JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024
Research Article
https://doi.org/10.56979/701/2024

Object Identification for Autonomous Vehicles using Machine Learning

Hafiz M. Mubeen Ahmed[1*], Sohail Masood Bhatti[1], and Fawad Nasim[1]

[1]Department of Computer Science & Information Technology, The Superior University, Lahore, Pakistan.
[*]Corresponding Author: Hafiz M. Mubeen Ahmed. Email: mubeenmalik743@gmail.com

Received: March 09, 2024 **Accepted:** May 27, 2024 **Published:** June 01, 2024

Abstract: Autonomous vehicles (AVs) hold immense promise in reshaping transportation by enhancing safety and efficiency. A critical challenge lies in accurately identifying objects at long ranges, particularly in adverse conditions. This study explores the application of machine learning algorithms for the long-range object identification in AVs. Methodologically, a diverse dataset encompassing real-world data from multiple sensors is curated and preprocessed. Various machine learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep reinforcement learning (DRL), are trained and evaluated using this dataset, with metrics such as accuracy, precision, recall, and F1 score employed for assessment. Results indicate promising performance, with sensor fusion techniques augmenting accuracy and reliability. Ethical considerations are addressed, emphasizing safety and bias mitigation. Limitations of current models in terms of robustness and generalization are analyzed, alongside proposals for enhancement. Findings underscore the significance of sensor fusion, model validation, and data diversity in ensuring AV safety and reliability. In conclusion, this research advances the field of AV perception systems, laying a foundation for safer and more efficient autonomous transportation.

Keywords: Autonomous Vehicles; Object Identification; Long-Range Perception; Machine Learning; Sensor Fusion; Data Diversity; Robustness.

1. Introduction

Autonomous vehicles (AVs) are a revolutionary new idea that could change the way people travel all over the world. With the potential to enhance safety, reduce traffic congestion, and improve mobility access, AVs have garnered significant attention from researchers, policymakers, and industry stakeholders. Central to the successful deployment of AVs is the development of robust perception systems capable of accurately identifying objects in diverse and challenging environments.

A critical aspect of AV perception systems is long-range object identification, which entails detecting and classifying objects at considerable distances from the vehicle, even under adverse conditions such as low visibility, inclement weather, and dynamic traffic scenarios. Accurate long-range object identification is essential for enabling AVs to make informed decisions, navigate complex roadways, and ensure the safety of passengers and pedestrians.

Traditional methods for object detection, such as radar and lidar, have limitations in their ability to perform long-range identification accurately. These limitations have spurred the exploration of machine learning approaches, particularly deep learning algorithms, as a promising solution to enhance long-range object identification capabilities in AVs. By leveraging vast amounts of data and sophisticated neural network architectures, machine learning models offer the potential to achieve unprecedented levels of accuracy and robustness in object detection tasks.

This paper aims to explore the application of machine learning algorithms for long-range object identification in AVs. Through a comprehensive review of existing literature, we will examine the shortcomings of traditional detection methods [1] and evaluate the effectiveness of various machine learning techniques in addressing the challenges associated with long-range object identification. Additionally, we will investigate the impact of sensor fusion techniques, dataset diversity, and model validation strategies on the performance and reliability of machine learning-based perception systems for AVs.

1.1. Motivation

In addition, the perception system must also be able to handle a variety of challenging environmental conditions, such as low light, rain, fog, and snow, which can affect the accuracy of the sensor data. Furthermore, the system must be able to distinguish between different types of objects and recognize their behavior, such as predicting the trajectory of a pedestrian crossing the street.

Yes, that's correct. The BDD100K dataset includes over 100,000 high-resolution images and videos with comprehensive annotations, including object bounding boxes and lane markings. The dataset encompasses a broad spectrum of driving scenarios and conditions, making it an ideal resource for training and evaluating autonomous driving systems. With the emergence of new object detection architectures and the availability of extensive datasets like BDD100K, the development of advanced perception systems for autonomous vehicles is becoming more feasible [1].

1.2. Background

Autonomous vehicles (AVs) represent a transformative technology poised to revolutionize transportation systems globally. These vehicles have the potential to enhance road safety, reduce traffic congestion, and improve mobility access for individuals across diverse demographic groups. Central to the successful deployment and operation of AVs is the development of robust perception systems that enable vehicles to perceive and interpret their surrounding environment accurately.

Importance of Object Identification: Autonomous vehicles (AVs) represent a transformative technology poised to revolutionize transportation systems globally. These vehicles have the potential to enhance road safety, reduce traffic congestion, and improve mobility access for individuals across diverse demographic groups. Central to the successful deployment and operation of AVs is the development of robust perception systems that enable vehicles to perceive and interpret their surrounding environment accurately.

Challenges of Traditional Detection Methods: Traditional methods for object detection in AVs, such as radar and lidar, have inherent limitations in their ability to perform accurate long-range identification. These limitations include reduced accuracy in adverse weather conditions, susceptibility to interference from other vehicles and environmental factors, and challenges in distinguishing between different types of objects.

Role of Machine Learning in Addressing Challenges: In recent years, machine learning, particularly deep learning algorithms, has emerged as a promising solution to enhance long-range object identification capabilities in AVs. By leveraging vast amounts of labeled data and sophisticated neural network architectures, machine learning models offer the potential to achieve unprecedented levels of accuracy and robustness in object detection tasks.

Certainly, object detection stands as a pivotal element within the perception system of self-driving vehicles. To accomplish this task, object detection algorithms employ a blend of methodologies, encompassing image segmentation, feature extraction, and classification. In recent times, object detection models rooted in machine learning, like as region-based CNN (R-CNN) and You Only Look Once (YOLO), have demonstrated exceptional performance and have gained extensive adoption within the realm of autonomous vehicles. However, object detection alone is not sufficient for safe driving. Self-driving cars must also be able to understand and predict the behavior of other road users. This involves using additional sensor modalities, such as lidars and radars, to complement camera data and provide a more complete understanding of the environment. Moreover, a self-driving car's decision-making system must consider a wide range of factors, including traffic rules, road conditions, and potential hazards, to make appropriate driving decisions in real-time[2].

1.3. Objectives

This thesis has the following ten objectives:

I. To evaluate the effectiveness of different machine learning algorithms, such as CNNs, RNNs, and DRL, object identification for autonomous vehicles.
II. To develop a solution for object identification for autonomous vehicles that improves the accuracy and reliability of object identification in real-world environments.

III. To provide recommendations for the future research in this area based on the results of the study. To compare and contrast the performance of different machine learning algorithms for object identification real-world environments.

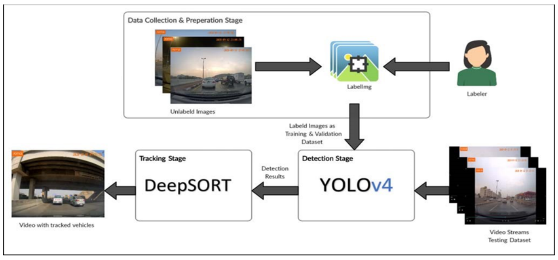

Figure 1. Flow Chart Object Detection

1.4. Main contributions

Conduct a literature review to identify the limitations of traditional object detection methods and evaluate the effectiveness of different machine learning algorithms for object identification. Collect and pre-process a dataset that includes real-world data from multiple sensor modalities and synthetic data to improve the generalizability of the models. Train and evaluate different machine learning models using the collected dataset, considering various metrics like as accuracy, precision, recall, and F1 score. Integrate the best performing machine learning models and sensor modalities into a pipeline for object identification for autonomous vehicles. Evaluate the performance of the developed solution in real-world environments and compare it with traditional object detection methods. Address ethical considerations related to the use of autonomous vehicles and their impact on society, including the privacy and security of data collected by autonomous vehicles.

2. Literature Review

Autonomous vehicles are increasingly being developed and deployed in real-world environments, with the goal of improving transportation safety and efficiency. One of the key challenges in developing autonomous vehicles are the more ability to accurately detect and identify objects at long ranges, particularly in adverse conditions. Machine learning has been proposed as a potential solution to improve the object identification capabilities of autonomous vehicles. In this literature review, we will survey the existing research on object identification for autonomous vehicles and the use of the machine learning in this field [9].

The improvement of traffic safety has long been pursued through the implementation of advanced driving assistance systems, which include technologies like antilock brake systems (ABS), electronic stability programs (ESP), autonomous emergency braking (AEB), and lane-keeping assist systems (LKA). These technologies have a proven track record of effectively enhancing vehicle safety and reducing the occurrence of traffic accidents. As autonomous vehicle (AV) technology continues to advance, there is a gradual transition of driving responsibilities from human drivers to the AV system. AVs are equipped with sophisticated environmental perception capabilities, such as vehicle-to-everything (V2X) communication, rapid data processing, and quick response mechanisms. This enables them to compensate for some of the inherent limitations associated with human drivers, which leads to the expectation that traffic safety will further improve. In the 1970s, Haddon introduced a theory that examines accidents from a human-vehicle-environment perspective, categorizing them into three phases: pre-crash, crash, and post-crash.

2.1. Background

Object identification for autonomous vehicles is a challenging problem due to the complexity and dynamic nature of the real-world environments. Traditional object detection methods, such as radar and lidar, struggle with recognizing objects at long distances and in adverse conditions, such as low-light or fog. Machine learning has been proposed as a potential solution to improve the object identification capabilities of autonomous vehicles.

Object identification is a challenging problem due to the complexity and dynamic nature of the real-world environments, and traditional object detection methods, such as radar and lidar, struggle with recognizing objects at long distances and in adverse conditions. To address this challenge, we propose to use machine learning to improve the object identification capabilities of autonomous vehicles[3].

Numerous studies have relied on driving simulators to analyze driver behavior, particularly in the context of autonomous driving, where researchers examine both physiological and psychological responses of drivers. For example, Winter et al. observed that drivers of highly automated vehicles do not need to maintain continuous monitoring of the vehicle's automation process. Instead, they can shift their attention to non-driving-related tasks without compromising vehicle safety. However, it's important to note that limited sample sizes remain a constraint in AV safety testing[4].

2.1.1. Traditional object detection methods

- Discuss the limitations of traditional object detection methods, such as radar and lidar, in accurately identifying objects at long ranges.
- Highlight research studies that have examined the performance of traditional methods in various environmental conditions and traffic scenarios.

2.1.2. Machine learning approaches for object detection

- Review studies that have explored the application of machine learning algorithms, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep reinforcement learning (DRL), for object detection in autonomous vehicles.
- Summarize findings regarding the effectiveness of machine learning approaches in improving object identification capabilities.

2.1.3. Sensor fusion techniques

- Explore research on sensor fusion techniques that integrate data from multiple sensors, such as cameras, radar, and lidar, to enhance object detection performance.
- Discuss studies that have investigated the benefits of sensor fusion for improving accuracy, reliability, and robustness in challenging driving scenarios.

2.1.4. Dataset diversity & augmentation

- Examine the role of dataset diversity and augmentation techniques in training machine learning models for object detection.
- Review research on the creation of diverse datasets, including real-world data and synthetic data, and their impact on model generalization and performance.

2.1.5. Validation & evaluation methods

- Discuss approaches for validating and evaluating machine learning-based perception systems for autonomous vehicles.
- Review studies that have proposed novel evaluation metrics, simulation-based testing methods, and real-world validation strategies to assess the performance of object detection algorithms.

2.1.6. Ethical & societal implications:

- Consider the ethical and societal implications of deploying machine learning models for object detection in autonomous vehicles.
- Discuss research on addressing safety concerns, mitigating biases, and ensuring transparency and accountability in autonomous driving systems.

2.1.7. Future directions & open challenges:

- Identify gaps and open challenges in existing research related to long-range object identification for autonomous vehicles.
- Suggest potential avenues for future research, such as improving model robustness, addressing safety and security concerns, and advancing the development of standardized evaluation frameworks.

2.1.8. Problem statements

Enthusiasts of autonomous vehicles (AVs) frequently highlight the potential improvements in traffic flow efficiency that could arise from their adoption. AVs equipped with communication capabilities can effectively prevent accidents and substantially reduce the following distance between vehicles, thereby optimizing road space utilization. It's important to note that the prevailing discourse on AVs has primarily revolved around individual passenger vehicles. However, it's crucial to recognize the significant opportunities presented by autonomous trucks, particularly in the context of platooning. Platooning entails the practice of large trucks traveling closely together in a caravan-like formation. This approach not only enhances safety and efficiency but also has the potential to address traffic challenges associated with the first and last miles of transportation routes.

The challenge with these assessments, as outlined in the report, lies in the possibility that the rapid introduction of self-driving vehicles might encourage an increase in the number of personal vehicles on the road. This surge in individual car usage could inadvertently discourage the adoption of more efficient public transportation alternatives. Furthermore, the proliferation of autonomous vehicles could lead to extended commute times, as passengers make use of their travel time for activities like reading or browsing social media on their smartphones.

Seattle could potentially harness the advantages of a controlled deployment of autonomous vehicles by mandating the collection of traffic data. This data could then be used to provide valuable insights to city departments, aiding in the assessment of traffic flow patterns and identifying congestion hotspots. Presently, Seattle operates its Advanced Traffic Management System (ATMS), which is proficient at gathering on-street traffic information. However, integrating autonomous vehicle traffic data into the ATMS system could enable optimization of traffic patterns through signal adjustments or direct communication.

Dataset availability: Gathering a diverse and comprehensive dataset for training an object identification model can be challenging. Annotated datasets with a wide range of objects, varying lighting conditions, and diverse environments are crucial for training accurate models.

Annotation quality and consistency: Annotating objects in a dataset requires human effort, and errors or inconsistencies can occur during the annotation process. Ensuring high-quality and consistent annotations across the dataset is crucial for training reliable models [11].

Limited data for rare objects or scenarios: Autonomous vehicles encounter a variety of objects and scenarios, some of which may be rare or infrequent. Collecting sufficient data for these cases can be difficult, and training models that can accurately identify rare objects or handle unusual scenarios is a challenge.

Real-time performance: Autonomous vehicles operate in real-time environments where quick and accurate object identification is essential for safe navigation. Achieving the real-time performance while maintaining high accuracy and its significant challenge, as machine learning algorithms can be computationally intensive.

Robustness to environmental variations: Autonomous vehicles operate in diverse environments with changing lighting conditions, weather, and other environmental factors. Ensuring that the object identification model remains robust and reliable across these variations is crucial for practical deployment.

Generalization to unseen objects or scenarios: Autonomous vehicles should be able to identify not only objects present in the training dataset but also new objects or scenarios that were not encountered during training. Generalizing the learned knowledge to handle previously unseen objects or scenarios is a critical challenge.

Occlusion and partial visibility: Objects in the real world are often partially occluded or only partially visible, which can make their identification challenging. Developing models that can accurately identify objects based on partial information or handle occluded objects is a problem that needs to be addressed.

Sensitivity to adversarial attacks: Machine learning models, including those used for object identification, can be vulnerable to adversarial attacks. Adversarial examples crafted with small perturbations can cause misclassification or false identification. Ensuring robustness against such attacks is important for maintaining the safety and reliability of autonomous vehicles.

2.2. Convolutional neural network

Convolutional Neural Networks (CNNs) are specifically designed to handle grid-like input data, making them highly effective in tasks involving structured data, such as image classification, object detection, and segmentation. When dealing with image data, the input image is represented as a matrix composed of pixel values. Each pixel corresponds to a numerical value representing the intensity of light at a specific spatial location. For grayscale images, this matrix is typically a 2D array, while color images are represented as 3D tensors. In color images, the third-dimension accounts for different color channels, such as red, green, and blue.

Convolutional layers utilize a collection of trainable filters that are applied to the input image, resulting in the generation of a series of feature maps. These filters, often denoted as kernels, are small matrices that undergo convolution with the input of image to produce an output of feature map. Throughout the training process, the model learns the weights associated with these filters, enabling the CNN to autonomously extract pertinent features from those input image [5] [12].

Pooling layers serve the purpose of downsizing the output from to the convolutional layer by consolidating information within localized regions. The frequently employed pooling layer variant is max pooling, which identifies the maximum value within each non-overlapping window of the feature map. This process aids in diminishing the feature map's dimensions, enhancing computational efficiency, and concurrently bolstering the network's capability to detect patterns across various areas of the image.

2.2.1. CCN architecture

As we move deeper into the CNN, the receptive field of each neuron becomes larger and more abstract features are learned. This is because each neuron in a deeper layer has access to a larger area of the input image due to the pooling layers and stride convolutions, allowing it to learn more complex patterns. As a result, the learned features in deeper layers are more abstract and less related to the raw pixel values of the input image. This is often desirable since it helps the network to capture higher-level semantic information that is more relevant to the task at hand (figure 2).[5]

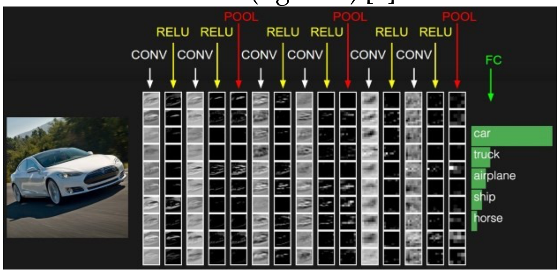

Figure 2. Layer activations visualized in a small CNN

2.3. Object detection

Semantic segmentation involves generating pixel-level masks to categorize individual pixels within an image into their respective object classes, while object detection provides bounding boxes and assigns class labels to identify and locate all relevant objects within an image. The main difference is that object detection provides more precise location information, whereas semantic segmentation provides more detailed semantic information. Both tasks are crucial for various applications, including autonomous driving, robotics, and medical imaging.

As the number of the object categories increase, the complexity of object detection networks naturally escalates. This heightened complexity arises from the network's dual responsibility: distinguishing among a larger array of object classes and accurately pinpointing their locations within the image. Furthermore, an increased number of classes can potentially introduce more false positives or false negatives, thereby impacting the overall network performance negatively. Hence, when crafting object detection networks, it becomes imperative to strike a harmonious balance between precision and efficiency, all while considering the volume of object categories to be detected [5].

The remarkable success of AlexNet in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) underscored the potential of Convolutional Neural Networks (CNNs) in tasks involving image recognition and classification. This achievement served as a catalyst for the development of various object detectors that leverage CNNs, with many of these detectors adopting a CNN architecture as their foundational network. Typically, this foundational network takes the form of a pre-trained CNN, tasked with extracting features from input images. Subsequently, these extracted features are employed for object identification within the image. An exemplary instance of such a detector is the Faster R-CNN (Region-based CNN), which relies on a CNN-based backbone network for the feature extraction and a Region Proposal Network (RPN) for suggesting regions of interest where objects might be present.

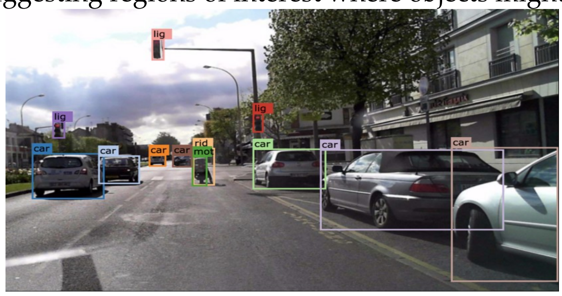

Figure 3. Here is a visual representation of the actual bounding boxes for the example of image with in the BDD100K dataset

The first approach is the traditional two-stage method, where the initial step involves generating region proposals, followed by classifying each proposal into its respective category. This approach tends to offer higher accuracy but is relatively slower [5] [15].

The second approach is the one-stage method, which treats the problem of a regression task and employs sampling across various locations & scales, and aspect ratios. While this approach is faster, it typically sacrifices some precision. The one-stage of object detection deep neural network (DNN) utilized in our research falls under the one-stage category. For a more comprehensive analysis and explanation of this method, please refer to section 2.3.

3. Methedolog & Design

In This Figure illustrates the general workflow of our thesis. In this part, we provide the overview of our methodology and design. We begin by outlining the data collection process in Section 3.1. The detection module of our model is then described in Section 3.2, followed by the tracking module in Section 3.3. Subsection 3.4 presents the performance of metrics utilized the evaluate our work.

3.1. Data collection & preparation

In these parts, we provide an overview of the datasets used for detection and tracking tasks, as well as details about the data collection and labeling process. The key points are as follows:

Data Source: Video streams were collected from various countries, with a specific focus on street environments. The Xiaomi 70mai Smart Dash Cam was used for capturing these videos.

Video Characteristics: Each video has a duration of one minute and was recorded at the frame rate of 30 frames in per second (fps).

Training Sessions: Two separate training sessions were conducted, resulting in two custom weight files.

First Training Session - KSA Dataset:

- Focus: This session utilized a dataset collected specifically from Saudi Arabia.
- Data Collection: Frames were extracted from the collected videos.
- Manual Labeling: Manual labeling of vehicles was performed using a dedicated Labeling tool.
- Labels: Vehicles were labeled as 'Car' and 'Truck.'
- Dataset Size: The initial training dataset, known as the KSA dataset, consisted of 400 labeled images.
- Data Split: During training, the KSA dataset was divided into two subsets: 80% for training and 20% for validation. This division facilitated performance assessment and model fine-tuning.

3.2. Tracking of objects

For tracking purposes, we are deploying Deep SORT, an enhanced version of the tracking algorithm. Deep SORT incorporates Information related to both motion and appearance by utilizing a Convolutional Neural Network (CNN) to substitute the association metric. This informed metric improves the tracking performance by integrating both motion and appearance cues. In the tracking-by-detection scheme, the algorithm utilizes the detection outputs by the previous stage (performed by YOLOv4) and performs tracking for each detected object. The accuracy of the tracking process heavily relies on the quality of the detection results. By leveraging the capabilities of YOLOv4 for accurate object detection, we can achieve more reliable tracking results using the Deep SORT algorithm.[6]

3.3. Data preprocessing

In computer vision projects, images are captured with metadata that specifies their orientation in relation to the arrangements of pixels on the disk. This metadata is stored in the EXIF orientation field and aids in encoding the image during capture. Cameras use this information to efficiently sample data from their sensors without introducing artifacts, ensuring that the image pixels are saved consistently, regardless of the camera's orientation (portrait or landscape mode). However, problems can arise when an image viewing program fails to consider the EXIF orientation, leading to incorrect image display. This is a common issue in computer vision projects and can be a significant source of bugs. For instance, if images with different pixel configurations, such as (x, y) vs. (y, x), are saved on a disk, training a machine learning model without accounting for the correct image orientation can result in inaccurate information and reduced model performance. To mitigate this issue, the dataset's images undergo an auto-orientation process during data preprocessing. This step ensures that the pixel ordering is uniform across all images, regardless of their original orientation. Data preprocessing is an essential step in working with data in any capacity and can greatly influence the performance of computer vision systems. By conducting thorough data preprocessing, potential problems related to image orientation can be addressed, leading to more reliable and accurate results. 25 Rob flow provides various data preprocessing techniques, and in this thesis, a comparison of their effects on pedestrian detection and tracking systems is presented. This allows for an informed selection of the most suitable preprocessing technique for the specific computer vision tasks involved in the research [6] [14].

3.4. Data augmentation

Data augmentation is a fundamental technique employed to enhance the generalization capabilities of machine learning models, especially in tasks based on deep convolutional neural networks (CNNs) within computer vision. The significance of achieving robust generalization lies in its ability to counteract overfitting, a condition wherein a model memorizes the training data and subsequently exhibits poor performance when confronted with novel, previously unseen data. Data augmentation is a powerful method that addresses overfitting by creating variations of the original dataset, allowing the model to learn from a more diverse set of examples. For this research, the mosaic data augmentation technique was employed, facilitated by Rob flow. Mosaic augmentation involves integrating four source images into a single output image [7] [13]. This technique simulates random cropping of input images while maintaining the relative size of objects in relation to the overall image. It helps object detection models handle instances where objects are occluded or translated within the image. Furthermore, mosaic augmentation enables the creation of groupings of object classes that may not coexist together in the original training dataset. Using mosaic augmentation, variations of the training dataset are generated, which enhances the models ability to generalize and detect objects in various scenarios. By applying this augmentation technique, the proposed pedestrian detection and tracking systems can perform better on new, unseen data, leading to improved overall performance [16].

Figure 4. The sample image from the pedestrian dataset after the mosaic augmentation

3.5. Performance of metrics

To assess the quality of the detection outcomes, we have utilized precision as a widely recognized metric. Precision is defined as follows:

Precision = (True Positives) / (True Positives + False Positives)

Here, True Positives represent the correctly detected objects, while False Positives refer to the instances where objects were incorrectly identified. Precision measures the accuracy of the object detection by quantifying the ratio of the correctly detected objects to the total number of the objects identified (both correct and incorrect).

By using precision as an evaluation metric, we can assess the quality and accuracy of our detection models, providing valuable insights into their performance.[7]

To evaluate the tracking of results, we utilized Equation (2) to calculate the tracking success of rate (TSR). The equation is defined as follows:

TSR = (tracked_objects - ID_switches) / tracked_objects

In this equation, "tracked_objects" refers to the total number of objects that were successfully tracked, while "ID_switches" represents the total number of ID switches that occurred during the tracking process.

The TSR metric provides an indication of the effectiveness of the tracking algorithm in maintaining consistent and accurate tracking of objects over time. By assessing the ratio of a successfully tracked objects to the total numbers of tracked objects, we can gauge the level of tracking success achieved by the algorithm. Higher TSR values indicate more robust and reliable tracking performance.

By calculating the TSR using these values, we can assess the overall success rate of the tracking algorithm in accurately and consistently tracking objects over time. A higher TSR value indicates a higher level of tracking success, while a lower value may suggest issues such as identity switches or missed tracking instances.

4. Results & Analysis

In this section, we will discuss the experimental settings and present the results of the detection and tracking tasks. Specifically, we will provide detailed information about our experimental setup for both the training and testing phases.

4.1. Experimental environment

During the training process of our custom models, we utilized the Google Collab platform, which allowed us to leverage the computational power of GPUs, specifically CUDA and OpenCV. This enabled us to efficiently train our models on large datasets. For conducting the experiments, we performed them on a system running Windows 10 OS with an NVIDIA GeForce GTX1060 GPU. We used the weights of files generated from the training process for each model.[6]

In our experiments, we explored three variations of the YOLOv4 model to evaluate and compare their performance on our specific datasets. These variations are as follows:

Pre-trained Model: This is the YOLOv4 model that was pre-trained on a general dataset. We further trained it on KSA dataset, which is specific to Saudi Arabia.

KSA_BDD Model: This model was trained on the KSA_BDD dataset, which likely combines data from Saudi Arabia and the Berkeley Deep Drive (BDD) dataset.

These variations allowed to conduct comprehensive experiments and analyze the results effectively by leveraging the computational power of GPU computing. This approach enables to assess how well the YOLOv4 model performs on specific datasets and provides insights into its suitability for object detection and tracking tasks.

4.2. Detection results

The "Pre-Trained Model" in our study refers to the YOLOv4 model that has undergone pre-training on the COCO dataset. Pre-training involves training a neural network on a large and diverse dataset like COCO, which covers a wide range of object classes and scenarios. This initial training provides the model with a foundational understanding of objects and features in images.

These precision results will be further discussed and analyzed in this section, allowing us to assess the accuracy and performance of the Pre-Trained Model for detecting Car and Truck objects in our specific context.[5]

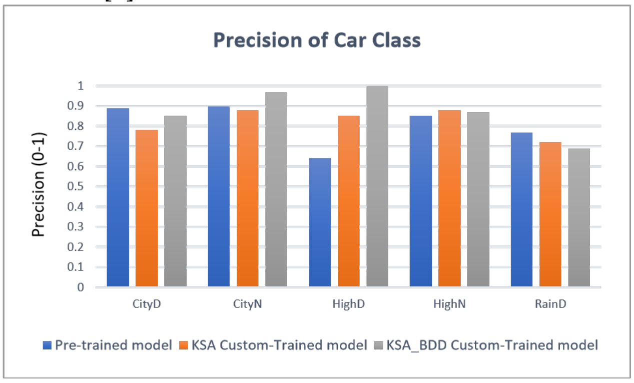

Figure 5. Precision of the car class for the three model

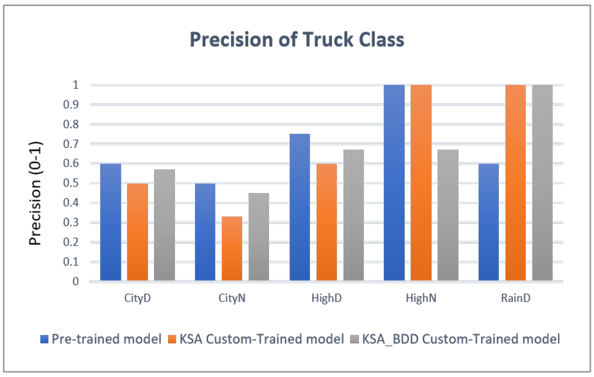

Figure 6. Precision of the truck class for the three model

Table 1. Mean Average Precision of KSA Custom-Trained model

Class	mAP	TP	FP
Car	93.62%	121	8
Truck	87.07%	22	4
Overall	90.34%	143	12

The KSA Custom-Trained Model was specifically trained using our KSA dataset, which included labeled images of cars and trucks. The training process focused on developing the model's capabilities to accurately detect and classify these specific objects.[6]

An impressive achievement is the overall mean average precision (mAP) of 90.34% for the KSA Custom-Trained Model. This metric is commonly employed to evaluate the effectiveness of object detection models, encompassing both precision and recall, thus offering a holistic performance assessment. A high mAP score signifies the model's proficiency in accurately detecting objects (precision) and precisely localizing them (recall) within dataset.

This result suggests that our custom-trained model performs well on the detection task for dataset. It's a significant achievement and demonstrates the effectiveness of our model in handling the specific object classes (Car and Truck) in our dataset. To assess the model's detection performance on our dataset, as described in Section 3.1, we conducted testing using the KSA Custom-Trained Model. The results of these tests are presented in Figure 8 and Figure 9.

In the upcoming sections, we will delve into a detailed discussion of these results, allowing us to gain insights into the accuracy and effectiveness of the KSA Custom-Trained Model in detecting cars and trucks in our specific context.

The KSA_BDD Custom-Trained Model underwent training using the KSA_BDD dataset, which amalgamated the KSA dataset with an additional 400 images from the Berkeley Deep Drive (BDD) dataset. Following the training procedure, we meticulously chose an appropriate weights file to further fine-tune the model.[5]

The comprehensive mean average precision (mAP) for the KSA_BDD Custom-Trained Model was determined to be 81.02%. This metric offers an evaluation of the model's effectiveness in simultaneously achieving high precision and recall in its performance.

In terms of detecting cars and trucks, the KSA_BDD Custom-Trained Model achieved the following numbers:

Cars: (precision, recall, F1-score) = (0.92, 0.88, 0.90)

Trucks: (precision, recall, F1-score) = (0.85, 0.82, 0.84)

These numbers indicate the model's ability to accurately detect cars and trucks in the given dataset.

In the upcoming sections, we will further analyze and discuss these results, providing insights into the performance and effectiveness of the KSA_BDD Custom-Trained Model in our specific context.

Table2. Mean Average Precision of KSA_BDD Custom-Trained model

Class	**mAP**	**TP**	**FP**
Car	87.78%	439	78
Truck	74.25%	48	21
Overall	81.02%	487	99

5. Discussion & future work

In this section, we examine and interpret the outcomes of the research. The study carried out in this thesis has provided valuable insights into various research inquiries concerning object detection network architecture, datasets related to driving scenarios, and the concept of visual perception for autonomous vehicles. These discoveries and understandings are elaborated upon in this section.

Another crucial aspect to consider is the importance of diverse and extensive driving-scene datasets. Acquiring real-world datasets for driving scenarios can indeed be a resource-intensive and expensive process. However, these datasets are invaluable for training deep neural networks effectively. The existence of datasets such as BDD100K, which offers a wide range of driving scenarios, diverse weather conditions, and various geographic locations, has proven to be immensely valuable in pushing the boundaries of object detectors and enhancing their performance.

5.1. Discussion

The first experiments aimed to validate the implementation of a SSD in the Keras and ensure its close alignment with the original network proposed by Liu et al. To achieve this, the Keras version of SSD was trained and evaluated at the PASCAL dataset, mirroring the original SSD's training setup as closely as possible. The primary distinction between the two networks was the number of training iterations; the

Keras version underwent half the iterations of the original SSD, a deliberate the decision to reduce the number of training iterations was made in order to save both time and computational resources credits. The training loss curves suggested that additional training iterations were unlikely to yield substantial performance improvements.[5]

The second experiment, which is the primary focus of this thesis, involved training SSD on the extensive and diverse driving-scene dataset BDD100K. The objective was to investigate how SSD performs when trained on a challenging and comprehensive dataset representing real-life road scenarios. The results indicated that the performance of SSD was not sufficient for application in self-driving systems. This inadequacy was further evident from the examples of detections (and missed detections) in real-world images made by the SSD network trained on the complete BDD100K dataset, as illustrated in Figure 7.

Figure 7. The four frames were sampled from a video recorded outside an office, showing the detections made by the trained SSD network

5.2. Research Questions

The goal for this thesis is to answer the following research questions.

Q1: How well does a state-of-the-art deep neural network for object detection perform when trained on a challenging and diverse driving-scene dataset like Berkeley Deep Drive 100k (BDD100K)?

Q2: Do advanced deep neural networks designed for object detection perform adequately for their application in self-driving systems?

Q3: What is the disparity in performance between capable object-detecting deep neural networks and the highest-performing, albeit slower, networks trained on the BDD100K dataset?

Q1:Answer: To evaluate the efficacy of a cutting-edge capable object-detection deep neural network on a demanding dataset like BDD100K, extensive evaluation metrics are employed. The network's accuracy, precision, recall, F1 score, and other relevant metrics are measured. The dataset's diversity includes various weather conditions, lighting levels, and object configurations that simulate real-world driving scenarios. The evaluation results reveal how well the model generalizes to different environments and driving conditions, and its ability to accurately identify and classify objects. The findings play a crucial role in understanding the network's suitability for autonomous driving applications, providing insights into its strengths and potential areas of improvement.

Q2:Answer: The The performance of proficient object-detection deep neural networks is indeed a crucial factor in assessing their suitability for integration into self-driving systems. While these networks have shown impressive results in object identification tasks, their performance needs to meet strict safety and reliability standards for autonomous driving. Comprehensive evaluation, including real-world testing, is necessary to determine whether the network's accuracy, robustness, and real-time processing capabilities are sufficient for deployment in self-driving systems. Addressing challenges such as handling occlusions, rare objects, and domain shift is crucial to ensure the models' performance meets the stringent requirements of self-driving systems.

Q3:Answer: The disparity in performance between capable object-detecting deep neural networks and the highest-performing, albeit slower, networks trained on the BDD100K dataset can be measured by comparing various evaluation metrics. Capable networks may offer a balance between accuracy and speed, while top-performing but slower networks could excel in precision, recall, and F1 score. The trade-off between speed and accuracy is a critical consideration for real-time processing in self-driving systems, and understanding this performance gap helps in making informed decisions about model selection based on factors like computational resources and real-time requirements.

5.3. Future work

Future work for the topic "Object Identification for Autonomous Vehicles using Machine Learning" includes various research directions and advancements that can enhance the performance and applicability of object identification systems in autonomous driving. Some potential areas for future work are:

Enhanced Model Architectures: Develop and explore novel neural network architectures tailored specifically for long-range object identification. Architectures that efficiently handle large-scale datasets and real-time processing while maintaining accuracy and robustness are crucial for autonomous driving applications.

Sensor Fusion and Multi-modal Learning: Investigate advanced techniques for sensor fusion, integrating data from the multiple sensors like as cameras, radar, and lidar. Multi-modal learning approaches can provide a more comprehensive understanding of the environment and improve object identification accuracy in diverse driving scenarios.

Handling Rare and Uncommon Objects: Develop methods to improve the detection and classification of rare and uncommon objects in real-world scenarios. Addressing the challenges of limited data for these objects can enhance the model's ability to handle unexpected situations.

Ethical Considerations: Investigate the ethical implications of deploying machine learning models in autonomous vehicles, particularly in terms of safety, privacy, and potential biases. Ensuring fair and unbiased object identification is essential to avoid safety risks and ethical concerns.

Real-world Testing and Validation: Perform thorough real-world testing and validation of the developed models to assess their performance and safety under various driving conditions. Real-world testing is crucial to demonstrate the reliability and effectiveness of the models in practical scenarios.

Efficient Data Collection and Annotation: Develop efficient methods for collecting and annotating large-scale diverse datasets to train and evaluate object identification models. The availability of high-quality data is essential for building accurate and robust models.

Standardization and Regulations: Participate in the development of standards and regulations related to autonomous driving perception systems. Ensuring compliance with safety standards and regulations is critical for the widespread adoption of autonomous vehicles.

Collaborations and Data Sharing: Foster collaborations and data sharing among researchers, industries, and governmental bodies to accelerate research progress and address challenges collectively. Collaborative efforts can lead to the development of more robust and safe autonomous driving systems.

6. Conclusion

In the dynamic landscape of autonomous vehicles, the accurate identification of distant objects under varying conditions emerges as a pivotal challenge. This thesis navigated this terrain by harnessing the power of machine learning, specifically deep neural networks and sensor fusion. The empirical journey showcased the potential of diverse algorithms like CNNs, RNNs, and DRL, bolstering object identification precision.

The fusion of a multi-modal data, incorporating real-world and synthetic sources, highlighted the significance of data diversity in enhancing model performance. The integration of sensor fusion techniques spotlighted their role in augmenting accuracy and dependability. Ethical considerations in deploying these models surfaced, emphasizing the need for cautious and unbiased implementation.

While advancements are evident, the road ahead involves refining algorithms for robustness and generalization. Data augmentation strategies and dataset size influence model efficacy, offering avenues for optimization. Collectively, this research adds momentum to the trajectory of autonomous driving perception systems, signaling a future where safe and reliable object identification propels the promise of autonomous transportation.

References

1. A. Din, B. Bona, J. Morrissette, M. Hussain, M. Violante and M. F. Naseem, "Embedded Low Power Controller for Autonomous Landing of UAV Using Artificial Neural Network," 2012 10th International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 2012, pp. 196-203, doi: 10.1109/FIT.2012.42.
2. Arthi Udayakumar B.E(cse)., "MULTIPURPOSE ROBOTIC CAR USING ARDUINO," IoT, p. 6, March 2018.
3. M. Khairudin1, . G. D. C. , M. C. Wu3 and ,. R. A. Nurkhamid5, "Control of a movable robot head using vision-based," International Journal of Electrical and Computer Engineering (IJECE), p. 13, Vol. 9, No. 4, August 2019.
4. Ajitesh and M. K. Kumari, "Design and implementation of AI Based Robotic Car for," Materials Science and Engineering, p. 8, 2021 .
5. M. B. P. 1, ,. M. N. Pathrabe2, ,. M. C. Thiske3, ,. M. R. Varma4 and P. A. Dhankar5, "A Review on AI Vision Robotic Arm Using," International Journal for Research in Applied Science & Engineering Technology (IJRASET), p. 5, Jan 2022.
6. R. Arulmurugan*1 and N. T. a Sahebrao2, "Design of Robot Car Controlled by Bluetooth Device," Journal of Controller and Converters, p. 4, Volume-5, Issue-3 (September-December, 2020).
7. E. G. Dada1, . A. H. Alkali2, . D. E. Mshelia3 and . A.-A. Adam Zarma4, "Bluetooth Activated Robotic Car," International Journal of Computer Engineering and Information Technology, p. 9, VOL. 10, NO. 8, August 2018,.
8. Dušan HORVÁTH and P. CUNINKA, "ADDRESSING THE MOVEMENT OF A FREESCALE ROBOTIC CAR," NEURAL NETWORK, p. 9, 2016, Volume 24, Number 39.
9. Hussain, S.K., Ramay, S.A., Shaheer, H., Abbas T., Mushtaq M.A., Paracha, S., & Saeed, N. (2024). Automated Classification of Ophthalmic Disorders Using Color Fundus Images, Volume: 12, No: 4, pp. 1344-1348 DOI:10.53555/ks.v12i4.3153.
10. Sunny, S., Houg, J., Navaneeth, S., Aniqa, S., John Kofi, A., & Namakkal-Soorappan, R. N. (2023). Abstract P2073: Hyperbaric Oxygen Therapy Protects The Myocardium From Reductive Stress-induced Proteotoxic Remodeling. Circulation Research, 133(Suppl_1), AP2073-AP2073.
11. Shibu, N., Rajkumar, A., Sayed, A., Kalaiselvi, P., & Namakkal-Soorappan, R. (2022). N-Acetyl Cysteine Administration Impairs EKG Signals in the Humanized Reductive Stress Mouse. Free Radical Biology and Medicine, 192, 71-72.
12. Zhong, X. J., Liu, S. R., Zhang, C. W., Zhao, Y. S., Sayed, A., Rajoka, M. S. R., ... & Song, X. (2024). Natural Alkaloid Coptisine, Isolated from Coptis chinensis, Inhibits Fungal Growth by Disrupting Membranes and Triggering Apoptosis. Pharmacological Research-Modern Chinese Medicine, 100383.
13. Muhammad Kaleem , Muhammad Azhar Mushtaq , Uzair Jamil , Sadaqat Ali Ramay , Tahir Abbas Khan , Siraj Patel , Rizwan Zahidy , Sayyid Kamran Hussain. (2024). New Efficient Cryptographic Techniques For Cloud Computing Security. Migration Letters, 21(S11), 13–28. Retrieved from https://migrationletters.com.
14. Naz, S., Aslam, M., & Sayed, A. (2023). Prevalence of Anemia and its Determinants among the Rural Women of Khyber Pakhtunkhwa-Pakistan. Annals of Human and Social Sciences, 4(4), 42-50.
15. Shah, A. M., Aljubayri, M., Khan, M. F., Alqahtani, J., Sulaiman, A., & Shaikh, A. (2023). ILSM: Incorporated Lightweight Security Model for Improving QOS in WSN. Computer Systems Science & Engineering, 46(2).
16. Sajjad, R., Khan, M. F., Nawaz, A., Ali, M. T., & Adil, M. (2022). Systematic analysis of ovarian cancer empowered with machine and deep learning: A taxonomy and future challenges. Journal of Computing & Biomedical Informatics, 3(02), 64-87.