Deep Learning for Autonomous Vehicle Image and Video Processing

By Dr. Andrei Tonkoshkur

Associate Professor of Computer Science, Belarusian State University of Informatics and Radioelectronics (BSUIR)

1. Introduction to Deep Learning in Autonomous Vehicles

Autonomous vehicles have garnered increasing attention due to advancements in deep learning algorithms [1]. Vision-based driving assistance often uses computational architectures like recurrent neural networks due to the low cost of cameras. However, self-driving car systems requiring the low-latency operation benefit from well-optimized software and hardware stacks. Further, the training process of deep learning models is computationally expensive, often taking days, or even weeks, to finish [2]. Deep learning is setting new benchmarks in various fields almost daily. Owing to this reason, the automotive industry also has reached great heights with applications such as intelligent vehicle control, image stabilizing devices as an advanced version of Anti-lock Braking System, avoid object detection, automatic parking, traffic sign recognition and others. In this paper, the various applications of the deep learning technique viz vehicle count and control, vehicle detection and segmentation,lane detection, driver fatigue and drowsiness detection, etc., to deal with the difficulties of the vehicle monitoring and detecting the vehicle problems [3].

1.1. Overview of Autonomous Vehicles

To improve pre-processing step, model development, data assimilation, and to choose more suitable sensors for deep learning, in self-driving car systems, data and sensor noises have been considered by updating datasets with accurate dimensions. Autonomous systems must process lessons learned from the supervised learning (to convert the sentence to "Autonomous systems must learn lessons from different types of learning techniques including supervised learning"), which is a function to predict the outputs. In some cases, the objective is to detect the uncertainties from its own process and the learned concepts, manage them, and reduce sensors or models crosstalk to make the model robust in the worst cases. On the basis of the acquired raw information, autonomous systems can use neural networks (with

different topologies like CNNs or long short-term memory (LSTM) and time convolutions (TC)) to extract the lessons from the ambient by learning different features including. Without laboriously engineering programs that cannot handle boundary conditions very well, continuous learning algorithms (negative feedback loop algorithms instead of supervised learning of software loop of perception) must be constructed to learn in any situation as they interact with their ambient. [4]

The automotive industry is currently undergoing a significant evolutionary shift toward autonomous or self-driving vehicles, robotics, and alternative powertrains. Many studies suggest that these three major transformations will change the industry's momentum in terms of product development (vehicle and related services production), sales, and the market demand pattern. The automotive ecosystem is going through a real and immediate engineering shift toward the electrification of the entire propulsion system, new vehicle dynamics control and new vehicle AI systems, safe human–machine interactions and fully automatic vehicle management system with reliable sensor and sensor data processing algorithms[in this project section, we concentrate our analysis to sensor data decoding to get signal from image or video data]. Deep learning techniques have significantly improved the performance of automotive scene understanding systems, particularly in computer vision[in this project section, we concentrate our analysis to this branch of science]. In fact, the continuous actual improvements are the main rationale for quickly integrating deep learning in vehicles. [5] To this aim, image and video data are usually analyzed and processed to get required information for driving decisions, for example: 1) classifying the detected objects in an image (image processing or classification); 2) modeling (learning model) or reproducing detected vehicles and traffic objects in the scene to get information; 3) handling of images and videos to make them more robust, for instance by making them invariant to view points or noise perturbations (feature extraction); 4) aligning data for more efficient storage or to enable collective classification or drive action (condensation or clustering methods); 5) visual demonstration of driving skills, transfer information through camera to the driver or to the car's mates or get information with physical interference[citations[see the references in this article]]; 6) proving other learning tasks like representation learning or transfer learning through comparing this learning process with different similar learning algorithms (image matching and co-training); 7) performing an analysis of space coordination or temporal variation to recognize objects (spatial and/or temporal segmentation); and 8) finding the

simplest and all canonical distributions for clusters found with clustering methods[in this project, analysis and treatment of videos/images are from scene understanding, scenarios, and traffic sign detection through object detection and/or identification.]. Some additional details to consider are: A sensor data decoder translates sensor data raw images or videos to the required format and processes these formats by interpretation of other processing branches like image processing or classification, feature extraction, scene analysis, action detection, image matching, etc. [6]

1.2. Role of Deep Learning in Image and Video Processing

[7] The rise of deep learning methods significantly improved the inference performance and model sizes in object recognition, detection, and tracking in images and videos. These state-of-the-art recognition and detection systems surpassed the human level performance for various object categories in the ImageNet Large Scale Visual Recognition Challenge. Most of this success of deep learning was made possible through large-scale datasets with many annotated training images and millions to billions of model parameters, which was achieved through distributed computing in the last few years. Due to the recent fast development and open sourcing of deep learning methods, visual recognition accuracy levels were significantly improved across multiple benchmarks, datasets, and testbeds. In particular, deep learning significantly improved the accuracy and speed of traffic monitoring systems, especially for traffic surveillance networks. Most intelligent systems, such as autonomous driving, robotics, smart homes, and connected vehicles greatly benefit from deep learning-based computer vision methods. Deep learning techniques are used to detect, segment, recognize, classify, and track objects, scenes, and events in their sensory modalities of images and videos.[5] Data can come from multiple sensors such as multimodal image and video camera sensors, LIDAR, RADAR, GPS, and IMU. Learning robust and noisy features from camera input is a key challenge in unsupervised, self-supervised, and supervised learning methods, but multi-sensor fusion is typically employed to mitigate camera sensor noise and uncertainty. Passive sensors, such as RGB or intensity image cameras, provide semantic information of the objects and scenes in the world inside their field of view. Multi-modal or active sensors, such as LIDAR or RADAR, are typically used for obstacle localizations, 3D feature extractions, and segmentation tasks. Deep learning can play a significant role to solve the 3D localization, mapping, and environmental sensing and decision challenges of these sensors, as well. In this

chapter, we will focus solely on scene understanding and visual recognition tasks in deep learning applied to still images and videos.

2. Fundamentals of Deep Learning

The main objective of the proposed method is to introduce a fast U-Net architecture with a light-weight backbone structure that can detect objects with limited computational requirements in real traffic scenarios. Secondly, an interaction analysis module was proposed for the ablation study to verify the importance of the interaction information for the overall scene understanding [8]. The optimal network was compiled with a lightweight efficient encoder, and the interaction module provided improved semantic segmentation output, creating a pre-classification segmental label. Various models use in-vehicle applications for edge AI and frequently need to perform frame-by-frame semantic segmentation analyses with minimal parametric overall customer requirements [6].

With the prevailing U-net, Vision Transformer (ViT) and Resnet architectures, video and image semantic segmentation results in autonomous vehicle scenario analyses have notably improved recently. Although these models are powerful in understanding high-level video contents, the outputs are not robust enough for dense scenarios. These models tend to miss small objects, which is a crucial drawback in visual perception for autonomous driving [9]. Studies demonstrate that models are more successful in detecting large objects, such as pedestrians, but they are generally significantly less successful at detecting pedestrians compared to cars and buses. While object detection is a basic requirement for scene understanding, object interaction captures a different level of complexity for autonomous vehicle visual perception.

2.1. Neural Networks

All these architectures attempted to suppress noises coming from "close-to-vehicle" objects located in the surroundings of autonomous vehicle driving, while they usually prescribed to focus on approaching obstacles located at safe driving distances. Such architectures can generally decrease the DWS reliability. The recent literature considers metrics such as time to collision, level of danger, relative mass and relative acceleration/deceleration as properties that help to decide on driving under adversarial conditions. It is simple to configure such properties directly in a scenario from depth stripes, but DNNs can be considered as Part-Based

Classifiers as well, since they are able to ingest the entire RGB-D input frame and gather the necessary information to measure the advantages and/or disadvantages that there are for autonomous vehicle driving with respect to the central agents as well as non-deformable obstructions.

Since the Nvidia team and other pioneers demonstrated that DNNs could be successfully used in this field, several DNN architectures and models have been proposed for various DWS properties and/or tasks. [4] Depth sensors provide accurate depth information directly, while stereo cameras can obtain it by triangulating the light rays associated to 2D correspondences retrieved by matching the stereo pair of "images". General computer vision methods for 2D disparity maps estimation require significant computational resources and their performance is generally poor (limited depth range, outliers, etc.). For these reasons, different tasks such as the semi-dense image optical flow methods have been proposed as alternative state-of-the-art techniques. The latter methods were predominantly used for DWS as they are robust and require less computational resources (relative to the Dense SIFT-based methods), despite having limited accuracy. Several techniques have been reported for autonomous vehicle driving, such as environment understanding from maps and ad-hoc solutions implemented in simulations [10].

2.2. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) resemble the visual aspect of the human brain and they are more appropriate for applications in machine vision than other types of neural networks. As a practical aspect of autonomous platforms, CNNs have transformed the whole process of perceiving external environmental data. The key function of focusing on this method to process image perception is performing feature extraction via a convolutional layer to accomplish the desired output. The layers in CNNs are made up of a set of different filters. These filters are useful to make an efficient conversion. The convolutional layer's capability in optimizing various crucial areas of embedded vision algorithms and embedded machine vision application systems in autonomous vehicle designing is absolutely essential. The primary advantage of CNNs is that they can also achieve intelligent data interpretation. This layer ends up providing additional flexibility for image interpretation [11].

Deep Learning has become almost synonymous with Artificial Intelligence (AI) mainly due to the success achieved in the field of Image Processing. The most popular neural network

architectures for image processing are fully connected feedforward networks, convolutional neural networks (CNN), and recurrent neural networks [12] (RNN). However, CNNs have evolved as the most powerful to solve problems like image classification, image segmentation, detection, and localization, which are some of the major challenges for embedded systems vision in autonomous driving [13].

2.3. Recurrent Neural Networks (RNNs)

Zhu et al. proposed a coarse-to-fine attention mechanism based on a recurrent neural network (RNN) for vehicle reidentification by incorporating values of higher layer features in the attention scores [14]. Shen et al. proposed a multi-object tracking solution for driver assistance applications. First, they constructed the appearance-and-motion-based features for vehicle detection and tracking using a channel-wise spatial context module and a LSTM-based temporal context module, respectively.

Owing to the unique characteristics of RNN which incorporate previous information to process next input, they have widely been adopted in many sequential learning tasks and have shown good performance. Carmigniani et al. utilized a simple RNN for lateral control of vehicles and traffic speed prediction [15]. Du et al. used RNN to model the object-likeness of features at multiple levels of a deep architecture and integrated the prediction of all levels using a two-layer LSTM for vehicle re-identification.

2.4. Deep Learning Frameworks

Several deep learning frameworks have been developed for autonomous vehicle vision. The different components to be learned in an end-to-end autonomous vehicle pipeline are: localization, obstacle detection, object detection, and navigation. This high-level overview includes some pre-processing for proper scaling and format. The image data was used during developing and testing through the ROS framework. The data used can be nearly any source, but common datasets such as KITTI and Udacity's self-driving car data were used often. The large amounts of high-quality labeled data used to train has dramatically increased due to the widespread availability of Nvidia's GPUs. Perhaps the most significant development in autonomous vehicle technology over the last several years has been the extensive use of deep learning, especially Convolutional Neural Networks, a significant coloration in terms of Kitti dataset [7].

Autonomous driving technology has been steadily improving in recent decades. Volume-based solutions like motion sensors and GPS are common in forklifts and other vehicles in factories or airports. However, image-based solutions have emerged in recent years owing to the abundance of data available for localization, object detection, and obstacle avoidance without dedicated infrastructure. With the integration of deep learning technologies, high-resolution cameras, and robust control systems, the visual perception of vehicles is approaching human-level capabilities [16].

3. Data Preparation and Preprocessing

The datasets, collected using cameras and various sensors, are in unstructured format and are associated with various artifacts such as occlusions, shadows, noise, and other problems. Thus, it becomes essential to transform raw data in to more efficient format for the use of deep-learning inference at edge devices. The process may typically involve formatting images (resized to a same lengthed array) and converting grayscale images to three-channel images. Additionally, noise removal, image smoothening techniques, such as, applying a median filter, are applied to enhance the quality of the images. Finally, the last step of data pre-processing has involved class balancing or equalizing the number of pedestrian and vehicle datasets [17].

[18] Data preparation and preprocessing are the critical steps to build a high-performance deep learning-based autonomous vehicle image, and video processing ends-to-end model. In this step, the raw data, i.e., images and videos, is converted to the useful and efficient format. Generally, these steps significantly affect the effectiveness and the performance of the deep learning model. Specifically, preprocessing help in noise removal, quality improvement of videos and images and, also enhancing the performance of algorithms.

3.1. Data Collection and Annotation

In order to define the important points in the two images created as references, both images are transformed into a binary map. According to the moving upward and moving to the left with these two area definitions, x and y velocities are calculated. As a result, training data, validation data, and test data are separated using a set of motion craft pairs using random selection. In order to have balanced testing data, 1962 images are selected which include 984 images with vehicle-dependent region and 984 images with vehicle-independent region. After

all these processes were completed, our open data set, which we called "UAV-DIOR-V1-IR", consists of RGB images of vehicles, non-vehicle regions, and driven hang gliders. This data set has a resolution of 960 x 540 and 1400 different unique aircraft. These are the redescriptions in this data set. [19].

Vehicles, pedestrians, and construction zones in the initial images are labeled manually using LabelImg [20]. Images are annotated by drawing bounding boxes around instances of the vehicle category, and a corresponding text file containing information about the bounding box information and the class in the YOLO (You Only Look Once) format is saved in the same folder as the image [21]. Decompression and column operations are performed on this txt file to be compatible with the encoding by the area recommender of the network. Then, in the vehicle definition, an area with a width equal to 1/4 of the height around the label point is determined. A region definition is made according to this area. This region is added to the first 50% reference. A reference for the second region is created using Lucas Kanade method, which finds important things by tracking interesting points in the images.

3.2. Data Augmentation

Data augmentation (DA) methods have been widely used to improve the performance of deep learning systems by generating a variety of realistic transformed images. Synthetic DA methods have been demonstrated to be effective for bridging the gap between different domains. Specifically, RenderMe and PseudoPhotorealism both render synthetic images and add them to the training set to improve performance on the target dataset. The authors argue that these are less effective than removing the differences between the synthetic renderings and the real data. A concession for this due to computational time. To address these limitations, the authors have employed novel sim2real synthetic images that have the appearance of Unity Engine [22].

[23] Semantic segmentation has brought about the accurate and efficient driving decisions for AVs. It is the fundamental nerve centre that ensures deep learning based AVs response to the traffic environment and interacts with pedestrians, other vehicles and static objects. Deep semantic segmentation performs excellently but attains weaker generalization. By targeting real-world AV cameras, real dataset distribution is inevitable to contain typical data bias. Therefore, the corresponding deep learning models exhibit high vulnerability to a variety of domain shift environments [24]. Besides, the models trained on a single dataset generally

demonstrate clear performance deteriorations when tested on a different dataset since real dataset bias usually brings about discriminative visual features or the reliability of the vision tasks is overestimated.

3.3. Normalization and Standardization

In a machine learning perspective, we need to make sure that the input to the learning algorithm is well prepared and is stationary. This means the input data should have a specific distribution that can help it to converge faster during learning [25]. Once the inappropriate pixel values of an input image are adjusted, it will lead a network to learn faster due to well-prepared inputs. In contrast to other normalization types, a PCA whitening step is also included in the study to investigate effect of addi-tional removal of linear correlation between image channels in the study. Adjusted validation accuracy values suggested that z-score standardization provides best performance and its performance is followed by PCA whitened standard-ization step. Standardization is used to do feature-wise zero-center and divide by the standard deviation. So the learned filters are directly responsible to represent meaningful information in the input images.

When we talk about machine learning and neural networks, simply feeding input in the neuron layers is not enough. We need to preprocess these inputs [26]. When we talk about images inputs, these inputs are high dimensional and simply feeding them to the network is not effective. So they need to be converted back to a more learnable domain for network [20]. Normalization and standardization are generally used for this purpose. Standardization is used to make all input images in a same shape by resizing them, and Normalization can ensure input pixel data has a normalized distribution for improved training phase convergence.

4. Image Processing Techniques

[4] [27]The image processing techniques used in deep learning for autonomous vehicle image processing mainly consist of object detection, object classification, semantic segmentation, and depth estimation. For object detection, the goal is to detect the existence of certain types of objects within the input image or each frame of the video of a sequence. The object detection issues have been categorized as two types: detection with and without a given or constant observation distance. When simply detecting if one or more potential obstacles could be

detected from a 3D camera image, the distance of the obstacles can be simply estimated if the relation between the true physical dimensions of the objects and the object dimension in the camera image should be obtained by means of extrinsic and intrinsic camera calibration parameters. In the case of further detection and depth estimation of obstacles, by measuring the pixel dispersions and depth estimations of objects can be regarded as essential features for object classification of obstacles. There are four possible states of the pixel being occupied by an obstacle or being free, and this is why the classification for each state can be treated as a binary hypothesis test problem.[1]Segmentation of the input image is considered to be the most fundamental level for autonomous vehicle and robotic applications so that it has many important functions, such as for environment and object recognition or distance estimation. Therefore, many researchers have made an effort to develop different kinds of separation techniques. Semantic image segmentation examined by [article main idea="779aa6df-0883-41f7-82aa-3893d7c0eba2"] separates every pixel with an identical object class. Instance image segmentation separates every pixel into an individual class that belongs to individual instances. Real-time image segmentation puts forward the problem of detecting all environment or mobile objects using a single necessary camera image but not necessarily high expense stereo side image pairs with a current single step processing method. An evaluation of object detection and instance and semantic segmentation algorithms with some analysis and quantitative performance evaluation results have been presented to provide better insight and a better understanding about the potential of current deep learning models.

4.1. Edge Detection

Lane detection is crucial for accurate localization, route planning, and object detection in autonomous vehicles (AVs). In this paper, a Deep-Learning-based approach has been proposed and implemented in a real-time Lane Detection system, which has proved to outperforms a traditional rule-based decision method.. Some logical steps are included in preprocessing: chromatic absorption, edge and channel processing, normalization, and initial data system recognition. For edge detection and enhancement, the combined canny algorithm provides an effective solution. Different approaches like Sobel filter operators and Canny edge detection technique were studied and its comparative analysis were discussed. Canny edge detection technique was implemented and convoluted with an integral Kernal to further compact and segregate a combined edge map. This developed low-level edge detection was regionally partitioned to detect lane markings. In order to know what edges in the image

actually correspond to lane markings, we can reduce the search space only to the region where the lanes could be present, instead of searching for the lanes all over the image. It is called the Region of Interest (ROI). Canny edge detection is straightforward algorithm, that tracks intensity gradient in an image. Since it is not direction dependent, it cannot choose which direction the detected lane lies, or whether it would actually lie, in the ROI or not. Hence, the subsequent processes and techniques are proposed and discussed in results to enhance the lane markings domain in the Canny output-edge map.

Edge detection is one of the most fundamental and important areas in the field of Computer Vision. In this area, edge detection algorithms are used to identify the edges of objects in an image which deliver key information to the processing systems,"lanes" in lane detection respectively, for decision making during the movement of an autonomous vehicle (AV) . The lane detection system being a critical input to an AV's decision making, a robust and reliable approach is necessary.

4.2. Object Detection

In computer vision, object tracking is the problem of following which object/instance the object had recognized in the frame of origin of the object by finding it in the subsequent frames. Marking the detection and updating process in the map creates the map for the real-time sense data which is accessible by the driving agent. In spite of using blurred smaller detection boxes causes a real-time sense data map also to be updated with an enormously variation. A deep learning model like FairMOT (multiple objects tracking) which is applied to autonomous driving systems predicts detections (bounding boxes) from frames being updated according to the different level of object density in existing map and provides results with better accuracy and faster processing time. When detected items of the frame are separated into individual person cars and trucks, complete motion data of the detected vehicles is detected with appropriate width and height details in the class information of the corresponding text file. Motion prediction algorithm can help to resolve motion prediction and section features of urban IP containers transportation vehicles so robotics can make good predictions. Various detection models are implemented to improve the accuracy of object detection on fresh networks for FPN. The maps can be updated without correction with detected object denoting and false object actions/frames are detected. If the map correction is

selected for the training of the YOLOv5s-based motion prediction algorithm, the best motion prediction algorithm is acquired.

[7] [28] Having a precise understanding of what are the categories of objects (car, motorbike, bicycle, truck, bus, …) that are likely to be encountered in the environment by the sensors on the vehicle is a crucial task in the traffic-agnostic driving paradigm. This information is acquired from cities in which the target vehicles are driven, then annotated with labels and bounding box coordinates, according to the objects, and finally stored in a database. This database is searched to determine the category of objects during the driving process. In this context, deep learning, one of the most important techniques in the learning-based approach for visual perception on autonomous vehicles, is very successful when applied for object detection. In the detection and classification of the object in the image, deep learning models like YOLOv5, Faster-RCNN and FPN have been successfully used. On network update, YOLOv5s (YOLOv5-small) is enough to determine lighter vehicles. However, since the weights of the objects detected with it are less influential, both their sizes and weights can be underestimated in terms of motion prediction. When the Faster-RCNN network is updated, the motion map they generate does not make sense when the object separation is inadequate with the large object detection threshold value, and if it environmental density increases, all object separation may be incorrectly detected. Still, in order to correct the uncertainties mentioned above, it has to be used correct feature levels for object group separation as it is demonstrated. The Faster R-CNN has a two-stage model: the first stage called the region proposal network (RPN), which proposes a region from an input image and the second stage acts as a classification to classify and refine the proposed region. One of the most important features of the Faster R-CNN is that it has a capacity to deal with false detections and over-detections which might be a major source of traffic accidents in the future. At present it also offers good performance in terms of speed of detection and ability to distinguish small differences in objects, while working very effectively on small moving objects with a high detection ability provided by the model. The FPN can merge feature maps that come from different scales and feature levels, and use them to generate different density feature levels, and extract deep-lower level, the complex structure and the highly rich characteristics of the radials.

4.3. Semantic Segmentation

Artificial intelligence (AI) -based image processing, particularly deep learning systems, has seen significant advancements in various real-life applications including autonomous vehicles. AI techniques, such as CNNs, play an essential role in these successful applications, mainly because of their ability to interact with real-time situations. Their ability to model high-level abstractions of data using architectures that consist of multiple non-linear transformations and train them using large-scale distributed datasets has made CNNs quite robust in processing and utilising images for predicting near-futuristic scenarios, hence playing vital roles in real-world applications. The various applications of AI-based image processing using deep learning architectures in practical and realistic problems are found in the fields of autonomous and connected vehicles [7]. Deep-learning is not a drastically new concept: it maps data to the output class using complex transfer functions. Face alignment, pedestrian detection, or head-and-tail segmentation have been explored using various deep-learning models. Especially, for the automotive application domain, deep learning has proved to be of great worth in terms of obstacle detection, ego localization, or viability prediction [4].

The semantic segmentation process, also related to understanding the scene, concerns the subdivision of a digital image into multiple segments to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse. This allows for the selection of a subset of points, called superpixels, with the same color information. These can be further combined into coifiable units. Superpixels work well with an object-oriented coding system, one of the crucial reasons for using them. Associating superpixels from the image with one semantic class delivers a semantic image [9]. Deep learning enables fast, often real-time, semantic segmentation calculations and over the recent years, CNNs have become the backbone for this task. With the introduction of AlexNet, CNN has emerged as a mature technology for semantic segmentation in digital imaging, and various CNN architectures have since been developed specifically to handle this particular task of computer vision.

5. Video Processing Techniques

Behavior Modelling and Segmentation- Clustering methods like K-means classify segments of the testing data into categories. TensorFlow allows the agents to learn optimal policies exploiting transitions across the video segment boundaries. We have also implemented open-

loop and closed loop imitation learning models to easily exploit already trained behaviors of the system resulting in enhanced driver performances.

Driving agents perspective: the full frame is fed into a shared denoising autoencoder network ,which tries to denoise the input and decodes it as the saturated version of the image. This image is also passed to the 3d convolution blocks to finally infer the optimal steering and speed commands. In addition, the full frame information is also passed to a 2d convolution with softmax layer which try to predict the agents next steering and speed command to calculate the costs of the commands.

Object detection and tracking – We use a dedicated RPN to accurately detect and locate pedestrians and vehicles in each frame. Fast Moving Vehicles in the driving agents perspective are followed- simply tracking mechanisms

Frame level preprocessing – In our work it involves, a rotation (if frontal camera is facing left side), grayscaling, normalization of pixel values, and appropriate croppings to feed image to the deep learning networks. Multiple frames are also stacked and fed to the deep networks to account for motion.

[7], [2]This last section explains the video processing techniques necessary for an autonomous vehicle to efficiently derive meaningful interpretation from the visual data. 3+ years: In order to modularize the autonomous driver codebase, first it is necessary to image preprocessing techniques to extract useful information from the visual data streams. Then, object detection and tracking techniques are necessary to interpret the visual data in order to make the vehicle keep safe distance from other agents. seperated. Lastly, scene utilization techniques tries to understand the overall key information present in the visual data and help the driving agent's decision making process. Humans are able to perform complex video processing tasks seamlessly. This section explains various techniques that are necessary for an autonomous vehicle to efficiently interpret visual data in a meaningful manner.

5.1. Optical Flow

[29] Optical flow between consecutive frames provides a depth-free, pixel-wise motion field description of the scene. In the context of the automotive domain, optical flow serves as a fundamental cue for various tasks, such as pedestrian detection, satisfaction of motion consistency for moving object detection, accumulating motion magnitudes over time to infer

depth, and scene understanding using instant motion dynamics. Similar to single image methods, the transition from traditional handcrafted feature extraction methods to learned ones has had a major role in enhancing optical flow estimation. For the first time, Sun et al. learn a CNN for optical flow estimation (FlowNet [30]). Later, Dosovitskiy et al. propose a better architecture, which is larger and has a different connectivity structure. On the other hand, Ilg et al. introduce the first multi-task model that jointly learns inherent tasks, such as optical flow, disparity estimation and scene flow. Liuyuan Deng et al. also focus on the optical flow estimation of multiple objects using the FlowNet but refine it to eliminate target-dependent and object-unrelated information. Garg et al. take all shapes and solutions into consideration using a pipeline of models to generate synthetic data based on real flow, but results are ill. Several methods have tried to employ explicit or implicit motion segmentation to improve flow estimation. Sudipta N. Sinha et al. use unsupervised spatiotemporal feature learning for motion segmentation and flow estimation by decoupling it from camera ego-motion estimation. Friedrich Fraundorfer et al. focus on image pyramids for the FlowNet to robustly estimate optical flow in an incremental manner. Finally, Jaechul Kim et al. directly extract objects' motion features for focus-based optical flow. All the aforementioned methods consider the complexity of the optical flow problem assuming freely moving objects without any specific type, whereas the optical flow is perceptually dominated by rigid, moving objects for vehicle-centric, autonomous applications.

5.2. Video Classification

The key insight is to identify the benefits and limitations of using CNN-based perception methods in autonomous vehicles through real-time scenario based studies using official software (Torcs), Hardware-In-the-loop (HIL) simulators and cloud based applications like Carla. By considering CNNs implementation with different objective functions to classify frame per second (fps) and pixels using pointwise neural network (PNN) with a constant clockwise and anti-clockwise inertia for different fps is proposed. The accuracy of CNN- based perception is high when high level of different vehicle outside objects is present with more labelled samples in different skylights and cluttered conditions [6].

Deep Learning methods like Convolutional Neural Networks (CNNs) have shown state-of-the-art results with applications in image recognition and video analysis, especially in the domain of autonomous vehicles [11]. Methods to deploy CNNs in the architecture of

autonomous vehicles, and scenario based studies of challenges and advantages of deep CNNs over conventional methods for autonomous vehicles are mentioned in this section. With CNNs, real time visual perception is possible onboard the vehicle so that it can predict the steering angle by detecting turns and curve directions in continuous video frames [15].

5.3. Action Recognition

Action recognition of surrounding vehicles plays a pivotal role in predicting future movements, especially lane change events, which are generally difficult to be identified due to the fast displacement characteristics of traffic scenes. Hence, it is of significance to model the dynamical action information fully for the driving vehicles, which is important for forecasting future road conditions to support self-driving cars with the ability of decision making and path planning. Recently, the action recognition task in computer vision has shifted to complex long-term history sequences. Despite impressive improvements in the visual detectors and the temporal action recognizers, the recognition of long-term dynamic elaboration for the invisible context still remains a big challenge. Therefore, obtaining effective context representation that could capture the comprehensive spatial and temporal context for the interested action is essential, particularly in the anticipatory lane change prediction. [31]

The increasing demand for autonomous vehicles has led to a widespread discussion of technologies supporting their autonomous environment perception, control, and decision making, with deep learning as the most commonly used technology in this area. This chapter focuses on deep learning techniques used for autonomous vehicle image and video processing, and summarizes deep learning-based object detection, object recognition, depth estimation, object tracking, semantic segmentation, and action recognition [32].

6. Deep Learning Models for Autonomous Vehicles

Recent advances have introduced various probabilistic neural network to classify road signs in production only. cars. Additionally it has been shown that the rate of neural network models becomes much slower due to increasing data classes. A tracking mechanism based on deep learning is well studied. In this contribution it will be shown that the neural network approach is also well suited for trajectory prediction along with the subsequent probability estimation of 3D curves and surfaces based on the Long Short-Term Memory (LSTM) model.

Our awareness for lane and traffic sign recognition directly uses a CNN-based real-time tracking system.

The second sample idea of this contribution is the application of state-of-the-art object recognition models from the computer vision literature to production embedded systems. A traffic sign classier can recognize over 25 different signs, ranging from not only speed limits, parking and stopping signs over to hazardous situation warnings, but also to specify markings for the controlling of trac light positions. The last scenario in this paper concerns doubtful cases in real-time detection. This study shows a model embedded development for an automotive dataset which was used in the area of in-car vision. [2] Autonomous vehicles and self-learning robots are essential topics of the ongoing research. Computation intelligence, especially deep learning with neural networks, has had a great impact on these systems. A recurrent model proposed by Lipton et al. (2015) is known for predicting time series. The network generally is able to detect and recognize lanes with up to 95% accuracy, because of detailed information in frame-by-frame or sequential image streams. All our achieved results show that recurrent networks are able to build a dense lane model in the short term from the road cockpit view with low-cost sensors.

[33] Recent samples of using deep learning models for end-to-end image processing in typical visual modules in an autonomous vehicle are provided in this case of study. Deep learning techniques are nowadays extensively used in many disciplines; the automotive field is no exception. They can substantially increase the processing speed of an autonomous vehicle and significantly decrease the supplier development time. This chapter focuses on the use of long short-term memory (LSTM) neural networks for lane departure warning, traffic sign recognition as well as an in-car vision system, and on the probability estimation of 3D curves and surfaces. The first sample of using recurrent models for road situation analysis in autonomous cars is a lane detection system. The principal idea of this method is to find the most probable real-world states satisfying distance and angle constraints from given camera states at the current states, Our system estimates the probability density of street lines based on different orientation features in the recorded data by means of a supervised learning algorithm with a trained LSTM network.

6.1. End-to-End Learning

Instead of fixing the end-to-end solution to directly predict the six degrees of freedom (6 DoF) trajectory from a camera sensor, we propose to enrich the raw image with additional, highly informative, and interpretable visual features. These features are generated using well-known image processing methods from the area of automotive computer vision. The main advantage of this approach is that the generated visual features can be easily interpreted and controlled by the engineers in order to guarantee, for example, that the correct information was fed to the network. This significantly improves the trust in the predictor system and hence its reliability and interpretability. Furthermore, the generated visual features can also be used for a dedicated analysis of the prediction results, which can be addressed in specific post-processing tasks.

Using machine learning methods to predict control signals in autonomous driving has a well-established background. This includes using Deep Learning methods such as Convolutional Neural Networks for predicting vehicle controls. Yet, the majority of works assume that the visual features extracted from the camera feed are directly fed to the network input. At the same time, manipulating images for human interpretation prior to feeding the network is a common practice in the automotive industry. Just to give an example, high-dynamic-range (HDR) processing and gamma equalization are applied at different stages of computation and visualization onboard advanced driver assistance system (ADAS) vehicles. Surround view systems are an example that includes a number of image manipulation methods ranging from perspective transformations to put camera images into a common ground plane to fish-eye-lens compensation to image stitching. [34]

[1]

6.2. Behavioral Cloning

Some previous work directly generate those rare simulated query scenarios randomly or by imitation learning, most of them can improve the model performances noticeably (Dosovitskiy et al., 2017; Jain et al., 2019; Ha and Schmidhuber, 2018; Osendorfer et al., 2016). But the backward problem still occurs in them: there are still potentially risky events or scenarios in the real driving datasets which are not covered by the learned behaviours. Many previous works aim to solve this problem offlined in the imitaion learning framework (Ross et al., 2010; Piot et al., 2014; Igl et al., 2020). But such approach is still considered to be suboptimal because it may need to collect a considerable amount of expert-picked well-

behaved data (Avtov, 2021), which can be difficult and time-consuming to obtain. In actual scenarios, we suspect that almost all training data are dirty. Because AdaBEL/BF is directly trained with the important partitions instead of dirty data, it almost avoids the backward problem and keeps high score on average. That is the reason why we get the observation 1.

[35] When {}\mathrm{CL}_{\mathrm{Dagger}} is used, the model uses the same minimization objective, but in the experiments it tries to collapse from the simulated query scenarios in order to obtain an evenly spread knowledge that can be applicable to the real test scenarios. We use the training splitting protocol in Figure 2 and in Section B.2. It reduces the average driving penalty by 75% on real world data. We also give implementations with optimizations and detailed experiences of training in Section 5. Even with the optimizations like curriculum learning (Bengio et al., 2009) and exploration schedule (Osband et al., 2016b), training neural network controllers in environments with hundreds of rare yet critically significant situations is still a challenge with reinforcement learning algorithms (Zhao et al., 2021; Jin et al., 2020).

6.3. Reinforcement Learning

Therefore, we introduce a deep reinforcement learning framework Dyna-Q for autonomous driving tasks. In Dyna-Q, we consider another component of the agent network, and we replace the temporal difference error with the predictions of target and online networks. The purpose of reward function is to encourage the vehicle to drive safely, in a straight line, and at a constant average speed. In addition, in contrast with traditional DRL approach, we do not make the trade-off between safety and comfort, and the weighting for different behaviors of the agent are obtained dynamically [36].

Deep reinforcement learning (DRL) has been successfully applied to intelligent vehicle navigation due to its superior performance in complex tasks. Deep Q-Network is a popular method in DRL where the action is chosen from a separate target and online network. But negative experience bias may occur since the target is updated with the same network that is used to derive it. When the policy diverges a cascade of errors may occur as new repeated experiences are all duplicated and added to the replay buffer [37].

7. Challenges and Limitations

[38] However, the authors acknowledge the instability, heavy training data and infrastructure requirements, ethical considerations, and flawed training-based learning in AI and machine learning systems, the opacity of AI systems, the challenges of vehicle routing and form-to-function interaction, and how the aforementioned limitations of machine learning-based AI solutions can pose potential commercial, legal, and privacy issues and damage consumer trust as serious technological and ethical challenges of AI applications. The same authors and others also pointed to issues of vehicular cybersecurity, system defenses, and blockchain-based solutions, especially when it comes to connected, autonomous, and electric vehicle (CAV) security [arg: 16b92040-a1ad-44c8-9180-42de851bb8bb]; in particular, LiDARs are reported to be fooled by only a few photons of maliciously created light that force an AI-assisted vehicle to treat a potentially fatal traffic hazard as merely a harmless object, thus making CAV systems vulnerable to tricky adversarial attacks, and vehicle-to-infrastructure (V2I) protocols are exposed to third-party information leakage [arg: 181098fd-d8a7-41ec-89df-4a605083f958].[39] Achieving an 100% feasible vision-based solution that takes into account the potential future loss of feature extraction effectiveness, the highly non-linear relation between vehicle movement and acquired distractor-free image distribution, the high processing costs (high-energy consumption, high response time, and high real-time implementation costs), the simultaneous deployment of distractor removal network and object detection and segmentation algorithms in one system, the possibility of the distractor removal network itself becoming distractor-prone, the requirement for a convolutional filter with a much larger net capacity in single-shot based localization approaches, the only use of bounding box and orientation information, the real-time requirements and architectural difficulties of implementing image-based self-localization algorithms, occlusion issues that prevent the system from finding the proper matching between the main ego-vehicle and target object, the recent tendency of developing algorithms that do not take into account the feature context and the static information in sequential images, distillation approaches that cut down the classification task, only use of LiDAR, radar, ultrasounds, inertial measurement units, and global positioning system aided systems, the absence of certainty information in most non-LiDAR based systems, the requirement of a high-precision map, and the need for efficient edge processing and convolutive separability, among others. All of these have to still be researched first to make object detection, classification, recognition, image segmentation, scene understanding, trajectory estimation, navigating, sensor fusion, safety, ethics and privacy, and adas, and a vs, and FOTA, and factory, and battery, and supply-chain, and mtc,

and keypad, and jt, and M2M, and low, and packet, and V2V, and smart, and fo-g, and co-an, and cl, and tc, and cnn, clipboard, and basic, and aeeseueepr, and cve, and t, and h, and AI, and rvrscdcwabotffncdp, all tasks efficiently and effectively handled by AI, ML, DL, graphs, h-d-e-es, wsns, automotive and industrial wireless communications and data-analytically empowered, and convolutive, non-linearity, non-stationarity, sequence-based, periodically and aperiodically over neural-networked, as the robotic-ready and autonomous vehicles' vision and audio-visual algorithm pandemonium is simply too much complex for N2N, N2VN, N2SN, NPDR, M, SDR, and PDCN, among all-Ns and other algorithms and systems to solve.

7.1. Data Quality and Quantity

In addition to the quality, on account of the temporal and spatial dynamics of a frame, the quantity of frames in addition to the variety of scenarios need to be considered. In autonomous driving datasets, the vast majority of labelled frames have been similarly distributed due to consideration of the relevancy of modern traffic scenes. Consequently, models training on these kinds of datasets learn different routines with different effectiveness. Regulation is needed to prevent a situation where a higher number of unsafe scenes appear within training data and the model learns become imprecise. The behavior of training sequence distributions may be decoded by mistakes happening due to lack of information to tackle specific difficulties. For instance, frames that model has never seen while training may cause it to make mistakes leading to accidents. However, on account of data from differentiating incident conditions by dint of using Vulkan Ray Tracing, it is ensured that the trained model can maintain age in its perception to observe new patterns overall.

Deep learning-based computer vision models rely on large amounts of labeled data and, accordingly, the quality of the data, as well as the number of available samples, is crucial. However, in real-world problems like autonomous driving, labeled data is expensive and challenging to acquire [1]. To solve this challenge, that is, to acquire data with correct quality and consistency, distinct collection mechanisms as well as simulators have been developed. For instance, [40] trained an end-to-end autonomous driving model by employing simulated data, as well as real-life data. Alongside this, crowd-sourcing of large-scale datasets, like BDD100K, apolloScape, and KITTI, make observing general patterns of perception tasks

feasible. However, there still exist concerns about whether these openly accessible datasets contain enough general patterns to improve realworld autonomous driving systems.

7.2. Interpretability and Explainability

Whereas these kinds of methods seem to centre on a convenient approach to gainfully display model interpretability, interpretable models present a more proper way to explicitly embody explanations into the model. [41] In this context, when a model compounds these two facets, it shall be called entirely interpretable. This would be able to provide robust insights about both its working mechanism and its predictions. Hence, a partial and simple machine learning system, such as decision trees, would be the potential interpretation of a predictive model. On the other hand, a non-interpretable approach would imply having a complex and non-transparent system such as the neural network. At this juncture, the very pragmatic next aspect is to hybrid these two elements of the journey.

Interpretability represents the capability to deduce valuable information from a model, thus ensuring the delivery of a transparent explanation about its internal mechanisms. Explainability is the ability to provide high-level interpretation of the prediction enhancing the transparent clarify of the internal decision-making process. [42] Among the wide variety of work dealing with explanation methods of deep learning models, those oriented towards images are the most relevant. Seeing that driving decisions are made based on camera sensors, developing explicability techniques in the automotive sector has a significant potential. Visual melting of the areas during the analysis of the reasons behind a model's predictions constitutes an ideal complement to the different regards. This technique favours the extraction of visual supporting evidence for making a specific decision.

7.3. Adversarial Attacks

To test a real-world DNN model that is used for controlling an autonomous vehicle—specifically, an NVIDIA Jetson AGX Xavier platform for lateral and longitudinal control—we investigated the robustness of the resulting perception and control models to a range of physical adversarial attacks [43]. Physical adversarial attacks, which include sticker attacks, recognized shortly after the work of Eykholt et al. , and raindrop attacks , have already shown that many DNN-based perception and control systems can easily be outmaneuvered by a range of small but noticeable input distortions [44].

Although the deep neural networks (DNNs) have the capability to recognize real objects accurately in real-world scenarios, they are vulnerable to adversarial perturbations_Law1 caused by small input changes. AdFrob is the first end-to-end evaluation framework for sensor-based DNN perception in automated vehicles to measure and improve the robustness of sensor-based DNN perception in automated vehicles under real-world conditions [45]. Autonomous vehicles (AVs) are becoming increasingly popular, and they are expected to revolutionize the transport sector. However, the reliance on DNN-based perception and control algorithms brings new security and safety pitfalls, as these DNNs can be easily manipulated like other types of neural networks.

8. Case Studies and Applications

Finally, benchmark prototypes can easily be developed to characterize performance. All this development promises a swift transfer of Deep Learning AI-trained models to a large number of such INA-chip capable platforms. To discover how shock-resistant real automotive AI training data can be, the project team measured their capacity for generalizing to foreign vehicles and road condition mosaics. All subject video mosaics were AI-trainset-cloned with disturbing state-of-the-art deep convnets and then subjected to various new TVT101 and TVT102 tests to measure successful transfer from CHIL Urban, CHIL Highways and the CHIL Town track to other CHIL tracks [1].

A significant drive towards autonomous systems have led to the development and release of low-cost chips for AI in Autonomous Vehicles. As for the Amazon developed AWS NeoAI guides for efficient porting of trained models to the most relevant degree of reduced precision (e.g. 8-bit for Conv2D neural network layers), a recent Horizon 2020 EU project, CrowdLabs, has released porting placements from a variety of different academic institutions. A second release, INAT, pertains to quantization-aware training to simplify porting while maintaining final model accuracy [6].

8.1. Tesla Autopilot

Paradoxically, a new market segment will arise, like the leisure industry; it will become more attractive to visit places without having to worry about reaching the destination safely. A central computing eco-system may take over physical hardware in the advance autonomous from inside and outside the vehicle.

This environmental-friendly mobility with various advantages brings a lot of side effects. It will change entire transportation infrastructure in general, increasing or reshaping public parking spaces which require less cars, and affect other industries like insurance businesses, drivers (taxi and bus drivers will lose their job), voracious car parking, fines, EMI, some industries will be ablated such as petrol business, service stations or re-modeled in a different way such as workshops.

In the future it will wreck the car ownership. On a long-term horizon, most of the population will prefer advanced autonomous shared vehicles to car ownership. Many autonomous shared vehicles will replace car ownership.

Soon the Autopilot from Tesla will further enhance drivers' capabilities. It will offer GPS, sensors, machine learning mobile cocoon, and other connected vehicle services. [15] Neural networks and deep learning have been integrated into autonomous vehicles to simplify the complex decision-making task by predicting future paths. This emerging technology will reduce delight in driving. In future days, ML technology will take over the prevalent road user by achieving 85% (migration to advanced autonomous) market share of SV (shared vehicle). Shared advanced autonomous will mainly be used for ride-sharing or smart bus, temporarily individual transportation as taxi alternative for specific purposes, like family lunches, bad weather, urgent time and delivering fresh food, for deputies.

[1] Tesla is known for its innovation both in autonomous vehicles and with its technology. The electrical manufacturer was the first company to roll out lane-change support and parking assist features on the highway. Tesla has been developing its autopilot hardware for a number of years and strongly believes that it will soon achieve full selfdriving. Its main ASP is that the Tesla camera and sensor technology is more than just a system that provides the driver with additional driving comfort; it is intended to be the main sensor for autonomous driving, thus fulfilling all safety-critical tasks. Other manufacturers such as Mercedes, Toyota, Honda, and Lexus are the premium car companies known for developing a traditional approach using LIDAR, camera technology, and other sensors to establish their HMI and communication with a classical OBD adapter. However, as simple vehicles do not furnish their OBD sensors with enough data to implement advanced driver assistance eco-systems, they are available only in newer vehicles. Promising accuracy rates in the domain of LIDAR technology are invisible deep or up to 200 m ahead of the vehicle, despite suboptimal visibility, light, or weather.

8.2. Waymo's Self-Driving Cars

[46] Waymo is an autonomous vehicle operating company of Alphabet Inc. (the parent of Google), known for its numerous initiatives for efficient deployment of its fully autonomous vehicles (Fridman and Dai 2020). The dataset released by Waymo has played a significant role in current advancements of autonomous vehicle deployment. The Waymo dataset consists of labeled data collected by Waymo self-driving cars and has over 10 million miles (16 million kilometers) with 25 cities, which is the largest dataset at the time of writing these lines. The data consists of Lidar point clouds, vision images, and meticulous annotations for vehicles, pedestrians, cyclists, and signs. There are more than 12 million 3D annotations in the training set and 1.2 million 2D annotations in the training set, which is larger than the largest publicly available dataset in 2020 (Geiger et al. 2012). This dataset has rendered end-to-end learning invaluable to Autonomous Vehicle Image and Video Processing community.[47] Autonomous vehicles is an upcoming technology that holds the promise of enhancing road transportation efficiency, safety, and accessibility. However, to date, in practice, there are still some significant challenges and concerns to be addressed. To help the automotive and AI (Artificial Intelligence) communities further understand the potential of AI techniques in autonomous vehicle safety, this paper provides a systematic literature review on the impact of AI on autonomous vehicle safety. The study reveals two dominant research lines: Those in the intelligent perception field are focused on building pedestrian detection, lane detection, object recognition, turn signal recognition, sign recognition, road marking recognition models, etc.; while those in end-to-end learning are mainly devoted to building models to map images/maps/videos to the direction values/velocity values/gas brake values and steering angle values that the autonomous vehicle should move towards (Li et al. 2021).

8.3. Uber ATG

[48] Uber ATG constitutes one of the largest players in AV research. The SurroundView dataset was introduced by its researchers to train models for estimating the distances to objects in the scene. In addition to object detection tasks, the dataset could be used for depth estimation, and the development of Advanced Driver Assistance Systems (ADAS) in general. It consists of over 150k images from 1440x810px or 1440x1280px 360° panoramic cameras (mounted in the car front), which give a high-voltage range between 6-80 volt in a resolution of 160x90 depth pixels.[5] Moreover, its high dynamic range introduces an ideal benchmark

for further high-dynamic-range imaging tasks at this quality, e.g., style transfer. The images contain the actual object orientation and are densely annotated with ego-motion—every key frame is provided with the images taken from six base-mounted cameras in 720x540 RGB resolution, as well. Note that this work by Uber determined the architectural basis for Bosch´s neural network architecture AV Laboratory Neural Network (AVarCN), featuring a base task Aggregated inference from box classification and distance.

9. Ethical and Safety Considerations

On the other hand, negative externalities created by TEVs must be considered. An immediate social concern is that as the number of autonomous vehicles increases, these systems will lead to the feared scenario of mass-unemployment and mental health issues (e.g. mental stress due to unemployment among transport service workers), unfair moral decisions, and a reduced understanding of human capabilities. In general, researchers need to focus more on the social implications of TEVs, so that the innovation of autonomous vehicles can fully integrate humans into the interaction processes [7].

To safely and responsibly deploy Deep Learning in the automotive industry, it is fundamental to address legal, ethical, and social aspects of these technology-enabled vehicles (TEVs). In the EU, TEVs are expected to adhere to the recently released regulatory framework on data protection, explained by [1]. Moreover, it is important to consider the moral implications of using Deep Learning to make decisions and to implement correct decision-making models. Interpretability is one promising approach to enhance the decision-making process by explaining the intermediate behavior of a Deep Learning model to decision-makers and users, thereby promoting transparency and trust in automated systems. This approach is useful for revealing potential defects in the adopted models and could contribute towards the optimization of decision-making using omission of non-essential features. Since there is no comprehensive visual attention mapping method for autonomous vehicles on the market, reliability and explainability remain important themes in this research [49]. Moreover, from a broader ethical perspective, the application of Deep Learning in TEVs can help individuals pay better attention to road conditions, reduce driver workload, reduce fuel consumption, increase road capacity, reduce the need to build new roads, and enhance traffic flow levels by decreasing passenger travel time, traffic accidents and congestion.

9.1. Bias and Fairness

In my deep learning works on autonomous driving, I have before attempted three different types of bias mitigation: on the one hand, color bias which is, specifically for urban driving, a canonically occurring issue [Just Drive]; on the other, I have made use of a possibility to identify confounding variables in visual data and to train (relatively simple) forecasting models, generating counterfactual predictions and, through these "proof-of-concept" interventions, make the respective learning models temperate with their learnings of these spurious relationships; finally, last but not least, by reweighing examples with the discrepancy substitution method [Fair Facial Attribute Classification].

The integration of deep learning into autonomous vehicle (AV) image and video processing is often troubled with unintentional, or human-induced, biases [Just Drive] [Fair Facial Attribute Classification]. These biases, sourced by biased training data, sub-optimally fit the data they were trained on and consequently make it underperform on other, unseen, data. This likely causes adverse effects in a safety-critical context such as [Just DriveHigh resolution semantic segmentation] autonomous driving, where the processing of computer vision algorithms feeds directly into action-critical decisions about the vehicle's environmental perception, its path planning and its interaction with traffic participants. Specifically, urban driving represents a paradigmatic example of a setting that is characterized by great diversity and where biased models often underperform; depending on the model can completely fail to generalize to the visual conditions accompanying driving in UK whole seasons and across varying times of the day.

9.2. Safety-Critical Systems

The original SHAP idea is prepared to work on interpretable system design yet different tasks including face recognition shalline or for the robust design of the adversarially learnt deep generative models based visual recognition [50]. Fung et al. [51] proposed an approach for safety case formulation for the AI critical system in the general form on the autonomous vehicle. A lot of ongoing research work is toward (1) providing the assistance such as continuous safety case formulation, particularly for the autonomous driving example, (2) integrating the critical safety and the artificial intelligence, particularly of the deep learning-based systems, and last but not the least, ensuring the safe verification over the safety-critical systems in the safe operations.

Deep learning (DL) and artificial intelligence (AI) are gaining significant attention over the past few years, improving safety, reliability, and performance in autonomous vehicles. A critical need for safety in autonomous vehicles is to develop AI safety standards to ensure state-of-the-art AI system performance. Supervised learning of deep neural network (DL) architectures is often difficult to manage for working well with images that do not follow the distribution of the training images [52]. Especially in the semantic scene classification where protected output, such as vehicles, roads, building, and traffic signs, at the inferencing episode, does not see a clean and smooth flow in the testing in autonomous vehicle use cases. Small proportion of the total pixels (less than 2%) would carry the significant result of the 70% of the total probability of the pixels of the interestingness for the nuScenes benchmark.

10. Future Trends and Research Directions

The perception part of the TAS might well take its inspiration from unsupervised training (like contrastive learning, CT-CLOSUP, SwAV, BYOL) and/or from human-understandable audio and speech phenomena like speech enhancement and enhancement, exemplar algorithm, hidden Markov model, convolutional LSTM, resnet DNN, CRNN, KeOps in the form of causal and acausal models,. Through these two ideas, Enhanced contrastive self-supervised learning (ECSSL) considers the contrastive alignment of all three modalities associated with vehicle. It introduces two weighted cross-modal similarity-based kinetic energy functional, terms maximizing the distance between mutually exclusive representations and minimizing the distance between transformations as well as distance integrating the two with self-supervised loss, and the singular value decomposition-based Riemannian snake wavelet-based computer-based security. It also suggests that extrinsic auditory cues are better suited to learn representations useful for converting modalities. Combined with the residual backbone epoch self-supervised encoder, this updated cascade resiliently allows the ECL filter (by which it differs from the more basic CL filter) to be augmented by residual modules to the acausal duration causal plus null modelsasterisk.

Referring to the article [7] that is declared in the context of Deep Learning for Autonomous Vehicle Image and Video Processing, the novel direct perception formulation given by Tesla through its chaotic flow-based neural network (FSD) mapping images directly to control torques, would be of fundamental significance. Additionally, the resultant knowledge on 1000 classes of objects in a large number of diversified environments (ImageNet dataset) and 3D

objects in a controlled environment (KITTI dataset) through its AlexNet and ResNet image classifiers, and FasterRCNN, YOLO, Sin_fov versioned single-stage image localization based 2.5D object detectors can be leveraged right away. However, as declared earlier the perception part of Tesla's self-driving AI is discovers some substantial shortcomings in view of the 1.43 M traffic related casualties occurred only in developed countries over the last two decades, which itself speaks to the maturity level of the perception part and hence, tackling these issues would be of prime importance.

10.1. Multi-Sensor Fusion

The vision system based on computer vision has restrictions on working in some conditions like dark roads and this might affect the performance of autonomous vehicles [53]. Low luminance, occlusions, and noises due to weather conditions affect the perception accuracy of the vision-based systems. This is why, the vision sensor cameras in autonomous vehicles are used in alternative with LIDAR sensors for additional support in these environments [54]. On the other hand, LIDAR sensors might have noise problems and these noisy data might affect the detection accuracy of the autonomous vehicle. For this reason, it is important to integrate the data from the cameras and LIDAR sensors to have a reliable and efficient perception system. In this study, the data from the camera sensor and the LIDAR sensors are integrated in feature level to help in object detection and especially to increase the feature map density. Feature maps convey important spatial information and in deep learning studies, while spatial information is important for object detection, it might be insufficient in some cases [55]. By increasing the density of the feature maps, the situations that the spatial information is insufficient or inadequate will be prevented and potential object detection failures will be eliminated.

10.2. Self-Supervised Learning

The main limitation of supervised learning is the need for large-scale labeled data. In autonomous vehicles, labeling an image or a video frame is at least as complex as the set of pairs object detection/task definition required locations/labels. It might be even more time-consuming if the labeling requires contextual understanding of the scene [48]. Given a number of training samples in some domain, self-supervised learning consists in the generation of many auxiliary tasks that automatically extract supervision signals out of the domain. These tasks use cheap annotations, resulting in feasible labels generation at scale. The cost of such

broader Label Universe with Apache Santuario includes the increase in the number of training parameters and the hyper-parameter tuning, which can be more delicate. Nevertheless, overcoming these challenges with more generic text tasks in an unsupervised domain will be, for all of these reasons, a key question for the learning community in the coming years. In addition, self-supervised learning methods are gaining in popularity. But many non-optimally selected auxiliary tasks, visual or textual, offer arbitrarily learned scenarios not sharing common mechanisms between them. It can have at best a very weak range of correlation with any specific image processing task. To make good use of all the data we have in self-supervised settings, the auxiliary tasks must be selected optimally to always derive helpful representations such as shared linear factors for different recognition tasks. Some widely used strategies to learn shared representations of corresponding features from images and words are common visual words or patch representations leading to clustering features. Contextual information of words has been useful for analysis of social media and the internet in recent years but there is no such method in the field of RV with visual context, which we call Visual Contextual Information. We propose a new method to mine visual contextual information (Cross-modal Contextual information) in the visual Encoding of the image field, and Image Region Contextual Information [56]. Cross-modal methods exploit the commonalities among different data modalities and avoid the complexity of generating or deriving new tasks or patterns. Cross-modal methods have generally been used in multi-modal learning settings, facilitating the learning process and the inference due to complementary features in each modality. Diachronic has become a very important Information Retrieval (IR) research topic due to the increasing interest in automatic phonetically-based systems. Diachronic systems are aware of the fact that spoken or written documents are not valid for the whole of their lifespan. Diachronic behaviour may be caused by changes in state-of-the-art technologies, changes in users' needs, etc. have recently seen a dramatic increase in the use of Sub-word Units (SUs) in many Natural Language Processing (NLP) tasks such as automatic speech recognition and spoken term detection. Although Crossmodal Visual and Language Model Pre-training (CLIP) initializes a single concept-based and data-independent model that is effectively performant on diverse unimodal and multimodal task spaces, the pretraining process can be applied to other domains only if the two modalities have been hypernymically related in a large-scale external textual corpus.

10.3. Edge Computing

Despite the advances in the computational resources in autonomous vehicles, the use of edge computing still has significant importance. Real-time data processing for many scenario-dependent tasks can be managed at the edge of the network (vehicle side) to meet the strict delay in perception to decision process. The real-time system at vehicle side benefits from proximity to sensors, thus reducing the impact of interconnection latency. If the autonomous vehicle system itself is not fast enough in processing the data, the on-vehicle AI module can be augmented via the edge computing systems. The challenge includes deploying edge computing resources such that the allocated communication and computation resources are best utilized at the intersections of two continuously dynamically changing topologies of sensors and connected vehicle nodes [57].

Given the stringent response time requirements from the autonomous vehicle systems, real-time computing systems are highly critical. Real-time systems processing the data collected at the vehicle's level play an important role in the decision-making process in autonomous vehicle architecture. Such systems run on-vehicle at the sensors' side and process raw video data or image streams in almost real time. They offer a low-latency acceleration for critical tasks that impacts the vehicle's safety.

[58] [59]

Reference:

1. 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.
2. Ponnusamy, Sivakumar, and Dinesh Eswararaj. "Navigating the Modernization of Legacy Applications and Data: Effective Strategies and Best Practices." Asian Journal of Research in Computer Science 16.4 (2023): 239-256.
3. Shahane, Vishal. "Security Considerations and Risk Mitigation Strategies in Multi-Tenant Serverless Computing Environments." *Internet of Things and Edge Computing Journal* 1.2 (2021): 11-28.

4. Tomar, Manish, and Vathsala Periyasamy. "Leveraging advanced analytics for reference data analysis in finance." *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)* 2.1 (2023): 128-136.
5. Abouelyazid, Mahmoud, and Chen Xiang. "Machine Learning-Assisted Approach for Fetal Health Status Prediction using Cardiotocogram Data." *International Journal of Applied Health Care Analytics* 6.4 (2021): 1-22.
6. 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.
7. 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.