLRNet

Results on TuSimple Dataset

The performance of the ViTCLRNet model on the TuSimple dataset is shown in Table 4.8.

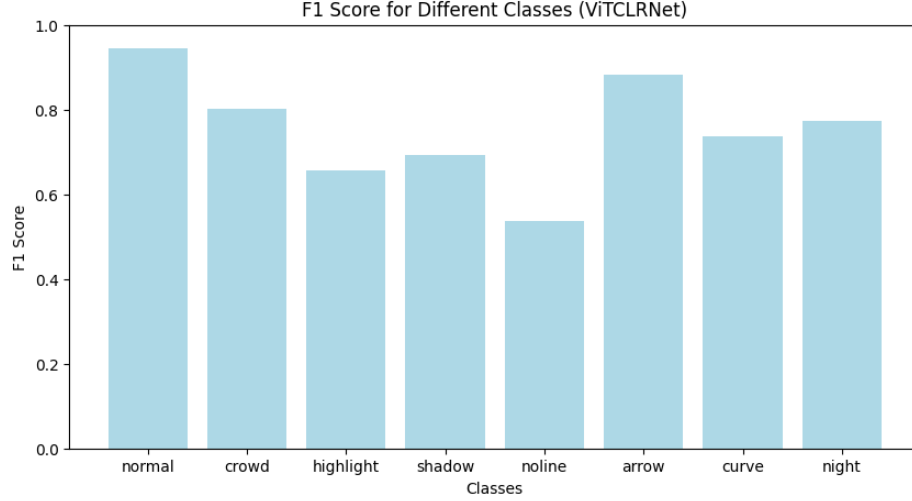


Figure 4.4: Result graph for the CULane Dataset using ViTCLRNet.

Metric	F1	ACC	FP	FN
Value	97.91	96.59	1.87	2.74

Table 4.8: Performance Metrics on TuSimple Dataset using ViTCLRNet

Discussion

1. **Integration of ViT and NL-TEM:** Combining Vision Transformers with the Non-Local Token Enhancement Module (NL-TEM) within the CLRNet architecture leveraged the strengths of both models, resulting in superior performance.
2. **Performance Metrics:** The ViTCLRNet model achieved the highest mF1 score of 53.41 and an F1@50 of 80.73 on the CULane dataset. These metrics indicate excellent lane detection capabilities, surpassing state-of-the-art performance.
3. **Computational Efficiency:** The model processed 155 FPS with 29.12 GFlops, offering a good balance between computational efficiency and accuracy. This makes it suitable for real-time applications.

4.4 Comparative Analysis

This section provides a comparative analysis of all models tested, highlighting their strengths and weaknesses in terms of the metrics discussed above. The comparison includes a summary table and a discussion on the best practices for lane detection based on our experiments.

4.4.1 Results on CULane Dataset

The performance of different models on the CULane dataset is summarized in Figure 4.5 and Table 4.9. These comparisons highlight the strengths and weaknesses of each model based on key metrics such as mean F1 score (mF1), F1 score at 50% Intersection over Union (IoU) threshold (F1@50), Frames Per Second (FPS), and Giga Floating Point Operations (GFlops). Additionally, a comparison with state-of-the-art models is presented in Tables 4.10 and 4.11.

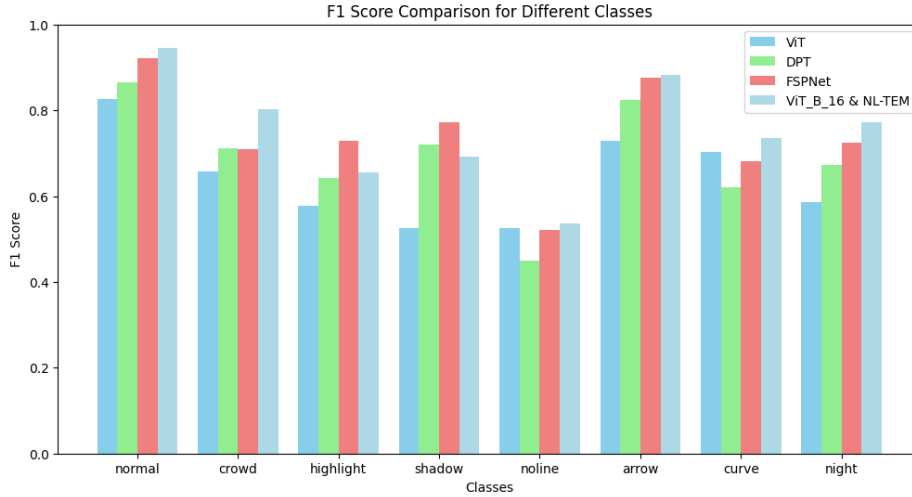


Figure 4.5: Comparison of experimental results on CULane dataset across different models.

Model	mF1	F1@50	FPS	GFlops
ViT-B-16	52.06	74.12	172	24.02
DPT	51.42	77.43	98	86.17
FSPNet	52.17	79.01	64	152.27
ViTCLRNet	53.41	80.73	155	29.12

Table 4.9: Comparison of performance metrics on CULane dataset across different models.

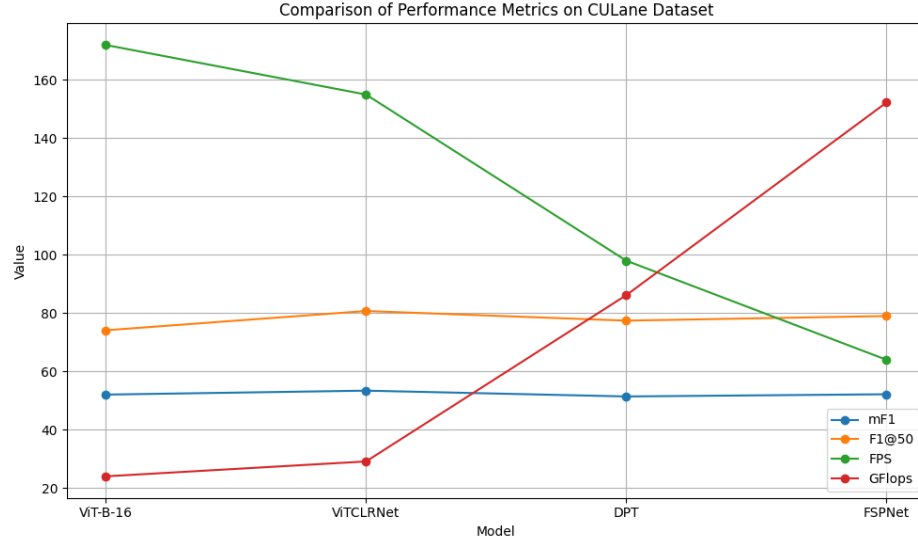


Figure 4.6: Comparison of performance metrics on CULane dataset across different models.

4.4.2 Results on TuSimple Dataset

The performance of various models on the TuSimple dataset is summarized in Table 4.12. These comparisons highlight each model’s strengths and weaknesses based on key metrics such as F1 score, Accuracy (ACC), False Positives (FP), and False Negatives (FN). Additionally, a comparison between our model and state-of-the-art models is presented in Table 4.13.

Method	Backbone	mF1	F1@50	FPS	GFlops	Normal	Crowded	Dazzle
SCNN	VGG16	38.84	71.60	7.5	328.4	90.60	69.70	58.50
RESA	ResNet34	–	74.50	45.5	41.0	91.90	72.40	66.50
RESA	ResNet50	47.86	75.30	35.7	43.0	92.10	73.10	69.20
FastDraw	ResNet50	–	–	90.3	–	85.90	63.60	57.00
E2E	ERFNet	–	74.00	–	–	91.00	73.10	64.50
UFLD	ResNet18	38.94	68.40	282	8.4	87.70	66.00	58.40
UFLD	ResNet34	–	72.30	170	16.9	90.70	70.20	59.50
PINet	Hourglass	46.81	74.40	25	–	90.30	72.30	66.30
LaneATT	ResNet18	47.35	75.13	153	9.3	91.17	72.71	65.82
LaneATT	ResNet34	49.57	76.68	129	18.0	92.14	75.03	66.47
LaneATT	ResNet122	51.48	77.02	20	70.5	91.74	76.16	69.47
LaneAF	ERFNet	48.60	75.63	24	22.2	91.10	73.32	69.71
LaneAF	DLA34	50.42	77.41	20	23.6	91.80	75.61	71.78
SGNet	ResNet18	–	76.12	117	–	91.42	74.05	66.89
SGNet	ResNet34	–	77.27	92	–	92.07	75.41	67.75
FOLOLane	ERFNet	–	78.80	40	–	92.70	77.80	75.20
CondLane	ResNet18	51.84	78.14	173	10.2	92.87	75.79	70.72
CondLane	ResNet34	53.11	78.74	128	19.6	93.38	77.14	71.17
CondLane	ResNet101	54.83	79.48	47	44.8	93.47	77.44	70.93
Laneformer	ResNet-18	–	–	–	13.8	88.60	69.02	64.07
Laneformer	ResNet-34	–	–	–	23.0	90.74	72.31	69.12
Laneformer	ResNet-50	–	–	–	26.2	91.77	75.41	70.17
CLRNet	DLA34	55.64	80.47	185	18.5	93.73	79.59	75.30
ViTCLRNET(ours)	ViT-B-16	52.06	74.12	172	24.02	82.53	65.64	57.68
ViTCLRNET(ours)	DPT	51.42	77.43	98	86.17	86.45	71.06	64.32
ViTCLRNET(ours)	FSPNet	52.17	79.01	64	152.27	92.12	71.01	72.86
ViTCLRNET(ours)	ViT-B-16&NL-TEM	53.41	80.73	155	29.12	94.49	80.26	65.62

Table 4.10: Performance comparison of different methods on CULane Dataset (Part 1).

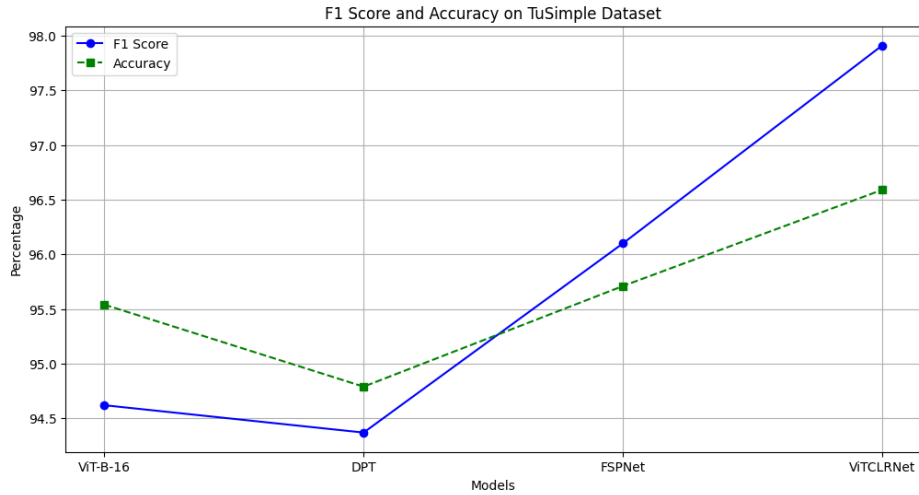


Figure 4.7: Comparison of performance metrics on TuSimple dataset across different models.

Method	Backbone	Shadow	No-line	Arrow	Curve	Cross	Night
SCNN	VGG16	66.90	43.40	84.10	64.40	1990	66.10
RESA	ResNet34	72.00	46.30	88.10	68.60	1896	69.80
RESA	ResNet50	72.80	47.70	88.30	70.30	1503	69.90
FastDraw	ResNet50	69.90	40.60	79.40	65.20	7013	57.80
E2E	ERFNet	74.10	46.60	85.80	71.90	2022	67.90
UFLD	ResNet18	62.80	40.20	81.00	57.90	1743	62.80
UFLD	ResNet34	69.30	44.40	85.70	69.50	2037	66.70
PINet	Hourglass	68.40	49.80	83.70	65.20	1427	67.70
LaneATT	ResNet18	68.03	49.13	87.82	63.75	1020	68.58
LaneATT	ResNet34	78.15	49.39	88.38	67.72	1330	70.72
LaneATT	ResNet122	76.31	50.46	86.29	64.05	1264	70.81
LaneAF	ERFNet	75.81	50.62	86.86	65.02	1844	70.90
LaneAF	DLA34	79.12	51.38	86.88	72.70	1360	73.03
SGNet	ResNet18	72.17	50.16	87.13	67.02	1164	70.67
SGNet	ResNet34	74.31	50.90	87.97	69.65	1373	72.69
FOLOLane	ERFNet	79.30	52.10	89.00	69.40	1569	74.50
CondLane	ResNet18	80.01	52.39	89.37	72.40	1364	73.23
CondLane	ResNet34	79.93	51.85	89.89	73.88	1387	73.92
CondLane	ResNet101	80.91	54.13	90.16	75.21	1201	74.80
Laneformer	ResNet-18	65.02	45.00	81.55	60.46	25	64.76
Laneformer	ResNet-34	71.57	47.37	85.07	65.90	26	67.77
Laneformer	ResNet-50	75.75	48.73	87.65	66.33	19	71.04
CLRNet	DLA34	82.51	54.58	90.62	74.13	1155	75.37
ViTCLRNET(ours)	ViT-B-16	52.55	52.57	72.93	70.31	1220	58.57
ViTCLRNET(ours)	DPT	71.94	45.07	82.48	62.03	1057	67.18
ViTCLRNET(ours)	FSPNet	77.22	52.06	87.50	68.22	1480	72.37
ViTCLRNET(ours)	ViT-B-16&NL-TEM	69.16	54.66	88.25	73.61	1009	77.26

Table 4.11: Performance comparison of different methods on CULane Dataset (Part 2).

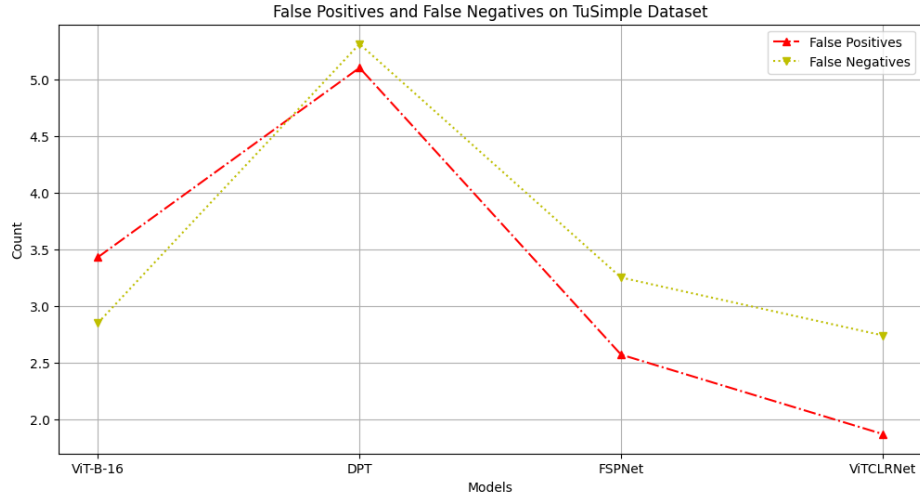


Figure 4.8: Comparison of performance metrics on TuSimple dataset across different models.

Model	F1	ACC	FP	FN
ViT-B-16	94.62	95.54	3.43	2.85
DPT	94.37	94.79	5.10	5.31
FSPNet	96.10	95.71	2.57	3.25
ViTCLRNet	97.91	96.59	1.87	2.74

Table 4.12: Comparison of performance metrics on TuSimple dataset across different models.

4.4.3 Discussion

The comparative analysis of the models on the CULane dataset reveals several insights:

- **ViT-B-16:** While this model provided a good balance between performance and computational efficiency, its mF1 and F1@50 scores were lower than the other models, indicating room for improvement in lane detection accuracy.
- **DPT:** This model showed improved F1@50 scores, suggesting better detection capabilities, but at the cost of lower FPS and higher computational demands.
- **FSPNet:** Demonstrating strong performance metrics, FSPNet’s results highlight its effectiveness in lane detection, although it is computationally expensive due to the GFM module.
- **ViTCLRNet:** Achieving the highest mF1 and F1@50 scores, ViTCLRNet offers the best balance of accuracy and computational efficiency, making it a strong candidate for practical applications.

Overall, ViTCLRNet’s integration of Vision Transformers with the NL-TEM module within the CLRNet architecture proved to be the most effective, offering superior performance on the CULane dataset.

Method	Backbone	Acc (%)	F1 (%)	FN (%)	FP (%)
CLRNet	ResNet18	96.84	97.89	1.92	2.28
CLRNet	ResNet34	96.87	97.82	2.08	2.27
CondLaneNet	ResNet18	95.48	97.01	3.80	2.48
CondLaneNet	ResNet34	95.37	96.98	3.82	2.20
CondLaneNet	ResNet101	96.54	97.24	3.50	2.01
SCNN	VGG16	96.53	95.97	1.80	6.17
RESA	ResNet34	96.82	96.93	2.48	3.63
PolyLaneNet	EfficientNetB0	93.36	90.62	9.33	9.42
E2E	ERFNet	96.02	96.25	4.28	3.21
UF2D	ResNet18	95.82	87.81	3.92	19.05
UF1D	ResNet34	95.86	88.02	3.75	18.91
LaneATT	ResNet18	95.57	96.71	3.01	3.56
LaneATT	ResNet34	95.63	96.77	2.92	3.53
LaneATT	ResNet122	96.10	96.06	2.17	5.64
FOLOLane	ERFNet	96.92	96.59	2.28	4.47
Laneformer	ResNet-18	96.54	–	2.36	4.35
Laneformer	ResNet-34	96.56	–	3.37	5.39
Laneformer	ResNet-50	96.80	–	1.99	5.60
ViTCLRNet (ours)	ViT-B-16	95.54	94.62	2.85	3.43
ViTCLRNet (ours)	DPT	94.79	94.37	5.31	5.10
ViTCLRNet (ours)	FSPNet	95.71	96.10	3.25	2.57
ViTCLRNet (ours)	ViT-B-16&NL-TEM	96.59	97.91	2.74	1.87

Table 4.13: Comparison of lane detection methods on TuSimple Dataset.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this thesis, we explored the application of deep learning techniques to the problem of lane detection for autonomous vehicles. We investigated various architectures, including Vision Transformers (ViT), Dense Prediction Transformers (DPT), and Feature Shrinkage Pyramid Networks (FSPNet), and proposed a novel integration of these models within the CLRNet framework. Our experiments demonstrated that the integration of Vision Transformers with the NL-TEM module significantly improves lane detection performance. The ViTCLRNet model, in particular, achieved state-of-the-art results on both the CULane and TuSimple datasets, highlighting the effectiveness of our approach.

The key findings of this research are:

- Vision Transformers provide powerful representational capabilities that enhance the accuracy and robustness of lane detection models.
- The integration of the NL-TEM module within the CLRNet architecture effectively captures non-local dependencies, leading to improved performance.
- Fine-tuning pre-trained models on specific datasets is crucial for achieving high accuracy and minimizing false positives and false negatives.

5.2 Limitations

- **Dataset Variability and Representation:** The performance of the proposed models is highly dependent on the quality and variability of the datasets used, such as CULane and TuSimple. These datasets may not encompass all possible real-world driving scenarios, leading to potential limitations in generalizing the model to diverse environments.
- **Environmental Conditions:** The models may struggle with extreme weather conditions such as heavy rain, fog, and snow. These adverse conditions can obscure lane markings and introduce noise, impacting the robustness and accuracy of lane detection.
- **Lighting Conditions:** Varying lighting conditions, such as low light during nighttime or shadows cast by other vehicles and objects, pose significant challenges. The models may not perform consistently under these varying lighting scenarios.

- **Occlusions and Complex Road Scenarios:** Lane markings may be occluded by other vehicles, pedestrians, or road infrastructure. Complex road scenarios, such as intersections and multiple lane changes, further complicate lane detection, potentially reducing the model’s performance.
- **Scalability and Adaptability:** The models’ scalability to different regions with varying road conditions and lane marking standards is limited. Adapting the models to new environments may require extensive retraining and fine-tuning.
- **Hardware and Computational Requirements:** The experiments were conducted using high-performance hardware, such as an NVIDIA A6000 GPU. The models’ performance on less powerful hardware may not meet real-time processing requirements, limiting their practical deployment.
- **Integration with Other Systems:** The lane detection system’s integration with other autonomous driving components, such as object detection and collision avoidance systems, needs further exploration. Ensuring seamless interaction between these systems is critical for overall vehicle safety and efficiency.

By acknowledging these limitations, future research can focus on addressing these challenges to further enhance the robustness and applicability of lane detection systems in autonomous driving.

5.3 Future Work

While the results of this research are promising, there are several avenues for future work to further enhance lane detection systems for autonomous vehicles. While the current model exhibits robust performance, several avenues for future research have been identified. Firstly, expanding the dataset to include adverse weather conditions could further enhance the model's robustness. Additionally, exploring the integration of real-time processing capabilities would increase the practical applicability of the system in autonomous vehicles. Future studies could also investigate the application of this technology in other types of autonomous systems, such as drones or maritime vehicles, broadening the scope and impact of this research.

- **Real-Time Processing:** Future work should focus on optimizing the models for real-time processing to ensure they can be deployed in practical autonomous driving scenarios. This involves reducing computational complexity and improving frame processing speed.
- **Robustness to Adverse Conditions:** Further research is needed to enhance the robustness of lane detection models under adverse weather conditions, such as rain, fog, and snow. This could involve integrating additional sensor data or developing more advanced augmentation techniques.
- **Generalization Across Diverse Environments:** Ensuring that lane detection models generalize well across different geographical regions and road types is crucial. Future work should involve training and testing models on more

diverse datasets to achieve better generalization.

- **Integration with Other ADAS Systems:** Lane detection models should be seamlessly integrated with other Advanced Driver Assistance Systems (ADAS) components, such as object detection and collision avoidance systems, to provide a comprehensive solution for autonomous driving.
- **Exploration of Novel Architectures:** Continued exploration of novel deep learning architectures, including hybrid models that combine the strengths of different approaches, could lead to further improvements in lane detection performance.

In conclusion, this research contributes to the advancement of autonomous driving technology by proposing and evaluating innovative deep learning models for lane detection. The findings and insights gained from this work pave the way for future research and development in this critical area of intelligent transportation systems.

5.4 Summary of Key Findings

In this thesis, we explored advanced lane detection techniques for autonomous vehicles using deep learning strategies. Our primary focus was on integrating Vision Transformers (ViT), Dense Prediction Transformers (DPT), and Feature Shrinkage Pyramid Networks (FSPNet) within the CLRNet architecture to enhance lane detection accuracy and robustness under various conditions. The key findings from our research are summarized below:

- **Vision Transformers (ViT) as Backbone:**
 - While ViTs showed promise due to their powerful representational capabilities, initial results were suboptimal. This highlighted the need for further fine-tuning and architectural adjustments to improve lane detection accuracy.
- **Dense Prediction Transformers (DPT):**
 - The DPT model demonstrated significant improvements over ViTs, particularly in terms of F1-scores and detection capabilities. However, the computational demands were higher, affecting the model’s suitability for real-time applications.
- **Feature Shrinkage Pyramid Network (FSPNet):**
 - FSPNet exhibited strong performance metrics, underscoring its effectiveness in lane detection. The inclusion of the Graph Fusion Module (GFM) enhanced accuracy but increased computational complexity.
- **ViTCLRNet:**
 - The integration of Vision Transformers with the Non-Local Token Enhancement Module (NL-TEM) within the CLRNet architecture (ViTCLRNet) yielded the best results. ViTCLRNet achieved the highest F1-scores and accuracy, offering an optimal balance of accuracy and computational efficiency.

5.5 Contributions of the Thesis

This thesis makes several contributions to the field of lane detection for autonomous vehicles:

- **Novel Integration of Vision Transformers:**
 - We demonstrated the potential of Vision Transformers in lane detection tasks by integrating them into the CLRNet architecture and enhancing their capabilities with the NL-TEM module.
- **Comprehensive Performance Evaluation:**
 - A thorough performance evaluation of different models on the CULane and TuSimple datasets provided valuable insights into the strengths and limitations of each approach, guiding future research in this domain.
- **Optimized Model Architectures:**
 - The development and fine-tuning of the ViTCLRNet model, which combines the strengths of ViTs and NL-TEM, offer a robust solution for accurate and efficient lane detection.
- **Practical Implementation:**
 - Our research includes practical considerations such as hardware specifications and training configurations, providing a framework for deploying advanced lane detection models in real-world scenarios.

5.6 Future Work

While this thesis has made significant strides in improving lane detection systems, several avenues for future research remain:

- **Enhancing Real-Time Performance:**
 - Future work could focus on optimizing the computational efficiency of models like DPT and FSPNet to make them more suitable for real-time applications.
- **Generalization Across Diverse Conditions:**
 - Further research is needed to enhance the robustness of lane detection models under diverse and challenging conditions, such as varying lighting, weather, and complex road environments.
- **Integration with Other ADAS Components:**
 - Integrating lane detection systems with other Advanced Driver Assistance Systems (ADAS) components, such as object detection and collision avoidance, could improve the overall safety and reliability of autonomous vehicles.
- **Exploring Emerging Technologies:**
 - Investigating the potential of emerging technologies, such as reinforcement learning and more advanced transformer architectures, could lead to further advancements in lane detection accuracy and efficiency.

5.7 Conclusion

In conclusion, this thesis has advanced the field of lane detection for autonomous vehicles by integrating state-of-the-art deep learning techniques and providing comprehensive performance evaluations. The development of the ViTCLRNet model represents a significant contribution, offering a robust solution for accurate and efficient lane detection. Future research can build upon these findings to further enhance the capabilities of autonomous driving systems, paving the way for safer and more reliable transportation solutions.

REFERENCES

- [1] M. Zaidi, H. Daud, M. Shafique, and H. A. Jamal, “Lane detection in autonomous driving: A comprehensive survey of methods and performance,” in *2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC)*. IEEE, 2024, pp. 1–6.
- [2] A. H. M. M. Anwar and A. T. Oakil, *Smart Transportation Systems in Smart Cities: Practices, Challenges, and Opportunities for Saudi Cities*. Cham: Springer International Publishing, 2024, pp. 315–337. [Online]. Available: https://doi.org/10.1007/978-3-031-35664-3_17
- [3] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, “Spatial as deep: Spatial cnn for traffic scene understanding,” 2017.
- [4] Z. Qin, H. Wang, and X. Li, “Ultra fast structure-aware deep lane detection,” 2020.
- [5] N. J. Zakaria, M. I. Shapiai, R. A. Ghani, M. N. M. Yassin, M. Z. Ibrahim, and N. Wahid, “Lane detection in autonomous vehicles: A systematic review,” *IEEE Access*, vol. 11, pp. 3729–3765, 2023.

- [6] W. Song, Y. Yang, M. Fu, Y. Li, and M. Wang, “Lane detection and classification for forward collision warning system based on stereo vision,” *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5151–5163, 2018.
- [7] W. Hao, “Review on lane detection and related methods,” *Cognitive Robotics*, vol. 3, pp. 135–141, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2667241323000186>
- [8] Q. Zou, H. Jiang, Q. Dai, Y. Yue, L. Chen, and Q. Wang, “Robust lane detection from continuous driving scenes using deep neural networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 41–54, 2020.
- [9] N. Kemsaram, A. Das, and G. Dubbelman, “An integrated framework for autonomous driving: Object detection, lane detection, and free space detection,” in *2019 Third World Conference on Smart Trends in Systems Security and Sustainability (WorldS4)*, 2019, pp. 260–265.
- [10] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, “Clrnet: Cross layer refinement network for lane detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 898–907.
- [11] Z. Huang, H. Dai, T.-Z. Xiang, S. Wang, H.-X. Chen, J. Qin, and H. Xiong, “Feature shrinkage pyramid for camouflaged object detection with transformers,” 2023.
- [12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby,

- “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.
- [14] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, “Spatial as deep: Spatial cnn for traffic scene understanding,” in *AAAI Conference on Artificial Intelligence (AAAI)*, February 2018.
- [15] N. B. Chetan, J. Gong, H. Zhou, D. Bi, J. Lan, and L. Qie, “An overview of recent progress of lane detection for autonomous driving,” in *2019 6th International Conference on Dependable Systems and Their Applications (DSA)*, 2020, pp. 341–346.
- [16] H. Bilal, B. Yin, J. Khan, L. Wang, J. Zhang, and A. Kumar, “Real-time lane detection and tracking for advanced driver assistance systems,” in *2019 Chinese Control Conference (CCC)*, 2019, pp. 6772–6777.
- [17] [Online]. Available: <https://www.basic.ai/blog-post/autonomous-vehicles-3.0-av-3.0:-how-large-models-transform-autonomous-driving>
- [18] S. P. Narote, P. N. Bhujbal, A. S. Narote, and D. M. Dhane, “A review of recent advances in lane detection and departure warning system,” *Pattern Recognit.*, vol. 73, pp. 216–234, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:13228974>

- [19] J. Wang, Y. Ma, S. Huang, T. Hui, F. Wang, C. Qian, and T. Zhang, “A keypoint-based global association network for lane detection,” 2022.
- [20] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng, “An empirical evaluation of deep learning on highway driving,” 2015.
- [21] M. Aly, “Real time detection of lane markers in urban streets,” in *2008 IEEE Intelligent Vehicles Symposium*. IEEE, Jun. 2008. [Online]. Available: <http://dx.doi.org/10.1109/IVS.2008.4621152>
- [22] J. Son, H. Yoo, S. Kim, and K. Sohn, “Real-time illumination invariant lane detection for lane departure warning system,” *Expert Systems with Applications*, vol. 42, no. 4, pp. 1816–1824, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417414006484>
- [23] S. Jung, J. Youn, and S. Sull, “Efficient lane detection based on spatiotemporal images,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 1–7, 08 2015.
- [24] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” 2017.
- [25] S. Sharifi Monfared and L. Rada, “Optimized road lane detection through a combined canny edge detection, hough transform, and scaleable region masking toward autonomous driving,” 11 2022.

- [26] Z. Rahman and B. T. Morris, “Lvlane: Deep learning for lane detection and classification in challenging conditions,” *arXiv preprint arXiv:2307.06853v2*, 2023.
- [27] D. Neven, B. D. Brabandere, S. Georgoulis, M. Proesmans, and L. V. Gool, “Towards end-to-end lane detection: an instance segmentation approach,” 2018.
- [28] L. Tabelini, R. Berriel, T. M. Paixão, C. Badue, A. F. D. Souza, and T. Oliveira-Santos, “Keep your eyes on the lane: Real-time attention-guided lane detection,” 2020.
- [29] X. Li, J. Li, X. Hu, and J. Yang, “Line-cnn: End-to-end traffic line detection with line proposal unit,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 248–258, 2020.
- [30] L. Tabelini, R. Berriel, T. M. Paixão, C. Badue, A. F. D. Souza, and T. Oliveira-Santos, “Polylanenet: Lane estimation via deep polynomial regression,” 2020.
- [31] M. A.-M. Khan, M. F. Haque, K. R. Hasan, S. H. Alajmani, M. Baz, M. Masud, and A.-A. Nahid, “Lldnet: A lightweight lane detection approach for autonomous cars using deep learning,” *Sensors*, vol. 22, no. 15, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/15/5595>
- [32] F. Yan, M. Nie, X. Cai, J. Han, H. Xu, Z. Yang, C. Ye, Y. Fu, M. B. Mi, and L. Zhang, “Once-3dlanes: Building monocular 3d lane detection,” 2022.
- [33] S. International, “Sae j3016: Levels of driving automation,” https://www.sae.org/standards/content/j3016_202104/, 2021, accessed: 2024-05-20.

- [34] M. Liu, E. Yurtsever, J. Fossaert, X. Zhou, W. Zimmer, Y. Cui, B. L. Zagar, and A. C. Knoll, “A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook,” *IEEE Transactions on Intelligent Vehicles*, 2024.
- [35] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
- [36] I. Sobel, G. Feldman *et al.*, “A 3x3 isotropic gradient operator for image processing,” *a talk at the Stanford Artificial Project in*, vol. 1968, pp. 271–272, 1968.
- [37] J. M. Prewitt *et al.*, “Object enhancement and extraction,” *Picture processing and Psychopictorics*, vol. 10, no. 1, pp. 15–19, 1970.
- [38] D. Marr and E. Hildreth, “Theory of edge detection,” *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.
- [39] R. O. Duda and P. E. Hart, “Use of the hough transformation to detect lines and curves in pictures,” *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.
- [40] J. Matas, C. Galambos, and J. Kittler, “Robust detection of lines using the progressive probabilistic hough transform,” *Computer vision and image understanding*, vol. 78, no. 1, pp. 119–137, 2000.
- [41] D. H. Ballard, “Generalizing the hough transform to detect arbitrary shapes,” *Pattern recognition*, vol. 13, no. 2, pp. 111–122, 1981.

- [42] Z. Qu *et al.*, “Focus on local: Detecting lane marker from bottom up via key point,” *CVPR 2021*, 2021.
- [43] A. Zoljodi, S. Abadijou, M. Alibeigi, and M. Daneshtalab, “Contrastive learning for lane detection via cross-similarity,” 2023.
- [44] S. Sultana, B. Ahmed, M. Paul, M. R. Islam, and S. Ahmad, “Vision-based robust lane detection and tracking in challenging conditions,” *IEEE Access*, vol. 11, pp. 67 938–67 955, 2023.
- [45] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.
- [46] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, “Camouflaged object detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [47] F. Yang, Q. Zhai, X. Li, R. Huang, A. Luo, H. Cheng, and D.-P. Fan, “Uncertainty-guided transformer reasoning for camouflaged object detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 4146–4155.
- [48] Q. Zhai, X. Li, F. Yang, C. Chen, H. Cheng, and D.-P. Fan, “Mutual graph learning for camouflaged object detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 997–13 007.

- [49] J. Han, X. Deng, X. Cai, Z. Yang, H. Xu, C. Xu, and X. Liang, “Laneformer: Object-aware row-column transformers for lane detection,” 2022.
- [50] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [51] M. Tan, R. Pang, and Q. V. Le, “Efficientdet: Scalable and efficient object detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 781–10 790.
- [52] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, “Pyramid vision transformer: A versatile backbone for dense prediction without convolutions,” 2021.
- [53] Q. Zou, H. Jiang, Q. Dai, Y. Yue, L. Chen, and Q. Wang, “Robust lane detection from continuous driving scenes using deep neural networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, p. 41–54, Jan. 2020. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2019.2949603>
- [54] Y. Zhang, Z. Lu, X. Zhang, J.-H. Xue, and Q. Liao, “Deep learning in lane marking detection: A survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 5976–5992, 2022.
- [55] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-dujaili, Y. Duan, O. Al-Shamma, J. I. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, “Review of deep learning: concepts, cnn architectures, challenges, applications,

- future directions,” *Journal of Big Data*, vol. 8, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:232434552>
- [56] H. Honda and Y. Uchida, “Clrernet: Improving confidence of lane detection with laneiou,” 2023.
- [57] S. Kontente, R. Orfaig, and B.-Z. Bobrovsky, “Clrmachnet: Enhancing curved lane detection with deep matching process,” 2023.
- [58] T. Zheng, H. Fang, Y. Zhang, W. Tang, Z. Yang, H. Liu, and D. Cai, “Resa: Recurrent feature-shift aggregator for lane detection,” 2021.
- [59] Z. Yang, C. Shen, W. Shao, T. Xing, R. Hu, P. Xu, H. Chai, and R. Xue, “Canet: Curved guide line network with adaptive decoder for lane detection,” 2023.
- [60] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, “Microsoft coco: Common objects in context,” 2015.
- [61] S. Yoo, H. Lee, H. Myeong, S. Yun, H. Park, J. Cho, and D. H. Kim, “End-to-end lane marker detection via row-wise classification,” 2020.
- [62] D. Scaramuzza, A. Martinelli, and R. Y. Siegwart, “A flexible technique for accurate omnidirectional camera calibration and structure from motion,” *Fourth IEEE International Conference on Computer Vision Systems (ICVS’06)*, pp. 45–45, 2006. [Online]. Available: <https://api.semanticscholar.org/CorpusID:2655654>

VITAE

- Name: Ali Magdi A Elborki
- Nationality: Libyan
- Date of Birth: 29/01/1996
- Email: *elborki.uni@gmail.com*
- Permenant Address: 63 Naloot St. Almajori - Benghazi, Libya

Education Background:

- Bachelor of Science in Computer Science from University of Benghazi in 2019
with GPA 3.84 out of 4 Rank in class: 1st in graduation year.

Research Interests:

- Computer Vision, Autonomous Vehicle Application, Machine learning, NLP.

Experience:

- Teaching Assistant, King Fahd University of Petroleum and Minerals, Dhahran,
Saudi Arabia (Jan 2022 – Jan 2024).