Overtaking Mechanisms based on Augmented Intelligence for Autonomous Driving: Datasets, Methods, and Challenges

Vinay Chamola, *Senior Member, IEEE*, Amit Chougule, Aishwarya Sam, Amir Hussain, *Senior Member, IEEE*, and F. Richard Yu, *Fellow, IEEE*

***Abstract*—The field of autonomous driving research has made significant strides towards achieving full automation, endowing vehicles with self-awareness and independent decision-making. However, integrating automation into vehicular operations presents formidable challenges, especially as these vehicles must seamlessly navigate public roads alongside other cars and pedestrians. An intriguing yet relatively underexplored domain within autonomous driving is overtaking. Overtaking involves a dynamic interplay of complex tasks, including precise steering and speed control, rendering it one of the most intricate operations for implementing augmented intelligence driving technologies. Surprisingly, the overtaking of autonomous vehicles remains largely uncharted territory in the context of augmented intelligence for autonomous systems. This void in knowledge beckons researchers to embark on explorations and investigations in this nascent field. Our review paper systematically synthesises overtaking methodologies hinging on computer vision techniques tailored for augmented intelligence autonomous driving scenarios in response to this pressing need. Our analysis encompasses an array of domains central to overtaking in augmented intelligence autonomous vehicles, encompassing Object Detection, Lane/Line Detection, Depth Estimation, Obstacle Detection, Segmentation, and Pedestrian Detection. We meticulously analyze each domain using well-established Multimodal datasets. We assess different models' performance across various parameters by employing graphical structures, enabling visual comparative analyses. In object detection, YOLOv4 achieves a top performance with 0.90 mAP on the BDD100K dataset. For lane detection, CLRNET excels with the highest F1 score of around 0.96 on the LLAMAS dataset. ViT-Adapter-L leads in segmentation tasks, boasting an impressive mIoU score of 83 on Cityscapes. The Hierarchical Model achieves a superior mAP of 0.90 in road sign detection on the Tsinghua-Tencent Dataset. Steering angle computation sees InterFuser as the standout, achieving the highest driving score of approximately 74.0. This paper's primary contributions include a comprehensive assessment of diverse models for each Multimodal dataset, aiding future research in this evolving domain.**

***Index Terms*—Autonomous driving, Overtaking mechanism, Computer vision, Augmented Intelligence.**

This work is supported by the SICI SICRG grant received by Vinay Chamola and F. Richard Yu.

Vinay Chamola and Amit Chougule are with the Department of Electrical and Electronics Engineering & APPCAIR, BITS-Pilani, Pilani Campus, 333031, India. (e-mail: amitchougule121@gmail.com, vinay.chamola@pilani.bits-pilani).

Aishwarya Sam is from the Department of Computer Science and Information Systems, BITS-Pilani, Pilani Campus, 333031, India. (e-mail: aishwarya.sam25@gmail.com).

Amir Hussain is with School of Computing, Edinburgh Napier University, Scotland, UK (email: A.Hussain@napier.ac.uk)

F. Richard Yu is with the Department of Information Technology, Carleton University, Ottawa, ON, Canada. (e-mail: richard.yu at carleton.ca).

Fig. 1: **Top Left:** Off-road overtaking scenario. **Top Right:** Overtaking scenario on Highway. **Bottom Left:** Night time overtaking. **Bottom Right:** Accident while doing overtaking [1]

I. INTRODUCTION

The automotive industry has undergone a remarkable evolution over the years, driven by advancements in technology and changing consumer demands. From the early days of steam-powered vehicles to the mass production of internal combustion engine cars and now the emergence of electric and augmented intelligence-based vehicles, the industry has constantly adapted and innovated. The integration of electronics, connectivity, and augmented intelligence has transformed automobiles into sophisticated and intelligent machines, offering enhanced safety features, improved fuel efficiency, and personalized driving experiences. This ongoing evolution is paving the way for a future of sustainable and intelligent mobility solutions, revolutionizing the way we travel and interact with vehicles.

As the automobile industry evolved, human driver behaviour also changed with each transition [2]. The evolution of cars from public transportation modes such as buses to personal human cars has significantly impacted the relationship between individuals and their vehicles. With the advent of personal cars, individuals gained the freedom to travel at their convenience, allowing for greater autonomy and flexibility in daily commuting. This shift from shared transportation to personal vehicles transformed the way people perceive and interact with their mode of transportation, providing a sense of ownership, privacy, and control over their journeys. As cars continue to advance with technological innovations like electric and

augmented intelligence, the relationship between humans and their vehicles is evolving further, emphasizing convenience, sustainability, and a more personalized driving experience [3, 4]. Augmented intelligence seamlessly intertwines with the foundation of self-driving cars, enhancing the prowess of artificial intelligence (AI) through symbiotic collaborations with human expertise. In the realm of sensor fusion and perception, engineers leverage augmented intelligence to amalgamate data from diverse sensors, such as cameras, LiDAR, radar, and GPS, crafting a holistic understanding of the vehicle's environment [5, 6]. Human involvement is pivotal in designing algorithms that interpret sensor data, making nuanced decisions in complex scenarios. Data annotation and machine learning benefit from human experts who meticulously label datasets, enhancing the system's ability to recognize and respond to various situations. In the development of user interfaces and human-machine interaction, augmented intelligence aids in creating intuitive systems, while during the testing phase, human safety drivers or operators provide critical oversight, ensuring the system's safety and reliability. Moreover, human experts contribute to defining ethical guidelines, regulatory compliance, and standards, guiding the ethical decision-making processes of self-driving systems. This Augmented intelligence approach harnesses the strengths of both artificial and human intelligence, propelling the advancement of safe and effective autonomous driving technologies. When augmented intelligence is integrated into vehicles, humans benefit from enhanced safety through advanced driver assistance systems, reduced human error, improved traffic management, and a more comfortable and convenient driving experience [7, 8]. This technology empowers vehicles to assist and collaborate with drivers, making transportation safer, more efficient, and enjoyable. Also, the Driving experience also changed, including various activities, such as high-speed driving, comfortable driving, power steering, and overtaking. The act of overtaking during driving enhances the overall driving experience for human beings. It provides a sense of control and excitement as drivers navigate and surpass slower-moving vehicles, allowing them to engage in the dynamics of the road actively. Overtaking manoeuvres requires skill, judgment, and situational awareness, highlighting the connection between human decision-making and the art of driving. The successful execution of overtaking manoeuvres contributes to a satisfying driving experience, adding an element of challenge and skill to the journey. In the case of autonomous driving, Due to various complexities and implementation challenges, the concept of overtaking in autonomous vehicles has been a topic that has yet to be worked on in-depth, which will strengthen the future of self-driving car technology [9, 10].

The overtaking mechanisms in traditional cars and self-driving cars share certain commonalities, yet diverge significantly in their technological underpinnings and decision-making processes. In the realm of traditional cars, the onus falls entirely on the human driver during overtaking maneuvers. This involves a multifaceted analysis of on-road activities, encompassing the detection of front vehicles, objects, estimation of free space for overtaking, path determination, maintenance of safe distances with other cars, acceleration or deceleration, and potential lane changes [11, 12]. Human drivers rely on visual cues and mirrors for perception during overtaking in traditional cars [13]. In stark contrast, self-driving cars leverage an array of technologies, including sensors, cameras, Lidar, and artificial intelligence systems, to execute overtaking maneuvers [14, 15]. The decision-making process is shifted from the human driver to the autonomous system. Perception in self-driving cars is facilitated through cameras and Lidar sensors [16]. Vehicle-to-vehicle (V2V) communication becomes instrumental in the overtaking process for self-driving cars [17, 18], allowing them to inform surrounding vehicles of their intentions without the need for manual signals or hand gestures. Figure 1 represents some of the overtaking cases. The present autonomous driving research has become closer to level 5 automated driving, but overtaking is still an uninvestigated research area. This realm of autonomous driving has not been explored yet, and research in this area is in its infancy. Overtaking is a crucial aspect of driving in general and autonomous driving in particular [19]. Since steering and speed control is dynamic and challenging, overtaking is one of the most daunting and complex operations for implementing automated driving technologies. Automating the overtaking manoeuvre necessitates high-precision environment perception technologies that track the state of the surrounding traffic and anticipate probable time-sensitive decision-making and action implementation [20, 21]. Lane-keeping, lane-changing, and overtaking are complicated driving actions necessary for everyday driving conditions and can improve road efficiency, particularly on one- and two-lane highways. According to NHTSA statistics, motor vehicle crashes claimed the lives of 28,190 people in the first nine months of 2020. The most common causes of accidents were lane changes or overtaking attempts by the driver. The outcome of an overtaking operation is affected by several factors, such as the skills, roadway capacities, and information processing abilities of those involved [22, 23]. Overtaking requires skill and judgment from the driver. They have to gauge the speed and distance of other vehicles, verify the space available for passing, and complete the maneuver safely and efficiently. A driver who does not possess these skills may encounter difficulties in overtaking successfully, leading to potential hazards and dangers on the road. Overtaking maneuvers also require sufficient roadway capacities for safety purposes. If the road is narrow, has bad visibility, or does not have appropriate passing zones, it becomes harder to overtake other vehicles. Insufficient roadway capacities can heighten the risks of overtaking, making it harder for drivers to assess the situation and accomplish the maneuver safely. Additionally, overtaking requires the ability to quickly and accurately process and interpret relevant information. This includes factors such as the speed of the vehicles involved, the actions of other drivers, road conditions, traffic signs, and signals [24, 25]. If a driver finds it hard to process this information effectively, it can affect their decision-making abilities and make it harder to judge the feasibility and safety of an overtaking maneuver. The driver's failure to identify hazards causes almost 75% of road traffic accidents [26, 27]. There have been many unfortunate instances of accidents in the past that are linked to overtaking [28]. Hence,

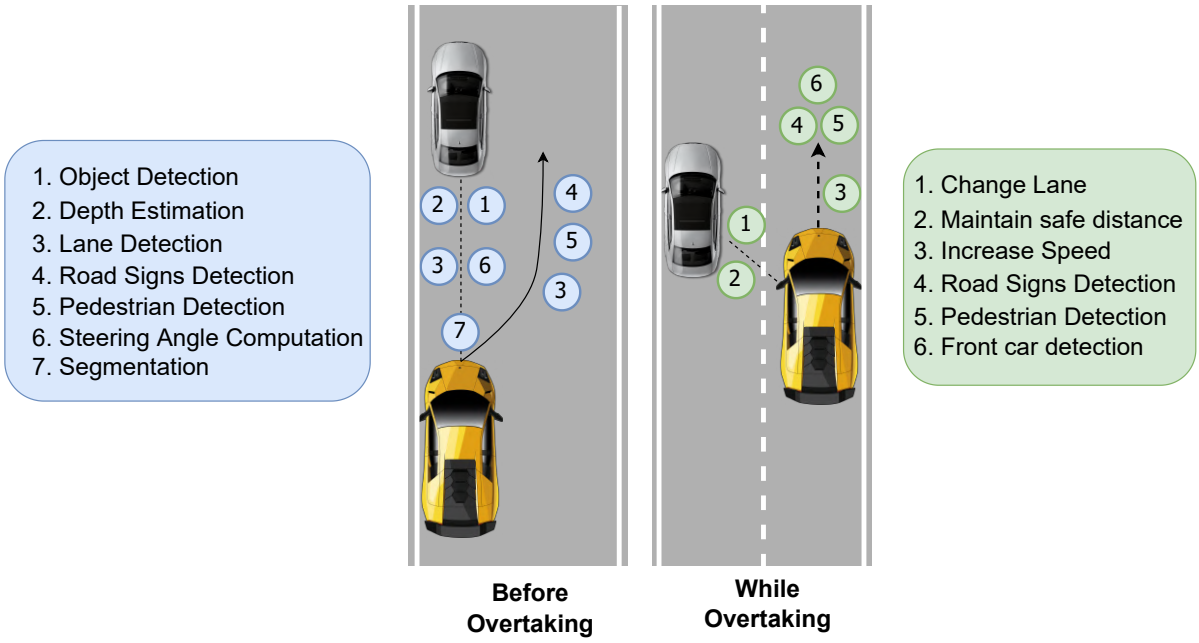

Fig. 2: Overtaking mechanism

developing and applying automated overtaking systems can reduce human error, which can help prevent accidents and injuries.

Keeping all this in mind, we are compelled to consider the question: How do various present autonomous driving cars, such as Tesla and Waymo, perform overtaking maneuvers? When Tesla originally introduced Autopilot, like most other automakers today, the vehicle could maintain its lane and a reasonable distance from the car in front of it [29, 30]. It could not overtake, change lanes, prevent collisions, etc. Similarly, with the new peripheral vision system, Waymo can lessen blind spots from parked big vehicles. With the help of these side cameras, we can glimpse the truck in front of us to determine whether it can be safely overtaken or whether we need to wait. The ultimate goal of an autonomous vehicle is to be as safe as or even more secure than human driving. Having discussed all these aspects, overtaking in autonomous vehicles has to be dealt with in greater depth to take autonomous driving to the next level [4, 31]. This paper deals with the overtaking aspect of autonomous driving using computer vision technology. Figure 2 illustrates the mechanism and various tasks involved before and during overtaking.

Before overtaking, the self-driving car must assess the nearby surroundings. The vehicle must conduct scene understanding for this. The self-driving car first does object detection to determine what objects are in front of it, such as on-road front cars and other obstacles. The autonomous driving vehicle performs depth estimation based on detected objects and evaluates the space between the autonomous vehicle and the front vehicles and objects. This estimated distance helps it overtake and maintain a safe distance from objects in front of the vehicle. At the same time, the self-driving car needs to identify the road lanes. When the route is narrow, autonomous vehicles can not perform overtaking due to safety considerations. This lane detection operation is also used when performing lane changes. Another important activity is the detection of road signs. The road sign indicates upcoming road activities. A model for detecting road signs can identify such signs, interpret their significance, and execute the necessary actions. Some factors, such as narrow roads, bridges, or speed limitations, prohibit cars from safely overtaking [32]. It is also critical that while overtaking, the pedestrian detection algorithm is employed to avoid on-road pedestrian collisions. Augmented Intelligent, self-driving cars must be able to move while remaining inside the drivable area of the road, which is a crucial task [33]. Steering angle computation is vital for meeting safety-critical requirements during the overtaking maneuver and improving the safety and interpretability of end-to-end driving. It keeps the car in the center of the road or inside the lane boundaries. Segmentation aids self-driving cars in detecting which areas of an image are driveable. The image can be accurately evaluated for semantic and instance content using segmentation. The segmentation data from the perception system can be used by the planning and control modules to inform autonomous driving decisions better. To overtake the front vehicle, the self-driving car must first change lanes and maintain a safe distance from the front cars and surrounding cars. The self-driving car must also accelerate faster than the front car in order to pass it. Additionally, the vehicle should recognize road signs, pedestrians, and front cars while performing these operations to avoid a collision. The main contributions of this review are as follows:

- We systematically identified and curated diverse datasets pertinent to computer vision, specifically tailored for the

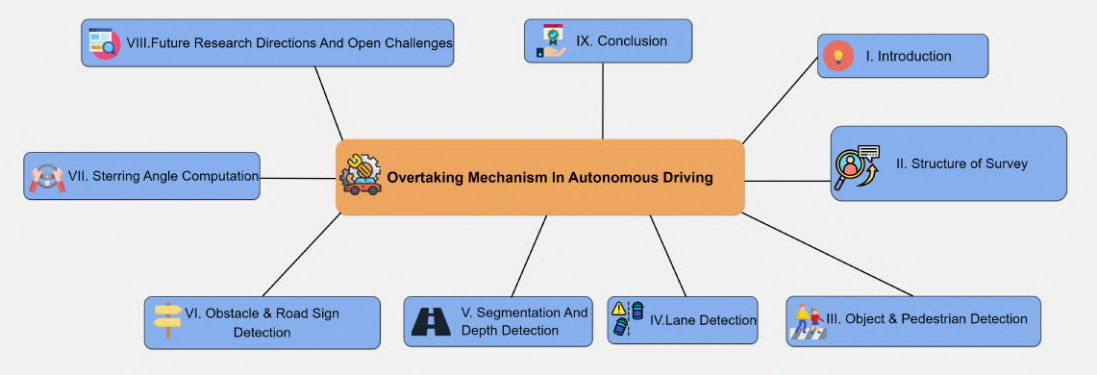

Fig. 3: Structure of our Survey

nuanced task of overtaking manoeuvres in autonomous driving scenarios.

- Employing the identified datasets, we conducted an exhaustive comparative analysis of various models. Our evaluation encompassed performance metrics, and the results were graphically presented for enhanced clarity and visual comprehension.
- Delving into the intricacies of overtaking challenges within autonomous vehicles, we initiated a comprehensive discussion on existing limitations. Our exploration highlights current challenges and serves as a foundation for delineating promising avenues for future research in this dynamic field.

II. Structure of Survey

We begin by discussing existing surveys in overtaking manoeuvres of autonomous driving and comparing them to the current study. Table I shows a collection of significant survey and review works on autonomous driving done in the previous few years.

Section I introduces the notion of autonomous driving and the significance of the overtaking mechanisms in self-driving cars. Section II provides an overview of this paper, which includes a list of various sections that are included in the paper. Additionally, this section explores the existing surveys related to overtaking manoeuvres for autonomous driving. Section III dives deep into the object and pedestrian detection domain, demonstrating the significance of detection capabilities for performing overtaking and illustrating various datasets and models. Section IV provides insights into lane detection functionality and how it aids in steering control and in various decisions. Section V gives insights into segmentation and depth estimation along with its various available datasets and models. Section VI illustrates obstacle and road sign detection, which enables the self-driving car to interpret the meaning of road symbols and perform required actions. Section VII gives various datasets, models, and steering angle computation task functionality. Section VIII summarizes future directions and open challenges in autonomous driving. Section IX concludes the study with a brief review of our contribution. Figure 3 shows the structure of our survey.

III. Object and Pedestrian Detection

Overtaking is one of the most challenging tasks for a self-driving car, and it requires its ability to detect and track objects. It is used to detect and track vehicles, pedestrians, cyclists, and other objects on the road. The object detection model uses visual details to analyze, detect and localize objects in an image. There are many potential benefits of using object detection in car overtaking scenarios. For example, object detection could be used to automatically identify and track vehicles in the overtaking lane, allowing the driver to focus their attention on other tasks. Object detection could also help to determine when a vehicle in front of another vehicle has begun to change lanes so that the second vehicle can take evasive action if necessary. Another example is using object detection to identify obstacles in the path of a vehicle so that the car can avoid them. Additionally, object detection could be used to warn the driver of potential hazards, such as oncoming traffic [40], obstacles in the road, or animals crossing the street. While object detection and pedestrian detection share similarities, they diverge significantly in various aspects. The object detection mechanism is designed to identify a multitude of diverse objects within a frame, encompassing vehicles, bicycles, animals, and environmental elements like trees, barricades, and potholes. This broader scope necessitates the recognition of a diverse range of objects. In contrast, the pedestrian detection mechanism is specifically tailored to focus solely on human beings and their activities. Unlike the object detection system, the pedestrian detection system is characterized by a singular class, exclusively addressing the identification and tracking of individuals within the scene. Therefore, using the object and pedestrian detection, self-driving cars can make various decisions while performing overtaking. If the AV detects a pedestrian or an approaching car, or if there is insufficient free space to perform overtaking, the vehicle can slow down or change lanes and postpone the overtaking process for some time.

In the realm of object and pedestrian detection in autonomous vehicles (AVs), the distinction between 3D object detection and 2D object detection is crucial, as they serve different purposes and provide varying levels of information. 2D object detection primarily focuses on identifying objects within a 2D image plane, which provides essential

Year	Author	Contribution
2018	Dixit, Shilp, *et al.* [34]	Review on trajectory planning and tracking for autonomous overtaking
2019	Ritchie, Owain T., *et al.* [35]	Review on overtaking strategies of autonomous vehicles in relation to other drivers
2020	Hegedűs *et al.* [36]	Survey on overtaking strategies for autonomous vehicles
2021	Perumal, P. Shunmuga, *et al.* [37]	Overview of crash avoidance and overtaking advice systems for Autonomous Vehicles
2021	Sourelli *et al.* [38]	Review on objective and perceived risk in overtaking
2023	Lodh *et al.* [39]	Review on autonomous vehicular overtaking maneuver
2023	This Survey	Reviews the Overtaking mechanism based on computer vision for autonomous vehicle

TABLE I: Related Surveys on Autonomous Driving

information such as bounding box coordinates (x, y, width, height) and class labels for objects present on the road. Typically relying on data from 2D sensors like RGB cameras, 2D object detection constructs a representation of the scene without considering the depth or distance of objects from the camera. It assumes a flat, two-dimensional perspective of the environment, offering a basic understanding of the objects present. On the other hand, 3D object detection goes beyond mere identification by estimating the three-dimensional properties of detected objects. This includes determining their position in 3D space, dimensions (length, width, height), and orientation (yaw, pitch, roll). Unlike 2D object detection, 3D object detection provides depth information, offering a more accurate spatial understanding of the arrangement of objects in the scene. This depth of information is crucial for tasks like navigation and collision avoidance [41]. Furthermore, while 2D object detection is often limited to data from RGB cameras, 3D object detection leverages information from both RGB cameras and 3D sensors such as LiDAR (Light Detection and Ranging). The integration of LiDAR data enhances the ability to capture not only the appearance of objects but also their precise depth information. The popularity of YOLO (You Only Look Once) in the field of object detection can be attributed to its superior Frames Per Second (FPS) performance, particularly suitable for tracking moving objects [42, 43]. Additionally, YOLO boasts better detection accuracy, measured by Intersection over Union (IoU), making it a favored choice among researchers in detection-related domains [43]. Recent research endeavors have extensively utilized YOLO and its variants for object detection, showcasing its efficacy in various applications [44, 45]. Moreover, YOLO has been employed for pedestrian detection, exemplified by studies such as [46, 47]. The efficiency of CNN-based models for both object and pedestrian detection is also notable in the literature [48]. The integration of novel computer vision techniques, such as transformers, has further advanced object detection models. Despite the initial prominence of transformers in recurrent neural network (RNN) applications, they have found adaptability in object detection as well [49, 50].

This section examines the most recent state-of-the-art computer vision-based object detection datasets and models and provides a brief performance analysis. They are reviewed, categorized, and compared with each other. Table II represents available well-known datasets for object detection. To evaluate the object detection performance of the model, most of the models used datasets such as KITTI-3D Object detection [51], BDD100K [52], and Waymo [53]. Therefore, we solely used these datasets to compare models.

Dataset	Year	Images
KITTI-3D Object detection [51]	2012	15,000
BDD100K [52]	2018	100,000
Waymo Dataset [53]	2020	250,000
UA-DETRAC [54]	2020	140,000

TABLE II: Available Datasets for Object detection tasks

Andreas Geiger *et al.* [51] introduced the KITTI dataset, which is suitable for 3D object detection applications. The dataset consists of around 15,000 images that include point clouds. There are 80,256 labeled objects in the collection. The dataset was created by using 22 stereo sequences totaling 39.2 kilometers of on-road driving. Figure 4 shows sample images of the KITTI [51] 3D Object detection dataset in different scenarios.

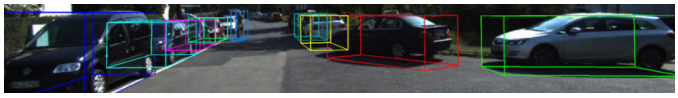

Fig. 4: Sample images of KITTI [51] 3D object detection Dataset

Fig. 5 shows various models' performances on the KITTI 3D object detection dataset using the mAP (mean Average Precision) as the evaluation parameter, whereas the higher the mAP, the better the model performance. The range for mAP

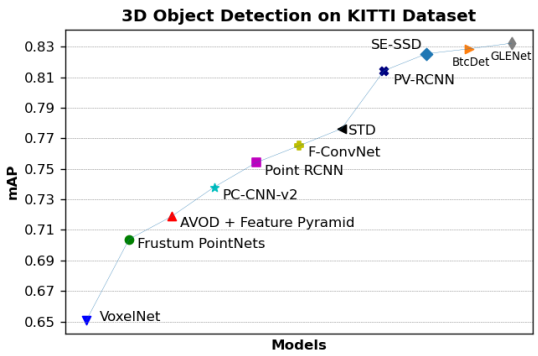

Fig. 5: Performance of models on KITTI [51] 3D object detection Dataset

Fig. 6: Sample images of the BDD100k [52] Dataset

is from 0 to 1. mAP is calculated as,

$$\text{mAP} = \frac{1}{NC}\sum_{i=1}^{NC}\text{AP}_i \tag{1}$$

Where:

NC is the total number of classes,
AP is Average Precision.

Various models like VoxelNet [55], Frustum PointNets [56], AVOD + Feature Pyramid [57], PC-CNN-v2 [58], Point RCNN [59], F-ConvNet [60], STD [61], PR-RCNN [62], SE-SSD [63], BtcDet [64] and GLENet [65] show their 3D object detection performance on the KITTI [51] dataset. Out of all these models, GLENet [65] has excellent performance on the KITTI [51] dataset. GLENet [65] is a generative framework based on conditional variational autoencoders that use latent variables to simulate the one-to-many connection between a specific 3D object and its associated ground-truth bounding boxes. GLENet [65] employs a novel uncertainty-aware quality estimator (UAQE) to facilitate (Intersection over Union) IoU-branch training, influenced by the high correlation between localization quality and predicted uncertainty in probabilistic detectors. Fisher Yu *et al.* [52] built BDD100K, a driving video dataset, including 100K pictures and ten tasks to assess the performance of object detection algorithms on autonomous driving. The dataset has geographic, environmental, and climatic diversity, which is helpful for training models that will be less affected by unexpected situations. The dataset is collected by driving a car on-road, which covers New York, San Francisco Bay Area, and other regions. Figure 6 shows sample images of the BDD100k [52] dataset. Similarly, Fig. 7 shows various models' performances on the BDD100K [52] dataset using the mAP (mean Average Precision) as the evaluation parameter, whereas the higher the mAP, the better the model performance.

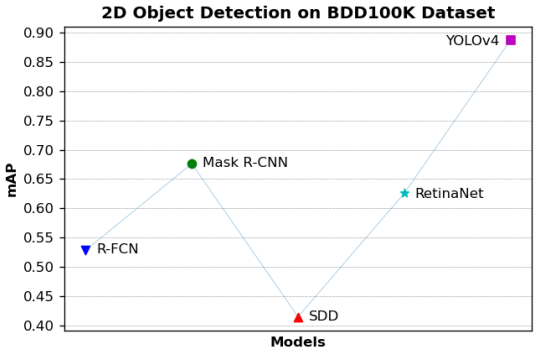

Fig. 7: Performance of models on BDD100k [52] Dataset

Several models like R-FCN [66], Mask R-CNN [67], SDD [68], RetinaNet [69] and YOLOv4 [70] show performance on the BDD100K [52] dataset in terms of object detection capabilities. Out of all these models, YOLOv4 [70] has outstanding performance on the BDD100K [52] dataset. It utilizes Bag-of-Freebies (BoF) and Bag-of-Specials (BoS) methods for the detector, which improve the training performance. Google's Waymo [53] Dataset is a large-scale dataset for self-driving vehicles. The Waymo Open Dataset is a massive collection of LiDAR point clouds, images, and bounding boxes that have been annotated with 3D tracking information. The dataset is intended to train computer vision and machine learning algorithms to recognize and track vehicles in a self-driving vehicle. The collection will also aid researchers in developing new 3D object detection and tracking technologies. It contains 250,000 images from 1150 scenes. Figure 8 shows sample images of the Waymo [53] dataset in different scenarios.

Fig. 9 shows various models' performances on the Waymo [53] dataset using the mAP (mean Average Precision) as the evaluation parameter, whereas the higher the mAP, the better the model performance.

Fig. 8: Samples of the Waymo [53] Dataset

Various models like Lenovo_LR_PCIE_ [71], dereyly_alex_2 [72], YOLOR_P6_TRT [73], DIDI MapVision [72], SPNAS-Noah [72], HorizonDet [72], LeapMotor_Det [73, 74], Noah CV Lab [72], RW-TSDet [75] and Noah Octopus [76]. Out of all these models, Noah Octopus [76] achieves the superior results over other models when evaluated on the Waymo dataset [53].

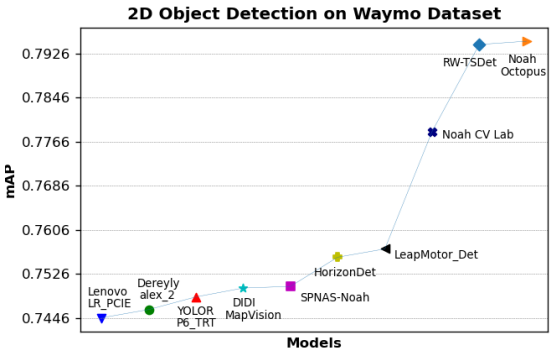

Fig. 9: Performance of models on Waymo [53] Dataset

Recent car safety research has discovered various features for autonomous vehicles that make them more secure and safer for passengers. However, another significant factor that must be considered is on-road pedestrian safety. As self-driving vehicles are driven by machine learning and computer vision algorithms, all decisions and planning are made solely by these algorithms and models. Any errors or incorrect decisions made by these models have significant repercussions, such as accidents and deaths. One of the crucial duties of self-driving cars is on-road vehicle and pedestrian detection, which is inextricably linked to public safety. Pedestrian detection entails recognizing pedestrians in pictures or videos and determining their on-road position. This technology enhances the safety of self-driving cars by preventing collisions with pedestrians. The system learns to recognize the shape and movement of a pedestrian. Self-driving cars identify pedestrians on the road and avoid crashes by slowing down or switching lanes. Using camera data [77], a color histogram [78], skin tone histogram [79], or a face detection method [80] they may classify an object as a pedestrian. In the case of overtaking mechanisms, if a car detects any person on the road before or while overtaking the self-driving car should take appropriate action, such as changing lanes, slowing down, or canceling the overtaking operation, and ensuring that on-road public safety is prioritized. This section provides a comprehensive overview of the most recent datasets and models for autonomous driving for on-road pedestrian detection. Several models are also examined and compared in terms of detection efficiency and performance. Table III shows available datasets for Pedestrian Detection.

Dataset	Year	Images
Caltech [81]	2011	250,000
CityPersons [82]	2017	5000
TJU-DHD [83]	2020	115,354
LLVIP [84]	2021	30,976

TABLE III: Available Datasets for Pedestrian Detection

Piotr Dollar *et al.* [81] provided a large real monocular dataset for pedestrian detection. The dataset contains 350,000 pedestrians who were annotated using bounding boxes in 250,000 frames of the collection. The data was acquired by utilizing 10 hours of 30 Hz video with a resolution of 640 x 480 from a vehicle traveling in an urban environment. Various models used the Caltech dataset for performing pedestrian detection tasks. Figure 10 illustrates sample images of the Caltech [81] dataset.

Fig. 10: Samples of the Caltech [81] Dataset

Figure 11 depicts the performance of various models using the evaluation metric Reasonable Miss Rate (RMR). The lower the Reasonable Miss Rate, the better the model performance in terms of model performance ranking. While AP concentrates on the precision-recall trade-off and measures how well the model detects objects at different confidence levels, RMR provides insights into the performance of the model with respect to missing objects that are considered ”reasonable” to detect. The RMR is calculated using,

$$\text{RMR} = \frac{\left\{\begin{array}{c}\text{Number of missed objects}\\ \text{with confidence below threshold}\end{array}\right\}}{\left\{\begin{array}{c}\text{Total number of objects}\\ \text{with confidence below threshold}\end{array}\right\}} \quad (2)$$

It includes models such as F2DNet [85], Pedestron [86], FRCNN-FPN-Crowdhuman [87], RepLoss [88], Zhang *et al.* [89], RPN + BF [90], SA-FastRCNN [91], CompACT-Deep

[92], Checkerboards+ [93], TA-CNN [94], AlexNet [95] and LDCF [96]. Out of these models, F2DNet [85] shows the lowest Reasonable Miss Rate, which represents excellent performance. F2DNet [85] proposed a novel two-stage detection architecture that replaces the region proposal network with a focal detection network along with a bounding box head with a fast suppression head, which handles false positives. Additionally, it proposed a focal detection network as a classification and bounding box regression head, which improves the results.

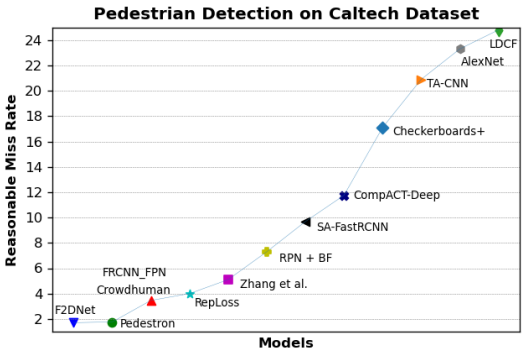

Fig. 11: Performance of models on Caltech [81] Dataset

Shanshan Zhang *et al.* [82] introduced the CityPersons dataset for pedestrian detection. It is comprised of a diverse group of stereo video sequences shot in several cities throughout Germany and adjacent countries. There are 5000 photos divided into 30 classes representing 35k individuals and 19,654 unique individuals. Figure 12 demonstrates sample images of the CityPersons [82] dataset.

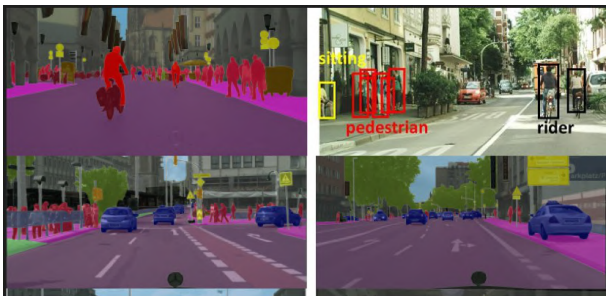

Fig. 12: Samples of the CityPersons [82] Dataset

Many models, notably Pedestron [86], F2DNet [85], ACSP [97], CrowdHuman [87], NOH-NMS [98], CSP (with offset) + ResNet-50 [99], ALFNet [100], OR-CNN [101], RepLoss [88], TL+MRF [102] and FRCNN-FPN-POS [103] used the CityPersons dataset for human recognition. These models performed pedestrian detection, and their performance is shown in figure 13. The performance is evaluated using Reasonable MR^{-2} as an evaluation parameter. Reasonable MR^{-2} is considered as an evaluation parameter, where the lower the Reasonable MR^{-2} value, the better the model performance. Pedestron [86] outperforms all other models in terms of pedestrian detection tasks. It employs a progressive training pipeline to enhance pedestrian detection capability.

Fig. 14: Samples of the TJU-DHD [83] Dataset. **1st Row:** illumination variance(day,night,front light and backlight) **2nd Row:** scene variance (urban road, rural road and highway) **3rd Row:** weather variance (sunny, cloudy, and rain) **4th Row:** season variance (spring, summer, autumn, and winter)

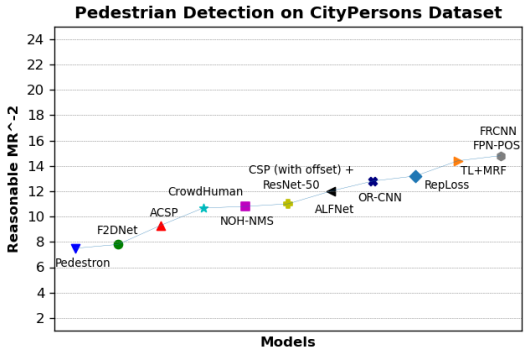

Fig. 13: Performance of models on CityPersons [82] Dataset

Existing well-known public datasets such as MS COCO do not concentrate on distinctive scenarios as well as other datasets like CityPersons [82], and KITTI [51] are limited in the number of images, instances, the resolution, and the diversity. Due to these aspects, Yanwei Pang *et al.* [83] introduced the TJU-DHD dataset, an extensive collection of rich and diverse High-Resolution pedestrian datasets. The dataset contains 115,354 high-resolution images, which contain 709,330 labeled objects. Figure 14 demonstrate sample images of the TJU-DHD [83] dataset. Furthermore, the dataset images cover a variety of circumstances, such as lighting and climate. Due to this rich diversity, various models, such as EGCL [104], CrowdDet [105], FPN [106], FCOS [107] and RetinaNet [69] utilized TJU-DHD dataset for pedestrian identification. Figure 15 shows the model performance comparison in respective with R (Miss rate) [83] as evaluation parameter. While considering the models' performance ranking, the lower the R (Miss rate), the better the model performance. Among all modes, EGCL (Exemplar-Guided Contrastive Learning) [104] outperforms all other models in terms of performance. Detection of pedestrians with significant appearance differences, such as various pedestrian shapes, various angles, or diverse attire, remains a critical difficulty. EGCL (Exemplar-Guided

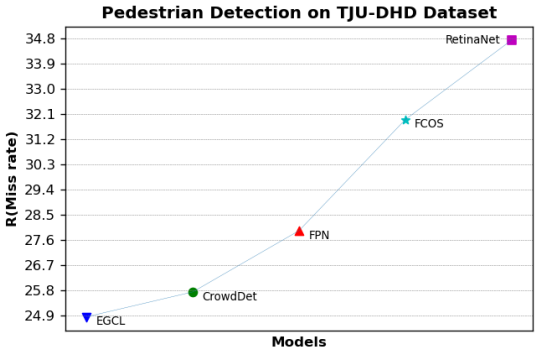

Fig. 15: Performance of models on TJU-DHD [82] Dataset

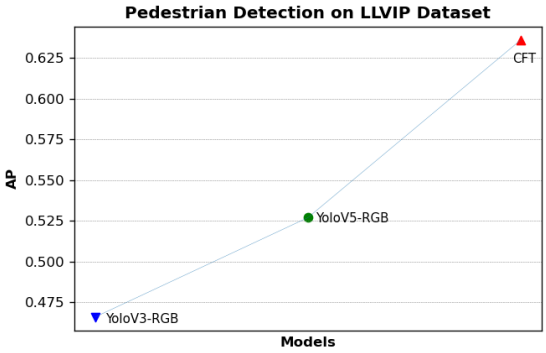

Fig. 17: Performance of models on LLVIP [84] Dataset

Fig. 16: Samples of the LLVIP [84] Dataset. **1st Column:** Pedestrian detection on infrared images **2nd Column:** Pedestrian detection on visible images

Contrastive Learning) [104] envisioned contrastive learning to assist feature learning so that the semantic distance between pedestrians with distinct attire in the learned feature space is minimized to reduce appearance diversities.

It is challenging to detect pedestrians when environmental circumstances are like low light image conditions. Xinyu Jia *et al.* [84] presented a low-light visible-infrared paired (LLVIP) dataset for pedestrian detection. This dataset provides an extensive collection of infrared and visible images that can be used jointly to provide rich, precise details and effective target areas. This dataset contains 30,976 images or 15,488 paired images captured in shady surroundings. Figure 16 exhibits sample images from LLVIP [84] dataset. Various models used the LLVIP dataset to perform pedestrian detection in low-light occurrences, which includes YoloV3-RGB [108], YoloV5-RGB and CFT [109]. All these models' performances are compared with each other, and their results are shown in figure 17, where Average precision (AP) is considered the evaluation parameter. The Average Precision (AP) is calculated as:

$$\text{AP} = \sum_{r} \text{precision}(r) \cdot \Delta\text{recall}(r) \tag{3}$$

For model performance ranking, the higher the Average precision (AP) value, the better the model performance. Out of these all models, CFT [109] shows excellent performance over all other models. The Cross-Modality Fusion Transformer (CFT) employs a straightforward but efficacious cross-modality feature fusion method in the feature extraction stage that understands long-range dependencies and combines global contextual information guided by the Transformer architecture. The model can intuitively perform intra-modality and inter-modality fusion concurrently and reliably incorporates the hidden connections between RGB and thermal sectors by utilizing the Transformer's self-attention, which enhances the accuracy of multispectral pedestrian identification.

IV. Lane Detection

Modern vehicles are equipped with a variety of Advanced Driver Assistance Systems (ADAS), one of which is lane detection and tracking (LDaT) [110]. It is used in the process of autonomous driving and driver assistance. The goal of LDaT is to identify lanes on the roads and assist the driver in following them or entirely taking over the steering. Although most of the contemporary LDaT solutions are effective, there remains scope for development. They frequently have a variety of problems due to other objects, poor vision, and strange road shapes.

Since LDaT is such a hot issue, several papers [111, 112] outlining various solutions are produced every year. Depending on their solution types, these are generally classified into two groups: *traditional computer vision-based* [113] and *Deep Learning based* [110]. Both have advantages as well as cons. Denis Vajak *et al.* [114] conducted a survey that demonstrated that most earlier research approaches were based on traditional computer vision. Even deep neural network-based ones must rely on traditional computer vision to make the picture or data used for the deep neural network (DNN). Despite tremendous advances in the use of DNNs for many tasks, including LDaT, traditional computer vision-based lane identification remains the most often employed technique in most of the reviewed solutions. This is mainly because computer vision does not take a significant amount of resources, can be effectively implemented even on lower-spec hardware, and has long been used in sectors other than the automotive. Because of this, we have concentrated on computer vision-based solutions rather than Deep Learning-based approaches in this section. This section examines the most recent state-of-the-art computer vision-based LDaT datasets and models and provides a brief performance review. They are analyzed, categorized, and compared with each other. Table IV represents available datasets for lane detection.

The allocation of lanes on roads is contingent upon the road type, with variations reflecting distinct transportation needs

Dataset	Year	Images
CULane [115]	2018	133,235
TuSimple	2017	6,408
LLAMAS [116]	2019	100,042
CurveLanes [117]	2020	150,000

TABLE IV: Available Datasets for Lane detection tasks

[118]. Highways, characterized by their function of facilitating swift and high-volume traffic, typically boast four or more lanes per direction to optimize transportation efficiency [119]. This abundance of lanes in highway configurations caters to the imperative of managing heavy traffic volumes. Conversely, urban areas exhibit a nuanced lane distribution influenced by factors such as traffic density and urban development. The lanes in these settings may vary from two to single lanes per direction. In rural areas, a more streamlined infrastructure prevails, typically featuring single lanes. Many countries prescribe on-road driving guidelines that prescribe specific lane usage based on vehicle speed. These guidelines stipulate that as a vehicle accelerates beyond a designated speed threshold, a lane change becomes advisable. Similarly, when decelerating, vehicles are prompted to transition to an appropriate lane.

Xingang Pan *et al.* suggested the CULane [115] dataset to interpret traffic scenes and recognize traffic lanes. The dataset was created by driving six vehicles throughout Beijing by different drivers. The dataset contains 55 hours of footage, from which 133,235 on-road images were collected. Figure 18 demonstrate sample images of the CULane [115] dataset at different scenarios.

Fig. 18: Samples of the CULane [115] Dataset

Fig. 19 shows various models' performances on the CULane dataset using the F1 score as the evaluation parameter, wherein the higher the F1 score, the better the model performance. The F1 score is calculated as,

$$\text{F1} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{4}$$

Various models like UFAST [120], ENet-SAD [121], SCNN [122], PINet [123], CurveLane-L [124], RESA [125], LaneATT [126], SGNet [127], CondLaneNet-L [128] and CLRNet [129] utilized the CULane dataset for lane detection. Cross-Layer Refinement Network - CLRNet [129], which is based on DLA34 [130] delivers the most promising results with evaluation parameter as F1 score on the CULane dataset. Due to its model architecture which extracts high-level semantic features and then distills them using low-level features, it gives better results as compared to other models. Additionally, it provides more contextual details to detect lanes while maximizing detailed local road attributes to enhance localization accuracy. TuSimple is another popular

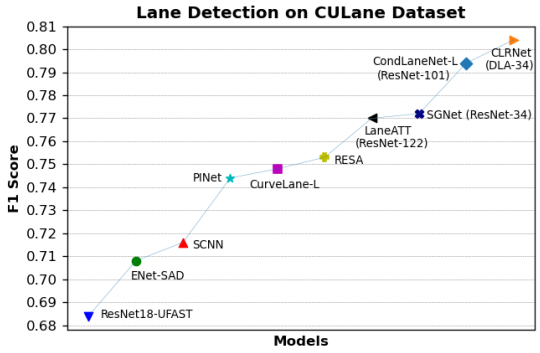

Fig. 19: Performance of models on CULane [115] Dataset

dataset for lane detection. It consists of 1280*720 resolution road photos taken on US roads. Dataset accomplishes various conditions, including weather conditions, different daytime, and various traffic conditions along with 2-lane, 3-lane, 4-lane, and highway road footage. Figure 20 illustrate sample images of the TuSimple dataset.

Fig. 20: Sample images of the TuSimple Dataset

The performance of several models on the TuSimple dataset is shown in Fig. 21 using accuracy as the evaluation parameter, whereas the higher the accuracy, the better the model performance. On TuSimple, several models, including PolyLaneNet [131], End-to-end ERFNet [132], ENet-SAD [121], RESA [125], FOLOLane [133] and SCNN-UNet-ConvLSTM2 [134] showcase their performances. Due to unique hybrid spatial-temporal (ST) sequence-to-one architecture, SCNN-UNet-ConvLSTM2 [134] outperforms all other models. The design takes full advantage of the spatial-temporal features in multiple continuous picture frames to identify lanes.

Karsten Behrendt and Ryan Soussan suggested the LLAMAS [116] dataset of large and diversified unsupervised labeled lane markers. The LLAMAS dataset contains 100,042 annotated lane marker images collected while driving across 14 highways of 25 km each, totaling 350 kilometers. Figure 22 show sample images of the LLAMAS [116] dataset. On the LLAMAS dataset, Fig. 23 illustrates the performance of various models using the F1 score as the evaluation parameter, whereas the higher the F1 score, the better the model performance. Using the LLAMAS dataset, different models

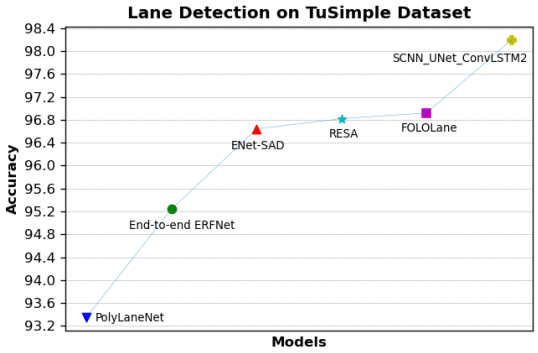

Fig. 21: Performance of models on TuSimple Dataset

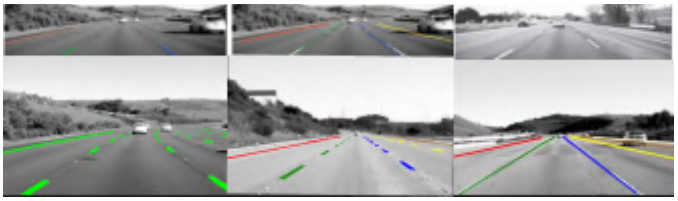

Fig. 22: Samples of the LLAMAS [116] Dataset

such as PolyLaneNet [131], LaneATT [126], BezierLaneNet [135], LaneAF [136] and CLRNet [129] demonstrates their lane-detecting abilities. CLRNet [129] surpasses the other models when the evaluation parameter is chosen as an F1 score. CLRNet uses DLA34 [130] as a backbone. The author also suggested ROIGather and Line Intersection over Union (LIoU), which enhances the effectiveness of lane identification.

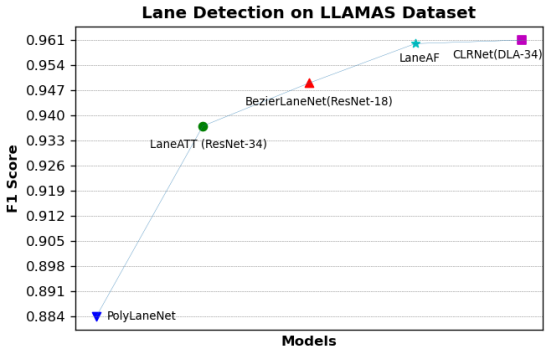

Fig. 23: Performance of models on LLAMAS[116] Dataset

Hang Xu *et al.* proposed CurveLanes [117] dataset, which consists of 150K images with 680K labels. Figure 24 demonstrate sample images of the CurveLanes [117] dataset. On the CurveLanes dataset, Fig. 25 presents the lane detection results of several models using the F1 score as the evaluation parameter, whereas the higher the F1 score, the better the model performance.

Fig. 24: Samples of the CurveLanes [117] Dataset

Several models, including Enet-SAD [121], SCNN [115, 122], PointLaneNet [137], CurveLane-L [117] and CondLaneNet-L [128] demonstrate their lane-detecting abilities using the CurveLane dataset and the F1 score as assessment criteria. The CondLaneNet-L [128] has the best performance of all of these models. It is a top-to-down lane detection framework that first recognizes the instances of lanes before dynamically predicting the shape of the lines for each occurrence using conditional convolution and row-wise formulation. Recurrent Instance Module (RIM) is also used to solve the issue of recognizing lane lines with complicated topologies, such as thick lines and fork lines.

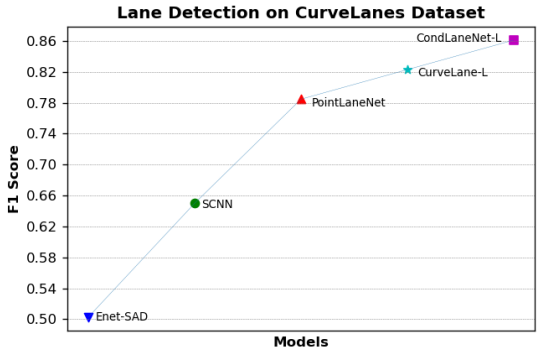

Fig. 25: Performance of models on CurveLanes[117] Dataset

V. Segmentation and Depth Estimation

To navigate correctly, self-driving cars must be able to recognize diverse objects on the road [138]. The world's perspective from a self-driving vehicle frequently comprises bounding boxes – cars, humans, and road signs neatly contained in rectangles. In reality, however, not everything fits into a box. It's advantageous for the vehicle's perception system to provide a more profound understanding of its surroundings in extremely complicated driving conditions, such as a construction zone marked by traffic cones or a pedestrian unloading a moving van with cargo sticking out the back. Segmentation helps autonomous driving cars identify the driveable regions of an image by dividing the visual information into pixels and assigning them labels. Using segmentation, the image may be accurately analyzed for semantic content and instance

content. The perception system's segmentation data can be used by planning and control modules to inform autonomous driving decisions better. For example, complex object shape and silhouette information aid in object tracking, resulting in more accurate steering and acceleration input. It can also be used in conjunction with dense (pixel-level) distance-to-object estimate algorithms to estimate a scene's 3D depth. The more accurately and quickly we accomplish segmentation, the more the vehicle knows the surrounding environment and makes the proper decision every time.

The segmentation task in the realm of autonomous vehicles is intricately stratified into distinct categories, each tailored to specific requisites and scenarios [139, 140]. Semantic segmentation, a fundamental classification paradigm, assigns each pixel to predefined categories such as road, car, or pedestrian [141]. This facilitates a high-level comprehension of the scene, transcending the need to distinguish between individual instances within a class. In autonomous vehicles, semantic segmentation proves instrumental for image classification, enriching scene understanding [139]. Instance segmentation, an evolutionary stride beyond semantic segmentation, not only categorizes pixels but also individuates between specific instances within the same category. For instance, when confronted with a scenario featuring a group of pedestrians, instance segmentation delineates each person separately, furnishing a nuanced understanding imperative for navigating intricate interactions [140]. This granularity is particularly vital in scenarios involving complex interactions with pedestrians and other objects, augmenting the autonomous vehicle's capacity for detailed environmental comprehension [140]. Panoptic segmentation amalgamates the virtues of both semantic and instance segmentation, affording a comprehensive perceptual lens. For on-road driving scenarios, in panoptic segmentation, every pixel undergoes dual characterization, receiving a class label denoting its semantic category (e.g., road, sidewalk) and a distinctive identifier for individual instances (e.g., cars, pedestrians). This dual annotation methodology equips the system with the capacity to extend its discernment beyond the mere recognition of specific objects, facilitating a nuanced understanding of the broader scene [141].

However, segmentation is difficult because of the complex interaction among pixels in each image frame and between succeeding frames. Despite the rapid development of new technologies such as deep learning, which have made segmentation more efficient, conducting accurate segmentation in real-time remains a hot topic in current research. This section briefly overviews the most recent datasets and models available for autonomous driving on-road segmentation. Furthermore, several models have been tested and compared based on segmentation efficacy and capacity. Table V represents available datasets for segmentation. Out of Cityscapes [142, 143], Mapillary [144], COCO [145], KITTI [51], ApolloScape [146] and BDD100K [52] datasets, several models utilized Cityscapes [142, 143], Mapillary [144], COCO [145] and KITTI [51] datasets only for segmentation tasks. Therefore, we solely used these datasets to compare models. Marius Cordts *et al.* proposed a cityscapes dataset for visual understanding of complex urban street scenes. The cityscapes dataset [142, 143] contains diverse stereo video sequences recorded in streets from 50 different cities in Germany and neighboring countries. It includes 5000 images that have high-quality pixel-level annotations and 20,000 images with coarse annotations promote techniques that leverage large volumes of weakly-labeled data. Figure 26 shows sample images of the cityscapes [142, 143] dataset.

Dataset	Year	Images
Cityscapes [142, 143]	2016	25,000
Mapillary [144]	2017	25,000
COCO [145]	2017	328,000
KITTI [51]	2012	15,000
ApolloScape Dataset [146]	2019	140,000
BDD100K [52]	2018	100,000

TABLE V: Available Datasets for Segmentation tasks

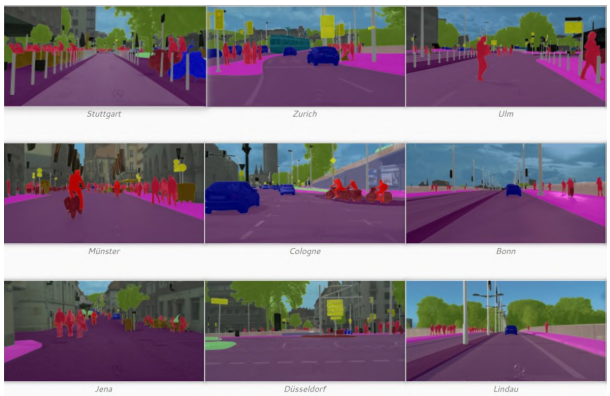

Fig. 26: Samples Images of the cityscapes [142, 143] Dataset. **1st Row:** Stuttgart, Zurich, Ulm **2nd Row:** Munster, Cologne, Bonn **3rd Row:** Jena, Dusseldorf, Lindau

Many models have utilized cityscapes dataset [142, 143] to perform segmentation tasks. This includes models such as DeepLab [147, 148], Context [149], LRR-4x [150], RefineNet [151], PSPNet++ [152], DeepLabv3 [153], SSMA [154], MRFM [155], GALDNet [156] and ViT-Adapter-L [157], which utilizes Mean IoU (Intersection over Union) as an evaluation parameter. The Mean Intersection over Union (IoU) is calculated as:

$$\text{Mean IoU} = \frac{1}{NC}\sum_{i=1}^{NC}\frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i + \text{FN}_i} \tag{5}$$

where:

NC is the number of classes.

TP_i is the number of true positives for class i,

FP_i is the number of false positives for class i,

FN_i is the number of false negatives for class i.

Figure 27 represents the segmentation performance of these models. Among these models, the ViT-Adapter-L [157] out-

performs all others. ViT-Adapter-L achieves the greatest Mean IoU (Intersection over Union) value (85.2) for segmentation on cityscapes test dataset [142, 143]. ViT-Adapter-L uses a spatial prior module and two feature interaction operators, which adopt the required local continuity characteristics of ViT and rearrange fine-grained multi-scale features.

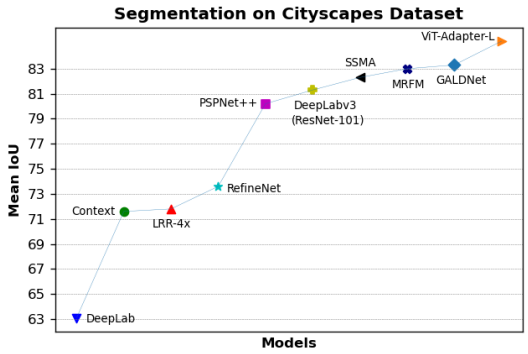

Fig. 27: Performance of models on Cityscapes [142, 143] Dataset

Another popular and diverse segmentation dataset is the Mapillary Vistas Dataset [144]. This is a substantial street-level image dataset featuring 25,000 high-resolution photographs of urban, rural, and off-road situations that are annotated in 66 different categories and 37 classes. Compared to the cityscapes dataset, the Mapillary Vistas Dataset has more fine-grained annotations. The mapillary dataset was collected from portions of Europe, North and South America, Asia, Africa, and Oceania. Many researchers have used the Maplillary dataset for their research because of the enormous volume of annotations and the high-quality on-road images. Figure 28 exhibit sample images of the Maplillary [144] dataset.

Fig. 28: Sample images of the Maplillary [144] Dataset

Several models, like JSIS-Net [158], AdaptIS [159], Panoptic-DeepLab (X71) [160], EfficientPS [161], Axial-DeepLab-L [162] and SWideRNets [163] have employed Maplillary datasets to conduct segmentation. Figure 29 depicts the performance of various models using panoptic quality (PQ) as an evaluation criterion, wherein higher panoptic quality (PQ) is considered as better the model performance. The Panoptic Quality (PQ) is calculated as:

$$\mathrm{PQ} = \mathrm{PQ}_{\text{semantic}} \times \mathrm{PQ}_{\text{instance}} \tag{6}$$

where:

$$\mathrm{PQ}_{\text{semantic}} = \frac{\sum_i \mathrm{IoU}_{\text{semantic}}(i) \cdot \mathrm{TP}_{\text{semantic}}(i)}{(\sum_i \mathrm{TP}_{\text{semantic}}(i) + \mathrm{FP}_{\text{semantic}}(i) + \mathrm{FN}_{\text{semantic}}(i))},$$

$$\mathrm{PQ}_{\text{instance}} = \frac{\sum_i \mathrm{IoU}_{\text{instance}}(i) \cdot \mathrm{TP}_{\text{instance}}(i)}{\left\{ \begin{array}{c} (\sum_i (\mathrm{TP}_{\text{instance}}(i) + 0.5 \cdot \mathrm{FP}_{\text{instance}}(i)) \\ +0.5 \cdot \mathrm{FN}_{\text{instance}}(i)) \end{array} \right\}}$$

Out of these all models, the SWideRNets [163] model provides excellent results on the Maplillary dataset as compared to other models. Scaling Wide Residual Networks, aka SWideRNets utilizes Wide Residual Networks (WR-41) [164, 165] as a base model on top, it integrates the simplified Squeeze-and-Excitation (SE) module [166, 167] and Switchable Atrous Convolution (SAC) [168].

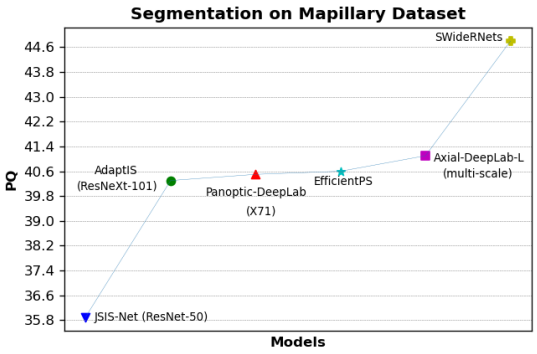

Fig. 29: Performance of models on Maplillary [144] Dataset

Another popular dataset in the segmentation domain is MS COCO (Microsoft Common Objects in Context) dataset [145]. It includes 328,000 highly annotated images containing 91 labels. Figure 30 demonstrate sample images of the MS COCO [145] dataset.

Fig. 30: Sample images of the MS COCO [145] Dataset

Figure 31 demonstrates the performance of models such as MultiPath Network [169], FCIS++ +OHEM [170], Mask R-CNN [67], MaskLab+ [171], PANet [172], CBNet [173], SpineNet [174], DetectoRS [168], Swin-L [175], Soft Teacher + Swin-L [176], SwinV2-G [177] and FD-SwinV2-G [178]

that used the MS COCO dataset. These models used Mask AP (Average precision) as an evaluation measure. For model performance ranking for segmentation, the model which shows higher Mask AP, is considered as better performance as compared to all others. Out of these all models, FD-SwinV2-G [178] shows the best results over MS COCO dataset. Its unique architecture, feature distillation (FD) methods, and masked image modeling (MIM) algorithm improve fine-tuning performance.

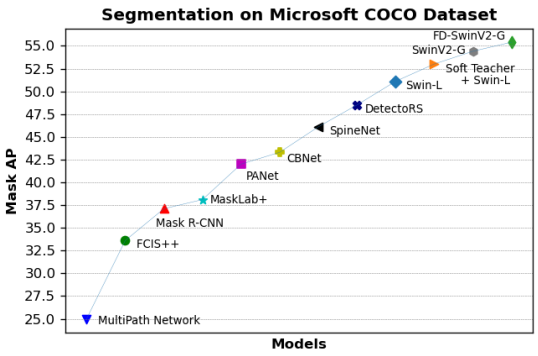

Fig. 31: Performance of models on MS COCO [145] Dataset

Andreas Geiger *et al.* [51] proposed the KITTI dataset, which is widely used in segmentation. It is formed up of hours of transport scenarios acquired using a range of sensor modalities, such as high-resolution RGB cameras, grayscale stereo cameras, and a 3D laser scanner. Despite its prominence, the dataset lacks in ground truth for semantic segmentation. However, other researchers have manually annotated portions of the dataset to meet their needs. Alvarez *et al.* [179, 180] created ground truth for 323 photos from the road detection challenge, categorizing them as road, vertical, and sky. Zhang *et al.* [181] annotated 252 RGB and Velodyne scans (140 for training and 112 for testing) from the tracking challenge for 10 object categories: building, sky, road, vegetation, sidewalk, car, pedestrian, bicycle, sign/pole, and fence. Ros *et al.* [182] assigned 11 classes to 170 training photos and 46 testing images from the visual odometry challenge: building, tree, sky, car, sign, road, pedestrian, fence, pole, sidewalk, and biker. Figure 32 illustrates sample images of the KITTI dataset.

Fig. 32: Sample images of the KITTI Dataset

Various models utilized KITTI dataset for segmentation purposes which includes APMoE-seg [183], SegStereo [184], AHiSS [185], MapillaryAI [186] and DeepLabV3Plus + SDCNetAug [187]. These models used mean IOU (intersection over union) as an evaluation parameter. Figure 33 shows all model performance comparisons with respect to each other. Out of these all models, DeepLabV3Plus + SDCNetAug [187] shows excellent segmentation results over the KITTI dataset. The authors of model [187] established a unique framework for a video prediction-based approach to scale training datasets by synthesizing fresh training samples to increase the accuracy of semantic segmentation networks.

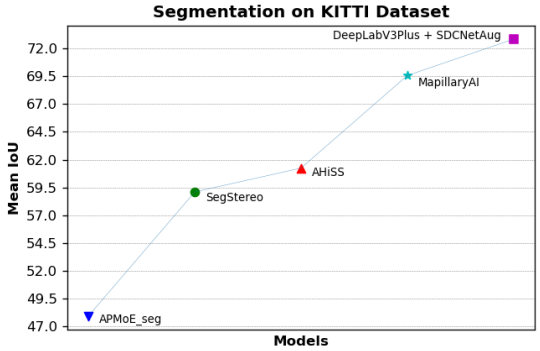

Fig. 33: Performance of models on KITTI [51] Dataset

One of the essential tasks in the autonomous driving overtaking process is estimating an accurate depth map from a camera RGB picture. It is crucial in scene understanding. It is possible to predict the depth of front objects such as cars, barriers, or pedestrians using computer vision algorithms. Using depth prediction, an autonomous vehicle can measure the distance from the front car and maintain a safe distance while executing overtaking with regard to the front vehicle.
In depth estimation domain, a significant amount of research has been done, but there are still many challenges, and issues are open because of its inherent ambiguity [188]. When compared to traditional feature-based approaches [189], supervised [190] and stereo self-supervised [191] learning have been shown to outperform them. However, these solutions need either a significant quantity of high-quality annotated ground truth that is hard to obtain or sophisticated stereo calibration. As a result, this section covers the most recent cutting-edge computer vision-based depth estimation datasets and models and provides a brief performance evaluation. They are evaluated, classified, and compared to one another also. Table VI represents available datasets for depth estimation. Out of KITTI Dataset [192], Cityscapes [142, 143], DIML/CVL [193], DrivingStereo [194] and DDAD [195] dataset most models utilized KITTI [192] dataset for depth estimation tasks. Therefore, we solely used KITTI Dataset [192] dataset to compare models. One of the most used depth estimation datasets for autonomous driving is KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) [192]. It comprises hours' worth of driving conditions captured using a range of sensor modalities, such as high-resolution RGB, grayscale stereo, and 3D laser scanner cameras. It has about 94 thousand depth maps aligned with the raw data of the KITTI dataset and includes accompanying raw LiDAR scans and RGB pictures. This dataset will enable the construction

Dataset	Year	Images
KITTI Dataset [192]	2017	94,000
Cityscapes [142, 143]	2016	25,000
DIML/CVL Dataset [193]	2021	55,577
DrivingStereo [194]	2019	180,000
DDAD [195]	2020	99600

TABLE VI: Available Datasets for Depth estimation tasks

of sophisticated deep-learning algorithms for depth completion and single-picture depth prediction objectives. Figure 34 exhibits sample images from KITTI [192] dataset.

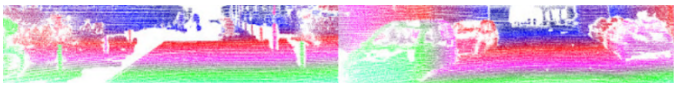

Fig. 34: Samples of the KITTI [192] Dataset

The performance of several models, such as BinsFormer [196], Depthformer [197], MonoDELSNet [198], SfM-Revisited [199], AdaBins [200], DORN [201], DPT-Hybrid [202], GCNDepth [203] SC-Depth [204] and DNET [205] that employed the KITTI dataset for depth estimation is shown in figure 35. The RMSE (Root Mean Square Error) measure is used to assess the model's performance. In performance evaluation, the lower the RMSE value, the better the model performance. As a result, BinsFormer [196], a classification-regression-based depth estimation model, beats all other models. It mainly concentrates on appropriate adaptive bin generation and adequate relationships between probability distribution and predictions of adaptive bins. It utilizes a transformer decoder to produce bins. In order to fully comprehend spatial geometry details and generate depth maps in a coarse-to-fine way, it also incorporates a multi-scale decoder structure.

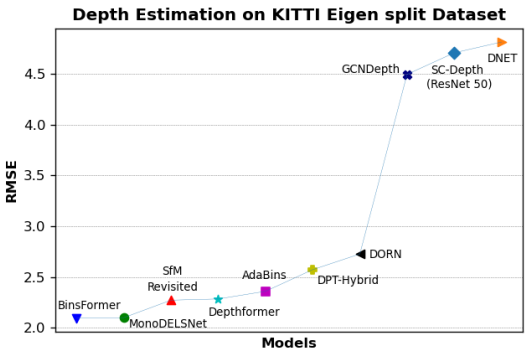

Fig. 35: Performance of models on KITTI dataset [192]

VI. Obstacle and Road Sign Detection

The detection and identification of road indicators are significant in overtaking procedures. While performing overtaking, the autonomous driving car must recognize the road signboard and determine if it is appropriate to execute overtaking or not. For example, overtaking on a straight road is usually safer than on any curve/turn road or bridge. So, whenever there is a turn ahead, the road sign board indicates it so self-driving cars can avoid overtaking on curved roads.

In the realm of autonomous driving, obstacle detection and road sign detection are fundamental components, each serving distinct roles despite their shared objective of enhancing the safety and efficiency of the overtaking process [206, 207]. The key differentiators lie in their viewpoints, functionalities, and the information they provide to the autonomous vehicle (AV) [208, 209]. Obstacle detection primarily focuses on on-road activities, ensuring the AV's ability to navigate safely and avoid collisions with obstacles in its path [206]. This system is crucial for maintaining a safe distance from vehicles in the front, thereby preventing potential collisions. Additionally, obstacle detection plays a pivotal role in free path following planning, enabling the AV to plan overtaking maneuvers by identifying clear paths and ensuring collision-free navigation [207, 208]. Conversely, road sign detection operates with a broader perspective, scanning the sides and tops of roads to identify various traffic signs. Its primary function is to recognize and interpret the meaning or indication of these signs, providing essential information about upcoming on-road activities [209]. This includes identifying signs indicating narrow roads, pedestrian crossings, or speed limits. The information gleaned from road sign detection is integral in the decision-making process for overtaking maneuvers. For instance, when road signs indicate a narrow road or a specific speed limit, the AV adjusts its strategy, refraining from overtaking even if a clear path is available. This aligns with the vehicle's adherence to on-road driving guidelines and ensures a judicious approach to overtaking based on the prevailing road conditions. Currently, there exist an approximate repertoire of 200 distinct traffic signs [210]. These signs are broadly categorized into three classes based on their priority and the degree of adherence required from the driver, encompassing mandatory signs, cautionary signs, and informatory signs [211, 212]. Mandatory signs mandate specific instructions or prohibitions and include indications such as Straight Prohibitor, Pedestrian Prohibited, No Parking, Speed Limit, among others. Cautionary signs, conversely, forewarn road users about potential hazards or dangers, incorporating symbols like Right-Hand Curve, Left-Hand Curve, Steep Ascent, Steep Descent, Narrow Road Ahead, Pedestrian Crossing, School Ahead, and others. Simultaneously, informatory signs fulfill an informative role, offering details about locations, distances, and directions to specific destinations, featuring representations like Petrol Pump, Hospital, First Aid Place, Park This Side, and additional variants.

Many real-world services rely on traffic sign recognition systems (TSRS), including autonomous driving, traffic monitoring, safe driving and assist, and traffic scene analysis. A TSRS typically addresses two related topics: traffic sign detection (TSD) [213] and traffic sign recognition (TSR) [214, 215]. Traffic sign detection concentrates on the positioning of objects in images, whereas the traffic sign identification algorithm classifies in order to determine the type of targets discovered. However, due to real-world unpredictability, such as scale variations, poor visibility, motion blur, fading colors,

occlusions, and lightning circumstances, developing a solid real-time TSRS remains a difficult challenge [216, 217]. Several algorithms have been presented, and sophisticated driver assistance systems that detect and recognize traffic signals have hit the market. However, despite the numerous competing techniques, no clear consensus exists on the state-of-the-art in this sector [216]. This can be attributed to a lack of resources and datasets. Consequently, this section gives a brief performance review and covers the most significant computer vision-based traffic sign datasets and models [214]. They are also assessed, ranked, and compared to one another.

Table VII represents available datasets for Obstacle and Road Signs Detection. Out of Tsinghua-Tencent [218], GTSRB [219], Bosch Small Traffic Lights [220], Swedish traffic-sign dataset (STSD) [221], and European Dataset [222] several models utilized Tsinghua-Tencent [218] and GTSRB [219] datasets only for obstacle and road signs detection tasks. Therefore, we solely used these datasets to compare models. Zhe Zhu *et al.* [218] suggested a Tsinghua-Tencent dataset

Dataset	Year	Images
Tsinghua-Tencent [218]	2021	100,000
GTSRB [219]	2013	50,000
Bosch Small Traffic Lights Dataset [220]	2017	13,427
Swedish traffic-sign dataset (STSD) [221]	2011	20,000
The European Dataset [222]	2018	80,000

TABLE VII: Available Datasets for Obstacle and Road Signs Detection tasks

for traffic-sign detection and classification. It collects 100000 panoramas comprising 30000 traffic-sign instances from the Tencent Data Center and covers ten areas from 5 distinct cities in China (covering both downtown regions and suburbs for each city). These images cover a wide range of lighting and weather situations. Figure 36 exhibits sample images of the Tsinghua-Tencent [218] dataset.

Fig. 36: Sample images of the Tsinghua-Tencent [218] Dataset

Figure 37 depicts the performance of several models, such as Hierarchical Model [223], Hierarchical + Background Threshold Model [223], Background Threshold Model [223], TSR-SA (without receptive field block-cross (RFB-C)) [224] and TSR-SA (with receptive field block-cross (RFB-C)) [224]. The mAP (mean Average Precision) measure is used to assess the model's performance. In performance evaluation, the higher the mAP value, the better the model performance. In all these models, Hierarchical Model [223] shows better performance over all other models.

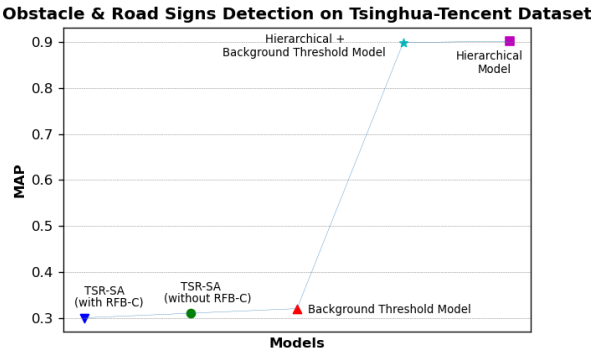

Fig. 37: Performance of models on Tsinghua-Tencent Dataset [218]

Another most popular dataset for depth estimation is GTSRB [219]. Sebastian Houben *et al.* [219] proposed a German Traffic Sign Detection Benchmark (GTSRB) dataset containing various traffic signs. The dataset was captured near Bochum, Germany. The author captured urban and rural surroundings as well as roads during daytime and dusk featuring weather conditions. The images in the dataset have a resolution of 1360 x 1024 pixels, and the traffic sign sizes range from 16 to 128 pixels. Figure 38 exhibit sample images of the GTSRB [219] dataset.

Fig. 38: Sample images of the GTSRB [219] Dataset

Figure 39 displays the models' individual performance on the GTSRB dataset. Besides that, it demonstrates the evaluation results of all the models by considering accuracy as the evaluation criteria. The model's performance is deemed to be superior when it displays a higher level of accuracy. Out of SEER [225], MicronNet [226], MCDNN [227], SillNet [228], CNN with 3 Spatial Transformers [229] which are models that utilized GTSRB dataset for depth estimation, CNN with 3 Spatial Transformers [229] shows the best performance. It is based on a spatial transformer network (STN) equipped convolutional neural network (CNN).

VII. Steering Angle Computation

Intelligent, self-driving cars must be able to move without leaving the drivable portion of the road, which is a vital

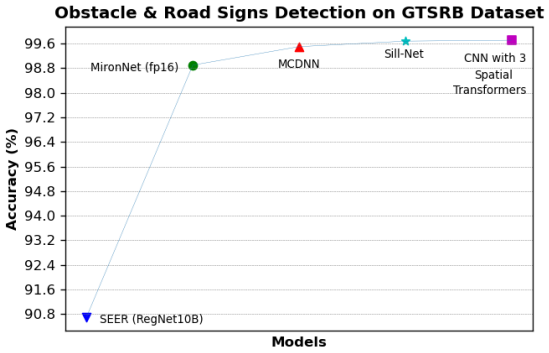

Fig. 39: Performance of models on GTSRB Dataset [219]

necessity. Steering angle computation is crucial to meet safety-critical requirements while conducting the overtaking maneuver and improve the safety and interpretability of end-to-end driving. It keeps the car in the middle of the road or within the boundary lanes. Computer vision-based steering angle calculation approaches [230, 231] are incredibly productive and affordable in these circumstances. The scope of this review is limited to the more recent and relevant techniques, models, and datasets based on the computer vision approach. This section briefly overviews the different datasets, models, and computer vision-based steering angle calculation approaches. Furthermore, many models are evaluated and contrasted against one another depending on the effectiveness and capacity of steering angle computation. Table VIII represents available datasets for steering angle computation. Out of CARLA [232], Steering angle computation dataset [233] and Steering angle dataset [234], numerous models utilized CARLA [232] dataset only for steering angle computation tasks. Therefore, we solely used CARLA [232] dataset to compare models.

Dataset	Year
CARLA Dataset [232]	2017
Steering angle computation dataset [233]	2016
Steering angle dataset [234]	2021

TABLE VIII: Available Datasets for steering angle computation tasks

Alexey Dosovitskiy *et al.* [232] proposed a simulator for autonomous driving research. CARLA is an open-source simulator that can simulate various use cases of self-driving. It also provides urban layouts, buildings, and vehicles that create a completely realistic environment for autonomous driving. Using CARLA, the [232] generated steering angle computation. Using this dataset many models, such as LBC [235], MaRLn [236], World on Rails [237], GRIAD [238], Latent TransFuser [239], TransFuser [239], LAV [240], TCP [241] and InterFuser [242] calculated steering angle computation. The performance of all these models is shown in Fig. 40. The driving score is taken into consideration as an evaluation factor when determining how well each of these models performs. While ranking the model performance, the higher the driving score, the better the model ranking. Out of the mentioned models, InterFuser [242] shows excellent results among all these models. InterFuser, an interpretable sensor fusion transformer, promotes reasoning and global contextual perception across various modalities. Additionally, by exposing intermediate aspects of the model and limiting actions to safe sets, the InterFuser framework also improves the safety and interpretability of end-to-end driving.

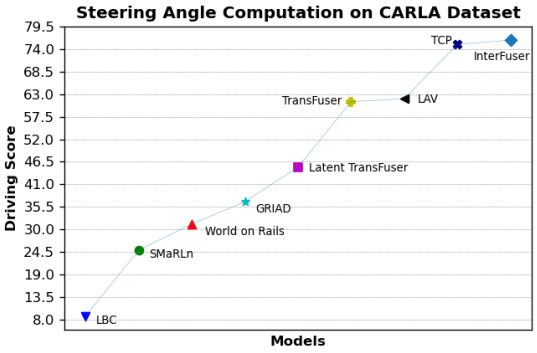

Fig. 40: Performance of models on CARLA [232] Dataset

VIII. Future Research Directions and Open Challenges

Although autonomous driving has seen humongous progress, there is always scope for improvement that can only be taken through research.

- The first area that can be improved involves sensors. To help the vehicle's control system decide where to steer or when to brake, sensors in autonomous vehicles map the environment [243]. Accurate sensors are necessary while overtaking manoeuvres to detect objects, distances, speeds, and other factors in any situation [244, 245]. The camera's detecting capacity can be significantly impacted by bad weather, heavy traffic, and confusing road signs [246, 247].
- The potential and efficacy of machine learning algorithms play a crucial role in the overtaking process [248, 249]. Many self-driving cars rely on machine learning (ML) and computer vision techniques to analyze sensor data, enhance object classification, detect distances and movements, and make informed decisions. These aspects are vital for executing successful overtaking manoeuvres. However, it is essential to note that the evaluation of machine learning algorithms in real-world scenarios is currently limited. While machine learning models demonstrate outstanding performance on synthetic datasets or in simulation environments, their effectiveness tends to falter when deployed in real-world conditions. Consequently, there is a pressing need to enhance the efficiency and accuracy of machine learning algorithms in real-world settings. By addressing these limitations, we can

ensure that overtaking manoeuvres are executed reliably, without failures, and with higher levels of accuracy.
- There is need for developing laws for specific operations of autonomous vehicles, like automated lane-keeping systems. Allowing autonomous vehicles to operate on public roads without established rules and guidelines is dangerous. There should be ethics and regulations for self-driving cars [250, 251].

Hence, there are areas that need attention and stand as open challenges for future research on overtaking manoeuvres.

IX. Conclusion

Augmented intelligence seamlessly intertwines with the foundation of self-driving cars, enhancing the prowess of artificial intelligence (AI) through symbiotic collaborations with human expertise. In the realm of sensor fusion and perception, engineers leverage augmented intelligence to amalgamate data from diverse sensors, such as cameras, LiDAR crafting a holistic understanding of the vehicle's environment. This augmented intelligence approach harnesses the strengths of both artificial and human intelligence, propelling the advancement of safe and effective autonomous driving technologies. The autonomous driving field has attracted a lot of research, with various successful efforts to turn the idea into reality. However, there have been various unfortunate events concerning autonomous vehicle accidents. When traced back, it is revealed that a lack of accuracy in overtaking mechanisms has contributed to a more significant proportion of such cases. It makes sense to initiate and facilitate research into overtaking in autonomous vehicles. For this very purpose, this review paper touches upon the various domains of autonomous driving in general and overtaking in particular. In this survey, we thoroughly analyzed several crucial computer vision domains while performing overtaking. We have analyzed the different datasets and state-of-the-art models available for computer vision-based overtaking tasks for autonomous vehicles. We provided model comparisons using the dataset and standard evaluation parameters to get a clear insight into each model's performance over other models. Finally, based on the survey, we list the significant challenges and future research directions in this domain. This survey will guide the researchers and professionals venturing into the research and development of overtaking solutions based on computer vision for autonomous driving.

References

[1] Unsplash. Accessed: May 13, 2024. [Online]. Available: https://unsplash.com/

[2] N. Merat, A. H. Jamson, F. C. Lai, M. Daly, and O. M. Carsten, "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," *Transportation research part F: traffic psychology and behaviour*, vol. 27, pp. 274–282, 2014.

[3] O. Carsten and M. H. Martens, "How can humans understand their automated cars? hmi principles, problems and solutions," *Cognition, Technology & Work*, vol. 21, no. 1, pp. 3–20, 2019.

[4] C. Ding, C. Li, Z. Xiong, Z. Li, and Q. Liang, "Intelligent identification of moving trajectory of autonomous vehicle based on friction nano-generator," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[5] X. Zhao, Y. Fang, H. Min, X. Wu, W. Wang, and R. Teixeira, "Potential sources of sensor data anomalies for autonomous vehicles: An overview from road vehicle safety perspective," *Expert Systems with Applications*, p. 121358, 2023.

[6] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and P. Zhao, "An incentive mechanism of incorporating supervision game for federated learning in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[7] H. Yang, Z. Li, and Y. Qi, "Predicting traffic propagation flow in urban road network with multi-graph convolutional network," *Complex & Intelligent Systems*, pp. 1–13, 2023.

[8] Y. Shi, J. Xi, D. Hu, Z. Cai, and K. Xu, "Raymvsnet++: learning ray-based 1d implicit fields for accurate multi-view stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[9] B. Cao, Z. Li, X. Liu, Z. Lv, and H. He, "Mobility-aware multiobjective task offloading for vehicular edge computing in digital twin environment," *IEEE Journal on Selected Areas in Communications*, 2023.

[10] Y. Chen, "Research on collaborative innovation of key common technologies in new energy vehicle industry based on digital twin technology," *Energy Reports*, vol. 8, pp. 15 399–15 407, 2022.

[11] Z. Lin, H. Wang, and S. Li, "Pavement anomaly detection based on transformer and self-supervised learning," *Automation in Construction*, vol. 143, p. 104544, 2022.

[12] H. Zhang, G. Luo, J. Li, and F.-Y. Wang, "C2fda: Coarse-to-fine domain adaptation for traffic object detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12 633–12 647, 2021.

[13] J. Chen, Q. Wang, W. Peng, H. Xu, X. Li, and W. Xu, "Disparity-based multiscale fusion network for transportation detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18 855–18 863, 2022.

[14] J. Xu, X. Zhang, S. H. Park, and K. Guo, "The alleviation of perceptual blindness during driving in urban areas guided by saccades recommendation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16 386–16 396, 2022.

[15] J. Xu, S. H. Park, X. Zhang, and J. Hu, "The improvement of road driving safety guided by visual inattentional blindness," *IEEE transactions on intelligent transportation systems*, vol. 23, no. 6, pp. 4972–4981, 2021.

[16] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Perception task offloading with collaborative computation for autonomous driving," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 2, pp. 457–473, 2022.

[17] R. Fukatsu and K. Sakaguchi, "Automated driving with cooperative perception using millimeter-wave v2v communications for safe overtaking," *Sensors*, vol. 21, no. 8, p. 2659, 2021.

[18] Y. Fang, H. Min, X. Wu, W. Wang, X. Zhao, and G. Mao, "On-ramp merging strategies of connected and automated vehicles considering communication delay," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 298–15 312, 2022.

[19] Y. Yao, F. Shu, Z. Li, X. Cheng, and L. Wu, "Secure transmission scheme based on joint radar and communication in mobile vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[20] X. Zhang, S. Wen, L. Yan, J. Feng, and Y. Xia, "A hybrid-convolution spatial–temporal recurrent network for traffic flow prediction," *The Computer Journal*, p. bxac171, 2022.

[21] H. Yang, X. Zhang, Z. Li, and J. Cui, "Region-level traffic prediction based on temporal multi-spatial dependence graph convolutional network from gps data," *Remote Sensing*, vol. 14, no. 2, p. 303, 2022.

[22] H. Tehrani, Q. H. Do, M. Egawa, K. Muto, K. Yoneda, and S. Mita, "General behavior and motion model for automated lane change," in *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2015, pp. 1154–1159.

[23] H. Min, Y. Li, X. Wu, W. Wang, L. Chen, and X. Zhao, "A measurement scheduling method for multi-vehicle cooperative localization considering state correlation," *Vehicular Communications*, vol. 44, p. 100682, 2023.

[24] J. Chen, M. Xu, W. Xu, D. Li, W. Peng, and H. Xu, "A flow feedback traffic prediction based on visual quantified features," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[25] J. Chen, Q. Wang, H. H. Cheng, W. Peng, and W. Xu, "A review of vision-based traffic semantic understanding in itss," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[26] X. Ma, Z. Dong, W. Quan, Y. Dong, and Y. Tan, "Real-time assessment of asphalt pavement moduli and traffic loads using monitoring data from built-in sensors: Optimal sensor placement and identification algorithm," *Mechanical Systems and Signal Processing*, vol. 187, p.

109930, 2023.
[27] J. Xu, K. Guo, and P. Z. Sun, "Driving performance under violations of traffic rules: Novice vs. experienced drivers," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 908–917, 2022.
[28] Y.-C. Lin, C.-L. Lin, S.-T. Huang, and C.-H. Kuo, "Implementation of an autonomous overtaking system based on time to lane crossing estimation and model predictive control," *Electronics*, vol. 10, no. 18, p. 2293, 2021.
[29] S. Jiang, C. Zhao, Y. Zhu, C. Wang, Y. Du *et al.*, "A practical and economical ultra-wideband base station placement approach for indoor autonomous driving systems," *Journal of Advanced Transportation*, vol. 2022, 2022.
[30] H. Min, Y. Fang, X. Wu, X. Lei, S. Chen, R. Teixeira, B. Zhu, X. Zhao, and Z. Xu, "A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis," *Expert Systems with Applications*, vol. 224, p. 120002, 2023.
[31] Z. Xiao, J. Shu, H. Jiang, G. Min, J. Liang, and A. Iyengar, "Toward collaborative occlusion-free perception in connected autonomous vehicles," *IEEE Transactions on Mobile Computing*, 2023.
[32] X. Zhang, S. Fang, Y. Shen, X. Yuan, and Z. Lu, "Hierarchical velocity optimization for connected automated vehicles with cellular vehicle-to-everything communication at continuous signalized intersections," *IEEE Transactions on Intelligent Transportation Systems*, 2023.
[33] J. Xu, S. Pan, P. Z. Sun, S. H. Park, and K. Guo, "Human-factors-in-driving-loop: Driver identification and verification via a deep learning approach using psychological behavioral data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3383–3394, 2022.
[34] S. Dixit, S. Fallah, U. Montanaro, M. Dianati, A. Stevens, F. Mccullough, and A. Mouzakitis, "Trajectory planning and tracking for autonomous overtaking: State-of-the-art and future prospects," *Annual Reviews in Control*, vol. 45, pp. 76–86, 2018.
[35] O. T. Ritchie, D. G. Watson, N. Griffiths, J. Misyak, N. Chater, Z. Xu, and A. Mouzakitis, "How should autonomous vehicles overtake other drivers?" *Transportation research part F: traffic psychology and behaviour*, vol. 66, pp. 406–418, 2019.
[36] T. Hegedűs, B. Németh, and P. Gáspár, "Challenges and possibilities of overtaking strategies for autonomous vehicles," *Periodica Polytechnica Transportation Engineering*, vol. 48, no. 4, pp. 320–326, 2020.
[37] P. S. Perumal, M. Sujasree, S. Chavhan, D. Gupta, V. Mukthineni, S. R. Shimgekar, A. Khanna, and G. Fortino, "An insight into crash avoidance and overtaking advice systems for autonomous vehicles: A review, challenges and solutions," *Engineering applications of artificial intelligence*, vol. 104, p. 104406, 2021.
[38] A.-M. Sourelli, R. Welsh, and P. Thomas, "Objective and perceived risk in overtaking: The impact of driving context," *Transportation research part F: traffic psychology and behaviour*, vol. 81, pp. 190–200, 2021.
[39] S. S. Lodh, N. Kumar, and P. K. Pandey, "Autonomous vehicular overtaking maneuver: A survey and taxonomy," *Vehicular Communications*, p. 100623, 2023.
[40] S. Yu, C. Zhao, L. Song, Y. Li, and Y. Du, "Understanding traffic bottlenecks of long freeway tunnels based on a novel location-dependent lighting-related car-following model," *Tunnelling and Underground Space Technology*, vol. 136, p. 105098, 2023.
[41] Y. Zheng, P. Liu, L. Qian, S. Qin, X. Liu, Y. Ma, and G. Cheng, "Recognition and depth estimation of ships based on binocular stereo vision," *Journal of Marine Science and Engineering*, vol. 10, no. 8, p. 1153, 2022.
[42] G. Al-refai and M. Al-refai, "Road object detection using yolov3 and kitti dataset," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, 2020.
[43] M. Haris and A. Glowacz, "Road object detection: A comparative study of deep learning-based algorithms," *Electronics*, vol. 10, no. 16, p. 1932, 2021.
[44] Z. Lian, Y. Nie, F. Kong, and B. Dai, "Oro-yolo: An improved yolo algorithm for on-road object detection," in *International Conference on Autonomous Unmanned Systems*. Springer, 2022, pp. 3653–3664.
[45] R. Ghosh, "A modified yolo model for on-road vehicle detection in varying weather conditions," *Intelligent Computing and Communication Systems*, pp. 45–54, 2021.
[46] W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian detection based on yolo network model," in *2018 IEEE international conference on mechatronics and automation (ICMA)*. IEEE, 2018, pp. 1547–1551.
[47] W.-Y. Hsu and W.-Y. Lin, "Ratio-and-scale-aware yolo for pedestrian detection," *IEEE transactions on image processing*, vol. 30, pp. 934–947, 2020.
[48] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object detection using convolutional neural networks," in *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018, pp. 2023–2027.
[49] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*. Springer, 2020, pp. 213–229.
[50] J. Beal, E. Kim, E. Tzeng, D. H. Park, A. Zhai, and D. Kislyuk, "Toward transformer-based object detection," *arXiv preprint arXiv:2012.09958*, 2020.
[51] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
[52] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.
[53] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.
[54] L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M.-H. Yang, and S. Lyu, "Ua-detrac: A new benchmark and protocol for multi-object detection and tracking," *Computer Vision and Image Understanding*, vol. 193, p. 102907, 2020.
[55] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4490–4499.
[56] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum pointnets for 3d object detection from rgb-d data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 918–927.
[57] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3d proposal generation and object detection from view aggregation," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1–8.
[58] X. Du, M. H. Ang, S. Karaman, and D. Rus, "A general pipeline for 3d detection of vehicles," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3194–3200.
[59] S. Shi, X. Wang, and H. Li, "Pointrcnn: 3d object proposal generation and detection from point cloud," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 770–779.
[60] Z. Wang and K. Jia, "Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3d object detection," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1742–1749.
[61] Z. Yang, Y. Sun, S. Liu, X. Shen, and J. Jia, "Std: Sparse-to-dense 3d object detector for point cloud," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1951–1960.
[62] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "Pv-rcnn: Point-voxel feature set abstraction for 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 529–10 538.
[63] W. Zheng, W. Tang, L. Jiang, and C.-W. Fu, "Se-ssd: Self-ensembling single-stage object detector from point cloud," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 494–14 503.
[64] Q. Xu, Y. Zhong, and U. Neumann, "Behind the curtain: Learning occluded shapes for 3d object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, 2022, pp. 2893–2901.
[65] Y. Zhang, Q. Zhang, Z. Zhu, J. Hou, and Y. Yuan, "Glenet: Boosting 3d object detectors with generative label uncertainty estimation," *arXiv preprint arXiv:2207.02466*, 2022.
[66] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, vol. 29, 2016.
[67] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
[68] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.
[69] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[70] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[71] Y. Li, A. W. Yu, T. Meng, B. Caine, J. Ngiam, D. Peng, J. Shen, Y. Lu, D. Zhou, Q. V. Le, A. Yuille, and M. Tan, "Deepfusion: Lidar-camera deep fusion for multi-modal 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 17 182–17 191.

[72] S. Chen, Y. Wang, L. Huang, R. Ge, Y. Hu, Z. Ding, and J. Liao, "2nd place solution for waymo open dataset challenge–2d object detection," *arXiv preprint arXiv:2006.15507*, 2020.

[73] Y. Zhang, X. Song, B. Bai, T. Xing, C. Liu, X. Gao, Z. Wang, Y. Wen, H. Liao, G. Zhang *et al.*, "2nd place solution for waymo open dataset challenge–real-time 2d object detection," *arXiv preprint arXiv:2106.08713*, 2021.

[74] X. Du, W.-C. Hung, and T.-Y. Lin, "Optimizing anchor-based detectors for autonomous driving scenes," *arXiv preprint arXiv:2208.06062*, 2022.

[75] Z. Huang, Z. Chen, Q. Li, H. Zhang, and N. Wang, "1st place solutions of waymo open dataset challenge 2020–2d object detection track," *arXiv preprint arXiv:2008.01365*, 2020.

[76] C. Jiang, H. Xu, W. Zhang, X. Liang, and Z. Li, "Sp-nas: Serial-to-parallel backbone search for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 863–11 872.

[77] Z. Chen and X. Huang, "Pedestrian detection for autonomous vehicle using multi-spectral cameras," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 211–219, 2019.

[78] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Computer Vision–ACCV 2010: 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers, Part IV 10*. Springer, 2011, pp. 501–512.

[79] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *2010 IEEE Computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 1030–1037.

[80] B. Jun, I. Choi, and D. Kim, "Local transform features and hybridization for accurate face and human detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1423–1436, 2012.

[81] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 4, pp. 743–761, 2011.

[82] S. Zhang, R. Benenson, and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in *CVPR*, 2017.

[83] Y. Pang, J. Cao, Y. Li, J. Xie, H. Sun, and J. Gong, "Tju-dhd: A diverse high-resolution dataset for object detection," *IEEE Transactions on Image Processing*, 2020.

[84] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "Llvip: A visible-infrared paired dataset for low-light vision," *arXiv preprint arXiv:2108.10831*, 2021.

[85] A. H. Khan, M. Munir, L. van Elst, and A. Dengel, "F2dnet: Fast focal detection network for pedestrian detection," *arXiv preprint arXiv:2203.02331*, 2022.

[86] I. Hasan, S. Liao, J. Li, S. U. Akram, and L. Shao, "Generalizable pedestrian detection: The elephant in the room," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 328–11 337.

[87] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, and J. Sun, "Crowdhuman: A benchmark for detecting human in a crowd," *arXiv preprint arXiv:1805.00123*, 2018.

[88] X. Wang, T. Xiao, Y. Jiang, S. Shao, J. Sun, and C. Shen, "Repulsion loss: Detecting pedestrians in a crowd," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7774–7783.

[89] S. Zhang, R. Benenson, and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3213–3221.

[90] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster r-cnn doing well for pedestrian detection?" in *European conference on computer vision*. Springer, 2016, pp. 443–457.

[91] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast r-cnn for pedestrian detection," *IEEE transactions on Multimedia*, vol. 20, no. 4, pp. 985–996, 2017.

[92] Z. Cai, M. Saberian, and N. Vasconcelos, "Learning complexity-aware cascades for deep pedestrian detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3361–3369.

[93] S. Zhang, R. Benenson, B. Schiele *et al.*, "Filtered channel features for pedestrian detection." in *CVPR*, vol. 1, no. 2, 2015, p. 4.

[94] Y. Tian, P. Luo, X. Wang, and X. Tang, "Pedestrian detection aided by deep learning semantic tasks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5079–5087.

[95] J. Hosang, M. Omran, R. Benenson, and B. Schiele, "Taking a deeper look at pedestrians," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4073–4082.

[96] W. Nam, P. Dollár, and J. H. Han, "Local decorrelation for improved pedestrian detection," *Advances in neural information processing systems*, vol. 27, 2014.

[97] W. Wang, "Adapted center and scale prediction: more stable and more accurate," *arXiv preprint arXiv:2002.09053*, 2020.

[98] P. Zhou, C. Zhou, P. Peng, J. Du, X. Sun, X. Guo, and F. Huang, "Noh-nms: Improving pedestrian detection by nearby objects hallucination," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1967–1975.

[99] W. Liu, I. Hasan, and S. Liao, "Center and scale prediction: Anchor-free approach for pedestrian and face detection," *Pattern Recognition*, p. 109071, 2022.

[100] W. Liu, S. Liao, W. Hu, X. Liang, and X. Chen, "Learning efficient single-stage pedestrian detectors by asymptotic localization fitting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 618–634.

[101] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Occlusion-aware r-cnn: detecting pedestrians in a crowd," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 637–653.

[102] T. Song, L. Sun, D. Xie, H. Sun, and S. Pu, "Small-scale pedestrian detection based on somatic topology localization and temporal feature aggregation," *arXiv preprint arXiv:1807.01438*, 2018.

[103] J. Wan, J. Deng, X. Qiu, and F. Zhou, "Body-face joint detection via embedding and head hook," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2959–2968.

[104] Z. Lin, W. Pei, F. Chen, D. Zhang, and G. Lu, "Pedestrian detection by exemplar-guided contrastive learning," *IEEE transactions on image processing*, 2022.

[105] X. Chu, A. Zheng, X. Zhang, and J. Sun, "Detection in crowded scenes: One proposal, multiple predictions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 214–12 223.

[106] T.-Y. Lin, P. Dollár, 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*, 2017, pp. 2117–2125.

[107] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9627–9636.

[108] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[109] F. Qingyun, H. Dapeng, and W. Zhaokui, "Cross-modality fusion transformer for multispectral object detection," *arXiv preprint arXiv:2111.00273*, 2021.

[110] J. Tang, S. Li, and P. Liu, "A review of lane detection methods based on deep learning," *Pattern Recognition*, vol. 111, p. 107623, 2021.

[111] G. Kaur and D. Kumar, "Lane detection techniques: A review," *International Journal of Computer Applications*, vol. 112, no. 10, 2015.

[112] S. Waykole, N. Shiwakoti, and P. Stasinopoulos, "Review on lane detection and tracking algorithms of advanced driver assistance system," *Sustainability*, vol. 13, no. 20, p. 11417, 2021.

[113] H. Zhou and H. Wang, "Vision-based lane detection and tracking for driver assistance systems: A survey," in *2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. IEEE, 2017, pp. 660–665.

[114] D. Vajak, M. Vranješ, R. Grbić, and D. Vranješ, "Recent advances in vision-based lane detection solutions for automotive applications," in *2019 International Symposium ELMAR*. IEEE, 2019, pp. 45–50.

[115] P. L. X. W. Xingang Pan, Jianping Shi and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in *AAAI Conference on Artificial Intelligence (AAAI)*, February 2018.

[116] K. Behrendt and R. Soussan, "Unsupervised labeled lane markers using maps," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[117] X. C. W. Z. X. L. Z. L. Hang Xu, Shaoju Wang, "Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending," in *ECCV*, 2020.

[118] T. Yuan, F. Alasiri, Y. Zhang, and P. A. Ioannou, "Evaluation of integrated variable speed limit and lane change control for highway traffic flow," *IFAC-PapersOnLine*, vol. 54, no. 2, pp. 107–113, 2021.

[119] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1812–1823, 2016.

[120] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," in *European Conference on Computer Vision*. Springer, 2020, pp. 276–291.

[121] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning lightweight lane detection cnns by self attention distillation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1013–1021.

[122] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[123] Y. Ko, Y. Lee, S. Azam, F. Munir, M. Jeon, and W. Pedrycz, "Key points estimation and point instance segmentation approach for lane detection," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[124] H. Xu, S. Wang, X. Cai, W. Zhang, X. Liang, and Z. Li, "Curvelanenas: Unifying lane-sensitive architecture search and adaptive point blending," in *European Conference on Computer Vision*. Springer, 2020, pp. 689–704.

[125] T. Zheng, H. Fang, Y. Zhang, W. Tang, Z. Yang, H. Liu, and D. Cai, "Resa: Recurrent feature-shift aggregator for lane detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, 2021, pp. 3547–3554.

[126] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Keep your eyes on the lane: Real-time attention-guided lane detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 294–302.

[127] J. Su, C. Chen, K. Zhang, J. Luo, X. Wei, and X. Wei, "Structure guided lane detection," *arXiv preprint arXiv:2105.05403*, 2021.

[128] L. Liu, X. Chen, S. Zhu, and P. Tan, "Condlanenet: a top-to-down lane detection framework based on conditional convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3773–3782.

[129] 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.

[130] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2403–2412.

[131] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Polylanenet: Lane estimation via deep polynomial regression," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 6150–6156.

[132] F. Pizzati, M. Allodi, A. Barrera, and F. García, "Lane detection and classification using cascaded cnns," in *International Conference on Computer Aided Systems Theory*. Springer, 2019, pp. 95–103.

[133] Z. Qu, H. Jin, Y. Zhou, Z. Yang, and W. Zhang, "Focus on local: Detecting lane marker from bottom up via key point," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 122–14 130.

[134] Y. Dong, S. Patil, B. van Arem, and H. Farah, "A hybrid spatial–temporal deep learning architecture for lane detection," *Computer-Aided Civil and Infrastructure Engineering*, 2022.

[135] Z. Feng, S. Guo, X. Tan, K. Xu, M. Wang, and L. Ma, "Rethinking efficient lane detection via curve modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 062–17 070.

[136] H. Abualsaud, S. Liu, D. B. Lu, K. Situ, A. Rangesh, and M. M. Trivedi, "Laneaf: Robust multi-lane detection with affinity fields," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7477–7484, 2021.

[137] Z. Chen, Q. Liu, and C. Lian, "Pointlanenet: Efficient end-to-end cnns for accurate real-time lane detection," in *2019 IEEE intelligent vehicles symposium (IV)*. IEEE, 2019, pp. 2563–2568.

[138] R. Simhambhatla, K. Okiah, S. Kuchkula, and R. Slater, "Self-driving cars: Evaluation of deep learning techniques for object detection in different driving conditions," *SMU Data Science Review*, vol. 2, no. 1, p. 23, 2019.

[139] M. Siam, M. Gamal, M. Abdel-Razek, S. Yogamani, M. Jagersand, and H. Zhang, "A comparative study of real-time semantic segmentation for autonomous driving," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 587–597.

[140] S. Mohanapriya *et al.*, "Instance segmentation for autonomous vehicle," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 9, pp. 565–570, 2021.

[141] O. Elharrouss, S. Al-Maadeed, N. Subramanian, N. Ottakath, N. Almaadeed, and Y. Himeur, "Panoptic segmentation: A review," *arXiv preprint arXiv:2111.10250*, 2021.

[142] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[143] M. Cordts, M. Omran, S. Ramos, T. Scharwächter, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset," in *CVPR Workshop on the Future of Datasets in Vision*, vol. 2. sn, 2015.

[144] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4990–4999.

[145] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[146] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The apolloscape open dataset for autonomous driving and its application," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 10, pp. 2702–2719, 2019.

[147] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[148] L.-C. Chen, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[149] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3194–3203.

[150] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *European conference on computer vision*. Springer, 2016, pp. 519–534.

[151] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1925–1934.

[152] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[153] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[154] A. Valada, R. Mohan, and W. Burgard, "Self-supervised model adaptation for multimodal semantic segmentation," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1239–1285, 2020.

[155] J. Yuan, Z. Deng, S. Wang, and Z. Luo, "Multi receptive field network for semantic segmentation," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2020, pp. 1883–1892.

[156] X. Li, L. Zhang, A. You, M. Yang, K. Yang, and Y. Tong, "Global aggregation then local distribution in fully convolutional networks," *arXiv preprint arXiv:1909.07229*, 2019.

[157] Z. Chen, Y. Duan, W. Wang, J. He, T. Lu, J. Dai, and Y. Qiao, "Vision transformer adapter for dense predictions," *arXiv preprint arXiv:2205.08534*, 2022.

[158] D. De Geus, P. Meletis, and G. Dubbelman, "Panoptic segmentation with a joint semantic and instance segmentation network," *arXiv preprint arXiv:1809.02110*, 2018.

[159] K. Sofiiuk, O. Barinova, and A. Konushin, "Adaptis: Adaptive instance selection network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7355–7363.

[160] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 475–12 485.

[161] R. Mohan and A. Valada, "Efficientps: Efficient panoptic segmenta-

tion," *International Journal of Computer Vision*, vol. 129, no. 5, pp. 1551–1579, 2021.
[162] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation," in *European Conference on Computer Vision*. Springer, 2020, pp. 108–126.
[163] L.-C. Chen, H. Wang, and S. Qiao, "Scaling wide residual networks for panoptic segmentation," *arXiv preprint arXiv:2011.11675*, 2020.
[164] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.
[165] Z. Wu, C. Shen, and A. Van Den Hengel, "Wider or deeper: Revisiting the resnet model for visual recognition," *Pattern Recognition*, vol. 90, pp. 119–133, 2019.
[166] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
[167] Y. Lee and J. Park, "Centermask: Real-time anchor-free instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13 906–13 915.
[168] S. Qiao, L.-C. Chen, and A. Yuille, "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10 213–10 224.
[169] S. Zagoruyko, A. Lerer, T.-Y. Lin, P. O. Pinheiro, S. Gross, S. Chintala, and P. Dollár, "A multipath network for object detection," *arXiv preprint arXiv:1604.02135*, 2016.
[170] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2359–2367.
[171] L.-C. Chen, A. Hermans, G. Papandreou, F. Schroff, P. Wang, and H. Adam, "Masklab: Instance segmentation by refining object detection with semantic and direction features," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4013–4022.
[172] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.
[173] Y. Liu, Y. Wang, S. Wang, T. Liang, Q. Zhao, Z. Tang, and H. Ling, "Cbnet: A novel composite backbone network architecture for object detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 653–11 660.
[174] X. Du, T.-Y. Lin, P. Jin, G. Ghiasi, M. Tan, Y. Cui, Q. V. Le, and X. Song, "Spinenet: Learning scale-permuted backbone for recognition and localization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 592–11 601.
[175] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 012–10 022.
[176] M. Xu, Z. Zhang, H. Hu, J. Wang, L. Wang, F. Wei, X. Bai, and Z. Liu, "End-to-end semi-supervised object detection with soft teacher," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3060–3069.
[177] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong *et al.*, "Swin transformer v2: Scaling up capacity and resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 009–12 019.
[178] Y. Wei, H. Hu, Z. Xie, Z. Zhang, Y. Cao, J. Bao, D. Chen, and B. Guo, "Contrastive learning rivals masked image modeling in fine-tuning via feature distillation," *arXiv preprint arXiv:2205.14141*, 2022.
[179] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, "Road scene segmentation from a single image," in *European Conference on Computer Vision*. Springer, 2012, pp. 376–389.
[180] G. Ros and J. M. Alvarez, "Unsupervised image transformation for outdoor semantic labelling," in *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2015, pp. 537–542.
[181] R. Zhang, S. A. Candra, K. Vetter, and A. Zakhor, "Sensor fusion for semantic segmentation of urban scenes," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 1850–1857.
[182] G. Ros, S. Ramos, M. Granados, A. Bakhtiary, D. Vazquez, and A. M. Lopez, "Vision-based offline-online perception paradigm for autonomous driving," in *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2015, pp. 231–238.
[183] S. Kong and C. Fowlkes, "Pixel-wise attentional gating for parsimonious pixel labeling," *arXiv preprint arXiv:1805.01556*, 2018.
[184] G. Yang, H. Zhao, J. Shi, Z. Deng, and J. Jia, "Segstereo: Exploiting semantic information for disparity estimation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 636–651.
[185] P. Meletis and G. Dubbelman, "Training of convolutional networks on multiple heterogeneous datasets for street scene semantic segmentation," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1045–1050.
[186] S. R. Bulo, L. Porzi, and P. Kontschieder, "In-place activated batchnorm for memory-optimized training of dnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5639–5647.
[187] Y. Zhu, K. Sapra, F. A. Reda, K. J. Shih, S. Newsam, A. Tao, and B. Catanzaro, "Improving semantic segmentation via video propagation and label relaxation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8856–8865.
[188] K. Wang, S. Ma, J. Chen, F. Ren, and J. Lu, "Approaches, challenges, and applications for deep visual odometry: Toward complicated and emerging areas," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 1, pp. 35–49, 2020.
[189] K. Karsch, C. Liu, and S. B. Kang, "Depth extraction from video using non-parametric sampling," in *European conference on computer vision*. Springer, 2012, pp. 775–788.
[190] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," *Advances in neural information processing systems*, vol. 27, 2014.
[191] M. Goldman, T. Hassner, and S. Avidan, "Learn stereo, infer mono: Siamese networks for self-supervised, monocular, depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
[192] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
[193] J. Cho, D. Min, Y. Kim, and K. Sohn, "Deep monocular depth estimation leveraging a large-scale outdoor stereo dataset," *Expert Systems with Applications*, vol. 178, p. 114877, 2021.
[194] G. Yang, X. Song, C. Huang, Z. Deng, J. Shi, and B. Zhou, "Drivingstereo: A large-scale dataset for stereo matching in autonomous driving scenarios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 899–908.
[195] V. Guizilini, R. Ambrus, S. Pillai, A. Raventos, and A. Gaidon, "3d packing for self-supervised monocular depth estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
[196] Z. Li, X. Wang, X. Liu, and J. Jiang, "Binsformer: Revisiting adaptive bins for monocular depth estimation," *arXiv preprint arXiv:2204.00987*, 2022.
[197] A. Agarwal and C. Arora, "Depthformer: Multiscale vision transformer for monocular depth estimation with local global information fusion," *arXiv preprint arXiv:2207.04535*, 2022.
[198] A. Gurram, A. F. Tuna, F. Shen, O. Urfalioglu, and A. M. López, "Monocular depth estimation through virtual-world supervision and real-world sfm self-supervision," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
[199] J. Wang, Y. Zhong, Y. Dai, S. Birchfield, K. Zhang, N. Smolyanskiy, and H. Li, "Deep two-view structure-from-motion revisited," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8953–8962.
[200] S. F. Bhat, I. Alhashim, and P. Wonka, "Adabins: Depth estimation using adaptive bins," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4009–4018.
[201] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep ordinal regression network for monocular depth estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2002–2011.
[202] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 179–12 188.
[203] A. Masoumian, H. A. Rashwan, S. Abdulwahab, J. Cristiano, and D. Puig, "Gcndepth: Self-supervised monocular depth estimation based on graph convolutional network," *arXiv preprint arXiv:2112.06782*, 2021.
[204] J.-W. Bian, H. Zhan, N. Wang, Z. Li, L. Zhang, C. Shen, M.-M. Cheng, and I. Reid, "Unsupervised scale-consistent depth learning from video," *International Journal of Computer Vision*, vol. 129, no. 9, pp. 2548–2564, 2021.
[205] F. Xue, G. Zhuo, Z. Huang, W. Fu, Z. Wu, and M. H. Ang, "Toward hierarchical self-supervised monocular absolute depth estimation for autonomous driving applications," in *2020 IEEE/RSJ International*

Conference on Intelligent Robots and Systems (IROS). IEEE, 2020, pp. 2330–2337.

[206] N. B. Romdhane, M. Hammami, and H. Ben-Abdallah, "A generic obstacle detection method for collision avoidance," in *2011 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2011, pp. 491–496.

[207] K. Chu, M. Lee, and M. Sunwoo, "Local path planning for off-road autonomous driving with avoidance of static obstacles," *IEEE transactions on intelligent transportation systems*, vol. 13, no. 4, pp. 1599–1616, 2012.

[208] H. Laghmara, M.-T. Boudali, T. Laurain, J. Ledy, R. Orjuela, J.-P. Lauffenburger, and M. Basset, "Obstacle avoidance, path planning and control for autonomous vehicles," in *2019 IEEE intelligent vehicles symposium (IV)*. IEEE, 2019, pp. 529–534.

[209] A. De La Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE transactions on industrial electronics*, vol. 44, no. 6, pp. 848–859, 1997.

[210] D. Tabernik and D. Skočaj, "Deep learning for large-scale traffic-sign detection and recognition," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 4, pp. 1427–1440, 2019.

[211] A. Møgelmose, D. Liu, and M. M. Trivedi, "Detection of us traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3116–3125, 2015.

[212] M. G. Lay, "Design of traffic signs," *The human factors of transport signs*, pp. 25–48, 2004.

[213] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, "Traffic sign detection and analysis: Recent studies and emerging trends," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2012, pp. 1310–1314.

[214] S. Bouaafia, S. Messaoud, A. Maraoui, A. C. Ammari, L. Khriji, and M. Machhout, "Deep pre-trained models for computer vision applications: traffic sign recognition," in *2021 18th International Multi-Conference on Systems, Signals & Devices (SSD)*. IEEE, 2021, pp. 23–28.

[215] B. Sanyal, R. K. Mohapatra, and R. Dash, "Traffic sign recognition: a survey," in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2020, pp. 1–6.

[216] S. B. Wali, M. A. Abdullah, M. A. Hannan, A. Hussain, S. A. Samad, P. J. Ker, and M. B. Mansor, "Vision-based traffic sign detection and recognition systems: Current trends and challenges," *Sensors*, vol. 19, no. 9, p. 2093, 2019.

[217] S. B. Wali, M. A. Hannan, A. Hussain, and S. A. Samad, "Comparative survey on traffic sign detection and recognition: a review," *Przeglad Elektrotechniczny*, vol. 1, no. 12, pp. 40–44, 2015.

[218] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[219] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *International Joint Conference on Neural Networks*, no. 1288, 2013.

[220] K. Behrendt and L. Novak, "A deep learning approach to traffic lights: Detection, tracking, and classification," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE.

[221] F. Larsson and M. Felsberg, "Using fourier descriptors and spatial models for traffic sign recognition," in *Scandinavian conference on image analysis*. Springer, 2011, pp. 238–249.

[222] C. G. Serna and Y. Ruichek, "Classification of traffic signs: The european dataset," *IEEE Access*, vol. 6, pp. 78 136–78 148, 2018.

[223] A. Pon, O. Adrienko, A. Harakeh, and S. L. Waslander, "A hierarchical deep architecture and mini-batch selection method for joint traffic sign and light detection," in *2018 15th Conference on Computer and Robot Vision (CRV)*. IEEE, 2018, pp. 102–109.

[224] J. Chen, K. Jia, W. Chen, Z. Lv, and R. Zhang, "A real-time and high-precision method for small traffic-signs recognition," *Neural Computing and Applications*, vol. 34, no. 3, pp. 2233–2245, 2022.

[225] P. Goyal, Q. Duval, I. Seessel, M. Caron, M. Singh, I. Misra, L. Sagun, A. Joulin, and P. Bojanowski, "Vision models are more robust and fair when pretrained on uncurated images without supervision," *arXiv preprint arXiv:2202.08360*, 2022.

[226] A. Wong, M. J. Shafiee, and M. S. Jules, "Micronnet: a highly compact deep convolutional neural network architecture for real-time embedded traffic sign classification," *IEEE Access*, vol. 6, pp. 59 803–59 810, 2018.

[227] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3642–3649.

[228] H. Zhang, Z. Cao, Z. Yan, and C. Zhang, "Sill-net: Feature augmentation with separated illumination representation," *arXiv preprint arXiv:2102.03539*, 2021.

[229] Á. Arcos-García, J. A. Alvarez-Garcia, and L. M. Soria-Morillo, "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods," *Neural Networks*, vol. 99, pp. 158–165, 2018.

[230] J. Sujatha *et al.*, "Computer vision based novel steering angle calculation for autonomous vehicles," in *2018 Second IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2018, pp. 143–146.

[231] D.-H. Choi, S.-Y. Oh, and K.-I. Kim, "Enhancing reliability of a vehicle steering algorithm by combining computer vision and neural vision," in *Proceedings of ICNN'95-International Conference on Neural Networks*, vol. 5. IEEE, 1995, pp. 2703–2708.

[232] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[233] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[234] T.-D. Phan, T.-S. Nguyen, M.-T. Duong, M.-H. Le *et al.*, "Steering angle estimation for self-driving car based on enhanced semantic segmentation," in *2021 International Conference on System Science and Engineering (ICSSE)*. IEEE, 2021, pp. 32–37.

[235] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl, "Learning by cheating," in *Conference on Robot Learning*. PMLR, 2020, pp. 66–75.

[236] M. Toromanoff, E. Wirbel, and F. Moutarde, "End-to-end model-free reinforcement learning for urban driving using implicit affordances," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7153–7162.

[237] D. Chen, V. Koltun, and P. Krähenbühl, "Learning to drive from a world on rails," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 590–15 599.

[238] R. Chekroun, M. Toromanoff, S. Hornauer, and F. Moutarde, "Gri: General reinforced imitation and its application to vision-based autonomous driving," *arXiv preprint arXiv:2111.08575*, 2021.

[239] K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz, and A. Geiger, "Transfuser: Imitation with transformer-based sensor fusion for autonomous driving," *arXiv preprint arXiv:2205.15997*, 2022.

[240] D. Chen and P. Krähenbühl, "Learning from all vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 222–17 231.

[241] P. Wu, X. Jia, L. Chen, J. Yan, H. Li, and Y. Qiao, "Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline," *arXiv preprint arXiv:2206.08129*, 2022.

[242] H. Shao, L. Wang, R. Chen, H. Li, and Y. Liu, "Safety-enhanced autonomous driving using interpretable sensor fusion transformer," *arXiv preprint arXiv:2207.14024*, 2022.

[243] H. Wan, L. Gao, M. Su, Q. You, H. Qu, and Q. Sun, "A novel neural network model for traffic sign detection and recognition under extreme conditions," *Journal of Sensors*, vol. 2021, pp. 1–16, 2021.

[244] T. Sharma, B. Debaque, N. Duclos, A. Chehri, B. Kinder, and P. Fortier, "Deep learning-based object detection and scene perception under bad weather conditions," *Electronics*, vol. 11, no. 4, p. 563, 2022.

[245] C.-Y. Fang, S.-W. Chen, and C.-S. Fuh, "Road-sign detection and tracking," *IEEE transactions on vehicular technology*, vol. 52, no. 5, pp. 1329–1341, 2003.

[246] N. Yaghoobi Ershadi, J. M. Menéndez *et al.*, "Vehicle tracking and counting system in dusty weather with vibrating camera conditions," *Journal of Sensors*, vol. 2017, 2017.

[247] M. Bijelic, T. Gruber, and W. Ritter, "Benchmarking image sensors under adverse weather conditions for autonomous driving," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1773–1779.

[248] J. Luo, G. Wang, G. Li, and G. Pesce, "Transport infrastructure connectivity and conflict resolution: a machine learning analysis," *Neural Computing and Applications*, vol. 34, no. 9, pp. 6585–6601, 2022.

[249] L. Sun, J. Liang, C. Zhang, D. Wu, and Y. Zhang, "Meta-transfer metric learning for time series classification in 6g-supported intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[250] P. Lin, "Why ethics matters for autonomous cars," *Autonomous driving: Technical, legal and social aspects*, pp. 69–85, 2016.

[251] X. Sun, J. Li, P. Tang, S. Zhou, X. Peng, H. N. Li, and Q. Wang, "Exploring personalised autonomous vehicles to influence user trust," *Cognitive Computation*, vol. 12, pp. 1170–1186, 2020.