Deep Learning for Autonomous Vehicle Object Tracking and Recognition

By Dr. Xiaobo Li

Associate Professor of Electrical Engineering, Tsinghua University, China

1. Introduction

[1]Current tracking systems for autonomous vehicles are mostly reliant on object detection with detection-based tracking (DBT) (Li et al., 2016). However, to successfully track objects in an image series, these systems typically require an object detection step in the first frame, which could be challenging in cases of highly dynamic backgrounds or difficult object orientations. Additionally, they are susceptible to noise with sensor actors (e.g., pedestrians) after detection, failing to incorporate new target detections when detecting signals, and losing sight of targets (e.g., occlusions). The related task of visual multi-object tracking (V MOT) seeks to address these problems, yet it is also prone to significant issues, such as detection drifting, new object detections, as well as object occlusion and dynamic scenarios. As such, more robust algorithms that work well in different real-world dynamics and camera setups are required in the application of autonomous vehicles.[2]Autonomous vehicles rely on key perception capabilities including object detection, tracking and recognition to operate safely in the traffic ecosystem. Standard visual object detection techniques have been able to obtain high accuracy across different public benchmarks. In autonomous vehicles, the need to approach these problems under various difficulties has been addressed by researchers from multiple areas. The growing demand to understand real-world dynamic scenarios indicate that approaches towards lifelong tracking and recognition that are comparative to the best detectors are increasingly essential. Visual object recognition systems using dashcam images have been generally successful; however, external factors have been shown to affect performance, especially in regions with endemic snow. Meanwhile, Off-the-shelf commercial object recognition systems have poor real-world object tracking due to sensor dynamics. Sensor-aimed physical vision data, preconditions, and external auxiliary data supervision are rarely incorporated in the existing recognition and tracking.

1.1. Background and Motivation

Traffic accidents, fatalities, pollution, road congestion, and driver comfort are important motivators for autonomous vehicles. Verification of safety-critical systems has been an essential challenge for many decades. Uncountable hours have been spent on verification and validation processes of the autonomous vehicle. Verification of a fully autonomous vehicle is even more complicated, particularly since they will be sharing the environment in real time with other human-driven vehicles. Real world driving plays a major role in the verification and validation of the controllers for autonomous vehicles. Approaches like reinforcement or imitation learning are popular in this direction [3]. Advanced neural network architectures referred to as deep learning has found its importance in computer vision for numerous tasks such as object recognition, detection, tracking, and various applications in the autonomous vehicle domain. This paper offers a collection of open-access records of vehicle object detection and tracking in the city and highway scenarios. In summary, deep learning, particularly convolutional neural networks (CNN), has revolutionized the field of image and video analysis. Among autonomous vehicle-related published work in this area, the use of human-generated visual data is well known. NYC Taxi-Vision, KITTI, and various datasets in the wild domain act as important vehicles for training and testing object detection and tracking models [4].

AI and machine learning techniques, such as supervised, unsupervised, and semi-supervised learning, are widely used in autonomous driving for tasks like perception, object classification, detection, and localization [5]. However, deep learning, with its advanced neural network architectures, has been exploited increasingly more in recent years and is particularly popular with image and vision applications like autonomous vehicle object tracking and image recognition. Typically used sensors like high-definition (HD) cameras and Light Detection and Ranging (LiDAR) provide abundant information on both object localization and 3D spatial separation in the environment. In autonomous vehicle object tracking tasks, 2D object detection for cars and pedestrians are crucial. A typical pipeline includes 2D object detection and object association based on detection object identity (ID). The well-known RCNN architectures like Faster RCNN, and its improvement Mask RCNN, have been widely exploited for 2D object detection tasks. As for object tracking tasks, re-identification (ReID) based data association is explored, usually based on trained deep ReID models from DeepSORT paper.

1.2. Research Objectives

Objective 2: A deep-Siamese neural network addresses each highlighted task with the best trade-off between accuracy and computational speed [6].Einrasib et al., 2020. A Siamese network may be trained to highlight the existence of differences between a given template and all the items available at test-time. And additionally, its attack is relevant in different circumstances, a typical situation when working with sensor data from an unmanned aerial vehicle (UAV) systems.

Prompt detection is required to follow and recognize vehicles that cross the camera's field of view (FoV) for a reduced time. Visual Perception techniques need such a prompt and very weakly supervised inspection by their nature.

Objective 1: An efficient and cross viewpoint vehicle detection algorithm will be developed thanks to the introduction of a lightweight deep neural network architecture inflated over a generic convolutional neural network (CNN) pre-trained model as underlined in [ref: 9893f252-73ac-41d6-b316-3a6963971eed, 9893f252-73ac-41d6-b316-3a6963971eed].

Definition: The research is framed in the design and development of a non-contact vehicle traffic investigating system. Its main objective is monitoring multiple vehicles by computer vision; in particular by using a convolutional neural network (CNN) based tracking algorithm that recognize and keep track of vehicles within the camera's field of view.

Deep Learning for Autonomous Vehicle Object Tracking and Recognition 1.2. Research Objectives

2. Fundamentals of Deep Learning

Explaining how to classify and identify objects in images, I make a comparison with the standard approach using handcrafted features depending on classifiers, Support vector machines, and KNN, where the features are designed and so named by humans. Also, these standard ways face the problems that they are not able to recognize new objects and typically not always able to detect the object where the surrounding environment is complex. However, deep learning not only solves the problems of the standard approach but also it minimizes the difficulties to obtain a successful object detection algorithm [7]. For instance, this road marking allows a person to exceed the traffic rules or to avoid some important things. Thus,

road marking is an essential component for driver assistants. From the progress in the area of deep learning, road marking detectors are developed using very different CNN structures having several modifications and improvements in training and test practices. Lane detection using deep learning has been actively studied in recent years [8]. Drive Assist System (DAS) requires a camera system for road and lane detection. For eyes and a camera, a lane and road markings are the most important to know where the car is, and deep learning based lane and road mark detectors are currently being developed and tested for series production. Only when deep learning algorithms achieve a certain percentage of recognition success rates for certain miniatures, they are translated to the serious drives. In the real life, garbage clouds coming from garbage collection vehicles affect the lower level of the urban blocks. Each garbage cloud envelops the 4th and 5th levels of the blocks. The Bottom Up approach contains Motion detection, foreground partition, COB, Siamese range, blob finding, object identification and local moving coordinate system. Similarly, Top Down approach contains Direct Object detection, road detection, and Virtual moving coordinate system. Both Effective object recognition has been developed during the research study, and both a new garbage robot and a new autonomous driving garbage truck having the hybrid autonomous driving system were developed and constructed during the research study. Results of the research studies showed that our approaches will be solution of the garbage problem in the urban areas. [9].

2.1. Neural Networks

For 3D detection, we followed the ideas of 2D image detection concepts and applied them to 3D objects. Besides the bird-eye view (BEV) and front view (FV), we also used the LIDAR range and height information to construct voxel-based feature volumes. The 2D/3D detection branch was for generating a high resolution of feature maps to keep the accuracy of small or medium-sized objects. Associating with lower resolution maps, the final prediction of 3D bounding boxes was also generated. From our experiments, we see that the feature enrichment has a positive impact. In particular, the VH1+ branch was trained for 3.5improving the vehicle category. With the adapted backprojection layer, we can now do a simple and efficient mapping of the feature back to the point cloud in a tensor-friendly manner. This feature map can be acquired for every point (supporting the “partial supervision” during testing time) in parallel very efficiently. With 3D multi-modalities fully utilized, our feature maps enable even higher accuracy 3D detection by capturing more details

(Table II). We observe that by aggregating multiple cues across channels in a supervised manner and integrating it efficiently into the network graph, we make our network infrastructure very effective and efficient across modality signals. With the offset feature map from LIDAR FV, the FAST-EnglishNet tracks the moving object almost perfect even without stably detected 3D boxes.

[10] [11] [6]Sensor data and the availability of high-quality training data have become the two most important factors for the success of machine learning systems. GPS sensors are widely used to acquire the movement information of autonomous vehicles. LiDAR sensors provide rich environmental information, such as the size, shape, and distance of surrounding objects. In the little information, it is easy for autonomous vehicles to judge the correct movement strategy. And the historical data is important additional information for predicting future trajectories. Combing the environment information, movement state information, and historical information, making the system of one car obtains the environmental information of other vehicles. Also, during real-time communication with other vehicles, this information could be supplemented by other sensors to reduce uncertainty. At the same time sensor fusion technologies are expected to be widely used in autonomous driving, further study of this field of technology is needed.

2.2. Convolutional Neural Networks (CNNs)

The structure of these networks contains multiple layers using linear and non-linear functions, and therefore result in the overall successful representation of the image [7]. The architecture of CNN model consists of various layers, from convolutional layer, activation layer, pooling layer, to fully connected layer and finally to softmax layer. The Convolutional layer applies a specified number of convolution filters to the image. Because image pixels are quite correlated to themselves and quite different from the pixels of nearby pixels, this operation encodes detailed local spatial information of the image because the operations in the convolution produce feature maps which represents the functional connectivity between local patches in the input image and the neurons of the next layer. Therefore, the successive layers in the deep network encode increasing higher layer abstraction.

Convolutional Neural Networks (CNNs) are a type of neural networks that are mainly used for image recognition purposes. These networks have also gained attention in object detection and visual tracking tasks [12]. When using CNNs for object detection, instead of direct

localizing and detecting the object, the image is divided into grid cells, and these cells are used to predict probability of various classes and how these predicted probabilities are correlated with the anchor boxes. This way, object detection and tracking is performed as a classification problem over the grid cells, and also as a regression problem from the anchor boxes to the ground truth values of some transformation of the bounding boxes.

2.3. Recurrent Neural Networks (RNNs)

RNNs are a natural fit for sequential data because they model data as a series of iterations and can learn information over time. RNNs address vanishing/exploding gradient problems by introducing internal memory states that allow information to persist beyond a single time step. Both LSTMs and GRUs are examples of RNN architectures with an ability to learn dependencies over time while mitigating vanishing gradients. In an LSTM [13], the input, forget, and output gates permit control of information flow in and out of the integral unit. Instead of replacing the internal states with simple non-linearity as RNNs do, GRUs use a reset gate to forget information and an update gate to update the value in each neuron in an effort to mitigate the vanishing gradient issue by using limited interactions in time. Selective usage of such information can enhance the performance of a network in both the object detection and tracking stages of a visual recognition system.

Recurrent Neural Networks (RNNs) are a class of artificial neural networks that contain a cyclic directed graph. In a classic feed-forward neural network, the data flows in a single direction, while in RNNs, the output is used as input for the next time step [14]. RNNs have universal approximator ability, graphs of any width and depth, and the ability to achieve any real number arithmetic. Simple Recurrent Neural Network (SRNN), a type of RNN, decides the value for Hidden State function by using both an input value and a previous neuron's state. It can reduce the number of required connections by unrolling a loop into a chain of identical neurons, enhancing the training efficiency [15]. Nowadays, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), a type of so called "gated" neural network, are widely employed to solve time series problems especially in the natural language processing, where it can automatically extract the significance of wording and learn the vehicle control of an autonomous vehicle.

2.4. Deep Learning Architectures for Object Tracking and Recognition

Moreover, the AV should be capable of interpreting the urban environment around it. Weissman et al. proposed two practical tools for this purpose: one to create a database with the locations, velocities, and directions of all objects detected; the other to build a set of local hypotheses (i.e., sequence of object states and tracking confidence) from this database that are relevant to the immediate future and limited to relevant spatial areas (full-occupancy cells), and estimated how likely they are.REFERENCES According to Reference, most of the localisation projects focus on designing a system to localise a robot on a natural environment. There are numerous precision localisation systems for verification, education, as well as for environmental modeling in digital structures in indoor environments. For instance, building indoor maps helps users to keep track of where they are in the building and navigate using an augmented reality application.

In this work, we analyze domain and non-target objects that are not a focus of interest of the AV (e.g., pedestrians and cars for a vehicle that is tracking people elsewhere in the field of view) and refrain from identifying the entire video stream as a dynamic object. Finally, existing tracking algorithms that are compatible with various vision systems, including onboard platforms, distributed systems, and high-definition camera systems, are also part of the considered AV technologies [8]. Here, most of the current trackers send their outputs to a single 'oracle' that decides on the action to take based on the tracking results combined with the object detection and about the scene [16]. Hence, the goal of employing an Object/Person Tracker, i.e., to provide the vehicle's onboard AV with the flow of detected.Location associated with the domain object allows the AV to classify and interpret it, based on the sensed information and its prior knowledge about the location of domain objects in the vicinity of the AV.

3. Autonomous Vehicles: Technologies and Challenges

Deep learning has gained enormous popularity in last few years. The most prevalent types of deep learning neural network architectures are convolutional neural network (CNN), long short-term memory (LSTM), recurrent neural network (RNN), open pose model, and cycle GANs. CNN architecture has been widely used for image and video frame-based object recognition and detection tasks [5]. Significant research is also reported for object detection with 3D data representation. In the context of autonomous vehicles object recognition survivability is another important perspective that has not been much addressed in the

literature yet. Some preliminary work is reported for object recognition survivability for UAVs, which is still an open challenge for this field.

Object detection and recognition are major tasks in traffic monitoring systems. Currently, the performance of video-based traffic monitoring systems largely relies on object detection and recognition. During recent years, deep learning architectures have significantly improved the performance of traffic monitoring systems [16]. Object tracking is crucial in many application, such as surveillance systems, autonomous vehicle systems, and medical imaging. For traffic monitoring systems, tracking of on-road vehicles is crucial, for example, for tracking their movement is order to understand congestion behaviour. Additional to tracking objects, a second crucial task in traffic monitoring is re-identification of an already tracked object over different scenes [9]. Discriminative and generative photometric colour models are now widely used as the base for visual tracking systems. The discriminative models like support vector machine (SVM), and deep learning-based methods are devoted to regression of target states directly from pixel values. Generative models model a probability distribution over the pixel values directly based on the given target position. Recently, generative models have been successful for visual tracking.

3.1. Sensors and Perception Systems

The information collected using cameras has dimensions beyond 2D, but also includes color values of the pixels. Modern cameras can achieve 60fps or higher and they are relatively low-cost sensor options. However, visual perception can be easily influenced by light, weather, and environmental elements. Further, the light condition is very important to maintaining the accuracy of visual perception. When weather gets worse, various road or environmental elements can be affected by visual lighting and these will be detrimental to the perception of these cameras [17]. The point cloud map obtained by LiDAR is relatively good, while the pixel point cloud map of MmWave radar and stereo camera has unsatisfactory quality. MmWave radar can provide strong resistance to various weather conditions. Obstructions of hail, weather and rain have little influence on radar measurements. consistent high performance, rapid update rate, and multi-target measurement. It is impossible to generate high-resolution flood maps in low-visibility ranges, and it is challenging to distinguish pedestrians from complicated backgrounds in the inspection. In environmental perception systems, we use the radar sensor to detect potential objects in the vehicle's driving environment.

Autonomous driving vehicles should react to the external environment in real-time. They should identify and track objects accurately that includes pedestrians, vehicles, two-wheeled vehicles and so on. High-quality perception is essential for advanced maneuvers like driving at intersections or parking without collisions. Currently, there are four primary types of sensors in the field of autonomous vehicle perception [6]. LiDAR has a high spatial resolution and accurate range measurements. LiDARs can achieve high levels of performance at night or in ambient lighting, and the reflectivity of objects has minimal impact on the measurement accuracy. LiDAR can provide 3D point clouds for object detection in the surrounding environment. However, LiDAR is still expensive to manufacture and buy. In some weather conditions and on some surfaces, the measurement accuracy of LiDAR is reduced, making it unsuitable [18].

3.2. Localization and Mapping

Localization and mapping (SLAM) are tasks essential for every autonomous driving application because they represent the way in which the vehicle identifies its position and the track to follow. The inherent connection between the company's inputs and the related goals is one hot topic of deep learning renowned in analytics and system identification. Automatic characteristic extraction during manufacturing data acquisition improves the data-driven process knowledge by providing users with explainable advanced analytics.

[5] The perception module in an autonomous vehicle requires real-time information for promptly observing and recognizing a wide range of road objects. Due to the importance of data interpretation, high speeds, and dynamic scenes, high resolution detection and tracking are used by state-of-the-art autonomous driving systems. There are several publications that discuss object tracking and detection mechanisms in great detail [16]. Including unsupervised learning, supervised learning, and semi-supervised learning, AI/machine learning methods such as artificial neural networks (ANN) are widely used in the existing systems for many autonomous driving tasks. Detection is a key task in traffic monitoring, including for detecting pedestrians and on-road vehicles, and is essential in object tracking; many deep learning techniques have been applied to object monitoring and pedestrian monitoring to develop sophistication and velocity to approach human-level tracking.[19] Autonomous vehicles need to use their sensors to know their position and orientation, map the environment around them, and navigate through it.

3.3. Control Systems and Decision Making

Recall the input-output sequence (I, O, I, O, etc.) representation for a controller. This can be interpreted as a control system that given a vector I to output O when actuated in an environment that also returns transitions between I and O, responding to unpredictable control systems. Historically, most control systems forked from viewpoints that were strongly rooted into optimal control with planning world models for making informed decisions under uncertainty in future states [20]. Almost invariably, the input-output sequence representation is obtained by parsing the generated trajectories in a batch manner, with all the information (£>! £ being the time steps of the trajectory) being used in the construction of I and 0. However, [21] were able to uncover that by following this strategy very closely an RNN policy only needs the latest observation at each time step to approximately mimic the optimal control.

The success of end-to-end training [22] indicates the feasibility of a new control infrastructure in which a single end-to-end neural network learns the entire control policy of the vehicle. However, learning an entire, high-performance control policy in a fully reactive setting can be difficult. It is not clear that learning can operate at the same level of performance as current control systems for the foreseeable future, especially considering the practical and ethical challenges of such a strategy. Instead, we are likely to see a series of domain-specific control policies form different vehicle-related companies or organizations informed by the successes of end-to-end and behavioral cloning.

3.4. Challenges in Object Tracking and Recognition

In summary, the recognition, classification, and tracking of vehicles and other objects moving on a freeway is a crucial task for the safe operation of autonomous vehicles, as well as for the development of intelligent transportation systems. The task entails the recognition or classification of static and moving vehicles visible in images based on reasoning on their size, shape, and changes in shape. In this section, we want to discuss the state of the art in the use of visual odometry information as a metric visual perception task [23]. If this choice is some limitation as most official datasets and benchmarks do not give enough importance to motion cues for localization and visual odometry, we believe this is still a major aspect to be considered to furtherly improve the robustness of the vehicle tracking and recognition in this challenging environment.

There are several key challenges in object tracking and recognition that dictate critical requirements and principles for adopting robust algorithms for the intended application. Given that traffic scenarios are varied and complex, deep learning models and algorithms require a quality data source and a lot of training, leading to potential adverse effects on the environment and on drivers [16]. Moreover, performance evaluation of the algorithms is often biased and not conclusive. If the diversity of test scenarios is insufficient, the generalization performance is over-estimated, thereby leading to the generalization performance being unable to meet the actual application requirements. If the diversity of test scenarios is sufficient, the test environment can be guaranteed to cover the application range of the model as much as possible, enabling the actual generalization performance to be closer to the generalization requirement of the model.

4. Object Detection and Localization

object detection models have improved results in recent years, for example R-CNN family (Girshick 2015, Ren et al. 2015), YOLO family (Redmon et al. 2015, 2016) and SSD (Liu et al. 2016). We will assume this vehicle detector is well trained, and apply respective detector on the data we have. There have been many ways to filter FALSE- positives, e.g. Kalman filter (Li et al., 2016), matching the lost target with candidates, and re-detection with the detector frame by frame.

[24] Deep learning for object tracking is the latest gold standard and it shows robustness to variations in object appearance and background and to the tracker's viewpoint. Liu, Lin, and He (2018) use deep learning for pedestrian detection and tracking, for scene representation and for areas localization in a surveillance context. In particular, scene representation combines an LSTM recurrent architecture to capture time dependencies in the sequence and a CNN architecture for spatial information (Liu, Lin & He, 2018). Two other contributions propose deep object detectors and deep tracking algorithms to localize and track vehicles in urban scenes (Deng, Ye & Jiao, 2018; Xu et al. 2018). These approaches consider occlusions, different camera placements and the non-stationary dynamic of the scene, examples of the internal and external challenges of the task (Smeulders et al., 2014).[16] In the context of autonomous vehicles, models based on deep learning have been preponderant in image processing and for events and objects recognition and localization, as they have been, among others illustrations, in (Alirezaie et al. 2018; Xu et al. 2018). For the simple and efficient

recognition and localization of vehicles in urban scenes treated in this paper, the deep learning approach seems to be relevant. State of the art

4.1. Single Shot Detectors (SSDs)

Should we use a detection layer to solve this kind of problems? To answer this question, we have to briefly describe the most common detectors for object tracking. The Single-Shot Multibox Detector (SSD) is a fast but less accurate object detector that discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location (v) [25]. Using small convolutional kernels to predict the bounding boxes, the method predicts many possible boxes of different ratios and scales from one single layer. Hence, this is called a one-stage detector. SSDe's detection time is not region-based as in R-CNN, which needs a region proposal network to propose the RoIs. The encoder and the decoder are built into the network architecture instead. Nevertheless, looking up at every feature in the prediction layer to match it with every possible bounding box could be computationally expensive but still allows for the great increase of frames per second speed.

These detectors have had significant improvements in the context of vehicle tracking and classifying objects for Deep Learning algorithms [26]. Based on their results, practitioners and researchers have approached vehicle tracking using DL in different ways: using a R-CNN/R-FCN base approach or starting from a dense regression architecture. They also proposed taking as proxy the instant performance of popular architectures for general object recognition such as YOLO9000 or SSD with either fine-tuning or from scratch training. Among the many proposed techniques, single frame multiple object localization through bounding box prediction, and their subsequent tracking, have remain one of the most valuable and used by industry for vehicle detection. In the vehicle specific tracking literature a newer and faster tracking algorithm named re-detection by re-called worth investigating was presented [27].

4.2. Region-based CNNs (R-CNNs)

The VFC-Net consists of two stages: a very fast R-CNN (RCNN) and a deep vehicle detection and classification network (DVDC-Net) [28]. In Haque and Karray, six objects (target) detection and classification networks (ODCNs) were investigated for moving object detection in fast-moving traffic scenes. Rezatofighi et al. (2017) discussed the performance improvement of the early dramatics with feature pyramid networks in selective search for region-based

detectors. Redmon et al. (2015) introduced you only look once (YOLO), one of the first region-based CNNs for object detection with fast moving vehicles in traffic scenes. The darknet-19 and tiny-yolo architecture were used as the VFC-Net for this research. The VFC-Net was implemented in python using Keras, TensorFlow, and opencv from MATLAB code of Haque and Karray and trained using the training data and testing data from Haque and Karray. The R-CNN architecture consists of two main procedures, i.e., proposal, classification, and regression () [29].

Integrating object detection and a target recognition in fast-moving traffic scenes, such as the tracking of surrounding vehicles on a freeway with an autonomous vehicle, is a critical challenge, because of the limited resource and computation requirement in embedded systems used for autonomous vehicles. Haque and Karray introduced region-based convolutional neural networks (R-CNNs) as region-based deep learning methods associated with target recognition based on deep learning for moving object detection in fast-moving traffic scenes [30]. Then, a very fast vehicle classifier network (VFC-Net) was introduced based on Sobel et al. (2015) by Wang et al.

4.3. YOLO (You Only Look Once)

YOLOv5 is implemented in this study for its integrated 3D tracking capabilities. In a robotic reinforcement learning driven autonomous navigation set up, it can be very beneficial. Specifically, the visual area in the autonomous robotic platform provides the vision system in the YOLOv5 a still single frame of different objects therefore its possible to provide a 99 % accuracy rate with utilizing fps * # of objects of objects where fps is the frame per second that the model is running on the output screen (article).ustum CSPNet, driven by a novel full network with incorporated 3D object tracking algorithms to enable real-time data extraction performance. YOLOv5 follows the same principle methodology as YOLOv3, which has 5 discrete staggered self-monitored streams so that high-level features to the lanet loss can be provided by smallPhi input's cascaded to the output. With feature pyramid network and using a listing of cascaded as the features generated by the CSPNet within the stream, the greatest method combining process comes in a single prediction mechanical system. The YOLOv5 model's adoption allows for hard real-time management with the necessary localization, and monitoring of moving objects with introspective, the output is a flat and very

simple 3D object tracking interface, in this work serving as heading states for the robotic reinforcement learning algorithm.

Object Detection has been further improved in the workflow of the recognition module using a further object detection model You Only Look Once (YOLO), a forerunner in the class of single-stage architecture models (article [31]). Today there have been several versions of YOLO. The YOLO family of models, inspired by the successful convolutional architecture of the VGG-16 ImageNet model from the problem of the image classification using the so-called Darknet-19 architecture, YOLOv1 utilizes ten additional layers employing address transformations, including fully-connected layers, while further improvements, for example, were implemented in version v3 which YOLOv4 and the most recent version YOLOv5 in this work focus on. YOLOv5 is being implemented in this project for its incorporated 3D tracking capabilities for great control system that offers object recognition and localization with state of the art predictions with reasonable speed. The model is utilizing CSPNet (Cross Stage Partial Network) with a novel frozen and pre-mixout technique.

4.4. Performance Metrics for Object Detection

The detection accuracy for different viewpoints and distances around the vehicle for different objects, compared to the baseline pretrained object detector, was achieved using a five-class Keras model [32]. The mean average precision (mAP) was used to validate different glare-reduction methods that were used to process input data [33]. Safe and reliable ground truth was defined as the union of the annotations from all annotators . The labyrinth game involves navigating through various randomly generated labyrinths and drivers have a similar horizontal field of view with moving head enabled. The laps end when a vehicle reaches a zero distance (collision), drifts off-road, or crashes at a barrier. Drift and crash events have the annotation collision [34]. For safety purposes, we added the same annotations to the labyrinth game as the real-life data because we are interested in sharing the most critical information with the method under investigation. Defining these two critical error modes as well as the relaxation of the scoring metrics, we tested these attention-based and the saliency-based techniques on the four data subsets that we designed by including test settings . We designed a training subset shared in both the attention and collective laser signal experiments. The remaining data were used for fine-tuning the model to their special characteristics, and the relaxation tests are performed on the disjoint part, as already described .

5. Object Tracking Techniques

One important research direction is to break the perfect perception into several subtasks, each with well-defined quality and a clear grading matrix. The quality of 3D object detection can be graded by the discrepancy between the detected shape and the true 3D shape. 2D or joint 2D3D tracking quality can be graded by the computation of tracking IOU in addition to 3D static or dynamic shapes differences. The quality of long-term pedestrian motion prediction can be graded by the computation of trajectory FDE. As well, understanding the knowledge heading the learnable deep prediction models can guide 4D constitutive curve estimation, which can have applications not only in computer vision, but also in bioenergetics, traffic jamming, and network robustness, as it will be detailed in future works. [35]Therefore, other work should include understanding very deeply, and ultimately breaking down/ mitigating, these types of misorientations. Here, we develop a dynamic object tracking framework for an autonomous UAV monitoring and/or control system. In this framework, we first find a fast 3D propulsion network that is capable of producing the fast state of the decision-making of moving objects. The UAV behaviors, mostly moving from a shop-like proposal to another, and developing a manner of computational speed dating in a practical way. Then, after the retrieval of 3D vehicles, we develop new sampling strategies to enrich the relevant database in context for predictor networks. This final passive sensor describes a full life cycle and very good prediction accuracy. Allowing the UAV to successfully guide its motion on the basis of a vision-based 3D object tracking and decide new stations from target tracking positions for a very good 3D object die mixology.

Passive-based system and camera networks are robust against occlusions from other vehicles and lights and signs, therefore, it is very popular. [5] In order to maintain high quality of vehicle detection and tracking outputs even in such real-world challenging conditions, we use Michael and Adam's framework of end-to-end 3D object detection, which naturally tackles sensor dropout and varying speeds challenges. In order to obtain long-term predictions or tracking trajectories, we develop a new alert-based tracking technique, which is able to accurately track vehicles over long distances and time spans, while also being capable of predicting future maneuvers it intends to make. It takes both the 3D bounding box and the image as inputs and directly outputs the next step into the future. Notice that the image is always the image cropped in the 2D skating area defined around the 3D vehicle's bounding box so that the camera perspective will not change. Moreover, the network does not estimate

kinematics at all, and hence has no explicit velocity or acceleration/ jerk outputs. Nevertheless, thanks to convolutional state compression (CSC) following and network outputs, the network learns a vector representing a complex state space that encapsulates both physical dimensions and final causes. This CSC is derived and explained in supplementary therefrom, simplifying various physical dimensions estimation, while taking efficiently into account natural concomitant occlusions that can happen in urban scenes.

5.1. Kalman Filters

The demand for an alternative tracking mechanism has been addressed by deep learning, recently receiving increasing attention, because of its superior capabilities for handling large amounts of complex data, and desirable applications like object recognition, image segmentation, etc. It has recently also been successfully applied in the domain of tracking and localization, demonstrating improvements in accuracy, real-time performance and generalization to complex cases. Orthogonally, due to: (i) increasing awareness in the (in)security of deep learning, (ii) finite-size effects of deep neural models, (iii) the huge amounts of data necessary to train such models, a huge and re-emerging interest exists on probabilistic models over deep learning models, especially in the reduced-sample regime, obtaining huge success. Also, work on employing probabilistic techniques to deep architectures is also gaining momentum, e.g. Variational Bayesian Learning to quantitatively represent and handle uncertainty in deep networks, or unsupervised learning to design deep architectures for compression and generative models.

Autonomous driving demands accurate and real-time estimation of vehicle position and velocity, typically carried out using a variant of the Kalman Filter (KF) [36]. These filters are a popular class of probabilistic estimators, providing a smooth estimate of the posterior distribution of the internal state of a system from the observed measurements dataset [37]. However, the success of the KF approach hinges upon system linearity, Gaussian noise, and prior knowledge of system state-space dynamics, which are very restrictive assumptions for the myriad complex, highly and non-linear systems, like autonomous vehicles under different weather and driving scenarios. In these situations, accurate modeling and accurate prediction become challenging, leading to poor localization and tracking, necessitating improved and alternative tracking approaches [9].

5.2. Particle Filters

After obtaining the preliminary tracking results, the long-term consensus data association is applied to associate the tracked object with the trajectory model [38]. Finally, the detection confidence score and trajectory consistency score are used for final data association. The optical flow and depth information can also be employed to refine particle proposals in the particle filter, so as to handle the scale, spatial location and 3D shape deformations as well as occlusions. In [39], the appearance information is learned and employed to select the best frame representation, which makes the model suitable for multiple movements of the contender as well as for object tracking in urban city conditions. The relevant frames help to build a better model which improves the tracking in cluttered and occluded scenes.

Particle filters are often chosen as a tracking method to solve the filtering problem when the system dynamics of the object representation observed are nonlinear, or the object initiated distribution is uncertain. As a kind of Monte-Carlo- based algorithm, the particle filter approximates a tracking distribution with weighted particles, which are generated by sampling from the proposal distribution and re-weighting according to the likelihood function. So the performance of particle filter is influenced by the proposal distribution. In [37], Nie et al. combines a fully convolutional network with a particle filter with importance sampling for tracking. A series of short-term proposal candidates are generated by sampling from the generative model P (z v t | y 1:t), and the likelihood scores for each candidate are evaluated and computed through the observation model P (y t | z v t), based on the cross-correlation features.

5.3. Deep Learning-based Tracking

Based on a person's facial videos, we can automatically quantify the facial and ocular indicators, and then in conjunction with intelligent models, estimate his/her heart rate, heart rate variability (HRV), and blood pressure (BP) changes, among other parameters. SRAGE utilizes facial videos due to the fact that (1) we can take facial videos with our own smartphone without any additional & external devices, (2) we can keep monitoring the changes in physiological parameters over a long period, from morning to night, in various postures, with different garments or skin colors, and (3) facial and ocular indicators are considered to be promising noninvasive parameters for estimating the onset of hypertension and risks [17]. Along with the large convenience to users, SRAGE enhances continuous and ambient

monitoring enormously when COVID-19 keeps people at home and restricts their outdoor time and activities.

Hypertension is a major public health challenge worldwide, and it is estimated that more than 10 million people die prematurely due to uncontrolled elevated blood pressure every year [40]. Hence, measuring blood pressure, especially monitoring it during daily activities, is critical to prevent severe and preventable consequences such as stroke. In this work, we present a mobile application, SRAGE, which is capable of tracking and monitoring the onset of hypertension and risks by extracting essential physiological parameters from videos of the human face taken by a smartphone camera [41]. This application is aimed at addressing the challenges caused by current wearable and nonwearable monitoring devices, such as inconspicuous wearing, discomfort, and the need to purchase new devices and replace them constantly for new versions.

6. Deep Learning for Object Recognition

The immense progress of deep learning has enabled the development of various deep-learning-based object tracking approaches, which are more accurate and robust compared to the traditional methods. [42] The traditional object-tracking algorithm has limited capacities when they are deployed in complex scenes including occlusion, illumination variation, affineness, and similar objects. Here, we present a comprehensive review of the current state of the art in deep-learning-based discovery for object tracking, following mainly the deep-learning-based visual tracking (DVT) field within computer vision, as shown in Figure 1.

[43] Deep learning has rapidly advanced the state of the art in object recognition, with deep learning techniques now achieving or surpassing human performance on standard benchmarks for image recognition and video detection. It has been reported that deep-neural-network-based approaches are now the dominant method for object detection and recognition tasks in the area of computer vision. [41] This method, in general is called the deep-learning-based object recognition. In an autonomous vehicle, for performing any classification and detection task, the vehicle perceives its surrounding environment through different sensors such as LIDAR, radar, and camera. However, imaging sensors are more popular in object detection and classification tasks, so the focus of this article will be on purely image-based deep-learning-based object recognition methods. The area of computer vision and deep learning has gone through highly significant changes during the last few years, thanks to

extraordinary improvements in processing power—particularly the ready availability of high-performance parallel hardware in the form of graphical processing units (GPUs).

6.1. Image Classification

There are two types of KNN models; KNN for test and radar data tracking and correlation analysis (firm reflective points in test segments from the neural network and radar data). In two separate neural network models of LIDAR and radar sensor data, two Neural Network models (NN) were designed to classify pedestrians and vehicles in a mixed neural network model. obvious correlation between the distance to the test side of the architecture of the network layer nearest LIDAR sensor and radar sensors. This so-called LIDAR system includes sensor layer, fusion layer, LIDAR data LIDAR convolutional neural network and closest Radar convolutional neural network. Although the proposed hybrid method between LIDAR and radar systems to examine collected data from TNO can fail in many connections and inside lanes similar to the second segment connected to the conventional systems. The advantage of this new method and radar system is robust object classification for pedestrians, vehicles and cyclists in different types of parking .

Object recognition in autonomous vehicles refers to the task of recognizing objects, such as vehicles, pedestrians, and motorcyclists, whether they are static or moving. This task usually integrates object detection with object tracking, which is generally referred to as MOT (multiple-object tracking). After each dimensionality reduction technique is applied to transform the data points into zero-mean and unity-quantified scales, the k-means clustering algorithm is exploited to group closest points together. The main goal of this research is to be able to visualize and locate the detected objects [44], and it is usually expressed by the object centroids in 3D space. The proposed algorithm is successful in subsequence object tracking and classification. The capabilities of the proposed algorithm in classifying vehicles, pedestrians, and cyclers are both quantitatively and qualitatively very well, which are compared to the well-known detection and classification algorithms such as the Faster R-CNN, YOLOv5, MobileNetV2, and ResNet50.

6.2. Object Detection and Recognition

Pedestrian detection and tracking needs to work under varying conditions of pedestrian density, distance, shadow, lighting, occlusion, scale, pose, and motion (relative to the moving

vehicle) [45]. We review methods that detect, track, or both detect and track pedestrians in driving video. The current best performing multi-object tracker (MOT) trackers are effective when objects are visible all or most of the time, but as yet no tracking method removes the need to also use detector outputs as a high confidence alert. Exercise-induced variations also add another layer of challenges. In contrast, some object detectors specialize in low-resolution, partly occluded object detection, for instance by using of rectangles from detection at the last timestep for tracking. Single-object trackers trained on specific challenges like KITTI have an opportunity to perform very well because they can focus and adapt to the specific types of false positives, or misplacement errors that are made by the driving video specific detectors. In the light of the positive results presented, we aim to investigate whether a dedicated deep learning –based tracker can achieve improvements over the detection outputs from generic Frontier-RCNN object detectors on specific KITTI pedestrian detection and tracking tasks.

Robust detection of objects in video, especially in the context of autonomous driving, is crucial [46]. In the recent years, the best performing deep learning-based object detectors are trained and evaluated as image classifiers one frame at a time. Here we develop the DeepSort method to simultaneously track and detect bounding boxes around objects in video. The key idea is that the object tracks are used to create a high-confidence "detection" where the objects are known to be located at the previous time step; this high-confidence detection is essential for robustly classifying small partially occluded, and difficult to see objects such as pedestrians from a moving vehicle. Our approach is simple, efficient, and straightforward to use, runs at frame rate on a laptop, and does not need to be re-learned for each new paradigm. We show it consistently performs better than previous methods in pedestrian detection and tracking on the KITTI dashboard camera driving dataset, the Caltech pedestrian dataset, and on the MOT16 benchmark. We also show qualitative improvements in tracking in a range of other object tracking contexts.

6.3. Fine-grained Recognition

To demonstrate the effectiveness of our approach in fine-grained dataset chosen synthetic Car196-3D dataset [47]. We also compare the performance of our boxGPT-based approach with pose-based relational transformer-based (PoRT) and regular transformers. PoRT. Regular transformers. We start by comparing the performance based on the original 2D car images, where the EV model has multiple fine-tuning heads for different canonical views. We

get the fine-tuning following the top-l KNN protocol, where the features are easily replaced with the distance to linear normalized features. Concat represents flatten+Concat for all positions relating to the target classes listed on bottom of car pictures in top-l and 2D. here, 2D 2D and 2D V 3D denotes different experiment settings about original and 3D rotated EV models. AutoDiscovery. When comparing the performance based on these discovered object parts on MVD from rig shod, for relic bodies, while it is hard to observe substantial quality gain from these hyper synchronous regional images except from the view of 3D–2D projection, one can substantially improve the performance to the point with different wearing in the rigid parts.

Fine-grained object recognition corresponds to the identification of objects of interest with high intra-class differences [48]. Traditional object recognition often employs a classification-decadent-detection pose, which suffers from the large spatial context that is poorly preserved during region-based classification results. In this section, we aim to improve the performance of our transformer-based object recognition system by making it more robust to distractors in the surrounding region. We feed this feature vector into the final fully connected linear layer to compute the fine-tuned class prediction. In each box, we only show the top-3 object classes, due to visual clarity. For objects close to each other, they may share similar neighbors and hence may be less similar to each other, which may hurt the performance of feature-based approaches. By directly taking the auto-discovered spatial relationship directly compared from different objects, spatial systems such as STN can have better generalization ability on image-based fine-grained problems. The advantages of this pose-based approach is mostly obvious for genetically resealable medium objects, where auto-discovery of proper detailed parts benefits the overall performance.

7. Datasets and Benchmarks for Autonomous Vehicle Object Tracking and Recognition

As we mentioned in section–4, there are several important datasets created by different countries. These datasets can be broadly categorized: BDD100K, PASCAL VOC, COCO, Waymo [49]. BDD100K Traffic light (dangerous area) is replaced by the fire brigade which is usually parked at the roads on the encouraging side. Inspired by the test dataset of BDD100K (BDD), our dataset is annotated in terms of horizontal and vertical lengths, colored, and shapes so that relevant dimensions of objects have been obtained. objects are, in fact, small-

scale objects. Every object is automatically shortlisted in the best possible matched class group, in the case of incorrect offering; it is listed as a new class.

As of today, DNN is utilized for different applications such as simultaneous object detection and recognition as well as target tracking, context learning/object tracking, target reacquisition, and temporal fusion. [50]. In this section, it is mainly focused on the vehicle object tracking and recognition challenges. We introduce the most important and widely used datasets and benchmarks for obtaining depth in this field and proving the actual capabilities of different methods. There are several different benchmarks and suites which can be employed in order to highlight effective and ineffective parts of the algorithms. For challenges and several startups which motivate specialists to design and develop new improved algorithms by combining the best parts of different methods and use the strengths of different methods to solve the difficulties with high precision, such as Central Processing Unit (CPU), Linux, and file systems implementation; different sources of Wi-Fi, GPS, and cell phone locationwhile driving have noise and/or low accuracy [51]. The most popular public datasets are employed and analyzed in this chapter

8. Applications of Deep Learning in Autonomous Vehicles

The safety and reliability requirements of object tracking in an autonomous vehicle result in strict requirements of robustness and real-time predictability. In this chapter, we focus on ways to model various objects in the scene and develop a robust and efficient object tracking system based on deep learning architectures. The evolution of neural network-based object detection algorithms stems from the use of feature-based methods in the tracking data [52]. We focus on studies that combine sensor data for object tracking and recognition which are essential for autonomous vehicles.

Machine learning enables the process of training an algorithm to establish patterns and connections from sensory input [3]. This technique can be used in tandem with sensory data to identify and classify various objects in the real world. Once the sequence of events is processed and all objects are identified in a scene, this culmination of the process is object tracking. These techniques can bring together various types of sensor data for effective tracking in the real world [41]. Additionally, the use of neural networks for object tracking is considered as an inherent part of autonomous vehicle systems.

9. Ethical and Safety Considerations in Autonomous Vehicle Object Tracking and Recognition

Security and privacy are no longer seen as two conflicting ethics and are given equal importance. AVs and other autonomous machines still have to generate usable safety messages that are difficult enough for attackers to generate human imperceptible adversarial perturbation, but once it is perturbed, the detection/recognition system either fails to detect the traffic signal or detect a changed traffic sign. As part of to solve this problem stated they have proposed two methods to generate adversarial attacks such that, a) they are challenging for machine to generate an adversarial example which are hard to be detected by machine, and b) human generated driving safety messages are not detected by human either. A machine and human cannot always find efficient algorithms to generate adversarial examples in parallel [53].

The increased popularity of deep learning in computer vision and autonomous vehicles has brought out a great concern of adhering to ethical and safety considerations in developing such technologies. Many studies have started addressing ethical and safety considerations in object tracking and recognition in autonomous vehicles to ensure its socially and legally accepted use [54]. However, research on ethics related to fully data-driven perception systems employed in autonomous vehicles (AVs) and its impacts on object tracking and recognition is still an open area. Notably, there is a recent work which highlights that safety (ethics) and perfomance do not have to be defined independently in AV decision making. Security and safety are conflicts ethics disguised [5]. The study stated that adversarial examples can be effective attacks on the perception systems of the AVs to accomplish safety-related goals. They created human adversarial attacks by only using perturbation of the human generated road signs using 4 different real world scenes. Then using a state of the art traffic object tracking system, they showed that the traffic signs could achieve the desired target size.

10. Future Directions and Emerging Trends in Deep Learning for Autonomous Vehicles

(Deep Learning and AI technologies: In transforming the mobility sector, manufacturers and technology carriers have critically recognized the need for enhanced perception in which the vehicle could detect, track and recognize a variety of objects that come across its way. To prevent near misses with bus or animal, to expedite the correct speed control earlier than getting on the icy road or crossing speed bump without damaging the vehicle, a perfect notion

is prerequisite. Precise perception and localization is crucial to continuous vehicle safety. LiDAR can provide dense depth maps even at long-range and works well in poor lighting conditions but is still an expensive technology on the road to deployment. Imagery, associated with pixel-wise wise depth predictions offer an interesting alternative, facing the crucial challenge of an accurate handling of object edges in 3D [55]. In AD, deep learning has taken an important role, and as especially in object detection, one-stage and two-stage algorithms have fascinated our attention. Progress in object detection arises from network architecture, backbone, reasoning, bounding box proposals, among others. However, for AD, visual object detection shows great difficulties so far as mobile obstacles have diverse scenarios in which they appear with difficulties if associated with obstacle attention or parking sensor activations in AD. Existing detectors are examined in scenarios of a few obstacles, few scenarios and day light, with positive impacts associated with few points of interest or Boxs till the assistant of context or consultancy. Lastly, radar signals also need to be involved to reach plenitude detection, context understanding and the target distinction.)

The increased focus on improving deep learning-based autonomous vehicles has seen greater collaboration and integration between the vision and radar systems. This approach to multitask, sensor fusion has found success in improvements in concurrent measurement and detection of objects [17]. This method possesses the advantage of real-time detection and tracking and improved reliability, consistency, and accuracy over single-sensor systems. Technology shifts in radar sensor systems, newer driver assistance options (ADAS) adoption, and upcoming modifications in legislation are expected to influence industry demand for radar sensors. It is evident that deep learning has achieved extensive excellence in the domain of object detection and tracking as evident from massive monetization of different types of vision systems, suggesting the improvement of further systems based on deep learning models gained popularity [56].

References:

[1] G. Zhang, J. Yin, P. Deng, Y. Sun et al., "Achieving Adaptive Visual Multi-Object Tracking with Unscented Kalman Filter," 2022. ncbi.nlm.nih.gov

[2] Tatineni, Sumanth, and Venkat Raviteja Boppana. "AI-Powered DevOps and MLOps Frameworks: Enhancing Collaboration, Automation, and Scalability in Machine Learning Pipelines." *Journal of Artificial Intelligence Research and Applications* 1.2 (2021): 58-88.

[3] Shahane, Vishal. "Harnessing Serverless Computing for Efficient and Scalable Big Data Analytics Workloads." *Journal of Artificial Intelligence Research* 1.1 (2021): 40-65.

[4] Abouelyazid, Mahmoud, and Chen Xiang. "Architectures for AI Integration in Next-Generation Cloud Infrastructure, Development, Security, and Management." *International Journal of Information and Cybersecurity* 3.1 (2019): 1-19.

[5] Prabhod, Kummaragunta Joel. "Utilizing Foundation Models and Reinforcement Learning for Intelligent Robotics: Enhancing Autonomous Task Performance in Dynamic Environments." *Journal of Artificial Intelligence Research* 2.2 (2022): 1-20.

[6] Tatineni, Sumanth, and Anirudh Mustyala. "AI-Powered Automation in DevOps for Intelligent Release Management: Techniques for Reducing Deployment Failures and Improving Software Quality." Advances in Deep Learning Techniques 1.1 (2021): 74-110.

[7] Y. Azadvatan and M. Kurt, "MelNet: A Real-Time Deep Learning Algorithm for Object Detection," 2024. [PDF]

[8] J. W. Pyo, S. H. Bae, S. H. Joo, M. K. Lee et al., "Development of an Autonomous Driving Vehicle for Garbage Collection in Residential Areas," 2022. ncbi.nlm.nih.gov

[9] J. Zhang, Y. Liu, Q. Li, C. He et al., "Object Relocation Visual Tracking Based on Histogram Filter and Siamese Network in Intelligent Transportation," 2022. ncbi.nlm.nih.gov

[10] H. Wang, Y. Cai, and L. Chen, "A Vehicle Detection Algorithm Based on Deep Belief Network," 2014. ncbi.nlm.nih.gov

[11] W. Luo, B. Yang, and R. Urtasun, "Fast and Furious: Real Time End-to-End 3D Detection, Tracking and Motion Forecasting with a Single Convolutional Net," 2020. [PDF]

[12] H. Bang, J. Min, and H. Jeon, "Deep Learning-Based Concrete Surface Damage Monitoring Method Using Structured Lights and Depth Camera," 2021. ncbi.nlm.nih.gov

[13] Y. Ed-Doughmi, N. Idrissi, and Y. Hbali, "Real-Time System for Driver Fatigue Detection Based on a Recurrent Neuronal Network," 2020. ncbi.nlm.nih.gov

[14] W. Haider Bangyal, R. Qasim, N. ur Rehman, Z. Ahmad et al., "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," 2021. ncbi.nlm.nih.gov

[15] Z. Zhang, G. Li, Y. Xu, and X. Tang, "Application of Artificial Intelligence in the MRI Classification Task of Human Brain Neurological and Psychiatric Diseases: A Scoping Review," 2021. ncbi.nlm.nih.gov

[16] Y. Zhu, M. Wang, X. Yin, J. Zhang et al., "Deep Learning in Diverse Intelligent Sensor Based Systems," 2022. ncbi.nlm.nih.gov

[17] Z. Wei, F. Zhang, S. Chang, Y. Liu et al., "MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review," 2022. ncbi.nlm.nih.gov

[18] D. Fernandes, T. Afonso, P. Girão, D. Gonzalez et al., "Real-Time 3D Object Detection and SLAM Fusion in a Low-Cost LiDAR Test Vehicle Setup," 2021. ncbi.nlm.nih.gov

[19] C. Chen, B. Wang, C. Xiaoxuan Lu, N. Trigoni et al., "A Survey on Deep Learning for Localization and Mapping: Towards the Age of Spatial Machine Intelligence," 2020. [PDF]

[20] R. Gandikota, "Computer Vision for Autonomous Vehicles," 2018. [PDF]

[21] A. Biglari and W. Tang, "A Review of Embedded Machine Learning Based on Hardware, Application, and Sensing Scheme," 2023. ncbi.nlm.nih.gov

[22] H. Cao, W. Zou, Y. Wang, T. Song et al., "Emerging Threats in Deep Learning-Based Autonomous Driving: A Comprehensive Survey," 2022. [PDF]

[23] Z. Wu, F. Li, Y. Zhu, K. Lu et al., "Design of a Robust System Architecture for Tracking Vehicle on Highway Based on Monocular Camera," 2022. ncbi.nlm.nih.gov

[24] A. Md Niamul Taufique, B. Minnehan, and A. Savakis, "Benchmarking Deep Trackers on Aerial Videos," 2021. [PDF]

[25] M. H. Sheu, S. M. Salahuddin Morsalin, J. X. Zheng, S. C. Hsia et al., "FGSC: Fuzzy Guided Scale Choice SSD Model for Edge AI Design on Real-Time Vehicle Detection and Class Counting," 2021. ncbi.nlm.nih.gov

[26] L. Arab Marcomini and A. Luiz Cunha, "Truck Axle Detection with Convolutional Neural Networks," 2022. [PDF]

[27] V. Thakar, W. Ahmed, M. M Soltani, and J. Yuan Yu, "Ensemble-based Adaptive Single-shot Multi-box Detector," 2018. [PDF]

[28] L. Liu, Z. Pan, and B. Lei, "Learning a Rotation Invariant Detector with Rotatable Bounding Box," 2017. [PDF]

[29] P. Srimuk, A. Boonpoonga, K. Kaemarungsi, K. Athikulwongse et al., "Implementation of and Experimentation with Ground-Penetrating Radar for Real-Time Automatic Detection of Buried Improvised Explosive Devices," 2022. ncbi.nlm.nih.gov

[30] N. Ghatwary, M. Zolgharni, and X. Ye, "Early esophageal adenocarcinoma detection using deep learning methods," 2019. ncbi.nlm.nih.gov

[31] N. Adiuku, N. P. Avdelidis, G. Tang, and A. Plastropoulos, "Advancements in Learning-Based Navigation Systems for Robotic Applications in MRO Hangar: Review," 2024. ncbi.nlm.nih.gov

[32] R. Bloomfield, G. Fletcher, H. Khlaaf, L. Hinde et al., "Safety Case Templates for Autonomous Systems," 2021. [PDF]

[33] M. Z. Alam, Z. Kaleem, and S. Kelouwani, "How to deal with glare for improved perception of Autonomous Vehicles," 2024. [PDF]

[34] Y. Zou, W. Zhang, W. Weng, and Z. Meng, "Multi-Vehicle Tracking via Real-Time Detection Probes and a Markov Decision Process Policy," 2019. ncbi.nlm.nih.gov

[35] L. Y. Lo, C. Hao Yiu, Y. Tang, A. S. Yang et al., "Dynamic Object Tracking on Autonomous UAV System for Surveillance Applications," 2021. ncbi.nlm.nih.gov

[36] F. Leon and M. Gavrilescu, "A Review of Tracking, Prediction and Decision Making Methods for Autonomous Driving," 2019. [PDF]

[37] S. Scheidegger, J. Benjaminsson, E. Rosenberg, A. Krishnan et al., "Mono-Camera 3D Multi-Object Tracking Using Deep Learning Detections and PMBM Filtering," 2018. [PDF]

[38] R. Caiand Peng Zhu, "Occlusion-aware Visual Tracker using Spatial Structural Information and Dominant Features," 2021. [PDF]

[39] L. Wang and D. Sng, "Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey," 2015. [PDF]

[40] S. M. Marshall, A. R. G. Murray, and L. Cronin, "A Probabilistic Framework for Quantifying Biological Complexity," 2017. [PDF]

[41] M. S. Bahraini, A. B. Rad, and M. Bozorg, "SLAM in Dynamic Environments: A Deep Learning Approach for Moving Object Tracking Using ML-RANSAC Algorithm," 2019. ncbi.nlm.nih.gov

[42] E. Khatab, A. Onsy, and A. Abouelfarag, "Evaluation of 3D Vulnerable Objects' Detection Using a Multi-Sensors System for Autonomous Vehicles," 2022. ncbi.nlm.nih.gov

[43] A. Luckow, M. Cook, N. Ashcraft, E. Weill et al., "Deep Learning in the Automotive Industry: Applications and Tools," 2017. [PDF]

[44] F. Shafiei Dizaji, "Lidar based Detection and Classification of Pedestrians and Vehicles Using Machine Learning Methods," 2019. [PDF]

[45] M. Córdova, A. Pinto, C. Carrozzo Hellevik, S. Abdel-Afou Alaliyat et al., "Litter Detection with Deep Learning: A Comparative Study," 2022. ncbi.nlm.nih.gov

[46] M. Carranza-García, P. Lara-Benítez, J. García-Gutiérrez, and J. C. Riquelme, "Enhancing Object Detection for Autonomous Driving by Optimizing Anchor Generation and Addressing Class Imbalance," 2021. [PDF]

[47] J. Sochor, J. Špaňhel, and A. Herout, "BoxCars: Improving Fine-Grained Recognition of Vehicles using 3-D Bounding Boxes in Traffic Surveillance," 2017. [PDF]

[48] K. Valev, A. Schumann, L. Sommer, and J. Beyerer, "A Systematic Evaluation of Recent Deep Learning Architectures for Fine-Grained Vehicle Classification," 2018. [PDF]

[49] J. Lian, Y. Yin, L. Li, Z. Wang et al., "Small Object Detection in Traffic Scenes Based on Attention Feature Fusion," 2021. ncbi.nlm.nih.gov

[50] Z. Wei, F. Zhang, S. Chang, Y. Liu et al., "MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review," 2021. [PDF]

[51] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art," 2017. [PDF]

[52] Y. Li and J. Ibanez-Guzman, "Lidar for Autonomous Driving: The principles, challenges, and trends for automotive lidar and perception systems," 2020. [PDF]

[53] J. Kaur and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: a review," 2022. ncbi.nlm.nih.gov

[54] D. Garikapati and S. Sudhir Shetiya, "Autonomous Vehicles: Evolution of Artificial Intelligence and Learning Algorithms," 2024. [PDF]

[55] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A Survey of Deep Learning Techniques for Autonomous Driving," 2019. [PDF]

[56] J. Fayyad, M. A. Jaradat, D. Gruyer, and H. Najjaran, "Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review," 2020. ncbi.nlm.nih.gov