Deep Learning-Based Computer Vision Methods for Complex Traffic Environments Perception: A Review

Talha Azfar[1] · Jinlong Li[2] · Hongkai Yu[2] · Ruey L. Cheu[3] · Yisheng Lv[4] · Ruimin Ke[1]

Abstract

Computer vision applications in intelligent transportation systems (ITS) and autonomous driving (AD) have gravitated towards deep neural network architectures in recent years. While performance seems to be improving on benchmark datasets, many real-world challenges are yet to be adequately considered in research. This paper conducted an extensive literature review on the applications of computer vision in ITS and AD, and discusses challenges related to data, models, and complex urban environments. The data challenges are associated with the collection and labeling of training data and its relevance to real-world conditions, bias inherent in datasets, the high volume of data needed to be processed, and privacy concerns. Deep learning (DL) models are commonly too complex for real-time processing on embedded hardware, lack explainability and generalizability, and are hard to test in real-world settings. Complex urban traffic environments have irregular lighting and occlusions, and surveillance cameras can be mounted at a variety of angles, gather dirt, and shake in the wind, while the traffic conditions are highly heterogeneous, with violation of rules and complex interactions in crowded scenarios. Some representative applications that suffer from these problems are traffic flow estimation, congestion detection, autonomous driving perception, vehicle interaction, and edge computing for practical deployment. The possible ways of dealing with the challenges are also explored while prioritizing practical deployment.

Keywords Deep learning · Intelligent Transportation systems · Computer vision · Autonomous driving · Complex traffic environment

Introduction

Video cameras have been used to monitor traffic and provide valuable information to traffic management center (TMC) operators. The manual process of having TMC operators observe numerous video screens has given way to automated and semi-automated computer vision approaches for faster processing and response times with some humans in the loop to interpret and verify the data. Artificial neural networks are being increasingly used in computer vision in ITS and autonomous driving (AD) applications, showing benefits in traffic monitoring, traffic flow estimation, incident detection, etc. However, the use of deep neural networks (DNN) brings some issues and concerns that should be studied in further detail as they need to be more accurate, more reliable, and practical enough for large-scale deployment as part of ITS infrastructure or in autonomous vehicles.

Deep learning (DL) refers to machine learning architectures of multiple layers such as large (deep) neural networks spanning many layers in a variety of configurations.

✉ Ruimin Ke
ker@rpi.edu

Talha Azfar
azfart@rpi.edu

Jinlong Li
j.li56@vikes.csuohio.edu

Hongkai Yu
h.yu19@csuohio.edu

Ruey L. Cheu
rcheu@utep.edu

Yisheng Lv
yisheng.lv@ia.ac.cn

[1] Rensselaer Polytechnic Institute, Troy, NY 12180, USA

[2] Cleveland State University, Cleveland, OH 44115, USA

[3] The University of Texas at El Paso, El Paso, TX 79968, USA

[4] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

🖄 Springer

Advances in computational hardware and algorithmic efficiency have made DL popular in every field that deals with big data. DNNs have been applied to solve computer vision problems such as object detection, classification, motion tracking, and prediction (O'Mahony et al. 2019). ITS and AD researchers have adapted these models, training them for specific use cases with specially curated datasets and benchmarks such as KITTI (Geiger et al. 2012) and AI City (Naphade et al. 2019).

Automated analysis of traffic surveillance videos in ITS systems is important for incident and congestion management, while perception in autonomous vehicles is critical for vehicle control and navigation. Computer vision algorithms used for these purposes must therefore be scrutinized in detail and all of the possible problems should be addressed in advance before real-world deployment. In the course of this literature review, a number of recurrent issues were discovered related to the data, models, and complex urban environments which are detailed in this paper. Large quantities of data are necessary for training DNNs and evaluating their performance, but this poses issues such as over-representation of common events or classes, time and effort required to label and select data, and a lack of consistent benchmarks for a fair evaluation. Complex DL models can be trained to infer more accurately but this comes at the cost of efficiency, lack

of explainability, and difficulty in adapting the solution to diverse or unseen use cases (not present in the training set). Real-world uncertainties involved in complex urban environments like shadow, lighting, and occlusion are common issues, while variable surveillance camera angles and heterogeneous traffic conditions present further challenges to DNNs even after training on these conditions.

While these issues have been mentioned in some of the literature, only a few approaches have been developed to address them, and even fewer real-world implementation examples were found. Computer vision in transportation is a very active research field, and over 200 papers were selected and reviewed for this article. Figure 1 gives an overview of the applications and challenges for quick reference, while Table 1 summarizes the methods used in each application and associated challenges. The following Sections 2, 3, 4 discuss the specific challenges for data, models, and complex traffic environments. A number of representative applications and solutions to meet the challenges are explained in Section 5. This is followed by Section 6, a collection of future directions that research in this area should take. Finally, Section 7 presents some concluding remarks.

The contributions of this paper are:

- Classification of common challenges faced by computer vision DL methods in complex traffic environments.
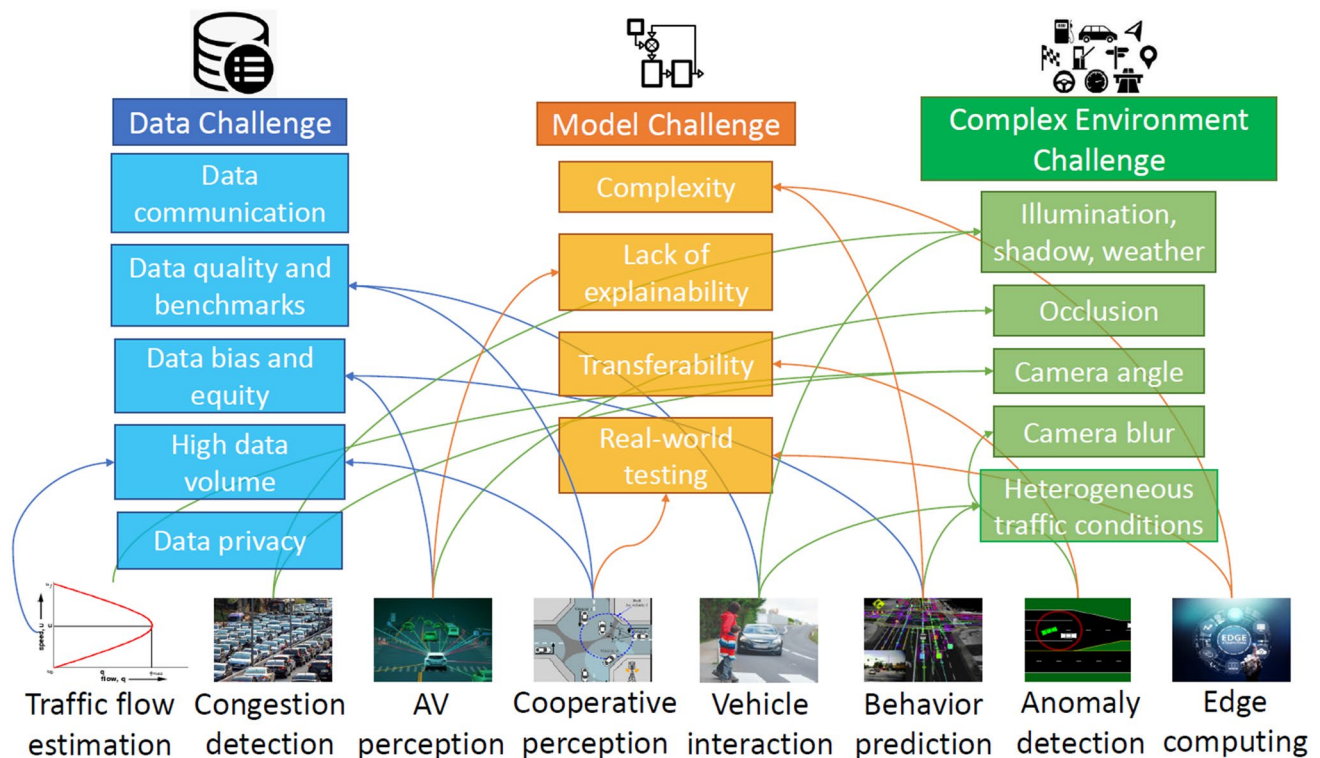
Fig. 1 Applications of deep learning-based computer vision methods in transportation and associated challenges in real-world deployment

- A review of DL models used for some representative computer vision applications susceptible to the challenges.
- Specific techniques are already being used to mitigate the challenges.
- Future directions of research to improve DL models for real world complex traffic environments.

Data Challenges

Data Communication

Data communication, while not considered in most ITS and AV computer vision studies in the lab, is critical in practical applications. Individual-camera-based deep learning tasks in practice commonly require data communication between the camera and the cloud server at TMC. Video data entails greater network utilization, which can cause potential data communication issues, such as transmission delay and package loss. In a cooperative camera-sensing environment, there are not only data communications with the server but also among different sensors. Therefore, two additional issues are multi-sensor calibration and data synchronization.

Calibration in a cooperative environment aims to determine the perspective transformation between sensors to be able to merge acquired data from several views at a given frame (Caillot et al. 2022). This task is quite challenging in a multi-user environment because the transformation matrix between sensors constantly changes as the vehicles move. In a cooperative context, calibration relies on the synchronization of the elements in a background image to determine the transformation between static or mobile sensors (Yang et al. 2021). There are multiple sources of desynchronization, such as an offset between the clocks or variable communication delays. Although clocks may be synchronized, it is difficult to ensure the data acquisitions are triggered at the same moment which adds uncertainty towards merging the acquired data. Similarly, different sampling rates require interpolation between acquired or predicted data, also adding uncertainty.

Quality of Training Data and Benchmarks

Traffic cameras are widely deployed on roadways and vehicles (Ke 2020). TMCs at DOTs and cities constantly collect network-wide traffic camera data, which are required for various ITS applications, such as event recognition and vehicle detection. However, labeled training data is much less common than unlabeled data (Halevy et al. 2009; Luo et al. 2018). The lack of annotated datasets for many applications is slowly being overcome with synthetic data, as graphical fidelity and simulated physics have become more

and more realistic. For example, ground truth 3D information in Hu et al. (2019) needs high accuracy during training for monocular 3D detection and tracking, so video game data was used. In addition to realistic appearance, simulated scenarios do not need to be manually labeled as the labels are already generated by the simulation, and can support a wide variety of illuminations, viewpoints, and vehicle behaviors (Yao et al. 2020). The 2020 AI City challenge for vehicle re-identification winner utilized a hybrid dataset to significantly improve the performance (Zheng et al. 2020) by generating examples from real-world data and adding other simulated views and environments. However, if using synthetic data, additional learning procedures, e.g., domain adaptation, are still needed for real-world applications. Low-fidelity simulated data were used to train a real-world object detector with domain randomization transfer learning (Tobin et al. 2017).

The lack of good quality crash and near-crash data is often cited as a practical limitation (Taccari et al. 2018). More crash data will update the attention guidance in AD, allowing it to capture long-term crash characteristics, thereby improving crash risk estimation (Li et al. 2021b). There is also a lack of representation in the literature regarding bicycles as the ego vehicle as mentioned in Ibrahim et al. (2021). A near-miss incident database was developed in Kataoka et al. (2018) to compensate for the unavailability, however, it is private because of copyright issues. A review of vehicle behavior prediction methods (Mozaffari et al. 2022) discusses the lack of a benchmark for evaluating existing studies, preventing a fair comparison of different DL techniques, or classical methods like Bayesian or Markov decision process. It also highlights that faulty or limited sensors, constrained computational resources, and generalizability to any driving scenario are current barriers to practical deployment and represent a significant research gap. Some of these issues can be addressed by sensor fusion, internet of vehicles (IoV), and edge computing (Wang et al. 2020a).

Data Bias

Although current vehicle detection algorithms perform well on balanced datasets, they suffer from performance degradation on tail classes when facing imbalanced datasets. In real-world scenarios, data tends to obey the Zipfian distribution (Reed 2001) where a large number of tail categories have fewer samples. A typical example of this can be seen in the histogram in Fig. 2. In long-tail datasets, a few head classes (frequent classes) contribute most of the training samples, while tail classes (rare classes) are underrepresented. Most DL models trained with such data minimize empirical risk on long-tail training data, and are biased towards head categories since they contribute most

Fig. 2 Illustration of representative challenges associated with data in computer vision applications for transportation
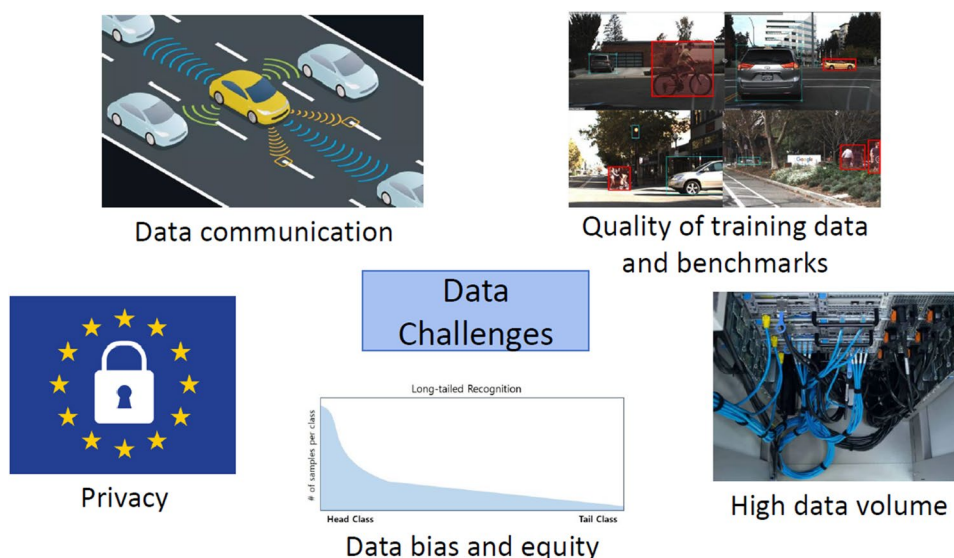
of this training data (Wang et al. 2022; Fu et al. 2021). Some methods like data re-sampling (Mahajan et al. 2018), loss re-weighting (Wang et al. 2021b), and cost-sensitive learning (Formosa et al. 2023), can compensate the underrepresented classes. However, they need to partition the categories into several groups based on their category frequency prior. Such hard division between head and tail classes brings two problems: training inconsistency between adjacent categories and lack of discriminative power for rare categories (Wang et al. 2021c).

An object detection study focusing on construction vehicles found that training a deep model with a huge general training dataset did not perform as well as a smaller model trained specifically on construction vehicles (Arabi et al. 2020). Another model based on YOLOv2 for vehicle size estimation performed well on commonly seen sizes but varied considerably with uncommon sizes (Wu et al. 2019). The dataset used by Carranza-García et al. (2021) for autonomous vehicle object detection had severe class imbalance with only 1% cyclists represented. A number of weight-based learning strategies were employed to address this, giving higher weight to underrepresented classes, showing significant improvements.

General object detectors can be improved using transfer learning with the underrepresented data for task-specific performance benefits (Zhao et al. 2019). In addition, it is noted in Ras et al. (2018) that model bias may not always be apparent from just the training set, and explanability methods are needed to address the problem.

High Data Volume

Visual data is composed of over 90% of the Internet traffic, and video transmission, computation, and storage pose

increasing challenges in ITS and AV fields (Ke 2020). The high volume of traffic and vehicular-based video data from the roadside and onboard sensors via the traffic camera network or the Internet of Vehicles (IoV) network poses computational and bandwidth bottlenecks that cannot be solved by using more powerful equipment (Xu et al. 2018). As many applications in connected or autonomous vehicles rely on DL, vehicle-cloud architecture is emerging as an effective distributed computing technique (Wang et al. 2011). With the integration of Road Side Units (RSU), these edge nodes can process faster and provide low communication latency.

Security and Privacy

Privacy concerns are an important human factor that cannot be overlooked in the design and operation of ITS applications (Fries et al. 2012). Observing and tracking the massive amounts of pedestrian and vehicle information causes security and privacy concerns in ITS environments. For example, UAVs are capable of collecting traffic data (through onboard video cameras). However, privacy concerns restrict them from being a regular part of the ITS sensor network (Khan et al. 2021). Video surveillance systems constantly collect human faces and license plates. Personal privacy is exchanged for security or safety services provided by the surveillance (Martínez-Ballesté et al. 2013). Systems deployed in practice might need to de-identify faces and license plates in real-time if raw video data is being sent or stored (Martínez-Ballesté et al. 2012). Any processing would ideally be done on the local edge unit, limiting the propagation of private information. Full anonymity is difficult to guarantee, for example, an uncommon model car with a distinctive paint pattern can be traced to its owner

by correlating with other information, even with a blurred license plate.

Model Challenges

Complexity

DL computer vision models have a high complexity with respect to neural network structures and training procedures. Many DL models are designed to run on high-performance cloud centers or AI workstations, and a good model requires weeks or months of training as well as high power consumption driven by modern Graphical Processing Units (GPUs) or Tensor Processing Units (TPUs).

In ITS and AV field applications, many have a requirement for real-time or near real-time operations (Ke 2020) for the sake of functionality and traffic safety. The DL model complexity adds a high cost to the training and inference in real-time applications; particularly, the trend of ITS and AV is towards large-scale on-device processing closer to where the traffic data is generated, e.g., crowdsensing. Three popular embedded devices are compared in Arabi et al. (2020), with the Nvidia Jetson Nano yielding the highest inference efficiency but it is too little computation power for complex applications.

Real-time applications usually make some modifications like resizing video to lower resolution or model quantization and pruning which can lead to loss of performance. The model complexity of the state-of-the-art DL methods needs to be reduced in many practical applications to meet the efficiency and accuracy requirements. For example, multi-scale deformable attention has been used with vision transformer neural networks in object detection for high performance and fast convergence leading to faster training and inference (Zhu et al. 2021).

Lack of Explainability

DNNs are largely seen as black boxes with many layers of processing, the working of which can be examined using statistics, but the learned internal representations of the network are based on millions or billions of parameters, making analysis extremely difficult (Ras et al. 2018). This means that the behavior is essentially unpredictable, and very little explanation can be given of the decisions. It also makes system validation and verification impossible for critical use cases like autonomous driving (Samek et al. 2017).

The common assumption that a complex black box is necessary for good performance is being challenged (Rudin 2019). Recent research is attempting to make DNNs more explainable. A visualization tool for vision transformers is presented in Aflalo et al. (2022), which can be used to see the inner mechanisms, such as hidden parameters, and gain insight into specific parts of the input that influenced the predictions. A framework for safety, explainability, and regulations for autonomous driving was evaluated in post-accident scenarios (Atakishiyev et al. 2021). The results showed many benefits including transparency and debugging. A convolutional neural network (CNN)-based architecture is proposed to detect action-inducing objects for autonomous vehicles, while also providing explanations for the actions (Xu et al. 2020).

Transferability and Generalizability

Generalization to out-of-distribution data is natural to humans yet challenging for machines because most learning algorithms strongly rely on the independent and identically distributed (i.i.d.) assumption training at testing data, which is often violated in practice due to domain shift. Domain generalization aims to generalize models to new domains without knowledge of the target distribution during training. Different methods have been proposed for learning generalizable and transferable representations (Dou et al. 2019).

Most existing approaches belong to the category of domain alignment, where the main idea is to minimize the difference between source domains for learning domain-invariant representations. Features that are invariant to the source domain shift should also be robust to any unseen target domain shift. Data augmentation has been a common practice to regularize the training of machine learning models to avoid overfitting and improve generalization (LeCun et al. 2015), which is particularly important for over-parameterized DNNs.

Visual attention in CNNs can be used to highlight the regions of the image involved in a decision, with causal filtering to find the most relevant parts (Kim and Canny 2018). The importance of individual pixels is estimated in Petsiuk et al. (2018) using randomly masked versions of images and comparing the output predictions. This approach does not apply to spatio-temporal methods or those that consider relationships between objects in complex environments.

Real-World Testing

In general, DL methods have been shown to be prone to underspecification, a problem that appears regardless of model type or application. Among other domains, underspecification of computer vision is analyzed in D'Amour et al. (2020), specifically for DL models such as the commonly used ResNet-50 and a scaled-up transfer learning image classification model, Big Transfer (BiT) Kolesnikov et al. (2019). It is shown that while benchmark scores improved with more model complexity and training data, testing with real-world distortions results in poor and highly

varied performance that depends strongly on the random seeds used to initialize training.

Practical systems need to be efficient in terms of memory and computation for real-time processing on a variety of low-cost hardware (Bai et al. 2021). Some approaches towards efficient and low-cost computation include parameter pruning, network quantization, low-rank factorization, and model distillation. Approaches like Cui et al. (2019) are efficient and capable of real-time trajectory prediction but are not end-to-end because they assume the prior existence of an object-tracking system to estimate the states of surrounding vehicles.

Vulnerable road users (VRU) such as pedestrians and bicyclists present a unique problem, since they can change their direction and speeds very rapidly, and interact with the traffic environment differently than vehicles (Saleh et al. 2017).

Some of the major barriers to the practical deployment of computer vision models in ITS are the heterogeneity of data sources and software, sensor hardware failure, and extreme or unusual sensing cases (Zhou et al. 2021). Furthermore, recent frameworks such as those based on edge computing directly expose the wireless communication signals of a multitude of heterogeneous devices with various security implementations, creating an ever-increasing potential attack surface for malicious actors (Contreras-Castillo et al. 2018; Haghighat et al. 2020). Deep learning models have been developed to detect these attacks, however real-time application and online learning are still areas of active research (Chen et al. 2019a).

IoV faces fundamental practical issues arising from the fact that moving vehicles will present highly variable processing requirements on the edge nodes, while each vehicle can also have many concurrent edge and cloud-related applications running, along with harsh wireless communication environments (Zhang and Letaief 2020). Other challenges related to edge computing for autonomous vehicles include cooperative sensing, cooperative decisions, and cybersecurity (Liu et al. 2019). Attackers can use lasers and bright infrared light to interfere with cameras and LiDAR, change traffic signage, and replay attacks over the communication channel. A visual depiction of model challenges can be seen in Fig. 3.

Complex Traffic Environments

Shadow, Lighting, Weather

Situations like shadows, adverse weather, similarity between background and foreground, strong or insufficient illumination in the real world are cited as common issues (Lin et al.

2019; Song et al. 2020). The appearance of camera images is known to be affected by adverse weather conditions, such as heavy fog, sleeting rain, snowstorms, and dust storms (Hassaballah et al. 2020).

A real-time crash detection method in Jiansheng (2014) utilizes foreground extraction using the Gaussian Mixture Model, then tracks vehicles using a mean shift algorithm. The position, speed, and acceleration of the vehicles are passed through a threshold functions to determine the detection of a crash. While computationally efficient, such methods suffer significantly in the presence of noise, complex traffic environment, and change in weather.

In harsh weather conditions, vehicles captured by traffic surveillance cameras exhibit issues such as underexposure, blurring, and partial occlusion. At the same time, raindrops, and snowflakes that appear in traffic scenes add difficulty for the algorithm to extract vehicle targets (Yang and Pun-Cheng 2018). At night, or in tunnels with vehicles driving towards the camera, the scene may be masked completely because of the high beam glare (Sonnleitner et al. 2020).

Occlusion

Occlusion is one of the most challenging issues, where a target object is only partially visible to the camera or sensor due to obstruction by another foreground object. Occlusion exists in various forms ranging from partial occlusion to heavy occlusion (Gilroy et al. 2019). In AD, target objects can be occluded by static objects such as buildings and lampposts. Dynamic objects such as moving vehicles or other road users may occlude one another, such as in crowds. Figure 4 shows how a single bus can occlude multiple vehicles. Occlusion is also a common issue in object tracking (Nowosielski et al. 2016) because once the tracked vehicle disappears from view and reappears, it is considered a different vehicle causing tracking and trajectory information to be inaccurate. In fact, when the vehicle reappears, it is double counted by detection and tracking algorithms regardless of the model used resulting in exaggerated counts (Mandal and Adu-Gyamfi 2020). Data imputation and post-processing for error correction are important steps for practical applications involving tracking through occlusion but these often require a manual analysis of results (Dhatbale and Chilukuri 2021).

Camera Angle

In the applications of transportation infrastructure, the diversity of surveillance cameras and their viewing angles pose challenges to DL methods trained on limited types of camera views (Buch et al. 2011; Santhosh et al. 2020). While the queue length estimation in Albiol et al. (2011) is computationally efficient and can work in varying lighting conditions and traffic density scenarios, lower-pitch camera views
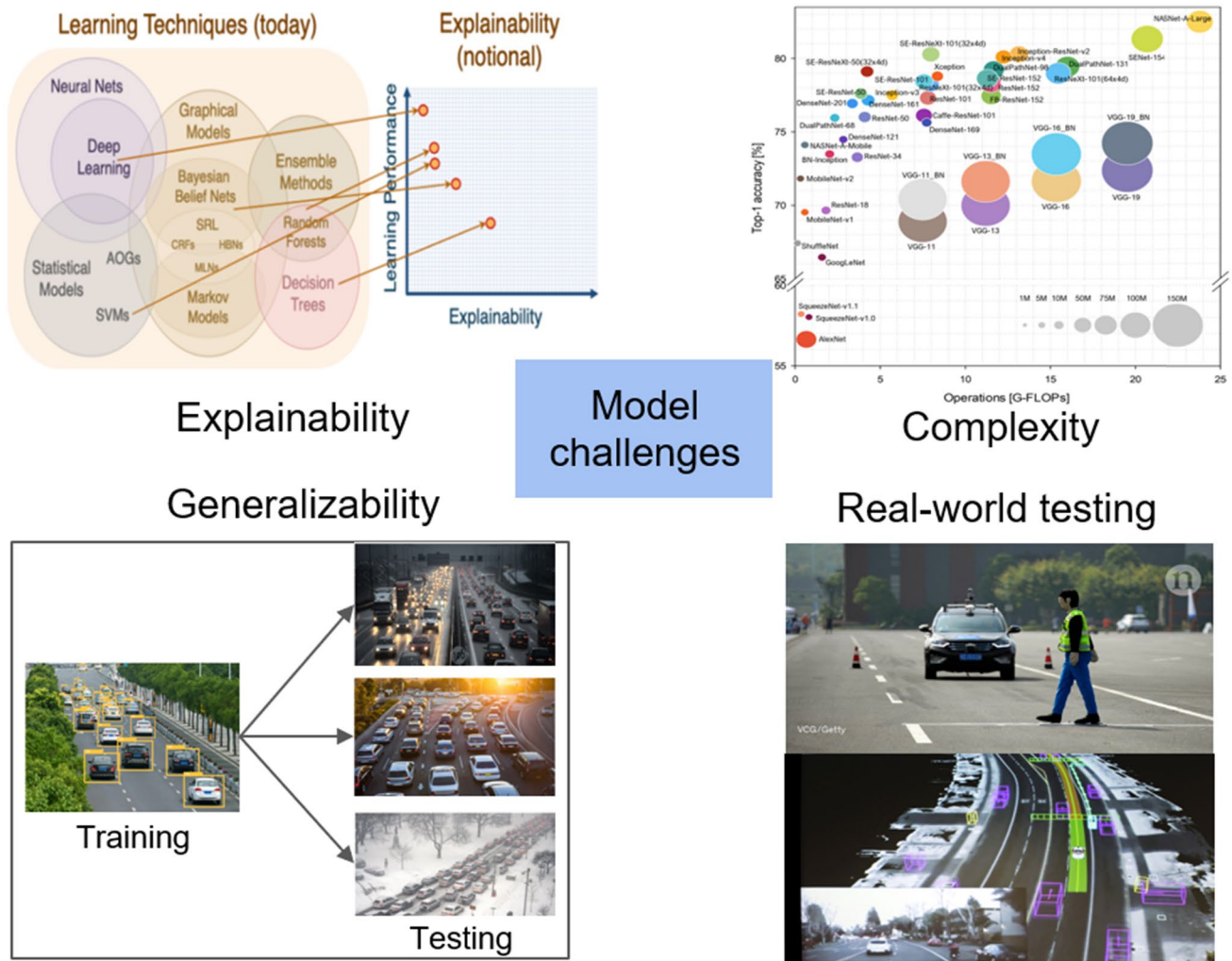
Fig. 3 Illustration of representative model challenges. Some demo images are adopted from Bianco et al. (2018); Bornstein (2016)

and road-marking corners can introduce significant errors. Similarly, cost-effective city-wide real-time vehicle count solutions are scalable, but accuracy drops due to shallow traffic camera angles (Huang and Sharma 2020). The model in Aboah et al. (2021) can identify anomalies near the camera, including their start and end times, but is not accurate for anomalies in the distance since the vehicles occupy only a few pixels.

An earlier survey on anomaly detection from surveillance video concluded that illumination, camera angle, heterogeneous objects, and a lack of real-world datasets are the major challenges (Santhosh et al. 2020). The methods used for sparse and dense traffic conditions are different and lack generalizability. Matching objects in different views is another major problem in a multi-view vision scene, as multi-view ITS applications need to process data across the different images captured by different cameras at the same time (Xie et al. 2021).

Camera Blur and Degraded Images

Surveillance cameras are subject to weather elements. Water, dust, and particulate matter can accumulate on the lens causing image quality degradation. Strong wind can cause a camera to shake, resulting in motion blur in the whole image. Front-facing cameras on autonomous vehicles also face this, as insects can smash onto the glass, causing blind spots in the camera's field of view. Specifically, object detection and segmentation algorithms suffer greatly (Vasiljevic et al. 2017), and unless preparations are made in the model, false detections can cause serious safety issues in AD and miss important events in surveillance applications. Some approaches to address this include using degraded images for training, image restoration preprocessing, and fine-tuning pre-trained networks to learn from degraded images. For example, Dense-Gram networks are used in Guo et al. (2019) which
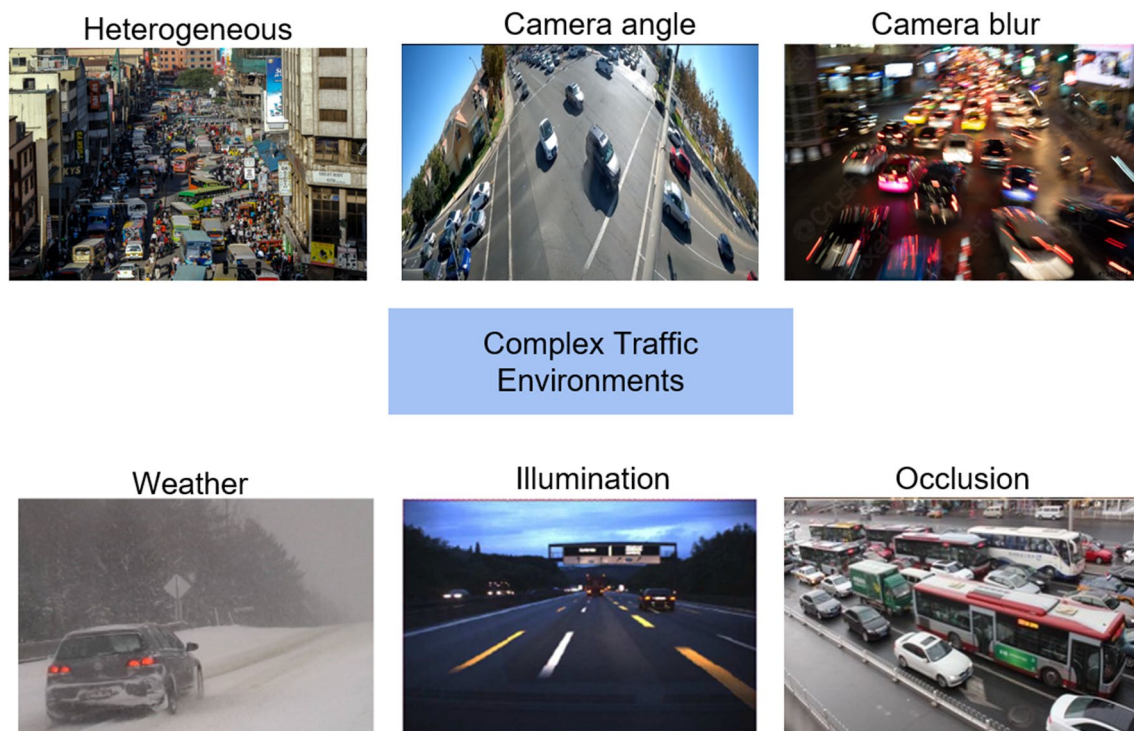
Fig. 4 Illustration of representative scenarios in complex traffic environments. Some demo images are adopted from Yang and Pun-Cheng (2018)

improve image segmentation performance in degraded images.

Heterogeneous, Urban Traffic Conditions

Dense urban traffic scenarios are full of complex visual elements, not only in quantity but also in the variety of different vehicles and their interactions, as shown in Fig. 4. The presence of cars, buses, bicycles, and pedestrians in the same intersection is a significant problem for autonomous navigation and trajectory computation (Ma et al. 2018). The different sizes, turning radii, speeds, and driver behaviors are further compounded by the interactions between these road users. From a DL perspective, it is easy to find videos of heterogeneous urban traffic, but labeling for ground truth is very time-consuming. Simulation software usually cannot capture the complex dynamics of such scenarios, especially the traffic rule-breaking behaviors seen in dense urban centers. In fact, a specific dataset was created to represent these behaviors in Chandra et al. (2019a). A simulator for unregulated dense traffic was created in Cai et al. (2020) which is useful for autonomous driving perception and control but does not represent the trajectory and interactions of real-world road users.

Applications

Traffic Flow Estimation

Models and Algorithms

Traffic flow variables include traffic volume, density, speed, and queue length. The algorithms and models to detect and track objects to estimate traffic flow variables from videos may be classified into one-stage and two-stage methods. In one-stage methods, the variables are estimated from detection results and there is no further classification and location optimization, for example: (1) YOLOv3 + Feature stitching (Hong et al. 2020; (2) YOLOv2 + spatial pyramid pooling (Kim et al. 2019; (3) AlexNet + optical flow + Gaussian mixture model (Ke et al. 2018a; (4) CNN + optical flow based on UAV video (Ke et al. 2018b; (5) SSD (single shot detection) based on UAV video (Tang et al. 2017).

Two-stage methods first generate region proposals that contain all potential targets in the input images and then conduct classification and location optimization. Examples of two-stage methods are: (1) Faster R-CNN + SORT tracker (Fedorov et al. 2019; (2) Faster R-CNN (Peppa et al. 2018; Mhalla et al. 2018; (3) Faster R-CNN based on UAV video (Peppa et al. 2021; Brkić et al. 2020).

Table 1 Computer vision applications in transportation

Application	Methods	Main challenges
Traffic flow parameter estimation	YOLOv3 + DeepSORT on surveillance videos (Hong et al. 2020), YOLOv2 + SPP (Kim et al. 2019), AlexNet + optical flow + Gaussian mixture model Ke et al. (2018a), Faster R-CNN + SORT (Fedorov et al. 2019)	Single metric for performance evaluation, weather and illumination, realtime video processing, diversity of surveillance camera view angles
Congestion detection and prediction	AlexNet + YOLO (Chakraborty et al. 2018), AlexNet + VGGNet (Wang et al. 2020d), YOLO + Mask R-CNN (Impedovo et al. 2019), YOLOv3 (Rashmi and Shantala 2020), YOLOv4 (Sonmleitner et al. 2020), Faster R-CNN Gao et al. (2021a)	Multi-lane highways with exits or barriers, parked vehicles, lack of lane markings, weather and illumination, realtime processing, variety of surveillance camera angles
Autonomous driving perception: detection	Sparse R-CNN (Cao et al. 2021), YOLOv3 (Wan et al. 2021a), SegNet + U-Net (Kamal et al. 2019), Faster R-CNN + Inception-Resnet-v2 Kim et al. (2018), CNN + ConvLSTM (Zou et al. 2019), GAN (Ghafoorian et al. 2018)	Weather, shadows, complex road environment, occlusion, small objects, degraded lane markings, motion blur
Autonomous driving perception: segmentation	YOLACT (Bolya et al. 2019), CenterMask (Lee and Park 2020), Path-Aggregation Network (Liu et al. 2018b), ESE-Seg (Xu et al. 2019)	Accuracy vs. processing speed trade-off, occlusion, illumination, shadows, weather, lack of generalizability
Cooperative perception	Early fusion (Chen et al. 2019d), DiscoNet (Li et al. 2021d), Intermediate fusion F-Cooper (Chen et al. 2019b), V2V intermediate fusion (Wang et al. 2020b), Attentive intermediate fusion (Xu et al. 2021), V2X + Vision transformer (Xu et al. 2022), Late fusion (Rauch et al. 2012)	Delay, packet loss, calibration, synchronization, non-overlapping points of view, computational cost
Vehicle interaction	YOLOv3 + optical flow (Taccari et al. 2018), Dynamic Spatial Assistance RNN (Chan et al. 2016), YOLOv3 + Gaussian velocity field + hidden semi-Markov (Zhang et al. 2019), spatio-temporal attention guidance + YOLOv3 (Li et al. 2021b), Domain adaptation + GAN (Li et al. 2021a), Gated recurrent unit + CNN (Choi et al. 2021a), Retinex + YOLOv3 Wang et al. (2020c), optical flow + CNN + LSTM (Ibrahim et al. 2021), ResNet (Suzuki et al. 2017)	Lack of near-crash data or other specific situations, underrepresentation of bicycles, complex urban traffic, weather, illumination, occlusion
Road user behavior prediction	LSTM + CNN (Chandra et al. 2019a), LSTM + SVM (Tripicchio and D'Avella 2022), Switching Linear Dynamical System + bayesian network (Kooij et al. 2019), YOLO + LSTM + random forest (Choi et al. 2021b), Transformer (Liu et al. 2021b)	Lack of benchmark to evaluate methods, realtime computation, faulty sensors, generalizability to different driving scenarios, occlusion
Traffic anomaly detection	Gaussian mixture model + Optical flow (Shirazi and Morris 2015), GMM background modeling + AdaBoost classifier + Modified Hausdorff Distance (Nowosielski et al. 2016), Locality-constrained Linear Coding (Li et al. 2016a), YOLOv5 + decision tree + adaptive thresholding (Aboah et al. 2021), ResNet50 + spatial-temporal matrix discriminating module (Bai et al. 2019), spatio-temporal convolutional autoencoder + LSTM (Pawar and Attar 2021)	Vehicles far from camera, camera angles, illumination, lack of generalizability between sparse and dense conditions, camera shaking by wind

Table 1 (continued)		
Application	Methods	Main challenges
Edge computing	Integer linear programming + fast heuristics (Cui et al. 2020), SSD + SORT (Ke et al. 2020), Deep Deterministic Policy Gradient + V2V networking (Dai et al. 2019), cloud-edge hybrid + Global foreground modeling + Gaussian mixture model (Liu et al. 2021a), YOLOv3 Wan et al. (2022), federated learning (Kairouz et al. 2019), spectral clustering compression (Chen et al. 2021a)	Power consumption, heterogeneous data sources, cybersecurity, wireless noise, scalability, neural network pruning and model compression, installation and maintenance costs

In both cases the focus is on effectively detecting vehicles allowing for accurate counts in a given road segment, while tracking enables the estimation of average speed and movement directions from traffic video.

Current Methods to Overcome Challenge

A DL method at the edge of the ITS that performs real-time vehicle detection, tracking, and counting in traffic surveillance video has been proposed in Chen et al. (2021b). The neural network detects individual vehicles at the single-frame level by capturing appearance features with the YOLOv3 object-detection method, deployed on edge devices to minimize bandwidth and power consumption, which are major practical hurdles in deployment. A vehicle detection and tracking approach in adverse weather conditions that achieves the best trade-off between accuracy and detection speed in various traffic environments is discussed in Hassaballah et al. (2020). Also, a novel dataset called DAWN (Kenk and Hassaballah 2020) is introduced for vehicle detection and tracking in adverse weather conditions like heavy fog, rain, snow, and sandstorms, to make training less biased. Meanwhile, low resolution and slow framerate issues are specifically addressed in Wei et al. (2019) to allow large-scale implementation on existing urban traffic surveillance systems using SSD-Mobilenet for detection and VGG16 features for tracking.

Traffic Congestion Detection

Models and Algorithms

The methods that detect traffic congestion based on computer vision may also be divided into one-stage methods and multi-step methods. The one-stage methods identify vehicles from the video images and directly perform traffic congestion detection. Among the one-stage methods are: (1) AlexNet and YOLO (Chakraborty et al. 2018) to distinguish congestion and non-congestion, (2) AlexNet and VGG-Net (Wang et al. 2020d) which classify 'jam' and 'no jam'; and (3) YOLO and Mask R-CNN (Impedovo et al. 2019) recognize light, medium, and heavy congestion (identifying the number of vehicles in each frame and then classify). The multi-step methods first apply traffic flow estimation models to measure traffic variables and then use the traffic flow variables to infer congestion. Examples of two-stage traffic congestion detection models are: (1) YOLOv3 (Rashmi and Shantala 2020) and YOLOv4 (Sonnleitner et al. 2020) for vehicle detection and counting, (2) counting vehicles using Faster R-CNN (Gao et al. 2021a) and applying regression for traffic congestion. Beside these, the traffic congestion can be evaluated by the traffic flow detection algorithms (Kumar and Raubal 2021) using vehicle detection and tracking.

Current Methods to Overcome Challenge

Congestion detection performance can be improved using multiple sensors based solutions including radar, lasers, and sensor fusion since it is hard to achieve ideal performance and accuracy using a single sensor in real-world scenarios. There is a wide use of decision-making algorithms for processing fusion data acquired from multiple sensors (Muhammad et al. 2020). A CNN-based model trained with bad weather condition datasets can improve the detection performance (Sharma et al. 2022), while generative adversarial network (GAN) based Style Transfer methods have also been applied (Lin et al. 2020; Li et al. 2021a). These approaches help to minimize the model challenges related to generalizability, which in turn improves real-world performance in a variety of environments.

Autonomous Driving Perception: Detection

Models and Algorithms

Common detection tasks that assist in AD are categorized into traffic sign detection, traffic signal detection, road/lane detection, pedestrian detection, and vehicle detection.

Traffic signs There are two tasks in a typical traffic sign recognition system: finding the locations and sizes of traffic signs in natural scene images (traffic sign detection) and classifying the traffic signs into their specific sub-classes (traffic sign classification) (Yang et al. 2015). An improved Sparse R-CNN was used for traffic sign detection in Cao et al. (2021), while an efficient algorithm based on YOLOv3 model for traffic sign detection was implemented in Wan et al. (2021a). SegU-Net, formed by merging the state-of-the-art segmentation architectures SegNet and U-Net to detect traffic signs from video sequences has been proposed (Kamal et al. 2019). Several adaptations to Mask R-CNN were tested in Tabernik and Skočaj (2019) for detection and recognition with end-to-end learning in the domain of traffic signs. They also proposed a data augmentation technique based on the distribution of geometric and appearance distortions.

A method that uses an encoder-decoder DNN with focal regression loss to detect small traffic signals is proposed in Lee and Kim (2019). It is shown in Kim et al. (2018) that Faster R-CNN with Inception-Resnet-v2 model is more suitable for traffic light detection than others. A practical traffic light detection system in Ouyang et al. (2019) combines CNN classifier model and heuristic region of interest (ROI) candidate detection on self-driving hardware platform Nvidia Jetpack Tx1/2 that can handle high-resolution images. The recognition accuracy and processing speed are improved by combining detection and tracking in Wang et al. (2021a) to enhance the practicality of the traffic signal recognition system in autonomous vehicles using CNN and integrated channel feature tracking to determine the coordinates and color for traffic lights.

Lane detection aims to identify the left and right lane boundaries from a processed image and apply an algorithm to track the road ahead. A novel hybrid neural network combining CNN and recurrent neural network (RNN) for robust lane detection in driving scenes has been proposed (Zou et al. 2019). Features on each frame of the input video were first abstracted by a CNN encoder and the sequential encoded features were processed by a ConvLSTM. The outputs were fed into the CNN decoder for information reconstruction and lane prediction. Another lane detection method is an anchor-based single-stage lane detection model called LaneATT (Tabelini et al. 2021). It uses a feature pooling method with a relatively lightweight backbone CNN while maintaining high accuracy. A novel anchor-based attention mechanism to aggregate global information was also proposed. A new method to impose structure on badly posed semantic segmentation problems is proposed in Ghafoorian et al. (2018) using a generative adversarial network architecture with a discriminator that is trained on both predictions and labels at the same time.

Pedestrian detection A two-stage detector SDS-RCNN (Brazil et al. 2017) jointly learned pedestrian detection and bounding-box aware semantic segmentation, thus encouraging model learning on pedestrian regions. RPN+BF (Zhang et al. 2016b) used a boosted forest to replace second-stage learning and leveraged hard mining for proposals. However, involving such downstream classifiers could bring more training complexity. AR-Ped (Brazil and Liu 2019) exploited sequential labeling policy in the region proposal network to gradually filter out better proposals. The work of Chen et al. (2018) employed a two-stage pretrained person detector (Faster R-CNN) and an instance segmentation model for person re-identification. Each detected person is cropped out from the original image and fed to another network. Wang et al. (2018) introduced repulsion losses that prevent a predicted bounding box from shifting to neighboring overlapped objects to counter occlusions. Two-stage detectors need to generate proposals in the first stage and thus are slow for inference in practice. One-stage detector GDFL (Lin et al. 2018) included semantic segmentation, which guided feature layers to emphasize pedestrian regions. Liu et al. (2018a) extended the single-stage architecture with an asymptotic localization fitting module storing multiple predictors to evolve default anchor boxes. This improves the quality of positive samples while enabling hard negative mining with increased thresholds. Similar to pedestrian detection, vehicle detection in ITS also is a popular and challenging computer vision task (Zhao et al. 2019).

Vehicle detection Current generic vehicle detectors are divided into two categories: CNN-based two-stage detectors

and CNN-based one-stage detectors. the representative two-stage detectors include Faster R-CNN (Zhang et al. 2016), spatial pyramid pooling (SPP)-net (He et al. 2015), feature pyramid networks (FPN) (Lin et al. 2017), and Mask R-CNN (He et al. 2017). The representative one-stage detectors include YOLO (Redmon et al. 2016), Single Shot Multi-Box Detector (SSD) (Liu et al. 2016), and deeply supervised object detectors (DSOD) (Shen et al. 2017). The two-step framework is a region proposal-based method, giving a coarse scan of the whole image first and then focusing on regions of interest (RoIs). While one-step frameworks are based on global regression/ classification, mapping straightly from image pixels to bounding box coordinates and class probabilities. Based on these two frameworks, most of the works get promising results in vehicle detection by combining other methods, such as multitask learning (Brahmbhatt et al. 2017), multi-scale representation (Bell et al. 2016), and context modeling (Kong et al. 2016).

Current Methods to Overcome Challenge

In traffic sign detection, existing traffic sign datasets are limited in terms of the type and severity of challenging conditions. Metadata corresponding to these conditions is unavailable and it is not possible to investigate the effect of a single factor because of the simultaneous changes in numerous conditions. To overcome this, Temel et al. (2019) introduced the CURE-TSDReal dataset, based on simulated conditions that correspond to real-world environments. An end-to-end traffic sign detection framework Feature Aggregation Multipath Network (FAMN) is proposed in Ou et al. (2019). It consists of two main structures named Feature Aggregation and Multipath Network structure to solve the problems of small object detection and fine-grained classification in traffic sign detection.

A vehicle highlight information-assisted neural network for vehicle detection at night is presented in Mo et al. (2019), which included two innovations: establishing the label hierarchy for vehicles based on their highlights and designing a multi-layer fused vehicle highlight information network. Real-time vehicle detection for nighttime situations is presented in Bell et al. (2021), where images include flashes that occupy large image regions, and the actual shape of vehicles is not well defined. By using a global image descriptor along with a grid of foveal classifiers, vehicle positions are accurately and efficiently estimated. AugGAN Lin et al. (2020) is an unpaired image-to-image translation network for domain adaptation in vehicle detection. It quantitatively surpassed competing methods for achieving higher nighttime vehicle detection accuracy because of better image-object preservation. A stepwise domain adaptation (SDA) detection method is proposed to further improve the performance of CycleGAN by

minimizing the divergence in cross-domain object detection tasks in Li et al. (2022). In the first step, an unpaired image-to-image translator is trained to construct a fake target domain by translating the source images to similar ones in the target domain. In the second step, to further minimize divergence across domains, an adaptive Center-Net is designed to align distributions at the feature level in an adversarial learning manner.

Autonomous Driving Perception: Segmentation

Models and Algorithms

Image segmentation contains three sub-tasks: semantic segmentation, instance segmentation, and panoptic segmentation. Semantic segmentation is a fine prediction task to label each pixel of an image with a corresponding object class, instance segmentation is designed to identify and segment pixels that belong to each object instance, while panoptic segmentation unifies semantic segmentation and instance segmentation such that all pixels are given both a class label and an instance ID (Gu et al. 2022).

YOLACT (Bolya et al. 2019) splits instance segmentation into two parallel sub-architectures. Protonet architecture extracts spatial information by generating a certain number of prototype masks, and Head architecture generates the mask coefficients and object locations. In addition, it employs Fast NMS rather than traditional NMS to reduce post-processing time. Path Aggregation Network (PANet) (Liu et al. 2018b) is proposed to integrate comprehensively low-level location information and high-level semantic information. Based on Feature Pyramid Networks (FPN) (Lin et al. 2017), PANet designs a bottom-up context information aggregation structure, which can integrate different levels of features. Hybrid Task Cascade (HTC) is proposed for instance segmentation in Chen et al. (2019c). It interweaves box and mask branches for joint multi-stage processing, adopts a semantic segmentation branch to provide spatial context, and integrates complementary features together in each stage.

In Dong et al. (2020), a novel real-time segmentation is proposed consisting of a convolutional attention module, spatial pyramid pooling, and a feature fusion network. It was evaluated on benchmark datasets Cityscapes and CamVid, which specifically target complex urban scenarios.

Moving objects viewed from a moving platform pose a unique challenge for segmentation, which is addressed in Zhou et al. (2017) using time-consecutive stereo images. Motion likelihood estimates for each pixel aids in ego-motion estimation, while segmentation is performed using a graph-cut algorithm. However, computational complexity is a major limitation of this method.

Current Methods to Overcome Challenge

Recent directions in segmentation include weakly-supervised semantic segmentation (Wang et al. 2020e; Sun et al. 2020c), domain adaptation (Chen et al. 2019e; Liu et al. 2021c), multi-modal data fusion (Feng et al. 2020; Cortinhal et al. 2021), and real-time semantic segmentation (Yu et al. 2018; Nirkin et al. 2021; Gao et al. 2021b).

TS-Yolo Wan et al. (2021b) is a CNN-based model for accurate traffic detection under severe weather conditions using new samples from data augmentation. The data augmentation was conducted using copy–paste strategy, and a large number of new samples were constructed from existing traffic-sign instances. Based on YoloV5, MixConv was also used to mix different kernel sizes in a single convolution operation so that patterns with various resolutions can be captured. Detecting and classifying real-life small traffic signs from large input images is difficult due to them occupying fewer pixels relative to larger targets. To address this, Dense-RefineDet (Sun et al. 2020) applies a single-shot, object-detection framework to maintain a suitable accuracy-speed trade-off. An end-to-end traffic sign detection framework Feature Aggregation Multipath Network is proposed in Ou et al. (2019) to solve the problems of small object detection and fine-grained classification in traffic sign detection.

Cooperative Perception

Models and Algorithms

In connected autonomous vehicles (CAV), cooperative perception can be performed at three levels depending on the type of data: early fusion (raw data), intermediate fusion (preprocessed data), where intermediate neural features are extracted and transmitted, and late fusion (processed data), where detection outputs (3D bounding box position, confidence score) are shared. Cooperative perception studies how to leverage visual cues from neighboring connected vehicles and infrastructure to boost the overall perception performance (Xu et al. 2022).

(1) Early fusion: Chen et al. (2019d) fuses the sensor data collected from different positions and angles of connected vehicles using raw-data level LiDAR 3D point clouds, and a point cloud-based 3D object detection method is proposed to work on a diversity of aligned point clouds. DiscoNet (Li et al. 2021d) leverages knowledge distillation to enhance training by constraining the corresponding features to the ones from the network for early fusion.

(2) Intermediate fusion: F-Cooper (Chen et al. 2019b) provides both a new framework for applications On-Edge, servicing autonomous vehicles as well as new strategies for 3D fusion detection. Wang et al. (2020b) proposed a vehicle-to-vehicle (V2V) approach for perception and prediction that transmits compressed intermediate representations of the P &P neural network. Xu et al. (2021) proposed an Attentive Intermediate Fusion pipeline to better capture interactions between connected agents within the network. A robust cooperative perception framework with vehicle-to-everything (V2X) communication using a novel vision Transformer is presented in Xu et al. (2022).

(3) Late fusion: Car2X-based perception (Rauch et al. 2012) is modeled as a virtual sensor to integrate it into a high-level sensor data fusion architecture.

Current Methods to Overcome Challenge

To reduce the communications load and overhead, an improved algorithm for message generation rules in collective perception is proposed (Thandavarayan et al. 2020), which improves the reliability of V2X communications by reorganizing the transmission and content of collective perception messages. This paper (Yoon et al. 2021) presents and evaluates a unified cooperative perception framework containing a decentralized data association and fusion process that is scalable with respect to participation variances. The evaluation considers the effects of communication losses in the ad-hoc V2V network and the random vehicle motions in traffic by adopting existing models along with a simplified algorithm for individual vehicle's on-board sensor field of view. AICP is proposed in Zhou et al. (2022), the first solution that focuses on optimizing informativeness for pervasive cooperative perception systems with efficient filtering at both the network and application layers. To facilitate system networking, they also use a networking protocol stack that includes a dedicated data structure and a lightweight routing protocol specifically for informativeness-focused applications.

Vehicle Interaction

Models and Algorithms

Computer vision methods can be used to detect and classify crash and near-crash events based on motion and trajectories. CNN, in the form of modified YOLOv3, is used to detect objects and extract semantic information about the road scene from onboard camera data from the SHRP2 dataset in Taccari et al. (2018). Optical flow is calculated from consecutive frames to track objects and to generate features (hard deceleration, the maximum area of the largest vehicle, time to collision, etc.) that are combined with telematics data (speed and 3-axis acceleration) to train a random forest classifier on safe, near-crash, and crash events.

Dashcam videos were used to train a Dynamic Spatial Assistance (DSA) network to distribute attention to objects and model temporal dependencies in Chan et al. (2016). The method was able to predict crashes around 2 s in advance. Understanding multi-vehicle interaction in urban environments is challenging, and model-based methods may require prior knowledge, so a more general approach is explored in Zhang et al. (2019), where YOLOv3 is used for object tracking from traffic camera video from the NGSIM dataset and a Gaussian velocity field is used to describe the interaction behaviors between multiple vehicles. From this, an 11-layer deep autoencoder learns the latent low-dimensional representations for each frame, followed by a hidden semi-Markov model with a hierarchical Dirichlet process, which optimizes the number of interaction patterns, to cluster representations into traffic primitives corresponding to the interaction patterns. The pipeline can be used to analyze complex multi-agent interactions from traffic video. DL methods are able to extract semantic descriptions from video, like in Li et al. (2021b), which can give advance warning of risky situations. A scenario-wise spatio-temporal attention guidance system was created by data mining from descriptive semantic variables in fatal crash data to support the design of a model based on YOLOv3 for evaluating crash risk from dashcam footage. The attention guidance extracted semantic descriptions like "pedestrian", "school bus" and "atmospheric condition", followed by DL to optimize attention on these variables to identify clusters and associate scene features with a crash features.

Current Methods to Overcome Challenges

While most vehicle interaction methods reviewed thus far make little mention of the practical challenges in variable weather and lighting, Zhang et al. (2018a) highlights a background learning method specifically to adapt to changing lighting conditions and headlight illumination in surveillance footage and even utilizes a threshold-based noise removal for rainy conditions to detect near-miss events at grade crossings. Domain adaptation, an example of transfer learning, was employed in Li et al. (2021a) to make use of labeled daytime footage for vehicle detection in unlabeled nighttime images by a generative adversarial network called CycleGAN (Zhu et al. 2017), which can be used with many real-world deep learning computer vision applications. YouTube dashcam footage was used for crash detection in an ensemble multimodal DL method, based on the gated recurrent unit (GRU) and CNN, which uses both video and audio data (Choi et al. 2021a). The real-world data consists of positive clips containing crashes and negative clips containing normal driving. A crowd-sourced dashcam video dataset was also contributed by Chan et al. (2016) for accident anticipation containing scenarios like crowded streets,

complicated road environments, and diversity of accidents. To address low-visibility conditions like rain, fog, and nighttime footage, Wang et al. (2020c) used Retinex image enhancement algorithm for preprocessing and YOLOv3 for object detection, followed by a decision tree to classify crashes. It balances dynamic range and enhances edges, but congested mixed-flow traffic, lower-quality video, and fast vehicles are still major sources of error. The use of deep convolutional autoencoders for representation learning complemented with vehicle tracking is used to detect accidents from surveillance footage in Singh and Mohan (2019). The testing was performed on data collected during bright sunlight, night, early morning, and also from a variety of cameras and angles. However, there are significant false alarms caused by low visibility, occlusions, and large variations in traffic patterns. The lack of near-miss data can be met by combining vehicle event recorder data and object detection from an onboard camera as proposed in Yamamoto et al. (2022). By extracting two deep feature representations that consider the car status and the surrounding objects, the deep learning method can label near-miss events. While the method does not claim to be real-time, it can generate large volumes of labeled training data for near-crash events.

A method to detect cycling near-misses from front view video is developed in Ibrahim et al. (2021) using optical flow, CNN, LSTM, and a fully connected prediction stage. The method was trained with complex urban environments and also contributes to a large dataset containing labeled near-miss events. A ResNet-based model was used to detect pedestrians and evaluate risk from a near-miss dataset in Suzuki et al. (2017). The dataset contains videos from different vehicles, places (intersections, city, major roads), day and night time, and weather conditions. However, the model suffers from overfitting as a result of having only near-miss data for training.

Road User Behavior Prediction

Models and Algorithms

Trajectory prediction from videos is useful for autonomous driving, traffic forecasting, and congestion management. Older works in this domain focused on homogeneous agents such as cars on a highway or pedestrians in a crowd, whereas heterogeneous agents were only considered in sparse scenarios with certain assumptions like lane-based driving. A long short-term memory (LSTM) and CNN hybrid network, that learns the relationship between pairs of heterogeneous agents, was developed in Chandra et al. (2019a) to extract agent shape, velocity, and traffic concentration, which are passed through LSTMs to generate horizon and neighborhood maps, which then go through convolution networks to produce latent representations that are passed through a final

LSTM to predict the trajectory. It can perform accurately in dense, heterogeneous, urban traffic conditions in real time. The paper also contributes a new labeled dataset captured from crowded Asian cities. To be useful, trajectory prediction needs to take into account the motion of surrounding objects and inter-object interactions in real time. Therefore, a different approach to motion prediction is discussed in Li et al. (2019) based on the graph convolutional model, which takes trajectory data as input and represents the interactions of nearby objects and extracts features. The graph model output is then passed into an encoder-decoder LSTM model for robust predictions that can consider the interaction between vehicles. The method enables 30% higher prediction accuracy in addition to 5x faster execution. The algorithm uses trajectory data that has already been extracted from surveillance video data like NGSIM Colyar and Halkias (2007). In Tripicchio and D'Avella (2022), vehicle trajectory of vehicles is calculated using Lucas-Kanade algorithm on dashcam video. Synthetic data was also used for augmenting the dataset to train an LSTM network to predict future motion and an SVM is used to classify the action, for eg. changing lanes. The method predicts the next 6 seconds of motion on highways with 92% accuracy.

Current Methods to Overcome Challenges

The dynamics of vulnerable road users are described by a Switching Linear Dynamical System (SLDS) in Kooij et al. (2019) and extended with a dynamic bayesian network using context from features extracted from vehicle-mounted stereo cameras focusing on both static and dynamic cues. The approach can work in real-time, providing accurate predictions of road user trajectories. It can be improved by the inclusion of more context such as traffic lights and pedestrian crossings. The use of onboard camera and LiDAR along with V2V communication is explored in Choi et al. (2021b) to predict trajectories using the random forest and LSTM architecture. YOLO is used to detect cars and provide bounding boxes, while LiDAR provides subtle changes in position, and V2V communication transmits raw values like steering angles to reduce the uncertainty and latency of predictions.

The TRAF dataset was used in Chandra et al. (2019b) for robust end-to-end real-time trajectory prediction from still or moving cameras. Mask R-CNN and reciprocal velocity obstacles algorithm are used for multi-vehicle tracking. The last 3 seconds of tracking are used to predict the next 5 seconds of trajectory as in Chandra et al. (2019a), with the added advantage of being end-to-end trainable and not requiring annotated trajectory data. The paper also contributes TrackNPred, a python-based library that contains implementations of different trajectory prediction methods. It is a common interface for many trajectory prediction

approaches and can be used for performance comparisons using standard error measurement metrics on real-world dense and heterogeneous traffic datasets.

Most DL methods for trajectory prediction do not uncover the underlying reward function, instead, they only rely on previously seen examples, which hinders generalizability and limits their scope. In Fernando et al. (2021), inverse reinforcement learning is used to find the reward function so that the model can be said to have a tangible goal, allowing it to be deployed in any environment. Transformer-based motion prediction is performed in Liu et al. (2021b) to achieve state-of-the-art multimodal trajectory prediction in the Agroverse dataset. The network models both the road geometry and interactions between the vehicles. Pedestrian intention in complex urban scenarios is predicted by graph convolution networks on spatio-temporal graphs in Liu et al. (2020a). The method considers the relationship between pedestrians waiting to cross and the movement of vehicles. While achieving 80% accuracy on multiple datasets, it predicts intent to cross one second in advance. On the other hand, pedestrians modeled as automatons, combined with SVM without the need for pose information, result in longer predictions but lack the consideration of contextual information (Jayaraman et al. 2020).

Traffic Anomaly Detection

Models and Algorithms

Traffic surveillance cameras can be used to automatically detect traffic anomalies like stopped vehicles and queues. The detection of low-level image features like corners of vehicles has been used by Albiol et al. (2011) to demonstrate queue detection and queue length estimation without object tracking or background removal in different lighting conditions. Tracking methods based on optical flow can not only provide queue length, but also speed, vehicle count, waiting time, and time headway. In Shirazi and Morris (2015), the authors use optical flow assuming constant short-term brightness to detect vehicle features and successfully track them even with occlusions. The speed of individual vehicles can be estimated, allowing the detection of stopped vehicles or queue formation. Trajectory analysis has also been deployed to identify illegal or dangerous movements (Nowosielski et al. 2016). The background subtraction-based approaches are, however, limited to favorable scenarios and do not generalize well.

An interesting method is applied in Li et al. (2016a) involving partitioning the video into spatial and temporal blocks, local invariant features are then learned from traffic footage to create a visual codebook of the image descriptors using Locality-constrained Linear Coding. Then, a Gaussian distribution model is trained to learn the probabilities

corresponding to normal traffic, which can be used to detect anomalies. The image description makes it more robust to lighting, perspective, and occlusions. Aboah et al. (2021) proposed a decision tree-based DL approach for anomaly detection using YOLOv5 for vehicle detection, followed by background estimation, then a decision tree considers factors such as vehicle size, likelihood, and road feature mask to eliminate false positives. Adaptive thresholding allows for robustness under variable illumination and weather conditions. A perspective map approach is discussed by Bai et al. (2019), which models the background using road segmentation based on a traffic flow frequency map, then the perspective is detected from linear regression of object sizes based on ResNet50. Finally, a spatial-temporal matrix discriminating module performs thresholding on consecutive frames to detect anomalous states.

Current Methods to Overcome Challenges

Anomaly detection relies on surveillance cameras which usually provide a view far along the road, but vehicles in the distance occupy only a few pixels which make detection difficult. Thus, Li et al. (2020) uses pixel-level tracking in addition to box-level tracking for multi-granularity. The key idea is mask extraction based on frame difference and vehicle trajectory tracking based on the Gaussian Mixture Model to eliminate moving vehicles combined with segmentation based on frame changes to also eliminate parking zones. Anomaly fusion uses the box and pixel-level tracking features with backtracking optimization to refine predictions. Surveillance cameras are prone to shaking in the wind, so video stabilization preprocessing was performed before using two-stage vehicle detection in the form of Faster R-CNN and Cascade R-CNN (Zhao et al. 2021b). An efficient real-time method for anomaly detection from surveillance video decouples the appearance and motion learning into two parts (Li et al. 2021c). First, an autoencoder learns appearance features, then 3D convolutional layers can use latent codes from multiple past frames to predict features for future frames. A significant difference between predicted and actual features indicates an anomaly. The model can be deployed on edge nodes near the traffic cameras, and the latent features appear to be robust to illumination and weather changes compared to pixel-wise methods.

To shed reliance on annotated data for anomalies, an unsupervised one-class approach in Pawar and Attar (2021) applies spatio-temporal convolutional autoencoder to get latent features, stacks them together, and a sequence-to-sequence LSTM learns the temporal patterns. The method performs well on multiple real-world surveillance footage datasets, but not better than supervised training methods. The advantage is that it can be indefinitely trained on normal traffic data without any labeled anomalies.

Edge Computing

Models and Algorithms

Computer vision in ITS requires efficient infrastructure architecture to analyze data in real time. If all acquired video streams are sent to a single server, the required bandwidth and computation would not be able to provide a usable service. For example, edge computing architecture for real-time automatic failure detection using a video usefulness metric was explored in (Sun et al. 2020a). Only video deemed to be useful is transmitted to the server, while malfunction of the surveillance camera, or obstruction of view, is automatically reported. Edge-cloud-based computing can implement DL models, not just for computer vision tasks, but also for resource allocation and efficiency (Xie et al. 2021). Passive surveillance has now been superseded in literature by the increasing availability of sensor-equipped vehicles that can perform perception and mapping cooperatively (Zhang and Letaief 2020).

Onboard computing resources in vehicles are often not powerful enough to process all sensor data in real time, and applications like localization and mapping can be very computationally intensive. The internet of things (IoT) architecture allows for edge nodes to offload that computation and provide results at low latency to nearby users (Ferdowsi et al. 2019). This approach can avoid multiple cars doing the same computation with similar inputs. One technique to offload computation tasks is discussed in Cui et al. (2020), combining integer linear programming for offline scheduling optimization and heuristics for online, real-world deployment. The authors compress 3D point cloud LIDAR data collected from the vehicle's sensor and send it to the edge node for classification and feature extraction. A deep reinforcement learning algorithm known as Deep Deterministic Policy Gradient is proposed in Dai et al. (2019), which can dynamically allocate computing and caching resources throughout the network. Future work in this direction will handle multiple communication channels, interference management, forecast handover, and bandwidth allocation. In the macro scale, V2V communication can be used for traffic parameter estimation and management with sparse connectivity, while higher connected vehicle market penetration will allow safety applications like collision avoidance (Dey et al. 2016).

Applications for vehicles can include near-crash detection, navigation, video streaming, and smart traffic lights. The onboard unit can also be used as a mobile cache, and communicate with other vehicles via V2V networking. Real-time near-crash detection using edge computing was developed in Ke et al. (2020). The system uses dashcam video for SSD vehicle and pedestrian detection, followed by SORT for tracking to estimate the time to collision (TTC). It was

tested on online datasets and on real cars and buses. The detected events, along with CAN bus messages were used to filter irrelevant data, saving bandwidth for data collection. A practical deployment of parking surveillance using edge-cloud computing was presented in Ke et al. (2021), the edge device performs detection and transmits the bounding box and object types to the server, which uses this information for labeling and tracking. A different approach by Bura et al. (2018) focused on vehicle tracking from top view cameras and number plate recognition from ground-level cameras for real-time occupancy information and to automatically charge a vehicle for the time it was parked. Large-scale traffic monitoring using computer vision and edge computing was detailed in Liu et al. (2021a) where edge nodes close to surveillance cameras can process low-resolution videos to monitor traffic, detect congestion, and detect speed if the available bandwidth is low. If high bandwidth to the server is available, high-quality video will be sent for similar processing. Edge computing for vehicle detection is examined in Wan et al. (2022). The algorithm divides the traffic video into segments of interest and then uses YOLOv3 for vehicle detection in real-time on the edge node, and the extracted clips are used as training data for the edge server.

Current Methods to Overcome Challenges

One problem with large-scale DL is that the huge quantity of data produced cannot be sent to a cloud computer for training. Federated learning (Konečný et al. 2015) has emerged as a solution to this problem, especially considering the heterogeneous data sources, bandwidth, and privacy issues (Zhou et al. 2021). Training can be performed on edge nodes or edge servers, with the results being sent to the cloud to aggregate in the shared deep-learning model (Zhang and Letaief 2020). Federated learning is also robust to failure of individual edge nodes (Kairouz et al. 2019). Concerns of bandwidth, data privacy, and power requirements are addressed in Song et al. (2018) by transferring only inferred data from edge nodes to the cloud, in the form of incremental and unsupervised learning. In general, the processing of data on the edge to reduce bandwidth has the pleasant side effect of anonymizing the transmitted data (Barthélemy et al. 2019). Another effort to reduce bandwidth requirements employs spectral clustering compression performed on spatio-temporal features needed for traffic flow prediction (Chen et al. 2021a).

Deep learning models cannot be directly exported to mobile edge nodes, as they are usually too computationally intensive. Direct adaptation for vehicle counting resulted in 1–4 fps for models in AI city challenge if they ran at all (Anastasiu et al. 2020). Neural network pruning both in terms of storage and computation was introduced in Han et al. (2015), while implementation of the resulting sparse

network on hardware is discussed in Zhang et al. (2016a), achieving multiple orders of magnitude increase in efficiency. A general lightweight CNN model was developed for mobile edge units in Zhou et al. (2019), matching or outperforming AlexNet and VGG-16 while being a fraction of the size and computation cost. Edge computing-based traffic flow detection using deep learning was deployed by Chen et al. (2021b) where YOLOv3 was trained and pruned, along with DeepSORT, to be deployed on the edge device for real-time performance. A thorough review of deploying compact DNNs on low-power edge computers for IoT applications can be found in Zhang et al. (2021). They note that the diversity and quantity of DNN applications require an automated method for model compression beyond traditional pruning techniques.

Future Directions

For Solving Data Challenges

While a large quantity of data is essential for training deep learning models, often the quality is the limiting factor in training performance. Data curation is a necessary process to include edge cases and train the model on representative data from the real world. Labeling vision data, especially in complex urban environments is a labor-intensive task performed by humans. It can be sped up by first using existing object detection or segmentation algorithms based on the relevant task to automatically label the data. Then this can be further checked by humans to eliminate errors by the machine, thus creating a useful labeled dataset. This approach has greatly improved the quality of naturalistic driving datasets (Miao et al. 2022). There is also a need for datasets that include multiple sensors from different views for training cooperative perception algorithms. Collecting such data is bound to be challenging because of hardware requirements and synchronization issues but it is possible to achieve with connected vehicles and instrumented intersections similar to the configuration that will be deployed. Crowd-sourcing via smartphone apps is also a viable method of producing high-quality reliable data (Aboah et al. 2022). Some examples of useful datasets are collected in Table 2.

The problems associated with poor quality or viewing angle of real-world cameras can be mitigated by using realistic CCTV benchmarks and datasets that include a wide variety of surveillance footage, including synthetic video (Revaud and Humenberger 2021). Data-driven simulators like Amini et al. (2021) use high-fidelity datasets to simulate cameras and LiDAR, which can be used to train DL models with data that is hard to capture in the real world (Azfar et al. 2022). Such an approach has shown promise in end-to-end reinforcement learning of autonomous vehicle

Table 2 Summary of datasets

Dataset	Description	Application
Argoverse 2 (Wilson et al. 2021)	High-definition maps, sensor data (lidar, camera, radar, GPS), 3D object annotations, tracking data, semantic segmentation, and vehicle ego-motion information	Detection, segmentation, cooperative perception, sensor fusion, motion estimation
Cityscapes (Cordts et al. 2016)	High-resolution images in the urban setting with a focus on street scenes containing roads, sidewalk, cars, pedestrians in various weather, light, and landscapes	Detection, semantic segmentation, instance-level annotations, fine and coarse levels of segmentation
CamVid (Brostow et al. 2008)	Dashcam view, semantic segmentation of 32 classes, urban street scenes	Fine grained semantic segmentation, scene understanding, road user behavior prediction, anomaly detection
DAWN (Kenk and Hassaballah 2020)	Fog, snow, rain and sandstorms with urban, highway, and freeway traffic; bounding boxes and vehicle labels	Vehicle detection, weather condition classification, robust perception
MIO-TCD (Luo et al. 2018)	786,702 images, 11 road user types, different times of day and year	Classification and localization of moving vehicles, anomaly detection
NGSIM (Colyar and Halkias 2007)	Multiple locations, trajectories (position, speed, acceleration), high spatial and temporal resolution, no images	Trajectory prediction, vehicle interaction, simulation validation
TRAF (Chandra et al. 2019a)	Dense and heterogeneous traffic, traffic rule violations, dashcam and top view, different times of day	Multi-vehicle trajectory prediction, road user behavior, vehicle interaction
SHRP2 Naturalistic Driving Study (Dingus et al. 2015)	Driver face view, travel behavior, traffic flow, safety (crash, near crash, and baseline), accelerometer and vehicle network data	Congestion detection, anomaly detection, near-crash detection, driver behavior
Massachusetts Roads Dataset (Mnih 2013)	Aerial view, 2600 square kilometers of urban, suburban, and rural regions	Instance segmentation, UAV based detection and classification
Waymo Open Dataset (Perception) (Sun et al. 2020b)	5 cameras and top mounted LiDAR, 2030 scenes, 6.4 h	2D and 3D detection, panoptic segmentation, pedestrian keypoints

control (Amini et al. 2020). Domain adaptation techniques are expected to be further extended to utilize synthetic data and conveniently collected data.

Sub-fields in transfer learning, especially few-shot learning and zero-shot learning, will be extensively applied with expert knowledge to address the lack of data challenges, such as corner case recognition in ITS and AD. Likewise, new unsupervised learning and semi-supervised learning models are expected in the general field of real-world computer vision. Future work in vision transformer explainability will allow for more comprehensive insights based on aggregated metrics over multiple samples (Aflalo et al. 2022). Interpretability research is also expected to evaluate differences between model-based and model-free reinforcement learning approaches (Atakishiyev et al. 2021).

Data decentralization is a well-recognized trend in ITS. To address issues like data privacy, large-scale data processing, and efficiency, crowdsensing (Ning et al. 2021) and federated learning on vision tasks (Liu et al. 2020b) are unavoidable future directions in ITS and AD. Additionally, instead of the traditional way of training a single model for a single task, multiple downstream tasks learning with a generalized foundation model, e.g., Florence Yuan et al. (2021), is a promising trend to deal with various data challenges. Another mechanism is data processing parallelism in ITS coupled with edge computing for multi-task (e.g., traffic surveillance and road surveillance) learning (Ke et al. 2022).

For Solving Model Challenges

Deep learning models are trained until they achieve good accuracy, but real-world testing often reveals weaknesses in edge cases and complex environmental conditions. There is a need for online learning for such models to continue to improve and adapt to real-world scenarios otherwise they cannot be of practical use. If online training is not possible due to a lack of live feedback on the correctness of the predictions, the performance must be analyzed periodically with real data stored and labeled by humans. This can serve as a sort of iterative feedback loop, where the model does not need to be significantly changed, just incrementally retrained based on the inputs it finds most challenging. One possible way to partially automate this would be to have multiple different redundant architectures using the same input data to make predictions along with confidence scores. If the outputs do not agree, or if the confidence scores are low for a certain output, that data point can be manually labeled and added to the training set for the next training iteration.

Complex deep learning models deployed to edge devices need to be more efficient through methods such as pruning (Han et al. 2015). Simple pruning methods can improve CNN performance by over 30% (Li et al. 2016b). Depending on the specific architecture, the models may also be split into different functional blocks deployed on separate edge units to minimize bandwidth and computation time (Sufian et al. 2021). A foreseeable future stage of edge AI is "model training and inference both on the edge," without the participation of cloud datacenters.

In recent years much work has been done towards explainable AI, especially in computer vision. CNNs have been approached with three explainability methods: gradient-based saliency maps, Class Activation Mapping, and Excitation Backpropagation (Zhang et al. 2018b). These methods were extended for graph convolutional networks in Pope et al. (2019), pointing out patterns in the input that correspond with the classification. Generic solutions for explainability have been presented in Chefer et al. (2021) for both self-attention and co-attention transformer networks. While it is not straightforward to apply these methods to transportation applications, some efforts have been made to understand deep spatio-temporal neural networks dealing with video object segmentation and action recognition quantifying the static and dynamic information in the network and giving insight into the models and highlighting biases learned from datasets (Kowal et al. 2022).

Cooperative sensing model development is a necessary future direction for better perception in 3D, in order to mitigate the effects of occlusion, noise, and sensor faults. V2X networks and vision transformers have been used for robust cooperative perception, which can support sensing in connected autonomous vehicle platforms (Xu et al. 2021, 2022). Connected autonomous vehicles will also host other deep-learning models that can learn from new data in a distributed manner. Consensus-driven distributed perception is expected to make use of future network technologies like 6G V2X, resulting in low-latency model training that can enable true level 5 autonomous vehicles (Barbieri et al. 2022).

For Solving Complex Traffic Environment Challenges

Multimodal sensing and cooperative perception are necessary future avenues of practical research. Different modalities like video, LiDAR, and audio can be used in combinations to improve the performance of methods purely based on vision. Audio is especially useful for detecting anomalies earlier among pedestrians such as fights or commotions, and for vehicles in crowded intersections where the visual chaos may not immediately reveal problems like mechanical faults, or minor accidents. Cooperative perception will allow multiple sensor views of the same environment from different vehicles to build a common picture that contains more information than any single agent can perceive thus solving problems of occlusion and illumination.

There is an increasing trend of using transfer learning to improve model performance in real-world tasks. Initially training the model on synthetic data and fine-tuning with

task-specific data reduces the reliability on complex, single-use deep learning models and improve real-world performance by retraining on challenging urban scenarios. As aforementioned, domain adaptation, zero-shot learning, few-shot learning, and foundation models are expected transfer learning areas that serve this purpose.

The results of unsupervised methods like in Pawar and Attar (2021) can be further improved by online learning in crowded and challenging scenarios after deployment on embedded hardware, as there is an unlimited supply of unlabeled data. The lack of theoretical performance analysis regarding the upper bound on false alarm rate in complex environments is discussed as an important aspect of deep learning methods for anomaly detection in Doshi and Yilmaz (2021). Future research is recommended to include this analysis as well. It is hard to imagine complete reliance on surveillance cameras for robust, widespread, and economical traffic anomaly detection. The method in Parsa et al. (2020) includes traffic, network, demographic, land use, and weather data sources to detect traffic. Such ideas can be used in tandem with computer vision applications for better overall performance.

Future directions in the application of edge computing in ITS will consider multi-source data fusion along with online learning (Xie et al. 2021). Many factors like unseen shapes of vehicles, new surrounding environments, variable traffic density, and rare events can be too challenging for DL models (Ferdowsi et al. 2019). This new data could be used for online training of the system. Traditional applications can be extended using edge computing and IoV/IoT frameworks. Vehicle re-identification from video is emerging as the most robust solution to occlusion (Zhao et al. 2021a). However, the inclusion of more spatio-temporal information for learning leads to greater memory and computational usage. Tracklets from one camera view can be matched with other views at different points in time using known features. Instead of using a fixed window, adaptive feature aggregation based on similarity and quality, can be generalized to many multi-object tracking tasks (Qian et al. 2020).

Transformers are good at learning dynamic interactions between heterogenous agents which will be particularly useful in crowded urban environments for detection and trajectory prediction. They can also be used for the detection of anomalies and the prediction of potentially hazardous situations like collisions in a multi-user heterogeneous scenario.

Conclusions

In real-world scenarios, most of the DL computer vision methods suffer from severe performance degradation when facing different challenges. In this paper, we review the specific challenges for data, models, and complex environments

in ITS and autonomous driving. Many related deep learning-based computer vision methods are reviewed, summarized, compared, and discussed. Furthermore, a number of representative deep learning-based applications of ITS and autonomous driving are summarized and analyzed. Based on our analysis and review, several potential future research directions are provided. We expect that this paper could provide useful research insights and inspire more progress in the community.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval Not applicable.

References

Aboah A, Shoman M, Mandal V, Davami S, Adu-Gyamfi Y, Sharma A (2021) A vision-based system for traffic anomaly detection using deep learning and decision trees. In: CVPR

Aboah A, Boeding M, Adu-Gyamfi Y (2022) Mobile sensing for multi-purpose applications in transportation. J Big Data Analyt Transp 4(2–3):171–183

Aflalo E, Du M, Tseng S-Y, Liu Y, Wu C, Duan N, Lal V (2022) Vl-interpret: an interactive visualization tool for interpreting vision-language transformers. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 21406–21415

Albiol A, Albiol A, Mossi JM (2011) Video-based traffic queue length estimation, pp 1928–1932 . https://doi.org/10.1109/ICCVW.2011.6130484

Amini A, Gilitschenski I, Phillips J, Moseyko J, Banerjee R, Karaman S, Rus D (2020) Learning robust control policies for end-to-end autonomous driving from data-driven simulation. IEEE Robot Autom Lett 5(2):1143–1150. https://doi.org/10.1109/LRA.2020.2966414

Amini A, Wang T-H, Gilitschenski I, Schwarting W, Liu Z, Han S, Karaman S, Rus D (2021) VISTA 2.0: an open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles. arXiv. https://doi.org/10.48550/ARXIV.2111.12083 . https://arxiv.org/abs/2111.12083

Anastasiu DC, Gaul J, Vazhaeparambil M, Gaba M, Sharma P (2020) Efficient city-wide multi-class multi-movement vehicle counting: a survey. J Big Data Analyt Transp 2:235–250

Arabi S, Haghighat A, Sharma A (2020) A deep-learning-based computer vision solution for construction vehicle detection. Comput-Aided Civ Infrastruct Eng 35(7):753–767

Atakishiyev S, Salameh M, Yao H, Goebel R (2021) Towards safe, explainable, and regulated autonomous driving. arXiv. https://

doi.org/10.48550/ARXIV.2111.10518. https://arxiv.org/abs/2111.10518

Azfar T, Weidner J, Raheem A, Ke R, Cheu RL (2022) Efficient procedure of building university campus models for digital twin simulation. IEEE J Radio Freq Identif 6:769–773

Bai S, He Z, Lei Y, Wu W, Zhu C, Sun M, Yan J (2019) Traffic anomaly detection via perspective map based on spatial-temporal information matrix. In: CVPR Workshops, pp 117–124

Bai X, Wang X, Liu X, Liu Q, Song J, Sebe N, Kim B (2021) Explainable deep learning for efficient and robust pattern recognition: a survey of recent developments. Pattern Recogn 120:108102. https://doi.org/10.1016/j.patcog.2021.108102

Barbieri L, Savazzi S, Brambilla M, Nicoli M (2022) Decentralized federated learning for extended sensing in 6g connected vehicles. Veh Commun 33:100396

Barthélemy J, Verstaevel N, Forehead H, Perez P (2019) Edge-computing video analytics for real-time traffic monitoring in a smart city. Sensors. https://doi.org/10.3390/s19092048

Bell S, Zitnick CL, Bala K, Girshick R (2016) Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 2874–2883

Bell A, Mantecón T, Díaz C, del-Blanco CR, Jaureguizar F, García N (2021) A novel system for nighttime vehicle detection based on foveal classifiers with real-time performance. IEEE Trans Intell Transp Syst 23(6):5421–5433

Bianco S, Cadene R, Celona L, Napoletano P (2018) Benchmark analysis of representative deep neural network architectures. IEEE Access 6:64270–64277

Bolya D, Zhou C, Xiao F, Lee YJ (2019) Yolact: real-time instance segmentation. In: Proceedings of the IEEE/CVF International Conference on computer vision, pp 9157–9166

Bornstein AM (2016) Is artificial intelligence permanently inscrutable? Nautilus https://nautil.us/is-artificial-intelligence-permanently-inscrutable-236088/. Accessed 6 Jan 2024

Brahmbhatt S, Christensen HI, Hays J (2017) Stuffnet: using 'stuff' to improve object detection. In: 2017 IEEE Winter Conference on applications of computer vision (WACV), pp 934–943. IEEE

Brazil G, Liu X (2019) Pedestrian detection with autoregressive network phases. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 7231–7240

Brazil G, Yin X, Liu X (2017) Illuminating pedestrians via simultaneous detection & segmentation. In: Proceedings of the IEEE International Conference on computer vision, pp 4950–4959

Brkić I, Miler M, Ševrović M, Medak D (2020) An analytical framework for accurate traffic flow parameter calculation from uav aerial videos. Remote Sens 12(22):3844

Brostow GJ, Fauqueur J, Cipolla R (2008) Semantic object classes in video: a high-definition ground truth database. Pattern Recognit Lett 30(2):88–97

Buch N, Velastin SA, Orwell J (2011) A review of computer vision techniques for the analysis of urban traffic. IEEE Trans Intell Transp Syst 12(3):920–939. https://doi.org/10.1109/TITS.2011.2119372

Bura H, Lin N, Kumar N, Malekar S, Nagaraj S, Liu K (2018) An edge based smart parking solution using camera networks and deep learning. In: 2018 IEEE International Conference on cognitive computing (ICCC), pp 17–24. https://doi.org/10.1109/ICCC.2018.00010

Cai P, Lee Y, Luo Y, Hsu D (2020) Summit: a simulator for urban driving in massive mixed traffic. In: 2020 IEEE International Conference on robotics and automation (ICRA), pp 4023–4029. https://doi.org/10.1109/ICRA40945.2020.9197228

Caillot A, Ouerghi S, Vasseur P, Boutteau R, Dupuis Y (2022) Survey on cooperative perception in an automotive context. IEEE Trans Intell Transp Syst 23:14204–14223

Cao J, Zhang J, Jin X (2021) A traffic-sign detection algorithm based on improved sparse r-cnn. IEEE Access 9:122774–122788

Carranza-García M, Lara-Benítez P, García-Gutiérrez J, Riquelme JC (2021) Enhancing object detection for autonomous driving by optimizing anchor generation and addressing class imbalance. Neurocomputing 449:229–244. https://doi.org/10.1016/j.neucom.2021.04.001

Chakraborty P, Adu-Gyamfi YO, Poddar S, Ahsani V, Sharma A, Sarkar S (2018) Traffic congestion detection from camera images using deep convolution neural networks. Transp Res Rec 2672(45):222–231

Chan F-H, Chen Y-T, Xiang Y, Sun M (2016) Anticipating accidents in dashcam videos. In: Asian Conference on computer vision, pp 136–153. Springer

Chandra R, Bhattacharya U, Bera A, Manocha D (2019a) Traphic: trajectory prediction in dense and heterogeneous traffic using weighted interactions, vol. 2019-June, pp 8475–8484. IEEE Computer Society. https://doi.org/10.1109/CVPR.2019.00868

Chandra R, Bhattacharya U, Roncal C, Bera A, Manocha D (2019b) Robusttp: End-to-end trajectory prediction for heterogeneous road-agents in dense traffic with noisy sensor inputs. In: ACM Computer Science in Cars Symposium. CSCS '19. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3359999.3360495

Chefer H, Gur S, Wolf L (2021) Generic attention-model explainability for interpreting bi-modal and encoder-decoder transformers. In: Proceedings of the IEEE/CVF International Conference on computer vision, pp 397–406

Chen D, Zhang S, Ouyang W, Yang J, Tai Y (2018) Person search via a mask-guided two-stream cnn model. In: Proceedings of the European Conference on computer vision (ECCV), pp 734–750

Chen Y, Zhang Y, Maharjan S, Alam M, Wu T (2019a) Deep learning for secure mobile edge computing in cyber-physical transportation systems. IEEE Netw 33(4):36–41. https://doi.org/10.1109/MNET.2019.1800458

Chen Q, Ma X, Tang S, Guo J, Yang Q, Fu S (2019b) F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds. In: Proceedings of the 4th ACM/IEEE Symposium on edge computing, pp 88–100

Chen K, Pang J, Wang J, Xiong Y, Li X, Sun S, Feng W, Liu Z, Shi J, Ouyang W, et al (2019c) Hybrid task cascade for instance segmentation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 4974–4983

Chen Q, Tang S, Yang Q, Fu S (2019d) Cooper: cooperative perception for connected autonomous vehicles based on 3d point clouds. In: 2019 IEEE 39th International Conference on distributed computing systems (ICDCS), pp 514–524. IEEE

Chen M, Xue H, Cai D (2019e) Domain adaptation for semantic segmentation with maximum squares loss. In: Proceedings of the IEEE/CVF International Conference on computer vision, pp 2090–2099

Chen C, Liu Z, Wan S, Luan J, Pei Q (2021a) Traffic flow prediction based on deep learning in internet of vehicles. IEEE Trans Intell Transp Syst 22(6):3776–3789. https://doi.org/10.1109/TITS.2020.3025856

Chen C, Liu B, Wan S, Qiao P, Pei Q (2021b) An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. IEEE Trans Intell Transp Syst 22(3):1840–1852. https://doi.org/10.1109/TITS.2020.3025687

Choi JG, Kong CW, Kim G, Lim S (2021a) Car crash detection using ensemble deep learning and multimodal data from dashboard cameras. Expert Syst Appl 183:115400. https://doi.org/10.1016/j.eswa.2021.115400

Choi D, Yim J, Baek M, Lee S (2021b) Machine learning-based vehicle trajectory prediction using v2v communications and

on-board sensors. Electronics 10(4):1. https://doi.org/10.3390/electronics10040420

Colyar J, Halkias J (2007) Us highway 101 dataset. Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-07-030, 27–69

Contreras-Castillo J, Zeadally S, Guerrero-Ibañez JA (2018) Internet of vehicles: architecture, protocols, and security. IEEE Internet Things J 5(5):3701–3709. https://doi.org/10.1109/JIOT.2017.2690902

Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B (2016) The cityscapes dataset for semantic urban scene understanding. In: Proc. of the IEEE Conference on computer vision and pattern recognition (CVPR)

Cortinhal T, Kurnaz F, Aksoy EE (2021) Semantics-aware multi-modal domain translation: from lidar point clouds to panoramic color images. In: Proceedings of the IEEE/CVF International Conference on computer vision, pp 3032–3048

Cui H, Radosavljevic V, Chou F-C, Lin T-H, Nguyen T, Huang T-K, Schneider J, Djuric N (2019) Multimodal trajectory predictions for autonomous driving using deep convolutional networks. In: 2019 International Conference on robotics and automation (ICRA), pp 2090–2096. https://doi.org/10.1109/ICRA.2019.8793868

Cui M, Zhong S, Li B, Chen X, Huang K (2020) Offloading autonomous driving services via edge computing. IEEE Internet Things J 7:10535–10547. https://doi.org/10.1109/JIOT.2020.3001218

Dai Y, Xu D, Maharjan S, Qiao G, Zhang Y (2019) Artificial intelligence empowered edge computing and caching for internet of vehicles. IEEE Wirel Commun 26:12–18. https://doi.org/10.1109/MWC.2019.1800411

D'Amour A, Heller K, Moldovan D, Adlam B, Alipanahi B, Beutel A, Chen C, Deaton J, Eisenstein J, Hoffman MD, et al (2020) Underspecification presents challenges for credibility in modern machine learning. arXiv preprint arXiv:2011.03395

Dey KC, Rayamajhi A, Chowdhury M, Bhavsar P, Martin J (2016) Vehicle-to-vehicle (v2v) and vehicle-to-infrastructure (v2i) communication in a heterogeneous wireless network - performance evaluation. Transp Res Part C Emerg Technol 68:168–184. https://doi.org/10.1016/j.trc.2016.03.008

Dhatbale R, Chilukuri BR (2021) Deep learning techniques for vehicle trajectory extraction in mixed traffic. J Big Data Analyt Transp 3:141–157

Dingus TA, Hankey JM, Antin JF, Lee SE, Eichelberger L, Stulce KE, McGraw D, Perez M, Stowe, L (2015) Naturalistic driving study: technical coordination and quality control vol. SHRP 2 Report S2-S06-RW-1

Dong G, Yan Y, Shen C, Wang H (2020) Real-time high-performance semantic image segmentation of urban street scenes. IEEE Trans Intell Transp Syst 22(6):3258–3274

Doshi K, Yilmaz Y (2021) Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. Pattern Recogn 114:107865. https://doi.org/10.1016/j.patcog.2021.107865

Dou Q, Castro D, Kamnitsas K, Glocker B (2019) Domain generalization via model-agnostic learning of semantic features. Adv Neural Inform Process Syst, vol. 32, https://proceedings.neurips.cc/paper_files/paper/2019/file/2974788b53f73e7950e8aa49f3a306db-Paper.pdf. Accessed 6 Jan 2024

Fedorov A, Nikolskaia K, Ivanov S, Shepelev V, Minbaleev A (2019) Traffic flow estimation with data from a video surveillance camera. J Big Data 6(1):1–15

Feng D, Haase-Schütz C, Rosenbaum L, Hertlein H, Glaeser C, Timm F, Wiesbeck W, Dietmayer K (2020) Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges. IEEE Trans Intell Transp Syst 22(3):1341–1360

Ferdowsi A, Challita U, Saad W (2019) Deep learning for reliable mobile edge analytics in intelligent transportation systems: an

overview. IEEE Veh Technol Mag 14(1):62–70. https://doi.org/10.1109/MVT.2018.2883777

Fernando T, Denman S, Sridharan S, Fookes C (2021) Deep inverse reinforcement learning for behavior prediction in autonomous driving: accurate forecasts of vehicle motion. IEEE Signal Process Mag 38(1):87–96. https://doi.org/10.1109/MSP.2020.2988287

Formosa N, Quddus M, Man CK, Timmis A (2023) Appraising machine and deep learning techniques for traffic conflict prediction with class imbalance. Data Sci Transp 5(2):4

Fries RN, Gahrooei MR, Chowdhury M, Conway AJ (2012) Meeting privacy challenges while advancing intelligent transportation systems. Transp Res Part C Emerg Technol 25:34–45

Fu L, Yu H, Juefei-Xu F, Li J, Guo Q, Wang S (2021) Let there be light: improved traffic surveillance via detail preserving night-to-day transfer. IEEE Trans Circ Syst Video Technol 32:8217–8226

Gao Y, Li J, Xu Z, Liu Z, Zhao X, Chen J (2021a) A novel image-based convolutional neural network approach for traffic congestion estimation. Expert Syst Appl 180:115037

Gao G, Xu G, Yu Y, Xie J, Yang J, Yue D (2021b) Mscfnet: a lightweight network with multi-scale context fusion for real-time semantic segmentation. IEEE Trans Intell Transp Syst 23(12):25489–25499

Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? the Kitti vision benchmark suite. In: 2012 IEEE Conference on computer vision and pattern recognition, pp 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

Ghafoorian M, Nugteren C, Baka N, Booij O, Hofmann M (2018) El-gan: embedding loss driven generative adversarial networks for lane detection. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp 0

Gilroy S, Jones E, Glavin M (2019) Overcoming occlusion in the automotive environment—a review. IEEE Trans Intell Transp Syst 22(1):23–35

Gu W, Bai S, Kong L (2022) A review on 2d instance segmentation based on deep neural networks. Image Vis Comput 120:104401

Guo D, Pei Y, Zheng K, Yu H, Lu Y, Wang S (2019) Degraded image semantic segmentation with dense-gram networks. IEEE Trans Image Process 29:782–795

Haghighat AK, Ravichandra-Mouli V, Chakraborty P, Esfandiari Y, Arabi S, Sharma A (2020) Applications of deep learning in intelligent transportation systems. J Big Data Analyt Transp 2:115–145

Halevy A, Norvig P, Pereira F (2009) The unreasonable effectiveness of data. IEEE Intell Syst 24(2):8–12

Han S, Pool J, Tran J, Dally W (2015) Learning both weights and connections for efficient neural network. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett, R (eds) Advances in neural information processing systems, vol. 28. Curran Associates, Inc., https://proceedings.neurips.cc/paper/2015/file/ae0eb3eed39d2bcef4622b2499a05fe6-Paper.pdf

Hassaballah M, Kenk MA, Muhammad K, Minaee S (2020) Vehicle detection and tracking in adverse weather using a deep learning framework. IEEE Trans Intell Transp Syst 22(7):4230–4242

He K, Zhang X, Ren S, Sun J (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–1916

He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE International Conference on computer vision, pp 2961–2969

Hong F, Lu C-H, Liu C, Liu R-R, Wei J (2020) A traffic surveillance multi-scale vehicle detection object method base on encoder-decoder. IEEE Access 8:47664–47674

Hu H-N, Cai Q-Z, Wang D, Lin J, Sun M, Krahenbuhl P, Darrell T, Yu F (2019) Joint monocular 3d vehicle detection and tracking.

In: Proceedings of the IEEE/CVF International Conference on computer vision (ICCV)

Huang T, Sharma A (2020) Technical and economic feasibility assessment of a cloud-enabled traffic video analysis framework. J Big Data Analyt Transp 2:223–233

Ibrahim MR, Haworth J, Christie N, Cheng T (2021) Cyclingnet: detecting cycling near misses from video streams in complex urban scenes with deep learning. IET Intel Transp Syst 15(10):1331–1344

Impedovo D, Balducci F, Dentamaro V, Pirlo G (2019) Vehicular traffic congestion classification by visual features and deep learning approaches: a comparison. Sensors 19(23):5213

Jayaraman SK, Tilbury DM, Jessie Yang X, Pradhan AK, Robert LP (2020) Analysis and prediction of pedestrian crosswalk behavior during automated vehicle interactions. In: 2020 IEEE International Conference on robotics and automation (ICRA), pp 6426–6432 . https://doi.org/10.1109/ICRA40945.2020.9197347

Jiansheng F (2014) et al: Vision-based real-time traffic accident detection. In: Proceeding of the 11th World Congress on intelligent control and automation, pp 1035–1038. IEEE

Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, Bonawitz K, Charles Z, Cormode G, Cummings R, D'Oliveira RGL, Eichner H, Rouayheb SE, Evans D, Gardner J, Garrett Z, Gascón A, Ghazi B, Gibbons PB, Gruteser M, Harchaoui Z, He C, He L, Huo Z, Hutchinson B, Hsu J, Jaggi M, Javidi T, Joshi G, Khodak M, Konečný J, Korolova A, Koushanfar F, Koyejo S, Lepoint T, Liu Y, Mittal P, Mohri M, Nock R, Özgür A, Pagh R, Raykova M, Qi H, Ramage D, Raskar R, Song D, Song W, Stich SU, Sun Z, Suresh AT, Tramér F, Vepakomma P, Wang J, Xiong L, Xu Z, Yang Q, Yu FX, Yu H, Zhao S (2019) Advances and open problems in federated learning. arXiv. https://doi.org/10.48550/ARXIV.1912.04977. https://arxiv.org/abs/1912.04977

Kamal U, Tonmoy TI, Das S, Hasan MK (2019) Automatic traffic sign detection and recognition using segu-net and a modified Tversky loss function with l1-constraint. IEEE Trans Intell Transp Syst 21(4):1467–1479

Kataoka H, Suzuki T, Oikawa S, Matsui Y, Satoh Y (2018) Drive video analysis for the detection of traffic near-miss incidents. In: 2018 IEEE International Conference on robotics and automation (ICRA), pp 3421–3428. https://doi.org/10.1109/ICRA.2018.8460812

Ke R (2020) Real-time video analytics empowered by machine learning and edge computing for smart transportation applications. University of Washington, Seattle

Ke X, Shi L, Guo W, Chen D (2018a) Multi-dimensional traffic congestion detection based on fusion of visual features and convolutional neural network. IEEE Trans Intell Transp Syst 20(6):2157–2170

Ke R, Li Z, Tang J, Pan Z, Wang Y (2018b) Real-time traffic flow parameter estimation from uav video based on ensemble classifier and optical flow. IEEE Trans Intell Transp Syst 20(1):54–64

Ke R, Cui Z, Chen Y, Zhu M, Yang H, Wang Y (2020) Edge computing for real-time near-crash detection for smart transportation applications. arXiv . https://doi.org/10.48550/ARXIV.2008.00549. https://arxiv.org/abs/2008.00549

Ke R, Zhuang Y, Pu Z, Wang Y (2021) A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on iot devices. IEEE Trans Intell Transp Syst 22(8):4962–4974. https://doi.org/10.1109/TITS.2020.2984197

Ke R, Liu C, Yang H, Sun W, Wang Y (2022) Real-time traffic and road surveillance with parallel edge intelligence. IEEE J Radio Frequ Identif 6:693–696

Kenk MA, Hassaballah, M (2020) Dawn: vehicle detection in adverse weather nature dataset. arXiv preprint arXiv:2008.05402

Khan MA, Ullah I, Alkhalifah A, Rehman SU, Shah JA, Uddin MI, Alsharif MH, Algarni F (2021) A provable and privacy-preserving authentication scheme for uav-enabled intelligent transportation systems. IEEE Trans Ind Inf 18(5):3416–3425

Kim J, Canny J (2018) In: Escalante HJ, Escalera S, Guyon I, Baró X, Güçlütürk Y, Güçlü U, Gerven M (eds) Explainable deep driving by visualizing causal attention. Springer, Cham, pp 173–193. https://doi.org/10.1007/978-3-319-98131-4_8

Kim H-K, Park JH, Jung H-Y (2018) An efficient color space for deep-learning based traffic light recognition. J Adv Transp. https://doi.org/10.1155/2018/2365414

Kim K-J, Kim P-K, Chung Y-S, Choi D-H (2019) Multi-scale detector for accurate vehicle detection in traffic surveillance data. IEEE Access 7:78311–78319

Kolesnikov A, Beyer L, Zhai X, Puigcerver J, Yung J, Gelly S, Houlsby N (2019) Big Transfer (BiT): general visual representation learning. arXiv. https://doi.org/10.48550/ARXIV.1912.11370. https://arxiv.org/abs/1912.11370

Konečný J, McMahan B, Ramage D (2015) Federated optimization: distributed optimization beyond the datacenter. arXiv. https://doi.org/10.48550/ARXIV.1511.03575. https://arxiv.org/abs/1511.03575

Kong T, Yao A, Chen Y, Sun F (2016) Hypernet: Towards accurate region proposal generation and joint object detection. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 845–853

Kooij JF, Flohr F, Pool EA, Gavrila DM (2019) Context-based path prediction for targets with switching dynamics. Int J Comput Vis 127(3):239–262

Kowal M, Siam M, Islam MA, Bruce ND, Wildes RP, Derpanis KG (2022) A deeper dive into what deep spatiotemporal networks encode: Quantifying static vs. dynamic information. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 13999–14009

Kumar N, Raubal M (2021) Applications of deep learning in congestion detection, prediction and alleviation: a survey. Transp Res Part C Emerg Technol 133:103432

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Lee E, Kim D (2019) Accurate traffic light detection using deep neural network with focal regression loss. Image Vis Comput 87:24–36

Lee Y, Park J (2020) Centermask: Real-time anchor-free instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer vision and pattern recognition, pp 13906–13915

Li Y, Liu W, Huang Q (2016a) Traffic anomaly detection based on image descriptor in videos. Multimed Tools Appl 75:2487–2505. https://doi.org/10.1007/s11042-015-2637-y

Li H, Kadav A, Durdanovic I, Samet H, Graf HP (2016b) Pruning filters for efficient convnets. arXiv preprint arXiv:1608.08710

Li X, Ying X, Chuah MC (2019) Grip: graph-based interaction-aware trajectory prediction. In: IEEE Intelligent Transportation Systems Conference (ITSC)

Li Y, Wu J, Bai X, Yang X, Tan X, Li G, Wen S, Zhang H, Ding E (2020) Multi-granularity tracking with modularlized components for unsupervised vehicles anomaly detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition workshops, pp 586–587

Li J, Xu Z, Fu L, Zhou X, Yu H (2021a) Domain adaptation from daytime to nighttime: a situation-sensitive vehicle detection and traffic flow parameter estimation framework. Transp Res Part C Emerg Technol 124:102946. https://doi.org/10.1016/j.trc.2020.102946

Li Y, Karim MM, Qin R, Sun Z, Wang Z, Yin Z (2021b) Crash report data analysis for creating scenario-wise, spatio-temporal attention guidance to support computer vision-based perception of fatal crash risks. Acc Anal Prev 151:105962. https://doi.org/10.1016/j.aap.2020.105962

Li B, Leroux S, Simoens P (2021c) Decoupled appearance and motion learning for efficient anomaly detection in surveillance video. Comput Vis Image Underst 210:103249. https://doi.org/10.1016/j.cviu.2021.103249

Li Y, Ren S, Wu P, Chen S, Feng C, Zhang W (2021d) Learning distilled collaboration graph for multi-agent perception. Adv Neural Inform Process Syst 34:29541–29552

Li G, Ji Z, Qu X (2022) Stepwise domain adaptation (sda) for object detection in autonomous vehicles using an adaptive centernet. IEEE Trans Intell Transp Syst

Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 2117–2125

Lin C, Lu J, Wang G, Zhou J (2018) Graininess-aware deep feature learning for pedestrian detection. In: Proceedings of the European Conference on computer vision (ECCV), pp 732–747

Lin C-Y, Muchtar K, Lin W-Y, Jian Z-Y (2019) Moving object detection through image bit-planes representation without thresholding. IEEE Trans Intell Transp Syst 21(4):1404–1414

Lin C-T, Huang S-W, Wu Y-Y, Lai S-H (2020) Gan-based day-to-night image style transfer for nighttime vehicle detection. IEEE Trans Intell Transp Syst 22(2):951–963

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) Ssd: single shot multibox detector. In: European Conference on computer vision, pp 21–37. Springer

Liu W, Liao S, Hu W, Liang X, Chen X (2018a) Learning efficient single-stage pedestrian detectors by asymptotic localization fitting. In: Proceedings of the European Conference on computer vision (ECCV), pp 618–634

Liu S, Qi L, Qin H, Shi J, Jia J (2018b) Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 8759–8768

Liu S, Liu L, Tang J, Yu B, Wang Y, Shi W (2019) Edge computing for autonomous driving: opportunities and challenges. Proc IEEE 107(8):1697–1716. https://doi.org/10.1109/JPROC.2019.2915983

Liu B, Adeli E, Cao Z, Lee K-H, Shenoi A, Gaidon A, Niebles JC (2020a) Spatiotemporal relationship reasoning for pedestrian intent prediction. IEEE Robot Autom Lett 5(2):3485–3492. https://doi.org/10.1109/LRA.2020.2976305

Liu Y, Huang A, Luo Y, Huang H, Liu Y, Chen Y, Feng L, Chen T, Yu H, Yang Q (2020b) Fedvision: an online visual object detection platform powered by federated learning. In: Proceedings of the AAAI Conference on artificial intelligence 34:13172–13179

Liu G, Shi H, Kiani A, Khreishah A, Lee J, Ansari N, Liu C, Yousef MM (2021a) Smart traffic monitoring system using computer vision and edge computing. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2021.3109481

Liu Y, Zhang J, Fang L, Jiang Q, Zhou B (2021b) Multimodal motion prediction with stacked transformers. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 7577–7586

Liu Y, Zhang W, Wang J (2021c) Source-free domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 1215–1224

Luo Z, Branchaud-Charron F, Lemaire C, Konrad J, Li S, Mishra A, Achkar A, Eichel J, Jodoin P-M (2018) Mio-tcd: a new benchmark dataset for vehicle classification and localization. IEEE Trans Image Process 27(10):5129–5141

Mahajan D, Girshick R, Ramanathan V, He K, Paluri M, Li Y, Bharambe A, Van Der Maaten L (2018) Exploring the limits of weakly supervised pretraining. In: Proceedings of the European Conference on computer vision (ECCV), pp 181–196

Ma Y, Manocha D, Wang W (2018) Autorvo: Local navigation with dynamic constraints in dense heterogeneous traffic. arXiv preprint arXiv:1804.02915

Mandal V, Adu-Gyamfi Y (2020) Object detection and tracking algorithms for vehicle counting: a comparative analysis. J Big Data Analyt Transp 2(3):251–261

Martínez-Ballesté A, Rashwan HA, Puig D, Fullana AP (2012) Towards a trustworthy privacy in pervasive video surveillance systems. In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops, pp 914–919. IEEE

Martínez-Ballesté A, Pérez-Martínez PA, Solanas A (2013) The pursuit of citizens' privacy: a privacy-aware smart city is possible. IEEE Commun Mag 51(6):136–141

Mhalla A, Chateau T, Gazzah S, Amara NEB (2018) An embedded computer-vision system for multi-object detection in traffic surveillance. IEEE Trans Intell Transp Syst 20(11):4006–4018

Miao H, Zhang S, Flannagan C (2022) Driver behavior extraction from videos in naturalistic driving datasets with 3d convnets. J Big Data Analyt Transp 4(1):41–55

Mnih V (2013) Machine learning for aerial image labeling. PhD thesis, University of Toronto

Mo Y, Han G, Zhang H, Xu X, Qu W (2019) Highlight-assisted night-time vehicle detection using a multi-level fusion network and label hierarchy. Neurocomputing 355:13–23

Mozaffari S, Al-Jarrah OY, Dianati M, Jennings P, Mouzakitis A (2022) Deep learning-based vehicle behavior prediction for autonomous driving applications: a review. IEEE Trans Intell Transp Syst 23:33–47. https://doi.org/10.1109/TITS.2020.3012034

Muhammad K, Ullah A, Lloret J, Del Ser J, Albuquerque VHC (2020) Deep learning for safe autonomous driving: current challenges and future directions. IEEE Trans Intell Transp Syst 22(7):4316–4336

Naphade M, Tang Z, Chang M-C, Anastasiu DC, Sharma A, Chellappa R, Wang S, Chakraborty P, Huang T, Hwang J-N et al (2019) The 2019 ai city challenge. In: CVPR Workshops, vol. 8, p 2

Ning Z, Sun S, Wang X, Guo L, Guo S, Hu X, Hu B, Kwok R (2021) Blockchain-enabled intelligent transportation systems: a distributed crowdsensing framework. IEEE Trans Mob Comput 21(12):4201–4217

Nirkin Y, Wolf L, Hassner T (2021) Hyperseg: Patch-wise hypernetwork for real-time semantic segmentation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 4061–4070

Nowosielski A, Frejlichowski D, Forczmański P, Gościewska K, Hofman R (2016) Automatic analysis of vehicle trajectory applied to visual surveillance. In: Choraś RS (ed) Image processing and communications challenges 7. Springer, Cham, pp 89–96

O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, Riordan D, Walsh J (2019) Deep learning vs. traditional computer vision. In: Science and Information Conference, pp 128–144. Springer

Ou Z, Xiao F, Xiong B, Shi S, Song M (2019) Famn: feature aggregation multipath network for small traffic sign detection. IEEE Access 7:178798–178810

Ouyang Z, Niu J, Liu Y, Guizani M (2019) Deep cnn-based real-time traffic light detector for self-driving vehicles. IEEE Trans Mob Comput 19(2):300–313

Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian AK (2020) Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. Acc Anal Prev 136:105405. https://doi.org/10.1016/j.aap.2019.105405

Pawar K, Attar V (2021) Deep learning based detection and localization of road accidents from traffic surveillance videos. ICT Express. https://doi.org/10.1016/j.icte.2021.11.004

Peppa M, Bell D, Komar T, Xiao W (2018) Urban traffic flow analysis based on deep learning car detection from cctv image series. In:

SPRS TC IV Mid-term Symposium "3D Spatial Information Science–The Engine of Change". Newcastle University

Peppa MV, Komar T, Xiao W, James P, Robson C, Xing J, Barr S (2021) Towards an end-to-end framework of cctv-based urban traffic volume detection and prediction. Sensors 21(2):629

Petsiuk V, Das A, Saenko K (2018) Rise: randomized input sampling for explanation of black-box models. arXiv preprint arXiv:1806.07421

Pope PE, Kolouri S, Rostami M, Martin CE, Hoffmann H (2019) Explainability methods for graph convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 10772–10781

Qian Y, Yu L, Liu W, Kang G, Hauptmann AG (2020) Adaptive feature aggregation for video object detection. In: Proceedings of the IEEE/CVF Winter Conference on applications of computer vision workshops, pp 143–147

Ras G, Gerven M, Haselager P (2018) Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges. In: Escalante, H., et al. Explainable and Interpretable Models in Computer Vision and Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-319-98131-4_2

Rashmi C, Shantala C (2020) Vehicle density analysis and classification using yolov3 for smart cities. In: 2020 4th International Conference on electronics, communication and aerospace technology (ICECA), pp 980–986. IEEE

Rauch A, Klanner F, Rasshofer R, Dietmayer K (2012) ar2x-based perception in a high-level fusion architecture for cooperative perception systems. In: 2012 IEEE Intelligent Vehicles Symposium, pp 270–275. IEEE

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 779–788

Reed WJ (2001) The pareto, zipf and other power laws. Econ Lett 74(1):15–19

Revaud J, Humenberger M (2021) Robust automatic monocular vehicle speed estimation for traffic surveillance. In: Proceedings of the IEEE/CVF International Conference on computer vision (ICCV), pp 4551–4561

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1(5):206–215

Saleh K, Hossny M, Nahavandi S (2017) Intent prediction of vulnerable road users from motion trajectories using stacked lstm network. In: 2017 IEEE 20th International Conference on intelligent transportation systems (ITSC), pp. 327–332. https://doi.org/10.1109/ITSC.2017.8317941

Samek W, Wiegand T, Müller K-R (2017) Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv . https://doi.org/10.48550/ARXIV.1708.08296 . https://arxiv.org/abs/1708.08296

Santhosh KK, Dogra DP, Roy PP (2020) Anomaly detection in road traffic using visual surveillance: a survey. ACM Comput Surv 53(6):10. https://doi.org/10.1145/3417989

Sharma T, Debaque B, Duclos N, Chehri A, Kinder B, Fortier P (2022) Deep learning-based object detection and scene perception under bad weather conditions. Electronics 11(4):563

Shen Z, Liu Z, Li J, Jiang Y-G, Chen Y, Xue X (2017) Dsod: learning deeply supervised object detectors from scratch. In: Proceedings of the IEEE International Conference on computer vision, pp 1919–1927

Shirazi MS, Morris B (2015) Vision-based vehicle queue analysis at junctions. In: 2015 12th IEEE International Conference on advanced video and signal based surveillance (AVSS), pp 1–6. https://doi.org/10.1109/AVSS.2015.7301732

Singh D, Mohan CK (2019) Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. IEEE Trans Intell Transp Syst 20(3):879–887. https://doi.org/10.1109/TITS.2018.2835308

Song M, Zhong K, Zhang J, Hu Y, Liu D, Zhang W, Wang J, Li T (2018) In-situ ai: towards autonomous and incremental deep learning for iot systems. In: 2018 IEEE International Symposium on high performance computer architecture (HPCA), pp 92–103. https://doi.org/10.1109/HPCA.2018.00018

Song S, Miao Z, Yu H, Fang J, Zheng K, Ma C, Wang S (2020) Deep domain adaptation based multi-spectral salient object detection. IEEE Trans Multimed 4:128–140

Sonnleitner E, Barth O, Palmanshofer A, Kurz M (2020) Traffic measurement and congestion detection based on real-time highway video data. Appl Sci 10(18):6270

Sufian A, Alam E, Ghosh A, Sultana F, De D, Dong M (2021) Deep learning in computer vision through mobile edge computing for iot. In: Mukherjee A, De D, Ghosh SK, Buyya R (eds) Mobile Edge Computing. Springer, Cham. pp. 443–471. https://doi.org/10.1007/978-3-030-69893-5_18

Sun C, Ai Y, Wang S, Zhang W (2020) Dense-refinedet for traffic sign detection and classification. Sensors 20(22):6570

Sun H, Shi W, Liang X, Yu Y (2020a) Vu: edge computing-enabled video usefulness detection and its application in large-scale video surveillance systems. IEEE Internet Things J 7(2):800–817. https://doi.org/10.1109/JIOT.2019.2936504

Sun P, Kretzschmar H, Dotiwalla X, Chouard A, Patnaik V, Tsui P, Guo J, Zhou Y, Chai Y, Caine B, Vasudevan V, Han W, Ngiam J, Zhao H, Timofeev A, Ettinger S, Krivokon M, Gao A, Joshi A, Zhang, Y, Shlens J, Chen Z, Anguelov D (2020b) Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)

Sun G, Wang W, Dai J, Van Gool L (2020c) Mining cross-image semantics for weakly supervised semantic segmentation. In: European Conference on computer vision, pp 347–365. Springer

Suzuki T, Aoki Y, Kataoka H (2017) Pedestrian near-miss analysis on vehicle-mounted driving recorders. In: 2017 Fifteenth IAPR International Conference on machine vision applications (MVA), pp 416–419. https://doi.org/10.23919/MVA.2017.7986889

Tabelini L, Berriel R, Paixao TM, Badue C, De Souza AF, Oliveira-Santos T (2021) Keep your eyes on the lane: real-time attention-guided lane detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 294–302

Tabernik D, Skočaj D (2019) Deep learning for large-scale traffic-sign detection and recognition. IEEE Trans Intell Transp Syst 21(4):1427–1440

Taccari L, Sambo F, Bravi L, Salti S, Sarti L, Simoncini M, Lori A (2018) Classification of crash and near-crash events from dashcam videos and telematics. In: 2018 21st International Conference on intelligent transportation systems (ITSC), pp 2460–2465. IEEE

Tang T, Zhou S, Deng Z, Lei L, Zou H (2017) Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks. Remote Sens 9(11):1170

Temel D, Chen M-H, AlRegib G (2019) Traffic sign detection under challenging conditions: a deeper look into performance variations and spectral characteristics. IEEE Trans Intell Transp Syst 21(9):3663–3673

Thandavarayan G, Sepulcre M, Gozalvez J (2020) Generation of cooperative perception messages for connected and automated vehicles. IEEE Trans Veh Technol 69(12):16336–16341

Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P (2017) Domain randomization for transferring deep neural networks from simulation to the real world. In: 2017 IEEE/RSJ

International Conference on intelligent robots and systems (IROS), pp 23–30 IEEE

Tripicchio P, D'Avella S (2022) Modeling multiple vehicle interaction constraints for behavior prediction of vehicles on highways. Comput Electr Eng 98:107700. https://doi.org/10.1016/j.compeleceng.2022.107700

Vasiljevic I, Chakrabarti A, Shakhnarovich G (2017) Examining the impact of blur on recognition by convolutional networks. arXiv preprint arXiv:1611.05760

Wan J, Ding W, Zhu H, Xia M, Huang Z, Tian L, Zhu Y, Wang H (2021a) An efficient small traffic sign detection method based on yolov3. J Signal Process Syst 93(8):899–911

Wan H, Gao L, Su M, You Q, Qu H, Sun Q (2021b) A novel neural network model for traffic sign detection and recognition under extreme conditions. J Sens. https://doi.org/10.1155/2021/9984787

Wan S, Ding S, Chen C (2022) Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles. Pattern Recogn 121:108146. https://doi.org/10.1016/j.patcog.2021.108146

Wang J, Cho J, Lee S, Ma T (2011) Real time services for future cloud computing enabled vehicle networks. In: 2011 International Conference on wireless communications and signal processing (WCSP), pp 1–5. https://doi.org/10.1109/WCSP.2011.6096957

Wang X, Xiao T, Jiang Y, Shao S, Sun J, Shen C (2018) Repulsion loss: detecting pedestrians in a crowd. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 7774–7783

Wang Z, Wu Y, Niu Q (2020a) Multi-sensor fusion in automated driving: a survey. IEEE Access 8:2847–2868. https://doi.org/10.1109/ACCESS.2019.2962554

Wang T-H, Manivasagam S, Liang M, Yang B, Zeng W, Urtasun R (2020b) V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In: European Conference on computer vision, pp 605–621. Springer

Wang C, Dai Y, Zhou W, Geng Y (2020c) A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition. J Adv Transp. https://doi.org/10.1155/2020/9194028

Wang R, Kashinath K, Mustafa M, Albert A, Yu R (2020d) Towards physics-informed deep learning for turbulent flow prediction. In: Proceedings of the 26th ACM SIGKDD International Conference on knowledge discovery and data mining, pp 1457–1466

Wang Y, Zhang J, Kan M, Shan S, Chen X (2020e) Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 12275–12284

Wang K, Tang X, Zhao S, Zhou Y (2021a) Simultaneous detection and tracking using deep learning and integrated channel feature for ambient traffic light recognition. J Ambient Intell Humaniz Comput 13:271–281

Wang J, Zhang W, Zang Y, Cao Y, Pang J, Gong T, Chen K, Liu Z, Loy CC, Lin D (2021b) Seesaw loss for long-tailed instance segmentation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 9695–9704

Wang T, Zhu Y, Zhao C, Zeng W, Wang J, Tang M (2021c) Adaptive class suppression loss for long-tail object detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 3103–3112

Wang T, Zhu Y, Chen Y, Zhao C, Yu B, Wang J, Tang M (2022) C2am loss: chasing a better decision boundary for long-tail object detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 6980–6989

Wei P, Shi H, Yang J, Qian J, Ji Y, Jiang X (2019) City-scale vehicle tracking and traffic flow estimation using low frame-rate traffic cameras. Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, pp 602–610

Wilson B, Qi W, Agarwal T, Lambert J, Singh J, Khandelwal S, Pan B, Kumar R, Hartnett A, Pontes JK, Ramanan D, Carr P, Hays J (2021) Argoverse 2: Next generation datasets for self-driving perception and forecasting. In: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021)

Wu Z, Sang J, Zhang Q, Xiang H, Cai B, Xia X (2019) Multi-scale vehicle detection for foreground-background class imbalance with improved yolov2. Sensors. https://doi.org/10.3390/s19153336

Xie J, Zheng Y, Du R, Xiong W, Cao Y, Ma Z, Cao D, Guo J (2021) Deep learning-based computer vision for surveillance in its: evaluation of state-of-the-art methods. IEEE Trans Veh Technol 70(4):3027–3042. https://doi.org/10.1109/TVT.2021.3065250

Xu W, Zhou H, Cheng N, Lyu F, Shi W, Chen J, Shen X (2018) Internet of vehicles in big data era. IEEE/CAA J Autom Sin 5(1):19–35. https://doi.org/10.1109/JAS.2017.7510736

Xu W, Wang H, Qi F, Lu C (2019) Explicit shape encoding for real-time instance segmentation. In: Proceedings of the IEEE/CVF International Conference on computer vision, pp 5168–5177

Xu Y, Yang X, Gong L, Lin H-C, Wu T-Y, Li Y, Vasconcelos N (2020) Explainable object-induced action decision for autonomous vehicles. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)

Xu R, Xiang H, Xia X, Han X, Liu J, Ma J (2021) Opv2v: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. arXiv preprint arXiv:2109.07644

Xu R, Xiang H, Tu Z, Xia X, Yang M-H, Ma J (2022) V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. arXiv preprint arXiv:2203.10638

Yamamoto S, Kurashima T, Toda H (2022) Classifying near-miss traffic incidents through video, sensor, and object features. IEICE Trans Inform Syst E105.D(2):377–386. https://doi.org/10.1587/transinf.2021EDP7017

Yang Z, Pun-Cheng LS (2018) Vehicle detection in intelligent transportation systems and its applications under varying environments: a review. Image Vis Comput 69:143–154

Yang Y, Luo H, Xu H, Wu F (2015) Towards real-time traffic sign detection and classification. IEEE Trans Intell Transp Syst 17(7):2022–2031

Yang Q, Fu S, Wang H, Fang H (2021) Machine-learning-enabled cooperative perception for connected autonomous vehicles: challenges and opportunities. IEEE Netw 35(3):96–101. https://doi.org/10.1109/MNET.011.2000560

Yao Y, Zheng L, Yang X, Naphade M, Gedeon T (2020) Simulating content consistent vehicle datasets with attribute descent. In: European Conference on computer vision, pp 775–791. Springer

Yoon DD, Ayalew B, Ali GMN (2021) Performance of decentralized cooperative perception in v2v connected traffic. IEEE Trans Intell Transp Syst 23(7):6850–6863

Yuan L, Chen D, Chen Y-L, Codella N, Dai X, Gao J, Hu H, Huang X, Li B, Li C, et al (2021) Florence: a new foundation model for computer vision. arXiv preprint arXiv:2111.11432

Yu C, Wang J, Peng C, Gao C, Yu G, Sang N (2018) Bisenet: Bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European Conference on computer vision (ECCV), pp 325–341

Zhang J, Letaief KB (2020) Mobile edge intelligence and computing for the internet of vehicles. Proc IEEE 108(2):246–261. https://doi.org/10.1109/JPROC.2019.2947490

Zhang S, Du Z, Zhang L, Lan H, Liu S, Li L, Guo Q, Chen T, Chen Y (2016a) Cambricon-x: an accelerator for sparse neural networks. In: 2016 49th Annual IEEE/ACM International Symposium on

Microarchitecture (MICRO), pp 1–12. https://doi.org/10.1109/MICRO.2016.7783723

Zhang L, Lin L, Liang X, He K (2016b) Is faster r-cnn doing well for pedestrian detection? In: European Conference on computer vision, pp 443–457. Springer

Zhang Z, Trivedi C, Liu X (2018a) Automated detection of grade-crossing-trespassing near misses based on computer vision analysis of surveillance video data. Saf Sci 110:276–285. https://doi.org/10.1016/j.ssci.2017.11.023

Zhang J, Bargal SA, Lin Z, Brandt J, Shen X, Sclaroff S (2018b) Top-down neural attention by excitation backprop. Int J Comput Vis 126(10):1084–1102

Zhang C, Zhu J, Wang W, Zhao D (2019) A general framework of learning multi-vehicle interaction patterns from video. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp 4323–4328. https://doi.org/10.1109/ITSC.2019.8917212

Zhang K, Ying H, Dai H-N, Li L, Peng Y, Guo K, Yu H (2021) Compacting deep neural networks for internet of things: methods and applications. IEEE Internet Things J 8(15):11935–11959. https://doi.org/10.1109/JIOT.2021.3063497

Zhao Z-Q, Zheng P, Xu S-T, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30(11):3212–3232

Zhao Z-Q, Zheng P, Xu S-T, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30(11):3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zhao J, Qi F, Ren G, Xu L (2021a) Phd learning: learning with Pompeiu-Hausdorff distances for video-based vehicle re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 2225–2235

Zhao Y, Wu W, He Y, Li Y, Tan X, Chen S (2021b) Good practices and a strong baseline for traffic anomaly detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 3993–4001

Zheng Z, Jiang M, Wang Z, Wang J, Bai Z, Zhang X, Yu X, Tan X, Yang Y, Wen S, et al: (2020) Going beyond real data: a robust visual representation for vehicle re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp 598–599

Zhou D, Frémont V, Quost B, Dai Y, Li H (2017) Moving object detection and segmentation in urban environments from a moving platform. Image Vis Comput 68:76–87

Zhou J, Dai H-N, Wang H (2019) Lightweight convolution neural networks for mobile edge computing in transportation cyber physical systems. ACM Trans Intell Syst Technol 10(6):10. https://doi.org/10.1145/3339308

Zhou X, Ke R, Yang H, Liu C (2021) When intelligent transportation systems sensing meets edge computing: vision and challenges. Appl Sci (Switzerland). https://doi.org/10.3390/app11209680

Zhou P, Kortoçi P, Yau Y-P, Finley B, Wang X, Braud T, Lee L-H, Tarkoma S, Kangasharju J, Hui P (2022) Aicp: augmented informative cooperative perception. IEEE Trans IntellTransp Syst 23(11):22505–22518

Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on computer vision, pp 2223–2232

Zhu X, Su W, Lu L, Li B, Wang X, Dai J (2021) Deformable detr: deformable transformers for end-to-end object detection. In: International Conference on learning representations. https://openreview.net/forum?id=gZ9hCDWe6ke. Accessed 6 Jan 2024

Zou Q, Jiang H, Dai Q, Yue Y, Chen L, Wang Q (2019) Robust lane detection from continuous driving scenes using deep neural networks. IEEE Trans Veh Technol 69(1):41–54

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;

2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;

3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;

4. use bots or other automated methods to access the content or redirect messages

5. override any security feature or exclusionary protocol; or

6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com